Centre de Recherche en Économie et Statistique
CREST

Série des Documents de Travail

n° 2012-17

Optimal Predictions of Powers of Conditionally Heteroskedastic Processes

C. FRANCQ[1]
J. M. ZAKOIAN[2]

August 21 2012

1 CREST and University Lille 3 (EQUIPPE).

2 CREST and University Lille 3 (EQUIPPE). *Address CREST; 15 Boulevard Gabriel Péri, 92245 Malakoff Cedex, France*. Email : zakoian@ensae.fr - Tel. : +33 1 41 17 78 25.

Optimal predictions of powers of conditionally heteroskedastic processes[1]

Christian Francq

CREST and Université Lille 3 (EQUIPPE)

Jean-Michel Zakoïan

CREST and Université Lille 3 (EQUIPPE)

August 21, 2012

Summary. In conditionally heteroskedastic models, the optimal prediction of powers, or logarithms, of the absolute value has a simple expression in terms of the volatility and an expectation involving the independent process. A natural procedure for estimating this prediction is to estimate the volatility in a first step, for instance by Gaussian quasi-maximum likelihood (QML) or by least-absolute deviations, 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 estimator, and establishes the asymptotic properties of the two approaches. Asymptotic comparisons and numerical experiments show that the differences in accuracy can be important, depending on the prediction problem and the innovations distribution. An application to indexes of major stock exchanges is given.

Keywords: Efficiency of estimators, GARCH, Least-absolute deviations estimation, Prediction, Quasi maximum likelihood estimation.

Address for correspondence: Jean-Michel Zakoïan, CREST, 15 Boulevard Gabriel Péri, 92245 Malakoff cedex.
Email: zakoian@ensae.fr

[1]A version of this article (without Appendix B) is forthcoming in the *Journal of the Royal Statistical Society, Series B.*
We are most thankful to the Joint Editor, Ingrid Van Keilegom, and to two referees for their constructive comments and suggestions. We are also grateful to the Agence Nationale de la Recherche (ANR), which supported this work via the Project ECONOM&RISK (ANR 2010 blanc 1804 03).

1. Introduction

Despite the considerable attention given to the autoregressive conditional heteroscedasticity (ARCH) model and its generalization, the GARCH model, 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 of the conditional mean in ARMA model with GARCH errors, and prediction of conditional variances in GARCH(p, q) models. 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 (2002) considered predicting the conditional mean and variance from an ARMA model with GARCH in mean effects. Pascual, Romo and Ruiz (2005) proposed a Bootstrap procedure to obtain prediction densities of returns and volatilities of GARCH processes. Pellegrini, Ruiz and Espasa (2012) analyze the effects of differencing GARCH with stochastic trend on the prediction intervals.

In this paper, our aim is to investigate the problem of predicting powers of the process (ϵ_t), defined as a solution of the general stochastic model

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

where (η_t) is a sequence of independent and identically distributed (iid) random variables, η_t being independent of $\{\epsilon_u, u < t\}$, $\theta_0 \in \mathbb{R}^m$ is a parameter belonging to a parameter space Θ, and $\sigma : \mathbb{R}^\infty \times \Theta \to (0, \infty)$. The variable σ_t^2 is generally referred to as the volatility of ϵ_t in the econometric literature.[2]

Most conditional volatility models can be embedded in Model (1). 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). Other specifications satisfying (1) are the GARCH(p, q), the asymmetric power GARCH(p, q) model proposed by Ding, Granger and Engle (1993), and the ARCH(∞) introduced by Robinson (1991). In GARCH models, it is generally assumed that $E\eta_t = 0$, but we do not make this assumption.

1.1. *Two approaches for predicting powers*

For any real number r such that $E|\eta_t|^r < \infty$, the best predictor in mean square of $|\epsilon_t|^r$ is the conditional expected value

$$E_{t-1}(|\epsilon_t|^r) = \sigma_t^r E|\eta_t|^r, \tag{2}$$

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

$$E_{t-1}(\log |\epsilon_t|) = \log \sigma_t + E \log |\eta_t|, \tag{3}$$

provided that $E \log |\eta_t|$ exists. This case can be seen as the limit of the case (2) when $r = 0$, via the Box-Cox transformation $\log |x| = \lim_{r \to 0}(|x|^r - 1)/r$.

[2]The choice of an appropriate GARCH model, including the orders, is clearly an important issue: see for instance Li (Chapter 6, 2004) for diagnostic tests in GARCH models, and Francq and Zakoïan (Chapter 5, 2010) for the selection of GARCH orders.

To estimate the volatility in Model (1) a scale constraint is required on the sequence (η_t), for evident identifiability reasons. The standard assumption $E\eta_t^2 = 1$ is required for the consistency of the Gaussian quasi-maximum likelihood estimator (QMLE), while the least absolute deviations estimator (LADE) requires the condition median$(\eta_t^2) = 1$. However, the constraint $E|\eta_t|^r = 1$, with $r \neq 0$, can be used as well. In view of this, we consider two approaches for predicting $|\epsilon_{n+1}|^r$, given observations $(\epsilon_1, \dots, \epsilon_n)$:

- A *fully parametric* one-step approach in which θ_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 (2) (resp. (3)) is then the estimated value of σ_{n+1}^r (resp. $\log \sigma_{n+1}$).
- A *mixed* (parametric and non parametric) two-step approach in which θ_0 is estimated by the practitioner's favorite estimator under a relevant identifiability assumption (for instance the QMLE under $E|\eta_t|^2 = 1$, or the LADE under median$(\eta_t^2) = 1$), 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 (2) (resp. (3)) is the estimated value of σ_{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 seems more natural. Many statistical procedures include two steps, involving the estimation of a characteristic of the error distribution in the second step. Examples include Generalized Least Squares (the errors variance), adaptive estimation (the errors density) and value-at-risk estimation in GARCH-type models (a quantile of the errors distribution). The fully parametric approach is novel, to our knowledge, and, as we shall demonstrate, it can provide efficiency gains over the more natural approach.

1.2. Non Gaussian QML

For the reparameterized model under the identifiability constraint $E|\eta_0|^r = 1$ with $r \neq 2$, the 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. Fan, Qi and Xiu (2010), and Francq, Lepage and Zakoian (2011) proposed two-stage procedures based on non Gaussian QMLE for estimating standard GARCH models under the standard identifiability condition. In the latter references, alternative QML criteria are introduced to achieve better efficiency than the Gaussian QMLE. By contrast, in the fully parametric approach of the present paper, the QML criterion will be imposed by the prediction problem under consideration.

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

The prediction of ϵ_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 moments of financial returns are an important measure of the market fluctuation. The conditional kurtosis, defined as the ratio of the fourth conditional moment over the squared volatility has drawn attention in the finance literature (see for instance Brooks, Burke, Heravi and Persand (2005) and the references therein). Conditional distributions of financial series typically display a sharp peak around zero as well as fat tails. To measure the fluctuations of tails, it is therefore sensible to evaluate how $E_{t-1}(\epsilon_t^4)/\{E_{t-1}(\epsilon_t^2)\}^2$ varies with time. If the GARCH model is correctly specified, this ratio should be constant. Estimation of the conditional kurtosis can thus be the basis of a formal specification test. In this paper we focus on the estimation of such conditional moments and leave this issue for further investigation.

ii) Interest of considering $0 \leq r < 2$. As argued by Taylor (2007, p. 398), "return outliers are amplified when they are squared and then forecast errors are typically very large compared with other times. Consequently another popular proxy [of volatility] is the absolute return." Indeed, when one suspects that second or fourth-order moments do not exist, it is sensible to consider predictions of absolute returns, or even smaller powers of returns as measures of the future price volatility.

iii) Interest of considering $r < 0$. For some applications, for instance duration time between events, it may be worth fitting a GARCH-type model to the inverses of the data. Autoregressive conditional duration (ACD) models were introduced by Engle and Russell (1998) for the analysis of durations. ACD 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: a GARCH effect is detected if the empirical correlations between present and past durations, $\widehat{\mathrm{Corr}}(x_t, x_{t-k})$ for $k > 0$, say, are significant. However, it may be of interest to capture clustering of small durations. Indeed, such small durations are likely to reflect high volatility of prices. If significant empirical correlations between present and past inverse durations, $\widehat{\mathrm{Corr}}(x_t^{-1}, x_{t-k}^{-1})$ for $k > 0$, are found, 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 ϵ_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 $r = -2$. In the supplementary document, we present an empirical example showing that such significant autocorrelations are detected for the inverses of durations between transactions.

1.4. Contributions of this paper

For the general Model (1), we study the aforementioned methods for predicting $|\epsilon_{n+1}|^r$ or $\log|\epsilon_{n+1}|$. For the first step of the mixed procedure, we focus on two methods: the Gaussian QMLE and the LADE.

Our main contributions are the following: 1) we introduce the one-step method, based on a generalized QML; 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 derived for Model (1); 4) the asymptotic properties of the mixed approach are derived; 5) for the standard GARCH model, we obtain surprisingly simple expressions for the Asymptotic Relative Efficiency (ARE) of the fully parametric method with respect to mixed methods in which either the QML or the LAD estimation is used in first step.

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). The choice of h is solved for the prediction problems (2)-(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 based on the Gaussian QML. For the standard GARCH(p,q), we show that the ARE of the one-step estimator only depends on the power r and moments of the iid process. Section 4 completes the comparison of the two approaches, by considering the LADE, instead of the Gaussian QMLE, for the two-step method. Section 5 proposes empirical applications based on financial data. The most technical assumptions and proofs are deferred to Appendix A. Complementary results, proofs and illustrations are collected in Appendix B.

2. Asymptotic properties 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 σ (for classical GARCH see Nelson (1990), Bougerol and Picard (1992)).

A0: (ϵ_t) is a strictly stationary and ergodic solution of (1).

Given observations $\epsilon_1, \dots, \epsilon_n$, and arbitrary initial values $\tilde{\epsilon}_i$ for $i \leq 0$, we define $\tilde{\sigma}_t(\theta) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_1, \tilde{\epsilon}_0, \tilde{\epsilon}_{-1}, \dots; \theta)$. 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

$$\tilde{Q}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} g(\epsilon_t, \tilde{\sigma}_t(\theta)), \qquad g(x,\sigma) = \log \frac{1}{\sigma} h\left(\frac{x}{\sigma}\right). \tag{4}$$

Let the QMLE

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

for some compact space Θ. This estimator is the standard Gaussian QMLE when h is the standard Gaussian density ϕ.

2.1. Identifiability conditions

To be able to identify the parameters in Model (1) it is necessary to impose a constraint on (η_t). For the sake of predicting $|\epsilon_t|^r$, a natural constraint in view of (2)-(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 η_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 σY where Y has the density h. Then **A3** can be written

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

This condition means that there is no way 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 η_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. Let $\mathcal{H}_0$ be the set of the instrumental densities $h > 0$ which are differentiable over $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$. For some constants $\delta \in \mathbb{R}$ and $C_0 > 0$, let

A4: $h \in \mathcal{H}_0$ with $|u\frac{h'(u)}{h(u)}| \leq C_0(1 + |u|^\delta)$ for all $u \in \mathbb{R}^*$ and $E|\eta_0|^\delta < \infty$.

A5: For any $x \in \mathbb{R}^m$, we have: $x' \left(\frac{\partial \sigma_t^2}{\partial \theta_i}\right)_{i=1,\ldots,m} = 0, \; a.s. \quad \Rightarrow \quad x = 0.$

A4 is a mild assumption which vanishes for instance when the instrumental density has the form $h(u) = K_1|u|^\lambda \exp\{K_2|u|^r\}$, for some constants λ, K_1, K_2. In this case, the inequality is satisfied with $\delta = r$ and the condition $E|\eta_0|^\delta < \infty$ thus follows from **A1**. Assumption **A5** is required for the invertibility of the matrix J involved in the asymptotic variance of the estimator. For specific forms of σ_t, for instance if the model is a standard GARCH, the condition reduces to standard assumptions on the lag polynomials of the volatility. For the reader's convenience, additional technical assumptions, **A6-A10**, are reported in the appendix. 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-A4** *and* **A6** *hold, for constants* $\delta \in \mathbb{R}$ *and* $C_0 > 0$*, then*

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

where θ_0 *is the true parameter value in Model (1) under the identifiability condition* **A1**. *If, in addition,* **A7-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

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

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

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

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

Remark 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). Straumann and Mikosch (2006) studied the Gaussian QMLE for conditionally heteroscedastic models where the volatility has the form $\sigma_t^2 = g(\epsilon_{t-1}, \ldots, \epsilon_{t-p}, \sigma_{t-1}^2, \ldots, \sigma_{t-q}^2; \theta)$. More recently Bardet and Wintenberger (2009) proved the asymptotic properties of the Gaussian QMLE for a general class of multidimensional causal processes encompassing (1). However, their conditions for consistency and AN require the existence of moments of orders 2 and 4 for ϵ_t, respectively, which we do not need for the class (1). Our sole moment assumptions are on η_t. For GARCH processes, the existence of moments for η_t does not imply finite moments for ϵ_t: see for instance Figures 2.8 and 2.10 in Francq and Zakoian (2010).

Remark 2. These results can also be compared with those, already mentioned in the introduction, using non Gaussian QMLE. Except in the case $r = 2$, our parameter θ_0 is not the GARCH parameter, denoted by θ_0^* in the sequel (see Equation (6)), which is estimated in the aforementioned articles. Fan, Qi and Xiu (2010) propose a method for estimating θ_0^* in the standard GARCH model, using a very general QML and two optimization procedures. For the same models and the same parameter θ_0^*, the approach of Francq, Lepage and Zakoian (2011) only requires one optimization procedure but uses more specific QMLs.

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* η_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 c and λ.

Proposition 1. *If the instrumental function h belongs to $\mathcal{C}(r)$ then the generalized QMLE is given by*

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

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

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

Table 2. Asymptotic efficiency of the ML with respect to the Generalized QML, for the Gaussian distribution, as a function of r.

r	0.01	0.1	0.25	0.5	1	1.5	2	2.5	3.5	4.5	9
$\tau^2_{h,f}/\tau^2_{f,f}$	2.43	2.12	1.78	1.44	1.14	1.03	1	1.02	1.19	1.53	9.04

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 ϵ_t^2 in $|\epsilon_t|^r$.[3] The following proposition shows that **A3** can be omitted in Theorem 2.1 when h is chosen in $\mathcal{C}(r)$.

PROPOSITION 2. *Let $h \in \mathcal{H}_0$ be such that* **A4** *holds. Then*

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

In view of Propositions 1 and 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, which is established in the supplementary document, characterizes the set of densities f of η_t for which a given h is optimal.

COROLLARY 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 η_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 $\tau^2_{f,f} < \tau^2_{h,f}$.

The loss of efficiency of the non Gaussian QML with respect to the ML is illustrated in Table 2 for Gaussian errors.

3. Mixed approach based on the Gaussian QML

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 (2) (or (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.

[3]When $r = 0$, the prediction of $\log|\epsilon_t|$ is equivalent to the prediction of the conditional mean in the regression model $y_t = \log\sigma_t(\theta_0) + e_t$, where $y_t = \log|\epsilon_t|$ and $e_t = \log|\eta_t|$. Gouriéroux, Monfort and Trognon (1984) showed that the consistent Pseudo maximum likelihood estimators (PMLE) are the minimizers of the objective functions $\sum_{t=1}^n \log h(y_t, \log\tilde{\sigma}_t(\theta))$, where $h(x, m)$ belongs to the family of the linear exponential densities of mean m. Such PMLE do not coincide with our QMLE because our estimator optimizes the likelihood of a scale parameter (see (4)) whereas the PMLE optimizes the likelihood of a location parameter.

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 $(x_i)_i$

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

Standard GARCH models obviously verify this assumption with

$$F(\theta, K) = (K^2\omega, K^2\alpha_1, \dots, K^2\alpha_q, \beta_1, \dots, \beta_p)' \tag{5}$$

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

$$\begin{cases} \epsilon_t = \sigma_t^* \eta_t^*, \quad E\eta_t^{*2} = 1, \\ \sigma_t^* = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_0^*) \end{cases} \tag{6}$$

The Gaussian QMLE of θ_0^*, denoted by $\hat{\theta}_n^*$, is defined as a maximizer over Θ of $n^{-1}\sum_{t=1}^n \log\left[\tilde{\sigma}_t^{-1}(\theta)\phi\left\{\tilde{\sigma}_t^{-1}(\theta)\epsilon_t\right\}\right]$. Let the rescaled residuals $\hat{\eta}_t^* = \epsilon_t/\hat{\sigma}_t^*$, 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 $s \neq 0$.

3.1. *Asymptotic distribution of* $(\hat{\theta}_n^*, \hat{\mu}_r^*)$

Theorem 3.1. *If* **A0**-**A2**, **B0**, **B1** *and, with* $\delta = \max(2, r)$, **A6**, **A9**-**A10** *hold, and* $\theta_0^* \in \overset{\circ}{\Theta}$, *then* $\hat{\theta}_n^* \to \theta_0^*$, $\hat{\mu}_r^* \to \mu_r^*$ *a.s. and*

$$\begin{pmatrix} \sqrt{n}\left(\hat{\theta}_n^* - \theta_0^*\right) \\ \sqrt{n}(\hat{\mu}_r^* - \mu_r^*) \end{pmatrix} \overset{\mathcal{L}}{\to} \mathcal{N}\left\{0, \Sigma_r := \begin{pmatrix} (\kappa_4 - 1)J_*^{-1} & -\lambda_r J_*^{-1}\Omega_* \\ -\lambda_r \Omega_*' J_*^{-1} & \sigma_{\mu_r^*}^2 \end{pmatrix}\right\}, \tag{7}$$

where

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

and

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

for $r \neq 0$, *and*

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

REMARK 3. In the proof, the following relation, of independent interest, is established:

$$\Omega_*' J_*^{-1} \Omega_* = 1. \quad (8)$$

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

REMARK 4. In the proof of (8) it is shown that $\hat{\mu}_2^* = \mu_2^* (= 1)$, *a.s.* This entails that, when $r = 2$, the two approaches for predicting ϵ_t^2 are the same. In this case, the asymptotic distribution for $\hat{\mu}_r^*$ is degenerate and $\hat{\theta}_n^*$ has the same limiting normal distribution as $\hat{\theta}_{n,\phi}$.

REMARK 5. In the centered Gaussian case, Σ_r is block-diagonal. Indeed, if η_t follows a $\mathcal{N}(0, N_r^{-2/r})$ distribution where $r > -1$ 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 $s > 1$. It follows that $\lambda_r = 0$.

3.2. Comparison of predictors

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 $\tilde{\sigma}^r(\epsilon_n, \epsilon_{n-1}, \ldots; \hat{\theta}_{n,h})$. By the mixed approach, based on the Gaussian QMLE $\hat{\theta}_n^*$ in Model (6), the same optimal prediction is estimated by $\tilde{\sigma}^r(\epsilon_n, \epsilon_{n-1}, \ldots; \hat{\theta}_n^*)\hat{\mu}_r^* = \tilde{\sigma}^r(\epsilon_n, \epsilon_{n-1}, \ldots; F(\hat{\theta}_n^*, \{\hat{\mu}_r^*\}^{1/r}))$. The optimal prediction $E_n \log|\epsilon_{n+1}|$ can similarly be estimated by $\log\tilde{\sigma}(\epsilon_n, \epsilon_{n-1}, \ldots; \hat{\theta}_{n,h})$ and $\log\tilde{\sigma}(\epsilon_n, \epsilon_{n-1}, \ldots; \hat{\theta}_n^*) + \hat{\mu}_0^* = \log\tilde{\sigma}(\epsilon_n, \epsilon_{n-1}, \ldots; F(\hat{\theta}_n^*, e^{\hat{\mu}_0^*}))$. 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 $G_0(\hat{\theta}_n^*, \hat{\mu}_0^*) = F(\hat{\theta}_n^*, e^{\hat{\mu}_0^*})$). Using the strong consistency of $(\hat{\theta}_n^*, \hat{\mu}_r^*)$, in Theorem 3.1, we have, under smoothness assumptions on the function F,

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

and

$$\sqrt{n}\left(\tilde{\theta}_n - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \left[\frac{\partial G_r(\theta_0^*, \mu_r^*)}{\partial(\theta', \mu)}\right] \Sigma_r \left[\frac{\partial G_r(\theta_0^*, \mu_r^*)'}{\partial(\theta', \mu)'}\right]\right). \quad (9)$$

The problem is thus to compare

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

The comparison can be explicitly done for the standard GARCH(p,q) model and some of its extensions. This is the object of the next section.

3.3. 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 admitting a finite fourth-order moment were obtained by Berkes, Horváth and Kokoszka (2003) and Francq and Zakoïan (2004). In this section, the results of Theorem 2.1 are applied to the standard GARCH(p,q) model

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t^2 = \omega_0 + \sum_{i=1}^q \alpha_{0i}\epsilon_{t-i}^2 + \sum_{j=1}^p \beta_{0j}\sigma_{t-j}^2 \end{cases} \quad (10)$$

where $\theta_0 = (\omega_0, \alpha_{01}, \dots, \beta_{0p})'$ satisfies $\omega_0 > 0, \alpha_{0i} \geq 0, \beta_{0j} \geq 0$. Let $\hat{\theta}_n^* = (\hat{\omega}^*, \hat{\alpha}_1^*, \dots, \hat{\beta}_p^*)$ be the Gaussian QMLE of $\theta_0^* = (\mu_2\omega_0, \mu_2\alpha_{01}, \dots, \mu_2\alpha_{0q}, \beta_{01}, \dots, \beta_{0p})'$. 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 (10) (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 theorem, which is proven in the supplementary document, provides the asymptotic distributions of the estimators of θ_0 involved in the two methods.

Theorem 3.2 (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* $\mathbf{C}$, *the one-step estimator of $\theta_0 \in \overset{\circ}{\Theta}$ satisfies*

$$\sqrt{n}\left(\hat{\theta}_{n,h} - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left\{0, \left(\frac{2}{r}\right)^2 \left(\frac{\kappa_{2r}}{\kappa_r^2} - 1\right) J^{-1}\right\}. \tag{11}$$

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

$$\sqrt{n}\left(\tilde{\theta}_n - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left\{0, (\kappa_4 - 1)J^{-1} + \left[\left(\frac{2}{r}\right)^2 \left(\frac{\kappa_{2r}}{\kappa_r^2} - 1\right) - (\kappa_4 - 1)\right] \overline{\theta}_0 \overline{\theta}_0'\right\} \tag{12}$$

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

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.

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

Corollary 2 (A criterion for efficiency comparison). *Under the assumptions of Theorem 3.2 the asymptotic variance matrices of the two estimators verify*

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

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

$$\left(\frac{2}{r}\right)^2 \left(\frac{\kappa_{2r}}{\kappa_r^2} - 1\right) \geq \kappa_4 - 1. \tag{13}$$

Remark 6. Surprisingly, the asymptotic efficiency comparison of the two approaches only depends on r and some moments of the iid process, not on θ_0. This result has importance for practical purposes. It gives a basis for selecting the (asymptotically) more efficient method, as a function of r and estimated moments of η_t. From the rescaled residuals of a standard GARCH estimation, one can estimate κ_r empirically for any value of r, and thus one should be able to infer which method is asymptotically the best. This will be illustrated in Section 5.

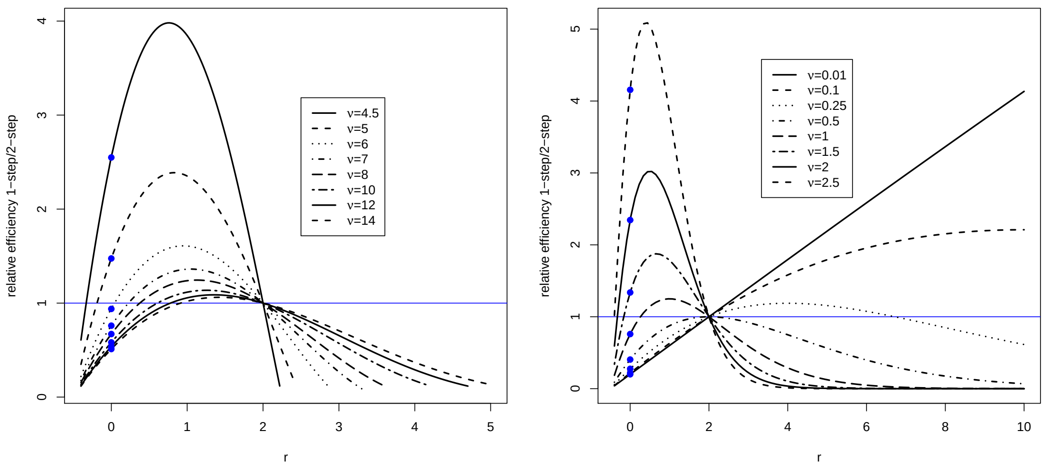

Fig. 1. ARE of the one-step QMLE relative to the two step QMLE for Student distributions (left panel) and GED (right panel) with parameter ν.

REMARK 7. An analogous of Theorem 3.2 and Corollary 2 is established for the case $r = 0$ in the supplementary document. The two-step estimator is asymptotically more accurate than the one-step estimator when

$$4\mathrm{Var}(\log|\eta_0|) \geq \kappa_4 - 1.$$

It is also shown that the asymptotic variances of the two estimators are the limits of the asymptotic variances in Theorem 3.2 when r goes to 0.

Figure 1 shows the ARE of the one-step QMLE relative to the two step QMLE as measured by the ratios[4]

$$(\kappa_4 - 1)/\left(\frac{2}{r}\right)^2\left(\frac{\kappa_{2r}}{\kappa_r^2} - 1\right) \quad \text{when } r \neq 0,$$

for Student distributions and Generalized Error Distributions (GED)[5], with parameter ν. For Student distributions (left panel), it is seen that the one-step method outperforms the indirect one when $r \in (k_\nu, 2)$ for some constant k_ν ranging from 0 (when $\nu \leq 6$) to 1 (when $\nu = 14$). On the contrary, for $r > 2$ and small or negative values of r, the two-step approach is preferable. The differences are particularly remarkable for small value of ν. As mentioned in Remark 7, the ARE's in the case $r = 0$, displayed as bullets in the graph, are the limits of the ARE's when r approaches zero. For the GED, similar conclusions can be drawn for $r \leq 2$, but the direct method can be superior to the two-step approach for $r > 2$.

[4]Note however that the term ARE does not refer here to the ratio of the asymptotic variances of two estimators.

[5]The density of η_0 is of the form $f(x) \propto e^{-0.5|x|^{1/\nu}}$.

3.4. 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 $\delta > 0$,

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t^\delta = \omega_0 + \sum_{i=1}^q \alpha_{0i+}(\epsilon_{t-i}^+)^\delta + \alpha_{0i-}(-\epsilon_{t-i}^-)^\delta + \sum_{j=1}^p \beta_{0j}\sigma_{t-j}^\delta \end{cases} \tag{14}$$

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).

The study conducted for the standard GARCH model can be reproduced for the Asymmetric Power GARCH. The most striking output of this study is that the conclusion of Corollary 2 holds true for Model (14): the estimator $\tilde{\theta}_n$ is asymptotically more efficient than $\hat{\theta}_{n,h}$ iff (13) holds. The proof is in Appendix B.

4. Mixed approach based on LAD estimation

The proposed one-step method can be compared not only to the Gaussian QML, but also to estimation procedures developed in the literature which are known to be relatively efficient in the case of Non-Gaussian noise; see for instance Davis, Knight and Liu (1992) and Ling (2005) for the study of M-estimators for autoregressions with infinite variance, Mukherjee (2008) for the study of M-estimators for GARCH. In this section, we consider a special case of M-estimation, the LAD estimator defined and studied by Peng and Yao (2003) in the context of GARCH models. For simplicity, we focus on the standard GARCH(p,q) model and $r \neq 0$.

To derive the LADE we reparameterize the model (10) in such a way that the median of the squared innovation be equal to 1. Assume that

C0: the density f_{η^2} of η_t^2 is continuous and positive at $M = \text{median}(\eta_t^2) > 0$ and $E|\log \eta_0^2| < \infty$,

and let $\eta_t^{**} = \frac{\eta_t}{\sqrt{M}}$. The LADE $\hat{\theta}_n^{**} = (\hat{\omega}^{**}, \hat{\alpha}_1^{**}, \dots, \hat{\beta}_p^{**})$ of $\theta_0^{**} = (M\omega_0, M\alpha_{01}, \dots, M\alpha_{0q}, \beta_{01}, \dots, \beta_{0p})'$, is defined as a minimizer over Θ of

$$\frac{1}{n}\sum_{t=1}^n |\log \epsilon_t^2 - \log \tilde{\sigma}_t^2(\theta)|. \tag{15}$$

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, for $r \neq 0$,

$$\hat{\mu}_r^{**} = \frac{1}{n}\sum_{t=1}^n |\hat{\eta}_t^{**}|^r, \quad \mu_r^{**} = E|\eta_t^{**}|^r.$$

The next theorem, which is proven in the supplementary document, provides the asymptotic distributions of the two-step LADE of θ_0 and its comparison with the one-step estimator.

THEOREM 4.1 (LAD FOR THE STANDARD GARCH(p,q)). *Let* $r \neq 0$, $h \in \mathcal{C}(r)$, $E|\eta_0|^r = 1$, $E|\eta_0|^{2r} < \infty$, $\theta_0^{**} \in \overset{\circ}{\Theta}$, *and let* **C** *and* **C0** *hold. Then, the two-step estimator based on LAD given by* $\tilde{\tilde{\theta}}_n = (\{\hat{\mu}_r^{**}\}^{2/r}\hat{\omega}^{**}, \{\hat{\mu}_r^{**}\}^{2/r}\hat{\alpha}_1^{**}, \ldots, \{\hat{\mu}_r^{**}\}^{2/r}\hat{\alpha}_q^{**}, \hat{\beta}_1^{**}, \ldots, \hat{\beta}_p^{**})$ *satisfies* $\tilde{\tilde{\theta}}_n \to \theta_0$ *a.s. Moreover, for any* $\sqrt{n}$*-consistent minimizer* $\hat{\theta}_n^{**}$ *of (15) we have*

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

where $\xi_{\eta^2} = \frac{1}{2Mf_{\eta^2}(M)}$. *Furthermore,*

$$Var_{as}\left\{\sqrt{n}\left(\hat{\theta}_{n,h} - \theta_0\right)\right\} \succeq Var_{as}\left\{\sqrt{n}\left(\tilde{\tilde{\theta}}_n - \theta_0\right)\right\} \tag{16}$$

if and only if

$$\left(\frac{2}{r}\right)^2\left(\frac{\kappa_{2r}}{\kappa_r^2} - 1\right) \geq \xi_{\eta^2}^2. \tag{17}$$

The conclusions are similar to those drawn for the two-step estimator based on the Gaussian QMLE. The ARE of the one-step QMLE relative to the two-step LADE only depends on the innovations distribution.

REMARK 8. Note that the ARE of the two-step QMLE relative to the two-step LADE, $\xi_{\eta^2}^2/(\kappa_4 - 1)$, does not depend on r. For example, when η_t follows the Student distribution (respectively the GED) with ν degrees of freedom, the two-step LADE is more efficient than the two-step QMLE for prediction of any power, *i.e.* $\xi_{\eta^2}^2 < \kappa_4 - 1$, if and only if $\nu < 5.52$ (respectively $\nu > 1.51$). More generally, for strongly heavy-tailed distributions the method based on LAD is often more efficient than that based on the QML. It is however possible to construct counter-examples for which ξ_{η^2} is arbitrarily large compared to $(\kappa_4 - 1)$, regardless of the tail thickness.

REMARK 9. An analogous of Theorem 4.1 is established for the case $r = 0$ in the supplementary document. The two-step LADE is asymptotically more accurate than the one-step QML estimator when $4\mathrm{Var}(\log|\eta_0|) \geq \xi_{\eta^2}^2$.

5. Numerical Illustrations

In this section we concentrate on the Threshold GARCH(1,1) model

$$\epsilon_t = \sigma_t\eta_t, \quad \sigma_t = \omega_0 + \alpha_{0+}(\epsilon_{t-1}^+) + \alpha_{0-}(-\epsilon_{t-1}^-) + \beta_0\sigma_{t-1}, \tag{18}$$

with the notation and the assumptions of (14). Suppose that $\epsilon_1, \ldots, \epsilon_n$ are observed, and consider three predictors of $|\epsilon_{n+1}|^r$ or $\log|\epsilon_{n+1}|$: (1) the two-step Gaussian QML predictor; (2) the two-step LAD predictor; (3) the one-step non-Gaussian QML predictor. In the following, we propose an "adaptive" procedure for determining the best method and for computing the appropriate prediction in practice. Then, this procedure will be illustrated on real data and compared with cruder prediction methods. Simulation experiments are available in Appendix B.

5.1. Implementation of the adaptive prediction method

The procedure is described for Model (18) in the case $r \neq 0$. It can be straightforwardly modified when $r = 0$. In view of Section 3.4 and its direct extension to the LAD estimator, the algorithm works the same way for any model belonging to the standard or Asymmetric Power GARCH classes.

Step 1. Fit the TGARCH(1,1) model (18) by Gaussian QML.

Step 2. Compute the rescaled residuals $\hat{\eta}_t^* = \frac{\epsilon_t}{\hat{\sigma}_t^*}$. Compute their empirical moments $\hat{\mu}_r^*$, $\hat{\mu}_{2r}^*$, $\hat{\mu}_4^*$. Compute the empirical median $\hat{M}^*$, and estimate the density $f_{\eta^{*2}}$ (for instance by the Kernel method) of the squared residuals $\hat{\eta}_t^{*2}$.

Step 3. Compute the quantities

$$c_0 = \frac{\hat{\mu}_4^*}{\hat{\mu}_2^{*2}} - 1, \qquad c_1 = \left(\frac{2}{r}\right)^2 \left(\frac{\hat{\mu}_{2r}^*}{\hat{\mu}_r^{*2}} - 1\right), \qquad c_2 = \frac{1}{\{2\hat{M}^* \hat{f}_{\eta^{*2}}(\hat{M}^*)\}^2}.$$

(i) If $c_0 = \min_{i=0,1,2} c_i$, then the Gaussian QMLE can be preferred for the prediction of $|\epsilon_{n+1}|^r$. The prediction is computed as $\hat{\sigma}_n^{*r} \hat{\mu}_r^*$.

(ii) If $c_2 = \min_{i=0,1,2} c_i$, then the LADE can be preferred. Reestimate the TGARCH(1,1) model by LAD,[6] and compute the $\hat{\eta}_t^{**} = \frac{\epsilon_t}{\hat{\sigma}_t^{**}}$. Compute their empirical moment $\hat{\mu}_r^{**}$. The prediction of $|\epsilon_{n+1}|^r$ is $\hat{\sigma}_n^{**r} \hat{\mu}_r^{**}$.

(iii) If $c_1 = \min_{i=0,1,2} c_i$, then the one-step estimator can be preferred. Reestimate the TGARCH(1,1) model by non-gaussian QML, by minimizing $\sum_{t=1}^n \log \tilde{\sigma}_t^r(\theta) + |\epsilon_t|^r / \tilde{\sigma}_t^r(\theta)$. The prediction of $|\epsilon_{n+1}|^r$ is $\tilde{\sigma}_n^r$.

Interestingly, the determination of the more efficient procedure in Step 3 can be based on the sole estimation by Gaussian QML. Note also that the numbers c_i are invariant by scale transformation of the residuals (which explains that c_3 is also an estimator of the number ξ_{η^2} of Theorem 4.1).

5.2. Empirical Illustration

We now consider prediction of powers 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 October 13, 2011, for the indices for which such historical data exist. Note that this period of time includes the recent sovereign-debt crises in Europe and US. We checked that the results are not qualitatively changed by suppressing the recent turbulent period or by replacing the TGARCH model (18) by a standard GARCH(1,1). Before applying our procedure to these data, it is of interest to determine, for each series and each power r, which method is asymptotically the best. This can be done by estimating the ARE of the one step-method with respect to the two-step methods, as derived in Section 3.2.

5.2.1. *Estimating the ARE of the one-step and two-step methods*

Figure 2 presents the estimated relative efficiencies of the one-step QMLE relative to the two step QMLE and LADE for the ten stock index returns. For the left panel, the TGARCH(1,1) model (18) 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, from which the ARE estimator, c_0/c_1, is computed (see Section 5.1). It is seen that, from an asymptotic point of view, the direct approach should

[6] For instance $(\hat{M}^* \hat{\omega}^*, \hat{M}^* \hat{\alpha}^*, \hat{\beta}^*)'$ can be used as initial value for the estimation of θ_0^{**}.

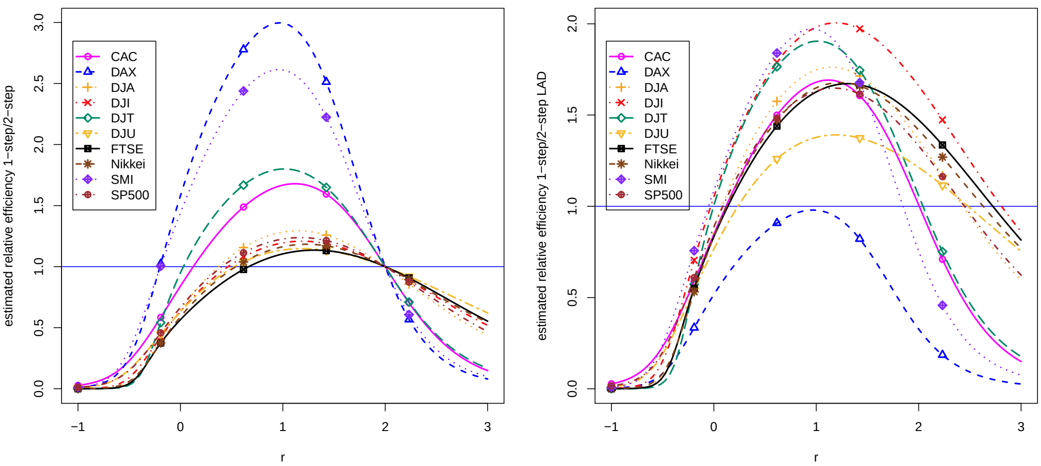

Fig. 2. Estimated ARE's of the one-step QMLE relative to the two-step QMLE (left panel) and LAD estimator (right panel) for stock index returns.

be superior for $r \in (0.5, 2)$; the indirect one should be preferable for $r > 2$ or $r < -0.5$. For $r \in (-0.5, 0.5)$ the results are more balanced. Turning to the right panel, the TGARCH(1,1) model (18) is now estimated by LAD in a first step, while the second step allows to compute the estimated AREs, c_2/c_1. The direct approach remains superior, for most series, when $r \in (0.5, 2)$. For some series, it is also superior for values of r larger than 2. Conversely, the two-step method is preferable for the DAX series, whatever r.

5.2.2. *Out-of-sample comparisons*

For predicting the power r of the next log-return, three competing methods are investigated: the "historical" prediction which predicts the next value by empirical means of observations to the power r; the "naive" method[7] which uses the power $r/2$ of the usual prediction of the squared return; the adaptive method described in Section 5.1.

Our numerical experiments setting is as follows, for each series ϵ_t of length n. Based on $\epsilon_{n_2-n_1+1}, \dots, \epsilon_{n_2}$, with $0 < n_1 \leq n_2 < n$, the historical prediction of $|\epsilon_{n_2+1}|^r$ is computed from the formula

$$\text{Historic}_{n_2+1} = \frac{1}{n_1} \sum_{t=n_2-n_1+1}^{n_2} |\epsilon_t|^r.$$

The Mean Square Prediction Error (MSPE) is thus

$$\frac{1}{n-n_1} \sum_{n_2=n_1}^{n-1} \left(|\epsilon_{n_2+1}|^r - \text{Historic}_{n_2+1} \right)^2.$$

[7]the method can be called naive because it targets $\{E_{t-1}(\epsilon_t^2)\}^{r/2}$ instead of $E_{t-1}|\epsilon_t|^r$.

Table 3. Percentages of MSPE losses with respect to the best method, for prediction of $|\epsilon_{n+1}|^r$.

r	0	0	0	0.5	0.5	0.5	1	1	1	2	2	2
method	N	H	A	N	H	A	N	H	A	N	H	A
CAC	28.6	7.3	0	19.2	14.9	0	6	21	0	0	19.3	0
DAX	27.8	6.6	0	17.8	13.7	0	4.5	20.8	0	0	21.7	0.1
DJA	33.3	7.2	0	24.3	16.8	0	7.4	25.1	0	7.2	29.3	0
DJI	34.8	7.8	0	25.3	17.8	0	7.7	26.9	0	0	31	0
DJT	30.2	2.8	0	21.4	6.9	0	6.6	9.1	0	0.5	4.8	0
DJU	30.9	5.1	0	21	13.3	0	5.7	23.5	0	0	28.8	0
FTSE	29.3	6.9	0	20.5	15.6	0	6	23.6	0	0	26.1	0
Nikkei	33.3	4.7	0	26.8	10.6	0	10	16.7	0	0	25.9	0.1
SMI	32.6	9.1	0	22.2	18.8	0	6.5	28.9	0	0	38.3	0.2
SP500	34	7.6	0	22.8	18	0	5.8	27.7	0	2.9	30.8	0

N denotes the naive method, H denotes the historic method, A denotes the adaptive method.

For the two other methods, n_1 observations are used to estimate the GARCH models and the appropriate moments. The same estimated model is kept for the computation of n_3 predictions.

For $n_1 = n_2 = n_3 = 300$, Table 3 displays the percentages of prediction losses with respect to the best method. For instance, for $r = 0$, the MSPEs of the CAC are 1.654 for the naive method, 1.380 for the historical method and 1.286 for the adaptive method. The adaptive method is thus the best in this case, implying a percentage of loss of 0, while the percentages of MSPE losses of the two other methods are $28.6 = 100 \times (1.654 - 1.286)/1.286$ and 7.3, respectively.

An outstanding feature of the results presented in Table 7 is that the adaptive method is superior in most cases to its competitors (as indicated by the presence of zeroes in the adaptive columns). Note that for $r = 2$, the naive method coincides with the two-stage Gaussian QML, and also with the generalized QML.[8] As expected, the naive method is spurious for predicting powers which are very different from the square (in particular $r = 0$). The performance of the historical method is satisfactory for small values of r but deteriorates as r increases. Experiments conducted with 30 observations (instead of 300) in the sample means of the historical method lead to slightly better results (see Appendix B) for this method, which is however still generally dominated by the adaptive approach. The overall conclusion of this study is that the adaptive method performs well, the MSPEs being always smaller than (or very close to) those of the two other methods.

6. Conclusion

We have shown that, in 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. 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. We also derived the asymptotic properties of alternative two-step approaches which combine Gaussian QML or LAD estimation in a first

[8]The adaptive method, in the case $r = 2$, sometimes chooses the LAD method which explains that the results for the adaptive method do not coincide with those of the naive method.

step, and estimation of the r-th order moment of the innovations in a second step.

In the case of finite-order standard and nonlinear GARCH models, we obtained a surprisingly simple, and easy to estimate, expression for the AREs of the one-step estimator with respect to the two-step QML and LAD estimators.

We suggest a procedure based on a sole Gaussian QML estimation of the model to determine which method should be used. Applied to a set of stock indices, this procedure suggests that the one-step approach should in general be used for moderate values of r. Conversely, for predicting $|\epsilon_t|^r$ with $r > 2$ or $r < 0.5$, the two-step methods should do a better job. We compared out-of-sample predictions obtained by the proposed adaptive procedure with more elementary approaches. The superiority of the adaptive method appears in a vast majority of cases, whatever the value of r.

A natural extension of this work would consider heteroskedastic models including a conditional mean. For instance, Ling (2004) introduced a class of double-autoregressive models and studied the properties of the QMLE, while Audrino and Bühlmann (2009) developed estimation procedures for a non-parametric class. This extension is left for future research.

A. Technical assumptions and proofs

Let $\Delta_t(\theta) = \tilde{\sigma}_t(\theta) - \sigma_t(\theta)$, $a_t = \sup_{\theta\in\Theta} |\Delta_t(\theta)|$. Constants $\delta \in \mathbb{R}$ and $C_0 > 0$ refer to Assumption **A4**. Let C and ρ be generic constants, whose values will be modified along the proofs, such that $C > 0$ and $0 < \rho < 1$. In assumptions **A6** and **A9** below, C is allowed to be a random variable which is measurable with respect to $\{\epsilon_u, u \leq 0\}$.

A6: For any real sequence (x_i), the function $\theta \mapsto \sigma(x_1, x_2, \ldots; \theta)$ is continuous. When $\delta > 0$ we have $E|\epsilon_0|^s < \infty$, when $\delta < -1$ we have $E \sup_{\theta\in\Theta} \sigma_0^s(\theta) < \infty$ for some $s > 0$. We have $a_t \leq C\rho^t$, a.s.

A7: θ_0 belongs to the interior $\overset{\circ}{\Theta}$ of Θ.

A8: h is twice differentiable at all $u \in \mathbb{R}^*$ with $|u^2 (h'(u)/h(u))'| \leq C_0(1 + |u|^\delta)$ and $E|\eta_0|^{2\delta} < \infty$.

A9: For any real sequence (x_i), the function $\theta \mapsto \sigma(x_1, x_2, \ldots; \theta)$ has continuous second-order derivatives. There exists a neighborhood $V(\theta_0)$ of θ_0 such that

$$b_t := \sup_{\theta\in V(\theta_0)} \left\| \frac{\partial \Delta_t(\theta)}{\partial\theta} \right\| \leq C\rho^t, \quad a.s.$$

A10: The following variables have finite expectation:

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

Assumptions **A6, A9, A10** are satisfied for standard GARCH models and many extensions. In particular, see Francq and Zakoïan (2004) for the bounds in $C\rho^t$ for a_t and b_t. Assumption **A7** is a standard assumption. Assumption **A8** reduces to $E|\eta_0|^{2r} < \infty$ for instrumental densities of the form $h(u) = K_1|u|^\lambda \exp\{K_2|u|^r\}$, for some constants λ, K_1, K_2.

A.1. Proof of Proposition 2

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. For $h \in \mathcal{H}_0$, we will use the convention that $xh'(x)/h(x)$ is equal to zero at $x = 0$. We then have

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

for $\sigma > 0$. Under **A4** we have $E \sup_{\sigma \in V(1)} |g_1(\eta_0, \sigma)| < \infty$, for some neighborhood $V(1)$ of 1. The dominated convergence theorem shows that **A3** entails the moment condition

$$E\left(\frac{h'(\eta_0)}{h(\eta_0)}\eta_0\right) = -1. \tag{19}$$

The problem is to find $h \in \mathcal{H}_0$ satisfying (19) for any distribution satisfying **A1**. The set of all possible densities h is thus,

$$\mathcal{H} = \left\{ h \in \mathcal{H}_0 \mid \quad \text{for any variable } \eta, \quad E|\eta|^r = 1 \Rightarrow E\left(\frac{h'(\eta)}{h(\eta)}\eta\right) = -1 \right\}.$$

We note that this set contains the set

$$\mathcal{H}' = \left\{ h \in \mathcal{H}_0 \mid \exists \lambda, \quad \frac{h'(x)}{h(x)}x + 1 = \lambda(|x|^r - 1) \right\}.$$

Now we prove that $\mathcal{H} \subset \mathcal{H}'$. Let $h \notin \mathcal{H}'$. If $h'(1)/h(1) \neq -1$ then $h \notin \mathcal{H}$ because if $\eta = 1$ a.s. then $E|\eta|^r = 1$ and $E\eta h'(\eta)/h(\eta) \neq -1$. Similarly $h'(-1)/h(-1) \neq 1$ entails $h \notin \mathcal{H}$. Now consider the case where $|x_1| > 1$, $|x_2| < 1$ 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 η 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

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

We thus have $h \notin \mathcal{H}$. 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. □

A.2. Proof of Theorem 2.1

The consistency is a consequence of the following intermediate results:

$i)$ $\lim_{n\to\infty} \sup_{\theta \in \Theta} |Q_n(\theta) - \tilde{Q}_n(\theta)| = 0$, *a.s.*, where $Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} g(\epsilon_t, \sigma_t(\theta))$,

$ii)$ if $\theta \neq \theta_0$, $\mathbb{E}g(\epsilon_1, \sigma_1(\theta)) < \mathbb{E}g(\epsilon_1, \sigma_1(\theta_0))$,

$iii)$ any $\theta \neq \theta_0$ has a neighborhood $V(\theta)$ such that

$$\limsup_{n\to\infty} \sup_{\theta^* \in V(\theta)} \tilde{Q}_n(\theta^*) < \limsup_{n\to\infty} \tilde{Q}_n(\theta_0) , \quad a.s.$$

The AN is proven by means of the following intermediate results: for some neighborhood $V(\theta_0)$ of θ_0 and for any θ^* between $\hat{\theta}_{n,h}$ and θ_0,

$$iv)\ \lim_{n\to\infty}\sqrt{n}\sup_{\theta\in V(\theta_0)}\left\|\frac{\partial}{\partial\theta}Q_n(\theta)-\frac{\partial}{\partial\theta}\tilde{Q}_n(\theta)\right\|=0\ ,\quad \text{in probability,}$$

$$v)\ \frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\theta^*)\to\frac{Eg_2(\eta_0,1)}{4}J\ ,\quad \text{in probability,}$$

$$vi)\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_0)\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,\frac{Eg_1^2(\eta_0,1)}{4}J\right),$$

To save place, we only give the proof of *i)*. This point, as well as *iv)*, which deal with the effect of the initial values, constitute the most delicate parts of the proof and illustrate the necessity of assumptions of the form **A4**, **A6** and **A9-A10**.

Using a Taylor expansion, almost surely

$$\begin{aligned}\sup_{\theta\in\Theta}|Q_n(\theta)-\tilde{Q}_n(\theta)| &\le n^{-1}\sum_{t=1}^n\sup_{\theta\in\Theta}|g_1(\epsilon_t,\sigma_t^*(\theta))||\Delta_t(\theta)|\\ &\le n^{-1}\sum_{t=1}^n a_t\sup_{\theta\in\Theta}\left|\frac{1}{\sigma_t^*}\frac{\epsilon_t}{\sigma_t^*}\frac{h'}{h}\left(\frac{\epsilon_t}{\sigma_t^*}\right)\right|+\frac{1}{\underline{\omega}}\,n^{-1}\sum_{t=1}^n a_t\\ &\le n^{-1}\sum_{t=1}^n a_t|\epsilon_t|^\delta\sup_{\theta\in\Theta}\left|\frac{1}{\sigma_t^*}\right|^{1+\delta}+\frac{C}{n}\sum_{t=1}^n a_t \qquad (20)\end{aligned}$$

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 $\delta\ge -1$. Then the supremum in (20) is bounded by C. If $\delta>0$, by the Markov inequality and **A6**, we deduce

$$\sum_{t=1}^\infty\mathbb{P}(a_t|\epsilon_t|^\delta>\varepsilon)\ \le\ \sum_{t=1}^\infty\frac{C\rho^{ts/\delta}\mathbb{E}|\epsilon_t|^s}{\varepsilon^{\frac{s}{\delta}}}<\infty$$

and thus $a_t|\epsilon_t|^\delta\to 0$ $a.s$ by the Borel-Cantelli lemma. The first term in (20) thus tends to zero $a.s.$, when $\delta>0$, by the Cesàro lemma. Now, if $\delta\in[-1,0]$, we note that $E|\epsilon_t|^\delta<\underline{\omega}^{-\delta}E|\eta_t|^\delta<\infty$ by **A2** and **A4**. Note also that, for $s\in(0,1)$, the c_r inequality (see Loève, 1977) entails

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

The last sum is $a.s.$ finite since its expectation is finite by **A6** and $E|\epsilon_t|^{\delta s}<\infty$ (because $s\in(0,1)$). Hence the first term in (20) tends to zero $a.s.$ when $\delta\in[-1,0]$. Now suppose $\delta<-1$. Observe that $\sup_{\theta\in\Theta}\sigma_t^*(\theta)\le\sup_{\theta\in\Theta}\sigma_t(\theta)+a_t$. It follows that, letting $\overline{\sigma}_t=\sup_{\theta\in\Theta}\sigma_t(\theta)$, the first term in (20) can be bounded by

$$\frac{C}{n}\sum_{t=1}^n a_t|\eta_t|^\delta\{\overline{\sigma}_t+a_t\}^{-(1+\delta)}\quad\le\quad\frac{C}{n}\sum_{t=1}^n\rho^t|\eta_t|^\delta\overline{\sigma}_t^{-(1+\delta)}+\frac{C}{n}\sum_{t=1}^n\rho^{-\delta t}|\eta_t|^\delta. \qquad (21)$$

Now by **A6**, there exists $s>0$ such that

$$\sum_{t=1}^\infty\mathbb{P}(\rho^t|\eta_t|^\delta\overline{\sigma}_t^{-(1+\delta)}>\varepsilon)\quad\le\quad\sum_{t=1}^\infty\frac{\left(\rho^{\frac{ts}{-(1+\delta)}}\mathbb{E}\overline{\sigma}_0^s\right)^{1/2}E|\eta_t|^{\frac{s\delta}{-2(1+\delta)}}}{\varepsilon^{\frac{s}{-2(1+\delta)}}}<\infty.$$

Thus, the first term in the right-hand side of (21) tends to zero $a.s.$ by the Cesàro Lemma. The second term is treated straightforwardly. We have shown that the first term in the right-hand side of (20) tends to zero $a.s.$ whatever the value of δ. By **A6**, the second term also tends to zero. Thus *i)* follows. $\square$

A.3. 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 η_0 replaced by η_0^*, is satisfied when h is the standard Gaussian density. Note also that **A4** and **A8** hold with $\delta = 2$. Thus, by Theorem 2.1, $\hat{\theta}_n^* \to \theta_0^*$ a.s. and

$$\sqrt{n}\left(\hat{\theta}_n^* - \theta_0^*\right) = -J_*^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(1 - \frac{\eta_t^2}{E\eta_t^2}\right)\frac{1}{\sigma_t^{*2}}\frac{\partial\sigma_t^2(\theta_0^*)}{\partial\theta} + o_P(1) \stackrel{\mathcal{L}}{\to} \mathcal{N}(0, (\kappa_4 - 1)J_*^{-1}). \tag{22}$$

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

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

We similarly define $\tilde{\mu}_r(\theta)$, by replacing $\eta_t(\theta)$ by $\tilde{\eta}_t(\theta)$. By **A6**, it can be shown that

$$\hat{\mu}_r^* = \tilde{\mu}_r(\hat{\theta}_n^*) = \mu_r(\hat{\theta}_n^*) + o_P(n^{-1/2}). \tag{23}$$

A Taylor expansion gives

$$\mu_r(\hat{\theta}_n^*) = \mu_r(\theta_0^*) + \frac{\partial\mu_r(\theta^*)}{\partial\theta'}(\hat{\theta}_n^* - \theta_0^*) \tag{24}$$

with θ^* between $\hat{\theta}_n^*$ and θ_0^* and

$$\left\|\frac{\partial\mu_r(\theta^*)}{\partial\theta'}\right\| \leq \frac{K}{n}\sum_{t=1}^{n}|\eta_t|^r\left(\frac{\sigma_t(\theta_0)}{\sigma_t(\theta^*)}\right)^r\frac{1}{\sigma_t^2(\theta^*)}\left\|\frac{\partial\sigma_t^2(\theta^*)}{\partial\theta'}\right\|.$$

Using Assumption **A10** and the Cauchy-Schwarz inequality, the left hand side has finite expectation. Thus, in view of the consistency of $\hat{\theta}_n^*$ to θ_0^*, the last term in (24) converges to 0 a.s. In view of the convergence of $\mu_r(\theta_0^*)$ to μ_r^*, and using (23), the strong consistency of $\hat{\mu}_r^*$ follows.

Now, by (22), **A10** and standard arguments, a Taylor expansion gives

$$\mu_r(\hat{\theta}_n^*) = \mu_r(\theta_0^*) + \frac{\partial\mu_r(\theta_0^*)}{\partial\theta'}(\hat{\theta}_n^* - \theta_0^*) + o_P(n^{-1/2}) \tag{25}$$

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

$$\begin{aligned}\sqrt{n}(\hat{\mu}_r^* - \mu_r^*) &= \sqrt{n}\{\mu_r(\theta_0^*) - \mu_r^*\} - \frac{r}{2}E|\eta_t^*|^r\Omega_*'\sqrt{n}(\hat{\theta}_n^* - \theta_0^*) + o_P(1) \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^n(|\eta_t^*|^r - \mu_r^*) - \frac{r}{2}E|\eta_t^*|^r\Omega_*'\sqrt{n}(\hat{\theta}_n^* - \theta_0^*) + o_P(1) \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^n\frac{(|\eta_t|^r - 1)}{\mu_2^{r/2}} - \frac{r}{2}\kappa_r\Omega_*'\sqrt{n}(\hat{\theta}_n^* - \theta_0^*) + o_P(1). \qquad (26)\end{aligned}$$

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 (26) that $\mathrm{Cov}\left(\sqrt{n}\left(\hat{\theta}_n^* - \theta_0^*\right), \sqrt{n}(\hat{\mu}_r^* - \mu_r^*)\right) = -\lambda_r J_*^{-1}\Omega_* + o_P(1)$. 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 (7), provided that (8) holds. Now we prove (8). 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 (4) 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. □

B. Complementary results

B.1. Proof of Corollary 1

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

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

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 seen in the proof of Proposition 2. A similar argument holds when $r = 0$. □

B.2. 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 *iv)-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 θ 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

$$\mathbb{E} \sup_{\theta^* \in V_k(\theta)\cap\Theta} g(\epsilon_t, \sigma_t(\theta^*)) < \log \frac{1}{\underline{\omega}} + C < \infty. \tag{27}$$

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 (27), 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 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

$$\begin{aligned}
&\sup_{\theta\in V(\theta_0)} \sqrt{n}\left\|\frac{\partial}{\partial\theta}Q_n(\theta)-\frac{\partial}{\partial\theta}\tilde{Q}_n(\theta)\right\| \\
\leq\ & \sup_{\theta\in V(\theta_0)} \frac{1}{\sqrt{n}}\sum_{t=1}^{n}|g_1(\epsilon_t,\sigma_t(\theta))-g_1(\epsilon_t,\tilde{\sigma}_t(\theta))|\left\|\frac{\partial\sigma_t(\theta)}{\partial\theta}\right\| \\
&+\sup_{\theta\in V(\theta_0)} \frac{1}{\sqrt{n}}\sum_{t=1}^{n}|g_1(\epsilon_t,\tilde{\sigma}_t(\theta))|\left\|\frac{\partial\sigma_t(\theta)}{\partial\theta}-\frac{\partial\tilde{\sigma}_t(\theta)}{\partial\theta}\right\|. \qquad (28)
\end{aligned}$$

Similarly to (20), the last term is bounded on $V(\theta_0)$ by

$$\begin{aligned}
&\frac{C}{\sqrt{n}}\sum_{t=1}^{n} b_t\left\{|\epsilon_t|^{\delta}\sup_{\theta\in V(\theta_0)}\left|\frac{1}{\tilde{\sigma}_t(\theta)}\right|^{1+\delta}+1\right\} \\
\leq\ & \frac{C}{\sqrt{n}}\sum_{t=1}^{n} b_t|\eta_t|^{\delta}\sup_{\theta\in V(\theta_0)}\left|\frac{\sigma_t(\theta_0)}{\tilde{\sigma}_t(\theta)}\right|^{\delta}+\frac{C}{\sqrt{n}}\sum_{t=1}^{n} b_t. \qquad (29)
\end{aligned}$$

We will prove that there exists $s>0$ such that

$$\sup_t E\sup_{\theta\in V(\theta_0)}\left\{\frac{\sigma_t(\theta_0)}{\tilde{\sigma}_t(\theta)}\right\}^{\delta s}<\infty. \qquad (30)$$

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

$$\begin{aligned}
\left\{\frac{\sigma_t(\theta_0)}{\tilde{\sigma}_t(\theta)}\right\}^{\delta s} &= \left\{\frac{\sigma_t(\theta_0)}{\sigma_t(\theta)}\right\}^{\delta s}-2\delta s\Delta_t(\theta)\{\sigma_t(\theta_0)\}^{\delta s}\left\{\frac{1}{\sigma_t^*(\theta)}\right\}^{\delta s+1} \\
&\leq \left\{\frac{\sigma_t(\theta_0)}{\sigma_t(\theta)}\right\}^{\delta s}+C\rho^t\{\sigma_t(\theta_0)\}^{\delta s}
\end{aligned}$$

since $\delta s>0$ for s small enough. The first term in the right-hand side admits a finite expectation when $s\leq 2$ using **A10**. The second term admits a finite expectation, hence (30) is proved.

We have $E|\eta_t|^{\delta s}<\infty$ for $s\in(0,1)$. Therefore, by (30),

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

and thus the random variable inside the bracket is a.s. finite. It follows that

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

which shows that the first term in the right-hand side of (29) goes to zero a.s. as n tends to infinity. The second term is handled in a straightforward way. Thus the last term in (28) converges to zero a.s. as n tends to infinity. Now note that

$$g_2(x,\sigma) := \frac{\partial g_1(x,\sigma)}{\partial \sigma} = \frac{1}{\sigma^2}\left[1 + 1_{\{x\neq 0\}}\frac{x}{\sigma}\left\{2\frac{h'}{h} + \frac{x}{\sigma}\left(\frac{h'}{h}\right)'\right\}\left(\frac{x}{\sigma}\right)\right]. \tag{31}$$

From **A4** and **A8**, the first term in the right-hand side of (28) is bounded by

$$\begin{aligned}
&\sup_{\theta\in V(\theta_0)}\frac{1}{\sqrt{n}}\sum_{t=1}^n |g_2(\epsilon_t,\sigma_t^*)||\Delta_t(\theta)|\left\|\frac{\partial\sigma_t(\theta)}{\partial\theta}\right\| \\
\leq\ & \frac{C}{\sqrt{n}}\sum_{t=1}^n a_t\left(1+|\epsilon_t|^\delta \sup_{\theta\in V(\theta_0)}\left|\frac{1}{\sigma_t^*}\right|^{2+\delta}\right)\sup_{\theta\in V(\theta_0)}\left\|\frac{1}{\sigma_t(\theta)}\frac{\partial\sigma_t(\theta)}{\partial\theta}\right\|
\end{aligned} \tag{32}$$

where $\sigma_t^* = \sigma_t^*(\theta)$ is between $\tilde{\sigma}_t(\theta)$ and $\sigma_t(\theta)$. For $\delta > 0$ there exists $s \in (0, 2\delta)$ such that, by the c_r and Cauchy-Schwarz inequalities

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

by **A6** and **A10**. For $\delta \in [-1, 0]$ and $s \in (0, 1)$ we have, similarly, using $E|\epsilon_0|^\delta \leq \underline{\omega}^\delta E|\eta_0|^\delta$,

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

The case $\delta < -1$ is treated in the same fashion, using an inequality similar to (21). By arguments already used, we conclude that the first term in the right-hand side of (28) goes to zero a.s. as n tends to infinity. Thus *iv)* is established. The invertibility of J follows from **A5**.

Now we establish *v)*. In view of **A4** and **A8**, 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, **A8** and **A10**. The ergodic theorem then implies that

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

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 (19),

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

and by (31), $g_2(\epsilon_t, \sigma_t(\theta_0)) = g_2(\eta_t, 1)\sigma_t^{-2}(\theta_0)$. 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

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

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

Now, from **A7** 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 θ^* is between $\hat{\theta}_{n,h}$ and θ_0. Applying *iv), v), vi)*, the proof of the asymptotic normality is complete.

B.3. Complementary results for the proof of Theorem 3.1

Proof of (23): Using **A2** and **A6**, we have

$$\begin{aligned}|\tilde{\mu}_r(\theta)-\mu_r(\theta)| &\leq \frac{1}{n}\sum_{t=1}^{n}|\epsilon_t|\frac{|\Delta_t(\theta)|}{\sigma_t(\theta)\tilde{\sigma}_t(\theta)}\\ &\leq \frac{1}{n\underline{\omega}^2}\sum_{t=1}^{n}a_t|\epsilon_t| \leq \frac{C}{n}\sum_{t=1}^{\infty}\rho^t|\epsilon_t|.\end{aligned}$$

By **A6**, $E|\epsilon_t|^s < \infty$ for some $s \in (0,1)$. By the c_r-inequality, $E\left(\sum_{t=1}^{\infty}\rho^t|\epsilon_t|\right)^s \leq \sum_{t=1}^{\infty}\rho^t E|\epsilon_t|^s < \infty$. It follows that

$$\sup_{\theta\in\Theta}|\tilde{\mu}_r(\theta)-\mu_r(\theta)| = O(1/n) \text{ a.s.}$$

which is a stronger result than (23).

Proof of (25): A Taylor expansion yields

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

where θ^* is between $\hat{\theta}_n^*$ and θ_0^* and

$$\begin{aligned}\frac{\partial\mu_r(\theta)}{\partial\theta'} &= \frac{-r}{2n}\sum_{t=1}^{n}|\eta_t(\theta)|^r\frac{1}{\sigma_t^2(\theta)}\frac{\partial\sigma_t^2(\theta)}{\partial\theta'}\\ \frac{\partial^2\mu_r(\theta)}{\partial\theta\partial\theta'} &= \left(\frac{r}{2}+\frac{r^2}{4}\right)\frac{1}{n}\sum_{t=1}^{n}|\eta_t(\theta)|^r\frac{1}{\sigma_t^4(\theta)}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\frac{\partial\sigma_t^2(\theta)}{\partial\theta'}\\ &\quad -\frac{r}{2n}\sum_{t=1}^{n}|\eta_t(\theta)|^r\frac{1}{\sigma_t^2(\theta)}\frac{\partial^2\sigma_t^2(\theta)}{\partial\theta\partial\theta'}\end{aligned}$$

Since

$$|\eta_t(\theta)|^{2r} = \frac{\sigma_t^{2r}(\theta_0)}{\sigma_t^{2r}(\theta)}|\eta_t|^{2r},$$

the ergodic theorem, the Cauchy-Schwarz inequality and **A10** show that

$$\lim_{n\to\infty}\sup_{\theta\in V(\theta_0)}\left\|\frac{\partial^2\mu_r(\theta)}{\partial\theta\partial\theta'}\right\| < \infty.$$

Since, almost surely, $\theta^* \in V(\theta_0)$ for n large enough, we have

$$\lim_{n\to\infty}\left\|\frac{\partial^2\mu_r(\theta^*)}{\partial\theta\partial\theta'}\right\| = O_P(1).$$

Noting that, in view of (22), we have $\hat{\theta}_n^*-\theta_0^* = O_P(n^{-1/2})$, we obtain

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

which implies (25).

B.4. Complementary results for the standard GARCH

Proof of Theorem 3.2. To prove (11) we note that Assumptions **A4** and **A8** are satisfied with $\delta = r$. Assumptions **A6** and **A9** 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 by Equations (4.6) and (4.33) in Francq and Zakoïan (2004). The latter paper also established the second part of **A2** and **A10**. The convergence (11) follows from Theorem 2.1.

The expression of the two-step estimator $\tilde{\theta}_n$ follows from (5). The convergence in distribution (12) follows from Theorem 3.1. Let Γ_r denote the asymptotic variance in (12). To derive an explicit expression for Γ_r we use (9) and the following calculations. Denote by L the lag operator. The derivatives of $\sigma_t^2(\theta)$ verify

$$\begin{aligned} \mathcal{B}_\theta(L)\frac{\partial\sigma_t^2}{\partial\omega}(\theta) &= 1, \qquad \mathcal{B}_\theta(L)\frac{\partial\sigma_t^2}{\partial\alpha_i}(\theta) = \epsilon_{t-i}^2, \quad i = 1,\dots,q, \\ \mathcal{B}_\theta(L)\frac{\partial\sigma_t^2}{\partial\beta_j}(\theta) &= \sigma_{t-j}^2(\theta), \quad j = 1,\dots,p. \end{aligned} \tag{33}$$

In view of (5), $\mathcal{B}_{\theta_0}(L) = \mathcal{B}_{\theta_0^*}(L)$. Moreover $\sigma_{t-j}^2(\theta_0^*) = \mu_2\sigma_{t-j}^2(\theta_0)$. Thus

$$\frac{\partial\sigma_t^2(\theta_0^*)}{\partial\theta} = A\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}, \qquad A = \begin{pmatrix} I_{q+1} & 0 \\ 0 & \mu_2 I_p \end{pmatrix}. \tag{34}$$

It follows that

$$J_* = \mu_2^{-2}AJA, \qquad \Omega_* = \mu_2^{-1}A\Omega, \tag{35}$$

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

$$\Sigma_r = \begin{pmatrix} (\kappa_4-1)\mu_2^2A^{-1}J^{-1}A^{-1} & -\lambda_r\mu_2A^{-1}J^{-1}\Omega \\ -\lambda_r\mu_2\Omega'J^{-1}A^{-1} & \sigma_{\mu_r^*}^2 \end{pmatrix} \tag{36}$$

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

$$\left[\frac{\partial G_r(\theta_0^*,\mu_r^*)}{\partial(\theta',\mu)}\right] = \left[\frac{1}{\mu_2}A \quad \frac{2}{r}\mu_2^{\frac{r}{2}}\overline{\theta}_0\right]. \tag{37}$$

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

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

Now we will show that

$$J^{-1}\Omega = \overline{\theta}_0, \qquad \Omega'J^{-1}\Omega = 1 \tag{38}$$

The second equality follows from (35) and (8) but we give a direct proof. In view of (33), 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 β_j, the roots of the polynomial $\mathcal{B}_\theta(L)$ are outside the unit circle, it follows that

$$\frac{\partial \sigma_t^2(\theta_0)}{\partial \theta^{[1:q+1]'}}\, \theta_0^{[1:q+1]} \quad = \quad \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta'}\, \overline{\theta}_0 = \sigma_t^2(\theta_0), \tag{39}$$

The first equality in (38) follows. We also have $\Omega'\overline{\theta}_0 = 1$. The second equality in (38) follows. Because $\mu_2^{r/2} = 1/\kappa_r$, we thus have, by (38)

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

which completes the proof of (12). The theorem is established. □

THEOREM B.1 (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*

$$\sqrt{n}\left(\hat{\theta}_{n,h} - \theta_0\right) \overset{\mathcal{L}}{\to} \mathcal{N}\left\{0, 4\,Var(\log|\eta_0|)J^{-1}\right\}. \tag{40}$$

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

$$\sqrt{n}\left(\tilde{\theta}_n - \theta_0\right) \overset{\mathcal{L}}{\to} \mathcal{N}\left\{0, (\kappa_4 - 1)J^{-1} + \left[4\,Var(\log|\eta_0|) - (\kappa_4 - 1)\right]\overline{\theta}_0\overline{\theta}_0'\right\}. \tag{41}$$

Proof. We note that (40) does not straightforwardly follow from Theorem 2.1 because Assumptions **A4** and **A8** 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 B.1.

To prove (41), we use the following expansion, similar to (26),

$$\sqrt{n}(\hat{\mu}_0^* - \mu_0^*) \quad = \quad \frac{1}{\sqrt{n}}\sum_{t=1}^n(\log|\eta_t| - E\log|\eta_t|) - \frac{1}{2}\Omega_*'\sqrt{n}(\hat{\theta}_n^* - \theta_0^*) + o_P(1).$$

Moreover, 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 3.2. □

The link between Theorems 3.2 and B.1 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 3 (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 3.2 and B.1, 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 following is the analogue of Corollary 2 for the case $r = 0$.

Corollary 3 (A criterion for efficiency comparison when $r = 0$). *Under the assumptions of Theorem B.1, the asymptotic variance matrices of the two estimators verify*

$$Var_{as}\left\{\sqrt{n}\left(\hat{\theta}_{n,h}-\theta_0\right)\right\} \succeq Var_{as}\left\{\sqrt{n}\left(\tilde{\theta}_n-\theta_0\right)\right\} \tag{42}$$

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

$$4\,Var(\log|\eta_0|) \geq \kappa_4 - 1. \tag{43}$$

Proof of Corollaries 2 and 3. It follows from Theorem 3.2 that, for $r \neq 0$,

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

A similar result holds for $r = 0$, by Theorem B.1. It remains to show that

$$J^{-1} \succeq \overline{\theta}_0\overline{\theta}_0'. \tag{44}$$

In view of (39),

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

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

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

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 (44) is proven. □

Note that (44) 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.

B.5. Complementary results for the LAD estimation

We start by proving the following result, giving the joint asymptotic distribution of the LADE and $\hat{\mu}_r^{**}$. Let $\overline{\theta_0}^{**}$, J_{**} and Ω_{**} be defined as $\overline{\theta_0}$, J_* and Ω_* but with θ_0^{**} instead of θ_0^*.

Theorem B.2 (LADE for the Standard GARCH(p, q)). *When $r \neq 0$, under the assumptions of Theorem 4.1, $\hat{\theta}_n^{**} \to \theta_0^{**}$, $\hat{\mu}_r^{**} \to \mu_r^{**}$ a.s. and for any $\sqrt{n}$-consistent minimizer $(\hat{\theta}_n^{**})$ of the criterion defined in (15), we have*

$$\begin{pmatrix} \sqrt{n}\left(\hat{\theta}_n^{**} - \theta_0^{**}\right) \\ \sqrt{n}(\hat{\mu}_r^{**} - \mu_r^{**}) \end{pmatrix} \overset{\mathcal{L}}{\to} \mathcal{N}(0, \Sigma_r^*), \tag{45}$$

where

$$\Sigma_r^* = \begin{pmatrix} \xi_{\eta^2}^2 J_{**}^{-1} & -a_r\overline{\theta_0}^{**} \\ -a_r\overline{\theta_0}^{**\prime} & \sigma^2_{\mu_r^{**}} \end{pmatrix},$$

and

$$\begin{aligned} a_r &= \frac{\xi_{\eta^2}}{M^{r/2}}\left\{\frac{r}{2}\xi_{\eta^2} + \delta_r\right\}, \quad \delta_r = E(|\eta_t|^r \mathbb{1}_{\eta_t^2 > M}) - E(|\eta_t|^r \mathbb{1}_{\eta_t^2 < M}), \\ \sigma^2_{\mu_r^{**}} &= \frac{1}{M^r}\left\{E|\eta_t|^{2r} - 1 + r\xi_{\eta^2}\left(\frac{r}{4}\xi_{\eta^2} + \delta_r\right)\right\}. \end{aligned}$$

When $r = 0$, under the assumptions of Theorem B.1, the previous results hold with

$$\begin{aligned} a_0 &= \xi_{\eta^2}\left\{\frac{1}{2}\xi_{\eta^2} + \delta_0\right\}, \quad \delta_0 = E(\log|\eta_t| \mathbb{1}_{\eta_t^2 > M}) - E(\log|\eta_t| \mathbb{1}_{\eta_t^2 < M}), \\ \sigma^2_{\mu_0^{**}} &= Var(\log|\eta_0|) + \xi_{\eta^2}\left(\frac{1}{4}\xi_{\eta^2} + \delta_0\right). \end{aligned}$$

Proof. Proceeding as in the proof of Theorem 2.1, to prove the strong consistency of $\hat{\theta}_n^{**}$ we check the intermediate results

$i)$ $\lim_{n\to\infty} \sup_{\theta\in\Theta} |Q_n(\theta) - \tilde{Q}_n(\theta)| = 0$, *a.s.*

$ii)$ if $\theta \neq \theta_0$, $\mathbb{E}|\log\epsilon_t^2 - \log\sigma_t^2(\theta)| < \mathbb{E}|\log\epsilon_t^2 - \log\sigma_t^2(\theta_0)|$,

$iii)$ any $\theta \neq \theta_0$ has a neighborhood $V(\theta)$ such that

$$\limsup_{n\to\infty} \sup_{\vartheta\in V(\theta)} \tilde{Q}_n(\vartheta) < \limsup_{n\to\infty} \tilde{Q}_n(\theta_0), \quad a.s.$$

with

$$\tilde{Q}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} |\log \epsilon_t^2 - \log \tilde{\sigma}_t^2(\theta)|, \quad Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} |\log \epsilon_t^2 - \log \sigma_t^2(\theta)|.$$

We have, using the elementary inequality $||z-y|-|z|| \leq |y|$ and already used arguments,

$$\sup_{\theta\in\Theta} |Q_n(\theta) - \tilde{Q}_n(\theta)| \leq \frac{1}{n}\sum_{t=1}^{n} \sup_{\theta\in\Theta} |\log \sigma_t^2(\theta) - \log \tilde{\sigma}_t^2(\theta)| \leq \frac{1}{n}\sum_{t=1}^{n} C\rho^t \leq \frac{C}{n},$$

which proves i). Now for $z \neq 0$,

$$|z-y| - |z| = -y\text{sgn}(z) + 2(y-z)(\mathbb{1}_{0<z<y} - \mathbb{1}_{y<z<0}) \geq -y\text{sgn}(z) \tag{46}$$

with equality only if $y = z$. Hence, for any $\sigma > 0$,

$$E|\log|\eta_0^{**}| - \log\sigma| - E|\log|\eta_0^{**}|| \geq -(\log\sigma)E\{\text{sgn}(\log|\eta_0^{**}|)\} = 0,$$

the inequality being strict unless if $\sigma = 1$, the equality following from the fact that median$(|\eta_0^{**}|) = 1$.

Hence, for any $\sigma > 0$,

$$E|\log|\eta_0^{**}| - \log\sigma| - E|\log|\eta_0^{**}|| \geq -(\log\sigma)E\{\text{sgn}(\log|\eta_0^{**}|)\} = 0,$$

the equality following from the fact that median$(|\eta_0^{**}|) = 1$. Now we show that the inequality is strict whence $\sigma \neq 1$. For instance, let $\sigma > 1$ and suppose $E|\log|\eta_0^{**}| - \log\sigma| - E|\log|\eta_0^{**}|| = 0$, that is, in view of (46), $E\left\{(\log\sigma - \log|\eta_0^{**}|)\mathbb{1}_{0<\log|\eta_0^{**}|<\log\sigma}\right\} = 0$. Then, because the variable under the expectation is nonnegative, $\log\sigma = \log|\eta_0^{**}|$ or $\mathbb{1}_{0<\log|\eta_0^{**}|<\log\sigma} = 0$. a.s. We are led to a contradiction, because $\log|\eta_0^{**}|$ has a positive density on some interval $(0,\epsilon)$ with $\epsilon > 0$. The case $\sigma < 1$ can be handled similarly. Result ii) straightforwardly follows and the proof of iii) being standard, it is omitted.

Now we turn to the asymptotic normality. Following the lines of proof of Davis, Knight and Liu (Lemma 2.2 and Remark 1, 1992), it can be shown that

$$\begin{aligned}\sqrt{n}\left(\hat{\theta}_n^{**} - \theta_0^{**}\right) &= -\xi_{\eta^2} J_{**}^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\mathbb{1}_{\eta_t^2 > M} - \mathbb{1}_{\eta_t^2 < M}) \frac{1}{\sigma_t^{**2}} \frac{\partial \sigma_t^2(\theta_0^{**})}{\partial\theta} + o_P(1) \\ &\stackrel{\mathcal{L}}{\to} \mathcal{N}(0, \xi_{\eta^2}^2 J_{**}^{-1}).\end{aligned} \tag{47}$$

This asymptotic distribution was obtained by Peng and Yao (2003) under similar assumptions, in particular **C0**. Similar to (26) we have

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

We have

$$\text{Cov}\left(\sqrt{n}\left(\hat{\theta}_n^{**} - \theta_0^{**}\right), \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (|\eta_t|^r - 1)\right) = -\delta_r \xi_{\eta^2} J_{**}^{-1} \Omega_{**} + o(1).$$

It follows that

$$\mathrm{Cov}\left(\sqrt{n}\left(\hat{\theta}_n^{**}-\theta_0^{**}\right),\sqrt{n}(\hat{\mu}_r^{**}-\mu_r^{**})\right) \quad = \quad -a_r J_{**}^{-1}\Omega_{**}+o(1).$$

We also have

$$\mathrm{Var}\left(\sqrt{n}(\hat{\mu}_r^{**}-\mu_r^{**})\right) \quad = \quad \frac{1}{M^r}\left\{E|\eta_t|^{2r}-1+r\xi_{\eta^2}\left(\frac{r\xi_{\eta^2}}{4}+\delta_r\right)\right\}+o(1).$$

using (8), which obviously holds with θ_0^* replaced by θ_0^{**}.

The case $r=0$ is handled similarly, using the expansion

$$\sqrt{n}(\hat{\mu}_0^{**}-\mu_0^{**}) \quad = \quad \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\log|\eta_t|-E\log|\eta_t|)-\frac{1}{2}\Omega_{**}'\sqrt{n}(\hat{\theta}_n^{**}-\theta_0^{**})+o_P(1).$$

□

Proof of Theorem 4.1 and Remark 9. The proof of Theorem 4.1 follows the lines of proof of Theorem 3.2. Note that, because $\eta_t^{**}=\frac{\eta_t}{\sqrt{M}}$, we have, similar to 36 and (37),

$$\Sigma_r \quad = \quad \begin{pmatrix} \xi_{\eta^2}^2 M^2(A^*)^{-1}J^{-1}(A^*)^{-1} & -a_r M(A^*)^{-1}J^{-1}\Omega \\ -a_r M\Omega' J^{-1}(A^*)^{-1} & \sigma^2_{\mu_r^{**}} \end{pmatrix},$$

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

Using (38), the asymptotic variance

$$\left[\frac{\partial G_r(\theta_0^{**},\mu_r^{**})}{\partial(\theta',\mu)}\right]\Sigma_r^*\left[\frac{\partial G_r(\theta_0^{**},\mu_r^{**})'}{\partial(\theta',\mu)'}\right]$$

of the reparameterized LADE of the two-step approach follows. Thus, using (11),

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

and the equivalence between (16) and (17) is deduced from (44).

Remark 9 follows from Theorem B.2 and

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

□

The ARE of the one-step QMLE relative to the two-step LADE only depends on the innovations distribution and is illustrated in Figure 3.

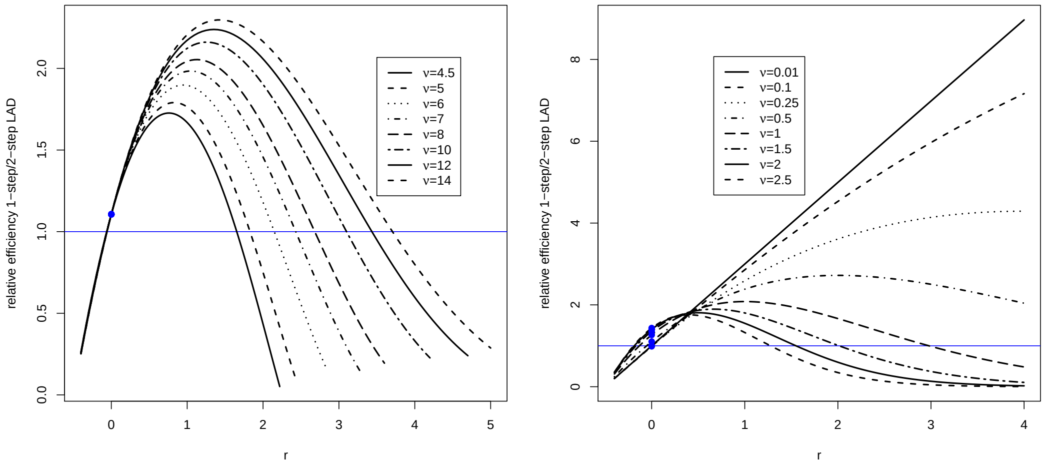

Fig. 3. Same as Figure 1 for the two-step LAD estimator.

B.6. Procedure in the case $r = 0$

When $r = 0$, that is for the prediction of $\log|\epsilon_{n+1}|$, the procedure can be modified as follows, starting from Step 2, in view of Remarks 7 and 9.

Step 2. Compute the rescaled residuals $\hat{\eta}_t^* = \frac{\epsilon_t}{\hat{\sigma}_t^*}$. Compute the empirical moment $\hat{\mu}_4^*$ of these residuals, and the empirical mean and variance of the log-absolute residuals, $\widehat{E}(\log|\eta|)$ and $\widehat{\mathrm{Var}}(\log|\eta|)$ respectively. Compute the empirical median $\hat{M}^*$, and estimate the density $f_{\eta^{*2}}$ of the $\hat{\eta}_t^{*2}$.

Step 3. Compute the quantities

$$c_0 = \frac{\hat{\mu}_4^*}{\hat{\mu}_2^{*2}} - 1, \qquad c_1 = 4\widehat{\mathrm{Var}}(\log|\eta|), \qquad c_2 = \frac{1}{\{2\hat{M}^* \hat{f}_{\eta^{*2}}(\hat{M}^*)\}^2}.$$

(i) If $c_0 = \min_{i=0,1,2} c_i$, then the Gaussian QMLE can be preferred for the prediction of $\log|\epsilon_{n+1}|$. The prediction is computed as $\log\hat{\sigma}_n + \widehat{E}(\log|\eta|)$.

(ii) If $c_2 = \min_{i=0,1,2} c_i$, then the LADE can be preferred. Reestimate the model by LAD, and compute the $\hat{\eta}_t^{**} = \frac{\epsilon_t}{\hat{\sigma}_t^{**}}$. Compute the empirical mean $\widehat{E}(\log|\eta^{**}|)$ of the log-absolute residuals. The prediction of $|\epsilon_{n+1}|^r$ is $\log\hat{\sigma}_n^{**} + \widehat{E}(\log|\eta^{**}|)$.

(iii) If $c_1 = \min_{i=0,1,2} c_i$, then the one-step estimator can be preferred. Reestimate the model by non-gaussian QML, by minimizing

$$\sum_{t=1}^{n} \left\{ \log \frac{|\epsilon_t|}{\tilde{\sigma}_t(\theta)} \right\}^2.$$

The prediction of $|\epsilon_{n+1}|^r$ is $\log\tilde{\sigma}_n$.

B.7. Comparison of the two-step QMLE and LADE

We have seen that the asymptotic relative efficiency of the two-step QML with respect to the two-step LAD method for predicting $|\epsilon_{n+1}|^r$ does not depend on r. To see how it depends

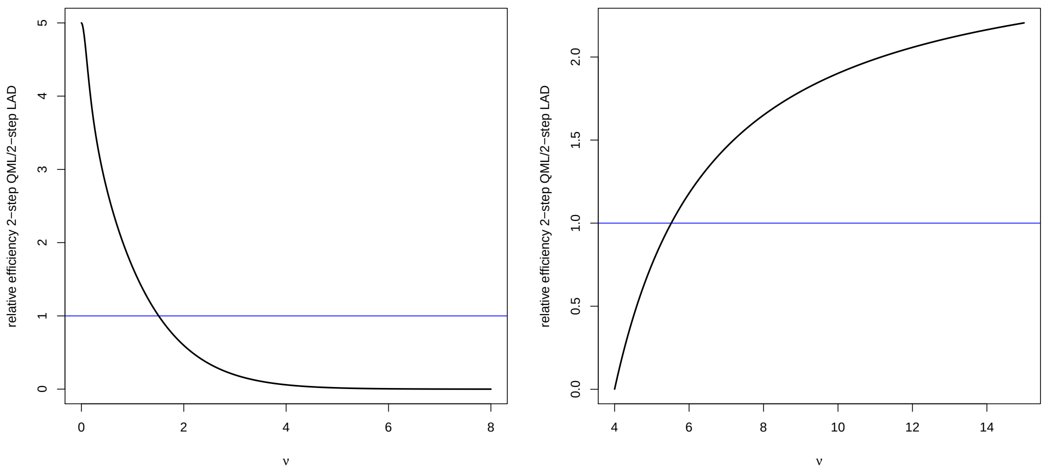

Fig. 4. ARE of the two-step QMLE relative to the two-step LAD for GED (left panel) and Student distributions (right panel) in function of ν.

on the distribution of η_t, we show the ARE for GED and Student distributions in Figure 4. As expected, the right graph shows that the LADE is preferable for fat tailed distributions. The left graph shows that for distributions close to the Gaussian or the Laplace, the QML is asymptotically better; when ν increases, the tails increase for the GED and the LADE tends to be the best.

B.8. Simulation experiments

The first set of simulation experiments aims to compare the effective relative efficiencies of the 1-step method based on the generalized QMLE, the 2-step method based on the QMLE and the 2-step method based on the LADE. We simulated $N = 500$ independent trajectories of size $n = 10200$ of GARCH(1,1) models $\epsilon_t = \sigma_t \eta_t$, in which η_t follows the generalized error distribution with ν degrees of freedom, GED(ν). The GARCH parameters have been chosen in such a way to obtain, for all values of r and ν, the usual parameterization

$$\begin{cases} \epsilon_t = \sigma_t^* \eta_t^*, \quad E\eta_t^{*2} = 1 \\ \sigma_t^{*2} = 1 + 0.05\epsilon_{t-1}^2 + 0.7\sigma_{t-1}^{*2}. \end{cases}$$

For each of the N simulations, the first $n_1 = 200$ values were used to estimate the GARCH parameters and the moments required by the last two methods. The last $n_2 = 10000$ values were used to compute the mean square prediction errors (MSPE) given, in the case $r \neq 0$, by

$$\frac{1}{n_2} \sum_{t=n_1}^{n-1} \left(|\epsilon_{t+1}|^r - P_{t,h} \right)^2, \quad \frac{1}{n_2} \sum_{t=n_1}^{n-1} \left(|\epsilon_{t+1}|^r - P_t^* \right)^2, \quad \frac{1}{n_2} \sum_{t=n_1}^{n-1} \left(|\epsilon_{t+1}|^r - P_t^{**} \right)^2$$

where $P_{t,h} = \tilde{\sigma}^r(\epsilon_n, \dots, \epsilon_1; \hat{\theta}_{n_1,h})$ are the one-step predictions and $P_t^* = \tilde{\sigma}^r(\epsilon_t, \dots, \epsilon_1; \hat{\theta}_{n_1}^*)\hat{\mu}_r^*$ and $P_t^{**} = \tilde{\sigma}^r(\epsilon_t, \dots, \epsilon_1; \hat{\theta}_{n_1}^{**})\hat{\mu}_r^{**}$ are the two-step predictions based respectively on the

Table 4. Percentages of MSPE losses with respect to the unfeasible optimal predictor, for predicting the power r of a GARCH(1,1) model with $\eta_t \sim GED(\nu)$.

r	$\nu = 0.1$				$\nu = 0.5$			
	Naive	1-step	2-QML	2-LAD	Naive	1-step	2-QML	2-LAD
0	21.3	0.8	0.6	0.8	33.1	0.9	0.7	0.9
0.5	17	1	0.8	1.2	27.1	1.1	1	1.3
1	8.6	1	0.9	1.6	13.4	1.2	1.1	1.6
2	1.2	1.2	1.2	3.4	1.3	1.3	1.3	2.8
r	$\nu = 2$				$\nu = 3$			
	Naive	1-step	2-QML	2-LAD	Naive	1-step	2-QML	2-LAD
0	90.4	1.1	1.9	1.3	132.8	1.2	3	1.2
0.5	95.7	1.6	3	3.7	168.5	1.4	7	1.8
1	38.3	1.7	2.8	2.5	162.7	5.3	19.1	3.4
2	5.7	5.7	5.7	6.4	45.4	45.4	45.4	8.3

Gaussian QMLE and the LADE. These MSPE have been compared to those obtained with the naive predictor $\tilde{\sigma}^r(\epsilon_t, \dots, \epsilon_1; \hat{\theta}^*_{n_1})$ and with the (approximated[9]) optimal predictor $\tilde{\sigma}^r(\epsilon_n, \dots, \epsilon_1; \theta_0)$. As expected, the minimal MSPE were always obtained with the optimal predictor. Table 4 gives the percentages of relative MSPE losses with respect to the optimal predictor over the N replications. As expected, the naive predictor entails important efficiency losses, except in the case $r = 2$ where this method coincides with the 1-step GQMLE and with the 2-step QMLE. Obviously, Table 4 confirms that the naive method must be avoided. In accordance with the asymptotic theory (see Remark 8), the ranking of the 2-step QML and 2-step LAD methods does not depend on r, the LADE is preferred when ν is large (*i.e.* $\nu = 3$) and the QML is slightly better when $\nu = 0.1$ or $\nu = 0.5$, whereas the two methods are almost equivalent when $\nu = 2$. As expected from Figures 1-3, the 1 step method is the best for all the values of ν and r considered in Table 4, except for $r = 2$ and $\nu = 3$ where the method based on the LAD is much more efficient.

In a second set of simulation experiments, we assess the effective performance of the adaptive method. We simulated $N = 500$ independent trajectories of size $n = 10500$ of a GARCH(1,1) model, with 3 different designs for the parameter $(\omega_0, \alpha_0, \beta_0)$ and for the distribution of the noise η_t. For each trajectory, the first 500 observations are used to estimate the GARCH models and the relevant moments of the noise, whereas the last 10000 values are used to compute the percentages of MSPE for predicting a given power r by the different methods. In Design A, η_t follows a multimodal distribution, a mixture of 3 normal distributions of the form

$$\eta_t \sim \pi\phi(x) + \frac{1-\pi}{2}\phi(x+m) + \frac{1-\pi}{2}\phi(x-m), \tag{48}$$

where $\phi(\cdot)$ denotes the standard gaussian density. We took $\pi = 1/2$ and $m = 10$, so that the distribution of η_t is such that c_2 is much greater than c_0 and c_1. Table 5 displays the percentages of MSPE losses with respect to the best prediction method among the four methods employed, that is, the naive, 1-step, 2-step and adaptive procedures. For Design A, we chose $\alpha_0 = 0.10$ and $\beta_0 = 0.8$ (the results are not sensitive to the value of ω_0), and we took $r = 1.5$. Of course, similar results are obtained for other choices of the parameters.

[9]This estimator is said to be an "approximation" of the optimal predictor because it is based on a finite number of past values. It is introduced as a benchmark but it can not be used in practice because θ_0 is unknown.

Table 5. Percentages of MSPE losses with respect to the asymptotically optimal prediction method.

	Naive	1-step	2-step QML	2-step LAD	Adaptive
Design A	2.9	0.2	0	66.3	0
Design B	418.4	0	10.2	1.9	1.7
Design C	1088.9	0	43	2.5	1.6

Table 6. Number of choices of each method by the adaptive method.

	1-step	2-step QML	2-step LAD
Design A	500	0	0
Design B	1	84	415
Design C	37	136	327

Table 5 confirms that, as expected the LADE is much less efficient than the other methods for that design. In Design B, the distribution of η_t is chosen to be the Cauchy distribution. We also chose $r = 0.45$, so that the moment of order $2r$ exist, as required with the 1-step method. Since the dispersion of the noise is high, the GARCH parameter should be chosen smaller than in Design A to obtain a strict stationary solution. We thus took $\alpha_0 = 0.005$ and $\beta_0 = 0.8$. This design should penalize the 2-step QML method because the required moments do not exist. Indeed, Table 5 shows that the 2-step QML method induces an important efficiency loss. In Design C keeps the parameters of Design B, except that η_t follows a mixture distribution of the form (48), where $\phi(\cdot)$ is replaced by the Cauchy density. As expected, the best method is the 1-step method in that case. Interestingly, the adaptive method is always close to the optimal method in terms of MSPE. Table 6 gives the number of times that each method is selected by the adaptive method. In the framework of Design A, the adaptive method always makes the right choice. In Designs B and C, the adaptive method often makes suboptimal choices, but the MSPE is however close to the optimal.

B.9. Complementary empirical results

In this section we come back to the prediction problem of the daily returns of the 10 world stock market indices of Section 5. We study the sensitivity of the results to i) a change of model, and to ii) a change of period for the data sets. Because stationarity is crucial for our results, we start by considering this issue.

B.9.1. Stationarity of returns

Figure 5 displays the sample paths of the CAC, DAX, FTSE, Nikkei, SMI and SP500 prices and returns. Similar graphs were obtained for the 4 other indices. While the non stationarity of prices is clear from these drawings, the sample paths of returns are compatible with stationarity. This is confirmed by Figure 6 showing the empirical autocorrelations of such returns. The significance bands computed for a GARCH(1,1)[10] show that these autocorrelations are compatible with a GARCH(1,1) model for the returns. In the GARCH(1,1) framework, a formal test of strict stationarity can be done. Applying the test developed by

[10]See Francq and Zakoian (2009). The R-code can be downloaded at `http://www.runmycode.org/CompanionSite/site.do?siteId=23`.

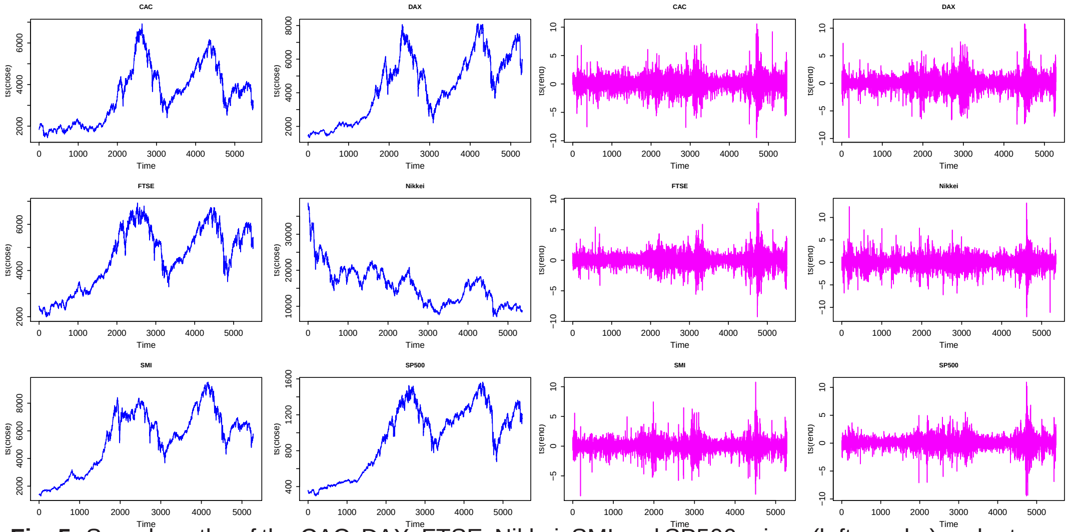

Fig. 5. Sample paths of the CAC, DAX, FTSE, Nikkei, SMI and SP500 prices (left graphs) and returns (right graphs).

Francq and Zakoian (2012), we conclude that the stationarity cannot be rejected, at any reasonable significance levels, for the returns.

B.9.2. Complement to Table 3

For $n_1 = n_2 = n_3 = 300$, Table 7 displays the percentages of prediction losses with respect to the best method, for $r = -0.5$, $r = 0$, $r = 0.5$, $r = 1$, $r = 1.5$ and $r = 2$.

B.9.3. Using a standard GARCH instead of a TARCH

We first re-estimate the ARE of the methods assuming a standard GARCH(1,1) model

$$\epsilon_t = \sigma_t \eta_t, \quad \sigma_t^2 = \omega_0 + \alpha_0 \epsilon_{t-1}^2 + \beta_0 \sigma_{t-1}^2, \tag{49}$$

instead of the TGARCH(1,1) model (49). Figure 7 is very similar to Figure 2, which leads again to the conclusion that the one-step method is often the most efficient when $r \in (0.5, 2)$, but is always dominated when $r > 2$ or r is small.

Note that Figures 2 and 7 do not directly compare the ARE of the 2-step LAD with respect to the 2-step Gaussian GMLE. These ARE do not depend on r, but just on the distribution of the iid noise. Table 8 indicates that the ranking of the 2 method may depend on the volatility model, but for the 5 European indices the LADE is often expected to be more efficient than the QMLE, whereas this is the opposite for the 5 other indices.

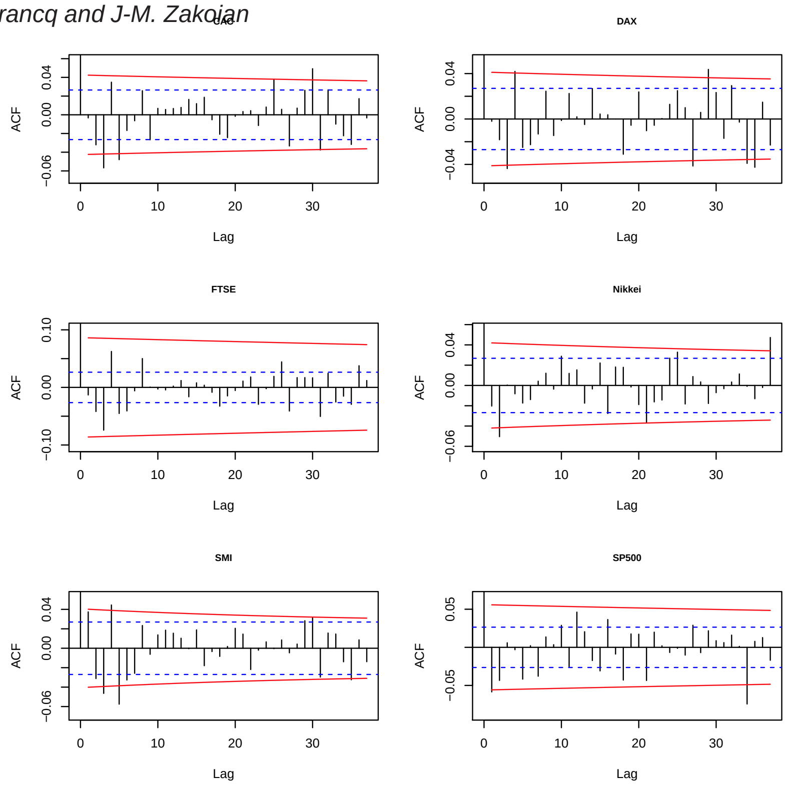

Fig. 6. Empirical autocorrelation functions of the CAC, DAX, FTSE, Nikkei, SMI and SP500 returns. Standard significance bands (at the level 95%) of a strong white noise are displayed in dotted lines, significance bands for a GARCH(1,1) process are displayed in solid lines.

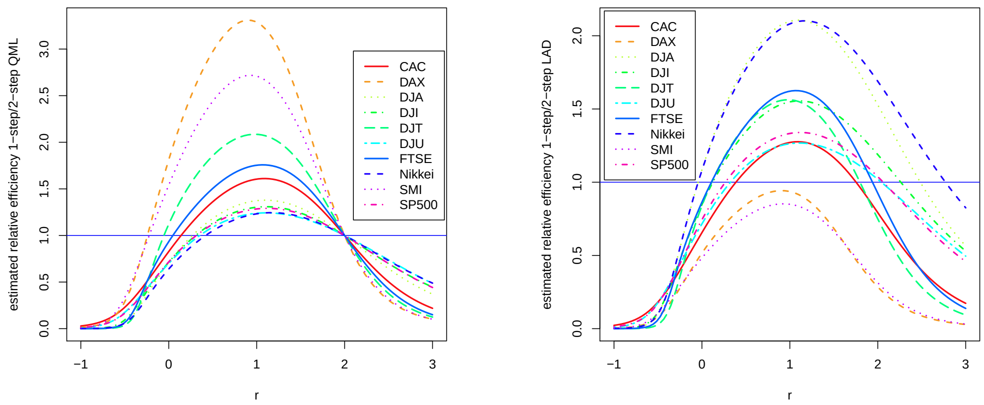

Fig. 7. Estimated ARE's of the one-step QMLE relative to the two-step QMLE (left panel) and LAD estimator (right panel) for stock index returns. As Figure 2, but the volatility model is the GARCH(1,1) model (49) instead of the TGARCH(1,1) model (49).

Table 7. Percentages of MSPE losses with respect to the best method, for prediction of $|\epsilon_{n+1}|^r$.

r	-0.5	-0.5	-0.5	0	0	0	0.5	0.5	0.5
	Naive	Historic	Adaptive	Naive	Historic	Adaptive	Naive	Historic	Adaptive
CAC	12.9	1.3	0	28.6	7.3	0	19.2	14.9	0
DAX	8.1	0.8	0	27.8	6.6	0	17.8	13.7	0
DJA	4.8	0.5	0	33.3	7.2	0	24.3	16.8	0
DJI	7.8	0.8	0	34.8	7.8	0	25.3	17.8	0
DJT	1.3	0	0.3	30.2	2.8	0	21.4	6.9	0
DJU	11.2	0.6	0	30.9	5.1	0	21	13.3	0
FTSE	5.4	0	0	29.3	6.9	0	20.5	15.6	0
Nikkei	4.8	0.4	0	33.3	4.7	0	26.8	10.6	0
SMI	12.5	1.7	0	32.6	9.1	0	22.2	18.8	0
SP500	11.3	1	0	34	7.6	0	22.8	18	0
r	1	1	1	1.5	1.5	1.5	2	2	2
	Naive	Historic	Adaptive	Naive	Historic	Adaptive	Naive	Historic	Adaptive
CAC	6	21	0	0.2	22.4	0	0	19.3	0
DAX	4.5	20.8	0	0	24.9	0.9	0	21.7	0.1
DJA	7.4	25.1	0	0.5	27.4	0	7.2	29.3	0
DJI	7.7	26.9	0	0	32.4	0.4	0	31	0
DJT	6.6	9.1	0	1	8.7	0	0.5	4.8	0
DJU	5.7	23.5	0	0	30.2	0.7	0	28.8	0
FTSE	6	23.6	0	0	28.7	0.5	0	26.1	0
Nikkei	10	16.7	0	0.7	22.5	0	0	25.9	0.1
SMI	6.5	28.9	0	0	35.9	0.7	0	38.3	0.2
SP500	5.8	27.7	0	0	33.1	0.8	2.9	30.8	0

Table 8. Estimates of the ARE's c_0/c_2 of the LAD method with respect to the Gaussian QML method for predicting powers of the stock index returns.

	CAC	DAX	DJA	DJI	DJT	DJU	FTSE	Nikkei	SMI	SP500
TGARCH	1.26	3.51	0.65	0.84	1.33	0.98	1.08	0.59	3.19	0.96
GARCH	0.99	3.05	0.74	0.60	0.94	0.83	0.68	0.70	1.32	0.76

Table 9. As Table 7, but the volatility is assumed to follow a standard GARCH(1,1) instead of a TGARCH(1,1).

r	-0.5	-0.5	-0.5	0	0	0	0.5	0.5	0.5
	Naive	Historic	Adaptive	Naive	Historic	Adaptive	Naive	Historic	Adaptive
CAC	12.7	0	0.4	25.8	0	0.2	18	3.4	0
DAX	8.8	0.6	0	33.5	5.9	0	23.9	11.8	0
DJA	4.9	0.3	0	35.6	5.3	0	26.8	12.9	0
DJI	7.7	0.3	0	34.1	4.5	0	26.4	12.7	0
DJT	1.3	0	0.1	33.1	3.1	0	25	7.4	0
DJU	12.5	0.5	0	33.6	5.7	0	25.1	13.8	0
FTSE	5.4	0	0.1	28.1	5	0	19.9	11.3	0
Nikkei	4.8	0.1	0	36.7	3.3	0	32.1	9.2	0
SMI	12.5	0.9	0	33.7	6.4	0	23.8	13.6	0
SP500	11.3	0.3	0	34.7	4.4	0	24.2	12.3	0
r	1	1	1	1.5	1.5	1.5	2	2	2
	Naive	Historic	Adaptive	Naive	Historic	Adaptive	Naive	Historic	Adaptive
CAC	5	7.1	0	0	9.8	0.1	0	9.6	0
DAX	8.5	16.7	0	1.1	18.1	0	0.2	12.7	0
DJA	9.3	20.1	0	0.7	22.3	0	0	19	0
DJI	8.1	15.7	0	1.8	15.9	0	0	18.5	1.1
DJT	9.9	11.1	0	1.8	9.9	0	0	5.2	0.7
DJU	8.9	23.5	0	0.9	28.3	0	0	24.2	0
FTSE	7.3	16.6	0	1.1	20.1	0	0	19.7	0.1
Nikkei	13.4	14.9	0	3	20.1	0	0.1	20.9	0
SMI	8.2	20.8	0	1.2	24.7	0	0.5	25	0
SP500	7.6	19.9	0	0.6	22.6	0	0.3	20.5	0

Table 9 and Table 7 are very similar, except that the relative MSPE losses of the Naive and Historic methods are globally more important for the TGARCH than for the standard GARCH. The Historic method being model-free, larger losses with respect to the adaptive method based on the TGARCH than with the one based on the standard GARCH is an indicator that the TGARCH model does a slightly better job for the predictions. For the two models, the adaptive method is clearly the most efficient.

B.9.4. Using a subperiod of the data set

Figure 8 and Tables 10 are respectively similar to Figure 7 and Table 9, except that the data cover the period from January 2, 1990, to January 22, 2009. The results are not much affected by the fact that the period does not include anymore the recent sovereign-debt crises in Europe and US.

Table 11 presents results similar to those of Table 10 but for empirical means based on 30 returns (instead of 250) in the historical method. For $r = -0.5$ the results worsen. For positive values of r the results are generally better with 30 observations, but the adaptive method remains superior (the percentages being always equal or very close to zero).

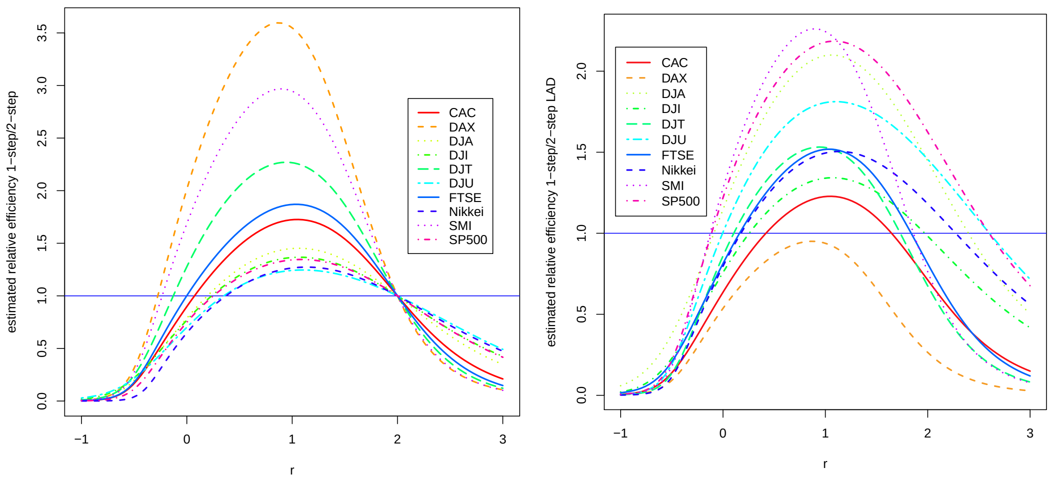

Fig. 8. As Figure 7, but for the period from January 2, 1990, to January 22, 2009.

Table 10. As Table 9, but for the period from January 2, 1990, to January 22, 2009.

r	-0.5	-0.5	-0.5	0	0	0	0.5	0.5	0.5
	Naive	Historic	Adaptive	Naive	Historic	Adaptive	Naive	Historic	Adaptive
CAC	10.3	0.2	0	28.9	1.6	0	22.2	7.1	0
DAX	9.5	0.9	0	32.5	6.5	0	23.1	14.5	0
DJA	17.9	0.1	0	34.2	3.9	0	19.6	8.1	0
DJI	15.2	0.1	0	32.3	2.6	0	19.1	6.7	0
DJT	9.5	0	0	32.1	1.8	0	19.8	4.5	0
DJU	9.2	0.3	0	30.5	3.2	0	21.1	7.6	0
FTSE	13.7	0.5	0	30.6	4.1	0	22.4	8.2	0
Nikkei	7.3	0.5	0	35.8	4	0	30.2	10.7	0
SMI	9.1	0.8	0	32.2	7.2	0	22.8	17.3	0
SP500	9.3	0	0.2	31.5	3.4	0	14.8	8.7	0

r	1	1	1	1.5	1.5	1.5	2	2	2
	Naive	Historic	Adaptive	Naive	Historic	Adaptive	Naive	Historic	Adaptive
CAC	6.3	12.3	0	0	15.1	0.3	0	11.5	0
DAX	8	19.9	0	1.1	19.3	0	0	12.4	0.1
DJA	3.5	10.2	0	0	10.1	1.1	0	6.7	0.2
DJI	2.8	5	0	0	2.7	0	0	5.2	0.5
DJT	6	6.7	0	1	6.1	0	0.2	2.5	0
DJU	6.7	15.6	0	0.4	22.7	0	0	21.8	0
FTSE	9.1	10.7	0	2.1	12.3	0	0	11.2	0.1
Nikkei	11.9	17.7	0	2.4	23.3	0	0.1	24.3	0
SMI	7.8	26.9	0	1.4	30.9	0	0.3	30.7	0
SP500	0.5	9.8	0	0	8.6	1.5	0	4.2	0.1

Table 11. As in Table 10 but for empirical means based on the last 30 returns (historical method).

r	-0.5	-0.5	-0.5	0	0	0	0.5	0.5	0.5
	Naive	Historic	Adaptive	Naive	Historic	Adaptive	Naive	Historic	Adaptive
CAC	10.3	2.2	0	30	0	0.9	23.1	0	0.8
DAX	9.5	2.6	0	32.5	1.1	0	23.1	0.7	0
DJA	17.9	1.6	0	34.2	1	0	20.7	0	0.9
DJI	15.2	1.5	0	34.4	0	1.5	21.4	0	1.9
DJT	9.5	2.3	0	32.1	0	0	22.1	0	1.9
DJU	9.2	2.6	0	31.3	0	0.6	23.2	0	1.8
FTSE	13.7	1.7	0	30.6	0.4	0	23.1	0	0.6
Nikkei	7.3	3	0	35.8	0.4	0	30.2	2	0
SMI	9.1	2.1	0	32.2	0.8	0	22.8	2.1	0
SP500	9.1	1.7	0	33	0	1.1	15.8	0	0.9
r	1	1	1	1.5	1.5	1.5	2	2	2
	Naive	Historic	Adaptive	Naive	Historic	Adaptive	Naive	Historic	Adaptive
CAC	6.5	0	0.1	0	1.5	0.3	0.1	0	0.1
DAX	8	1.9	0	1.1	2.8	0	0	0.2	0.1
DJA	4.8	0	1.3	0	0.5	1.1	0	0.3	0.2
DJI	6.9	0	4	4.3	0	4.3	0	0.6	0.5
DJT	8.3	0	2.2	3.1	0	2	2.5	0	2.3
DJU	6.7	2.3	0	0.4	8.1	0	0	9.4	0
FTSE	9.9	0	0.7	2.1	1.1	0	0	2.3	0.1
Nikkei	11.9	5.8	0	2.4	10.9	0	0.1	14.1	0
SMI	7.8	6.6	0	1.4	11.8	0	0.3	16.5	0
SP500	0.6	0	0.1	0	1.8	1.5	0	1.2	0.1

B.9.5. Duration models

The dynamics of duration between stock price changes has attracted much attention in the econometrics literature. Engle and Russell (1998) proposed the Autoregressive Conditional Duration (ACD) model, which assumes that the duration between price changes has the dynamics of the square of a GARCH:

$$\begin{cases} x_i = \psi_i \eta_i, \quad (\eta_i) \sim iid \\ \psi_i = \omega_0 + \sum_{k=1}^{q} \alpha_{0k} x_{i-k} + \sum_{j=1}^{p} \beta_{0j} \psi_{i-j} \end{cases} \tag{50}$$

with $\omega_0 > 0, \alpha_{0k} \geq 0$, $\beta_{0j} \geq 0$. An alternative specification that does not constrain the sign of the coefficients is the logarithmic ACD proposed by Bauwens and Giot (2000), which can be written as follows:

$$\begin{cases} x_i = e^{\phi_i} \eta_i, \quad (\eta_i) \sim iid \\ \phi_i = \omega_0 + \sum_{k=1}^{q} \alpha_{0k} \log x_{i-k} + \sum_{j=1}^{p} \beta_{0j} \phi_{i-j} \end{cases} \tag{51}$$

It is clear that both ACD models are of the form (1). Figure 9 displays the empirical autocorrelation functions for the absolute returns of the SP500, the inverse absolute returns of the SP500, IBM durations data, and the inverse IBM durations. For the stock index the absolute returns appear strongly autocorrelated, showing that a GARCH-type model is compatible with the series. For the inverse returns, a GARCH model would not be compatible with the autocorrelations.

B.10. The Asymmetric Power GARCH(p, q) case

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 (14). This condition entails the invertibility of the polynomial $\mathcal{B}_{\theta_0}(z)$ and allows to write the model under the form (1). It also ensures the existence of $E|\epsilon_t|^s$ for some $s > 0$.

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

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

Hamadeh and Zakoïan (2011) showed that the following assumption entails AN of the Gaussian QMLE of $\theta_0 = (\omega_0, \alpha_{01+}, \dots, \alpha_{0q-}, \beta_{01}, \dots, \beta_{0p})'$.

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 Γ, then Γ 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 B.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 (11).*

Under the same assumptions and $E\eta_0^4 < \infty$, the two-step estimator is given by $\tilde{\theta}_n = (\{\hat{\mu}_r^*\}^{\delta/r}\hat{\omega}^*, \{\hat{\mu}_r^*\}^{\delta/r}\hat{\alpha}_{1+}^*, \dots, \{\hat{\mu}_r^*\}^{\delta/r}\hat{\alpha}_{q-}^*, \hat{\beta}_1^*, \dots, \hat{\beta}_p^*)$ and satisfies

$$\sqrt{n}\left(\tilde{\theta}_n - \theta_0\right)$$

$$\overset{\mathcal{L}}{\to} \mathcal{N}\left\{0, (\kappa_4 - 1)J^{-1} + \left(\frac{\delta}{2}\right)^2 \left[\left(\frac{2}{r}\right)^2 \left(\frac{\kappa_{2r}}{\kappa_r^2} - 1\right) - (\kappa_4 - 1)\right] \overline{\theta}_0 \overline{\theta}_0'\right\}$$

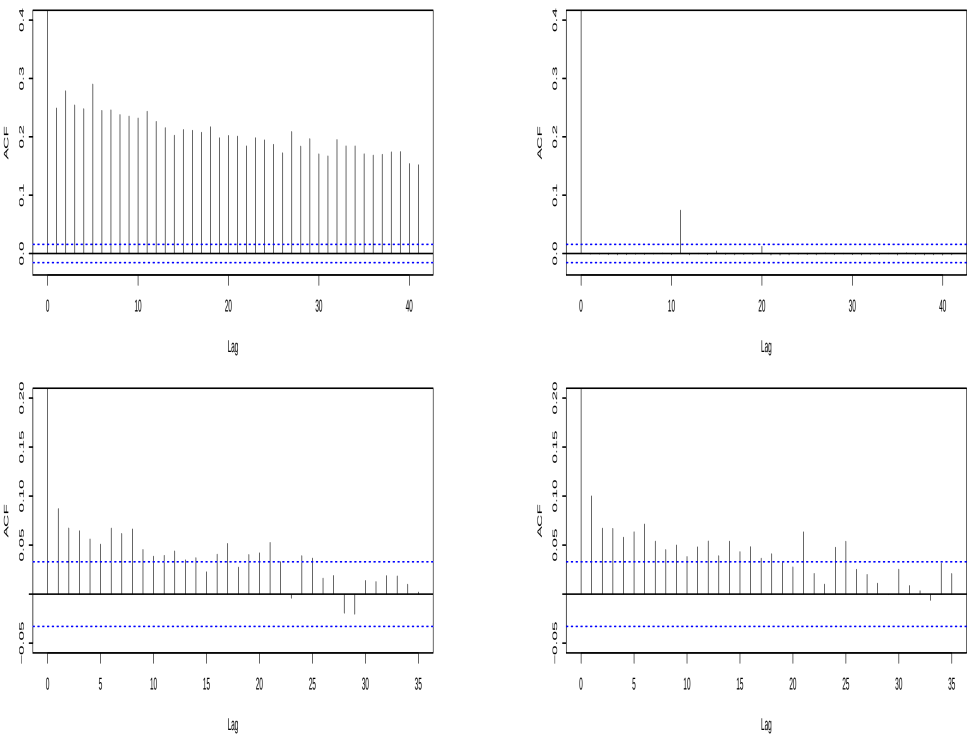

Fig. 9. Empirical autocorrelation functions for: i) absolute returns of the SP500 (top left), ii) inverse absolute returns of the SP500 (top right). iii) IBM durations (bottom left), i) inverse IBM durations (bottom right)

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

Moreover, the conclusion of Corollary 2 holds true for Model (14): the estimator $\tilde{\theta}_n$ *is asymptotically more efficient than* $\hat{\theta}_{n,h}$ *iff (13) holds.*

Proof. To prove the AN, we have already seen in the proof of Theorem 3.2 that Assumptions **A4** and **A7** are satisfied with $\delta = r$. Assumptions **A5** and **A8** are satisfied by the same arguments as in Theorem 3.2 and using Pan, Wang and Tong (2008), and Hamadeh and Zakoïan (2011). 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 (33), 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 (38)

$$J_\delta^{-1}\Omega_\delta = \overline{\theta}_0, \qquad \Omega_\delta' J_\delta^{-1}\Omega_\delta = 1 \tag{52}$$

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

$$J^{-1}\Omega = \frac{\delta}{2}\overline{\theta}_0, \qquad \Omega' J^{-1}\Omega = 1 \tag{53}$$

Moreover, similarly to (34), we have

$$\frac{\partial\sigma_t^2(\theta_0^*)}{\partial\theta} = \mu_2^{1-\frac{\delta}{2}} A_\delta \frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}. \tag{54}$$

It follows that, similar to (35),

$$J_* = \mu_2^{-\delta} A_\delta J A_\delta, \qquad \Omega_* = \mu_2^{-\delta/2} A_\delta \Omega. \tag{55}$$

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

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

In view of (53), the asymptotic variance follows.

Finally, the conclusion of Corollary 2 holds true for Model (14), 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

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

□

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.

Audrino, F. and P. Bühlmann (2009) Splines for financial volatility. *Journal of the Royal Statistical Society B*, 71, 655–670.

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.

Bauwens, L. and P. Giot (2000) The logarithmic ACD model: an application to the bid-ask quote process of three NYSE stocks. *Annales d'Economie et Statistique*, 60, 117–149.

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.

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

Brooks, C., Burke, S.P., Heravi, S. and G. Persand (2005) Autoregressive conditional kurtosis. *Journal of Financial Econometrics* 3, 399–421.

Davis, R.A., Knight, K. and J. Liu (1992) M-estimation for autoregressions with infinte variance. *Stochastic Processes and their Applications* 40, 145–180.

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.

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.

Fan, J., Qi, L. and D. Xiu (2010) Quasi Maximum Likelihood Estimation of GARCH Models with Heavy-Tailed Likelihoods. Discussion Paper, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1540363.

Francq, C., Lepage, G. and J-M. Zakoïan (2011) Two-stage non Gaussian QML estimation of GARCH Models and testing the efficiency of the Gaussian QMLE. *Journal of Econometrics* 165, 246–257.

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

Francq, C. and J.M. Zakoïan (2009) Bartlett's formula for a general class of non linear processes. *Journal of Time Series Analysis*, 30, 449–465.

Francq, C. and J-M. Zakoïan (2010) *GARCH Models: Structure, Statistical Inference and Financial Applications.* John Wiley.

Francq, C. and J.M. Zakoïan (2012) Strict stationarity testing and estimation of explosive and stationary GARCH models. *Econometrica*, 80, 821–861.

Gouriéroux, C., Monfort A. and A. Trognon (1984) Pseudo Maximum Likelihood Methods: Theory. *Econometrica* 52, 681–700.

Hamadeh, T. and J-M. Zakoïan (2011) Asymptotic properties of LS and QML estimators for a class of nonlinear GARCH Processes. *Journal of Statistical Planning and Inference*, 141, 488–507.

Karanasos, M. (2002) Prediction in ARMA models with GARCH in mean effects. *Journal of Time Series Analysis* 22, 555–576.

Li, W.K. (2004) *Diagnostic Checks in Time Series.* Chapman & Hall/CRC.

Ling, S. (2004) Estimation and testing stationarity for double-autoregressive models. *Journal of the Royal Statistical Society B*, 66, 63–78.

Ling, S. (2005). Self-weighted LAD estimation for infinite variance autoregressive models. *Journal of the Royal Statistical Society B*, 67, 381–393.

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

Mukherjee, K. (2008) M-estimation in GARCH models. *Econometric Theory* 24, 1530–1553.

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.

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

Pellegrini, S., Ruiz, E. and A. Espasa (2012) Prediction intervals in conditionally heteroscedastic time series with stochastic components. *International Journal of Forecasting,* In Press.

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

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.

Taylor, S.J. (2007) *Asset price dynamics, volatility and prediction*, Princeton: Princeton University Press.