# Optimal predictions of powers of conditionally heteroskedastic processes

Francq, Christian and Zakoian, Jean-Michel

17. April 2010

# Optimal predictions of powers of conditionally heteroskedastic processes

Christian Francq

University Lille 3, EQUIPPE-GREMARS

BP 60149, 59653 Villeneuve d'Ascq cedex, France*

Jean-Michel Zakoïan

Lille 3 and CREST, 15 Boulevard Gabriel Peri

92245 Malakoff Cedex, France†

**Abstract**

In conditionally heteroskedastic models, the optimal prediction of powers, or logarithms, of the absolute process has a simple expression in terms of the volatility process and an expectation involving the independent process. A standard procedure for estimating this prediction is to estimate the volatility by gaussian quasi-maximum likelihood (QML) in a first step, and to use empirical means based on rescaled innovations to estimate the expectation in a second step. This paper proposes an alternative one-step procedure, based on an appropriate non-gaussian QML estimation of the model, and establishes the asymptotic properties of the two approaches. Their performances are compared for finite-order GARCH models and for the ARCH($\infty$). For the standard GARCH($p, q$) and the Asymmetric Power GARCH($p, q$), it is shown that the ARE of the estimators only depends on the prediction problem and some moments of the independent process. An application to indexes of major stock exchanges is proposed.

**Keywords:** APARCH; ARCH($\infty$); Conditional Heteroskedasticity; Efficiency of estimators; GARCH; Prediction; Quasi Maximum Likelihood Estimation.

*Email: christian.francq@univ-lille3.fr

†Email: zakoian@ensae.fr

# 1 Introduction

The autoregressive conditional heteroscedasticity (ARCH) model introduced by Engle (1982) and its generalization, the GARCH model proposed by Bollerslev (1986) have received considerable attention in both applied and theoretical works. Following Engle's specification of the volatility as a linear combination of past squared innovations, an immense number of alternative formulations, generally motivated by specific features of the financial series, have been proposed. Most of them can be embedded in the model

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \ldots; \theta_0) \end{cases} \tag{1.1}$$

where $(\eta_t)$ is a sequence of independent and identically distributed (iid) random variables, with $\eta_t$ independent of $\{\epsilon_u, u < t\}$, $\theta_0 \in \mathbb{R}^m$ is a parameter belonging to a parameter space $\Theta$, and $\sigma : \mathbb{R}^\infty \times \Theta \to (0, \infty)$. The variable $\sigma_t^2$ is generally referred to as the volatility of $\epsilon_t$.

A leading model, the most widely used among practitioners, is the GARCH(1,1) model where $\sigma_t^2 = \omega_0 + \alpha_0 \epsilon_{t-1}^2 + \beta_0 \sigma_{t-1}^2$ and $\theta_0 = (\omega_0, \alpha_0, \beta_0)' \in (0, \infty) \times [0, \infty) \times [0, 1)$. For this model we have $\sigma_t^2 = \sum_{i=1}^\infty \beta^{i-1}(\omega_0 + \alpha_0 \epsilon_{t-i}^2)$ which is of the form (1.1). For evident identifiability reasons, a scale constraint is required on the sequence $(\eta_t)$. The standard assumption is $E\eta_t^2 = 1$ but any other constraint of the form $E|\eta_t|^r = 1$, with $r \neq 0$, can be used as well (provided that the $r$-th moment exists). Another class of conditionally heteroskedatic models is the ARCH($\infty$) introduced by Robinson (1991) and studied by many authors (see references below). This model is a particular case of (1.1) in which the function $\sigma^2$ is linear in the squares of the past values of $\epsilon_t$.

In GARCH models, it is generally assumed that $E\eta_t = 0$, but we do not make this assumption. Model (1.1) also includes the Autoregressive Conditional Duration (ACD) model for positively distributed $\eta_t$'s. ACD models have been introduced by Engle and Russell (1998) for the analysis of the duration time between events.

Although the literature on GARCH-type models is quite extensive, relatively few papers have examined the issue of forecasting. Engle and Kraft (1983) considered predictions of ARMA processes with ARCH errors. Engle and Bollerslev (1986) and Baillie and Bollerslev (1992) studied predictions in ARMA model with GARCH errors. Andersen and Bollerslev (1998) discussed the predictive qualities of GARCH, making a clear distinction between the prediction of volatility and that of the squared returns. Karanasos (1999) considered GARCH in mean models. Pascual, Romo and Ruiz (2005) proposed a Bootstrap procedure to obtain prediction densities of returns and volatilities of GARCH processes.

In this paper, our aim is to investigate the problem of predicting powers of the process $(\epsilon_t)$, defined as a solution of (1.1). For any real number $r$ such that $E|\eta_t|^r < \infty$, the optimal predictor, in the $L^2$ sense, of $|\epsilon_t|^r$ given its entire past is

$$E_{t-1}(|\epsilon_t|^r) = \sigma_t^r E|\eta_t|^r, \tag{1.2}$$

where $E_{t-1}$ denotes expectation conditional on the infinite past. We will also consider the optimal prediction of $\log|\epsilon_t|$ given by

$$E_{t-1}(\log|\epsilon_t|) = \log\sigma_t + E\log|\eta_t|, \tag{1.3}$$

provided that $E\log|\eta_t|$ exists. This case can be seen as the limit of the case (1.2) when $r=0$, via the Box-Cox transformation $\log|x| = \lim_{r\to 0}(|x|^r-1)/r$.

## 1.1 The two approaches

Given observations $(\epsilon_1,\dots,\epsilon_n)$, we consider two approaches for predicting powers of $|\epsilon_{n+1}|$:

- A *fully parametric* one-step approach in which $\theta_0$ is estimated under the assumption that $E|\eta_t|^r = 1$ when $r \neq 0$, and $E\log|\eta_t| = 0$ when $r=0$. The prediction of $|\epsilon_{n+1}|^r$ (resp. $\log|\epsilon_{n+1}|$) based on (1.2) (resp. (1.3)) is then the estimated value of $\sigma_{n+1}^r$ (resp. $\log\sigma_{n+1}$).
- A *mixed* (parametric and non parametric) two-step approach in which $\theta_0$ is estimated under the assumption that $E|\eta_t|^2 = 1$ and $E|\eta_t|^r$ (or $E\log|\eta_t|$ when $r=0$) is estimated non-parametrically. The prediction of $|\epsilon_{n+1}|^r$ (resp. $\log|\epsilon_{n+1}|$) based on (1.2) (resp. (1.3)) is the estimated value of $\sigma_{n+1}^r$ (resp. $\log\sigma_{n+1}$) multiplied by the estimate of $E|\eta_t|^r$ (resp. plus the estimate of $E\log|\eta_t|$).

The mixed approach is standard. The fully parametric approach is novel, to our knowledge.

## 1.2 Non Gaussian QML

For the reparameterized model under the identifiability constraint $E|\eta_0|^r = 1$ with $r \neq 2$, the Gaussian QML estimator (QMLE) is generally inconsistent. For our prediction problem with $r \neq 2$, we therefore consider a generalized QMLE based on an *instrumental* density $h$ different from the Gaussian. This QMLE coincides with the MLE when the error's distribution $f$ is correctly specified (that is when $h = f$). To keep the robustness of the standard QML, it should also be consistent for any error distribution $f$ satisfying $E|\eta_0|^r = 1$. This will imply a choice of $h$ depending on the prediction problem, that is on $r$. Newey and Steigerwald (1997) studied the identification conditions required for the consistency of non Gaussian QMLE's in general conditional heteroskedastic models. In the case of standard GARCH models, Berkes and Horváth (2004) derived the asymptotic distribution of such estimators.

## 1.3 Interest of predicting powers $r \neq 2$

The prediction of $\epsilon_t^2$, which is also the prediction of the volatility under the assumption that $E\eta_t^2 = 1$, is obviously important for financial applications but it does not appear to be sufficient.

*i) Interest of considering* $r > 2$. The conditional variance of the prediction errors of the squares involves the prediction of $(\epsilon_t^4)$. More precisely, in Model (1.1) under the standard assumption $E\eta_t^2 = 1$, the quadratic loss for the prediction of $\epsilon_t^2$, that is the conditional MSE, is

$$E_{t-1}(\epsilon_t^2 - \sigma_t^2)^2 = E_{t-1}(\epsilon_t^4) - \sigma_t^4.$$

Thus, the evaluation of the MSE requires prediction of $|\epsilon_t|^r$ for $r = 4$. Obviously, another loss function involving another power $r$ could be used.

*ii) Interest of considering* $0 \leq r < 2$. When one suspects that second or fourth-order moments do not exist, it is sensible to consider predictions of smaller powers of returns as measures of the future price volatility.

*iii) Interest of considering* $r < 0$. For some applications, it may be worth fitting a GARCH-type model to the inverses of the data. For instance, ACD-type models can be seen as squares of GARCH models applied to duration data, $x_t$ say (where $t$ denotes the $t$-th transaction, and $x_t$ the duration between the $(t-1)$-th and $t$-th transactions). Such models are appropriate to capture the clustering of large durations. However, it may be of interest to capture clustering between small durations. Indeed, such small durations are likely to reflect high volatility of prices. It is thus sensible to adjust a GARCH model to the inverse of such duration data, $\epsilon_t = 1/x_t$. The usual GARCH methodology allows to optimally predict $\epsilon_t^2$, but one is mostly interested in predicting $x_t$ or $x_t^2$. To this aim, we need to predict $|\epsilon_t|^r$ with $r = -1$ or $r = -2$.

## 1.4 Contributions of this paper

For the general Model (1.1), we consider the prediction of powers of any sign. The standard method is a two-step procedure which requires estimating the volatility and also moments of the iid process.

Our main contributions are the following: 1) as an alternative to the two-step method, we introduce a one-step method, based on a generalized QML, which is extremely simple to implement; 2) we obtain a complete characterization of the *omnibus* instrumental densities, that is those which render the generalized QMLE universally consistent; 3) the asymptotic properties of the generalized QMLE are studied in a quite general framework (including in particular the infinite ARCH) under conditions which coincide with the weakest conditions in the particular GARCH case; 4) for important subclasses, we obtain a surprisingly simple expression for the Asymptotic Relative Efficiency (ARE) of the two methods; 5) in practice it is simple to estimate this ARE, and therefore it is possible to determine what is the best method, asymptotically.

## 1.5 Organization of the paper

Section 2 is devoted to the strong consistency and asymptotic normality (AN) for generalized QMLE, based on an instrumental density $h$, in Model (1.1). The choice of $h$ is solved for the prediction problems (1.2)-(1.3), by characterizing the functions $h$ for which the consistency is achieved under a given condition $E|\eta_t|^r = 1$ or under the condition $E \log |\eta_t| = 0$. Section 3 is devoted to the asymptotic properties of the two-step approach. Section 4 proposes comparisons of the two approaches, for finite-order GARCH models and for the ARCH($\infty$). For the standard GARCH($p, q$) and the Asymmetric Power GARCH($p, q$), we show that the ARE of the estimators only depends on the power $r$ and the moments of the iid process. Section 5 proposes empirical applications based on financial data. The most technical proofs are given in Appendix A and Appendix B.

# 2 Asymptotic distribution of non-Gaussian QMLE

The asymptotic results of this paper will be established under the following assumption, which can be made more explicit for specific forms of the volatility function $\sigma$ (for classical GARCH see Nelson (1990), Bougerol and Picard (1992)).

**A0:** $(\epsilon_t)$ is a strictly stationary and ergodic solution of (1.1).

Given observations $\epsilon_1, \dots, \epsilon_n$, and arbitrary initial values $\tilde{\epsilon}_i$ for $i \leq 0$, we define, under assumptions given below

$$\tilde{\sigma}_t(\theta) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_1, \tilde{\epsilon}_0, \tilde{\epsilon}_{-1}, \dots; \theta).$$

This random variable will be used as a proxy of

$$\sigma_t(\theta) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_1, \epsilon_0, \epsilon_{-1}, \dots; \theta).$$

We choose an arbitrary integrable and positive function $h$, in general a density, which can be called *instrumental* density, and define the QML criterion

$$\tilde{Q}_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} g(\epsilon_t, \tilde{\sigma}_t(\theta)), \qquad g(x, \sigma) = \log \frac{1}{\sigma} h\left(\frac{x}{\sigma}\right). \tag{2.1}$$

Let the QMLE

$$\hat{\theta}_{n,h} = \arg\max_{\theta \in \Theta} \tilde{Q}_n(\theta)$$

for some compact space $\Theta$. This estimator is the standard gaussian QMLE when $h$ is the standard gaussian density $\phi$.

## 2.1 Identifiability conditions

To be able to identify the parameters in Model (1.1) it is necessary to impose a constraint on $(\eta_t)$. For the sake of predicting $|\epsilon_t|^r$, a natural constraint in view of (1.2)-(1.3), is

**A1:** $E|\eta_0|^r = 1$ when $r \neq 0$, and $E \log |\eta_0| = 0$ when $r = 0$.

We make the following assumption on the volatility function, for some $\underline{\omega} > 0$.

**A2:** Almost surely, $\sigma_t(\theta) \in (\underline{\omega}, \infty]$ for any $\theta \in \Theta$. Moreover, $\sigma_t(\theta_0)/\sigma_t(\theta) = 1$ $a.s.$ iff $\theta = \theta_0$.

For the consistency of the estimator $\hat{\theta}_{n,h}$, we assume that the function $\sigma \to Eg(\eta_0, \sigma)$ is valued in $[-\infty, +\infty)$ and has a unique maximum at 1:

**A3:** $Eg(\eta_0, \sigma) < Eg(\eta_0, 1) \qquad \forall \sigma > 0, \quad \sigma \neq 1.$

Let $f$ denote the density of the distribution of $\eta_0$, when existing. To interpret **A3**, denote by $K(f, f^*) = E \log(f/f^*)(\eta_0)$ the Kullback-Leibler "distance" between $f$ and a density $f^*$. Let $h_\sigma(x) = \sigma^{-1} h(\sigma^{-1} x)$, the density of $\sigma Y$ where $Y$ has the density $h$. Then **A3** can be written

$$K(f, h) < K(f, h_\sigma), \qquad \forall \sigma > 0, \quad \sigma \neq 1.$$

The condition thus signifies that it is impossible to obtain a density closer to $f$ by scaling $h$.

It is clear by the Jensen inequality that **A3** is always satisfied for the MLE, that is if $h = f$. However, in general $f$ is unknown and cannot be chosen as the instrumental density. When $h \neq f$, **A3** entails a moment condition on $\eta_0$, which may be incompatible with **A1**. For instance when $h = \phi$, we find that **A3** reduces to $E\eta_0^2 = 1$. This condition is compatible with **A1** only when $r = 2$. It is therefore important to characterize the functions $h$ which make **A1** and **A3** compatible. This will be done in Section 2.3.

## 2.2 Asymptotic properties of the generalized QMLE

Apart from identifiability assumptions, technical conditions are required for the asymptotic properties of the generalized QMLE.

For some constants $\delta \in \mathbb{R}$ and $C_0 > 0$, let

**A4:** $h$ is differentiable, for all $u \in \mathbb{R}$, $|uh'(u)/h(u)| \leq C_0(1 + |u|^\delta)$ and $E|\eta_0|^\delta < \infty$.

For the reader's convenience, additional technical assumptions, **A5-A10**, are reported in Appendix A. The following is an extension of results (Theorems 1.1 and 1.2) proven by Berkes and Horváth (2004) for the standard GARCH.

**Theorem 2.1.** *If* **A0-A5** *hold, for some constants $\delta \in \mathbb{R}$ and $C_0 > 0$, then*

$$\hat{\theta}_{n,h} \to \theta_0, \quad a.s.$$

*where* $\theta_0$ *is the true parameter value in Model (1.1) under the identifiability condition* **A1**. *If, in addition,* **A6-A10** *hold and* $Eg_2(\eta_0, 1) \neq 0$ *then*

$$\sqrt{n}\left(\hat{\theta}_{n,h} - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\tau_{h,f}^2 J^{-1})$$

*where*

$$J = J(\theta_0) = E\left(\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta'}(\theta_0)\right) \quad \text{and} \quad \tau_{h,f}^2 = \frac{Eg_1^2(\eta_0, 1)}{\{Eg_2(\eta_0, 1)\}^2}, \qquad (2.2)$$

*with* $g_1(x,\sigma) = \partial g(x,\sigma)/\partial\sigma$ *and* $g_2(x,\sigma) = \partial g_1(x,\sigma)/\partial\sigma$.

This result contains, as particular cases, the AN for the MLE (when $h = f$) and for the QMLE (when $r = 2$ and $h = \phi$). In the former case we have

$$\tau_{f,f}^2 = \left\{E\left(1 + \frac{f'(\eta_0)}{f(\eta_0)}\eta_0\right)\right\}^{-1}.$$

We also have $\tau_{\phi,f}^2 = (E\eta_0^4 - 1)/4$ when $r = 2$ and we retrieve the standard result.

**Remark 2.1.** The results of Theorem 2.1 can be compared with those obtained in other articles for the gaussian QMLE of general formulations similar to (1.1). Straumann and Mikosch (2006) studied the gaussian QMLE for conditionally heteroscedastic models where the volatility has the form $\sigma_t^2 = g(\epsilon_{t-1}, \dots, \epsilon_{t-p}, \sigma_{t-1}^2, \dots, \sigma_{t-q}^2; \theta)$. More recently Bardet and Wintenberger (2009) proved the asymptotic properties of the gaussian QMLE for a general class of multidimensional causal processes encompassing (1.1). However, their conditions for consistency and asymptotic normality require the existence of moments of orders 2 and 4, respectively, which we do not need for the class (1.1).

## 2.3 Choice of the instrumental density

A given function $h$ can be said to be *omnibus* for our prediction problem if Assumptions **A1** and **A3** are compatible *for any distribution of* $\eta_0$. In this section, we will show that under **A4**, the class of the omnibus functions $h$ reduces, for a given $r$, to the class $\mathcal{C}(r)$ of functions of the form

$$\begin{cases} c|x|^{\lambda-1}\exp\left(-\lambda|x|^r/r\right), & \text{if } r > 0,\\ c|x|^{-\lambda-1}\exp\left(\lambda|x|^r/r\right), & \text{if } r < 0,\\ \sqrt{\lambda/\pi}|2x|^{-1}\exp\left\{-\lambda(\log|x|)^2\right\}, & \text{if } r = 0,\end{cases}$$

for constants $\lambda, c > 0$. The following proposition, whose proof is straightforward, shows that, for a given $r$, the QMLE based on $h \in \mathcal{C}(r)$ does not depend on the constants $c$ and $\lambda$.

**Proposition 2.1.** *If the instrumental function $h$ belongs to $\mathcal{C}(r)$ then the generalized QMLE is given by*

$$\hat{\theta}_{n,h} = \begin{cases} \arg\min_{\theta\in\Theta} \sum_{t=1}^{n} \log \tilde{\sigma}_t^r(\theta) + \frac{|\epsilon_t|^r}{\tilde{\sigma}_t^r(\theta)}, & \text{if } r \neq 0, \\ \arg\min_{\theta\in\Theta} \sum_{t=1}^{n} \left\{\log \frac{|\epsilon_t|}{\tilde{\sigma}_t(\theta)}\right\}^2, & \text{if } r = 0. \end{cases}$$

The previous result shows that when $r \neq 0$, the non gaussian QMLE can be interpreted as a standard QMLE obtained by transforming the data $\epsilon_t^2$ in $|\epsilon_t|^r$. The following proposition shows that **A3** can be omitted in Theorem 2.1 when $h$ is chosen in $\mathcal{C}(r)$.

**Proposition 2.2.** *Let $h$ such that* **A4** *holds. Then*

**A3** *holds for any distribution of $\eta_0$ satisfying* **A1** *iff* $\quad h \in \mathcal{C}(r)$.

**Proof.** If $h \in \mathcal{C}(r)$ the implication can be obtained by direct verification, for $r > 0$, $r < 0$ and $r = 0$. For the converse, it will be sufficient to consider the case $r \neq 0$, the case $r = 0$ being treated along the same lines. Note that under **A4**,

$$g_1(x,\sigma) = \frac{\partial g(x,\sigma)}{\partial \sigma} = -\frac{1}{\sigma} - \frac{h'(x/\sigma)}{h(x/\sigma)}\frac{x}{\sigma^2}$$

exists for $\sigma > 0$, and that $E\sup_{\sigma\in V(1)} |g_1(\eta_0,\sigma)| < \infty$, for some neighborhood $V(1)$ of 1. The dominated convergence theorem shows that **A3** entails the moment condition

$$E\left(\frac{h'(\eta_0)}{h(\eta_0)}\eta_0\right) = -1. \tag{2.3}$$

The problem is to find $h$ satisfying (2.3) for any distribution satisfying **A1**. The set of all possible densities $h$ is thus,

$$\mathcal{H} = \left\{h \text{ density } \mid \quad \text{for any variable } \eta, \quad E|\eta|^r = 1 \Rightarrow E\left(\frac{h'(\eta)}{h(\eta)}\eta\right) = -1\right\}.$$

We note that this set contains the set

$$\mathcal{H}' = \left\{h \text{ density } \mid \exists\lambda, \quad \frac{h'(x)}{h(x)}x + 1 = \lambda(|x|^r - 1)\right\}.$$

Table 1: Choice of $h$ depending on the prediction problem.

| Problem | constraint | solution | instrumental density $h$ | $\tau^2_{h,f}$ |
|---|---|---|---|---|
| $E_{t-1}\|\epsilon_t\|^r$, $r>0$ | $E\|\eta_t\|^r=1$ | $\sigma_t^r$ | $c\|x\|^{\lambda-1}\exp(-\lambda\|x\|^r/r)$, $\lambda>0$ | $\frac{E\|\eta_t\|^{2r}-1}{r^2}$ |
| $E_{t-1}\|\epsilon_t\|^r$, $r<0$ | $E\|\eta_t\|^r=1$ | $\sigma_t^r$ | $c\|x\|^{-\lambda-1}\exp(\lambda\|x\|^r/r)$, $\lambda>0$ | $\frac{E\|\eta_t\|^{2r}-1}{r^2}$ |
| $E_{t-1}\log\|\epsilon_t\|$ | $E\log\|\eta_t\|=0$ | $\log\sigma_t$ | $\sqrt{\lambda/\pi}\|2x\|^{-1}\exp\{-\lambda(\log\|x\|)^2\}$ | $E(\log\|\eta_t\|)^2$ |

Now we prove that $\mathcal{H}\subset\mathcal{H}'$. If $h\notin\mathcal{H}'$ then for some $x_1,x_2$ with $|x_1|\neq 1$, and $\lambda_1\neq\lambda_2$

$$\frac{h'(x_i)}{h(x_i)}x_i+1=\lambda_i(|x_i|^r-1),\quad i=1,2.$$

Let $\eta$ such that $P(\eta=x_i)=p_i>0$ with $p_1+p_2=1$, and $(|x_1|^r-1)p_1+(|x_2|^r-1)p_2=0$. Then $E|\eta|^r=1$ and

$$E\left(\frac{h'(\eta)}{h(\eta)}\eta\right)+1=\lambda_1(|x_1|^r-1)p_1+\lambda_2(|x_2|^r-1)p_2=(\lambda_1-\lambda_2)(|x_1|^r-1)p_1\neq 0.$$

Then $h\notin\mathcal{H}$. We have proven that $\mathcal{H}=\mathcal{H}'$. It remains to verify that $\mathcal{H}=\mathcal{C}(r)$ by solving the differential equation involved in the definition of $\mathcal{H}'$, and the proposition follows. □

In view of Propositions 2.1 and 2.2, it is not restrictive to choose $h$ in the set $\mathcal{C}(r)$ with $\lambda=1$. The choice of the instrumental density $h$ is thus entirely determined by $r$, that is by the prediction problem. This is summarized in Table 1. The last column provides the factors $\tau^2_{h,f}$ which, by Theorem 2.1, measures the impact of $h$ on the asymptotic variance of the QMLE.

The next result characterizes the set of densities $f$ of $\eta_t$ for which a given $h$ is optimal.

**Corollary 2.1.** *Let the assumptions of Theorem 2.1 hold for some $h\in\mathcal{C}(r)$. Then the generalized QMLE based on $h$ coincides with the MLE when the density $f$ of $\eta_t$ belongs to $\mathcal{C}(r)$.*

*Conversely, when $f\notin\mathcal{C}(r)$ but is such that $\tau^2_{f,f}$ exists, any generalized QMLE based on $h\in\mathcal{C}(r)$ is asymptotically inefficient in the sense that $\tau^2_{f,f}<\tau^2_{h,f}$.*

**Proof.** The direct part is straightforward since we have seen that the QMLE does not depend on the choice of $h \in \mathcal{C}(r)$.

Now suppose $f \notin \mathcal{C}(r)$ for $r \neq 0$. Then, by Cauchy-Schwarz

$$\begin{aligned}
\frac{\tau^2_{h,f}}{\tau^2_{f,f}} &= \mathrm{Var}\left(1 + \frac{f'(\eta_0)}{f(\eta_0)}\eta_0\right)\mathrm{Var}\left(\frac{|\eta_0|^r - 1}{r}\right) \\
&\geq \left\{\mathrm{Cov}\left(\frac{f'(\eta_0)}{f(\eta_0)}\eta_0, \frac{|\eta_0|^r}{r}\right)\right\}^2 = \left\{E\left(\frac{f'(\eta_0)}{f(\eta_0)}\eta_0\frac{|\eta_0|^r}{r}\right) + \frac{1}{r}\right\}^2 = 1
\end{aligned}$$

where the last equality is obtained by integration by part. The inequality is strict except if

$$1 + \frac{f'(\eta_0)}{f(\eta_0)}\eta_0 = K(|\eta_0|^r - 1), \quad a.s.$$

for some constant $K$. The last equality is equivalent to $f \in \mathcal{C}(r)$, as already seen. A similar argument holds when $r = 0$. □

# 3 Asymptotic properties for the mixed approach

The mixed approach involves two steps. In a first step, the model is estimated by the standard QMLE and, in a second step, the expectation involved in (1.2) (or (1.3)) is estimated using the estimated rescaled innovations. To obtain the asymptotic properties of this method, it is necessary to derive the joint asymptotic distribution of the estimators of the two steps.

To be able to apply the standard gaussian QMLE, we need to reparameterize the model when **A1** holds. Assume that

**B0:** $E\eta_0^4 < \infty$ and $E|\eta_0|^{2r} < \infty$ when $r \neq 0$, $E\log^2|\eta_0| < \infty$ when $r = 0$,

and let

$$\eta_t^* = \frac{\eta_t}{\sqrt{E\eta_t^2}}.$$

The following assumption is required to reparameterize the model.

**B1:** There exists a function $F$ such that for any $\theta \in \Theta$, for any $K > 0$, and any $(x_i)_i$

$$K\sigma(x_1, x_2, \ldots; \theta) = \sigma(x_1, x_2, \ldots; \theta^*), \quad \text{where } \theta^* = F(\theta, K).$$

Standard GARCH models obviously verify this assumption with

$$F(\theta, K) = (K^2\omega, K^2\alpha_1, \dots, K^2\alpha_q, \beta_1, \dots, \beta_p)' \tag{3.1}$$

and usual notations. Let $\theta_0^* = F(\theta_0, \sqrt{\mu_2})$ where $\mu_s = E|\eta_t|^s$ for $s \neq 0$. The reparameterized model is

$$\begin{cases} \epsilon_t = \sigma_t^* \eta_t^*, & E\eta_t^{*2} = 1, \\ \sigma_t^* = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_0^*) \end{cases} \tag{3.2}$$

The gaussian QMLE of $\theta_0^*$, denoted by $\hat{\theta}_n^*$, is defined as a maximizer over $\Theta$ of

$$\frac{1}{n}\sum_{t=1}^n \log \frac{1}{\tilde{\sigma}_t(\theta)} \phi\left(\frac{\epsilon_t}{\tilde{\sigma}_t(\theta)}\right). \tag{3.3}$$

Let the rescaled residuals

$$\hat{\eta}_t^* = \frac{\epsilon_t}{\hat{\sigma}_t^*}, \quad \text{where } \hat{\sigma}_t^* = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots, \tilde{\epsilon}_0, \tilde{\epsilon}_{-1}, \dots; \hat{\theta}_n^*).$$

We define

$$\hat{\mu}_r^* = \frac{1}{n}\sum_{t=1}^n |\hat{\eta}_t^*|^r, \quad \mu_r^* = E|\eta_t^*|^r = \frac{1}{\mu_2^{r/2}}, \quad \text{for } r \neq 0,$$

$$\hat{\mu}_0^* = \frac{1}{n}\sum_{t=1}^n \log|\hat{\eta}_t^*|, \quad \mu_0^* = E\log|\eta_t^*| = -\frac{1}{2}\log\mu_2, \quad \text{for } r = 0,$$

and $\kappa_s = \frac{E|\eta_t|^s}{\mu_2^{s/2}}$ for any $s \neq 0$. The next result gives the joint asymptotic distribution of the QMLE and $\hat{\mu}_r^*$.

**Theorem 3.1.** *If* **A0**-**A2**, **B0**, **B1** *and, with* $\delta = 2$, **A5**, **A8**-**A10** *hold, and* $\theta_0^* \in \overset{\circ}{\Theta}$, *then*

$$\begin{pmatrix} \sqrt{n}\left(\hat{\theta}_n^* - \theta_0^*\right) \\ \sqrt{n}(\hat{\mu}_r^* - \mu_r^*) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_r), \tag{3.4}$$

*where*

$$\Sigma_r = \begin{pmatrix} (\kappa_4 - 1)J_*^{-1} & -\lambda_r J_*^{-1}\Omega_* \\ -\lambda_r \Omega_*' J_*^{-1} & \sigma_{\mu_r^*}^2 \end{pmatrix},$$

$$J_* = E\left(\frac{1}{\sigma_t^{*4}}\frac{\partial\sigma_t^2(\theta_0^*)}{\partial\theta}\frac{\partial\sigma_t^2(\theta_0^*)}{\partial\theta'}\right), \quad \Omega_*' = E\left(\frac{1}{\sigma_t^{*2}}\frac{\partial\sigma_t^2(\theta_0^*)}{\partial\theta'}\right)$$

*and*

$$\lambda_r = \frac{r}{2}\kappa_r(\kappa_4 - 1) - (\kappa_{2+r} - \kappa_r), \quad \sigma^2_{\mu_r^*} = \kappa_{2r} - \kappa_r^2 + \frac{r}{2}\kappa_r(\lambda_r - \kappa_{2+r} + \kappa_r)$$

*for* $r \neq 0$, *and*

$$\lambda_0 = \frac{\kappa_4 - 1}{2} - Cov\left(\log|\eta_t|, \frac{\eta_t^2}{\mu_2}\right), \quad \sigma^2_{\mu_0^*} = Var(\log|\eta_t|) + \lambda_0 - \frac{\kappa_4 - 1}{4}.$$

**Remark 3.1.** In the proof, the following relation, of independent interest, is established:

$$\Omega_*' J_*^{-1}\Omega_* = 1. \tag{3.5}$$

To show this equality we use an argument based on asymptotic results. A direct proof, based on algebra, will be given in the standard GARCH case (see the proof of Theorem 4.1).

**Remark 3.2.** In the proof of (3.5) it is shown that $\hat{\mu}_2^* = \mu_2^*(= 1)$, $a.s.$ This entails that, when $r = 2$, the two approaches for predicting $\epsilon_t^2$ are the same. In this case, the asymptotic distribution in (3.4) is degenerate.

**Remark 3.3.** In the Gaussian case, $\Sigma_r$ is block-diagonal. Indeed, assume that $\eta_t$ follows a $\mathcal{N}(0, N_r^{-2/r})$ distribution where $r > 0$ and $N_r = E|U|^r$ if $U$ is $\mathcal{N}(0,1)$ distributed. Then $\kappa_4 = 3$ and $\kappa_s = (s-1)\kappa_{s-2}$ for $s \geq 2$. It follows that $\lambda_r = 0$.

# 4 Comparison of the predictors based on the two approaches

By the direct approach, based on the generalized QMLE $\hat{\theta}_{n,h}$, the optimal prediction $E_n|\epsilon_{n+1}|^r$ is estimated by

$$P_{n,h} = \tilde{\sigma}^r(\epsilon_n, \epsilon_{n-1}, \ldots; \hat{\theta}_{n,h}).$$

By the mixed approach, based on the Gaussian QMLE $\hat{\theta}_n^*$ in Model (3.2), the same optimal prediction is estimated by

$$P_n^* = \tilde{\sigma}^r(\epsilon_n, \epsilon_{n-1}, \ldots; \hat{\theta}_n^*)\hat{\mu}_r^* = \tilde{\sigma}^r(\epsilon_n, \epsilon_{n-1}, \ldots; F(\hat{\theta}_n^{*\prime}, \{\hat{\mu}_r^*\}^{1/r})).$$

The optimal prediction $E_n \log|\epsilon_{n+1}|$ can similarly be estimated by

$$P_{n,h} = \log\tilde{\sigma}(\epsilon_n, \epsilon_{n-1}, \ldots; \hat{\theta}_{n,h})$$

and

$$P_n^* = \log\tilde{\sigma}(\epsilon_n, \epsilon_{n-1}, \ldots; \hat{\theta}_n^*) + \hat{\mu}_0^* = \log\tilde{\sigma}(\epsilon_n, \epsilon_{n-1}, \ldots; F(\hat{\theta}_n^{*'}, e^{\hat{\mu}_0^*})).$$

To compare the predictors it suffices to compare the asymptotic distributions of $\hat{\theta}_{n,h}$ and $\tilde{\theta}_n = G_r(\hat{\theta}_n^{*'}, \hat{\mu}_r^*)$ with $G_r(\hat{\theta}_n^{*'}, \hat{\mu}_r^*) = F(\hat{\theta}_n^{*'}, \{\hat{\mu}_r^*\}^{1/r})$ if $r \neq 0$, and $G_0(\hat{\theta}_n^{*'}, \hat{\mu}_0^*) = F(\hat{\theta}_n^{*'}, e^{\hat{\mu}_0^*}))$. Under smoothness assumptions on the function $F$ we have

$$\tilde{\theta}_n \to \theta_0 = G_r(\theta_0^{*'}, \mu_r^*), \quad a.s.$$

and

$$\sqrt{n}\left(\tilde{\theta}_n - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \left[\frac{\partial G_r(\theta_0^{*'}, \mu_r^*)}{\partial(\theta', \mu)}\right] \Sigma_r \left[\frac{\partial G_r(\theta_0^{*'}, \mu_r^*)'}{\partial(\theta', \mu)'}\right]\right). \tag{4.1}$$

The problem is thus to compare

$$4\tau_{h,f}^2 J^{-1} \quad \text{and} \quad \Gamma_r = \left[\frac{\partial G_r(\theta_0^{*'}, \mu_r^*)}{\partial(\theta', \mu)}\right] \Sigma_r \left[\frac{\partial G_r(\theta_0^{*'}, \mu_r^*)'}{\partial(\theta', \mu)'}\right].$$

## 4.1 The standard GARCH$(p, q)$ case

To our knowledge, the mildest assumptions for the $\sqrt{n}$ consistency and AN of the *gaussian* QMLE for standard GARCH with iid errors were obtained by Berkes, Horváth and Kokoszka (2003) and Francq and Zakoïan (2004) (see Escanciano (2009) for an extension to martingale differences, and Hall and Yao (2003) for the asymptotic behavior of the QMLE when $E\eta_t^4 = \infty$). In this section, the results of Theorem 2.1 are applied to the standard GARCH$(p, q)$ model

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t^2 = \omega_0 + \sum_{i=1}^q \alpha_{0i}\epsilon_{t-i}^2 + \sum_{j=1}^p \beta_{0j}\sigma_{t-j}^2 \end{cases} \tag{4.2}$$

where $\theta_0 = (\omega_0, \alpha_{01}, \ldots, \beta_{0p})'$ satisfies $\omega_0 > 0, \alpha_{0i} \geq 0$, $\beta_{0j} \geq 0$. Let $\hat{\theta}_n^* = (\hat{\omega}^*, \hat{\alpha}_1^*, \ldots, \hat{\beta}_p^*)$ be the Gaussian QMLE of $\theta_0^* = (\mu_2\omega_0, \mu_2\alpha_{01}, \ldots, \mu_2\alpha_{0q}, \beta_{01}, \ldots, \beta_{0p})'$. Let $\mathcal{A}_\theta(z) = \sum_{i=1}^q \alpha_i z^i$ and $\mathcal{B}_\theta(z) = 1 - \sum_{j=1}^p \beta_j z^j$. Let $\gamma(\mathbf{A_0})$ denote the top-Lyapunov exponent associated to Model (4.2) (see e.g. Francq and Zakoïan (2004)). For the standard GARCH, several assumptions of Section 2 can be made more explicit as follows.

**C:** $\gamma(\mathbf{A_0}) < 0$; $\forall\theta \in \Theta$, $\sum_{j=1}^{p} \beta_j < 1$ and $\omega > \underline{\omega}$ for some $\underline{\omega} > 0$; $|\eta_0|$ has a non degenerate distribution; 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 next two theorems provide the asymptotic distributions of the estimators of $\theta_0$ involved in the two methods.

**Theorem 4.1** (Standard GARCH$(p,q)$)**.** *Let $r \neq 0$. For $h \in \mathcal{C}(r)$, $E|\eta_0|^r = 1$, $E|\eta_0|^{2r} < \infty$ and under* **C***, the one-step estimator of $\theta_0 \in \overset{\circ}{\Theta}$ satisfies*

$$\sqrt{n}\left(\hat{\theta}_{n,h} - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left\{0, \left(\frac{2}{r}\right)^2 \left(\frac{\kappa_{2r}}{\kappa_r^2} - 1\right) J^{-1}\right\}. \tag{4.3}$$

*Under the same assumptions and $E\eta_0^4 < \infty$, the two-step estimator is given by $\tilde{\theta}_n = (\{\hat{\mu}_r^*\}^{2/r}\hat{\omega}^*, \{\hat{\mu}_r^*\}^{2/r}\hat{\alpha}_1^*, \ldots, \{\hat{\mu}_r^*\}^{2/r}\hat{\alpha}_q^*, \hat{\beta}_1^*, \ldots, \hat{\beta}_p^*)$ and satisfies*

$$\sqrt{n}\left(\tilde{\theta}_n - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left\{0, (\kappa_4 - 1)J^{-1} + \left[\left(\frac{2}{r}\right)^2 \left(\frac{\kappa_{2r}}{\kappa_r^2} - 1\right) - (\kappa_4 - 1)\right] \overline{\theta}_0 \overline{\theta}_0'\right\} \tag{4.4}$$

*where* $\overline{\theta}_0 = \begin{pmatrix} \theta_0^{[1:q+1]} \\ 0_p \end{pmatrix}$, $\theta_0^{[1:q+1]} = (\omega_0, \alpha_{01}, \ldots, \alpha_{0q})'$.

It is interesting to note that when applied to the Gaussian QML ($r = 2$), the assumptions of this theorem reduce to those of the aforementioned papers.

**Theorem 4.2** (Standard GARCH$(p,q)$ when $r = 0$)**.** *For $h \in \mathcal{C}(0)$, $E\log|\eta_0| = 0$, $E\log^2|\eta_0| < \infty$ and under* **C***, the one-step estimator of $\theta_0 \in \overset{\circ}{\Theta}$ satisfies*

$$\sqrt{n}\left(\hat{\theta}_{n,h} - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left\{0, 4\,Var(\log|\eta_0|)J^{-1}\right\}. \tag{4.5}$$

*Under the same assumptions and $E\eta_0^4 < \infty$, the two-step estimator is given by $\tilde{\theta}_n = (e^{2\hat{\mu}_0^*}\hat{\omega}^*, e^{2\hat{\mu}_0^*}\hat{\alpha}_1^*, \ldots, e^{2\hat{\mu}_0^*}\hat{\alpha}_q^*, \hat{\beta}_1^*, \ldots, \hat{\beta}_p^*)$ and satisfies*

$$\sqrt{n}\left(\tilde{\theta}_n - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left\{0, (\kappa_4 - 1)J^{-1} + \left[4\,Var(\log|\eta_0|) - (\kappa_4 - 1)\right] \overline{\theta}_0 \overline{\theta}_0'\right\}. \tag{4.6}$$

The link between the two preceding theorems is given by the following result, showing the continuity at $r = 0$ of the limiting distribution of the two estimators $\tilde{\theta}_n$ and $\hat{\theta}_{n,h}$.

**Proposition 4.1** (Continuity of the asymptotic variance at $r = 0$)**.** *Let $U$ denote a fixed variable (that is independent of $r$) and assume that*

$$\eta_0 \stackrel{d}{=} \frac{U}{(E|U|^r)^{1/r}}.$$

*Then, under the assumptions of Theorems 4.1 and 4.2, we have*

$$\lim_{r \to 0} \left(\frac{2}{r}\right)^2 \left(\frac{\kappa_{2r}}{\kappa_r^2} - 1\right) = 4\, Var(\log|\eta_0|).$$

**Proof.** Note that $EU^4 < \infty$ and $E(\log|U|)^2 < \infty$. Let $f(r) = E|U|^r$. Then, by application of the Lebesgue theorem, $f'(r) = E(|U|^r \log|U|)$ and $f''(r) = E(|U|^r\{\log|U|\}^2)$ for $r$ small enough. Hence

$$E|U|^r = 1 + rE(\log|U|) + \frac{r^2}{2}E(\{\log|U|\}^2) + o(r^2).$$

Thus

$$E|U|^{2r} - (E|U|^r)^2 = r^2 \mathrm{Var}(\log|U|) + o(r^2).$$

Because

$$\frac{\kappa_{2r}}{\kappa_r^2} - 1 = \frac{E|U|^{2r}}{(E|U|^r)^2} - 1,$$

the result straightforwardly follows. □

The next result allows for a very simple comparison of the efficiencies of the two methods.

**Corollary 4.1** (A criterion for efficiency comparison)**.** *Under the assumptions of Theorem 4.1 (resp. Theorem 4.2), the asymptotic variance matrices of the two estimators verify*

$$Var_{as}\left\{\sqrt{n}\left(\hat{\theta}_{n,h} - \theta_0\right)\right\} \succeq Var_{as}\left\{\sqrt{n}\left(\tilde{\theta}_n - \theta_0\right)\right\} \tag{4.7}$$

*in the sense of positive semi-definite matrices, if and only if*

$$\left(\frac{2}{r}\right)^2 \left(\frac{\kappa_{2r}}{\kappa_r^2} - 1\right) \geq \kappa_4 - 1 \qquad (\textit{resp.} \;\; 4\, Var(\log|\eta_0|) \geq \kappa_4 - 1). \tag{4.8}$$

**Proof.** It follows from Theorem 4.1 that, for $r \neq 0$,

$$\begin{aligned} & \text{Var}_{as}\{\sqrt{n}\left(\hat{\theta}_{n,h} - \theta_0\right)\} - \text{Var}_{as}\{\sqrt{n}\left(\tilde{\theta}_n - \theta_0\right)\} \\ = & \left[\left(\frac{2}{r}\right)^2 \left(\frac{\kappa_{2r}}{\kappa_r^2} - 1\right) - (\kappa_4 - 1)\right] (J^{-1} - \overline{\theta}_0 \overline{\theta}_0^{'}) \end{aligned}$$

A similar result holds for $r = 0$, by Theorem 4.2. It remains to show that

$$J^{-1} \succeq \overline{\theta}_0 \overline{\theta}_0^{'}. \tag{4.9}$$

In view of (A.17),

$$\overline{\theta}_0^{'} J = E(Z_t), \qquad Z_t = \frac{1}{\sigma_t^2(\theta_0)} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta}.$$

Thus $J - J\overline{\theta}_0 \overline{\theta}_0^{'} J = \text{Var}(Z_t)$ is positive semi-definite. It follows that

$$y' J (J^{-1} - \overline{\theta}_0 \overline{\theta}_0^{'}) J y = y'(J - J\theta_0 \theta_0^{'} J) y \geq 0, \qquad \forall y \in \mathbb{R}^{q+1}, y \neq 0.$$

Setting $x = Jy$, we thus have

$$x'(J^{-1} - \overline{\theta}_0 \overline{\theta}_0^{'}) x \geq 0, \qquad \forall x \in \mathbb{R}^{q+1}, x \neq 0$$

and (4.9) is proven. □

Note that (4.9) has interest beyond the proof. In particular, it can be used to obtain a simple lower bound for the asymptotic variance of the generalized QMLE.

**Remark 4.1.** It is worth noting that the asymptotic efficiency comparison of the two approaches only depends on $r$ and some moments of the iid process. We shall see below that this property also holds for a family of nonlinear GARCH. The fact that the comparison does not involve the parameter $\theta_0$ is surprising. This result may have crucial importance for practical purposes. It will allow to select straightforwardly the more efficient method, in function of $r$ and estimated moments of $\eta_t$. The latter can be obtained from standardized residuals of a standard GARCH estimation. From a single GARCH estimation, one should be able to decide which method is asymptotically the best for any value of $r$. This will be illustrated in Section 5. Of course, formal tests of the inequalities in (4.8) could be investigated, but this is left for further research.

Figure 1 shows the ARE of the one-step QMLE relative to the two step QMLE as measured by the ratios

$$(\kappa_4 - 1)/\left(\frac{2}{r}\right)^2\left(\frac{\kappa_{2r}}{\kappa_r^2} - 1\right) \quad \text{when } r \neq 0, \text{ and } \frac{\kappa_4 - 1}{4\text{Var}(\log|\eta_0|)} \quad \text{when } r = 0$$

for Student distributions. It is seen that the one-step method outperforms the indirect one when $r \in (0.5, 2)$. On the contrary, for $r > 2$ and small or negative values of $r$, the two-step approach is preferable. The differences are particularly spectacular for small value of $\nu$. The ARE's are displayed as dots in the case $r = 0$. The continuity property when $r$ approaches zero, established in Proposition 4.1, can thus be visualized on this graph.

Figure 2 displays the same ratios for the Generalized Error Distribution (GED) with parameter $\nu$ [1]. Contrary to the previous graph, it can be seen that the direct method can be superior to the two-step approach for $r > 2$.

## 4.2 The Asymmetric Power GARCH$(p, q)$ case

The following nonlinear GARCH$(p, q)$ model was introduced by Ding, Granger and Engle (1993). Letting $x^+ = \max(x, 0)$ and $x^- = \min(x, 0)$ we set, for a given $\delta > 0$,

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t^\delta = \omega_0 + \sum_{i=1}^q \alpha_{0i+}(\epsilon_{t-i}^+)^\delta + \alpha_{0i-}(-\epsilon_{t-i}^-)^\delta + \sum_{j=1}^p \beta_{0j}\sigma_{t-j}^\delta \end{cases} \tag{4.10}$$

where $\alpha_{0i+}, \alpha_{0i-}, \beta_{0j}$ are nonnegative coefficients, and $\omega_0 > 0$. This model allows to capture the so-called "leverage effect", and generalizes models introduced by Higgins and Bera (1992), and Zakoïan (1994).

Pan, Wang and Tong (2008) established that the strict stationarity condition writes $\gamma(\mathbf{B_0}) < 0$, where $\gamma(\mathbf{B_0})$ is the top-Lyapunov exponent associated to Model (4.10). This condition entails the invertibility of the polynomial $\mathcal{B}_{\theta_0}(z)$ and allows to write the model under the form (1.1). It also ensures the existence of $E|\epsilon_t|^s$ for some $s > 0$.

[1] The density of $\eta_0$ is of the form $f(x) \propto e^{-0.5|x|^{1/\nu}}$.

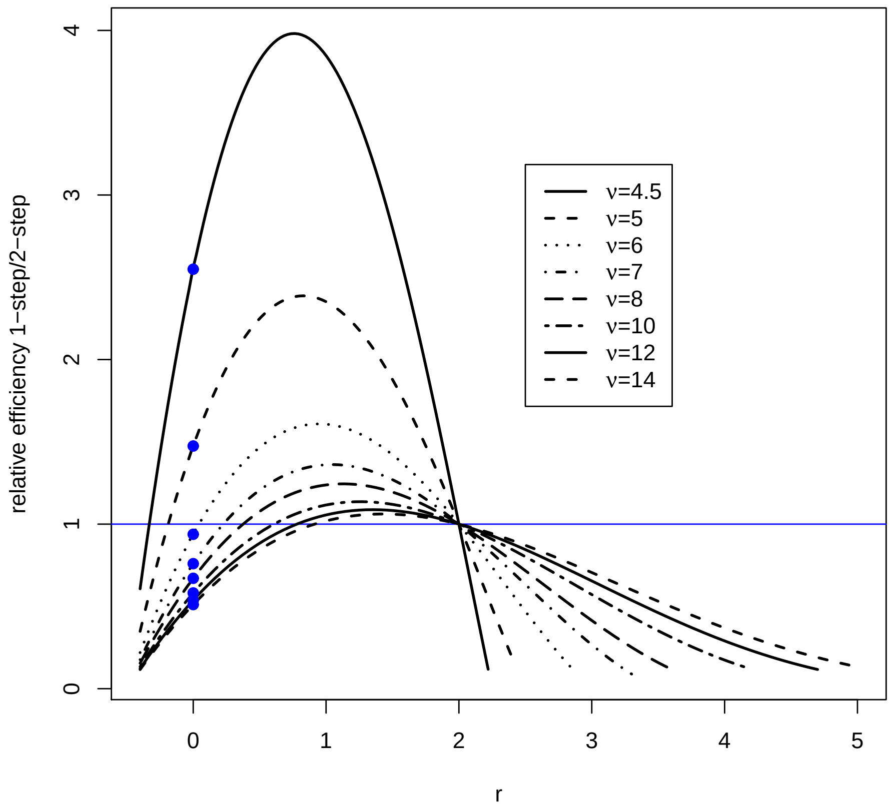


Figure 1: Relative efficiency of the one-step QMLE relative to the two step QMLE for Student distributions with parameter $\nu$.

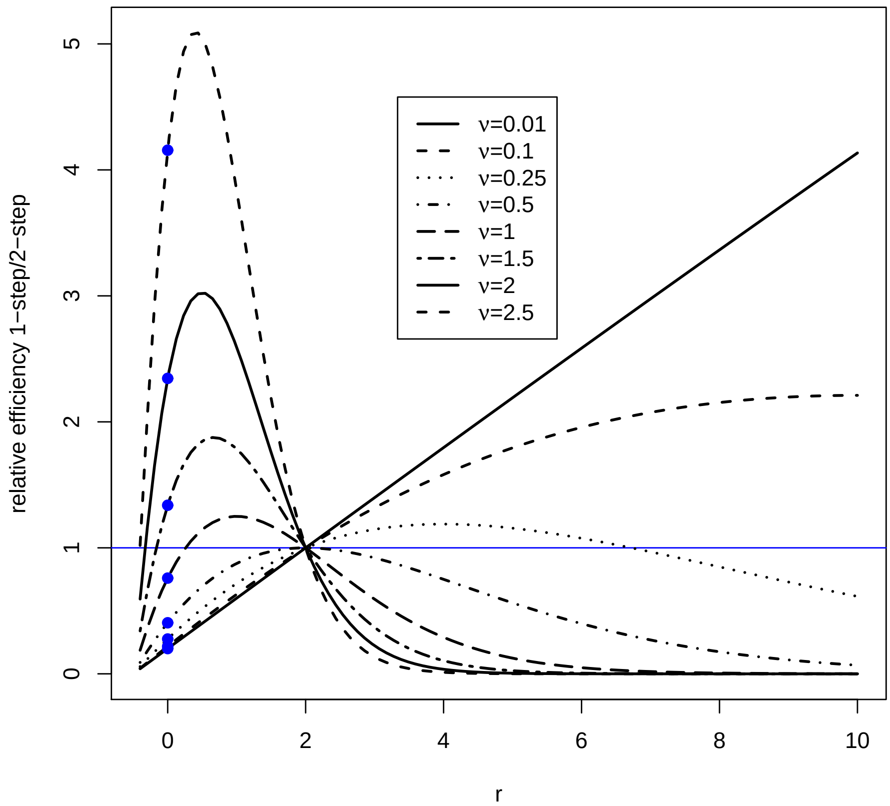


Figure 2: Same as Figure 1 for GED with parameter $\nu$.

With obvious notation, Assumption **B1** holds with

$$F(\theta, K) = (K^\delta \omega, K^\delta \alpha_{1+}, K^\delta \alpha_{1-}, \ldots, K^\delta \alpha_{q-}, \beta_1, \ldots, \beta_p)'.$$

Hamadeh and Zakoïan (2009) showed that the following assumption entails AN of the Gaussian QMLE of $\theta_0 = (\omega_0, \alpha_{01+}, \ldots, \alpha_{0q-}, \beta_{01}, \ldots, \beta_{0p})'$.

**D:** $\gamma(\mathbf{B_0}) < 0$; $\forall \theta \in \Theta$, $\sum_{j=1}^p \beta_j < 1$ and $\omega > \underline{\omega}$ for some $\underline{\omega} > 0$; if $P(\eta_t \in \Gamma) = 1$ for a set $\Gamma$, then $\Gamma$ has a cardinal $|\Gamma| > 2$; $P[\eta_t > 0] \in (0, 1)$; if $p > 0$, $\mathcal{B}_{\theta_0}(z)$ has no common root with $\mathcal{A}_{\theta_0+}(z)$ and $\mathcal{A}_{\theta_0-}(z)$. Moreover $\mathcal{A}_{\theta_0+}(1) + \mathcal{A}_{\theta_0-}(1) \neq 0$ and $\alpha_{0q,+} + \alpha_{0q,-} + \beta_{0p} \neq 0$.

**Theorem 4.3** (Asymmetric Power GARCH$(p, q)$)**.** *Let $r \neq 0$. For $h \in \mathcal{C}(r)$, $E|\eta_0|^r = 1$, $E|\eta_0|^{2r} < \infty$ and under* **D***, the one-step estimator of $\theta_0 \in \overset{\circ}{\Theta}$ satisfies (4.3).*

*Under the same assumptions and $E\eta_0^4 < \infty$, the two-step estimator is given by $\tilde{\theta}_n = (\{\hat{\mu}_r^*\}^{\delta/r} \hat{\omega}^*, \{\hat{\mu}_r^*\}^{\delta/r} \hat{\alpha}_1^*, \ldots, \{\hat{\mu}_r^*\}^{\delta/r} \hat{\alpha}_q^*, \hat{\beta}_1^*, \ldots, \hat{\beta}_p^*)$ and satisfies*

$$\sqrt{n}\left(\tilde{\theta}_n - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left\{0, (\kappa_4 - 1)J^{-1} + \left(\frac{\delta}{2}\right)^2 \left[\left(\frac{2}{r}\right)^2 \left(\frac{\kappa_{2r}}{\kappa_r^2} - 1\right) - (\kappa_4 - 1)\right] \overline{\theta}_0 \overline{\theta}_0'\right\}$$

*where* $\overline{\theta}_0 = \begin{pmatrix} \theta_0^{[1:2q+1]} \\ 0_p \end{pmatrix}$, $\theta_0^{[1:2q+1]} = (\omega_0, \alpha_{01+}, \ldots, \alpha_{0q-})'$.

*Moreover, the conclusion of Corollary 4.1 holds true for Model (4.10): the estimator $\tilde{\theta}_n$ is asymptotically more efficient than $\hat{\theta}_{n,h}$ iff (4.8) holds.*

## 4.3 The ARCH$(\infty)$ case

A process $(\epsilon_t)$ is called an ARCH$(\infty)$ if there exists a sequence of constants $\psi_{00} > 0$ and $\psi_{0i} \geq 0$, $i = 1, \ldots$, such that

$$\epsilon_t = \sigma_t \eta_t, \quad \sigma_t^2 = \psi_{00} + \sum_{i=1}^{\infty} \psi_{0i} \epsilon_{t-i}^2. \tag{4.11}$$

This class of models, which extends that of the GARCH, was introduced by Robinson (1991). Robinson and Zaffaroni (2006) and Douc, Roueff and

Soulier (2008) showed that there exists a strictly stationary and non anticipative solution to (4.11) satisfying $E|\epsilon_t|^{2s} < \infty$ if $A_s\mu_{2s} < 1$, for some $s \in (0,1]$, where $A_s = \sum_{i=1}^{\infty} \psi_{0i}^s$ and $\mu_{2s} = E|\eta_t|^{2s}$. The reader is referred to Giraitis, Kokoszka and Leipus (2000), Kazakevičius and Leipus (2007), Giraitis, Leipus and Surgailis (2008), and the references therein, for the other probabilistic properties of these models. Conditions ensuring consistency and asymptotic normality of the QMLE have been obtained by Robinson and Zaffaroni (2006). Parameterizing the coefficients as $\psi_{0i} = \psi_i(\theta_0)$ with known functions $\psi_i(\cdot) : \Theta \to [0,\infty)$, for $s \in (0,1]$ let us consider the following assumptions. In these assumptions and in the forthcoming derivations, the letter $C$ denotes any positive constant whose exact value is unimportant.

**E**$(s)$: $A_s\mu_{2s} < 1$; $\forall\theta \in \Theta$, we have $\psi_0(\theta) > \underline{\omega}$ for some $\underline{\omega} > 0$; if $\theta \neq \theta_0$ then $\{\psi_i(\theta)\} \neq \{\psi_i(\theta_0)\}$; $|\eta_0|$ has a non degenerate distribution; for all $i \geq 1$, $\sup_{\theta\in\Theta} \psi_i(\theta) \leq Ci^{-\underline{d}-1}$ for some $\underline{d}$ such that $s(\underline{d}+1) > 1$; for $k = 1,2,3$ and all $i \geq 0$, $\psi_i(\cdot)$ has a continuous $k$-th derivative on $\Theta$ such that, for all $i_1,\dots,i_k \in \{1,\dots,m\}$, $\left|\partial^k\psi_i(\theta)/\partial\theta_{i_1}\dots\partial\theta_{i_k}\right| \leq C\psi_i^{1-\iota}(\theta)$ for all $\iota > 0$ when $\psi_i(\theta) > 0$ and $\left|\partial^k\psi_i(\theta)/\partial\theta_{i_1}\dots\partial\theta_{i_k}\right| = 0$ when $\psi_i(\theta) = 0$; $\forall\iota > 0$, there exists a neighborhood $V(\theta_0)$ of $\theta_0$ such that $\sup_{\theta\in V(\theta_0)} \psi_i(\theta_0)/\psi_i(\theta) \leq C\psi_i^{-\iota}(\theta_0)$ and $\sup_{\theta\in V(\theta_0)} \psi_i(\theta)/\psi_i(\theta_0) \leq C\psi_i^{-\iota}(\theta_0)$; $K\psi_i(\theta) = \psi_i(\theta^*)$ where $\theta^* = H(\theta,K)$ and $H$ is continuously differentiable.

**F**$(s)$: There exist $V(\theta_0)$ neighborhood of $\theta_0$ and $d_0 > 1/2$ such that $\sup_{\theta\in V(\theta_0)} \psi_i(\theta) \leq Ci^{-d_0-1}$ and $s(d_0/2+3/4) > 1$.

**Theorem 4.4** (ARCH($\infty$))**.** *Let $r > 0$. When $r \in (0,2]$, assume* **E**$(s)$ *for $s = r/2$. When $r > 2$, assume $E|\epsilon_1|^r < \infty$ and* **E**$(s)$ *for some $s \in (0,1]$. For $h \in \mathcal{C}(r)$ and $E|\eta_0|^r = 1$ the one-step estimator $\hat{\theta}_{n,h}$ tends to $\theta_0$ a.s.*

*If, in addition, $E|\eta_0|^{2r} < \infty$, $\theta_0 \in \overset{\circ}{\Theta}$, and* **F**$(s)$ *holds, $\hat{\theta}_{n,h}$ has the asymptotic normal distribution given by (4.3).*

*Under the same assumptions and $E\eta_0^4 < \infty$, the two-step estimator is given by $\tilde{\theta}_n = H(\hat{\theta}_n^*, \hat{\mu}_r^{*2/r})$ and satisfies (4.1) with $G_r(\theta,\mu) = H(\theta,\mu^{2/r})$.*

Note that, in **F**$(s)$, the constraints on $d_0$ are exactly those of the Assumption H of Robinson and Zaffaroni (2006).

**Remark 4.2.** The simplest ARCH($\infty$) is obtained for $\psi_i(\theta) = c/i^{d+1}$ when $i \geq 1$ and $\psi_0(\theta) = 1$, with $\theta = (c,d) \in \Theta = [\underline{c},\overline{c}] \times [\underline{d},\overline{d}]$ where $0 < \underline{c} < \overline{c}$ and

$0 < \underline{d} < \overline{d}$. In this case $\mathbf{E}(s)$ reduces to

$$c_0 \sum_{i=1}^{\infty} i^{-(d_0+1)s} \mu_{2s} < 1 \text{ and } s(\underline{d}+1) > 1$$

and $\mathbf{F}(s)$ to

$$d_0 > 1/2 \text{ and } s(d_0/2 + 3/4) > 1.$$

# 5 Empirical Illustration

We now compare the two prediction methods on daily returns of 10 world stock market indices, namely the CAC, DAX, DJA, DJI, DJT, DJU, FTSE, Nikkei, SMI and SP500, from January 2, 1990, to January 22, 2009, for the indices for which such historical data exist. For each series $\epsilon_t$ of length $N$, and for $n$ varying from 250 to $N-1$, an "historical" prediction is computed by the formula

$$\text{Historic}_{n+1} = \frac{1}{250} \sum_{t=n-249}^{n} |\epsilon_{t+1}|^r.$$

The Mean Square Prediction Error (MSPE)

$$\frac{1}{N-250} \sum_{n=250}^{N-1} \left(|\epsilon_{n+1}|^r - \text{Historic}_{n+1}\right)^2$$

is reported in the column "Historical" of Table 2. The table also displays the MSPE's of the parametric and mixed prediction methods based on GARCH(1,1) models fitted on 250 past values:

$$\frac{1}{N-250} \sum_{n=250}^{N-1} \left(|\epsilon_{n+1}|^r - P_{n,h}\right)^2 \quad \text{and} \quad \frac{1}{N-250} \sum_{n=250}^{N-1} \left(|\epsilon_{n+1}|^r - P_n^*\right)^2$$

where $P_{n,h} = \tilde{\sigma}^r(\epsilon_n, \ldots, \epsilon_{n-249}; \hat{\theta}_{n,h})$ and $P_n^* = \tilde{\sigma}^r(\epsilon_n, \ldots, \epsilon_{n-249}; \hat{\theta}_n^*)\hat{\mu}_r^*$. In almost all cases, the historical method is outperformed by the approaches based on the GARCH(1,1) model. For $r = 1$, $r = 1.5$ and $r = 0.5$, the direct approach is superior to the two-step method (with some cases of equality, mostly for $r = 0.5$) in terms of MSPE. This is not surprising in view of

Corollary 2.1 and the efficiency comparisons displayed Figures 1 and 2. [2] For $r = 0$, that is for the prediction of $\log |\epsilon_t|$, the results are more balanced: for five assets the direct approach is better, while the indirect one is preferable for the remaining five assets. Finally, the two-step approach provides better results when $r = -0.5$, in all cases except two. Again, this is not surprising from the theoretical results of Section 4.

Another comparison of the two approaches can be done, based on the results of Section 4. Figure 3 presents the estimated relative efficiencies of the one-step QMLE relative to the two step QMLE for the ten stock index returns. A standard GARCH(1,1) model is estimated by Gaussian QML in a first step. In a second step, the standardized residuals $\hat{\eta}_t^* = \epsilon_t/\hat{\sigma}_t^*$ are computed. Finally, the generalized kurtosis coefficients are estimated to compute the ARE's

$$(\hat{\kappa}_4 - 1)/\left(\frac{2}{r}\right)^2\left(\frac{\hat{\kappa}_{2r}}{\hat{\kappa}_r^2} - 1\right), \qquad \hat{\kappa}_s = \frac{1}{n}\sum_{t=1}^{n}|\hat{\eta}_t^*|^s.$$

The conclusions are similar to those drawn from the MSPE: the direct approach is superior for $r \in (0.5, 2)$; the indirect one is preferable for $r > 2$ or $r < -0.5$. For $r \in (-0.5, 0.5)$ the results are more balanced.

# 6 Conclusion

We have shown that, in a general conditionally heteroskedastic models, the optimal predictions of powers (or the logarithm) of the observed process can be estimated in one step, using a non gaussian QML method applied to a reparameterization of the model. By comparison, the traditional approach requires estimating moments of the latent independent process.

We obtained a complete characterization of the omnibus instrumental densities $h$ which render the generalized QMLE universally consistent. The asymptotic properties of the generalized QMLE are studied in a quite general framework (including in particular the ARCH($\infty$)) under conditions which

[2] Of course, for these assets the underlying error distribution is unknown but many empirical studies have shown that distributions such as the Student or GED are plausible for stock returns.

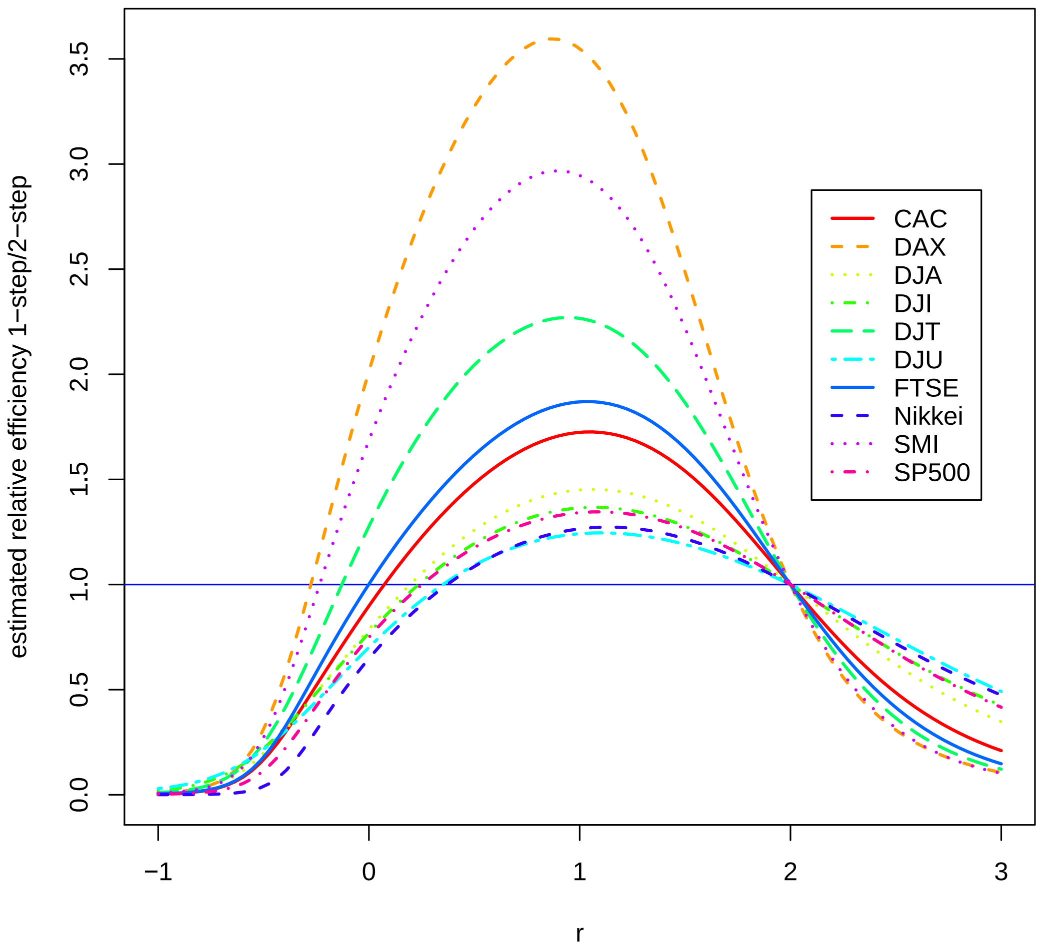


Figure 3: Estimated relative efficiency of the one-step QMLE relative to the two step QMLE for stock index returns.

Table 2: Mean square prediction error of $|\epsilon_{n+1}|^r$ for 10 series of daily returns $(\epsilon_t)$ and different powers $r$. For the column "Historic", the predictions are rolling averages of the $r$-powers the last 250 returns. The two other predictions methods are obtained from GARCH(1,1) models fitted on the last 250 values.

| $r$ | -0.5 | -0.5 | -0.5 | 0 | 0 | 0 | 0.5 | 0.5 | 0.5 |
|---|---|---|---|---|---|---|---|---|---|
| | Historic | $P_{n,h}$ | $P_n^*$ | Historic | $P_{n,h}$ | $P_n^*$ | Historic | $P_{n,h}$ | $P_n^*$ |
| CAC | 4.405 | 4.405 | 4.384 | 1.357 | 1.322 | 1.317 | 0.181 | 0.169 | 0.169 |
| DAX | 5.514 | 5.478 | 5.484 | 1.382 | 1.322 | 1.328 | 0.188 | 0.172 | 0.173 |
| DJA | 4.033 | 4.018 | 4.026 | 1.399 | 1.360 | 1.367 | 0.137 | 0.127 | 0.128 |
| DJI | 4.564 | 4.544 | 4.538 | 1.381 | 1.340 | 1.344 | 0.142 | 0.132 | 0.133 |
| DJT | 5.296 | 5.323 | 5.275 | 1.322 | 1.303 | 1.298 | 0.179 | 0.173 | 0.173 |
| DJU | 6.208 | 6.233 | 6.176 | 1.322 | 1.272 | 1.270 | 0.144 | 0.130 | 0.130 |
| FTSE | 3.899 | 3.888 | 3.878 | 1.312 | 1.267 | 1.263 | 0.143 | 0.131 | 0.131 |
| Nikkei | 6.459 | 6.454 | 6.431 | 1.431 | 1.381 | 1.379 | 0.210 | 0.194 | 0.195 |
| SMI | 5.896 | 5.883 | 5.864 | 1.340 | 1.271 | 1.274 | 0.156 | 0.139 | 0.140 |
| SP500 | 8.524 | 8.522 | 8.503 | 1.412 | 1.371 | 1.372 | 0.146 | 0.135 | 0.135 |

| $r$ | 1 | 1 | 1 | 1.5 | 1.5 | 1.5 | 2 | 2 | 2 |
|---|---|---|---|---|---|---|---|---|---|
| | Historic | $P_{n,h}$ | $P_n^*$ | Historic | $P_{n,h}$ | $P_n^*$ | Historic | $P_{n,h}$ | $P_n^*$ |
| CAC | 0.896 | 0.805 | 0.806 | 4.418 | 3.919 | 3.925 | 26.181 | 23.671 | 23.671 |
| DAX | 0.972 | 0.861 | 0.867 | 4.944 | 4.391 | 4.407 | 29.561 | 27.106 | 27.106 |
| DJA | 0.534 | 0.474 | 0.475 | 2.323 | 2.018 | 2.026 | 13.280 | 11.871 | 11.871 |
| DJI | 0.583 | 0.521 | 0.523 | 2.666 | 2.337 | 2.347 | 15.931 | 14.375 | 14.375 |
| DJT | 0.909 | 0.865 | 0.870 | 4.859 | 4.726 | 4.840 | 36.731 | 40.754 | 40.754 |
| DJU | 0.659 | 0.550 | 0.551 | 3.558 | 2.903 | 2.908 | 26.030 | 22.396 | 22.396 |
| FTSE | 0.600 | 0.524 | 0.526 | 2.623 | 2.252 | 2.257 | 14.311 | 12.487 | 12.487 |
| Nikkei | 1.119 | 0.989 | 0.991 | 5.950 | 5.060 | 5.069 | 40.025 | 33.775 | 33.775 |
| SMI | 0.686 | 0.584 | 0.584 | 3.022 | 2.526 | 2.525 | 15.985 | 13.588 | 13.588 |
| SP500 | 0.632 | 0.557 | 0.557 | 3.078 | 2.654 | 2.654 | 19.682 | 17.395 | 17.395 |

reduce to the weakest ones in the particular case of the standard GARCH. We also derived the asymptotic properties of the two-step approach. It is important to note that the technical assumptions required for the two methods are not exactly the same. In particular, the existence of a fourth moment for the iid errors is required for the AN of the gaussian QMLE. It follows that the validity of the standard two-step approach is questionable for predicting $|\epsilon_t|^r$ with $r < 2$ when $E\eta_t^4 = \infty$. The one-step approach allows to handle this situation.

In the case of finite-order GARCH models, we obtained a surprisingly simple expression for the ARE of the two methods. The latter does not depend on the parameter value, but vary with the power $r$ and some characteristics of the distribution of the iid process. In practice it is simple to estimate this ARE, and therefore it is possible to determine which method is asymptotically the best. Numerical comparisons for two classes of distributions and an empirical study showed that the one-step approach is in general preferable when $r$ is neither too large nor too small. Future work will propose tests for the superiority of one method over the other based on a single gaussian QMLE of the model.

# A Technical assumptions and proofs

Let $\Delta_t(\theta) = \tilde{\sigma}_t(\theta) - \sigma_t(\theta)$ and let $a_t = \sup_{\theta\in\Theta} |\Delta_t(\theta)|$.

**A5:** The function $\theta \mapsto \sigma(x_1, x_2, \dots; \theta)$ is continuous. When $\delta > 0$ we have $E|\epsilon_0|^s < \infty$ and $\sum_{t=1}^{\infty}(Ea_t^{s/\delta})^{1/2} < \infty$ for some $s > 0$, when $\delta \in [-1, 0]$ we have $\sum_{t=1}^{\infty}(Ea_t^{s})^{1/2} < \infty$ for some $s \in (0, 1)$, when $\delta < -1$ we have $\sum_{t=1}^{\infty}(Ea_t^{-s/(\delta+1)})^{1/2} < \infty$ and $E\sup_{\theta\in\Theta}\sigma_0^s(\theta) < \infty$ for some $s \in (0, \min\{-2(1+\delta), \delta(1+\delta)\})$.

**A6:** $\theta_0$ belongs to the interior $\overset{\circ}{\Theta}$ of $\Theta$.

**A7:** $h$ is twice differentiable with $|u^2\,(h'(u)/h(u))'\,| \le C_0(1 + |u|^\delta)$ for all $u \in \mathbb{R}$ and $E|\eta_0|^{2\delta} < \infty$.

**A8:** For any real sequence $(x_i)$, the function $\theta \mapsto \sigma(x_1, x_2, \dots; \theta)$ has continuous second-order derivatives. We have $Ea_t^{s_1} < \infty$ for some $s_1 \in (0, 1)$. There exists a neighborhood $V(\theta_0)$ of $\theta_0$ and a positive number $s_2$ such that $\sum_{t=1}^{\infty}(Eb_t^{s_1 s_2})^{1/2} < \infty$ and $E|\epsilon_0|^{2\delta s_1 s_2} < \infty$ where

$$b_t = \sup_{\theta\in V(\theta_0)} \left\| \frac{\partial \Delta_t(\theta)}{\partial \theta} \right\|, \quad s_1 s_2 \in (0, 1), \quad 2\delta s_2 > -1.$$

**A9:** There exists a neighborhood $V(\theta_0)$ of $\theta_0$ such that, the following variables

$$\sup_{\theta\in V(\theta_0)}\left\|\frac{1}{\sigma_t(\theta)}\frac{\partial\sigma_t(\theta)}{\partial\theta}\right\|^4,\quad \sup_{\theta\in V(\theta_0)}\left\|\frac{1}{\sigma_t(\theta)}\frac{\partial^2\sigma_t(\theta)}{\partial\theta\partial\theta'}\right\|^2,\quad \sup_{\theta\in V(\theta_0)}\left|\frac{\sigma_t(\theta_0)}{\sigma_t(\theta)}\right|^{2\delta}$$

have finite expectations.

**A10:** Let $c_t = \sup_{\theta\in V(\theta_0)}|\Delta_t(\theta)|$ where $V(\theta_0)$ is a neighborhood of $\theta_0$. When $\delta > 0$ we have $E|\epsilon_0|^s < \infty$ and $\sum_{t=1}^{\infty}(Ec_t^{s/2\delta})^{1/2} < \infty$ for some $s \in (0, 4\delta)$, when $\delta \in [-1, 0]$ we have $\sum_{t=1}^{\infty}(Ec_t^{s/2})^{1/2} < \infty$ for some $s \in (0, 1)$, when $\delta < -1$ we have $\sum_{t=1}^{\infty}(Ec_t^{-s/2(\delta+1)})^{1/2} < \infty$ and $E\sup_{\theta\in\Theta}\sigma_0^s(\theta) < \infty$ for some $s \in (0, \min\{-4(1+\delta), 2\delta(1+\delta)\})$.

Let $C$ and $\rho$ be generic constants, whose values will be modified along the proofs, such that $C > 0$ and $0 < \rho < 1$.

## A.1 Proof of Theorem 2.1.

The consistency is a consequence of the following intermediate results:

$$\begin{aligned}
&i)\ \lim_{n\to\infty}\sup_{\theta\in\Theta}|Q_n(\theta) - \tilde{Q}_n(\theta)| = 0\ , \quad a.s.\\
&ii)\ \text{if } \theta \neq \theta_0\ , \quad \mathbb{E}g(\epsilon_1, \sigma_1(\theta)) < \mathbb{E}g(\epsilon_1, \sigma_1(\theta_0))\ ,\\
&iii)\ \text{any } \theta \neq \theta_0 \text{ has a neighborhood } V(\theta) \text{ such that}\\
&\qquad \limsup_{n\to\infty}\sup_{\theta^*\in V(\theta)}\tilde{Q}_n(\theta^*) < \limsup_{n\to\infty}\tilde{Q}_n(\theta_0)\ , \quad a.s.
\end{aligned}$$

where

$$Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} g(\epsilon_t, \sigma_t(\theta)).$$

The asymptotic normality is proven by means of the following intermediate results: for some neighboorhood $V(\theta_0)$ of $\theta_0$,

$$\begin{aligned}
&iv)\ \lim_{n\to\infty}\sqrt{n}\sup_{\theta\in V(\theta_0)}\left\|\frac{\partial}{\partial\theta}Q_n(\theta) - \frac{\partial}{\partial\theta}\tilde{Q}_n(\theta)\right\| = 0\ , \quad \text{in probability},\\
&v)\ \frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\theta^*) \to \frac{Eg_2(\eta_0, 1)}{4}J\ , \quad \text{in probability},\\
&vi)\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{Eg_1^2(\eta_0, 1)}{4}J\right),
\end{aligned}$$

for any $\theta^*$ between $\hat{\theta}_{n,h}$ and $\theta_0$.

To save place, we only give the proof of *i)* and *iv)* which deal with the effect of the initial values. These points constitute the most delicate parts of the proof, and illustrate the necessity of assumptions of the form **A4**, **A5** and **A8-A10**.

We begin to show *i)*. Using a Taylor expansion, almost surely

$$\begin{aligned}
\sup_{\theta\in\Theta}|Q_n(\theta)-\tilde{Q}_n(\theta)| &\leq n^{-1}\sum_{t=1}^{n}\sup_{\theta\in\Theta}|g(\epsilon_t,\sigma_t(\theta))-g(\epsilon_t,\tilde{\sigma}_t(\theta))| \\
&\leq n^{-1}\sum_{t=1}^{n}\sup_{\theta\in\Theta}|g_1(\epsilon_t,\sigma_t^*(\theta))||\Delta_t(\theta)| \\
&\leq n^{-1}\sum_{t=1}^{n}a_t\sup_{\theta\in\Theta}\left|\frac{1}{\sigma_t^*}\frac{\epsilon_t}{\sigma_t^*}\frac{h'}{h}\left(\frac{\epsilon_t}{\sigma_t^*}\right)\right|+\frac{1}{\underline{\omega}}\,n^{-1}\sum_{t=1}^{n}a_t \\
&\leq n^{-1}\sum_{t=1}^{n}a_t|\epsilon_t|^{\delta}\sup_{\theta\in\Theta}\left|\frac{1}{\sigma_t^*}\right|^{1+\delta}+\frac{C}{n}\sum_{t=1}^{n}a_t \qquad \text{(A.1)}
\end{aligned}$$

where $\sigma_t^*(\theta)$ is between $\tilde{\sigma}_t(\theta)$ and $\sigma_t(\theta)$. The last two inequalities rest on Assumptions **A4** and **A2**. First suppose $\delta \geq -1$. Then the last supremum is bounded by $C$. If $\delta > 0$, by the Markov and Cauchy-Schwarz inequalities and **A5**, we deduce

$$\sum_{t=1}^{\infty}\mathbb{P}(a_t|\epsilon_t|^{\delta}>\varepsilon) \leq \sum_{t=1}^{\infty}\frac{\left(\mathbb{E}a_t^{s/\delta}\mathbb{E}|\epsilon_t|^{s}\right)^{1/2}}{\varepsilon^{\frac{s}{2\delta}}}<\infty \qquad \text{(A.2)}$$

and thus $a_t|\epsilon_t|^{\delta} \to 0$ *a.s* by the Borel-Cantelli lemma. The first term in (A.1) thus tends to zero *a.s.*, when $\delta > 0$, by the Cesàro lemma. Now, if $\delta \in [-1, 0]$, we note that $E|\epsilon_t|^{\delta} < \underline{\omega}^{-\delta}E|\eta_t|^{\delta} < \infty$. Note also that, for $s \in (0, 2)$, the $c_r$ inequality (see Loève, 1977) entails

$$\left(n^{-1}\sum_{t=1}^{n}a_t|\epsilon_t|^{\delta}\right)^{s/2} \leq n^{-s/2}\sum_{t=1}^{\infty}a_t^{s/2}|\epsilon_t|^{\delta s/2}.$$

The last sum is *a.s.* finite since its expectation is finite by **A5**, Cauchy-Schwarz's inequality and $E|\epsilon_t|^{\delta s} < \infty$ (because $s \in (0, 1)$). Hence the first term in (A.1) tends to zero *a.s.* when $\delta \in [-1, 0]$. Now suppose $\delta < -1$. Observe that $\sup_{\theta\in\Theta}\sigma_t^*(\theta) \leq \sup_{\theta\in\Theta}\sigma_t(\theta) + a_t$. It follows that the first term in (A.1) can be bounded by

$$\begin{aligned}
&\frac{C}{n}\sum_{t=1}^{n}a_t|\eta_t|^{\delta}\{\sup_{\theta\in\Theta}\sigma_t(\theta)+a_t\}^{-(1+\delta)} \\
\leq\ &\frac{C}{n}\sum_{t=1}^{n}a_t|\eta_t|^{\delta}\{\sup_{\theta\in\Theta}\sigma_t(\theta)\}^{-(1+\delta)}+\frac{C}{n}\sum_{t=1}^{n}a_t^{-\delta}|\eta_t|^{\delta}. \qquad \text{(A.3)}
\end{aligned}$$

Now by **A5**,

$$\begin{aligned}
&\sum_{t=1}^{\infty}\mathbb{P}(a_t|\eta_t|^{\delta}\{\sup_{\theta\in\Theta}\sigma_t(\theta)\}^{-(1+\delta)}>\varepsilon) \\
\leq\ &\sum_{t=1}^{\infty}\frac{\left(\mathbb{E}a_t^{\frac{s}{-(1+\delta)}}\mathbb{E}\sup_{\theta\in\Theta}\sigma_0^{s}(\theta)\right)^{1/2}E|\eta_t|^{\frac{s\delta}{-2(1+\delta)}}}{\varepsilon^{\frac{s}{-2(1+\delta)}}}<\infty. \qquad \text{(A.4)}
\end{aligned}$$

Thus, the first term in the right-hand side of (A.3) tends to zero $a.s.$ by the Cesàro Lemma. The second term is treated similarly. We have shown that the first term in the right-hand side of (A.1) tends to zero $a.s.$ whatever the value of $\delta$. By the Borel-Cantelli lemma, we show that the second term in (A.1) tends to zero $a.s.$ noting that, in **A5**, the powers of $a_t$ are positive whatever the value of $\delta$. Thus *i)* follows.

Now we prove *iv)*. We have

$$\frac{\partial}{\partial\theta}Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} g_1(\epsilon_t, \sigma_t(\theta))\frac{\partial\sigma_t(\theta)}{\partial\theta}, \quad \frac{\partial}{\partial\theta}\tilde{Q}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} g_1(\epsilon_t, \tilde{\sigma}_t(\theta))\frac{\partial\tilde{\sigma}_t(\theta)}{\partial\theta}.$$

It follows that

$$\begin{aligned} &\sup_{\theta\in V(\theta_0)} \sqrt{n}\left\| \frac{\partial}{\partial\theta}Q_n(\theta) - \frac{\partial}{\partial\theta}\tilde{Q}_n(\theta)\right\| \\ \leq\ & \sup_{\theta\in V(\theta_0)} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} |g_1(\epsilon_t, \sigma_t(\theta)) - g_1(\epsilon_t, \tilde{\sigma}_t(\theta))| \left\| \frac{\partial\sigma_t(\theta)}{\partial\theta}\right\| \\ &+ \sup_{\theta\in V(\theta_0)} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} |g_1(\epsilon_t, \tilde{\sigma}_t(\theta))| \left\| \frac{\partial\sigma_t(\theta)}{\partial\theta} - \frac{\partial\tilde{\sigma}_t(\theta)}{\partial\theta}\right\|. \end{aligned} \tag{A.5}$$

Similarly to (A.1), the last term is bounded on $V(\theta_0)$ by

$$\begin{aligned} &\frac{C}{\sqrt{n}}\sum_{t=1}^{n} b_t \left\{ |\epsilon_t|^{\delta} \sup_{\theta\in V(\theta_0)} \left|\frac{1}{\tilde{\sigma}_t(\theta)}\right|^{1+\delta} + 1\right\} \\ \leq\ & \frac{C}{\sqrt{n}}\sum_{t=1}^{n} b_t|\eta_t|^{\delta} \sup_{\theta\in V(\theta_0)} \left|\frac{\sigma_t(\theta_0)}{\tilde{\sigma}_t(\theta)}\right|^{\delta} + \frac{C}{\sqrt{n}}\sum_{t=1}^{n} b_t. \end{aligned} \tag{A.6}$$

We will prove that

$$\sup_t E \sup_{\theta\in V(\theta_0)} \left\{\frac{\sigma_t(\theta_0)}{\tilde{\sigma}_t(\theta)}\right\}^{\delta s_1 s_2} < \infty. \tag{A.7}$$

A Taylor expansion gives, for $\sigma_t^*(\theta)$ between $\sigma_t(\theta_0)$ and $\tilde{\sigma}_t(\theta)$,

$$\begin{aligned} \left\{\frac{\sigma_t(\theta_0)}{\tilde{\sigma}_t(\theta)}\right\}^{2\delta s_2} &= \left\{\frac{\sigma_t(\theta_0)}{\sigma_t(\theta)}\right\}^{2\delta s_2} - 2\delta s_2 a_t\{\sigma_t(\theta_0)\}^{2\delta s_2}\left\{\frac{1}{\sigma_t^*(\theta)}\right\}^{2\delta s_2+1} \\ &\leq \left\{\frac{\sigma_t(\theta_0)}{\sigma_t(\theta)}\right\}^{2\delta s_2} + Ca_t\{\sigma_t(\theta_0)\}^{2\delta s_2} \end{aligned}$$

since $2\delta s_2 + 1 > 0$. Hence, by the $c_r$-inequality

$$\left\{\frac{\sigma_t(\theta_0)}{\tilde{\sigma}_t(\theta)}\right\}^{\delta s_2 s_1} \leq \left\{\frac{\sigma_t(\theta_0)}{\sigma_t(\theta)}\right\}^{\delta s_2 s_1} + Ca_t^{s_1/2}\{\sigma_t(\theta_0)\}^{\delta s_2 s_1}$$

The first term in the right-hand side admits a finite expectation using **A9** and $s_2 s_1 < 1$. The second term admits a finite expectation by the Cauchy-Schwarz inequality and **A8**. Hence (A.7) is proved.

We have $E|\eta_t|^{\delta s_1 s_2} < \infty$ because $s_1 s_2 \in (0,1)$. Therefore

$$E\left(\sum_{t=1}^{\infty} b_t^{s_1 s_2/2} |\eta_t|^{\delta s_1 s_2/2} \sup_{\theta \in V(\theta_0)} \left|\frac{\sigma_t(\theta_0)}{\tilde{\sigma}_t(\theta)}\right|^{\delta s_1 s_2/2}\right) < \infty$$

by **A8**, (A.7) and Cauchy-Schwarz's inequality, and thus the random variable inside the bracket is finite almost surely. It follows that

$$\begin{aligned} &\left(n^{-1/2} \sum_{t=1}^{n} b_t |\eta_t|^{\delta} \sup_{\theta \in V(\theta_0)} \left|\frac{\sigma_t(\theta_0)}{\tilde{\sigma}_t(\theta)}\right|^{\delta}\right)^{s_1 s_2/2} \\ \leq\ & n^{-s_1 s_2/4} \sum_{t=1}^{\infty} b_t^{s_1 s_2/2} |\eta_t|^{\delta s_1 s_2/2} \sup_{\theta \in V(\theta_0)} \left|\frac{\sigma_t(\theta_0)}{\tilde{\sigma}_t(\theta)}\right|^{\delta s_1 s_2/2} \to 0 \end{aligned}$$

which shows that the first term in the right-hand side of (A.6) goes to zero a.s. as $n$ tends to infinity. The second term is handled in a straightforward way. Thus the last term in (A.5) converges to zero a.s. as $n$ tends to infinity. Now note that

$$g_2(x,\sigma) := \frac{\partial g_1(x,\sigma)}{\partial \sigma} = \frac{1}{\sigma^2}\left[1 + \frac{x}{\sigma}\left\{2\frac{h'}{h} + \frac{x}{\sigma}\left(\frac{h'}{h}\right)'\right\}\left(\frac{x}{\sigma}\right)\right]. \tag{A.8}$$

The first term in the right-hand side of (A.5) is bounded by

$$\begin{aligned} &\frac{1}{\sqrt{n}} \sum_{t=1}^{n} |g_2(\epsilon_t, \sigma_t^*)||\Delta_t(\theta)| \left\|\frac{\partial \sigma_t(\theta)}{\partial \theta}\right\| \\ \leq\ & \frac{C}{\sqrt{n}} \sum_{t=1}^{n} c_t \left(1 + |\epsilon_t|^{\delta} \sup_{\theta \in V(\theta_0)} \left|\frac{1}{\sigma_t^*}\right|^{1+\delta}\right) \sup_{\theta \in V(\theta_0)} \left\|\frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta}\right\| \end{aligned} \tag{A.9}$$

where $\sigma_t^* = \sigma_t^*(\theta)$ is between $\tilde{\sigma}_t(\theta)$ and $\sigma_t(\theta)$. For $\delta > 0$ and $s \in (0, 4\delta)$ we have, by the $c_r$ and Cauchy-Schwarz inequalities

$$\begin{aligned} &E\left(\sum_{t=1}^{\infty} c_t (1 + |\epsilon_t|^{\delta}) \sup_{\theta \in V(\theta_0)} \left\|\frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta}\right\|\right)^{s/4\delta} \\ \leq\ & \sum_{t=1}^{\infty} \{E c_t^{s/2\delta}\}^{1/2} \{E(1 + |\epsilon_0|^{s})\}^{1/4} \left\{E\left(\sup_{\theta \in V(\theta_0)} \left\|\frac{1}{\sigma_0(\theta)} \frac{\partial \sigma_0(\theta)}{\partial \theta}\right\|\right)^{s/\delta}\right\}^{1/4} < \infty \end{aligned}$$

by **A10** and **A9**. For $\delta \in [-1, 0]$ and $s \in (0,1)$ we have, similarly,

$$\begin{aligned} &E\left(\sum_{t=1}^{\infty} c_t (1 + |\epsilon_t|^{\delta}) \sup_{\theta \in V(\theta_0)} \left\|\frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta}\right\|\right)^{s/4} \\ \leq\ & \sum_{t=1}^{\infty} \{E c_t^{s/2}\}^{1/2} \{E(1 + |\epsilon_0|^{s\delta})\}^{1/4} \left\{E\left(\sup_{\theta \in V(\theta_0)} \left\|\frac{1}{\sigma_0(\theta)} \frac{\partial \sigma_0(\theta)}{\partial \theta}\right\|\right)^{s}\right\}^{1/4} < \infty. \end{aligned}$$

The case $\delta < -1$ is treated in the same fashion, using an inequality similar to (A.3). By arguments already used, we conclude that the first term in the right-hand side of (A.5) goes to zero a.s. as $n$ tends to infinity. Thus *iv)* is established.

## A.2 Proof of Theorem 3.1

It will be sufficient to derive the advanced results for $r \neq 0$. The same arguments can be used for $r = 0$.

Because $E\eta_t^{*2} = 1$, the identifiability condition **A3**, with $\eta_0$ replaced by $\eta_0^*$, is satisfied when $h$ is the standard gaussian density. Note also that **A4** and **A7** hold with $\delta = 2$. It follows that, by Theorem 2.1,

$$\begin{aligned}\sqrt{n}\left(\hat{\theta}_n^* - \theta_0^*\right) &= -J_*^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(1 - \frac{\eta_t^2}{E\eta_t^2}\right)\frac{1}{\sigma_t^{*2}}\frac{\partial\sigma_t^2(\theta_0^*)}{\partial\theta} + o_P(1) \\ &\xrightarrow{\mathcal{L}} \mathcal{N}(0, (\kappa_4 - 1)J_*^{-1}). \end{aligned} \tag{A.10}$$

Let

$$\eta_t(\theta) = \epsilon_t\sigma_t^{-1}(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta), \quad \tilde{\eta}_t(\theta) = \epsilon_t\sigma_t^{-1}(\epsilon_{t-1}, \dots, \epsilon_1, \tilde{\epsilon}_0, \tilde{\epsilon}_{-1}, \dots; \theta),$$

$$\mu_r(\theta) = \frac{1}{n}\sum_{t=1}^{n}|\eta_t(\theta)|^r, \quad \text{for } r \neq 0, \quad \mu_0(\theta) = \frac{1}{n}\sum_{t=1}^{n}\log|\eta_t(\theta)|, \quad \text{for } r = 0.$$

We similarly define $\tilde{\mu}_r(\theta)$, obtained by replacing $\eta_t(\theta)$ by $\tilde{\eta}_t(\theta)$. By **A5**, it can be shown that

$$\hat{\mu}_r^* = \tilde{\mu}_r(\hat{\theta}_n^*) = \mu_r(\hat{\theta}_n^*) + o_P(n^{-1/2}).$$

By (A.10) and arguments similar to those used to prove (4.26)-(4.29) in Francq and Zakoïan (2004), a Taylor expansion gives

$$\mu_r(\hat{\theta}_n^*) = \mu_r(\theta_0^*) + \frac{\partial\mu_r(\theta_0^*)}{\partial\theta'}(\hat{\theta}_n^* - \theta_0^*) + o_P(n^{-1/2})$$

with

$$\frac{\partial\mu_r(\theta_0^*)}{\partial\theta'} = \frac{-r}{2n}\sum_{t=1}^{n}|\eta_t^*|^r\frac{1}{\sigma_t^{*2}}\frac{\partial\sigma_t^2(\theta_0^*)}{\partial\theta'} = \frac{-r}{2}E|\eta_t^*|^r\Omega_*' + o_P(1).$$

It follows that

$$\begin{aligned}\sqrt{n}(\hat{\mu}_r^* - \mu_r^*) &= \sqrt{n}\{\mu_r(\theta_0^*) - \mu_r^*\} - \frac{r}{2}E|\eta_t^*|^r\Omega_*'\sqrt{n}(\hat{\theta}_n^* - \theta_0^*) + o_P(1) \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(|\eta_t^*|^r - \mu_r^*) - \frac{r}{2}E|\eta_t^*|^r\Omega_*'\sqrt{n}(\hat{\theta}_n^* - \theta_0^*) + o_P(1) \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{(|\eta_t|^r - 1)}{\mu_2^{r/2}} - \frac{r}{2}\kappa_r\Omega_*'\sqrt{n}(\hat{\theta}_n^* - \theta_0^*) + o_P(1). \end{aligned} \tag{A.11}$$

Noting that $\mathrm{Cov}(|\eta_t|^r, \eta_t^2) = \mu_2^{1+r/2}(\kappa_{2+r} - \kappa_r)$, we have

$$\mathrm{Cov}\left(\sqrt{n}\left(\hat{\theta}_n^* - \theta_0^*\right), \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{(|\eta_t|^r - 1)}{\mu_2^{r/2}}\right) = (\kappa_{2+r} - \kappa_r)J_*^{-1}\Omega_* + o_P(1).$$

It follows from (A.11) that

$$\mathrm{Cov}\left(\sqrt{n}\left(\hat{\theta}_n^* - \theta_0^*\right), \sqrt{n}(\hat{\mu}_r^* - \mu_r^*)\right) = -\lambda_r J_*^{-1}\Omega_* + o_P(1). \tag{A.12}$$

We also have

$$\mathrm{Var}\left(\sqrt{n}(\hat{\mu}_r^* - \mu_r^*)\right) = \kappa_{2r} - \kappa_r^2 + \frac{r}{2}\kappa_r\{\lambda_r - (\kappa_{2+r} - \kappa_r)\}\Omega_*' J_*^{-1}\Omega_* + o_P(1).$$

Finally, the CLT for martingale differences and the Wold-Cràmer device entail (3.4), provided that (3.5) holds.

Now we prove (3.5). First note that $\lambda_2 = 0$. Because $\mu_2^* = 1$, the previous expansion writes, when $r = 2$,

$$\mathrm{Var}\left(\sqrt{n}(\hat{\mu}_2^* - 1)\right) = (\kappa_4 - 1)(1 - \Omega_*' J_*^{-1}\Omega_*) + o_P(1).$$

Note that by **B1**, for any $c > 0$, $c\tilde{\sigma}(\hat{\theta}_n^*) = \tilde{\sigma}(F(\hat{\theta}_n^*, c))$. Then the maximum of the function $c \mapsto \tilde{Q}_n(F(\hat{\theta}_n^*, c))$, where $\tilde{Q}_n$ is defined in (2.1) with $h = \phi$, is uniquely obtained for $c = \hat{\mu}_2^*$. Because $c = 1$ also yields a maximum, by definition of the QMLE, we must have $\hat{\mu}_2^* = 1, a.s.$ The conclusion follows. □

## A.3 Proof of Theorem 4.1

To prove (4.3) we note that Assumptions **A4** and **A7** are satisfied with $\delta = r$. Assumptions **A5** and **A8** are satisfied because the strict stationarity implies the existence of a moment of order $s$, for some $s > 0$ (see Berkes et al (2003), Lemma 2.3), and because $a_t$, $b_t$ and $c_t$ decrease at an exponential rate when $t$ goes to infinity. More precisely, $\max\{a_t, b_t, c_t\} \leq K\rho^t$ where $K$ is a random variable, measurable with respect to $\{\epsilon_u, u \leq 0\}$, and $\rho \in (0, 1)$ is a constant (see Francq and Zakoïan (2004), Equations (4.6) and (4.33)). The latter paper also established the second part of **A2** and **A9**. The conclusion follows from Theorem 2.1.

The expression of the two-step estimator $\tilde{\theta}_n$ follows from (3.1). The convergence in distribution (4.4) follows from Theorem 3.1. Let $\Gamma_r$ denote the asymptotic variance in (4.4). To derive an explicit expression for $\Gamma_r$ we use (4.1) and the following calculations. Denote by $L$ the lag operator. The derivatives of $\sigma_t^2(\theta)$ verify

$$\begin{aligned}
\mathcal{B}_\theta(L)\frac{\partial\sigma_t^2}{\partial\omega}(\theta) &= 1, \qquad \mathcal{B}_\theta(L)\frac{\partial\sigma_t^2}{\partial\alpha_i}(\theta) = \epsilon_{t-i}^2, \quad i = 1, \ldots, q, \\
\mathcal{B}_\theta(L)\frac{\partial\sigma_t^2}{\partial\beta_j}(\theta) &= \sigma_{t-j}^2(\theta), \quad j = 1, \ldots, p.
\end{aligned} \tag{A.13}$$

In view of (3.1), $\mathcal{B}_{\theta_0}(L) = \mathcal{B}_{\theta_0^*}(L)$. Moreover $\sigma_{t-j}^2(\theta_0^*) = \mu_2\sigma_{t-j}^2(\theta_0)$. Thus

$$\frac{\partial\sigma_t^2(\theta_0^*)}{\partial\theta} = A\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}, \qquad A = \begin{pmatrix} I_{q+1} & 0 \\ 0 & \mu_2 I_p \end{pmatrix}. \tag{A.14}$$

It follows that

$$J_* = \mu_2^{-2} AJA, \qquad \Omega_* = \mu_2^{-1} A\Omega, \tag{A.15}$$

where $\Omega = E\left(\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\right)$. Hence, the asymptotic variance of Theorem 3.1 is given by

$$\Sigma_r = \begin{pmatrix} (\kappa_4-1)\mu_2^2 A^{-1}J^{-1}A^{-1} & -\lambda_r\mu_2 A^{-1}J^{-1}\Omega \\ -\lambda_r\mu_2\Omega' J^{-1}A^{-1} & \sigma^2_{\mu_r^*} \end{pmatrix}$$

Moreover, in view of $G_r(\theta_0^{*'},\mu_r^*) = \left((\mu_r^*)^{2/r}\omega_0^*,\dots,(\mu_r^*)^{2/r}\alpha_{0q}^*,\beta_{01}^*,\dots,\beta_{0p}^*\right)'$ we have

$$\left[\frac{\partial G_r(\theta_0^{*'},\mu_r^*)}{\partial(\theta',\mu)}\right] = \left[\frac{1}{\mu_2}A \quad \frac{2}{r}\mu_2^{\frac{r}{2}}\overline{\theta}_0\right].$$

Hence the asymptotic variance of the reparameterized QMLE of the two-step approach

$$\Gamma_r = (\kappa_4-1)J^{-1} - \lambda_r\frac{2}{r}\mu_2^{\frac{r}{2}}\left(\overline{\theta}_0\Omega' J^{-1} + J^{-1}\Omega\overline{\theta}_0'\right) + \sigma^2_{\mu_r^*}\left(\frac{2}{r}\mu_2^{\frac{r}{2}}\right)^2\overline{\theta}_0\overline{\theta}_0'.$$

Now we will show that

$$J^{-1}\Omega = \overline{\theta}_0, \qquad \Omega' J^{-1}\Omega = 1 \tag{A.16}$$

The second equality follows from (A.15) and (3.5) but we give a direct proof. In view of (A.13), we have

$$\mathcal{B}_\theta(L)\frac{\partial\sigma_t^2(\theta)}{\partial\theta^{[1:q+1]'}}\,\theta^{[1:q+1]} = \omega + \sum_{i=1}^{q}\alpha_i\epsilon_{t-i}^2 = \mathcal{B}_\theta(L)\sigma_t^2(\theta),$$

Because, by assumption **C** and the positivity of the $\beta_j$, the roots of the polynomial $\mathcal{B}_\theta(L)$ are outside the unit circle, it follows that

$$\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta^{[1:q+1]'}}\,\theta_0^{[1:q+1]} = \frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}\,\overline{\theta}_0 = \sigma_t^2(\theta_0), \tag{A.17}$$

The first equality in (A.16) follows. We also have $\Omega'\overline{\theta}_0 = 1$. The second equality in (A.16) follows. Because $\mu_2^{r/2} = 1/\kappa_r$, we thus have, by (A.16)

$$\begin{aligned}\Gamma_r &= (\kappa_4-1)J^{-1} + \left[\sigma^2_{\mu_r^*}\left(\frac{2}{r\kappa_r}\right)^2 - \frac{4}{r\kappa_r}\lambda_r\right]\overline{\theta}_0\overline{\theta}_0' \\ &= (\kappa_4-1)J^{-1} + \left(\frac{2}{r\kappa_r}\right)^2\left[\kappa_{2r}-\kappa_r^2+\frac{r}{2}\kappa_r(\lambda_r-\kappa_{2+r}+\kappa_r) - r\kappa_r\lambda_r\right]\overline{\theta}_0\overline{\theta}_0' \\ &= (\kappa_4-1)J^{-1} + \left(\frac{2}{r\kappa_r}\right)^2\left[\kappa_{2r}-\kappa_r^2-\frac{r}{2}\kappa_r\left\{\frac{r}{2}\kappa_r(\kappa_4-1)\right\}\right]\overline{\theta}_0\overline{\theta}_0',\end{aligned}$$

which completes the proof of (4.4). The theorem is established. □

# B Complementary proofs

## B.1 Complementary results for the proof of Theorem 2.1

For the consistency, it remains to show *ii)* and *iii)*, and for the asymptotic normality it remains to show *v)* and *vi)*.

To prove *ii)*, it suffices to use **A2-A3** and

$$g(\epsilon_t, \sigma_t(\theta)) = g\left(\eta_t, \frac{\sigma_t(\theta)}{\sigma_t(\theta_0)}\right) - \log \sigma_t(\theta_0).$$

Indeed, we have

$$\mathbb{E}\{g(\epsilon_1, \sigma_1(\theta)) - g(\epsilon_1, \sigma_1(\theta_0))\} \quad = \quad \mathbb{E}\left\{g\left(\eta_t, \frac{\sigma_t(\theta)}{\sigma_t(\theta_0)}\right) - g(\eta_t, 1)\right\} \leq 0,$$

with equality if and only if $\theta = \theta_0$.

Now we will show *iii)*. For any $\theta \in \Theta$ and any positive integer $k$, let $V_k(\theta)$ be the open ball with center $\theta$ and radius $1/k$. We have,

$$\begin{aligned} & \limsup_{n\to\infty} \sup_{\theta^* \in V_k(\theta)\cap\Theta} \tilde{Q}_n(\theta^*) \\ \leq\ & \limsup_{n\to\infty} \sup_{\theta^* \in V_k(\theta)\cap\Theta} Q_n(\theta^*) + \limsup_{n\to\infty} \sup_{\theta\in\Theta} |Q_n(\theta) - \tilde{Q}_n(\theta)| \\ \leq\ & \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} \sup_{\theta^* \in V_k(\theta)\cap\Theta} g(\epsilon_t, \sigma_t(\theta^*)) \quad a.s. \end{aligned}$$

where the second inequality comes from *i)*. Note that since $h$ is integrable and differentiable, $h$ is bounded. It follows, by **A2**, that

$$\mathbb{E} \sup_{\theta^* \in V_k(\theta)\cap\Theta} g(\epsilon_t, \sigma_t(\theta^*)) < \log \frac{1}{\underline{\omega}} + C < \infty. \tag{B.1}$$

Using an ergodic theorem for stationary and ergodic processes $(X_t)$ such that $\mathbb{E}(X_t)$ exists in $\mathbb{R} \cup \{-\infty, +\infty\}$ (see Billingsley, 1995, p. 284 and 495), it follows that

$$\limsup_{n\to\infty} \sup_{\theta^* \in V_k(\theta)\cap\Theta} \tilde{Q}_n(\theta^*) \leq \mathbb{E}X_{t,k}(\theta), \qquad X_{t,k}(\theta) = \sup_{\theta^* \in V_k(\theta)\cap\Theta} g(\epsilon_t, \sigma_t(\theta^*))\ .$$

When $k$ tends to infinity, the sequence $\{X_{t,k}(\theta)\}_k$ decreases to $X_t(\theta) = g(\epsilon_t, \sigma_t(\theta))$. Thus $\{X^-_{t,k}(\theta)\}_k$ increases to $X^-_t(\theta)$. By the Beppo-Levi theorem, $\mathbb{E}X^-_{t,k}(\theta) \uparrow \mathbb{E}_{\theta_0} X^-_t(\theta)$ when $k \uparrow +\infty$. By (B.1), the fact that the sequence $\{X^+_{t,k}(\theta)\}_k$ is decreasing, and the Lebesgue theorem, $\mathbb{E}X^+_{t,k}(\theta) \downarrow \mathbb{E}X^+_t(\theta)$ when $k \uparrow +\infty$. Thus we have shown that $\mathbb{E}X_{t,k}$ converges to $\mathbb{E}\{X_t(\theta)\}$ when $k \to \infty$. By *ii)*, *iii)* is proved.

As in the proof of Theorem 2.1 in Francq and Zakoïan (2004), the consistency is a consequence of a standard compactness argument and of the intermediate results **i-iii**.

Now we establish *v)*. In view of **A4** and **A7**, we have

$$\begin{aligned}\left\|\frac{\partial^2 Q_n(\theta)}{\partial\theta\partial\theta'}\right\| &= \left\|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 g(\epsilon_t,\sigma_t(\theta))}{\partial\theta\partial\theta'}\right\| \\ &= \left\|\frac{1}{n}\sum_{t=1}^{n} g_2(\epsilon_t,\sigma_t(\theta))\frac{\partial\sigma_t(\theta)}{\partial\theta}\frac{\partial\sigma_t(\theta)}{\partial\theta'} + g_1(\epsilon_t,\sigma_t(\theta))\frac{\partial^2\sigma_t(\theta)}{\partial\theta\partial\theta'}\right\| \\ &\leq \frac{C}{n}\sum_{t=1}^{n}\left(1+\left|\frac{\sigma_t(\theta_0)\eta_t}{\sigma_t(\theta)}\right|^{\delta}\right)\left(\left\|\frac{1}{\sigma_t(\theta)}\frac{\partial^2\sigma_t(\theta)}{\partial\theta\partial\theta'}\right\|\right. \\ &\qquad\left. + \left\|\frac{1}{\sigma_t^2(\theta)}\frac{\partial\sigma_t(\theta)}{\partial\theta}\frac{\partial\sigma_t(\theta)}{\partial\theta'}\right\|\right).\end{aligned}$$

Hence

$$E\sup_{\theta\in V(\theta_0)}\left\|\frac{\partial^2 Q_n(\theta)}{\partial\theta\partial\theta'}\right\| \leq C$$

by the Hölder inequality, **A7** and **A9**. The ergodic theorem then implies that

$$\begin{aligned}&\lim_{n\to\infty}\sup_{\theta\in V(\theta_0)}\left\|\frac{\partial^2 Q_n(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 Q_n(\theta_0)}{\partial\theta\partial\theta'}\right\| \\ \leq\ & E\sup_{\theta\in V(\theta_0)}\left\|\frac{\partial^2 g(\epsilon_t,\sigma_t(\theta))}{\partial\theta\partial\theta'} - \frac{\partial^2 g(\epsilon_t,\sigma_t(\theta_0))}{\partial\theta\partial\theta'}\right\|, \quad a.s.\end{aligned}$$

By the dominated convergence theorem, the last expectation tends to zero when the neighborhood $V(\theta_0)$ tends to the singleton $\{\theta_0\}$. The consistency of $\hat{\theta}_{n,h}$ thus entails

$$\lim_{n\to\infty}\left|\frac{\partial^2 Q_n(\theta^*)}{\partial\theta\partial\theta'} - \frac{\partial^2 Q_n(\theta_0)}{\partial\theta\partial\theta'}\right| = 0, \quad a.s.$$

In view of (2.3),

$$Eg_1(\epsilon_t,\sigma_t(\theta_0))\frac{\partial^2\sigma_t(\theta_0)}{\partial\theta\partial\theta'} = 0$$

and by (A.8), $g_2(\epsilon_t,\sigma_t(\theta_0)) = g_2(\eta_t,1)\sigma_t^{-2}(\theta_0)$. By the ergodic theorem

$$\lim_{n\to\infty}\frac{\partial^2 Q_n(\theta_0)}{\partial\theta\partial\theta'} = \frac{Eg_2(\eta_t,1)}{4}J, \quad a.s.$$

and *v)* is established.

To prove *vi)* it suffices to note that

$$\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_0) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} g_1(\eta_t,1)\frac{1}{2\sigma_t^2(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}$$

and to apply a CLT for square integrable stationary martingale differences (see Billingsley (1961)).

Now, from **A6** and the consistency of $\hat{\theta}_{n,h}$, a Taylor expansion yields

$$\begin{aligned} 0 &= \sqrt{n}\frac{\partial}{\partial\theta}Q_n(\hat{\theta}_{n,h}) + \sqrt{n}\frac{\partial}{\partial\theta}\tilde{Q}_n(\hat{\theta}_{n,h}) - \sqrt{n}\frac{\partial}{\partial\theta}Q_n(\hat{\theta}_{n,h}) \\ &= \sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_0) + \frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\theta^*)\sqrt{n}(\hat{\theta}_{n,h} - \theta_0) \\ &\quad + \sqrt{n}\left(\frac{\partial}{\partial\theta}\tilde{Q}_n(\hat{\theta}_{n,h}) - \frac{\partial}{\partial\theta}Q_n(\hat{\theta}_{n,h})\right), \end{aligned}$$

where $\theta^*$ is between $\hat{\theta}_{n,h}$ and $\theta_0$. Applying *iv), v), vi)*, the proof of the asymptotic normality is complete.

## B.2 Proof of Theorem 4.2

We note that (4.5) does not straightforwardly follow from Theorem 2.1 because Assumptions **A4** and **A7** are not satisfied when $r = 0$ and $h \in \mathcal{C}(0)$. However, tedious computation shows that the conclusion of Theorem 2.1 continues to hold under the assumptions of Theorem 4.2.

To prove (4.6), observe that

$$\left[\frac{\partial G_0(\theta_0^{*'}, \mu_0^*)}{\partial(\theta', \mu)}\right] = \left[\frac{1}{\mu_2}A \quad 2\overline{\theta}_0\right].$$

The conclusion follows along the same lines as in the proof of Theorem 4.1. □

## B.3 Proof of Theorem 4.3

To prove the AN, we have already seen in the proof of Theorem 4.1 that Assumptions **A4** and **A7** are satisfied with $\delta = r$. Assumptions **A5** and **A8** are satisfied by the same arguments as in Theorem 4.1 and using Pan, Wang and Tong (2008), and Hamadeh and Zakoïan (2009). The latter paper also established the second part of **A2** and **A9**. The AN follows from Theorem 2.1.

Because $G_r(\theta_0^{*'}, \mu_r^*) = \left((\mu_r^*)^{\delta/r}\omega_0^*, \dots, (\mu_r^*)^{\delta/r}\alpha_{0q-}^*, \beta_{01}^*, \dots, \beta_{0p}^*\right)'$ we have

$$\left[\frac{\partial G_r(\theta_0^{*'}, \mu_r^*)}{\partial(\theta', \mu)}\right] = \left[\mu_2^{-\frac{\delta}{2}}A_\delta \quad \frac{\delta}{r}\mu_2^{\frac{r}{2}}\overline{\theta}_0\right], \quad A_\delta = \begin{pmatrix} I_{2q+1} & 0 \\ 0 & \mu_2^{\frac{\delta}{2}}I_p \end{pmatrix}.$$

Similarly to (A.13), the derivatives of $\sigma_t^\delta(\theta)$ verify

$$\begin{aligned} \mathcal{B}_\theta(L)\frac{\partial\sigma_t^\delta}{\partial\omega}(\theta) &= 1, \\ \mathcal{B}_\theta(L)\frac{\partial\sigma_t^\delta}{\partial\alpha_{i+}}(\theta) &= (\epsilon_{t-i}^+)^\delta, \qquad \mathcal{B}_\theta(L)\frac{\partial\sigma_t^\delta}{\partial\alpha_{i-}}(\theta) = (-\epsilon_{t-i}^-)^\delta, \quad i = 1, \dots, q, \\ \mathcal{B}_\theta(L)\frac{\partial\sigma_t^\delta}{\partial\beta_j}(\theta) &= \sigma_{t-j}^\delta, \quad j = 1, \dots, p. \end{aligned}$$

It follows that, similarly to (A.16)

$$J_\delta^{-1}\Omega_\delta = \overline{\theta}_0, \qquad \Omega_\delta' J_\delta^{-1}\Omega_\delta = 1 \tag{B.2}$$

where

$$J_\delta = E\left(\frac{1}{\sigma_t^{2\delta}}\frac{\partial\sigma_t^\delta}{\partial\theta}\frac{\partial\sigma_t^\delta}{\partial\theta'}(\theta_0)\right) = \left(\frac{\delta}{2}\right)^2 J, \quad \Omega_\delta = E\left(\frac{1}{\sigma_t^\delta}\frac{\partial\sigma_t^\delta}{\partial\theta}(\theta_0)\right) = \frac{\delta}{2}\Omega.$$

Thus

$$J^{-1}\Omega = \frac{\delta}{2}\overline{\theta}_0, \qquad \Omega' J^{-1}\Omega = 1 \tag{B.3}$$

Moreover, similarly to (A.14), we have

$$\frac{\partial\sigma_t^2(\theta_0^*)}{\partial\theta} = \mu_2^{1-\frac{\delta}{2}} A_\delta \frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}. \tag{B.4}$$

It follows that, similar to (A.15),

$$J_* = \mu_2^{-\delta} A_\delta J A_\delta, \qquad \Omega_* = \mu_2^{-\delta/2} A_\delta \Omega. \tag{B.5}$$

Hence, the asymptotic variance of Theorem 3.1 is given by

$$\Sigma_r = \begin{pmatrix} (\kappa_4 - 1)\mu_2^\delta A_\delta^{-1} J^{-1} A_\delta^{-1} & -\lambda_r \mu_2^{\delta/2} A_\delta^{-1} J^{-1}\Omega \\ -\lambda_r \mu_2^{\delta/2}\Omega' J^{-1} A_\delta^{-1} & \sigma_{\mu_r^*}^2 \end{pmatrix}$$

Therefore, the asymptotic variance of the reparameterized QMLE of the two-step approach

$$\begin{aligned}\Gamma_r &= \begin{bmatrix}\mu_2^{-\frac{\delta}{2}} A_\delta & \frac{\delta}{r}\mu_2^{\frac{r}{2}}\overline{\theta}_0\end{bmatrix} \Sigma_r \begin{bmatrix}\mu_2^{-\frac{\delta}{2}} A_\delta' \\ \frac{\delta}{r}\mu_2^{\frac{r}{2}}\overline{\theta}_0'\end{bmatrix} \\ &= (\kappa_4 - 1)J^{-1} - \lambda_r \frac{\delta}{r}\mu_2^{\frac{r}{2}}\left(\overline{\theta}_0\Omega' J^{-1} + J^{-1}\Omega\overline{\theta}_0'\right) + \sigma_{\mu_r^*}^2\left(\frac{\delta}{r}\mu_2^{\frac{r}{2}}\right)^2 \overline{\theta}_0\overline{\theta}_0'.\end{aligned}$$

In view of (B.3), the asymptotic variance follows.

Finally, the conclusion of Corollary 4.1 holds true for Model (4.10), since

$$\begin{aligned}&\mathrm{Var}_{as}\left\{\sqrt{n}\left(\hat{\theta}_{n,h} - \theta_0\right)\right\} - \mathrm{Var}_{as}\left\{\sqrt{n}\left(\tilde{\theta}_n - \theta_0\right)\right\} \\ = &\left[\left(\frac{2}{r}\right)^2\left(\frac{\kappa_{2r}}{\kappa_r^2} - 1\right) - (\kappa_4 - 1)\right]\left(J^{-1} - \left(\frac{\delta}{2}\right)^2 \overline{\theta}_0\overline{\theta}_0'\right)\end{aligned}$$

and

$$J^{-1} \succeq \left(\frac{\delta}{2}\right)^2 \overline{\theta}_0\overline{\theta}_0'.$$

□

## B.4 Proof of Theorem 4.4

We verify the conditions of Theorem 2.1. The first condition in $\mathbf{E}(s)$ entails **A0** (see Theorem 1 in Douc et al. (2008)). The first part of **A2** is already in $\mathbf{E}(s)$. To show the second part, note that if

$$\sigma_t^2(\theta_0) - \sigma_t^2(\theta) = \psi_0(\theta_0) - \psi_0(\theta) + \sum_{i=1}^{\infty} \{\psi_i(\theta_0) - \psi_i(\theta)\} \epsilon_{t-i}^2 = 0 \qquad a.s.$$

and if $\psi_i(\theta_0) \neq \psi_i(\theta)$ for some $i \geq 1$, then for $i_0 \geq 1$, $\epsilon_{t-i_0}^2$ can be written as a linear combination of the $\epsilon_{t-i}^2$'s for $i > i_0$. In that case, $\eta_{t-i_0}^2$ would be measurable with respect to a $\sigma$-field independent of $\eta_{t-i_0}^2$. This is impossible because the distribution of $\eta_{t-i_0}^2$ is assumed to be nondegenerate. Thus **A2** is shown. Because $h \in \mathcal{C}(r)$, Assumption **A3** holds true, and Assumption **A4** is satisfied with $\delta = r$.

Note that Assumption **A5** is only used to show the point $i)$ in the proof of Theorem 2.1. We therefore directly prove $i)$ by showing that the right-hand side of (A.1) tends to zero $a.s.$ For simplicity, the proof is written with the initial values $\tilde{\epsilon}_i = 0$ for $i \leq 0$. We have

$$\begin{aligned} a_t &= \sup_{\theta\in\Theta} \left| \frac{\tilde{\sigma}_t^2(\theta) - \sigma_t^2(\theta)}{\tilde{\sigma}_t(\theta) + \sigma_t(\theta)} \right| \leq C \sum_{i=t}^{\infty} \sup_{\theta\in\Theta} \psi_i(\theta) \epsilon_{t-i}^2 \leq C\overline{a}_t, \\ \overline{a}_t &= \sum_{i=t}^{\infty} \frac{1}{i^{\underline{d}+1}} \epsilon_{t-i}^2 = \sum_{i=0}^{\infty} \frac{1}{(t+i)^{\underline{d}+1}} \epsilon_{-i}^2. \end{aligned}$$

Note that $\overline{a}_t$ is $a.s.$ finite because

$$E\left(\overline{a}_t\right)^s \leq \sum_{i=0}^{\infty} \frac{1}{(t+i)^{(\underline{d}+1)s}} E\left|\epsilon_{-i}\right|^{2s} \leq \frac{C}{t^{(\underline{d}+1)s-1}}. \tag{B.6}$$

Moreover, the dominated convergence theorem entails that the decreasing sequence $\overline{a}_t$, and thus $a_t$, converge to zero $a.s.$ as $t \to \infty$. By Cesàro's lemma, it follows that the second term in (A.1) tends to zero $a.s.$ The first term tends to zero $a.s.$ by Toeplitz's lemma because $\sup_{\theta\in\Theta} |\sigma_t^*|^{-1-r} < C$, $a_t \to 0$ $a.s.$ as $t \to \infty$ and $n^{-1} \sum_{t=1}^n |\epsilon_t|^r \to E|\epsilon_0|^r < \infty$ $a.s.$ as $n \to \infty$. For the last inequality, we note that $E\epsilon_1^{2s} < \infty$ under $\mathbf{E}(s)$ (see Robinson and Zaffaroni (2006) Page 1062 and Douc et al. (2008) Theorem 1). The proof of $i)$ follows. Thus, in view of Theorem 2.1 and its proof, the consistency of the one-step estimator is shown.

Turning to the asymptotic normality, note that Assumption **A6** is satisfied and **A7** is satisfied with $\delta = r$. In order to show **A9**, we first establish an inequality similar to that given in Robinson and Zaffaroni (2006) Page 1067. To lighten the notation, write $\psi_i = \psi_i(\theta)$ and $\psi_{0i} = \psi_i(\theta_0)$. We also use the convention $\psi_i^k = 0$ for all $k \in \mathbb{R}$ when $\psi_i = 0$. By Hölder's inequality, for all $k > s^*$ and all $s^* \in (0, 1]$ we have

$$\begin{aligned} \sigma_t^2(\theta_0) &= \psi_{00}\psi_0^{s^*/k-1} \times \psi_0^{1-s^*/k} + \sum_{i=1}^{\infty} \psi_{0i}\psi_i^{s^*/k-1} \epsilon_{t-i}^{2s^*/k} \times \psi_i^{1-s^*/k} \epsilon_{t-i}^{2(1-s^*/k)} \\ &\leq \left\{ \psi_{00}^{k/s^*} \psi_0^{1-k/s^*} + \sum_{i=1}^{\infty} \psi_{0i}^{k/s^*} \psi_i^{1-k/s^*} \epsilon_{t-i}^2 \right\}^{s^*/k} \left\{\sigma_t^2(\theta)\right\}^{1-s^*/k}. \end{aligned} \tag{B.7}$$

Since $\left\{\sigma_t^2(\theta)\right\}^{-s^*/k} \leq C$, we obtain

$$\left\{\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)}\right\}^k \leq C\left(1+\sum_{i=1}^{\infty}\psi_{0i}^k\psi_i^{s^*-k}\epsilon_{t-i}^{2s^*}\right).$$

Thus, for all $\iota > 0$ there exists a neighborhood $V(\theta_0)$ of $\theta_0$ such that for all $\theta \in V(\theta_0)$

$$\left\{\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)}\right\}^k \leq C\left(1+\sum_{i=1}^{\infty}\psi_i^{s^*}\psi_{0i}^{-k\iota}\epsilon_{t-i}^{2s^*}\right) \leq C\left(1+\sum_{i=1}^{\infty}\frac{1}{i^{(\underline{d}+1)s^*-k\iota}}\epsilon_{t-i}^{2s^*}\right)$$

and

$$\left\{\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)}\right\}^k \leq C\left(1+\sum_{i=1}^{\infty}\frac{1}{i^{(\underline{d}+1)s^*-k\iota}}\epsilon_{t-i}^{2s^*}\right).$$

The same inequality holds when the left-hand side is replaced by its inverse. Since $(\underline{d}+1)s > 1$ and $E|\epsilon_1^{2s}| < \infty$, it follows that for some neighborhood $V(\theta_0)$

$$E\sup_{\theta\in V(\theta_0)}\left\{\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)}\right\}^k < \infty, \qquad \forall k \in \mathbb{R}. \tag{B.8}$$

The last result of **A9** follows. Similarly to (B.7), for $k > s^*$ and $s^* \in (0,1]$ we have

$$\left|\frac{\partial\sigma_t^2(\theta)}{\partial\theta_j}\right| \leq \left(\left|\frac{\partial\psi_0}{\partial\theta_j}\right|^{k/s^*}\psi_0^{1-k/s^*}+\sum_{i=1}^{\infty}\left|\frac{\partial\psi_i}{\partial\theta_j}\right|^{k/s^*}\psi_i^{1-k/s^*}\epsilon_{t-i}^2\right)^{s^*/k}\left\{\sigma_t^2(\theta)\right\}^{1-s^*/k}.$$

Thus we have

$$\left|\frac{1}{\sigma_t(\theta)}\frac{\partial\sigma_t(\theta)}{\partial\theta_j}\right|^k \leq C\left(1+\sum_{i=1}^{\infty}\left|\frac{\partial\psi_i}{\partial\theta_j}\right|^k\psi_i^{s^*-k}\epsilon_{t-i}^{2s^*}\right) \leq C\left(1+\sum_{i=1}^{\infty}\frac{1}{i^{(\underline{d}+1)s^*-k\iota}}\epsilon_{t-i}^{2s^*}\right)$$

for all $\iota > 0$, all $k > s^*$ and all $s^* \in (0,1]$. Choosing $s^* = s$ and $\iota$ such that $(\underline{d}+1)s - k\iota > 1$ in the last sum, it follows that

$$E\sup_{\theta\in\Theta}\left\|\frac{1}{\sigma_t(\theta)}\frac{\partial\sigma_t(\theta)}{\partial\theta}\right\|_k < \infty, \qquad \forall k > 1. \tag{B.9}$$

The same result holds when the first-order derivatives are replaced by second-order derivatives, which completes the proof of **A9**.

Note that Assumptions **A8** and **A10** are only used to show the point *iv*) in the proof of Theorem 2.1. We therefore directly prove *iv*) by showing that the right-hand sides of

(A.6) and (A.9) tend to zero in probability. We have

$$\begin{aligned}
b_t &= \sup_{\theta \in V(\theta_0)} \left\| \frac{1}{2\sigma_t(\theta)} \frac{\partial \sigma_t^2(\theta)}{\partial \theta} - \frac{1}{2\tilde{\sigma}_t(\theta)} \frac{\partial \tilde{\sigma}_t^2(\theta)}{\partial \theta} \right\| \\
&\leq \sup_{\theta \in V(\theta_0)} |\sigma_t(\theta) - \tilde{\sigma}_t(\theta)| \sup_{\theta \in V(\theta_0)} \frac{\sigma_t(\theta)}{\tilde{\sigma}_t(\theta)} \sup_{\theta \in V(\theta_0)} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\| \\
&\quad + C \sup_{\theta \in V(\theta_0)} \left\| \frac{\partial \sigma_t^2(\theta)}{\partial \theta} - \frac{\partial \tilde{\sigma}_t^2(\theta)}{\partial \theta} \right\| \\
&\leq C b_t^{(0)} (b_t^{(0)} + 1) \sup_{\theta \in V(\theta_0)} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\| + C b_t^{(\iota)}, \\
b_t^{(\iota)} &= \sum_{i=0}^{\infty} \frac{1}{(t+i)^{(d_0+1)(1-\iota)}} \epsilon_{-i}^2
\end{aligned}$$

and

$$\begin{aligned}
\sup_{\theta \in V(\theta_0)} \left| \frac{\epsilon_t}{\tilde{\sigma}_t(\theta)} \right|^r &= \sup_{\theta \in V(\theta_0)} \left| \frac{\epsilon_t}{\sigma_t(\theta)} \right|^r \left| 1 + \frac{\sigma_t(\theta) - \tilde{\sigma}_t(\theta)}{\tilde{\sigma}_t(\theta)} \right|^r \\
&\leq 2^r (1 + c_t^r) \sup_{\theta \in V(\theta_0)} \left| \frac{\epsilon_t}{\sigma_t(\theta)} \right|^r \leq C \sup_{\theta \in V(\theta_0)} \left| \frac{\epsilon_t}{\sigma_t(\theta)} \right|^r,
\end{aligned}$$

where $C$ does not vary with $t$ and only depends on $\{\epsilon_u, u \leq 0\}$, since $c_t \leq c_0$. For the convergence of the right-hand side of (A.6) and (A.9), it therefore suffices that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} b_t^{(0)} \sup_{\theta \in V(\theta_0)} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\| \xrightarrow{P} 0, \tag{B.10}$$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} b_t^{(\iota)} \xrightarrow{P} 0, \tag{B.11}$$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} b_t^{(0)} \sup_{\theta \in V(\theta_0)} \left| \frac{\epsilon_t}{\sigma_t(\theta)} \right|^r \sup_{\theta \in V(\theta_0)} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\| \xrightarrow{P} 0, \tag{B.12}$$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} b_t^{(\iota)} \sup_{\theta \in V(\theta_0)} \left| \frac{\epsilon_t}{\sigma_t(\theta)} \right|^r \xrightarrow{P} 0. \tag{B.13}$$

The expectation of the left-hand side of (B.12) to the power $s^* \in (1/(d_0+1), s)$ is bounded by

$$\begin{aligned}
&\frac{1}{n^{s^*/2}} \sum_{t=1}^{n} E|\eta_t|^{rs^*} E \left\{ (b_t^{(0)})^{s^*} \sup_{\theta \in V(\theta_0)} \left| \frac{\sigma_t(\theta_0)}{\sigma_t(\theta)} \right|^{rs^*} \sup_{\theta \in V(\theta_0)} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\|^{s^*} \right\} \\
\leq\ &\frac{C}{n^{s^*/2}} \sum_{t=1}^{n} \|(b_t^{(0)})^{s^*}\|_{p_0} \left\| \sup_{\theta \in V(\theta_0)} \left| \frac{\sigma_t(\theta_0)}{\sigma_t(\theta)} \right|^{rs^*} \sup_{\theta \in V(\theta_0)} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\|^{s^*} \right\|_{q_0} \\
\leq\ &\frac{C}{n^{s^*/2}} \sum_{t=1}^{n} \|(b_t^{(0)})^{s^*}\|_{p_0} \leq \frac{C}{n^{s^*/2}} \sum_{t=1}^{n} \frac{1}{t^{(d_0+1)s^*-1}}
\end{aligned}$$

where $p_0^{-1} + q_0^{-1} = 1$, $p_0 > 1$ such that $E|\epsilon_t|^{2s^* p_0} < \infty$, by Hölder's inequality, (B.8), (B.9) and (B.6). If $(d_0 + 1)s > 2$, $s^*$ can be chosen sufficiently close to $s$ such that $\sum_{t=1}^n t^{-(d_0+1)s^*+1} < \infty$. If $(d_0 + 1)s \leq 2$, we have $n^{-s^*/2} \sum_{t=1}^n t^{-(d_0+1)s^*+1} \leq Cn^{-s^*/2+2-(d_0+1)s^*}$, which tends to zero because $s^*(d_0 + 3/2) > 2$ for $s^*$ sufficiently close to $s$. The convergence in (B.12) is thus established. The convergences in (B.10), (B.11) and (B.13) are obtained by the same arguments. Having shown iv), the asymptotic normality of the one-step estimator then follows from Theorem 2.1. For the last rest on the two-step estimator, it suffices to show that (3.4) holds true by a straightforward adaptation of the proof of Theorem 3.1. □

# References


**Andersen, T.G. and T. Bollerslev** (1998) Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39, 885–906.

**Baillie, R. and T. Bollerslev** (1992) Prediction in dynamic models with time-dependent conditional variance. *Journal of Econometrics* 52, 91–113.

**Bardet, J-M. and O. Wintenberger** (2009) Asymptotic normality of the Quasi-maximum likelihood estimator for multidimensional causal processes. *The Annals of Statistics* 37, 2730–2759.

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

**Berkes, I., Horváth, L. and P. Kokoszka** (2003) GARCH processes: structure and estimation. *Bernoulli* 9, 201–227.

**Billingsley, P.** (1961) The Lindeberg-Levy theorem for martingales. *Proceedings of the American Mathematical Society* 12, 788–792.

**Billingsley, P.** (1995) *Probability and Measure.* John Wiley & Sons, New York.

**Bollerslev, T.** (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.

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

**Ding, Z., Granger, C. and R.F. Engle** (1993) A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83–106.

**Douc, R., Roueff, F. and P. Soulier** (2008) On the existence of some ARCH($\infty$) processes. *Stochastic Processes and their Applications* 118, 755–761.

**Engle, R.F.** (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica* 50, 987–1007.

**Engle, R.F. and T. Bollerslev** (1986) Modeling the persistence of conditional variances. *Econometric Reviews* 5, 1–50.

**Engle, R.F. and D.F. Kraft** (1983) Multiperiod forecast error variances of inflation estimated from ARCH models. *Applied Time Series Analysis of Economic Data.* Washington, DC: Bureau of the Census.

**Engle, R.F. and J.R. Russell** (1998) Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. *Econometrica* 66, 1127–1162.

**Escanciano, J.C.** (2009) Quasi-maximum likelihood estimation of semi-strong GARCH models. *Econometric Theory* 25, 561–570.

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

**Giraitis, L., Kokoszka, P. and R. Leipus** (2000) Stationary ARCH models: dependence structure and central limit theorem. *Econometric Theory* 16, 3–22.

**Giraitis, L., Leipus, R., and D. Surgailis** (2008) ARCH($\infty$) and long memory properties. In *Handbook of Financial Time Series.* Eds. T. G. Andersen, R. A. Davis, J-P. Kreiss, T. Mikosch. Springer, Berlin Heidelberg New York.

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

**Hamadeh, T. and J-M. Zakoïan** (2009) Asymptotic properties of least-squares and quasi-maximum likelihood estimators for a class of nonlinear GARCH Processes. *Unpublished Document, University Lille 3.*

**Higgins, M.L. and A.K. Bera** (1992) A class of nonlinear ARCH models. *International Economic Review*, 33, 137–158.

**Karanasos, M.** (1999) Prediction in ARMA models with GARCH in mean effects. *Journal of Time Series Analysis* 22, 555–576.

**Kazakevičius, V. and R. Leipus** (2003) A new theorem on existence of invariant distributions with applications to ARCH processes. *Journal of Applied Probability* 40, 147–162.

**Loève, M.** (1977) Probability Theory I, 4th edition Springer, New-York.

**Nelson, D. B.** (1990) Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* 6, 318–334.

**Newey, W.K. and D.G. Steigerwald** (1997) Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models. *Econometrica* 65, 587–599.

**Pan, J., Wang, H., and H. Tong** (2008) Estimation and tests for power-transformed and threshold GARCH models. *Journal of Econometrics*, 142, 352–378.

**Pascual, L., Romo, J., and E. Ruiz** (2005) Bootstrap prediction for returns and volatilities in GARCH models. *Computational Statistics & Data Analysis* 50, 2293–2312.

**Robinson, P.M.** (1991) Testing for strong correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics*, 47, 67–84.

**Robinson, P.M. and P. Zaffaroni** (2006) Pseudo-Maximum Likelihood Estimation of Arch($\infty$) Models. *The Annals of Statistics*, 34, 1049–1074.

**Straumann, D. and T. Mikosch** (2006) Quasi-maximum likelihood estimation in conditionally heteroscedastic Time Series: a stochastic recurrence equations approach. *The Annals of Statistics* 5, 2449–2495.

**Teräsvirta, T.** (2007) An introduction to univariate GARCH models. Forthcoming in *Handbook of Financial Time Series*, ed. by T.G. Andersen, R.A. Davis, J.-P. Kreiss and T. Mikosch. New York: Springer.

**Zakoïan, J-M.** (1994) Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* 18, 931–955.