# Finite moments testing in a general class of nonlinear time series models

Francq, Christian and Zakoian, Jean-Michel

June 2024

# Finite moments testing in a general class of nonlinear time series models

Christian Francq and Jean-Michel Zakoïan*

June 2024

**Abstract**

We investigate the problem of testing the finiteness of moments for a class of semi-parametric time series encompassing many commonly used specifications. The existence of positive-power moments of the strictly stationary solution is characterized by the Moment Determining Function (MDF) of the model, which depends on the parameter driving the dynamics and on the distribution of the innovations. We establish the asymptotic distribution of the empirical MDF, from which tests of moments are deduced. Alternative tests based on estimation of the Maximal Moment Exponent (MME) are studied. Power comparisons based on local alternatives and the Bahadur approach are proposed. We provide an illustration on real financial data and show that semi-parametric estimation of the MME provides an interesting alternative to Hill's nonparametric estimator of the tail index.

*Keywords:* Efficiency comparisons of tests, maximal moment exponent, stochastic recurrence equation, tail index

## 1 Introduction

If a random variable $X$ does not have finite moments of any order, its distribution is said to be heavy-tailed. If the distribution of $|X|$ is regularly varying with tail index $\alpha > 0$, its distribution is heavy-tailed and it is often said to be fat-tailed. In this case, one can say that the maximal moment exponent (MME)—that is, the highest finite moment order—of $|X|$ is equal to $\alpha$ because $E|X|^u < \infty$ for $u < \alpha$ and $E|X|^u = \infty$ for $u > \alpha$.

Knowing the tail index (or MME) of the marginal distribution of a stationary time series model is obviously of interest. Kesten [40] is a primary reference for tail index characterization of general linear Stochastic Recurrence Equations (SREs). Basrak, Davis and Mikosch [3] gave conditions for the existence of a tail index for general SREs and showed that the marginal distribution of a GARCH process is regularly varying. Zhang and Ling [53] showed that, under mild additional assumptions, the MME is also the tail index of GARCH extensions.

Based on these advances in the probabilistic structure of stochastic processes, tail index and MME estimators have been proposed and studied by Berkes et al. [5], Chan et al. [14], and Zhang et al. [52] for particular GARCH-type time series models.

*CREST, 5 Avenue Henri Le Chatelier, 91120 Palaiseau, France.
E-mails: christian.francq@ensae.fr, zakoian@ensae.fr

Surprisingly, little attention has been paid to testing moment finiteness (see however [31]). Testing the existence of moments seems however crucial, in particular for the validity of many statistical tools commonly used for the analysis of such models. Even if the consistency of Quasi-Maximum Likelihood (QML) estimators may hold under strict stationarity without any extra moment assumption (Berkes, Horváth and Kokoszka [6], and Francq and Zakoïan [25]; see also [27] for a review), many applications rest on finite unconditional moments. [1] Moreover, the existence of moments for real time series (such as financial returns) is an interesting issue *per se*, which is a controversial subject in the empirical literature.

The present paper proposes new methods for testing the existence of moments for a general class of time series models.

### 1.1 Time series model

We consider the class of time series models defined, for some subsets $H, F$, and $Y$ of $\mathbb{R}$, by

$$\begin{cases} y_t &= g(f_t, \eta_t; \boldsymbol{\theta}_0), \\ f_t &= \varphi(\eta_{t-1}, f_{t-1}; \boldsymbol{\theta}_0), \end{cases} \tag{1}$$

where $\boldsymbol{\theta}_0 \in \mathbb{R}^d$ is a vector of parameters, $(\eta_t)_{t\geqslant 0}$ is a sequence of independent and identically distributed (i.i.d.) $H$-valued random variables and the functions $g : F \times H \mapsto Y$ and $\varphi : H \times F \mapsto F$ are measurable. The times series $(y_t)$ is observed, while the process $(f_t)$ is latent. Two important examples are: (i) the additive model $y_t = f_t + \eta_t$, in which the variable $f_t$ can be interpreted as a time-varying location parameter, and (ii) the multiplicative model $y_t = f_t\eta_t$, where the variable $f_t$ can be interpreted as a time-varying volatility. More specific models, such as first-order ARMA or GARCH-type models, belong to this class (examples will be provided below).

### 1.2 Two characterizations of the existence of moments

Under conditions given below (see Proposition 2.1), Model (1) admits a strictly stationary solution $(y_t)$, and $f_t$ is independent of $\eta_t$. We make the assumption that, for any $u > 0$,

$$(E|\eta_t|^u < \infty \quad \wedge \quad E|f_t|^u < \infty) \quad \Longrightarrow \quad E|y_t|^u < \infty. \tag{2}$$

Note that this implication holds true for the volatility and location models presented above, and is even an equivalence in the latter case. [2] Moreover, omitting $\boldsymbol{\theta}_0$ for ease of presentation, we will show that if $f \mapsto \varphi(\eta, f)$ is Lipschitz continuous for all $\eta \in H$, we have, for any $f^0 \in F$,

$$\left(E\left|\varphi(\eta_t, f^0) - f^0\right|^u < \infty \quad \wedge \quad E\left\{\Lambda^u(\eta_t; \boldsymbol{\theta}_0)\right\} < 1\right) \Longrightarrow E|f_t|^u < \infty, \tag{3}$$

where

$$\Lambda(\eta_t; \boldsymbol{\theta}_0) = \sup_{\substack{f_1, f_2 \in F \\ f_1 \neq f_2}} \left|\frac{\varphi(\eta_t, f_1) - \varphi(\eta_t, f_2)}{f_1 - f_2}\right|.$$

[1] For instance, the existence of the autocorrelation function of any transform (e.g. square or absolute values) of the returns requires appropriate moments; prediction of the squared returns over a long horizon requires a finite variance, and prediction confidence intervals require fourth-order moments.

[2] Indeed, in the location model we have, through the $C_r$ inequality, $E|y_t|^u \leqslant C_u(E|f_t|^u + E|\eta_t|^u)$ for some constant $C_u > 0$. Hence, (2) holds. Now, since $f_t$ and $\eta_t$ are independent, if $E|f_t|^u = \infty$, then $E|f_t + c|^u = \infty$ for all $c$ and, with obvious notations, it follows that $E|y_t|^u = \int E|f_t + c|^u dP_\eta(c) = \infty$. Similarly, $E|\eta_t|^u = \infty$ entails $E|y_t|^u = \infty$.

The behaviour of the function $u \mapsto E\{\Lambda^u(\eta_1; \boldsymbol{\theta}_0)\}$, referred to hereafter as the *Moment Determining Function* (MDF) of the model (1) is thus crucial for the existence of moments.

Under the conditions discussed below, there exists a unique $u_0 > 0$ such that $E\{\Lambda^u(\eta_1; \boldsymbol{\theta}_0)\} = 1$ and the moment condition can be written

$$\left(u < u_0 \quad \wedge \quad E\left|\varphi(\eta_t, f^0) - f^0\right|^u < \infty \quad \wedge \quad E|\eta_t|^u < \infty\right) \Longrightarrow E|y_t|^u < \infty.$$

Following Berkes et al.'s terminology [5], $u_0$ will be referred to as the *Maximal Moment Exponent* (MME). Under more restrictive assumptions, this coefficient will be related to the *tail index* of the distribution of $y_t$.

### 1.3 Testing the existence of moments

Our main contribution in this paper is to propose tests for the existence of moment of any (positive) order, based on empirical versions of the MDF and MME. Using a semi-parametric version of Model (1), in which the functions $g$ and $\varphi$ depend on a finite-dimensional parameter $\boldsymbol{\theta}_0$ but the distribution of $\eta_t$ is left unspecified, we will provide conditions for the existence and the consistency and asymptotic normality (CAN) of the empirical MDF and MME,

$$S_n^{(u)} = \frac{1}{n}\sum_{t=1}^{n} \Lambda^u(\widehat{\eta}_t; \widehat{\boldsymbol{\theta}}_n) \quad \text{where } u > 0, \qquad \widehat{u}_n = \sup\left\{u > 0;\ S_n^{(u)} \leqslant 1\right\}, \tag{4}$$

where $\widehat{\boldsymbol{\theta}}_n$ denotes any consistent estimator of $\boldsymbol{\theta}_0$, and $\widehat{\eta}_t$, for $t = 1, \ldots, n$, denote residuals. For the standard GARCH(1,1) model, these results were established by [5].

Building on this, we will derive tests for the existence of moments. Let the test statistics based on the empirical MDF and MME,

$$T_n^{(u)} = \frac{\sqrt{n}\left(S_n^{(u)} - 1\right)}{\widehat{v}_u} \quad \text{and} \quad U_n^{(u)} = \frac{\sqrt{n}\,(u - \widehat{u}_n)}{\widehat{w}_{\widehat{u}_n}},$$

where $\widehat{v}_u^2$ and $\widehat{w}_{\widehat{u}_n}^2$ denote consistent estimators of the asymptotic variances of $S_n^{(u)}$ and $\widehat{u}_n$, respectively. Tests of the moment condition $E|y_t|^u < \infty$ at the asymptotic level $\underline{\alpha} \in (0, 1)$, are defined by the rejection regions

$$C_T^{(u)} = \left\{T_n^{(u)} > \Phi^{-1}(1 - \underline{\alpha})\right\} \quad \text{and} \quad C_U^{(u)} = \left\{U_n^{(u)} > \Phi^{-1}(1 - \underline{\alpha})\right\},$$

where $\Phi$ is the $\mathcal{N}(0, 1)$ Cumulative Distribution Function (CDF). Assuming that $\eta_t$ has a known density, or a parametric density, parametric versions $V_n^{(u)}$ and $W_n^{(u)}$ of the statistic $U$ will also be introduced.

### 1.4 Contributions of the paper

We study the aforementioned tests for the existence of moments in Model (1). Since the model is semi-parametric, we will not restrict ourselves to the Maximum Likelihood (ML) estimation method or any specific method of estimation for the parameter $\boldsymbol{\theta}_0$. Our conditions allow for general consistent estimators admitting a Bahadur-type expansion, although some of our results are particular to the QML and ML methods.

Our contributions are as follows:

a) we discuss the existence and uniqueness of a solution to the SRE associated with Model (1); providing conditions for the existence of a unique MME;

b) we establish the weak convergence of the empirical MDF process, from which we deduce the asymptotic distribution of the estimator of the MME/tail index;

c) we propose new tests of moment finiteness;

d) cases where the error density is either known or parameterized are discussed;

e) for a class of GARCH(1,1)-type models, we provide power comparisons of semi-parametric and parametric tests under local alternatives or using the Bahadur approach.

### 1.5 Comparison with alternative approaches

Nonparametric procedures for checking the existence of finite moments have been developed previously in the statistical literature. Note that they usually require assumptions about the CDF of the observed variables (for instance a Pareto-type tail). The most widely used methods, arguably, are based on estimation of the tail index, as in Hill [38]. In particular, many papers have established the asymptotic properties of Hill's tail index estimator for both independent and stationary sequences of observations. The weaknesses of this estimator (in particular its extreme sensitivity to the choice of tuning parameters) are well known and have given rise to variants and improvements (see Embrechts et al. [23], Section 6.4, for a review). Other nonparametric approaches do not require tail index estimation (in particular, see Trapani [49] for a test based on the convergence versus divergence of sample moments, Ng and Yau [45] for a bootstrap procedure).

Those nonparametric approaches focus on the existence of moments *per se*. Within the semi-parametric framework of this study, once a model is chosen for a particular time series, we scrutinize its applicability for various objectives, such as prediction, or estimation of conditional risk measures. Reliable inference procedures generally require finiteness of some moments. Although the tests presented in this paper are susceptible to model misspecification, their advantage is that they have a parametric convergence rate. Our numerical simulations clearly demonstrate the superiority of our approach under correct model specification.

### 1.6 Structure of the paper

In Section 2, we develop the asymptotic theory for the empirical MDF and we derive a test based on the MDF. Section 3 derives parametric and semi-parametric tests based on the MME. In Section 4 we apply our results to GARCH-type processes. For these models, comparisons based on local alternatives are studied in Section 5. An empirical illustration is provided in Section 6. Technical assumptions, proofs, additional properties, and Monte-Carlo experiments are provided in appendix.

## 2 Estimating the MDF and testing the existence of moments

Let $\boldsymbol{\theta}$ denote a generic value of the parameter, which is assumed to belong to a compact parameter set $\Theta \subset \mathbb{R}^d$.

The second equation in (1) has the form of an SRE which enables us to study its probability properties. Assuming that $f \mapsto \varphi(\eta, f; \boldsymbol{\theta})$ is Lipschitz continuous for all $\eta \in H$ and $\boldsymbol{\theta} \in \Theta$, set

$$\Lambda(\eta; \boldsymbol{\theta}) = \sup_{\substack{f_1, f_2 \in F \\ f_1 \neq f_2}} \left| \frac{\varphi(\eta, f_1; \boldsymbol{\theta}) - \varphi(\eta, f_2; \boldsymbol{\theta})}{f_1 - f_2} \right|.$$

When the function $f \mapsto \varphi(\eta, f; \boldsymbol{\theta})$ is differentiable with respect to $f$, which is the case for all commonly used models, the supremum reduces to the supremum of the first derivative of this function. Otherwise, it has to be computed on a case-by-case basis.

The existence of a strictly stationary solution to Model (1) rests on Assumption

**A0:** $f \mapsto \varphi(\eta, f; \boldsymbol{\theta}_0)$ is Lipschitz continuous for all $\eta \in H$,
(i) $E \log^+ \left|\varphi(\eta_t, f^0; \boldsymbol{\theta}_0) - f^0\right| < \infty$ for some constant $f^0 \in F$;
(ii) $E \log^+ \Lambda(\eta_t; \boldsymbol{\theta}_0) < \infty$ and $E \log \Lambda(\eta_t; \boldsymbol{\theta}_0) < 0$.

**Proposition 2.1.** *Under* **A0**, *there exists a strictly stationary, ergodic and nonanticipative*[3] *solution* $(y_t)$ *to Model* (1). *Moreover, if* $E\left|\varphi(\eta_t, f^0; \boldsymbol{\theta}_0) - f^0\right|^r < \infty$ *and* $E\left\{\Lambda^r(\eta_t; \boldsymbol{\theta}_0)\right\} < \infty$ *for some* $r > 0$, *we have* $E|f_t|^s < \infty$ *for some* $s > 0$. *Finally, if* $E\left|\varphi(\eta_t, f^0; \boldsymbol{\theta}_0) - f^0\right|^u < \infty$ *and* $E\left\{\Lambda^u(\eta_t; \boldsymbol{\theta}_0)\right\} < 1$ *for some* $u > 0$, *we have* $E|f_t|^u < \infty$.

The proof follows straightforwardly from Bougerol [12] and Straumann and Mikosch [47] (see also Lemma 4.1 in Francq and Zakoian [30], and Lemmas 1 and 2 in Blasques et al. [9]).

## 2.1 Invertibility

Given observations $y_1, \ldots, y_n$, and arbitrary initial values $\tilde{y}_0 \in Y$ and $\tilde{f}_0 \in F$, we define recursively, for any $\boldsymbol{\theta}$, a sequence $\tilde{f}_t(\boldsymbol{\theta})$, which depends on the observations used to estimate $\boldsymbol{\theta}_0$. We make the following invertibility assumption.

**A1:** There exists a function $g^*$ such that, for all $(y, f, \eta) \in Y \times F \times E$ and $\boldsymbol{\theta} \in \Theta$,
$y = g(f, \eta; \boldsymbol{\theta}) \quad \Longleftrightarrow \quad \eta = g^*(f, y; \boldsymbol{\theta})$.

Define, for $t = 1, \ldots, n$ and any $\boldsymbol{\theta}$ belonging to $\Theta$,

$$\tilde{f}_t(\boldsymbol{\theta}) = \varphi\left[g^*\left\{\tilde{f}_{t-1}(\boldsymbol{\theta}), y_{t-1}; \boldsymbol{\theta}\right\}, \tilde{f}_{t-1}(\boldsymbol{\theta}); \boldsymbol{\theta}\right] := \psi\left\{y_{t-1}, \tilde{f}_{t-1}(\boldsymbol{\theta}); \boldsymbol{\theta}\right\} \tag{5}$$

where $\tilde{f}_0(\boldsymbol{\theta}) = \tilde{f}_0$ and $y_0 = \tilde{y}_0$. The above SRE raises the question of the *invertibility* of the model, which holds only if $\tilde{f}_t(\boldsymbol{\theta})$ does not depend asymptotically on the initialization (see Blasques et al. [10], and Straumann and Mikosch [47]). The sequence $(\tilde{f}_t(\boldsymbol{\theta}))_{t \geqslant 0}$ can be approximated by $(f_t(\boldsymbol{\theta}))$, the solution of the SRE

$$f_t(\boldsymbol{\theta}) = \varphi\left[g^*\left\{f_{t-1}(\boldsymbol{\theta}), y_{t-1}; \boldsymbol{\theta}\right\}, f_{t-1}(\boldsymbol{\theta}); \boldsymbol{\theta}\right] = \psi\left\{y_{t-1}, f_{t-1}(\boldsymbol{\theta}); \boldsymbol{\theta}\right\}, \ t \in \mathbb{Z}. \tag{6}$$

The existence of a strictly stationary solution to (6) is guaranteed by the following assumptions. Set, for any $\boldsymbol{\theta} \in \Theta$,

$$\Lambda_1(y; \boldsymbol{\theta}) = \sup_{\substack{f_1, f_2 \in F \\ f_1 \neq f_2}} \left|\frac{\psi(y, f_1; \boldsymbol{\theta}) - \psi(y, f_2; \boldsymbol{\theta})}{f_1 - f_2}\right|,$$

and assume

**A2:** $F$ is a closed subset of $\mathbb{R}$ and for any $(y, \boldsymbol{\theta}) \in Y \times \Theta$, the mapping $f \in F \mapsto \psi(y, f; \boldsymbol{\theta})$ is Lipschitz continuous. Moreover,
(i) $E \log^+ \left|\psi(y_t, f^0; \boldsymbol{\theta}) - f^0\right| < \infty$ for some constant $f^0 \in F$;
(ii) $E \log^+ \Lambda_1(y_t; \boldsymbol{\theta}) < \infty$ and $E \log \Lambda_1(y_t; \boldsymbol{\theta}) < 0$.

A uniform (in $\boldsymbol{\theta}$) version of **A2** is

[3] i.e. $y_t \in \mathcal{F}_t$, the $\sigma$-field generated by $(\eta_t, \eta_{t-1}, \ldots)$.

**A3:** **A2** holds with (i)-(ii) replaced by
(i') $E \log^+ \sup_{\boldsymbol{\theta}\in\Theta} \left|\psi(y_t, f^0; \boldsymbol{\theta}) - f^0\right| < \infty$ for some constant $f^0 \in F$;
(ii') $E \log^+ \sup_{\boldsymbol{\theta}\in\Theta} \Lambda_1(y_t; \boldsymbol{\theta}) < \infty$ and $E \log \sup_{\boldsymbol{\theta}\in\Theta} \Lambda_1(y_t; \boldsymbol{\theta}) < 0$.

*Lemma* 1. Under assumptions **A0-A2**, for any $\boldsymbol{\theta} \in \Theta$ there exists a stationary and ergodic solution $\{f_t(\boldsymbol{\theta})\}$ (with $f_t(\boldsymbol{\theta}) \in F$) to the SRE (6). If in addition, **A3** holds, for any starting value $\widetilde{f}_0 \in F$, there exists $\rho \in (0,1)$ such that $\rho^{-t} \sup_{\boldsymbol{\theta}\in\Theta} \left|\widetilde{f}_t(\boldsymbol{\theta}) - f_t(\boldsymbol{\theta})\right| \to 0$ a.s. as $t \to \infty$.

The latter results shows that the difference between the stationary ergodic sequence $f_t(\boldsymbol{\theta})$ and its feasible approximation $\widetilde{f}_t(\boldsymbol{\theta})$ tends to zero exponentially fast.

## 2.2 Asymptotic distribution of $S_n^{(u)}$

Let $u > 0$. Assuming $S_\infty^{(u)} := E\{\Lambda^u(\eta_1; \boldsymbol{\theta}_0)\}$ is finite, we now derive the asymptotic distribution of the estimator $S_n^{(u)}$ defined in (4), where $\widehat{\eta}_t = g^*(\widehat{f}_t, y_t; \widehat{\boldsymbol{\theta}}_n)$ with $\widehat{f}_t = \widetilde{f}_t(\widehat{\boldsymbol{\theta}}_n)$.

It will be useful to consider the quantities defined for $\boldsymbol{\theta} \in \Theta$ by

$$S_n^{(u)}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} \Lambda^u\{\eta_t(\boldsymbol{\theta}); \boldsymbol{\theta}\}, \quad \widetilde{S}_n^{(u)}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} \Lambda^u\{\widetilde{\eta}_t(\boldsymbol{\theta}); \boldsymbol{\theta}\},$$

where $\eta_t(\boldsymbol{\theta}) = g^*\{f_t(\boldsymbol{\theta}), y_t; \boldsymbol{\theta}\}$ and $\widetilde{\eta}_t(\boldsymbol{\theta}) = g^*\left\{\widetilde{f}_t(\boldsymbol{\theta}), y_t; \boldsymbol{\theta}\right\}$. We introduce the following high-level assumptions, which will be worked out in particular cases.

**HL1:** There exists a vector $\boldsymbol{g}_u \in \mathbb{R}^d$ such that

$$\sqrt{n}\left(S_n^{(u)} - S_\infty^{(u)}\right) = \sqrt{n}\left(S_n^{(u)}(\boldsymbol{\theta}_0) - S_\infty^{(u)}\right) + \boldsymbol{g}_u'\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_P(1).$$

**HL2:** $\boldsymbol{\theta}_0$ belongs to the interior $\overset{\circ}{\Theta}$ of $\Theta$, $\widehat{\boldsymbol{\theta}}_n \in \Theta$ is a strongly consistent estimator of $\boldsymbol{\theta}_0$ and the following Bahadur expansion holds

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{\Delta}_{t-1}\boldsymbol{V}(\eta_t) + o_P(1),$$

where $\boldsymbol{V}(\cdot)$ is a measurable function, $\boldsymbol{V}: H \mapsto \mathbb{R}^k$ for some positive integer $k$, and $\boldsymbol{\Delta}_{t-1}$ is a $\mathcal{F}_{t-1}$-measurable $d \times k$ matrix, $(\boldsymbol{\Delta}_t)$ being stationary. The variables $\boldsymbol{\Delta}_t$ and $\boldsymbol{V}(\eta_t)$ belong to $L^2$ with $E\{\boldsymbol{V}(\eta_t)\} = 0$, $\mathrm{var}\{\boldsymbol{V}(\eta_t)\} = \boldsymbol{\Upsilon}$ and $E(\boldsymbol{\Delta}_t) = \boldsymbol{\Delta}$. Moreover, $\boldsymbol{x}'\boldsymbol{\Delta}_t$ is a non-constant random vector for any non-zero vector $\boldsymbol{x} \in \mathbb{R}^d$ and $\boldsymbol{\Upsilon}$ is positive definite.

For a given model, **HL1** can be checked by: i) noting that $S_n^{(u)} = \widetilde{S}_n^{(u)}(\widehat{\boldsymbol{\theta}}_n)$; ii) showing the asymptotic irrelevance of the initial values (i.e. $\widetilde{S}_n^{(u)}$ can be replaced by $S_n^{(u)}$); iii) performing a Taylor expansion of $S_n^{(u)}(\boldsymbol{\theta})$ around $\boldsymbol{\theta}_0$. We follow this approach in the proof of Corollary 1 for a class of GARCH-type models. **HL2** is a mild assumption that is satisfied by many commonly used estimators, as illustrated in Corollary 7 in the the appendix.

The following result provides the asymptotic distribution of the empirical MDF $S_n^{(u)}$.

*Theorem* 1. Under **A0**, **A1**, **A3** and **HL1-HL2** and assuming $E\{\Lambda^s(\eta_t; \boldsymbol{\theta}_0)\} < \infty$ for $s > 0$ we have, for $0 < u \leqslant s/2$ such that $\boldsymbol{\xi}_u = \boldsymbol{\Delta} E\{\boldsymbol{V}(\eta_t)\Lambda^u(\eta_t; \boldsymbol{\theta}_0)\}$ exists,

$$\sqrt{n}\left(S_n^{(u)} - S_\infty^{(u)}\right) \overset{d}{\to} \mathcal{N}\left(0, v_u^2 := \boldsymbol{g}_u'\boldsymbol{\Sigma}\boldsymbol{g}_u + \psi_u + 2\boldsymbol{g}_u'\boldsymbol{\xi}_u\right), \tag{7}$$

where $\boldsymbol{\Sigma} = E(\boldsymbol{\Delta}_t\boldsymbol{\Upsilon}\boldsymbol{\Delta}_t')$ and $\psi_u = \mathrm{Var}\{\Lambda^u(\eta_1; \boldsymbol{\theta}_0)\}$. Moreover, if $\psi_u > 0$ we have $v_u^2 > 0$.

Figure 1: MDF for $\Lambda(\eta_t) = 0.1\eta_t^2 + 0.85$ and for Student errors with $\nu$ degrees of freedom. The values of the MME $u_0$ are displayed over the horizontal axis.

Theorem 1 does not require any moment assumption on the observed process $(y_t)$. The moment assumption on $\Lambda(\eta_t; \boldsymbol{\theta}_0)$ is in general very weak: for instance, in some models, this assumption is innocuous (as in the Beta-$t$-GARCH models of Harvey [36, Chapter 4] and Creal et al. [15], see the appendix Section D, where the variables $\Lambda(\eta_t, \boldsymbol{\theta}_0)$ are bounded).

### 2.3 Testing the existence of moments of a given order

In this section, we test the following assumptions, for a specific $u > 0$,

$$\boldsymbol{H}_{0,u}: \quad E\{\Lambda^u(\eta_t)\} \leqslant 1 \quad \text{against} \quad \boldsymbol{H}_{1,u}: \quad E\{\Lambda^u(\eta_t)\} > 1, \tag{8}$$

where $\Lambda(\eta_t) = \Lambda(\eta_t; \boldsymbol{\theta}_0)$. In view of (2) and (3), $E\{\Lambda^u(\eta_t)\} < 1$ entails the existence of a finite $u$-th order moment for $|y_t|$, provided $E\left|\varphi(\eta_t, f^0) - f^0\right|^u < \infty$ and $E|\eta_t|^u < \infty$. The reverse hypotheses could also be tested, that is $\boldsymbol{H}^*_{0,u}: E\{\Lambda^u(\eta_t)\} \geqslant 1$ against $\boldsymbol{H}^*_{1,u}: E\{\Lambda^u(\eta_t)\} < 1$, simply by reversing the inequalities in the critical regions that we are going to define.

By definition of the MME $u_0$, the null hypothesis can also be written $\boldsymbol{H}_{0,u}: u \leqslant u_0$. The next proposition gathers existing results on the existence of a finite MME.

**Proposition 2.2.** *Suppose $\gamma = E \log \Lambda(\eta_1) < 0$.*
*i) If $P\{\Lambda(\eta_1) \leqslant 1\} = 1$, then for all $u > 0$, $E\{\Lambda^u(\eta_1)\} < 1$.*
*ii) If $1 \leqslant E\{\Lambda^s(\eta_1)\} < \infty$ for some $s > 0$, then there exists a unique $u_0 > 0$ such that $E\{\Lambda^{u_0}(\eta_1)\} = 1$.*
*Moreover, if $E\{\Lambda^u(\eta_1)\} < 1$ and $E\{\Lambda^v(\eta_1)\} > 1$ for $0 < u < v$ then $u_0 \in (u, v)$.*

*Remark* 1. When $\Lambda(\eta_1)$ has unbounded support and admits moments of any order $m$, these moments tend to infinity when $m$ increases and the condition $1 \leqslant E\{\Lambda^s(\eta_1)\} < \infty$ for some $s > 0$ is satisfied. More generally, the condition is satisfied for most classical distributions with unbounded support. However, the following example shows that the condition is non trivial: suppose that the density $g$ of $\Lambda(\eta_1)$ is such that $g(x) \overset{x\to\infty}{\sim} K(x^2 \log^2 x)^{-1}$. Then we have $E\{\Lambda^s(\eta_1)\} = \infty$ for any $s > 1$ but $E\{\Lambda(\eta_1)\} < \infty$ (if, for instance, $g$ is bounded). It is clear that the latter expectation can be made smaller than 1 by scaling the function $\Lambda$. For these distributions, $u_0$ does not exist.

The shape of the MDF is illustrated in Figure 1 for the quadratic function $\Lambda(\eta_t) = \alpha\eta_t^2 + \beta$ (corresponding for instance to a standard GARCH(1,1) model) with Student error distributions. Under the assumptions of Proposition 2.2 case ii), this shape is general: as $u$ increases the function $u \mapsto S_\infty^{(u)}$ first decreases and then increases, crossing the horizontal line $f(u) = 1$ at $u = u_0$.

Define the test statistic based on the empirical MDF,

$$T_n^{(u)} = \frac{\sqrt{n}\left(S_n^{(u)} - 1\right)}{\widehat{\upsilon}_u}, \tag{9}$$

assuming $\widehat{\upsilon}_u^2 > 0$ is a consistent estimator of $\upsilon_u^2$. We introduce the following assumption.

**HL3:** For any sequence $(\boldsymbol{\theta}_n)$ such that $\boldsymbol{\theta}_n \to \boldsymbol{\theta}_0$ in probability, we have, for any $r \leqslant s$, $\left|\widetilde{S}_n^{(r)}(\boldsymbol{\theta}_n) - S_n^{(r)}(\boldsymbol{\theta}_0)\right| \to 0$, in probability as $n \to \infty$.

This assumption implies that the initial values and the estimation of $\boldsymbol{\theta}_0$ both have negligible effects on the asymptotic behaviour of the empirical MDF. This will be verified explicitly for the models of Section 4.

**Proposition 2.3.** *Under the assumptions of Theorem 1 and* **HL3***, a test of* $\boldsymbol{H}_{0,u}$ *at the asymptotic level* $\underline{\alpha} \in (0,1)$ *is defined by the rejection region*

$$C_T^{(u)} = \left\{ T_n^{(u)} > \Phi^{-1}(1-\underline{\alpha}) \right\}. \tag{10}$$

*Moreover, the tests is consistent: under* $\boldsymbol{H}_{1,u}$*, we have* $P\left(C_T^{(u)}\right) \to 1$ *as* $n \to \infty$.

The condition $u \leqslant s/2$ in Theorem 1 ensures the existence of $S_\infty^{(2u)}$—allowing the use of the CLT—and is crucial for $\underline{\alpha}$ to be the asymptotic frequency of rejection of $\boldsymbol{H}_{0,u_0}$. It also ensures the consistency of the test of the latter proposition, since $S_n^{(u)}$ converges in probability to $S_\infty^{(u)}$ under the alternative.

# 3 Estimating the MME and alternative tests

We now investigate the estimation of the MME $u_0$ and the corresponding test under three different settings.

## 3.1 Semi-parametric estimation of the MME

For $\boldsymbol{\theta} \in \Theta$ let

$$\gamma_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^n \log \Lambda\{\eta_t(\boldsymbol{\theta}); \boldsymbol{\theta}\}, \qquad \tilde{\gamma}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^n \log \Lambda\{\tilde{\eta}_t(\boldsymbol{\theta}); \boldsymbol{\theta}\}, \qquad \gamma_n = \tilde{\gamma}_n(\widehat{\boldsymbol{\theta}}_n).$$

The following result is the sample counterpart of Proposition 2.2.

**Proposition 3.1.** *Suppose* $\gamma_n < 0$.

*If* $\Lambda(\widehat{\eta}_t; \widehat{\boldsymbol{\theta}}_n) \leqslant 1$ *for all* $1 \leqslant t \leqslant n$*, then* $S_n^{(u)} < 1$*, for all* $u > 0$.

*Conversely, if* $\Lambda(\widehat{\eta}_t; \widehat{\boldsymbol{\theta}}_n) > 1$ *for at least one* $1 \leqslant t \leqslant n$*, then there exists a unique* $u_n > 0$ *such that* $S_n^{(u_n)} = 1$*. Moreover, if* $S_n^{(u)} < 1$ *and* $S_n^{(v)} > 1$ *for* $0 < u < v$ *then* $u_n \in (u, v)$.

Letting $\widehat{u}_n = \sup\{u > 0; \ S_n^{(u)} \leqslant 1\}$, we have $\widehat{u}_n = \infty$ when $\Lambda(\widehat{\eta}_t; \widehat{\boldsymbol{\theta}}_n) \leqslant 1$ for all $1 \leqslant t \leqslant n$, and $\widehat{u}_n = u_n$ (of Proposition 3.1) in the opposite case. Let $\Lambda(\eta) = \Lambda(\eta; \boldsymbol{\theta}_0)$. We will show the strong consistency of $\widehat{u}_n$ under the following assumption

**HL4:** $\gamma_n = \gamma_n(\boldsymbol{\theta}_0) + o(1), \quad S_n^{(u)} = S_n^{(u)}(\boldsymbol{\theta}_0) + o(1),$ almost surely (a.s.) for any $u \leqslant s$.

*Theorem* 2. Assume that **A0-A2** and **HL4** hold. Then $\gamma_n \to \gamma$, a.s. Moreover,

i) if $P\{\Lambda(\eta_1) \leqslant 1\} = 1$, then $\widehat{u}_n \to \infty$, a.s.

ii) if $1 < E\{\Lambda^s(\eta_1)\} < \infty$ for some $s > 0$, then $\widehat{u}_n \to u_0$, a.s., where $u_0 > 0$ satisfies $E\{\Lambda^{u_0}(\eta_1)\} = 1$.

In order to obtain the asymptotic distribution of $\widehat{u}_n$, we will now consider a functional extension of Theorem 1. For $u_1 < u_2$, let $\mathcal{C}[u_1, u_2]$ denote the space of continuous functions on $[u_1, u_2]$, and let $\Rightarrow$ denote weak convergence in the space $\mathcal{C}$ equipped with uniform distance. Let $\Gamma_n(u) = \sqrt{n}\left(S_n^{(u)} - S_\infty^{(u)}\right)$ and, in view of **HL1**, let

$$\Gamma_n^0(u) \quad = \quad \sqrt{n}\left(S_n^{(u)}(\boldsymbol{\theta}_0) - S_\infty^{(u)}\right) + \boldsymbol{g}_u' \frac{1}{\sqrt{n}} \sum_{t=1}^n \boldsymbol{\Delta}_{t-1} \boldsymbol{V}(\eta_t).$$

To prove a uniform extension of Theorem 1, we need to introduce the following assumptions:

**HL5:** For $[u_1, u_2] \subset (0, s/2)$, we have $\sup_{u \in (u_1,u_2)} |\Gamma_n(u) - \Gamma_n^0(u)| = o_P(1)$.

**HL6:** For any $0 < u, v \leqslant s/2$, we have $\|\boldsymbol{g}_u - \boldsymbol{g}_v\| \leqslant K|u - v|$.

*Theorem* 3. Under the assumptions of Theorem 1 and if **HL5**-**HL6** hold,

$$\sqrt{n}\left(S_n^{(u)} - S_\infty^{(u)}\right) \overset{\mathcal{C}[u_1,u_2]}{\Longrightarrow} \Gamma(u) \tag{11}$$

where $\Gamma(u)$ stands for a Gaussian process with $E\Gamma(u) = 0$ and $\mathrm{Cov}\{\Gamma(u), \Gamma(v)\} = \boldsymbol{g}_u' \boldsymbol{\Sigma} \boldsymbol{g}_v + \psi_{u,v} + \boldsymbol{g}_u' \xi_v + \boldsymbol{g}_v' \boldsymbol{\xi}_u$ where $\psi_{u,v} = \mathrm{Cov}\{\Lambda^u(\eta_1; \boldsymbol{\theta}_0), \Lambda^v(\eta_1; \boldsymbol{\theta}_0)\}$.

Let $D_\infty^{(u)} = E[\Lambda^u(\eta_1; \boldsymbol{\theta}_0) \log\{\Lambda(\eta_1; \boldsymbol{\theta}_0)\}]$ be the first-order derivative of the MDF $u \to S_\infty^{(u)}$, which is well-defined for $u < s$. Note that $D_\infty^{(u_0)}$ is positive (in view of the convexity of the MDF established in the proof of Proposition 2.2). The asymptotic distribution of the MME was derived for standard GARCH models by Mikosch and Stărică [44] and by Berkes et al. [5], for an AR(1)-ARCH(1) model by Chan et al. [14], and for both models using a least absolute deviation estimator by Zhang et al. [52]. For Model (1), we have the following result.

*Theorem* 4. Let the assumptions of Theorem 3 hold, and let $1 < E\{\Lambda^s(\eta_1)\} < \infty$ for some $s > 0$, with $u_0 \in (u_1, u_2)$. Then, we have

$$\sqrt{n}(\widehat{u}_n - u_0) \xrightarrow{d} \mathcal{N}\left(0, w_{u_0}^2\right), \quad w_{u_0}^2 = \left(D_\infty^{(u_0)}\right)^{-2} v_{u_0}^2,$$

where $v_{u_0}^2$ is the asymptotic variance defined in Theorem 1.

This result allows us to build asymptotic confidence intervals (CIs) for the MME $u_0$, as will be illustrated in the case of multiplicative models.

For a given $u > 0$, define $D_n^{(u)} = \widetilde{D}_n^{(u)}(\widehat{\boldsymbol{\theta}}_n)$ with

$$D_n^{(u)}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^n \Lambda^u\{\eta_t(\boldsymbol{\theta}); \boldsymbol{\theta}\} \log\Lambda\{\eta_t(\boldsymbol{\theta}); \boldsymbol{\theta}\}, \quad \widetilde{D}_n^{(u)}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^n \Lambda^u\{\widetilde{\eta}_t(\boldsymbol{\theta}); \boldsymbol{\theta}\} \log\Lambda\{\widetilde{\eta}_t(\boldsymbol{\theta}); \boldsymbol{\theta}\}.$$

Let the test statistic, assuming $\widehat{w}_{\widehat{u}_n} > 0$,

$$U_n^{(u)} = \frac{\sqrt{n}\,(u - \widehat{u}_n)}{\widehat{w}_{\widehat{u}_n}}, \quad \text{where} \quad \widehat{w}_u^2 = \left(\frac{\widehat{v}_u}{D_n^{(u)}}\right)^2.$$

**HL7:** For any sequence $(\boldsymbol{\theta}_n)$ such that $\boldsymbol{\theta}_n \to \boldsymbol{\theta}_0$, a.s. we have, for any $u \leqslant s$, $\left|\widetilde{D}_n^{(u)}(\boldsymbol{\theta}_n) - D_n^{(u)}(\boldsymbol{\theta}_0)\right| \to 0$, in probability as $n \to \infty$.

**Proposition 3.2.** *Under the assumptions of Theorem 4 with $w_{u_0}^2 > 0$,* **HL7**, *a test of $\boldsymbol{H}_{0,u}$ at the asymptotic level $\underline{\alpha} \in (0, 1)$ is defined by the rejection region*

$$C_U^{(u)} = \left\{U_n^{(u)} > \Phi^{-1}(1 - \underline{\alpha})\right\}, \tag{12}$$

*and an asymptotic $100(1 - \underline{\alpha})\%$ CI for $u_0$ is $\widehat{u}_n \pm n^{-1/2}\Phi^{-1}(1 - \underline{\alpha})\widehat{w}_{\widehat{u}_n}$.*

*Moreover, the test is consistent: under $\boldsymbol{H}_{1,u}$ we have $P\left(C_U^{(u)}\right) \to 1$ as $n \to \infty$.*

We will now consider situations where the errors have a density that is either known or known up to a finite-dimensional parameter, yielding alternative estimators of the MME.

### 3.2 Purely parametric estimators of the MME

We assume that $\eta_t$ has a density $h$ and we make the following assumption, which is satisfied for many distributions, including the Gaussian distribution.

**HL8:** $\boldsymbol{\theta} \mapsto \int \Lambda^{u_0}(x; \boldsymbol{\theta}) h(x) dx$ is continuously differentiable under the integral sign.

#### 3.2.1 When the error density $h$ is known

When the MME $u_0 = u_{0,h}(\boldsymbol{\theta}_0)$ exists, by definition it is the solution of the implicit equation

$$\int \Lambda^{u_0}(x; \boldsymbol{\theta}_0) h(x) dx = 1.$$

Under **HL8** this solution satisfies, through the implicit function theorem,

$$\frac{\partial u_{0,h}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = \frac{-1}{D_\infty^{(u_0)}} \boldsymbol{r}_{u_0}, \quad \boldsymbol{r}_{u_0} := \frac{\partial}{\partial \boldsymbol{\theta}} S_\infty^{(u_0)} = E\left( u_0 \Lambda^{u_0 - 1}(\eta_t; \boldsymbol{\theta}_0) \frac{\partial \Lambda(\eta_t; \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right).$$

Let $\widehat{u}_{n,h} = u_{0,h}(\widehat{\boldsymbol{\theta}}_{n,ML})$ where $\widehat{\boldsymbol{\theta}}_{n,ML}$ is the MLE of $\boldsymbol{\theta}_0$. This estimator satisfies

$$\int \Lambda^{\widehat{u}_{n,h}}(x; \widehat{\boldsymbol{\theta}}_{n,ML}) h(x) dx = 1.$$

Note that $\widehat{u}_{n,h}$ is the ML estimator of $u_0$ (because of the functional invariance of the ML estimator) unlike $\widehat{u}_n$ (even when $\widehat{\boldsymbol{\theta}}_n$ is the ML estimator of $\boldsymbol{\theta}_0$).

Suppose that the distribution of $\widehat{\boldsymbol{\theta}}_{n,ML}$ is asymptotically Gaussian, with variance $\boldsymbol{\Sigma}_{ML}$.

Let the test statistic

$$V_n^{(u)} = \frac{\sqrt{n}(u - \widehat{u}_{n,h})}{\widehat{\sigma}_h},$$

where $\widehat{\sigma}_h$ is a consistent estimator of

$$\sigma_h = \left( \frac{\partial u_0}{\partial \boldsymbol{\theta}'} \boldsymbol{\Sigma}_{ML} \frac{\partial u_0}{\partial \boldsymbol{\theta}} \right)^{1/2} = \frac{1}{D_\infty^{(u_0)}} \left( \boldsymbol{r}'_{u_0} \boldsymbol{\Sigma}_{ML} \boldsymbol{r}_{u_0} \right)^{1/2}.$$

**Proposition 3.3.** *Let the assumptions of Theorem 4 (with $\widehat{\boldsymbol{\theta}}_n$ replaced by $\widehat{\boldsymbol{\theta}}_{n,ML}$) and Assumption* **HL8** *hold, and let $\boldsymbol{r}_{u_0} \neq \mathbf{0}$. Then, a test of $\boldsymbol{H}_{0,u}$ at the asymptotic level $\underline{\alpha} \in (0,1)$ is defined by the rejection region*

$$C_V^{(u)} = \left\{ V_n^{(u)} > \Phi^{-1}(1 - \underline{\alpha}) \right\}, \tag{13}$$

*and an asymptotic $100(1 - \underline{\alpha})\%$ CI for $u_0$ is $\widehat{u}_{n,h} \pm n^{-1/2} \Phi^{-1}(1 - \underline{\alpha}) \widehat{\sigma}_h$. Moreover, the test is consistent: under $\boldsymbol{H}_{1,u}$, we have $P\left(C_V^{(u)}\right) \to 1$ as $n \to \infty$.*

#### 3.2.2 When the error density is parameterized

In practical situations, it is unrealistic to assume that the density $h$ of $\eta_t$ is known. Alternatively, the density can be assumed to be known up to some finite parameter: $h(\cdot) = h(\cdot, \boldsymbol{\nu}_0)$ where $\boldsymbol{\nu}_0 \in \mathbb{R}^m$ for $m \in \mathbb{N}$. Let $\boldsymbol{\varphi}_0 = (\boldsymbol{\theta}'_0, \boldsymbol{\nu}'_0)'$ and assume $\boldsymbol{\varphi} \in \boldsymbol{\Phi} \subset \mathbb{R}^{m+d}$. Given $\boldsymbol{\varphi}$, the MME, when it exists, is now the solution $u_0 = u_{0,h}(\boldsymbol{\varphi})$ of

$$\int \Lambda^{u_0}(x; \boldsymbol{\theta}) h(x, \boldsymbol{\nu}) dx = 1.$$

Under **HL8** and

**HL9:** The function $\boldsymbol{\nu} \mapsto \int \Lambda^{u_0}(x;\boldsymbol{\theta})h(x,\boldsymbol{\nu})dx$ is continuously differentiable under the integral sign,

we have

$$\frac{\partial u_{0,h}(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\theta}} = \frac{-1}{D_\infty^{(u_0)}}\boldsymbol{r}_{u_0}, \qquad \frac{\partial u_{0,h}(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\nu}} = \frac{-1}{D_\infty^{(u_0)}}\boldsymbol{s}_{u_0},$$

with $\boldsymbol{s}_{u_0} = E\left(\Lambda^{u_0}(\eta_t;\boldsymbol{\theta}_0)\frac{1}{h(\eta_t;\boldsymbol{\nu}_0)}\frac{\partial h(\eta_t;\boldsymbol{\nu}_0)}{\partial \boldsymbol{\nu}}\right)$. Let $\widehat{u}_{n,\widehat{h}} = u_{0,h}(\widehat{\boldsymbol{\varphi}}_{n,ML})$ where $\widehat{\boldsymbol{\varphi}}_{n,ML} = (\widehat{\boldsymbol{\theta}}_{n,ML}, \widehat{\boldsymbol{\nu}}_{n,ML})$ is the MLE of $\boldsymbol{\varphi}_0$, obtained by solving

$$\int \Lambda^{\widehat{u}_{n,\widehat{h}}}(x;\widehat{\boldsymbol{\theta}}_{n,ML})h(x,\widehat{\boldsymbol{\nu}}_{n,ML})dx = 1. \tag{14}$$

Suppose that the distribution of $\widehat{\boldsymbol{\varphi}}_{n,ML}$ is asymptotically Gaussian, with variance $\boldsymbol{\Phi}_{ML}$. Let the test statistic $W_n^{(u)} = \frac{\sqrt{n}(u-\widehat{u}_{n,\widehat{h}})}{\widehat{\varsigma}_h}$ where $\widehat{\varsigma}_h$ is a consistent estimator of

$$\varsigma_h = \frac{1}{D_\infty^{(u_0)}}\left\{\left(\boldsymbol{r}'_{u_0}, \boldsymbol{s}'_{u_0}\right)\boldsymbol{\Phi}_{ML}\left(\boldsymbol{r}'_{u_0}, \boldsymbol{s}'_{u_0}\right)'\right\}^{1/2}.$$

**Proposition 3.4.** *Let the assumptions of Proposition 3.3 hold, along with Assumption* **HL9** *(with $h(\cdot)$ replaced by $h(\cdot;\boldsymbol{\nu}_0)$ in* **HL8***), and let $(\boldsymbol{r}'_{u_0}, \boldsymbol{s}'_{u_0}) \neq \mathbf{0}$. Then, a test of $\boldsymbol{H}_{0,u}$ at the asymptotic level $\underline{\alpha} \in (0,1)$ is defined by the rejection region*

$$C_W^{(u)} = \left\{W_n^{(u)} > \Phi^{-1}(1-\underline{\alpha})\right\}, \tag{15}$$

*and an asymptotic $100(1-\underline{\alpha})\%$ CI for $u_0$ is $\widehat{u}_{n,\widehat{h}} \pm n^{-1/2}\Phi^{-1}(1-\underline{\alpha})\widehat{\varsigma}_h$.*

We now consider an important particular sub-class of Model (1).

## 4 Multiplicative/Garch-type models

Tests of the existence of *even-order moments* for standard GARCH models have been studied by Francq and Zakoïan [31]. In this setup, the problem reduces to the derivation of the joint asymptotic distribution of the QML estimator of the volatility parameter and of a vector of moments of the innovations process (see Heinemann [37] for a bootstrap-based approach). However, this approach cannot be extended to other GARCH formulations for which the moment conditions are less explicit. Moreover, it cannot be used for general moments, in particular non-even power moments. Here, we consider the class of augmented GARCH processes (see e.g. Aue et al. [1]), defined as

$$\begin{cases} \epsilon_t &= \sigma_t(\boldsymbol{\theta}_0)\eta_t, \\ \sigma_t^\delta(\boldsymbol{\theta}_0) &= \omega(\eta_{t-1};\boldsymbol{\theta}_0) + a(\eta_{t-1};\boldsymbol{\theta}_0)\sigma_{t-1}^\delta(\boldsymbol{\theta}_0), \end{cases} \tag{16}$$

where $E\eta_t^2 = 1$, and $\delta > 0$ is given. [4] Necessary and sufficient conditions for the finiteness of moments of GARCH and augmented GARCH models have been derived by Ling and McAleer [42], Aue et al. [1], and Hörmann [39]. This class includes most of the first-order GARCH-type specifications proposed in the literature. Examples of commonly used specifications are provided in the appendix Section D. Assume that, for any $\boldsymbol{\theta} \in \Theta$, the functions $\omega(\cdot;\boldsymbol{\theta})$ and $a(\cdot;\boldsymbol{\theta})$ are differentiable and satisfy $\omega(\cdot;\boldsymbol{\theta}) : \mathbb{R} \to [\underline{\omega}, +\infty)$ and $a(\cdot;\boldsymbol{\theta}) : \mathbb{R} \to [\underline{a}, +\infty)$ with $\underline{\omega} > 0$ and $\underline{a} \geqslant 0$. The detailed assumptions required for the multiplicative model (16) are listed in Appendix A (labelled **MM1-MM9**). .

[4] The case where $\delta$ is estimated will be considered in Section 4.3.

### 4.1 Estimating the MDF and related tests

It is clear that Model (16) is of the form (1) with $y_t = \epsilon_t$, $f_t = \sigma_t^\delta$ and, omitting $\boldsymbol{\theta}_0$ for simplicity, $\varphi(\eta, f) = \omega(\eta) + a(\eta)f$, and $g(f, \eta) = f^{1/\delta}\eta$. We also have $g^*(f, \epsilon) = f^{-1/\delta}\epsilon$ and $\Lambda(\eta) = a(\eta)$. Assumptions **A0-A1** are thus satisfied under **MM1**. Under condition **MM2**, there exists a stationary ergodic solution to the SRE

$$\sigma_t^\delta(\boldsymbol{\theta}) \;=\; \omega\left(\frac{\epsilon_{t-1}}{\sigma_{t-1}(\boldsymbol{\theta})}; \boldsymbol{\theta}\right) + a\left(\frac{\epsilon_{t-1}}{\sigma_{t-1}(\boldsymbol{\theta})}; \boldsymbol{\theta}\right)\sigma_{t-1}^\delta(\boldsymbol{\theta}), \qquad t \in \mathbb{Z}, \tag{17}$$

and **A2**-**A3** are satisfied. The sequence $\widetilde{\sigma}_t^\delta(\boldsymbol{\theta})$ satisfies the same SRE, but for $t \geqslant 1$ with initial values $\widetilde{\sigma}_0$ and $\widetilde{y}_0$ as in (5).

Theorem 1 takes the following simplified form for Model (16).

*Corollary* 1 (**Augmented GARCH models**). For a strongly consistent estimator of $\boldsymbol{\theta}_0$ satisfying **HL2**, if **MM1-MM5** and **MM6**$(u)$ with $u \in (0, s/2]$ hold, we have

$$\sqrt{n}\left(S_n^{(u)} - S_\infty^{(u)}\right) \xrightarrow{d} \mathcal{N}\left(0, v_u^2 = \boldsymbol{g}_u'\boldsymbol{\Sigma}\boldsymbol{g}_u + \psi_u + 2\boldsymbol{g}_u'\boldsymbol{\xi}_u\right), \tag{18}$$

with $S_n^{(u)} = \frac{1}{n}\sum_{t=1}^n a^u(\widehat{\eta}_t; \widehat{\boldsymbol{\theta}}_n)$, $S_\infty^{(u)} = E\{a^u(\eta_1; \boldsymbol{\theta}_0)\}$, $\boldsymbol{g}_u = E(\boldsymbol{g}_{u,t})$ and $\boldsymbol{g}_{u,t} = \left[\frac{\partial}{\partial\boldsymbol{\theta}} a^u\{\eta_t(\boldsymbol{\theta}); \boldsymbol{\theta}\}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$.

One example of an Equation (16)-type model is Ding et al.'s [19] APARCH (asymmetric power ARCH) model defined by $\omega(\eta) = \omega$, $a(\eta) = \alpha_+|\eta|^\delta \mathbb{1}_{\eta>0} + \alpha_-|\eta|^\delta \mathbb{1}_{\eta<0} + \beta$. For APARCH models estimated by Gaussian QML, the assumptions of Corollary 1 can be considerably reduced.

*Corollary* 2 (**APARCH models estimated by QML**). Under the assumptions:

i) $P(\eta_t > 0) \in (0, 1)$, the support of the distribution of $\eta_t$ contains at least three points, and $E(|\eta_t|^{s\delta}) < \infty$ with $s\delta \geqslant 4$,

ii) $\Theta \subset [\underline{\omega}, \infty) \times (0, \infty)^2 \times [0, 1)$ is compact and $\boldsymbol{\theta}_0 = (\omega_0, \alpha_{0,+}, \alpha_{0,-}, \beta_0)' \in \overset{\circ}{\Theta}$,

iii) $E \log a(\eta_1, \boldsymbol{\theta}_0) < 0$,

(18) holds when $\widehat{\boldsymbol{\theta}}_n$ is the QML estimator of $\boldsymbol{\theta}_0$ and $u \leqslant s/2$.

The GARCH(1,1) process is a particular case of this APARCH model, obtained for $\delta = 2$ and $a(\eta) = \alpha\eta^2 + \beta$. In the appendix (Corollary 7) we provide an explicit expression of the asymptotic variance $v_u^2$ in (18) when (16) is a GARCH(1,1) model and $\widehat{\boldsymbol{\theta}}_n$ is the ML or QML estimator.

Under the conditions

$$E|\eta_1|^{u\delta} < \infty, \quad E\omega^u(\eta_1) < \infty, \tag{19}$$

the testing problem (8) in Model (16) is equivalent to $\boldsymbol{H}_{0,u}:\ E(|\epsilon_t|^{u'\delta}) < \infty, \forall u' < u$, against $\boldsymbol{H}_{1,u}:\ E(|\epsilon_t|^{u\delta}) = \infty$ (see Ling and McAleer [42] and Aue et al. [1]).

*Corollary* 3 ($T_n$ **test for augmented GARCH models**). Under the assumptions of Corollary 1 with $v_u > 0$ and assuming (19), the conclusion of Proposition 2.3 holds: a test of $\boldsymbol{H}_{0,u}$ at the asymptotic level $\underline{\alpha} \in (0, 1)$ is defined by the rejection region $C_T^{(u)} = \left\{T_n^{(u)} > \Phi^{-1}(1 - \underline{\alpha})\right\}$.

This result is an extension of a test studied by Francq and Zakoian [31] in the case where $u$ is even and $(\epsilon_t)$ follows a standard GARCH process (see the appendix Section E for details).

## 4.2 Estimating the MME and related tests

First, we provide results complementing Proposition 2.2. The following result shows that the tail index of $\epsilon_t$ is closely related to $u_0$.

**Proposition 4.1 (Tail index).** *When $E \log a(\eta_1) < 0$ and $1 < E\{a^s(\eta_1)\} < \infty$ for some $s > 0$, if the law of $\log a(\eta_1)$ is nonarithmetic (i.e. not supported by any arithmetic progression $h\mathbb{Z}$), and if $Ea(\eta_1)^{u_0} \log^+ a(\eta_1) < \infty$, there exists $c > 0$ such that $P(\sigma_t > x) \sim cx^{-\delta u_0}$, and $P(|\epsilon_t| > x) \sim E|\eta_t|^{\delta u_0} P(\sigma_t > x)$, as $x \to \infty$.*

These tail properties–established for standard GARCH models by Mikosch and Stărică [44] and for augmented GARCH models by Zhang and Ling [53]–show that, under mild additional assumptions, the coefficient $\delta u_0$ is also the *tail index* of augmented GARCH processes. Conditions for the existence of a tail index for general SRE were derived by Basrak et al. [3], and Kesten [40] characterized this coefficient as the solution of an equation taking the form $E\{a^{u_0}(\eta_1)\} = 1$ in the case of an augmented GARCH(1,1) process.

In the case of multiplicative models, **HL4** is the consequence of more primitive conditions, allowing the assumptions of Theorem 2 to be simplified.

*Corollary* 4 (**Strong consistency of $\widehat{u}_n$**). For Model (16), under **MM1-MM5** and **MM7**, if $\widehat{\boldsymbol{\theta}}_n$ is a strongly consistent estimator of $\boldsymbol{\theta}_0$, the conclusions of Theorem 2 hold.

Now, we give more explicit conditions for the weak convergence in Theorem 3, and the asymptotic distribution in Theorem 4, to hold. We refer to Section 3.1 for the weak convergence notation.

*Corollary* 5 (**Asymptotic distribution of $\widehat{u}_n$**). Under the assumptions of Corollary 1 and if **MM6**($s$) and **MM7** hold, for any $[u_1, u_2] \subset (0, s/2)$, we have $\sqrt{n}\left(S_n^{(u)} - S_\infty^{(u)}\right) \overset{\mathcal{C}[u_1,u_2]}{\Longrightarrow} \Gamma(u)$.

If, in addition, $1 < E\{a^s(\eta_1)\} < \infty$ and $u_0 < s/2$, we have $\sqrt{n}(\widehat{u}_n - u_0) \xrightarrow{d} \mathcal{N}\left\{0, \left(D_\infty^{(u_0)}\right)^{-2} v_{u_0}^2\right\}$.

This result allows the calculation of asymptotic CIs for MME $u_0$ and thus, by Proposition 4.1, for the tail index of the distribution of $\epsilon_t$.

*Remark* 2 (**Comparison with Hill's estimator of the tail index**). It is well known that Hill's estimator ([38]) crucially depends on which part of the sample it is calculated on (see for instance Figure 1 in Zhu and Ling [54]). Moreover, Baek et al. [2] showed that the Hill estimator is extremely biased for estimating the tail index of ARCH-type models. Even for i.i.d. data and very large samples, estimating the tail index using Hill's estimator is very challenging unless the underlying data comes from a Pareto distribution[5] (this will be illustrated in Section 6 using Student distributions and real series). Deriving CIs for the tail index using Hill's estimator is even more challenging. By Proposition 4.1 and Corollary 5, one can estimate the tail index of augmented GARCH models at a parametric rate, instead of resorting to extreme value statistics. A similar situation occurs in estimating the density of a GARCH(1,1) models since, by exploiting the dynamic structure of the model, Delaigle et al. [18] managed to obtain a root-$n$ consistent estimator. Trapani [49] also noted that Hill's estimation of the tail index "is fraught with difficulties" and proposed a randomised testing procedure applied on sample moments for testing for (in)finite moments in a general nonparametric framework.

In the case of Model (16), the test based on $U_n^{(u)}$ takes the following form.

[5] According to Drees et al. [20], "One would have to be paranormal to discern with confidence the true value from the Hill plot."

*Corollary* 6 ($U_n$ **test for augmented GARCH models**). Under the assumptions of Corollary 5, if $w^2_{u_0} > 0$, and (19) holds, a test of $\boldsymbol{H}_{0,u}$ at the asymptotic level $\underline{\alpha}$ is defined by the rejection region $C_U^{(u)} = \left\{U_n^{(u)} > \Phi^{-1}(1-\underline{\alpha})\right\}$, and an asymptotic $100(1-\underline{\alpha})\%$ CI for $u_0$ is $\hat{u}_n \pm n^{-1/2}\Phi^{-1}(1-\underline{\alpha})\hat{w}_{\hat{u}_n}$.

Now we turn to purely parametric estimators of the MME. Under regularity assumptions (derived by Berkes and Horváth [4] for a standard GARCH($p, q$) model), the MLE of $\boldsymbol{\theta}_0$ satisfies the expansion given in **MM8**. Let

$$\sigma_h = \frac{1}{D_\infty^{(u_0)}}\left(\frac{4}{\iota_h}\boldsymbol{r}'_{u_0}\boldsymbol{J}^{-1}\boldsymbol{r}_{u_0}\right)^{1/2}, \quad \boldsymbol{r}_{u_0} = E\left\{u_0 a^{u_0-1}(\eta_t;\boldsymbol{\theta}_0)\frac{\partial a(\eta_t;\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right\},$$

where $\iota_h$, defined in **MM8**, refers to the Fisher information for scale (whose existence is guaranteed by Assumption FIS). For the classical GARCH(1,1) model, we have $\boldsymbol{r}_{u_0} = u_0\boldsymbol{m}_{u_0}$ where $\boldsymbol{m}_{u_0}$ is defined in the appendix, Corollary 7. Proposition 3.3 can be specialized as follows.

**Proposition 4.2** ($V_n$ **test for augmented GARCH models).** *Let for $s > 0$, $1 < E\{a^s(\eta_1)\} < \infty$, with $u_0 < s/2$. Let* **HL8** *(with $\lambda(\cdot) = a(\cdot)$), Assumptions (19),* **FIS** *and* **MM8** *hold, and let $\boldsymbol{r}_{u_0} \neq \mathbf{0}$. Then, a test of $\boldsymbol{H}_{0,u}$ at the asymptotic level $\underline{\alpha}$ is defined by the rejection region $C_V^{(u)} = \left\{V_n^{(u)} > \Phi^{-1}(1-\underline{\alpha})\right\}$, and an asymptotic $100(1-\underline{\alpha})\%$ CI for $u_0$ is $\hat{u}_{n,h} \pm n^{-1/2}\Phi^{-1}(1-\underline{\alpha})\hat{\sigma}_h$ where $\hat{\sigma}_h$ is a consistent estimator of $\sigma_h$.*

When the error density is parameterized, the asymptotic properties of the MLE of $\boldsymbol{\varphi}_0$ were established by Straumann (Chapter 6, [46]). For the sake of brevity, we defer to this reference for the precise assumptions underlying these properties. Assuming that the MLE satisfies the Bahadur expansion in **MM9**, the conclusions of Proposition 3.4 hold with $\boldsymbol{\Phi}_{ML}$ replaced by $\mathfrak{J}^{-1}$ (defined in **MM9**).

### 4.3 Selecting $\delta$ in augmented GARCH models

In practice, estimating $\delta > 0$ in Model (16) is very challenging. Even if asymptotic normality of the joint QML estimator of $\delta$ and $\boldsymbol{\theta}_0$ has been established, the value of $\delta$ can be extremely difficult to identify in finite samples (see Hamadeh and Zakoïan [35]). Since the quasi-likelihood is very flat in the direction of $\delta$[6], estimating this coefficient is extremely difficult in practice. For this reason, instead of treating $\delta$ as a real-valued parameter, practitioners tend to select $\delta$ from a finite set of values corresponding to well-known models such as standard or GJR-GARCH models ($\delta = 2$) or T-GARCH models ($\delta = 1$) (see the appendix Section D for definitions and references). To reflect the existence of several candidates for $\delta$, assume that the true value $\delta_0$ belongs to a finite set,

$$\delta_0 \in \mathcal{D} = \{\delta_1, \ldots, \delta_d\}, \quad \delta_i > 0, \quad i = 1, \ldots, d.$$

For the sake of illustration, we focus on APARCH models (see the appendix Section D) and the QML estimator.

Writing the vector of parameters $\boldsymbol{\vartheta} = (\delta, \boldsymbol{\theta}')'$ and assuming $\boldsymbol{\vartheta} \in \mathcal{D} \times \Theta$ where $\Theta$ is a compact subset of $(0,\infty) \times [0,\infty)^3 \times [0,1)$, the true parameter value is denoted $\boldsymbol{\vartheta}_0 = (\delta_0, \boldsymbol{\theta}'_0)'$. To define the QMLE of $\boldsymbol{\vartheta}$, we recursively define $\tilde{\sigma}_t$, for $t \geqslant 1$, by

$$\tilde{\sigma}_t = \tilde{\sigma}_t(\boldsymbol{\vartheta}) = \left\{\omega + \alpha_+(\epsilon^+_{t-1})^\delta + \alpha_-(-\epsilon^-_{t-1})^\delta + \beta\tilde{\sigma}^\delta_{t-1}\right\}^{1/\delta}.$$

[6] See Table 1 and Figure 1 in [35].

A QMLE of $\boldsymbol{\vartheta}$ is defined as any measurable solution $\widehat{\boldsymbol{\vartheta}}_n^{QML} = (\widehat{\delta}_n^{QML}, \widehat{\boldsymbol{\theta}}^{QML'})'$ of $\widehat{\boldsymbol{\vartheta}}_n^{QML} = \arg\min_{\boldsymbol{\vartheta}\in\mathcal{D}\times\Theta} \widetilde{\mathbf{l}}_n(\boldsymbol{\vartheta})$, $\widetilde{\mathbf{l}}_n(\boldsymbol{\vartheta}) = n^{-1}\sum_{t=1}^n \widetilde{\ell}_t$, $\widetilde{\ell}_t = \widetilde{\ell}_t(\boldsymbol{\vartheta}) = \frac{\epsilon_t^2}{\widetilde{\sigma}_t^2} + \ln \widetilde{\sigma}_t^2$.

Let $a(\eta, \boldsymbol{\vartheta}) = \alpha_+ |\eta|^\delta \mathbb{1}_{\eta>0} + \alpha_- |\eta|^\delta \mathbb{1}_{\eta<0} + \beta$ and let $S_n^{(u)} = \frac{1}{n}\sum_{t=1}^n a^u(\widehat{\eta}_t; \widehat{\boldsymbol{\vartheta}}_n)$.

**Proposition 4.3.** *Under the following assumptions: i) $\eta_t$ has a positive density in some neighborhood of zero, $E(|\eta_t|^{s\delta_0}) < \infty$ with $s\delta_0 \geqslant 4$; ii) $\boldsymbol{\theta}_0 \in \overset{\circ}{\Theta}$, and iii) $E \log a(\eta_1, \boldsymbol{\vartheta}_0) < 0$, we have $\widehat{\delta}_n^{QML} = \delta_0$ for $n$ large enough and the weak convergence in Corollary 5 holds. If $\eta_t$ has positive density over the real line, the asymptotic distribution of $\widehat{u}_n$ holds.*

It is because $\mathcal{D}$ is discrete that the effects of estimating $\delta$ do not appear in the asymptotic results.

## 5 Asymptotic power comparisons

In this section, we focus on multiplicative models. To compare the tests of $\boldsymbol{H}_{0,u}$ we first note that, under the assumptions of Theorem 4 and from the proof of this theorem,

$$U_n^{(u_0)} = T_n^{(u_0)} + o_P(1). \tag{20}$$

Thus the statistics are equivalent at the frontier of the null assumption and, from Le Cam's theory, they are also equivalent under local alternatives. In this section, we will compare these tests with the parametric tests and also provide non-local comparisons.

### 5.1 Asymptotic power under local alternatives

Conditional on $\epsilon_0$ and $\sigma_0$, the density of the observations $(\epsilon_1, \ldots, \epsilon_n)$ satisfying (16) is given by $L_{n,h}(\boldsymbol{\theta}_0) = \prod_{t=1}^n \sigma_t^{-1}(\boldsymbol{\theta}_0) h\left\{\sigma_t^{-1}(\boldsymbol{\theta}_0)\epsilon_t\right\}$. Around $\boldsymbol{\theta}_0 \in \overset{\circ}{\Theta}$, consider a sequence of local parameters of the form

$$\boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + \boldsymbol{\tau}/\sqrt{n}, \tag{21}$$

where $\boldsymbol{\tau} \in \mathbb{R}^d$. We denote $P_{n,\boldsymbol{\tau}}$ (resp. $P_0$) the distribution of the observations when the parameter is $\boldsymbol{\theta}_n$ (resp. $\boldsymbol{\theta}_0$). If $1 < E\left\{a^s(\eta_1)\right\} < \infty$ for some $s > 0$, with $u_0 \in (0, u_2)$, for given $h$ and $\boldsymbol{\theta}_0$, there exists a unique $u_0 := u_0(\boldsymbol{\theta}_0, h)$ such that $E\{a^{u_0}(\eta_1)\} = 1$. Without loss of generality, assume that $n$ is sufficiently large so that $\boldsymbol{\theta}_n \in \Theta$. Note that, under appropriate assumptions on $\boldsymbol{\tau}$, the parameter $\boldsymbol{\theta}_n$ belongs to the alternative for testing $\boldsymbol{H}_{0,u_0}$.

Drost and Klaassen [21] showed that for standard GARCH models, the log-likelihood ratio $\Lambda_{n,h}(\boldsymbol{\theta}_n, \boldsymbol{\theta}_0) = \log L_{n,h}(\boldsymbol{\theta}_n)/L_{n,h}(\boldsymbol{\theta}_0)$ satisfies the LAN property

$$\Lambda_{n,h}(\boldsymbol{\theta}_n, \boldsymbol{\theta}_0) = \boldsymbol{\tau}'\boldsymbol{\Delta}_{n,h}(\boldsymbol{\theta}_0) - \frac{1}{2}\boldsymbol{\tau}'\boldsymbol{\mathfrak{I}}_h\boldsymbol{\tau} + o_{P_{\boldsymbol{\theta}_0}}(1), \tag{22}$$

where $\boldsymbol{\mathfrak{I}}_h = \iota_h E\left(\frac{1}{\sigma_t^2}\frac{\partial\sigma_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\right)$ and $\boldsymbol{\Delta}_{n,h}(\boldsymbol{\theta}_0) = \frac{-1}{\sqrt{n}}\sum_{t=1}^n g_1(\eta_t)\frac{1}{\sigma_t}\frac{\partial\sigma_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\mathfrak{I}}_h)$ under $P_{\boldsymbol{\theta}_0}$ as $n \to \infty$. Note that the so-called central sequence $\boldsymbol{\Delta}_{n,h}(\boldsymbol{\theta}_0)$ is conditional on the initial values. It is shown in [22] and [43] that (22) continues to hold for extensions of standard GARCH models. Lee and Taniguchi [41] showed that the initial values have no influence on the LAN property. Together with Le Cam's third lemma, the LAN property allows us to derive the local asymptotic powers (LAPs) of our tests.

**Proposition 5.1.** *Under Assumptions* **FIS**, (22) *and the assumptions of Corollaries 3 and 6, respectively, the LAPs of the tests of $\boldsymbol{H}_{0,u_0}$ defined in (10) and (12) are given by*

$$\lim_{n\to\infty} P_{n,\boldsymbol{\tau}}\left(C_T^{(u_0)}\right) = \lim_{n\to\infty} P_{n,\boldsymbol{\tau}}\left(C_U^{(u_0)}\right) = \Phi\left\{c_{h,u_0}(\boldsymbol{\theta}_0) - \Phi^{-1}(1-\underline{\alpha})\right\}, \tag{23}$$

*where*

$$c_{h,u_0}(\boldsymbol{\theta}_0) = \frac{-\boldsymbol{\tau}'}{\upsilon_{u_0}}\left[E\left(\frac{1}{\sigma_t}\frac{\partial\sigma_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right)E\{a^{u_0}(\eta_1)g_1(\eta_1)\} + E\left(\frac{1}{\sigma_t}\frac{\partial\sigma_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\boldsymbol{g}'_{u_0}\boldsymbol{\Delta}_{t-1}\right)E\{\boldsymbol{V}(\eta_1)g_1(\eta_1)\}\right].$$

In standard GARCH(1,1) models estimated by QML and ML, the calculations reported in the appendix Section G show that, with obvious notations,

$$c^{QML}_{h,u_0}(\boldsymbol{\theta}_0) \leqslant c^{ML}_{h,u_0}(\boldsymbol{\theta}_0). \tag{24}$$

Again, in standard GARCH(1,1) models, $u_0$ decreases as $\alpha_1$ or $\beta_1$ increases: $u_0(\boldsymbol{\theta}_0+\boldsymbol{e},h) < u_0(\boldsymbol{\theta}_0,h)$ for all directions $\boldsymbol{e}=(0,e_2,e_3)'$ such that $e_2\geqslant 0$ and $e_3\geqslant 0$, with at least one inequality being strict. In the more general case where the power $u_0$ decreases when the parameter increases in a given direction $\boldsymbol{e}\in\mathbb{R}^d$, we are able to derive the power of asymptotically locally uniformly most powerful unbiased (UMPU) tests and provide conditions for the tests $T$ and $U$ to be optimal in this sense.

**Proposition 5.2.** *Assume that $u_0(\boldsymbol{\theta}_0+\frac{\boldsymbol{e}}{\sqrt{n}},h) < u_0(\boldsymbol{\theta}_0,h)$ for all $n$. Then, under the assumptions of Proposition 5.1, any asymptotically locally UMPU test for $\boldsymbol{H}_{0,u}:\ u_0(\boldsymbol{\theta}_0,h)>u$ against $\boldsymbol{H}_{1,n,u}:\ u_0(\boldsymbol{\theta}_0+\frac{\boldsymbol{e}}{\sqrt{n}},h)\leqslant u$ has asymptotic power bounded by*

$$\lim_{n\to\infty} P_{\mathbf{H_{1,n,u}}}(C) = \Phi\left\{c_{\boldsymbol{e}} - \Phi^{-1}\left(1-\underline{\alpha}\right)\right\}, \quad \textit{with} \quad c_{\boldsymbol{e}} = \frac{\iota_h^{1/2}\boldsymbol{e}'\boldsymbol{e}}{2\sqrt{\boldsymbol{e}'\boldsymbol{J}^{-1}\boldsymbol{e}}}. \tag{25}$$

The assumption on the MME of Proposition 5.2 is satisfied for any commonly used GARCH-type model where the volatility increases with any component of the parameter.

*Remark* 3 (**Testing the existence of the second-order moment in standard GARCH models**). For standard GARCH(1,1) models with $u_0=1$ and $\boldsymbol{e}=(0,1,1)'$, the tests $C_T^{(1)}$ and $C_U^{(1)}$ obtained by QML/ML estimation are optimal if and only if the density of $\eta_t$ has the form

$$h(y) = \frac{a^a}{\Gamma(a)}e^{-ay^2}|y|^{2a-1}, \quad a>0, \quad \Gamma(a)=\int_0^\infty t^{a-1}e^{-t}dt. \tag{26}$$

The following result gives the LAPs of the test assuming the density is known.

**Proposition 5.3.** *Under the assumptions of Propositions 4.2 and 5.1, the LAPs of the test of $\boldsymbol{H}_{0,u_0}$ defined in (13) is given by*

$$\lim_{n\to\infty} P_{n,\boldsymbol{\tau}}\left(C_V^{(u_0)}\right) = \Phi\left\{d_{h,u_0}(\boldsymbol{\theta}_0) - \Phi^{-1}(1-\underline{\alpha})\right\}, \quad d_{h,u_0}(\boldsymbol{\theta}_0) = \boldsymbol{r}'_{u_0}\boldsymbol{\tau}\Big/\sqrt{\frac{4}{\iota_h}\boldsymbol{r}'_{u_0}\boldsymbol{J}^{-1}\boldsymbol{r}_{u_0}}. \tag{27}$$

Under the assumptions of Proposition 5.2, the test $C_V^{(u_0)}$ is optimal if $d_{h,u_0}(\boldsymbol{\theta}_0)=c_{\boldsymbol{\tau}}$, that is if the vectors $\boldsymbol{r}_{u_0}$ and $\boldsymbol{\tau}$ are collinear in the same direction.

Next, we turn to the case of Section 3.2.2 where the error density is parameterized and estimated. Around $\boldsymbol{\varphi}_0=(\boldsymbol{\theta}'_0,\boldsymbol{\nu}'_0)'\in\overset{\circ}{\boldsymbol{\Phi}}$, we now consider a sequence of local parameters of the form

$$\boldsymbol{\theta}_n=\boldsymbol{\theta}_0+\boldsymbol{\tau}_1/\sqrt{n}, \quad \boldsymbol{\nu}_n=\boldsymbol{\nu}_0+\boldsymbol{\tau}_2/\sqrt{n}, \tag{28}$$

where $\boldsymbol{\tau}_1\in\mathbb{R}^d,\boldsymbol{\tau}_2\in\mathbb{R}^m$. We still denote $P_{n,\boldsymbol{\tau}}$ (resp. $P_0$) the distribution of observations when the parameter is $\boldsymbol{\varphi}_n=(\boldsymbol{\theta}'_0+\boldsymbol{\tau}'_1/\sqrt{n},\boldsymbol{\nu}'_0+\boldsymbol{\tau}'_2/\sqrt{n})' := \boldsymbol{\varphi}_0+\boldsymbol{\tau}/\sqrt{n}$ (resp. $\boldsymbol{\varphi}_0$). Let the log-likelihood ratio $\Lambda_n(\boldsymbol{\varphi}_0+\boldsymbol{\tau}/\sqrt{n},\boldsymbol{\varphi}_0)=\log L_{n,h}(\boldsymbol{\varphi}_n)/L_{n,h}(\boldsymbol{\varphi}_0)$.

As shown in Drost and Klaassen [21], the LAN property (22) holds when the density $h$ can be treated as an infinite-dimensional nuisance parameter. In Francq and Zakoïan [32], we showed that the LAN property also holds in the parametric framework of this section: a Taylor expansion around $\boldsymbol{\varphi}_0$ of the log-likelihood ratio yields

$$\Lambda_{n,h}(\boldsymbol{\varphi}_n, \boldsymbol{\varphi}_0) = \boldsymbol{\tau}' \boldsymbol{\Delta}_{n,h}(\boldsymbol{\varphi}_0) - \frac{1}{2} \boldsymbol{\tau}' \mathfrak{J}_n(\boldsymbol{\varphi}_0) \boldsymbol{\tau} + o_{P_{\boldsymbol{\theta}_0}}(1), \tag{29}$$

where $\mathfrak{J}_n(\boldsymbol{\varphi}_0)$ is a consistent estimator of $\mathfrak{J}$ and, under $P_0$,

$$\boldsymbol{\Delta}_{n,h}(\boldsymbol{\varphi}_0) = \left( \frac{-1}{\sqrt{n}} \sum_{t=1}^{n} g_1(\eta_t, \boldsymbol{\nu}_0) \frac{1}{\sigma_t} \frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}, \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{1}{h(\eta_t, \boldsymbol{\nu}_0)} \frac{\partial h(\eta_t, \boldsymbol{\nu}_0)}{\partial \boldsymbol{\nu}'} \right)' \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathfrak{J}). \tag{30}$$

The next result provides the LAPs of test $W$.

**Proposition 5.4.** *Under Assumptions* **FIS**, (29)-(30), *and the assumptions of Proposition 3.4, the LAPs of the test of $\boldsymbol{H}_{0,u_0}$ defined in (15) is given by*

$$\lim_{n\to\infty} P_{n,\boldsymbol{\tau}} \left( C_W^{(u_0)} \right) = \Phi \left\{ e_{h,u_0}(\boldsymbol{\theta}_0) - \Phi^{-1}(1-\underline{\alpha}) \right\}, \quad e_{h,u_0}(\boldsymbol{\theta}_0) = \frac{\boldsymbol{r}'_{u_0}\boldsymbol{\tau}_1 + \boldsymbol{s}'_{u_0}\boldsymbol{\tau}_2}{\sqrt{\left(\boldsymbol{r}'_{u_0}, \boldsymbol{s}'_{u_0}\right) \mathfrak{J}^{-1} \left(\boldsymbol{r}'_{u_0}, \boldsymbol{s}'_{u_0}\right)'}}.$$

Under the assumptions of Proposition 5.2, the test $C_W^{(u_0)}$ is optimal if $e_{h,u_0}(\boldsymbol{\theta}_0) = c_{\boldsymbol{\tau}}$, that is, if the vectors $\left(\boldsymbol{r}'_{u_0}, \boldsymbol{s}'_{u_0}\right)$ and $\boldsymbol{\tau}'$ are collinear in the same direction. An example of the calculation and comparison of the LAPs of tests $T, U, V$ and $W$ is given in the appendix Section H. It is shown that, for GARCH(1,1) models with Student innovations, tests $T$ and $U$ are dominated by test $W$. As expected, test $V$ is locally asymptotically more efficient than the other tests.

### 5.2 Comparisons based on Bahadur slopes

To be able to distinguish tests $T$ and $U$, we turn to the Bahadur approach. We will also compare them with tests $V$ and $W$, which require knowledge or estimation of the density. Recall that the Bahadur slope is defined as the almost sure limit of $-2/n$ times the logarithm of the $p$-value of the test. In Bahadur's sense, one test is more efficient than another if the slope of the first test is greater than the slope of the second test.

**Proposition 5.5.** *In Bahadur's sense, test $T_n^{(u)}$ is more efficient than $U_n^{(u)}$ if and only if*

$$\frac{\left(S_\infty^{(u)} - 1\right)^2}{(u-u_0)^2} \frac{\upsilon_{u_0}^2}{\left(E[a^{u_0}(\eta_1;\boldsymbol{\theta}_0)\log\{a(\eta_1;\boldsymbol{\theta}_0)\}]\right)^2 \upsilon_u^2} > 1,$$

*and test $W_n^{(u)}$ is more efficient than $U_n^{(u)}$ if and only if* $\upsilon_{u_0}^2 \left(\boldsymbol{r}'_{u_0}, \boldsymbol{s}'_{u_0}\right) \mathfrak{J}^{-1} \left(\boldsymbol{r}'_{u_0}, \boldsymbol{s}'_{u_0}\right)' > 1.$

Note that the latter condition does not depend on $u$, i.e. on the alternative. Examples and comparisons of asymptotic slopes are given in the appendix Section H, showing that test $T$ is in general less efficient than the others.

## 6 Empirical application

Davis and Mikosch [16] noted that *"In applications to real-life data one often observes that the sum of the estimated parameters $\widehat{\alpha}_1 + \widehat{\beta}_1$ is close to 1 implying that moments slightly larger than* *two might not exist for a fitted GARCH process."* Francq and Zakoïan [31] made a first attempt to check this intuition by considering the returns of the French energy company Total SA, one of the main constituents of the CAC40 index, over the period 2001-07-16 to 2018-09-21. On this series of 4418 observations, we fitted a standard GARCH(1,1) model and, using $T_n^{(u)}$ to test the existence of *even-order moments*, found strong evidence for the existence of the second-order marginal moment and suspicions of non-existence of the 8-th order moment. Given that (i) tests based on $T_n^{(u)}$ often turn out to be much less powerful than those based on $U_n^{(u)}$ and $W_n^{(u)}$; (ii) the finiteness of any *positive-order* moment can be tested, and (iii) our analysis is not restricted to standard GARCH models, it should be possible to improve on the results obtained in [31].

We thus re-investigated the same series with APARCH(1,1) models, using the QMLE for tests $T_n^{(u)}$, $U_n^{(u)}$ and $V_n^{(u)}$ (the QMLE is actually the Gaussian MLE in the latter case), and the MLE, assuming a standardized Student distribution with $\nu$ degrees of freedom for the i.i.d. innovations, for the $W_n^{(u)}$ test. We searched $\delta \in \{0.5, 1, 1.5, 2\}$, and estimated the optimal value $\delta = 1$ with both the QML and ML estimators. The volatility model estimated by QML is

$$\sigma_t = \underset{(0.006)}{0.037} + \underset{(0.010)}{0.018}|\epsilon_{t-1}|\mathbb{1}_{\epsilon_{t-1}>0} + \underset{(0.012)}{0.132}|\epsilon_{t-1}|\mathbb{1}_{\epsilon_{t-1}<0} + \underset{(0.009)}{0.916}\sigma_{t-1}$$

where the estimated standard deviations in brackets are obtained from the asymptotic distribution of the QMLE.

The model estimated by Student-ML is

$$\sigma_t = \underset{(0.007)}{0.033} + \underset{(0.010)}{0.016}|\epsilon_{t-1}|\mathbb{1}_{\epsilon_{t-1}>0} + \underset{(0.016)}{0.126}|\epsilon_{t-1}|\mathbb{1}_{\epsilon_{t-1}<0} + \underset{(0.013)}{0.922}\sigma_{t-1}, \ \eta_t \sim \mathrm{St}(\underset{(1.7)}{11.1})$$

where $\mathrm{St}(\nu)$ denotes the standardized Student distribution with $\nu$ degrees of freedom. Note that the volatilities estimated by QML and ML are almost the same. The results presented in the appendix Section J show no dependence in the QMLE and MLE residuals and that the distribution of the residuals is better reproduced by a Student distribution than by a Gaussian distribution (in particular the empirical kurtosis of the QMLE and MLE residuals are respectively 3.807 and 3.816, which is much closer to the kurtosis of the fitted Student distribution, $3+6/(\nu-4) = 3.848$, than to the Gaussian kurtosis). Table 1 shows that the tests based on $U_n^{(u)}$ and $W_n^{(u)}$ are much more conclusive than the test based on $T_n^{(u)}$. The test based on $V_n^{(u)}$ does not seem reliable since we have seen that the empirical distribution of the residuals is far from Gaussian. The estimated maximum moment order is $\widehat{u}_0 = 7.9$ with the $U_n^{(u)}$ statistic, and $\widehat{u}_0 = 7.8$ with the $W_n^{(u)}$ statistic. At an asymptotic confidence level of 95%, the estimated CIs for $u_0$ is $[4.5, 11.3]$ with the $U_n^{(u)}$ statistic and $[5.9, 9.6]$ with the $W_n^{(u)}$ statistic. The estimated value of $u_0$ based on $U_n^{(u)}$ is $\widehat{u}_0 = 7.9$. To evaluate the variability of this estimator without using asymptotic theory, we simulated APARCH(1,1) models with a parameter $\widehat{\boldsymbol{\theta}}_n$–the QMLE calculated on the Total series–and noise with a distribution equal to that of the QML residuals. The empirical 95% CI for $u_0$ obtained from 10000 bootstrap replications is $[5.7, 9.8]$, which is similar to the estimates based on asymptotic theory. The two estimation methods based on $U_n^{(u)}$ and $W_n^{(u)}$ therefore give similar estimated tail indexes but, as expected, the CI from the fully parametric method based on $W_n^{(u)}$ is tighter. These results strongly support the existence of finite moments of order 5 or 6, allowing the validation of certain statistical procedures, such as the construction of confidence intervals for the prediction of the squared returns at long horizons. In contrast, illustrating Remark 2 and footnote [5], Figure 2 shows that the conventional Hill estimator provides little information on the value of the tail index, both for the Total series (left graph) and for a simulation of the model for which $u_0 = 7.8$ is known to be the maximum moment order. Note that Figure 2 is in perfect agreement with Figures 2 and 3 in Baek et al. [2].

Table 1: Test statistics $T_n^{(u)}$, $U_n^{(u)}$, $V_n^{(u)}$ (assuming Gaussian innovations), $W_n^{(u)}$ (assuming Student innovations) based on a APARCH(1,1) model for the Total return series.

| $u$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $T_n^{(u)}$ | -4.71 | -4.18 | -3.50 | -2.72 | -1.91 | -1.15 | -0.49 | 0.04 | 0.46 | 0.76 | 0.97 | 1.11 |
| $U_n^{(u)}$ | -3.94 | -3.37 | -2.80 | -2.23 | -1.66 | -1.09 | -0.52 | 0.05 | 0.62 | 1.19 | 1.76 | 2.33 |
| $V_n^{(u)}$ | -7.25 | -6.23 | -5.21 | -4.19 | -3.18 | -2.16 | -1.14 | -0.12 | 0.89 | 1.91 | 2.93 | 3.95 |
| $W_n^{(u)}$ | -6.88 | -5.86 | -4.84 | -3.82 | -2.81 | -1.79 | -0.77 | 0.25 | 1.26 | 2.28 | 3.30 | 4.32 |

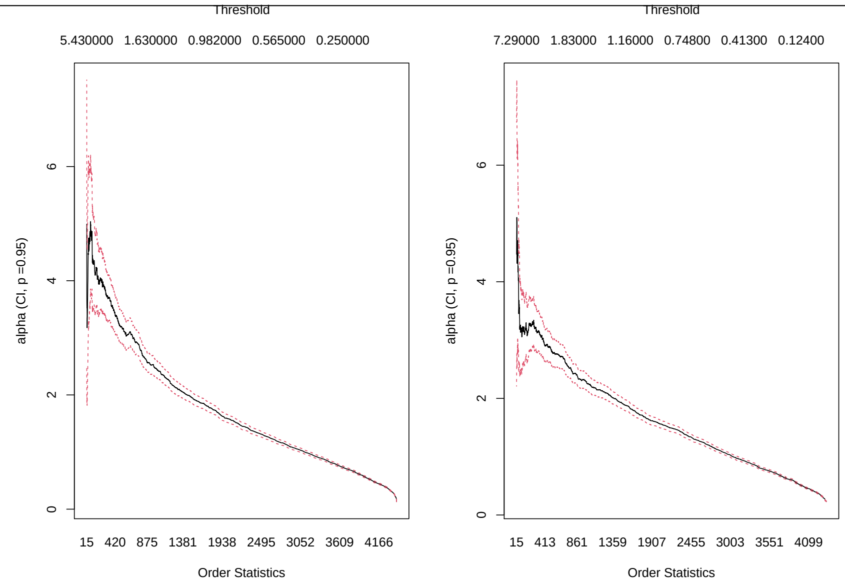


Figure 2: Hill plots of a simulation of an APARCH model with tail index 7.8 (left graph) and of the absolute value of the Total return series (right graph).

The conclusions drawn from this study are: 1) that the estimators proposed here are much more effective than the Hill estimator in assessing the value of the tail index of a GARCH-type model; 2) that estimating the maximum moment order is a difficult problem (since the CIs remain large, even in a fully parametric framework); 3) that at least for the Total series, moments seem to exist at orders much larger than two, moderating the overly pessimistic statement quoted at the beginning of this section.

# Appendix

This Appendix provides assumptions, proofs, examples of augmented GARCH models, complementary results for the MDF of GARCH models, power comparisons of tests, Monte-Carlo experiments and a supplement to the empirical application.

## A Assumptions

The following assumptions are used for the multiplicative model in Section 4. Let $\rho$ be a generic constant belonging to $(0,1)$.

**MM1:** $E\left\{\omega^s(\eta_1,\boldsymbol{\theta}_0)\right\}<\infty$, $E\log a(\eta_1,\boldsymbol{\theta}_0)<0$ and $E\left\{a^s(\eta_1,\boldsymbol{\theta}_0)\right\}<\infty$ for $s>0$.

**MM2:** For all $\boldsymbol{\theta}\in\Theta$ and $\epsilon\in\mathbb{R}$, the function $z\mapsto\omega\left(\frac{\epsilon}{z^{1/\delta}};\boldsymbol{\theta}\right)+a\left(\frac{\epsilon}{z^{1/\delta}};\boldsymbol{\theta}\right)z$ is differentiable over $[\underline{\omega},+\infty)$. There exists $z_0>\underline{\omega}$ such that $E\log^+\sup_{\boldsymbol{\theta}\in\Theta}\omega\left(\frac{\epsilon_t}{z_0^{1/\delta}};\boldsymbol{\theta}\right)<\infty$, $E\log^+\sup_{\boldsymbol{\theta}\in\Theta}a\left(\frac{\epsilon_t}{z_0^{1/\delta}};\boldsymbol{\theta}\right)<\infty$, and we have

$$E\log\sup_{z\geqslant\underline{\omega}}\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\partial}{\partial z}\left\{\omega\left(\frac{\epsilon_t}{z^{1/\delta}};\boldsymbol{\theta}\right)+a\left(\frac{\epsilon_t}{z^{1/\delta}};\boldsymbol{\theta}\right)z\right\}\right|<0.$$

**MM3:** The $\mathcal{F}_{t-1}$-measurable function $\boldsymbol{\theta}\to(\sigma_t(\boldsymbol{\theta}),\widetilde{\sigma}_t(\boldsymbol{\theta}))$ is a.s. twice continuously differentiable. Moreover, $\sup_{\boldsymbol{\theta}\in\Theta}|\sigma_t(\boldsymbol{\theta})-\widetilde{\sigma}_t(\boldsymbol{\theta})|+\left|\frac{\partial\sigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}-\frac{\partial\widetilde{\sigma}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|\leqslant K_t\rho^t$ where $K_t\in\mathcal{F}_{t-1}$ and $\sup_t E(K_t^r)<\infty$ for some $r>0$.

**MM4:** There exists a neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that $E\left(\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta})}\right)^r<\infty$ and $E\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\|\partial\sigma_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\|^r<\infty$ for some $r>0$.

**MM5:** For almost all $\epsilon$, the function $(\sigma,\boldsymbol{\theta})\mapsto a(\frac{\epsilon}{\sigma};\boldsymbol{\theta})$ is twice differentiable over $[\underline{\omega},+\infty)\times V(\boldsymbol{\theta}_0)$ and there exist $C,\tau>0$ such that, for any $(\epsilon,\sigma,\boldsymbol{\theta})\in\mathbb{R}\times[\underline{\omega},+\infty)\times V(\boldsymbol{\theta}_0)$,

$$\max\left\{a\left(\frac{\epsilon}{\sigma};\boldsymbol{\theta}\right),\left|\frac{\partial\log a(\frac{\epsilon}{\sigma};\boldsymbol{\theta})}{\partial\sigma}\right|,\left|\frac{\partial^2\log a(\frac{\epsilon}{\sigma};\boldsymbol{\theta})}{\partial\sigma^2}\right|,\left\|\frac{\partial\log a(\frac{\epsilon}{\sigma};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|,\right.$$
$$\left.\left\|\frac{\partial^2\log a(\frac{\epsilon}{\sigma};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\sigma}\right\|\right\}\leqslant C\left\{\left(\frac{|\epsilon|}{\sigma}\right)^\tau+1\right\}.$$

Let $\eta_t(\boldsymbol{\theta})=\epsilon_t/\sigma_t(\boldsymbol{\theta})$. For any $u>0$, we introduce the following assumption.

**MM6**$(u)$**:** There exist $p,q>0$ such that $\frac{1}{p}+\frac{2}{q}=1$ and

$$E\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\left(a^{up}(\eta_t(\boldsymbol{\theta});\boldsymbol{\theta})+\left\|\frac{\partial\log a(\eta_t(\boldsymbol{\theta});\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|^q+\left\|\frac{\partial^2\log a(\eta_t(\boldsymbol{\theta});\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\|^{q/2}\right)<\infty.$$

**MM7:** $E\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\left\|\frac{\partial}{\partial\boldsymbol{\theta}}\log a(\epsilon_t/\sigma_t(\boldsymbol{\theta});\boldsymbol{\theta})\right\|<\infty$.

**MM8:** Letting $g_1(y)=1+y\frac{h'}{h}(y)$ and assuming $\iota_h:=E\{g_1^2(\eta_t)\}<\infty$, we have

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{n,ML}-\boldsymbol{\theta}_0)=-\frac{2\boldsymbol{J}^{-1}}{\iota_h\sqrt{n}}\sum_{t=1}^n\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}g_1(\eta_t)+o_P(1).$$

**MM9:** $\sqrt{n}(\widehat{\boldsymbol{\varphi}}_{n,ML}-\boldsymbol{\varphi}_0) = \mathfrak{J}^{-1}\begin{pmatrix} \frac{1}{2\sqrt{n}}\sum_{t=1}^n \frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}} g_1(\eta_t) \\ \frac{-1}{\sqrt{n}}\sum_{t=1}^n \frac{1}{h(\eta_t;\boldsymbol{\nu}_0)}\frac{\partial h(\eta_t;\boldsymbol{\nu}_0)}{\partial\boldsymbol{\nu}} \end{pmatrix} + o_P(1)$, where

$$\mathfrak{J} = \begin{pmatrix} \frac{\iota_h}{4}\boldsymbol{J} & -\frac{1}{2}\boldsymbol{\Omega}E\left(\frac{g_1(\eta_t)}{h(\eta_t;\boldsymbol{\nu}_0)}\frac{\partial h(\eta_t;\boldsymbol{\nu}_0)}{\partial\boldsymbol{\nu}'}\right) \\ -\frac{1}{2}E\left(\frac{g_1(\eta_t)}{h(\eta_t;\boldsymbol{\nu}_0)}\frac{\partial h(\eta_t;\boldsymbol{\nu}_0)}{\partial\boldsymbol{\nu}}\right)\boldsymbol{\Omega}' & E\left(\frac{1}{h^2(\eta_t;\boldsymbol{\nu}_0)}\frac{\partial h(\eta_t;\boldsymbol{\nu}_0)}{\partial\boldsymbol{\nu}}\frac{\partial h(\eta_t;\boldsymbol{\nu}_0)}{\partial\boldsymbol{\nu}'}\right) \end{pmatrix}.$$

Assumption **MM1** ensures the existence of a strictly stationary and ergodic solution $(\epsilon_t)$ to Model (16) (see e.g. Brandt [13]), while **MM2** ensures the existence of a strictly stationary and ergodic solution to the SRE (17) by Lemma 1. Assumption **MM3** controls the effect of the initial values on the studied statistics as the sample size increases. Assumptions **MM4-MM7** are considerably weakened for particular specifications of the MDF, see for instance Corollary 2.

To derive the asymptotic distribution of the $V_n$ and $W_n$ tests, which are based on the ML estimation method, we use the Fisher information for scale $\iota_h$. Assuming that $h$ is everywhere positive, conditions for the existence of $\iota_h$ and its interpretation as the Fisher information for scale are (see e.g. Lehmann and Casella [?], Berkes and Horváth [4], and Francq and Zakoïan [26]):

**FIS:** $h$ has third-order derivatives and satisfies i) $\lim_{|y|\to\infty} yh(y) = \lim_{|y|\to\infty} y^2h'(y) = 0$, and ii) for some positive constants $K$ and $\varsigma$,

$$|y|\left|\frac{h'}{h}(y)\right| + y^2\left|\left(\frac{h'}{h}\right)'(y)\right| + y^2\left|\left(\frac{h'}{h}\right)''(y)\right| \leqslant K\left(1+|y|^{\varsigma}\right),\ E\left|\eta_1\right|^{2\varsigma} < \infty.$$

Assumption **FIS** is used for the purely parametric estimators of the MME in Propositions 4.2 and 5.1 and is satisfied by many distributions (including the Gaussian distribution).

# B Proofs of the main results

Proofs of key results in Sections 2 and 3 are provided below.

## B.1 Proof of Theorem 1

Noting that the sequence $\{(\boldsymbol{V}(\eta_t)'\boldsymbol{\Delta}_{t-1}', \Lambda^u(\eta_t;\boldsymbol{\theta}_0) - E\{\Lambda^u(\eta_t;\boldsymbol{\theta}_0)\}), \mathcal{F}_t\}$ is a second-order stationary martingale difference, the asymptotic distribution in (7) follows from **HL1-HL2** and the CLT of Billingsley[8]. Now assume $\upsilon_u^2 = 0$. Then

$$\Lambda^u(\eta_t;\boldsymbol{\theta}_0) - E\{\Lambda^u(\eta_t;\boldsymbol{\theta}_0)\} + \boldsymbol{g}_u'\boldsymbol{\Delta}_{t-1}'\boldsymbol{V}(\eta_t) = 0, \qquad \text{a.s.}$$

It follows that $\boldsymbol{g}_u \neq \boldsymbol{0}$, otherwise the random variable $\Lambda^u(\eta_t;\boldsymbol{\theta}_0)$ would be degenerate, in contradiction of $\psi_u > 0$. Because, from **HL2**, $\boldsymbol{g}_u'\boldsymbol{\Delta}_{t-1}$ is non-degenerate, and is independent of $\eta_t$, two $d\times k$ matrices exist, $\boldsymbol{\Delta}_1$ and $\boldsymbol{\Delta}_2$, with $\boldsymbol{g}_u'(\boldsymbol{\Delta}_1 - \boldsymbol{\Delta}_2) \neq \boldsymbol{0}$, such that $\Lambda^u(\eta_t;\boldsymbol{\theta}_0) - E\{\Lambda^u(\eta_t;\boldsymbol{\theta}_0)\} + \boldsymbol{g}_u'\boldsymbol{\Delta}_i'\boldsymbol{V}(\eta_t) = 0$, a.s., for $i = 1,2$. It follows that $\boldsymbol{g}_u'(\boldsymbol{\Delta}_1' - \boldsymbol{\Delta}_2')\boldsymbol{V}(\eta_t) = 0$, a.s., which is impossible because $\boldsymbol{\Upsilon}$ is positive definite. Thus we have shown that $\upsilon_u^2 > 0$.

## B.2 Proof of Theorem 2

The a.s. convergence of $\gamma_n$ follows from **HL4** by the ergodic theorem. Similarly,

$$S_n^{(u)} \to S_\infty^{(u)} \text{ a.s.,} \quad \text{for any } u \text{ such that} \quad S_\infty^{(u)} < \infty. \tag{31}$$

Now, we turn to case i). We have $S_\infty^{(u)} < 1$ by Proposition 2.2, thus $S_n^{(u)} < 1$ for $n$ large enough by (31). It follows, from Proposition 3.1, that $\hat{u}_n > u$ for all $u$ and $n$ large enough.

Turning to case ii), the consistency of $\hat{u}_n$ follows from the fact that, for $\varepsilon \in (0, \max\{u_0, s - u_0\})$,

$$\lim_{n\to\infty} \text{a.s. } S_n^{(u_0-\varepsilon)} = S_\infty^{(u_0-\varepsilon)} < 1, \quad \lim_{n\to\infty} \text{a.s. } S_n^{(u_0+\varepsilon)} = S_\infty^{(u_0+\varepsilon)} > 1.$$

### B.3 Proof of Theorem 3

Under **HL5**, it suffices to show that

$$\Gamma_n^0 \overset{\mathcal{C}[u_1,u_2]}{\Longrightarrow} \Gamma. \tag{32}$$

By the Cramér-Wold device, and the CLT of Billingsley [8] used in the proof of Theorem 1, it can be established that the finite-dimensional distributions of $\Gamma_n^0$ converge to those of $\Gamma$. By showing that

$$\text{the sequence} \quad \{\Gamma_n^0(u_1)\} \quad \text{is tight} \tag{33}$$

and, for some constant $K > 0$,

$$E\left\{\Gamma_n^0(u) - \Gamma_n^0(v)\right\}^2 \leqslant K(u-v)^2, \tag{34}$$

the tightness of the sequence $\{\Gamma_n^0\}$ will be established, according to Theorem 12.3 of Billingsley [7]. The weak convergence in (32) follows from Theorem 8.1 of Billingsley [7].

The convergence in distribution of $\{\Gamma_n^0(u_1)\}$ entails (33). We have

$$\begin{aligned}\Gamma_n^0(u) - \Gamma_n^0(v) =& \frac{1}{\sqrt{n}}\sum_{t=1}^n [\Lambda^u(\eta_t;\boldsymbol{\theta}_0) - E\{\Lambda^u(\eta_t;\boldsymbol{\theta}_0)\} - \Lambda^v(\eta_t;\boldsymbol{\theta}_0)\\ &+ E\{\Lambda^v(\eta_t;\boldsymbol{\theta}_0)\}] + (\boldsymbol{g}_u - \boldsymbol{g}_v)'\frac{1}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{\Delta}_{t-1}\boldsymbol{V}(\eta_t) := \Delta_{n,1}(u,v) + \Delta_{n,2}(u,v).\end{aligned}$$

Note that

$$E\Delta_{n,1}^2(u,v) = \text{Var}\{\Lambda^u(\eta_t) - \Lambda^v(\eta_t)\} \leqslant (u-v)^2 E\left(\{\Lambda^{2u_1}(\eta_t) + \Lambda^{2u_2}(\eta_t)\}\{\log\Lambda(\eta_t)\}^2\right) \leqslant K(u-v)^2.$$

Moreover $\text{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{\Delta}_{t-1}\boldsymbol{V}(\eta_t)\right) = \boldsymbol{\Sigma}$. It follows that, using **HL6**, $E\Delta_{n,2}^2(u,v) \leqslant K(u-v)^2$. This completes the proof of weak convergence in (32).

### B.4 Proof of Theorem 4

Writing $0 = S_n^{(\hat{u}_n)} - S_\infty^{(u_0)} = S_n^{(\hat{u}_n)} - S_\infty^{(\hat{u}_n)} + S_\infty^{(\hat{u}_n)} - S_\infty^{(u_0)}$, we deduce, by the mean-value theorem,

$$\sqrt{n}(\hat{u}_n - u_0) = -\frac{1}{D_\infty^{(u_n^*)}}\sqrt{n}\left(S_n^{(\hat{u}_n)} - S_\infty^{(\hat{u}_n)}\right) = -\frac{1}{D_\infty^{(u_n^*)}}\Gamma_n(\hat{u}_n)$$

where $u_n^*$ is between $\hat{u}_n$ and $u_0$. By continuity of $D_\infty^{(u)}$ we have $D_\infty^{(u_n^*)} \to D_\infty^{(u_0)}$ in probability (and also a.s.), and $\Gamma_n(\hat{u}_n) \xrightarrow{d} \Gamma(u_0)$. Indeed, for $\varsigma, \varsigma' > 0$,

$$P\{\Gamma_n(\hat{u}_n) > x\} \leqslant P\{\Gamma_n(\hat{u}_n) > x, |\hat{u}_n - u_0| \leqslant \varsigma\} + P(|\hat{u}_n - u_0| > \varsigma)$$

$$\leqslant P\left\{\sup_{|u-u_0|\leqslant\varsigma}\Gamma_n(u) > x\right\} + P\left(|\widehat{u}_n - u_0| > \varsigma\right)$$

$$\leqslant P\left\{\Gamma_n(u_0) > x - \varsigma'\right\} + P\left\{\sup_{|u-u_0|\leqslant\varsigma}|\Gamma_n(u) - \Gamma_n(u_0)| > \varsigma'\right\} + P\left(|\widehat{u}_n - u_0| > \varsigma\right).$$

Using the tightness property of the sequence $\Gamma_n\{\cdot\}$ (a consequence of (32) and **HL5**) and the a.s. convergence of $\widehat{u}_n$, the last two probabilities can be made arbitrarily small for $n$ sufficiently large and $\varsigma$ small enough. The other probability converges to $P\{\Gamma(u_0) > x - \varsigma'\}$ which is arbitrarily close to $P\{\Gamma(u_0) > x\}$ for $\varsigma'$ small enough. A similar upper bound can be obtained for $P\{\Gamma_n(\widehat{u}_n) < x\}$ from which the conclusion follows.

# C Additional proofs for the general model

## C.1 Proof of Lemma 1

This result can be seen as a consequence of Theorem 2.8 in Straumann and Mikosch [47]. We nevertheless give a direct proof.

For ease of presentation we omit $\boldsymbol{\theta}$ and set $\Gamma_t = \Lambda_1(y_t)$. For all $n \in \mathbb{N}$, $n > 0$, and $t \in \mathbb{Z}$, let

$$f_{t,n} = \varphi\left[g^*(f_{t-1,n-1}, y_{t-1}), f_{t-1,n-1}\right] = \psi(y_{t-1}, f_{t-1,n-1}),$$

where $f_{t-n,0} = f^0$. For fixed $n$, the sequence $(f_{t,n})_t$ is stationary and ergodic. We have

$$|f_{t,n} - f_{t,n-1}| \leqslant \Gamma_{t-1}|f_{t-1,n-1} - f_{t-1,n-2}| \leqslant \Gamma_{t-1}\Gamma_{t-2}\ldots\Gamma_{t-n+1}|\psi(y_{t-n}, f^0) - f^0|.$$

Thus, for $n < m$,

$$\begin{aligned}|f_{t,m} - f_{t,n}| &\leqslant \sum_{k=0}^{m-n-1}|f_{t,m-k} - f_{t,m-k-1}|\\ &\leqslant \sum_{k=0}^{m-n-1}\Gamma_{t-1}\ldots\Gamma_{t-m+k+1}|\psi(y_{t-m+k}, f^0) - f^0|\\ &\leqslant \sum_{j=n+1}^{\infty}\Gamma_{t-1}\ldots\Gamma_{t-j+1}|\psi(y_{t-j}, f^0) - f^0| \to 0 \quad \text{a.s. as} \quad n\to\infty. \end{aligned} \tag{35}$$

The latter convergence follows from the Cauchy rule applied to the infinite sum, using $E\log\Gamma_t < 0$ and $E\log^+|\psi(y_t, f^0) - f^0| < \infty$. We have shown that, a.s., $(f_{t,n})_{n\in\mathbb{N}}$ is a Cauchy sequence on the complete space $F$. Therefore $f_t(\boldsymbol{\theta}) = \lim_{n\to\infty} f_{t,n}$ provides the stationary solution of the SRE (6).

Now, note that

$$\sup_{\boldsymbol{\theta}\in\Theta}|f_t(\boldsymbol{\theta}) - \widetilde{f}_t(\boldsymbol{\theta})| \leqslant \overline{\Gamma}_{t-1}\overline{\Gamma}_{t-2}\cdots\overline{\Gamma}_0\sup_{\boldsymbol{\theta}\in\Theta}|f_0(\boldsymbol{\theta}) - \widetilde{f}_0|,$$

where $\overline{\Gamma}_t = \sup_{\boldsymbol{\theta}\in\Theta}\Gamma_t$ with $\Gamma_t = \Lambda_1(y_t;\boldsymbol{\theta})$. By (ii) of **A3** one can choose $\rho$ such that

$$1 > \rho > e^{E\log\overline{\Gamma}_1} > 0,$$

so that

$$\lim_{t\to\infty}\frac{1}{t}\ln\rho^{-t}\overline{\Gamma}_{t-1}\overline{\Gamma}_{t-2}\cdots\overline{\Gamma}_0 = -\ln\rho + E\ln\overline{\Gamma}_1 < 0.$$

Now, by letting $m \to \infty$ and taking $n = 0$ in (35), we have

$$|f_t(\boldsymbol{\theta}) - f^0| \leqslant \sum_{j=1}^{\infty} \Gamma_{t-1} \dots \Gamma_{t-j+1} |\psi(y_{t-j}, f^0) - f^0|.$$

Thus

$$\sup_{\boldsymbol{\theta} \in \Theta} |f_t(\boldsymbol{\theta}) - f^0| \leqslant \sum_{j=1}^{\infty} \overline{\Gamma}_{t-1} \dots \overline{\Gamma}_{t-j+1} \sup_{\boldsymbol{\theta} \in \Theta} |\psi(y_{t-j}, f^0) - f^0|.$$

By the arguments given in the fist part of the proof, $\sup_{\boldsymbol{\theta} \in \Theta} |f_0(\boldsymbol{\theta})|$ is almost surely finite under **A3** (i), and the conclusion follows.

## C.2 Proof of Proposition 2.2

Part i) is obvious: the condition $P\{\Lambda(\eta_1) \leqslant 1\} = 1$ entails $E\{\Lambda^u(\eta_1)\} \leqslant 1$ and the inequality is strict because $\gamma < 0$.
Now suppose $P\{\Lambda(\eta_1) \leqslant 1\} < 1$ and let $\epsilon > 0$ such that $P\{\Lambda(\eta_1) > 1 + \epsilon\} > 0$. Then we have $S_\infty^{(u)} = E\{\Lambda^u(\eta_1)\} > (1+\epsilon)^u P\{\Lambda(\eta_1) > 1 + \epsilon\} \to \infty$ as $u \to \infty$. For any $\eta > 0$, the function $u \mapsto \Lambda^u(\eta)$ is convex. Thus $u \mapsto E\{\Lambda^u(\eta_1)\}$ is convex on $(0, s]$. We consider two cases: a) when $P\{\Lambda(\eta_1) = 0\} = p > 0$ we have $S_\infty^{(0^+)} = 1 - p < S_\infty^{(0)} = 1$. In view of the convexity and the fact $S_\infty^{(s)} \geqslant 1$, the conclusion follows; b) when $P\{\Lambda(\eta_1) = 0\} = 0$, the right derivative of $u \mapsto S_\infty^{(u)}$ in the neighborhood of 0 is negative. Thus a value $0 < s_0 < s$ exists for which the function $u \mapsto E\{\Lambda^u(\eta_1)\}$ decreases over $(0, s_0)$ and increases over $(s_0, s]$. Since $S_\infty^{(s)} \geqslant 1$, it follows that there is a unique $u > 0$ such that $E\{\Lambda^u(\eta_1)\} = 1$. This completes the proof of Proposition 2.2.

## C.3 Proof of Proposition 2.3

Noting that $\widehat{\psi}_u = \widetilde{S}_n^{(2u)}(\widehat{\boldsymbol{\theta}}_n) - \{\widetilde{S}_n^{(u)}(\widehat{\boldsymbol{\theta}}_n)\}^2$, the strong consistency of $\widehat{\psi}_u$ follows from **HL3** and the strong consistency of $\widehat{\boldsymbol{\theta}}_n$. It follows that $\widehat{v}_u$ is a consistent estimator of $v_u$. Now, noting that $S_\infty^{(u)} \leqslant S_\infty^{(u_0)}$ for $u \leqslant u_0$, we have

$$\begin{aligned}
&P_{\boldsymbol{H}_{0,u}}(C_T^{(u)}) = P_{\boldsymbol{H}_{0,u}} \left\{ \widehat{v}_u^{-1} \sqrt{n} \left( S_n^{(u)} - 1 \right) > \Phi^{-1}(1 - \underline{\alpha}) \right\} \\
=&P_{\boldsymbol{H}_{0,u}} \left\{ \widehat{v}_u^{-1} \sqrt{n} \left( S_n^{(u)} - S_\infty^{(u)} \right) + \widehat{v}_u^{-1} \sqrt{n} \left( S_\infty^{(u)} - S_\infty^{(u_0)} \right) > \Phi^{-1}(1 - \underline{\alpha}) \right\} \\
\leqslant&P_{\boldsymbol{H}_{0,u}} \left\{ \widehat{v}_u^{-1} \sqrt{n} \left( S_n^{(u)} - S_\infty^{(u)} \right) > \Phi^{-1}(1 - \underline{\alpha}) \right\}
\end{aligned}$$

which tends to $\underline{\alpha}$ as $n \to \infty$ by Theorem 1. The conclusion under $\boldsymbol{H}_{0,u}$ follows.

Under $\boldsymbol{H}_{1,u}$, $T_n^{(u)} \sim \frac{\sqrt{n}(S_\infty^{(u)} - 1)}{v_u} \to \infty$, in probability as $n \to \infty$. The conclusion follows.

## C.4 Proof of Proposition 3.1

We apply Proposition 2.2, substituting the empirical distribution of $\{\Lambda(\widehat{\eta}_t; \widehat{\boldsymbol{\theta}}_n) : t = 1, \dots, n\}$ for the theoretical distribution of $\Lambda(\eta_1)$. The condition on the existence of $s > 0$ vanishes because moments exist at any order for the empirical distribution.

## C.5 Proof of Proposition 3.2

Under $\boldsymbol{H}_{0,u}$, the arguments are the same as those in the proof of Proposition 2.3, using **HL7** to show the consistency of $\widehat{w}_u^2$, and the asymptotic normality of $\sqrt{n}(\widehat{u}_n - u_0)$ established in Theorem 4. Under $\boldsymbol{H}_{1,u}$ we use the fact that $u - \widehat{u}_n \to u - u_0 > 0$ in probability as $n \to \infty$.

## C.6 Proof of Proposition 3.3

By the delta method we have,

$$\sqrt{n}\,(\hat{u}_{n,h} - u_0) = \frac{\partial u_0}{\partial \boldsymbol{\theta}'}\sqrt{n}(\hat{\boldsymbol{\theta}}_{n,ML} - \boldsymbol{\theta}_0) + o_P(1) \xrightarrow{d} \mathcal{N}\left(0, \sigma_h^2\right). \tag{36}$$

In view of the consistency of $\hat{\sigma}_h$, the conclusion under $\boldsymbol{H}_{0,u}$ follows. Under $\boldsymbol{H}_{1,u}$ we use the fact that $u - \hat{u}_{n,h} \to u - u_0 > 0$ in probability as $n \to \infty$.

## C.7 Proof of Proposition 3.4

The proof is similar to that of Proposition 3.3, based on the Taylor expansion

$$\frac{\sqrt{n}(u_0 - \hat{u}_{0,\hat{h}})}{\hat{\varsigma}_h} = \frac{-1}{\varsigma_h}\left(\frac{\partial u_0}{\partial \boldsymbol{\theta}'}\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + \frac{\partial u_0}{\partial \boldsymbol{\nu}'}\sqrt{n}(\hat{\boldsymbol{\nu}}_n - \boldsymbol{\nu}_0)\right) + o_P(1)$$

# D Examples of augmented GARCH models

The table below displays examples of models satisfying (16) (with $x^+ = \max(x, 0)$, $x^- = \max(-x, 0)$). GARCH models were introduced by Engle [24] and Bollerslev [11]. Taylor model was introduced by Taylor [48]. Threshold GARCH (TGARCH) models were introduced by Zakoïan [51]. GJR-GARCH models were introduced by Glosten et al. [33]. Asymmetric Power ARCH (APARCH) models were introduced by Ding et al. [19]. Beta-$t$-GARCH models were introduced by Harvey [36] and Creal et al. [15].

Table 2

| Model | $\boldsymbol{\theta}, \delta$ | $a(\eta_t, \boldsymbol{\theta})$ |
|---|---|---|
| GARCH[1] | $(\omega, \alpha, \beta), 2$ | $\alpha\eta^2 + \beta$ |
| Taylor model[2] | $(\omega, \alpha, \beta), 1$ | $\alpha\lvert\eta\rvert + \beta$ |
| TGARCH[3] | $(\omega, \alpha_+, \alpha_-, \beta), 1$ | $\alpha_+\eta^+ + \alpha_-\eta^- + \beta$ |
| GJR-GARCH[4] | $(\omega, \alpha_+, \alpha_-, \beta), 2$ | $\alpha_+\eta^{+2} + \alpha_-\eta^{-2} + \beta$ |
| APARCH[5] | $(\omega, \alpha, \xi, \beta), \delta$ | $\omega + \alpha(\lvert\eta\rvert - \xi\eta)^\delta + \beta$ |
| Beta-$t$-GARCH[6] | $(\omega, \alpha, \beta, \nu), 2$ | $\beta + \frac{\alpha(\nu+1)\eta^2}{(\nu-2)+\eta^2}$ |

[1] $\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$
[2] $\sigma_t = \omega + \alpha\lvert\epsilon_{t-1}\rvert + \beta\sigma_{t-1}$
[3] $\sigma_t = \omega + \alpha_+\epsilon_{t-1}^+ + \alpha_-\epsilon_{t-1}^- + \beta\sigma_{t-1}$
[4] $\sigma_t^2 = \omega + \alpha_+\epsilon_{t-1}^{+2} + \alpha_-\epsilon_{t-1}^{-2} + \beta\sigma_{t-1}^2$
[5] $\sigma_t^\delta = \omega + \alpha(\lvert\epsilon_{t-1}\rvert - \xi\epsilon_{t-1})^\delta + \beta\sigma_{t-1}^\delta$
[6] $\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\frac{(\nu+1)\epsilon_{t-1}^2}{(\nu-2)+\epsilon_{t-1}^2/\sigma_{t-1}^2}$

# E Empirical MDF of a GARCH process

The first-order GARCH process is a particular case of APARCH, obtained for $\delta = 2$ and $a(\eta) = \alpha\eta^2 + \beta$. The asymptotic variance of the empirical MDF has a more explicit form in GARCH models for two important estimation methods:

*Corollary* 7 (**GARCH estimated by ML or QML**). If i) the distribution of $\eta_t^2$ is non-degenerate and $E(|\eta_t|^4) < \infty$; ii) $\Theta \subset [\underline{\omega}, \infty) \times (0, \infty) \times [0, 1)$ is compact and $\boldsymbol{\theta}_0 = (\omega_0, \alpha_0, \beta_0)' \in \overset{\circ}{\Theta}$, iii) $E\log(\alpha_0\eta_t^2 + \beta_0) < 0$, then the conclusions of Corollary 1 hold for the QML and ML estimators with $u \leqslant 2$. Moreover, letting $M_{x,y} = E\left\{\eta_t^{2x}(\alpha_0\eta_t^2 + \beta_0)^y\right\}$, $x, y \in \mathbb{R}$, and

$$\boldsymbol{\Omega} = E\left(\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right), \qquad \boldsymbol{J} = E\left(\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\right), \tag{37}$$

we find that $\boldsymbol{g}_u = u\left(\boldsymbol{m}_u - \alpha_0 M_{1,u-1}\boldsymbol{\Omega}\right)$, where $\boldsymbol{m}_u = (0, M_{1,u-1}, M_{0,u-1})'$, and

$$\upsilon_u^2 = c_\eta u^2\left(\boldsymbol{m}_u'\boldsymbol{J}^{-1}\boldsymbol{m}_u - \alpha_0^2 M_{1,u-1}^2\right) + M_{0,2u} - M_{0,u}^2, \tag{38}$$

where $c_\eta = \kappa_4 - 1$ with $\kappa_4 = E\eta_t^4$ for the QMLE, and $c_\eta = 4/\iota_h$ for the MLE.

**Proof.** i) When the model is estimated by QML we have

$$\boldsymbol{V}(\eta_t) = \eta_t^2 - 1, \quad \boldsymbol{\Delta}_{t-1} = \boldsymbol{J}^{-1}\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}},$$

thus $\boldsymbol{\Sigma} = (\kappa_4 - 1)\boldsymbol{J}^{-1}$. It follows that

$$\begin{aligned}\upsilon_u^2 &= u^2(\kappa_4-1)\left(\boldsymbol{m}_u'\boldsymbol{J}^{-1}\boldsymbol{m}_u + \alpha_0^2 M_{1,u-1}^2 - 2\alpha_0 M_{1,u-1}\boldsymbol{m}_u'\boldsymbol{J}^{-1}\boldsymbol{\Omega}\right)\\ &\quad + M_{0,2u} - M_{0,u}^2 + 2u(M_{1,u} - M_{0,u})\left(\boldsymbol{m}_u'\boldsymbol{J}^{-1}\boldsymbol{\Omega} - \alpha_0 M_{1,u-1}\right).\end{aligned}$$

Noting that $\boldsymbol{J}^{-1}\boldsymbol{\Omega} = (\omega_0, \alpha_0, 0)'$ (see Francq and Zakoïan [29]), we obtain $\boldsymbol{g}_u'\boldsymbol{\xi}_u = 0$ and the formula for $\upsilon_u^2$ follows.

ii) If the model is estimated by ML we have

$$\boldsymbol{\Sigma} = \frac{4}{\iota_h}\boldsymbol{J}^{-1}, \quad \boldsymbol{V}(\eta_t) = g_1(\eta_t), \quad \boldsymbol{\Delta}_{t-1} = -\frac{2}{\iota_h}\boldsymbol{J}^{-1}\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}.$$

Noting that

$$\begin{aligned}Ea(\eta_t;\boldsymbol{\theta})g_1(\eta_t) &= \alpha + \beta + \int(\alpha x^2 + \beta)xh'(x)dx\\ &= \alpha + \beta - \int(3\alpha x^2 + \beta)h(x)dx = -2\alpha,\end{aligned}$$

we have, using $\boldsymbol{\Omega}'\boldsymbol{J}^{-1}\boldsymbol{\Omega} = 1$ (see Remark 3 in Francq and Zakoïan [29]) and $\boldsymbol{J}^{-1}\boldsymbol{\Omega} = (\omega_0, \alpha_0, 0)'$,

$$\boldsymbol{g}_u'\boldsymbol{\xi}_u = u\left\{\boldsymbol{m}_u - \alpha_0 M_{1,u-1}\boldsymbol{\Omega}\right\}'\frac{2}{\iota_h}\boldsymbol{J}^{-1}\boldsymbol{\Omega}Ea^u(\eta_t)g_1(\eta_t) = \boldsymbol{0}$$

and

$$\begin{aligned}\boldsymbol{g}_u'\boldsymbol{\Sigma}\boldsymbol{g}_u &= \frac{4u^2}{\iota_h}\left(\boldsymbol{m}_u - \alpha_0 M_{1,u-1}\boldsymbol{\Omega}\right)'\boldsymbol{J}^{-1}\left(\boldsymbol{m}_u - \alpha_0 M_{1,u-1}\boldsymbol{\Omega}\right)\\ &= \frac{4u^2}{\iota_h}\left(\boldsymbol{m}_u'\boldsymbol{J}^{-1}\boldsymbol{m}_u - \alpha_0^2 M_{1,u-1}^2\right).\end{aligned}$$

Thus

$$\upsilon_u^2 = \boldsymbol{g}_u'\boldsymbol{\Sigma}\boldsymbol{g}_u + \psi_u = \frac{4u^2}{\iota_h}\left(\boldsymbol{m}_u'\boldsymbol{J}^{-1}\boldsymbol{m}_u - \alpha_0^2 M_{1,u-1}^2\right) + M_{0,2u} - M_{0,u}^2.$$

The MLE is more efficient than the QMLE since $\kappa_4 - 1 \geqslant 4/\iota_h$ and, by the Cauchy-Schwarz inequality,

$$\begin{aligned}\boldsymbol{m}_u'\boldsymbol{J}^{-1}\boldsymbol{m}_u - \alpha_0^2 M_{1,u-1}^2 &= \boldsymbol{m}_u'\boldsymbol{J}^{-1}\boldsymbol{m}_u - (\boldsymbol{m}_u'\boldsymbol{J}^{-1}\boldsymbol{\Omega})^2\\ &\geqslant \boldsymbol{m}_u'\boldsymbol{J}^{-1}\boldsymbol{m}_u - (\boldsymbol{m}_u'\boldsymbol{J}^{-1}\boldsymbol{m}_u)(\boldsymbol{\Omega}'\boldsymbol{J}^{-1}\boldsymbol{\Omega}) = 0.\end{aligned}$$

□

In Francq and Zakoïan [31] we provided a test of finite moments of order $u$ in the case where $u$ is even and $(\epsilon_t)$ is a standard GARCH process. In this case, the moment condition is an explicit function of $\boldsymbol{\theta}_0$ and moments of $\eta_t$. The test statistic is thus computed differently, but is equivalent to the test $T_n^{(u)}$ of Corollary 3, as the next example illustrates.

*Example* (**2nd-order stationarity testing** ($u = 1$)). Consider a standard GARCH model ($\delta = 2$). We have $a(\eta, \boldsymbol{\theta}) = \alpha\eta^2 + \beta$. When the model is estimated by Gaussian QML we have, by Corollary 7, $v_1^2 = (\kappa_4 - 1)\boldsymbol{e}_0'\boldsymbol{J}^{-1}\boldsymbol{e}_0 + (\alpha_0 + \beta_0)^2 - 1$, where $\boldsymbol{e}_0' = (0, 1, 1)$. Thus under $\mathbf{H}_{0,1}$,

$$S_n^{(1)} \quad = \quad \frac{1}{n}\sum_{t=1}^{n}(\widehat{\alpha}_n\widehat{\eta}_t^2 + \widehat{\beta}_n) = \widehat{\alpha}_n + \widehat{\beta}_n + o_P(1), \quad v_1^2 = (\kappa_4 - 1)\boldsymbol{e}_0'\boldsymbol{J}^{-1}\boldsymbol{e}_0.$$

We retrieve the Wald-type test statistic for testing second-order stationarity,

$$T_n^{(1)} = \sqrt{n}\frac{(\widehat{\alpha}_n + \widehat{\beta}_n - 1)}{\{(\widehat{\kappa}_4 - 1)\boldsymbol{e}_0'\widehat{\boldsymbol{J}}^{-1}\boldsymbol{e}_0\}^{1/2}} + o_P(1).$$

# F Proofs for the multiplicative model

## F.1 Proof of Corollary 1

First note that $E\{\Lambda^s(\eta_t; \boldsymbol{\theta}_0)\} = Ea^s(\eta_t; \boldsymbol{\theta}_0) < \infty$ under **MM1**. The properties required in **A0-A3** are also satisfied by **MM1-MM2**. It therefore remains to demonstrate **HL1** and apply Theorem 1.

With $a(\frac{\epsilon}{\sigma}; \boldsymbol{\theta}) = b(\epsilon, \sigma; \boldsymbol{\theta})$ where $b : \mathbb{R} \times \mathbb{R}^+ \times \Theta \mapsto \mathbb{R}^+$, under **MM4**, for $b$ or $\log b$, $\nabla_\sigma$ (resp. $\nabla_{\boldsymbol{\theta}}$) denotes the partial derivative with respect to $\sigma$ (resp. $\boldsymbol{\theta}$), and $\nabla^2_{\sigma\sigma}$ (resp. $\nabla^2_{\sigma\boldsymbol{\theta}}$) denotes the unmixed (resp. mixed) second-order partial derivative with respect to $\sigma$ (resp. $\sigma$ and $\boldsymbol{\theta}$).[7] With this notation, we can write for instance

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\theta}}a\{\eta_t(\boldsymbol{\theta}); \boldsymbol{\theta}\} =& \frac{\partial}{\partial\boldsymbol{\theta}}b\{\epsilon_t, \sigma_t(\boldsymbol{\theta}); \boldsymbol{\theta}\}\\ =& \nabla_{\boldsymbol{\theta}}b\{\epsilon_t, \sigma_t(\boldsymbol{\theta}); \boldsymbol{\theta}\} + \nabla_\sigma b\{\epsilon_t, \sigma_t(\boldsymbol{\theta}); \boldsymbol{\theta}\}\frac{\partial}{\partial\boldsymbol{\theta}}\sigma_t(\boldsymbol{\theta}).\end{aligned}$$

We establish the following intermediate results:

**IR1:** There exists a neighborhood $V(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$ such that

$$\inf_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)} a\{\eta_t(\boldsymbol{\theta}); \boldsymbol{\theta}\} > 0 \quad \text{and} \quad \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)} n\left|\widetilde{S}_n^{(u)}(\boldsymbol{\theta}) - S_n^{(u)}(\boldsymbol{\theta})\right| = O(1) \quad \text{a.s.}$$

**IR2:** The function $\boldsymbol{\theta} \mapsto a\{\eta_t(\boldsymbol{\theta}); \boldsymbol{\theta}\}$ is continuously differentiable at $\boldsymbol{\theta}_0$. Moreover, for any sequence $(\boldsymbol{\theta}_n)$ such that $\boldsymbol{\theta}_n \to \boldsymbol{\theta}_0$ a.s., we have

$$\frac{\partial S_n^{(u)}(\boldsymbol{\theta}_n)}{\partial\boldsymbol{\theta}} - \frac{\partial S_n^{(u)}(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} \to 0, \quad \text{a.s. as } n \to \infty.$$

**IR3:** The expectation $\boldsymbol{g}_u = E\left[\frac{\partial}{\partial\boldsymbol{\theta}}a^u\{\eta_t(\boldsymbol{\theta}); \boldsymbol{\theta}\}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ exists in $\mathbb{R}^d$ and $\frac{\partial S_n^{(u)}(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} \to \boldsymbol{g}_u$, a.s. as $n \to \infty$.

[7] For instance in the standard GARCH(1,1) model with $\boldsymbol{\theta} = (\omega, \alpha, \beta)'$, we have $b(\epsilon, \sigma; \boldsymbol{\theta}) = \alpha\left(\frac{\epsilon}{\sigma}\right)^2 + \beta$, $\nabla_\sigma \log b(\epsilon, \sigma; \boldsymbol{\theta}) = \frac{-2}{\alpha\left(\frac{\epsilon}{\sigma}\right)^2 + \beta}\frac{\alpha}{\sigma}\left(\frac{\epsilon}{\sigma}\right)^2$ and $\nabla_{\boldsymbol{\theta}} \log b(\epsilon, \sigma; \boldsymbol{\theta}) = \frac{1}{\alpha\left(\frac{\epsilon}{\sigma}\right)^2 + \beta}\left(0, \left(\frac{\epsilon}{\sigma}\right)^2, 1\right)'$. In the ARCH(1) model, with $\boldsymbol{\theta} = (\omega, \alpha)'$, we have $b(\epsilon, \sigma; \boldsymbol{\theta}) = \alpha\left(\frac{\epsilon}{\sigma}\right)^2$, thus $\nabla_\sigma \log b(\epsilon, \sigma; \boldsymbol{\theta}) = \frac{-2}{\sigma}$ and $\nabla_{\boldsymbol{\theta}} \log b(\epsilon, \sigma; \boldsymbol{\theta}) = \left(0, \frac{1}{\alpha}\right)'$.

#### F.1.1 Proof of IR1

We have $\widetilde{\eta}_t(\boldsymbol{\theta}) = \epsilon_t \widetilde{\sigma}_t^{-1}(\boldsymbol{\theta})$ where $\widetilde{\sigma}_t^{\delta}(\boldsymbol{\theta})$ is defined after (17). For $\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)$ we have

$$\begin{aligned} a^u\{\widetilde{\eta}_t(\boldsymbol{\theta});\boldsymbol{\theta}\} - a^u\{\eta_t(\boldsymbol{\theta});\boldsymbol{\theta}\} &= b^u\{\epsilon_t, \widetilde{\sigma}_t(\boldsymbol{\theta});\boldsymbol{\theta}\} - b^u\{\epsilon_t, \sigma_t(\boldsymbol{\theta});\boldsymbol{\theta}\} \\ &= ub^u\{\epsilon_t, \sigma_t^*;\boldsymbol{\theta}\}\nabla_\sigma \log b(\epsilon_t, \sigma_t^*;\boldsymbol{\theta})\{\widetilde{\sigma}_t(\boldsymbol{\theta}) - \sigma_t(\boldsymbol{\theta})\} \end{aligned} \quad (39)$$

where $\sigma_t^*$ is between $\widetilde{\sigma}_t(\boldsymbol{\theta})$ and $\sigma_t(\boldsymbol{\theta})$.

Then, using **MM5** and the $c_r$ inequality, we deduce

$$|a^u\{\widetilde{\eta}_t(\boldsymbol{\theta});\boldsymbol{\theta}\} - a^u\{\eta_t(\boldsymbol{\theta});\boldsymbol{\theta}\}| \leqslant u2^u C^{u+1}\left\{\left(\frac{|\epsilon_t|}{\sigma_t^*}\right)^{\tau(u+1)} + 1\right\}|\widetilde{\sigma}_t(\boldsymbol{\theta}) - \sigma_t(\boldsymbol{\theta})|.$$

The r.h.s. of the inequality is bounded by a variable of the form $KX_t\rho^t$ where $X_t$ admits a small moment uniformly in $t$, $\rho \in (0,1)$ and $K$ is $\mathcal{F}_0$-measurable, using Lemma 1 and **MM3** and noting that

$$\frac{|\epsilon_t|}{\sigma_t^*} \leqslant |\eta_t|\frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta})}\frac{\sigma_t(\boldsymbol{\theta})}{\sigma_t^*} \leqslant |\eta_t|\left(1 + \frac{K\rho^t}{\underline{\omega}}\right)\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta})}.$$

Thus

$$n\left|\widetilde{S}_n^{(u)}(\boldsymbol{\theta}) - S_n^{(u)}(\boldsymbol{\theta})\right| \leqslant K\sum_{t=1}^n X_t\rho^t \leqslant K\sum_{t=1}^\infty X_t\rho^t,$$

where the latter sum admits a small moment by **MM4** and thus is finite a.s.

#### F.1.2 Proof of IR2

In view of **IR1**

$$\frac{\partial S_n^{(u)}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \frac{1}{n}\sum_{t=1}^n ua^{u-1}\{\eta_t(\boldsymbol{\theta});\boldsymbol{\theta}\}\frac{\partial a\{\eta_t(\boldsymbol{\theta});\boldsymbol{\theta}\}}{\partial\boldsymbol{\theta}}$$

is well-defined in $V(\boldsymbol{\theta}_0)$. We also have

$$\begin{aligned} \frac{\partial^2 S_n^{(u)}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} =& \frac{1}{n}\sum_{t=1}^n ub^u(\epsilon_t, \sigma_t(\boldsymbol{\theta});\boldsymbol{\theta})\left\{u\frac{\partial}{\partial\boldsymbol{\theta}}\log b(\epsilon_t,\sigma_t(\boldsymbol{\theta});\boldsymbol{\theta})\frac{\partial}{\partial\boldsymbol{\theta}'}\log b(\epsilon_t,\sigma_t(\boldsymbol{\theta});\boldsymbol{\theta})\right. \\ &\left. + \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\log b(\epsilon_t,\sigma_t(\boldsymbol{\theta});\boldsymbol{\theta})\right\}. \end{aligned}$$

From the Hölder inequality and **MM6**$(u)$

$$\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\left\|\frac{\partial^2 S_n^{(u)}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\| = O(1), \quad \text{a.s.}$$

The conclusions follows from a Taylor expansion of $\frac{\partial S_n^{(u)}(\boldsymbol{\theta}_n)}{\partial\boldsymbol{\theta}}$ around $\boldsymbol{\theta}_0$.

#### F.1.3 Proof of IR3

Noting that

$$\frac{\partial S_n^{(u)}(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} = \frac{1}{n}\sum_{t=1}^n ub^u(\epsilon_t,\sigma_t(\boldsymbol{\theta}_0);\boldsymbol{\theta}_0)\frac{\partial}{\partial\boldsymbol{\theta}}\log b(\epsilon_t,\sigma_t(\boldsymbol{\theta}_0);\boldsymbol{\theta}_0)$$

the result is a straightforward consequence of the Hölder inequality, **MM6**$(u)$ and the ergodic theorem.

Now, noting that $S_n^{(u)} = \widetilde{S}_n^{(u)}(\widehat{\boldsymbol{\theta}}_n)$, a Taylor expansion of $S_n^{(u)}(\widehat{\boldsymbol{\theta}}_n)$ around $\boldsymbol{\theta}_0$ yields

$$\sqrt{n}\left(S_n^{(u)} - S_\infty^{(u)}\right) = \sqrt{n}\left(\widetilde{S}_n^{(u)}(\widehat{\boldsymbol{\theta}}_n) - S_n^{(u)}(\widehat{\boldsymbol{\theta}}_n)\right) + \left\{\frac{\partial S_n^{(u)}(\boldsymbol{\theta}_n^*)}{\partial \boldsymbol{\theta}'} - \frac{\partial S_n^{(u)}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right\}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$$
$$+ \frac{\partial S_n^{(u)}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + \sqrt{n}\left\{S_n^{(u)}(\boldsymbol{\theta}_0) - S_\infty^{(u)}\right\},$$

where $\boldsymbol{\theta}_n^*$ is between $\widehat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$. In view of **IR1-IR3** and **HL2**, and since $\frac{\partial S_n^{(u)}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \to \boldsymbol{g}_u$ a.s. by the ergodic theorem, we can conclude that **HL1** holds true. Thus the conclusions of Theorem 1 apply.

### F.2 Proof of Corollary 2

Part iii) shows that **MM1** holds true. Noting that

$$\frac{\partial}{\partial z}\left\{\omega\left(\frac{\epsilon_t}{z^{1/\delta}};\boldsymbol{\theta}\right) + a\left(\frac{\epsilon_t}{z^{1/\delta}};\boldsymbol{\theta}\right)z\right\} = \beta$$

and that the strictly stationary solution admits a small-order moment, it can be check that **MM2** is satisfied. The CAN of the QMLE were established by Hamadeh and Zakoïan [35]. By Theorems 2.1 and 2.2 in Hamadeh and Zakoïan [35], assumption **HL2** holds with

$$\boldsymbol{V}(\eta_t) = \eta_t^2 - 1, \quad \boldsymbol{\Delta}_{t-1} = \frac{\delta}{2}\boldsymbol{J}_\delta^{-1}\frac{1}{\sigma_t^\delta}\frac{\partial \sigma_t^\delta(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}, \quad \boldsymbol{\Delta} = \frac{\delta}{2}\boldsymbol{J}_\delta^{-1}\boldsymbol{\Omega}_\delta,$$

where $\boldsymbol{\Omega}_\delta = E(\boldsymbol{D}_t)$, $\boldsymbol{J}_\delta = E(\boldsymbol{D}_t\boldsymbol{D}_t')$, $\boldsymbol{D}_t = \boldsymbol{D}_t(\boldsymbol{\theta}_0)$ and $\boldsymbol{D}_t(\boldsymbol{\theta}) = \sigma_t^{-\delta}(\boldsymbol{\theta})\partial\sigma_t^\delta(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. As an intermediate result to establish the CAN of the QMLE, [35] established **MM3**. By Equation (5.18) in Hamadeh and Zakoïan [35] the first moment condition in **MM4** is satisfied with $r = 2$. The second moment condition is also satisfied, using $\sup_{\boldsymbol{\theta}\in\Theta}|\beta| < 1$ and the existence of a small-order moment for $|\epsilon_t|$. Noting that $b(\epsilon, \sigma, \boldsymbol{\theta}) = \frac{|\epsilon|^\delta}{\sigma^\delta}(\alpha_+ \mathbb{1}_{\epsilon>0} + \alpha_- \mathbb{1}_{\epsilon<0}) + \beta$, it can be shown that **MM5** is satisfied for $\tau = \delta$. It has also been shown that

$$E\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\left\|\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})}\frac{\partial\sigma_t^\delta(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|^d < \infty, \quad E\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\left\|\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})}\frac{\partial^2\sigma_t^\delta(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\|^d < \infty$$

for any integer $d$ (by (5.20) in the aforementioned paper). It follows that **MM5**$(u)$ is satisfied for any $q > 0$ and for $p$ close enough to 1 when $u \leqslant s/2$. We conclude from Corollary 1.

### F.3 Proof of Corollary 3

The proof consists in checking the consistency of

$$\widehat{\boldsymbol{g}}_u = \frac{\partial \widetilde{S}_n^{(u)}(\widehat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} = \frac{1}{n}\sum_{t=1}^n \frac{\partial}{\partial \boldsymbol{\theta}} a^u\left\{\frac{\epsilon_t}{\widetilde{\sigma}_t}(\widehat{\boldsymbol{\theta}}_n); \widehat{\boldsymbol{\theta}}_n\right\}$$

and Assumption **HL3**. We start by showing that

$$\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\left\|\frac{\partial \widetilde{S}_n^{(u)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial S_n^{(u)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\| \to 0, \quad \text{in probability as } n \to \infty. \tag{40}$$

We have

$$\frac{\partial S_n^{(u)}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}-\frac{\partial \widetilde{S}_n^{(u)}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}=\frac{1}{n}\sum_{t=1}^{n}u\left[b^u\left\{\epsilon_t,\sigma_t(\boldsymbol{\theta});\boldsymbol{\theta}\right\}-b^u\left\{\epsilon_t,\widetilde{\sigma}_t(\boldsymbol{\theta});\boldsymbol{\theta}\right\}\right]$$

$$\times\left[\nabla_\sigma\log b\left\{\epsilon_t,\sigma_t(\boldsymbol{\theta});\boldsymbol{\theta}\right\}\frac{\partial\sigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}+\nabla_{\boldsymbol{\theta}}\log b\left\{\epsilon_t,\sigma_t(\boldsymbol{\theta});\boldsymbol{\theta}\right\}\right]$$

$$+\frac{1}{n}\sum_{t=1}^{n}ub^u\left\{\epsilon_t,\widetilde{\sigma}_t(\boldsymbol{\theta});\boldsymbol{\theta}\right\}\left[\nabla_\sigma\log b\left\{\epsilon_t,\sigma_t(\boldsymbol{\theta});\boldsymbol{\theta}\right\}-\nabla_\sigma\log b\left\{\epsilon_t,\widetilde{\sigma}_t(\boldsymbol{\theta});\boldsymbol{\theta}\right\}\right]\frac{\partial\sigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$$

$$+\frac{1}{n}\sum_{t=1}^{n}ub^u\left\{\epsilon_t,\widetilde{\sigma}_t(\boldsymbol{\theta});\boldsymbol{\theta}\right\}\nabla_\sigma\log b\left\{\epsilon_t,\widetilde{\sigma}_t(\boldsymbol{\theta});\boldsymbol{\theta}\right\}\left(\frac{\partial\sigma_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}-\frac{\partial\widetilde{\sigma}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)$$

$$+\frac{1}{n}\sum_{t=1}^{n}ub^u\left\{\epsilon_t,\widetilde{\sigma}_t(\boldsymbol{\theta});\boldsymbol{\theta}\right\}\left[\nabla_{\boldsymbol{\theta}}\log b\left\{\epsilon_t,\sigma_t(\boldsymbol{\theta});\boldsymbol{\theta}\right\}-\nabla_{\boldsymbol{\theta}}\log b\left\{\epsilon_t,\widetilde{\sigma}_t(\boldsymbol{\theta});\boldsymbol{\theta}\right\}\right]$$

$$:=\Delta_{1n}(\boldsymbol{\theta})+\Delta_{2n}(\boldsymbol{\theta})+\Delta_{3n}(\boldsymbol{\theta})+\Delta_{4n}(\boldsymbol{\theta}).$$

First consider $\Delta_{1n}(\boldsymbol{\theta})$. From the proof of Corollary 1 we have

$$\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\left|b^u\left\{\epsilon_t,\sigma_t(\boldsymbol{\theta});\boldsymbol{\theta}\right\}-b^u\left\{\epsilon_t,\widetilde{\sigma}_t(\boldsymbol{\theta});\boldsymbol{\theta}\right\}\right|\leqslant X_t\rho^t$$

where $X_t$ admits a small moment. By **MM3** and **MM4**, the other summands involved in $\Delta_1$ also admit small moments. It follows that $\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}|\Delta_{1n}(\boldsymbol{\theta})|\to 0$, in probability as $n\to\infty$.

Now we turn to $\Delta_{2n}$. Another Taylor expansion yields

$$\nabla_\sigma\log b\left\{\epsilon_t,\sigma_t(\boldsymbol{\theta});\boldsymbol{\theta}\right\}-\nabla_\sigma\log b\left\{\epsilon_t,\widetilde{\sigma}_t(\boldsymbol{\theta});\boldsymbol{\theta}\right\}=\nabla^2_{\sigma\sigma}\log b(\epsilon_t,\sigma_t^*;\boldsymbol{\theta})\{\widetilde{\sigma}_t(\boldsymbol{\theta})-\sigma_t(\boldsymbol{\theta})\},$$

where $\sigma_t^*$ is between $\widetilde{\sigma}_t(\boldsymbol{\theta})$ and $\sigma_t(\boldsymbol{\theta})$. By the same arguments, $\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}|\Delta_{2n}(\boldsymbol{\theta})|\to 0$, in probability as $n\to\infty$. The last two terms can be handled similarly. Hence, (40) is established. Now using **IR2-IR3** in the proof of Corollary 1, together with the consistency of $\widehat{\boldsymbol{\theta}}_n$, we conclude that $\widehat{\boldsymbol{g}}_u$ is a consistent estimator of $\boldsymbol{g}_u$.

We similarly show that Assumption **HL3** is satisfied, which completes the proof of Corollary 3.

## F.4 Proof of Proposition 4.1

The tail result for $\sigma_t$ is established using Theorem 4.1 in Goldie [34]. The tail result for $\epsilon_t$ follows by the arguments given by Mikosch and Stărică[44] in proving their Theorem 2.1.

## F.5 Proof of Corollary 4

It suffices to show that **HL4** holds true. Similar to (39) we have

$$\log b\left\{\epsilon_t,\sigma_t(\boldsymbol{\theta});\boldsymbol{\theta}\right\}-\log b\left\{\epsilon_t,\widetilde{\sigma}_t(\boldsymbol{\theta});\boldsymbol{\theta}\right\}=\nabla_\sigma\log b(\epsilon_t,\sigma_t^*;\boldsymbol{\theta})\left\{\widetilde{\sigma}_t(\boldsymbol{\theta})-\sigma_t(\boldsymbol{\theta})\right\},\tag{41}$$

thus, by arguments already used, $\gamma_n=\frac{1}{n}\sum_{t=1}^{n}\log b\left\{\epsilon_t,\sigma_t(\widehat{\boldsymbol{\theta}}_n);\widehat{\boldsymbol{\theta}}_n\right\}+o(1)$, a.s.

Moreover,

$$\log b\left\{\epsilon_t,\sigma_t(\boldsymbol{\theta});\boldsymbol{\theta}\right\}-\log b\left\{\epsilon_t,\sigma_t(\boldsymbol{\theta}_0);\boldsymbol{\theta}_0\right\}=\frac{\partial}{\partial\boldsymbol{\theta}'}\log b\left\{\epsilon_t,\sigma_t(\boldsymbol{\theta}^*);\boldsymbol{\theta}^*\right\}(\boldsymbol{\theta}-\boldsymbol{\theta}_0),\tag{42}$$

for $\boldsymbol{\theta}^*$ between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$. Using the consistency of $\widehat{\boldsymbol{\theta}}_n$ and **MM7**, we conclude that

$$\gamma_n=\frac{1}{n}\sum_{t=1}^{n}\log b\left\{\epsilon_t,\sigma_t(\boldsymbol{\theta}_0);\boldsymbol{\theta}_0\right\}+o(1),\ \text{a.s.}$$

The second convergence in **HL4** can be handled similarly.

### F.6 Proof of Corollary 5

In view of Theorem 3, we need to show **HL5** and **HL6**. First note that

$$\boldsymbol{g}_u - \boldsymbol{g}_v = E\left( \frac{\partial^2}{\partial u \partial \theta_i} b^u\{\epsilon_t, \sigma_t(\boldsymbol{\theta}); \boldsymbol{\theta}\}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0, u=u_i^*} \right)(u-v)$$

where the $u_i^*$ belong to $(u, v)$ and the existence of the expectation follows from **MM6**$(u_2)$. Thus **HL6** is established.

We will now show **HL5**. We have, for $\boldsymbol{\theta}^*$ between $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$,

$$\Gamma_n(u) - \Gamma_n^0(u) = \sqrt{n}\left\{S_n^{(u)} - S_n^{(u)}(\hat{\boldsymbol{\theta}}_n)\right\} + \frac{1}{n}\sum_{t=1}^n \left[\frac{\partial}{\partial \boldsymbol{\theta}} a^u\{\eta_t(\boldsymbol{\theta}^*); \boldsymbol{\theta}^*\} - \boldsymbol{g}_u\right]' \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$$
$$+ \boldsymbol{g}_u'\left\{\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) - \frac{1}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{\Delta}_{t-1}\boldsymbol{V}(\eta_t)\right\} := R_{n,1}(u) + R_{n,2}(u) + R_{n,3}(u). \quad (43)$$

The proof is thus divided into three steps.

i) We have, by **MM3**, with $\sigma_t^*(\boldsymbol{\theta})$ between $\tilde{\sigma}_t(\boldsymbol{\theta})$ and $\sigma_t(\boldsymbol{\theta})$

$$|R_{n,1}(u)| \leqslant \frac{1}{\sqrt{n}}\sum_{t=1}^n \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)} u b^u\{\epsilon_t, \sigma_t^*(\boldsymbol{\theta}); \boldsymbol{\theta}\} \left|\nabla_\sigma \log b\{\epsilon_t, \sigma_t^*(\boldsymbol{\theta}); \boldsymbol{\theta}\}\right| K_t \rho^t = o_P(1),$$

uniformly in $u \in [u_1, u_2]$, noting that, by **MM5**, the supremum admits a small-order moment.

ii) The second term, $R_{n,2}(u)$, can be handled by a Taylor expansion around $\boldsymbol{\theta}_0$ of $\frac{\partial}{\partial \boldsymbol{\theta}} a^u\{\eta_t(\boldsymbol{\theta}^*); \boldsymbol{\theta}^*\}$. Indeed, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^n \frac{\partial}{\partial \boldsymbol{\theta}} a^u\{\eta_t(\boldsymbol{\theta}^*); \boldsymbol{\theta}^*\} - \boldsymbol{g}_u \\
=&\frac{1}{n}\sum_{t=1}^n \frac{\partial}{\partial \boldsymbol{\theta}} a^u\{\eta_t(\boldsymbol{\theta}_0); \boldsymbol{\theta}_0\} - \boldsymbol{g}_u + \frac{1}{n}\sum_{t=1}^n \frac{\partial^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} a^u\{\eta_t(\boldsymbol{\theta}^{**}); \boldsymbol{\theta}^{**}\}(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0) \\
=&R_{n,4}(u) + R_{n,5}(u),
\end{aligned}$$

where $\boldsymbol{\theta}^{**}$ is between $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}_0$.

For any $u^* \in (u_1, u_2)$ and any positive integer $k$, let $V_k(u^*) = (u^* - \frac{1}{k}, u^* + \frac{1}{k})$. We have, for any $k$,

$$\begin{aligned}
&\sup_{u \in V_k(u^*)\cap[u_1,u_2]} \left|\frac{1}{n}\sum_{t=1}^n \frac{\partial}{\partial \boldsymbol{\theta}} a^u\{\eta_t(\boldsymbol{\theta}_0); \boldsymbol{\theta}_0\} - \boldsymbol{g}_u\right| \\
\leqslant\ & \frac{1}{n}\sum_{t=1}^n X_{t,k}(u^*) + \left|\frac{1}{n}\sum_{t=1}^n \frac{\partial}{\partial \boldsymbol{\theta}} a^{u^*}\{\eta_t(\boldsymbol{\theta}_0); \boldsymbol{\theta}_0\} - \boldsymbol{g}_{u^*}\right| + |\boldsymbol{g}_{u^*} - \boldsymbol{g}_u|, \quad (44)
\end{aligned}$$

where

$$X_{t,k}(u^*) := \sup_{u \in V_k(u^*)\cap[u_1,u_2]} \left|\frac{\partial}{\partial \boldsymbol{\theta}} a^u\{\eta_t(\boldsymbol{\theta}_0); \boldsymbol{\theta}_0\} - \frac{\partial}{\partial \boldsymbol{\theta}} a^{u^*}\{\eta_t(\boldsymbol{\theta}_0); \boldsymbol{\theta}_0\}\right|.$$

The last term tends to 0 as $k$ increases to infinity by continuity of $u \mapsto \boldsymbol{g}_y$. The second term converges to 0 a.s. by the ergodic theorem. Finally, the first sum in the r.h.s. of Equation (44) converges a.s. as $n \to \infty$ to $EX_{t,k}(u^*)$. Indeed, the ergodic theorem can be applied because the variables inside the absolute values are both continuous in $u$ and functions of the $\eta_{t-i}$'s for $i \geqslant 0$

(see Francq and Zakoian [30], Exercise 7.4). By the Beppo Levi theorem, $EX_{t,k}(u^*)$ decreases to 0 as $k \to \infty$. We have shown that the left-hand side of (44) converges to 0 as $k$ and $n \to \infty$. This conclusion is based on a compactness argument: for any cover of the compact set $[u_1, u_2]$ by sets of the form $V_k(u^*)$, there exists a finite subcover, of the form $V_k(u_1^*), \dots, V_k(u_d^*)$. We have

$$\begin{aligned}
&\sup_{u\in[u_1,u_2]} \left| \frac{1}{n}\sum_{t=1}^{n} \frac{\partial}{\partial \boldsymbol{\theta}} a^u\{\eta_t(\boldsymbol{\theta}_0);\boldsymbol{\theta}_0\} - \boldsymbol{g}_u \right| \\
&= \max_{i=1,\dots d} \sup_{u\in V_k(u_i^*)\cap[u_1,u_2]} \left| \frac{1}{n}\sum_{t=1}^{n} \frac{\partial}{\partial \boldsymbol{\theta}} a^u\{\eta_t(\boldsymbol{\theta}_0);\boldsymbol{\theta}_0\} - \boldsymbol{g}_u \right| = o_P(1).
\end{aligned}$$

We have shown that $\sup_{u\in[u_1,u_2]} R_{n,4}(u) = o_P(1)$. Now, since

$$\begin{aligned}
&\frac{\partial^2 a^u\{\eta_t(\boldsymbol{\theta});\boldsymbol{\theta}\}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \\
&= \left( \frac{\partial^2 \log a\{\eta_t(\boldsymbol{\theta});\boldsymbol{\theta}\}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} + \frac{\partial \log a\{\eta_t(\boldsymbol{\theta});\boldsymbol{\theta}\}}{\partial\boldsymbol{\theta}} \frac{\partial \log a\{\eta_t(\boldsymbol{\theta});\boldsymbol{\theta}\}}{\partial\boldsymbol{\theta}'} \right) a^u\{\eta_t(\boldsymbol{\theta});\boldsymbol{\theta}\},
\end{aligned}$$

we deduce from **MM6(s)** and the strong consistency of $\boldsymbol{\theta}^*$ to $\boldsymbol{\theta}_0$ that $\sup_{u\in[u_1,u_2]} R_{n,5}(u) = o_P(1)$. Thus $\sup_{u\in[u_1,u_2]} R_{n,2}(u) = o_P(1)$.

iii) The third term, $R_{n,3}(u)$, in the r.h.s. of (43) is an $o_P(1)$ uniformly in $u$ by **MM5** and using the fact that $\sup_{u\in(u_1,u_2)} \|\boldsymbol{g}_u\| < \infty$. This latter property follows from

$$\|\boldsymbol{g}_u\| \leqslant u \left\| E\left( \frac{\partial \log a\{\eta_t(\boldsymbol{\theta}_0);\boldsymbol{\theta}_0\}}{\partial\boldsymbol{\theta}} a^u\{\eta_t(\boldsymbol{\theta}_0);\boldsymbol{\theta}_0\} \right) \right\|,$$

whereas by Hölder's inequality, for each component $\boldsymbol{\theta}_i$ of $\boldsymbol{\theta}$,

$$\begin{aligned}
&\sup_{u\in(u_1,u_2)} E\left| \left( \frac{\partial \log a\{\eta_t(\boldsymbol{\theta}_0);\boldsymbol{\theta}_0\}}{\partial\boldsymbol{\theta}_i} a^u\{\eta_t(\boldsymbol{\theta}_0);\boldsymbol{\theta}_0\} \right) \right| \\
&\leqslant \left| \frac{\partial \log a\{\eta_t(\boldsymbol{\theta}_0);\boldsymbol{\theta}_0\}}{\partial\boldsymbol{\theta}_i} \right|_{q/2} \sup_{u\in(u_1,u_2)} |a^u\{\eta_t(\boldsymbol{\theta}_0);\boldsymbol{\theta}_0\}|_p \\
&\leqslant \left| \frac{\partial \log a\{\eta_t(\boldsymbol{\theta}_0);\boldsymbol{\theta}_0\}}{\partial\boldsymbol{\theta}_i} \right|_{q/2} \left( |a^{u_1}\{\eta_t(\boldsymbol{\theta}_0);\boldsymbol{\theta}_0\}|_p + |a^{u_2}\{\eta_t(\boldsymbol{\theta}_0);\boldsymbol{\theta}_0\}|_p \right) < \infty,
\end{aligned}$$

using **MM6(s)**. Thus **HL5** is established.

## F.7 Proof of Corollary 6

It can be shown that **HL7** holds by the arguments used to establish **HL3** in the proof of Corollary 3. The conclusion follows from Proposition 3.2.

## F.8 Proof of Proposition 4.2

In view of **MM8**, (36) and the consistency of $\widehat{\sigma}_h$, we deduce

$$V_n^{(u_0)} = \frac{\sqrt{n}(u_0 - \widehat{u}_{n,h})}{\widehat{\sigma}_h} = \frac{-2}{\sigma_h \iota_h \sqrt{n}} \sum_{t=1}^{n} \frac{\partial u_0}{\partial \boldsymbol{\theta}'} \boldsymbol{J}^{-1} \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}} g_1(\eta_t) + o_P(1), \tag{45}$$

from which the conclusion follows.

### F.9 Proof of Proposition 4.3

The strong consistency of $\widehat{\boldsymbol{\vartheta}}_n^{QML}$ follows from Theorem 3.1 in Hamadeh and Zakoïan [35]. Because $\mathcal{D}$ is discrete, it follows that $\widehat{\delta}_n^{QML} = \delta_0$ for sufficiently large $n$. By Corollary 2, the assumptions required for Corollary 5 are satisfied for $n$ large enough when $\delta$ is replaced by $\widehat{\delta}_n^{QML}$. If $\eta_t$ has a positive density over the real line, the condition $1 < E\{a^s(\eta_1)\} < \infty$ for $s > 0$ of Corollary 5 holds and the conclusion follows.

## G Proofs for the asymptotic power comparisons

### G.1 Proof of Proposition 5.1 and inequality (24)

In the proof of Corollary 1, we have seen that

$$T_n^{(u)} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{a^u(\eta_t) - 1}{\upsilon_u} + \boldsymbol{g}_u' \frac{1}{\upsilon_u \sqrt{n}} \sum_{t=1}^{n} \boldsymbol{\Delta}_{t-1} \boldsymbol{V}(\eta_t) + o_P(1), \tag{46}$$

where the first term is centered only for $u = u_0$. By (22), it follows that under $P_0$

$$\begin{pmatrix} T_n^{(u_0)} \\ \Lambda_{n,h}(\boldsymbol{\theta}_0 + \boldsymbol{\tau}/\sqrt{n}, \boldsymbol{\theta}_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left\{ \begin{pmatrix} 0 \\ -\frac{1}{2}\boldsymbol{\tau}'\boldsymbol{\mathfrak{I}}\boldsymbol{\tau} \end{pmatrix}, \begin{pmatrix} 1 & c_{h,u_0}(\boldsymbol{\theta}_0) \\ c_{h,u_0}(\boldsymbol{\theta}_0) & \boldsymbol{\tau}'\boldsymbol{\mathfrak{I}}\boldsymbol{\tau} \end{pmatrix} \right\}.$$

Le Cam's third lemma (see *e.g.* van der Vaart [50], page 90) shows that

$$T_n^{(u_0)} \xrightarrow{d} \mathcal{N}\left(c_{h,u_0}(\boldsymbol{\theta}_0), 1\right), \qquad \text{under } P_{n,\boldsymbol{\tau}}.$$

The conclusion of Proposition 5.1 easily follows for the two tests using (20).

With the notations used in the proof of Corollary 7, for the standard GARCH(1,1) model estimated by QML we have

$$E\left(\frac{1}{\sigma_t}\frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \boldsymbol{g}_{u_0}' \boldsymbol{\Delta}_{t-1}\right) = E\left(\frac{1}{\sigma_t}\frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \boldsymbol{\Delta}_{t-1}' \boldsymbol{g}_{u_0}\right) = \frac{1}{2}\boldsymbol{g}_{u_0}, \quad E\{\boldsymbol{V}(\eta_1) g_1(\eta_1)\} = -2,$$

while with the ML we have

$$E\left(\frac{1}{\sigma_t}\frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \boldsymbol{g}_{u_0}' \boldsymbol{\Delta}_{t-1}\right) = E\left(\frac{1}{\sigma_t}\frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \boldsymbol{\Delta}_{t-1}' \boldsymbol{g}_{u_0}\right) = \frac{-1}{\iota_h}\boldsymbol{g}_{u_0}, \quad E\{\boldsymbol{V}(\eta_1) g_1(\eta_1)\} = \iota_h.$$

Moreover,

$$\begin{aligned} &\frac{1}{2} E a_t^{u_0} \left\{1 + \eta_t \frac{h'}{h}(\eta_t)\right\} + \alpha_0 u_0 E \eta_t^2 a_t^{u_0 - 1} = \frac{1}{2} + \frac{1}{2}\int a^{u_0}(x) x h'(x) dx + \alpha_0 u_0 \int x^2 a^{u_0-1} = (x) h(x) d \\ =& \frac{1}{2} + \frac{1}{2}\int a^{u_0}(x) x h'(x) dx + [a^{u_0}(x)\frac{x}{2} h(x)] - \int a^{u_0}(x) \left(\frac{h(x)}{2} + \frac{x}{2} h'(x)\right) dx 0 = 0. \end{aligned}$$

Thus, for the standard GARCH(1,1) model,

$$c_{h,u_0}(\boldsymbol{\theta}_0) = -\frac{\boldsymbol{\tau}'}{\upsilon_{u_0}}\left[\frac{1}{2}\boldsymbol{\Omega} E\{a^{u_0}(\eta_1) g_1(\eta_1)\} - u_0\left(\boldsymbol{m}_{u_0} - \alpha_0 M_{1,u_0-1}\boldsymbol{\Omega}\right)\right] = \frac{u_0}{\upsilon_{u_0}}\boldsymbol{\tau}'\boldsymbol{m}_{u_0},$$

where the formulas for $\upsilon_{u_0}$ are displayed in (38) for the ML and QML estimators.

### G.2 Proof of Proposition 5.2

Relation (22) implies that

$$\Lambda_{n,h}(\boldsymbol{\theta}_0 + \boldsymbol{\tau}/\sqrt{n}, \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}\left(-\frac{1}{2}\boldsymbol{\tau}'\mathfrak{I}\boldsymbol{\tau}, \boldsymbol{\tau}'\mathfrak{I}\boldsymbol{\tau}\right) \qquad \text{under } P_0,$$

which is the distribution of the log-likelihood ratio in the statistical model $\mathcal{N}\left(\boldsymbol{\tau}, \mathfrak{I}^{-1}\right)$ of parameter $\boldsymbol{\tau}$. In other words, denoting by $\mathcal{T}$ a subset of $\mathbb{R}^d$ containing a neighborhood of $\mathbf{0}$, the so-called local experiments $\{L_{n,h}(\boldsymbol{\theta}_0 + \boldsymbol{\tau}/\sqrt{n}), \boldsymbol{\tau} \in \mathcal{T}\}$ converge to the Gaussian experiment $\left\{\mathcal{N}\left(\boldsymbol{\tau}, \mathfrak{I}^{-1}\right), \boldsymbol{\tau} \in \mathcal{T}\right\}$.

Under the assumption of the proposition on $u_0(\boldsymbol{\theta}_0, h)$, for a given $u$, testing $\boldsymbol{H}_{0,u}$ against $\boldsymbol{H}_{1,n,u}$, amounts to testing $\boldsymbol{H}_0: \quad \boldsymbol{\tau} = \mathbf{0}$ against $\boldsymbol{H}_1: \quad \boldsymbol{\tau} = \boldsymbol{e}$ in the limiting experiment. The UMPU test based on $\boldsymbol{X} \sim \mathcal{N}\left(\boldsymbol{\tau}, \mathfrak{I}^{-1}\right)$ is the test of rejection region

$$C = \left\{\boldsymbol{e}'\boldsymbol{X}/\sqrt{\boldsymbol{e}'\mathfrak{I}^{-1}\boldsymbol{e}} > \Phi^{-1}(1-\underline{\alpha})\right\}.$$

This UMPU test has the power given in (25).

In the case of the standard GARCH(1,1) model,

$$\begin{aligned} c_{h,1}^{QML}(\boldsymbol{\theta}_0) =& \frac{\boldsymbol{e}'\boldsymbol{e}}{\sqrt{(\kappa_4 - 1)\boldsymbol{e}'\boldsymbol{J}^{-1}\boldsymbol{e}}} \\ \leqslant& c_{h,1}^{ML}(\boldsymbol{\theta}_0) = \frac{\boldsymbol{e}'\boldsymbol{e}}{\sqrt{\frac{4}{\iota_h}\boldsymbol{e}'\boldsymbol{J}^{-1}\boldsymbol{e} + \alpha_0^2\left(\kappa_4 - 1 - \frac{4}{\iota_h}\right)}} \leqslant c_{\boldsymbol{e}} = \frac{\iota_h^{1/2}\boldsymbol{e}'\boldsymbol{e}}{2\sqrt{\boldsymbol{e}'\boldsymbol{J}^{-1}\boldsymbol{e}}}, \end{aligned}$$

by the Cauchy-Schwarz inequality, with equality only when $g_1(y) = K(1-y^2)$, that is if and only if the density of $\eta_t$ has the form (26) (see [28], Proposition 5.5).

### G.3 Proof of Proposition 5.3

By the arguments of the proof of Proposition 5.1, using (45), we obtain

$$d_{h,u_0}(\boldsymbol{\theta}_0) = -\frac{1}{\sigma_h}\frac{\partial u}{\partial \boldsymbol{\theta}'}\boldsymbol{\tau} = \frac{-\frac{\partial u}{\partial \boldsymbol{\theta}'}\boldsymbol{\tau}}{\sqrt{\frac{4}{\iota_h}\frac{\partial u}{\partial \boldsymbol{\theta}'}\boldsymbol{J}^{-1}\frac{\partial u}{\partial \boldsymbol{\theta}}}} = \frac{\boldsymbol{r}_{u_0}'\boldsymbol{\tau}}{\sqrt{\frac{4}{\iota_h}\boldsymbol{r}_{u_0}'\boldsymbol{J}^{-1}\boldsymbol{r}_{u_0}}}.$$

### G.4 Proof of Proposition 5.4

Follows by the arguments of the proof of Proposition 5.1, using (37) and the LAN property (29)-(30).

### G.5 Proof of Proposition 5.5

The statistics $T_n^{(u)}$, $U_n^{(u)}$ and $W_n^{(u)}$ are $\mathcal{N}(0,1)$ distributed under the null. The $p$-values of the tests based on $T_n^{(u)}$ and $U_n^{(u)}$ are thus $1 - \Phi\left(T_n^{(u)}\right)$ and $1 - \Phi\left(U_n^{(u)}\right)$ respectively. Under the alternative $\boldsymbol{H}_{1,u}: \quad u > u_0$ we have, almost surely, as $n \to \infty$,

$$T_n^{(u)} = \frac{\sqrt{n}\left(S_n^{(u)} - 1\right)}{\widehat{v}_u} \sim \frac{\sqrt{n}\left(S_\infty^{(u)} - 1\right)}{v_u}, \quad U_n^{(u)} = \frac{\sqrt{n}\,(u - \hat{u}_n)}{\widehat{w}_{\hat{u}_n}} \sim \frac{\sqrt{n}\,(u - u_0)}{w_{u_0}},$$

$$V_n^{(u)} = \frac{\sqrt{n}(u - \hat{u}_{n,\hat{h}})}{\hat{\sigma}_h} \sim \frac{\sqrt{n}(u - u_0)}{\sigma_h}, \quad W_n^{(u)} = \frac{\sqrt{n}(u - \hat{u}_{0,\hat{h}})}{\hat{\varsigma}_h} \sim \frac{\sqrt{n}(u - u_0)}{\varsigma_h}.$$

It can be shown that $\log\{1 - \Phi(x)\} \sim -x^2/2$ as $x \to +\infty$. The asymptotic slopes of the tests are thus

$$c_T(u) = \frac{\left(S_\infty^{(u)} - 1\right)^2}{\upsilon_u^2}, \quad c_U(u) = \frac{(u - u_0)^2}{w_{u_0}^2}, \quad c_V(u) = \frac{(u - u_0)^2}{\sigma_h^2}, \quad c_W(u) = \frac{(u - u_0)^2}{\varsigma_h^2}.$$

The test $T_n^{(u)}$ is more efficient than $U_n^{(u)}$, in Bahadur's sense, if and only if

$$\frac{c_T(u)}{c_U(u)} = \frac{\left(S_\infty^{(u)} - 1\right)^2}{(u - u_0)^2} \frac{\upsilon_{u_0}^2}{\left(E[a^{u_0}(\eta_1; \boldsymbol{\theta}_0) \log\{a(\eta_1; \boldsymbol{\theta}_0)\}]\right)^2 \upsilon_u^2} > 1,$$

and the test $W_n^{(u)}$ is more efficient than $U_n^{(u)}$ if and only if

$$\frac{c_W(u)}{c_U(u)} = \upsilon_{u_0}^2 \left(\boldsymbol{r}'_{u_0}, \boldsymbol{s}'_{u_0}\right) \mathfrak{J}^{-1} \left(\boldsymbol{r}'_{u_0}, \boldsymbol{s}'_{u_0}\right)' > 1.$$

# H Examples of asymptotic power comparisons

Propositions 5.1,5.3 and 5.4 (with $\boldsymbol{\tau}_2 = 0$) are illustrated in Figure 3 for Student distributions with $\nu = 5, 20, 30$ and $\infty$. For the GARCH(1,1) model, the LAPs of the tests $T, U, V$ and $W$ depend on $\boldsymbol{\tau}$ through $\boldsymbol{m}'_{u_0}\boldsymbol{\tau}$, which is therefore shown on the horizontal axis. As expected, the test $V$ is locally asymptotically more efficient than the other tests, especially when $u_0$ is small for the equivalent tests $T$ and $U$. The latter two tests are also dominated by the $W$ test.

Examples of asymptotic slopes for the standard GARCH(1,1) model with Gaussian and Student errors are displayed in Figure 4 and 5. It is clear from these graphs that, for the alternative $\boldsymbol{H}_{1,u}: \quad u > u_0$, test $U$ based on the MME is more efficient than is test $T$ based on the GMF, and that the ratio $c_U(u)/c_T(u)$ increases as $u$ departs from $u_0$. On the contrary, for the alternative $\boldsymbol{H}^*_{1,u}: \quad u < u_0$, the asymptotic slopes favor test $T$. Test $V$ is always more powerful than $U$, but may be outperformed by $T$ in the left-hand side of $u_0$. Interestingly, the left panel shows that the slope of test $T$ may decrease for large values of $u$, which can be explained by the fact that the numerator and denominator of this ratio both tend to infinity as $u$ increases. On the other hand, for small values of $u$ the moment condition $u < s/2$ required for the validity of test $T$ can be satisfied while the condition $u_0 < s/2$, required for the validity of test $U$, can be violated.

Monte-Carlo experiments displayed in the next section illustrate test $T$'s lack of power relative to the others, in agreement with Figures 4-5.

# I Monte Carlo experiments

We first performed 10,000 simulations of a standard GARCH(1,1) model with $(\alpha_0, \beta_0) = (0.10, 0.86)$ and Gaussian innovations such that $u_0 = 4$, for different sample sizes. The results are reported in Table 3. Concerning the tests, the most striking result is the low power of test $T$ relative to the others, in agreement with Figures 4-5. Even for large sample sizes, test $T$ is too conservative but the levels of tests $U$ and $V$ at the boundary of the null are correct. As expected, test $V$ is slightly more powerful than test $U$. The CIs based on the statistics $\hat{u}_n$ and $\hat{u}_{n,h}$ (lines $U_n^{(u)}$ and $V_n^{(u)}$) are similar and, as expected, slightly tighter with the fully parametric

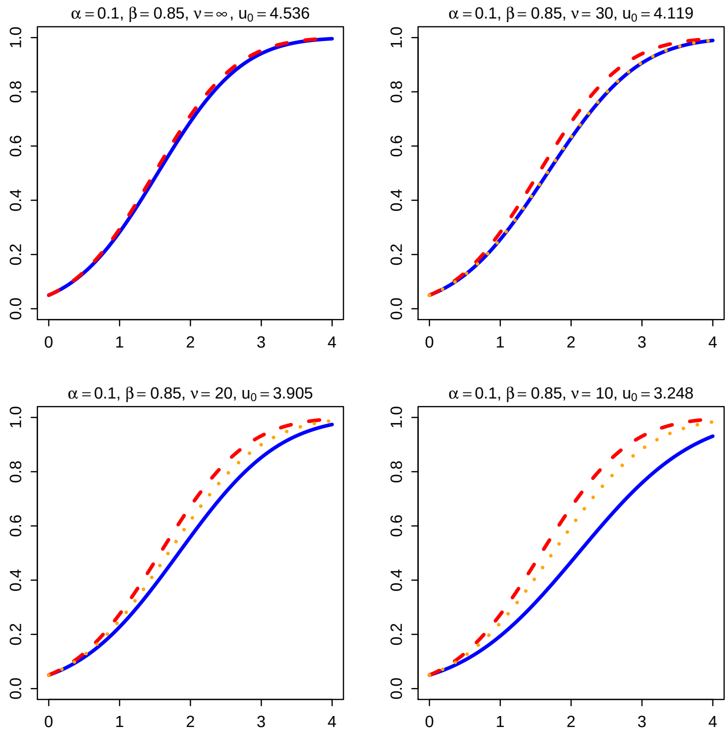


Figure 3: LAPs of tests $T$ and $U$ (blue line) based on the Gaussian QML, test $V$ (dotted red line), and test $W$ (dotted orange line) as functions of $\boldsymbol{m}'_{u_0}\boldsymbol{\tau}$, for a standard GARCH(1,1) model with $\alpha_0 = 0.10, \beta_0 = 0.85$ and for Student errors with $\nu$ degrees of freedom.

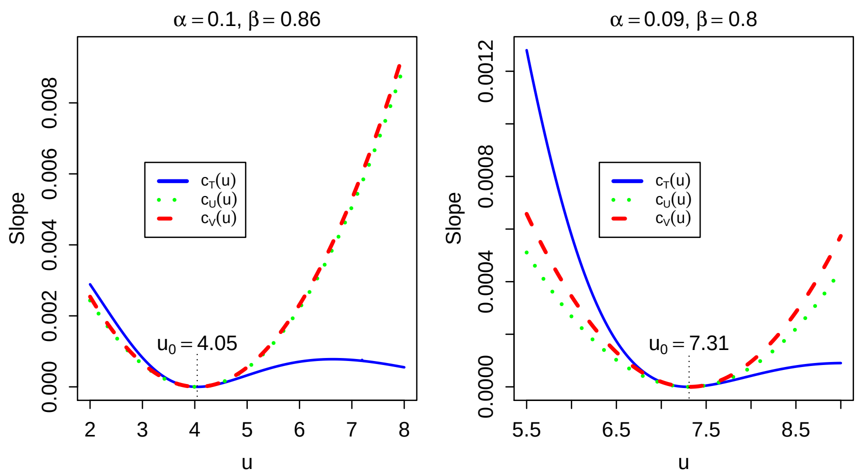


Figure 4: Asymptotic slopes of the tests $T, U$ and $V$ for Gaussian errors and the standard GARCH(1,1) models.

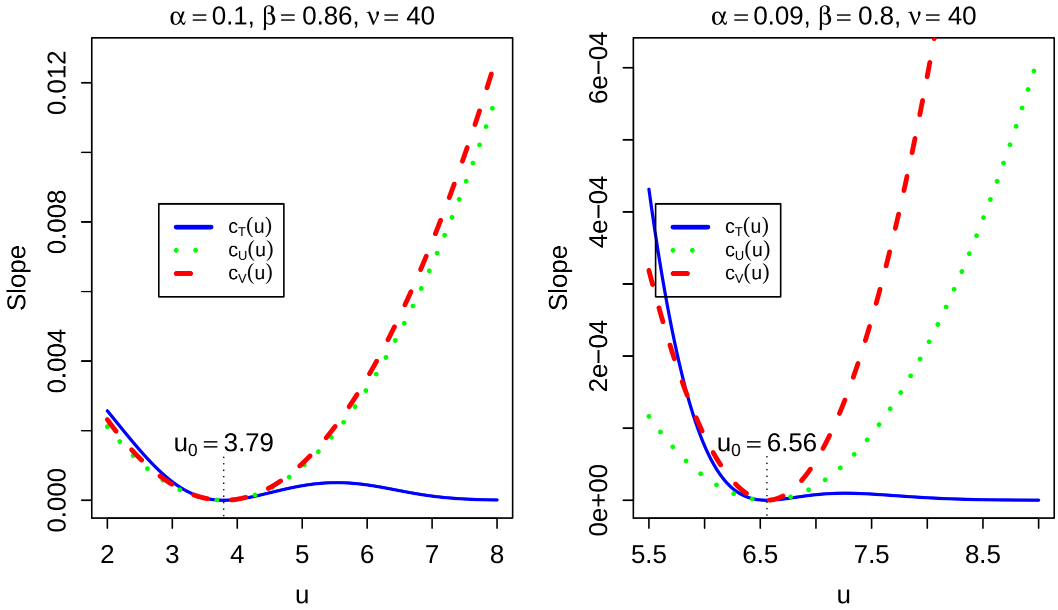


Figure 5: Asymptotic slopes of the tests $T, U$ and $V$ for Student errors ($\nu = 40$) and the standard GARCH(1,1) models.

method (the method based on $\widehat{u}_{n,h}$ with $f$ Gaussian). Note that the coverage probabilities are excellent (*i.e.* very close to nominal level $1-\alpha$) when $n = 4000$ or $n = 8000$. The results reported in Table 4 come from tests of $\boldsymbol{H}^*_{0,u}$, for the same experiments. In agreement with Figures 4-5 these results are more favorable to test $T$, even if the level is poorly controlled.

Next, we consider the Beta-$t$-GARCH models introduced by Harvey [36] and Creal et al. [15], such that

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\frac{(\nu+1)\epsilon_{t-1}^2}{(\nu-2)+\epsilon_{t-1}^2/\sigma_{t-1}^2},$$

and the rescaled innovations are Student's $t$ distributed with a degree of freedom $\nu$. This model is of the form (1) with $\delta = 2$, $\omega(\eta) = \omega$ and $a(\eta) = \beta + \frac{\alpha(\nu+1)\eta^2}{(\nu-2)+\eta^2}$. Note that for this model, even if the disturbances are $t$-distributed, we have $s = \infty$, i.e. $a(\eta_t)$ admits moments at any order. For the value of the parameter $\boldsymbol{\theta} = (\omega, \alpha, \beta, \nu)'$ used for the simulations, we have $u_0 = 3.5$. The results in Tables 5 and 6, obtained for simulations of this Beta-$t$-GARCH model, lead to similar conclusions as for standard GARCH models.

## J Complement to the empirical application

The QMLE and MLE residuals of the Total return series do not show any sign of dependence (in Figure 6, the aucorrelations of the squared residuals are not significantly non–zero). Moreover, it is seen that the distribution of the residuals is better represented by the Student than by the Gaussian distribution.

The empirical MDF $S_n^{(u)}$ is drawn in red in Figure 7. This curve crosses the horizontal line $y = 1$ at $\widehat{u}_0 = 7.9$, the estimated value of $u_0$ based on $U_n^{(u)}$. The MDF computed on the first 20 replications of the bootstrap simulation are plotted in Figure 7.

Table 3: For tests $T_n^{(u)}$ and $U_n^{(u)}$, relative frequency of rejection of $\boldsymbol{H}_{0,u}$ at the nominal level $\alpha\%$. The null hypothesis is true for $u \leqslant 4$ and false for $u > 4$. The last 3 columns are CIs for $u_0$ at the asymptotic confidence level $1-\alpha$. The column "mean" (resp. "median") gives the means (resp. medians) of the CI bounds. The column "coverage" gives the empirical coverage probability, that is the proportion of CIs that contain $u_0$ among the $N = 10,000$ replications.

| $n$ | $\alpha$ | Test | $u=2$ | $u=3$ | $u=4$ | $u=5$ | $u=6$ | $u=7$ | mean | median | coverage |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1000 | 1% | $T_n^{(u)}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | | | |
| | | $U_n^{(u)}$ | 0.00 | 0.02 | 1.32 | 8.44 | 22.15 | 38.96 | [0.49,9.04] | [0.62,8.03] | 0.99 |
| | | $V_n^{(u)}$ | 0.00 | 0.02 | 1.25 | 8.81 | 23.35 | 41.68 | [0.64,8.63] | [0.74,7.83] | 0.99 |
| | 5% | $T_n^{(u)}$ | 0.00 | 0.02 | 0.20 | 0.54 | 0.46 | 0.04 | | | |
| | | $U_n^{(u)}$ | 0.00 | 0.18 | 4.40 | 17.63 | 36.61 | 54.41 | [1.51,8.02] | [1.54,7.16] | 0.97 |
| | | $V_n^{(u)}$ | 0.00 | 0.17 | 4.71 | 18.62 | 39.07 | 57.89 | [1.60,7.68] | [1.61,6.98] | 0.97 |
| | 10% | $T_n^{(u)}$ | 0.00 | 0.12 | 2.07 | 6.25 | 9.24 | 8.12 | | | |
| | | $U_n^{(u)}$ | 0.00 | 0.55 | 8.02 | 25.04 | 46.13 | 63.58 | [2.03,7.50] | [2.02,6.72] | 0.95 |
| | | $V_n^{(u)}$ | 0.00 | 0.60 | 8.15 | 26.58 | 48.76 | 66.89 | [2.09,7.19] | [2.05,6.55] | 0.95 |
| 4000 | 1% | $T_n^{(u)}$ | 0.00 | 0.00 | 0.07 | 1.76 | 5.94 | 6.58 | | | |
| | | $U_n^{(u)}$ | 0.00 | 0.01 | 1.29 | 22.69 | 63.80 | 88.73 | [2.40,5.94] | [2.37,5.81] | 0.99 |
| | | $V_n^{(u)}$ | 0.00 | 0.00 | 1.27 | 23.83 | 66.49 | 90.86 | [2.43,5.86] | [2.40,5.76] | 0.99 |
| | 5% | $T_n^{(u)}$ | 0.00 | 0.01 | 1.94 | 21.05 | 52.55 | 72.22 | | | |
| | | $U_n^{(u)}$ | 0.00 | 0.03 | 4.95 | 40.63 | 79.93 | 95.47 | [2.82,5.51] | [2.79,5.40] | 0.95 |
| | | $V_n^{(u)}$ | 0.00 | 0.04 | 5.14 | 42.32 | 82.22 | 96.42 | [2.84,5.45] | [2.81,5.37] | 0.96 |
| | 10% | $T_n^{(u)}$ | 0.00 | 0.04 | 5.77 | 39.27 | 75.71 | 91.33 | | | |
| | | $U_n^{(u)}$ | 0.00 | 0.06 | 9.08 | 52.30 | 86.35 | 97.51 | [3.04,5.30] | [3.00,5.20] | 0.91 |
| | | $V_n^{(u)}$ | 0.00 | 0.05 | 9.39 | 54.00 | 88.53 | 98.10 | [3.05,5.24] | [3.01,5.16] | 0.91 |
| 8000 | 1% | $T_n^{(u)}$ | 0.00 | 0.00 | 0.21 | 14.62 | 56.87 | 79.55 | | | |
| | | $U_n^{(u)}$ | 0.00 | 0.00 | 1.26 | 40.82 | 90.13 | 99.38 | [2.87,5.30] | [2.84,5.25] | 0.99 |
| | | $V_n^{(u)}$ | 0.00 | 0.00 | 1.34 | 42.84 | 91.94 | 99.69 | [2.89,5.26] | [2.87, 5.22] | 0.99 |
| | 5% | $T_n^{(u)}$ | 0.00 | 0.00 | 2.67 | 47.84 | 90.75 | 98.53 | | | |
| | | $U_n^{(u)}$ | 0.00 | 0.00 | 5.30 | 62.47 | 96.47 | 99.92 | [3.16,5.01] | [3.13,4.96] | 0.95 |
| | | $V_n^{(u)}$ | 0.00 | 0.00 | 5.21 | 64.21 | 97.34 | 99.96 | [3.17,4.98] | [3.15,4.93] | 0.95 |
| | 10% | $T_n^{(u)}$ | 0.00 | 0.00 | 7.06 | 65.63 | 96.55 | 99.80 | | | |
| | | $U_n^{(u)}$ | 0.00 | 0.00 | 9.64 | 73.06 | 98.18 | 99.97 | [3.31,4.86] | [3.28,4.81] | 0.90 |
| | | $V_n^{(u)}$ | 0.00 | 0.00 | 9.87 | 75.25 | 98.67 | 99.99 | [3.31,4.83] | [3.29,4.79] | 0.90 |

Table 4: As first part of Table 3, but for the null $\boldsymbol{H}^*_{0,u}$, which is true for $u \geqslant 4$ and false for $u < 4$.

| $n$ | $\alpha$ | Test | $u=2$ | $u=3$ | $u=4$ | $u=5$ | $u=6$ | $u=7$ |
|---|---|---|---|---|---|---|---|---|
| 1000 | 1% | $T_n^{(u)}$ | 5.51 | 1.39 | 0.45 | 0.23 | 0.14 | 0.10 |
| | | $U_n^{(u)}$ | 0.51 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | | $V_n^{(u)}$ | 2.73 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | 5% | $T_n^{(u)}$ | 60.04 | 23.45 | 7.67 | 2.71 | 1.06 | 0.41 |
| | | $U_n^{(u)}$ | 51.54 | 9.13 | 0.04 | 0.00 | 0.00 | 0.00 |
| | | $V_n^{(u)}$ | 52.78 | 12.05 | 0.43 | 0.00 | 0.00 | 0.00 |
| | 10% | $T_n^{(u)}$ | 80.74 | 42.08 | 17.17 | 6.48 | 2.71 | 1.26 |
| | | $U_n^{(u)}$ | 76.11 | 30.85 | 7.78 | 1.34 | 0.00 | 0.00 |
| | | $V_n^{(u)}$ | 76.29 | 31.44 | 8.22 | 1.39 | 0.03 | 0.00 |
| 4000 | 1% | $T_n^{(u)}$ | 95.65 | 30.49 | 2.05 | 0.02 | 0.01 | 0.00 |
| | | $U_n^{(u)}$ | 92.01 | 16.43 | 0.26 | 0.01 | 0.00 | 0.00 |
| | | $V_n^{(u)}$ | 92.58 | 17.48 | 0.31 | 0.00 | 0.00 | 0.00 |
| | 5% | $T_n^{(u)}$ | 99.50 | 59.35 | 8.52 | 0.49 | 0.02 | 0.01 |
| | | $U_n^{(u)}$ | 99.22 | 49.96 | 4.33 | 0.07 | 0.01 | 0.00 |
| | | $V_n^{(u)}$ | 99.30 | 51.06 | 4.28 | 0.09 | 0.00 | 0.00 |
| | 10% | $T_n^{(u)}$ | 99.89 | 73.17 | 14.96 | 1.23 | 0.03 | 0.02 |
| | | $U_n^{(u)}$ | 99.85 | 68.10 | 10.93 | 0.55 | 0.02 | 0.01 |
| | | $V_n^{(u)}$ | 99.90 | 68.31 | 10.39 | 0.51 | 0.02 | 0.00 |
| 8000 | 1% | $T_n^{(u)}$ | 100.00 | 59.40 | 2.16 | 0.00 | 0.00 | 0.00 |
| | | $U_n^{(u)}$ | 99.97 | 46.04 | 0.53 | 0.00 | 0.00 | 0.00 |
| | | $V_n^{(u)}$ | 100.00 | 47.94 | 0.46 | 0.00 | 0.00 | 0.00 |
| | 5% | $T_n^{(u)}$ | 100.00 | 82.63 | 7.74 | 0.05 | 0.00 | 0.00 |
| | | $U_n^{(u)}$ | 100.00 | 77.97 | 4.95 | 0.01 | 0.00 | 0.00 |
| | | $V_n^{(u)}$ | 100.00 | 78.86 | 4.66 | 0.01 | 0.00 | 0.00 |
| | 10% | $T_n^{(u)}$ | 100.00 | 90.34 | 13.61 | 0.16 | 0.00 | 0.00 |
| | | $U_n^{(u)}$ | 100.00 | 88.31 | 10.54 | 0.06 | 0.00 | 0.00 |
| | | $V_n^{(u)}$ | 100.00 | 88.77 | 10.30 | 0.04 | 0.00 | 0.00 |

Table 5: Same results as presented in Table 3, but for $N = 1000$ replications of the Beta-t-GARCH model with $(\omega_0, \alpha_0, \beta_0, \nu_0) = (0.5, 0.1, 0.88, 7.78)$. The boundary of the null corresponds to $u = 3.5$.

| $n$ | $\alpha$ | Test | $u = 1.5$ | $u = 2.5$ | $u = 3.5$ | $u = 4.5$ | $u = 5.5$ | $u = 6.5$ |
|---|---|---|---|---|---|---|---|---|
| 2000 | 1% | $T_n^{(u)}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.10 |
| | | $U_n^{(u)}$ | 0.00 | 0.00 | 2.10 | 9.80 | 25.40 | 44.30 |
| | | $W_n^{(u)}$ | 0.00 | 0.00 | 1.10 | 9.80 | 28.20 | 50.10 |
| | 5% | $T_n^{(u)}$ | 0.00 | 0.00 | 0.60 | 3.80 | 6.10 | 7.60 |
| | | $U_n^{(u)}$ | 0.00 | 0.10 | 5.20 | 19.40 | 42.00 | 61.10 |
| | | $W_n^{(u)}$ | 0.00 | 0.10 | 4.30 | 19.60 | 45.70 | 64.20 |
| | 10% | $T_n^{(u)}$ | 0.00 | 0.10 | 4.20 | 11.70 | 20.70 | 27.60 |
| | | $U_n^{(u)}$ | 0.00 | 0.60 | 8.60 | 28.70 | 51.70 | 68.90 |
| | | $W_n^{(u)}$ | 0.00 | 0.60 | 7.40 | 29.70 | 53.90 | 71.10 |
| 4000 | 1% | $T_n^{(u)}$ | 0.00 | 0.00 | 0.00 | 0.50 | 2.30 | 3.10 |
| | | $U_n^{(u)}$ | 0.00 | 0.00 | 1.40 | 16.70 | 45.10 | 69.30 |
| | | $W_n^{(u)}$ | 0.00 | 0.00 | 1.20 | 18.40 | 50.90 | 76.40 |
| | 5% | $T_n^{(u)}$ | 0.00 | 0.00 | 2.10 | 13.60 | 29.10 | 41.40 |
| | | $U_n^{(u)}$ | 0.00 | 0.00 | 6.30 | 32.50 | 61.40 | 82.90 |
| | | $W_n^{(u)}$ | 0.00 | 0.00 | 5.30 | 33.60 | 68.40 | 84.90 |
| | 10% | $T_n^{(u)}$ | 0.00 | 0.00 | 6.90 | 27.60 | 52.30 | 69.00 |
| | | $U_n^{(u)}$ | 0.00 | 0.10 | 10.50 | 42.10 | 70.70 | 88.00 |
| | | $W_n^{(u)}$ | 0.00 | 0.10 | 9.30 | 44.60 | 76.90 | 89.70 |
| 8000 | 1% | $T_n^{(u)}$ | 0.00 | 0.00 | 0.00 | 5.60 | 23.00 | 42.70 |
| | | $U_n^{(u)}$ | 0.00 | 0.00 | 1.30 | 25.90 | 70.20 | 91.70 |
| | | $W_n^{(u)}$ | 0.00 | 0.00 | 1.00 | 32.30 | 78.10 | 95.60 |
| | 5% | $T_n^{(u)}$ | 0.00 | 0.00 | 2.90 | 29.00 | 68.80 | 87.50 |
| | | $U_n^{(u)}$ | 0.00 | 0.00 | 6.10 | 46.80 | 85.00 | 96.00 |
| | | $W_n^{(u)}$ | 0.00 | 0.00 | 5.60 | 54.00 | 89.40 | 98.60 |
| | 10% | $T_n^{(u)}$ | 0.00 | 0.00 | 7.20 | 48.80 | 84.40 | 95.20 |
| | | $U_n^{(u)}$ | 0.00 | 0.00 | 10.30 | 58.60 | 89.60 | 98.10 |
| | | $W_n^{(u)}$ | 0.00 | 0.00 | 10.40 | 65.20 | 93.60 | 99.30 |

Table 6: Same results as in Table 5, but for the null $\boldsymbol{H}^*_{0,u}$

| $n$ | $\alpha$ | Test | $u = 1.5$ | $u = 2.5$ | $u = 3.5$ | $u = 4.5$ | $u = 5.5$ | $u = 6.5$ |
|---|---|---|---|---|---|---|---|---|
| 2000 | 1% | $T_n^{(u)}$ | 13.80 | 2.40 | 0.00 | 0.00 | 0.00 | 0.00 |
| | | $U_n^{(u)}$ | 0.60 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | | $W_n^{(u)}$ | 7.80 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | 5% | $T_n^{(u)}$ | 62.40 | 21.70 | 6.40 | 1.50 | 0.50 | 0.10 |
| | | $U_n^{(u)}$ | 49.70 | 5.60 | 0.00 | 0.00 | 0.00 | 0.00 |
| | | $W_n^{(u)}$ | 61.40 | 6.90 | 0.00 | 0.00 | 0.00 | 0.00 |
| | 10% | $T_n^{(u)}$ | 82.50 | 39.40 | 13.40 | 4.40 | 1.20 | 0.50 |
| | | $U_n^{(u)}$ | 76.50 | 25.80 | 4.90 | 0.40 | 0.00 | 0.00 |
| | | $W_n^{(u)}$ | 86.50 | 29.80 | 1.60 | 0.00 | 0.00 | 0.00 |
| 4000 | 1% | $T_n^{(u)}$ | 62.60 | 10.50 | 0.90 | 0.00 | 0.00 | 0.00 |
| | | $U_n^{(u)}$ | 45.00 | 1.40 | 0.00 | 0.00 | 0.00 | 0.00 |
| | | $W_n^{(u)}$ | 68.50 | 1.80 | 0.00 | 0.00 | 0.00 | 0.00 |
| | 5% | $T_n^{(u)}$ | 90.40 | 36.10 | 6.80 | 0.80 | 0.00 | 0.00 |
| | | $U_n^{(u)}$ | 87.30 | 23.80 | 2.40 | 0.00 | 0.00 | 0.00 |
| | | $W_n^{(u)}$ | 96.00 | 31.90 | 0.60 | 0.00 | 0.00 | 0.00 |
| | 10% | $T_n^{(u)}$ | 96.00 | 51.80 | 11.90 | 2.80 | 0.20 | 0.00 |
| | | $U_n^{(u)}$ | 94.60 | 43.70 | 7.80 | 0.70 | 0.00 | 0.00 |
| | | $W_n^{(u)}$ | 99.10 | 53.70 | 6.90 | 0.00 | 0.00 | 0.00 |
| 8000 | 1% | $T_n^{(u)}$ | 96.20 | 25.90 | 1.30 | 0.00 | 0.00 | 0.00 |
| | | $U_n^{(u)}$ | 93.80 | 13.60 | 0.00 | 0.00 | 0.00 | 0.00 |
| | | $W_n^{(u)}$ | 99.10 | 23.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | 5% | $T_n^{(u)}$ | 99.70 | 58.00 | 6.50 | 0.10 | 0.00 | 0.00 |
| | | $U_n^{(u)}$ | 99.60 | 47.40 | 3.90 | 0.00 | 0.00 | 0.00 |
| | | $W_n^{(u)}$ | 100.00 | 62.60 | 2.20 | 0.00 | 0.00 | 0.00 |
| | 10% | $T_n^{(u)}$ | 99.70 | 72.40 | 11.30 | 0.80 | 0.00 | 0.00 |
| | | $U_n^{(u)}$ | 99.70 | 67.80 | 8.90 | 0.10 | 0.00 | 0.00 |
| | | $W_n^{(u)}$ | 100.00 | 78.60 | 7.80 | 0.00 | 0.00 | 0.00 |

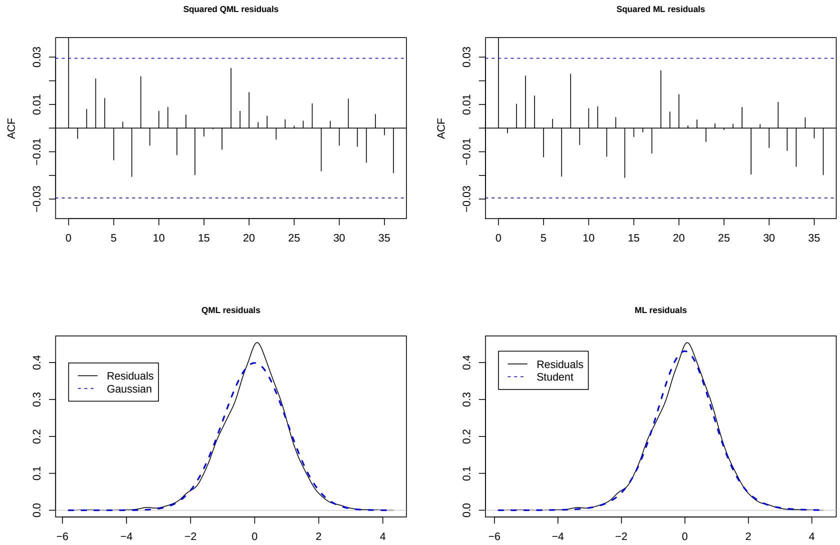


Figure 6: Autocorrelations of the squares of the QML and ML residuals, and empirical distributions of the QML and ML residuals, after fitting an APARCH model to the Total return series.

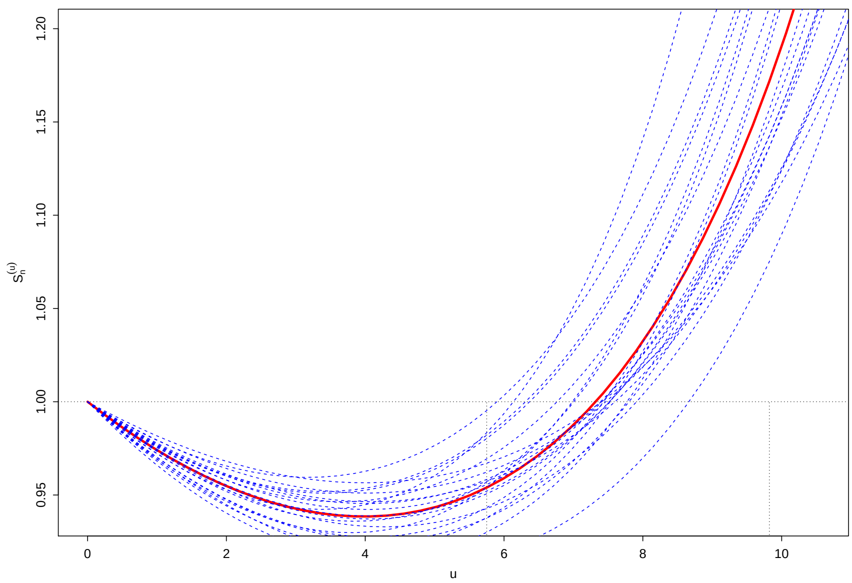


Figure 7: Empirical MDF for the APARCH(1,1) model fitted on the Total return series (red solid line), MDF of 20 bootstrap replications (blue dotted line), and 95% bootstrap interval (delimited by vertical dotted lines) over 10000 bootstrap replications.

## References


[1] Aue A., Berkes, I. and L. Horváth (2006). Strong approximation for the sums of squares of augmented GARCH sequences. *Bernoulli* **12** 583–608.

[2] Baek, C., Pipiras, V., Wendt, H. and P. Abry (2009). Second order properties of distribution tails and estimation of tail exponents in random difference equations. *Extremes* **12** 361–400.

[3] Basrak, B., Davis, R.A. and T. Mikosch (2002). Regular variation of GARCH processes. *Stochastic Process. Appl.* **99** 95–116.

[4] Berkes, I. and L. Horváth (2004). The efficiency of the estimators of the parameters in GARCH processes. *Ann. Statist.* **32** 633–655.

[5] Berkes, I., Horváth, L. and P.S. Kokoszka (2003). Estimation of the maximal moment exponent of a GARCH(1,1) sequence. *Econometric Theory* **19** 565–586.

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

[7] Billingsley P. (1968). *Convergence of Probability Measures* 1st ed. John Wiley, New York.

[8] Billingsley, P. (1961). The Lindeberg-Lévy theorem for martingales. *Proc. Amer. Math. Soc.* **12** 788–792.

[9] Blasques, F., Francq, C. and S. Laurent (2023). Quasi score-driven models. *J. Econometrics* **234** 251–275.

[10] Blasques, F., Gorgi, P., Koopman, S.J. and O. Wintenberger (2018). Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. *Electron. J. Stat.* **12** 1019–1052.

[11] Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31** 307–327.

[12] Bougerol, P. (1993). Kalman filtering with random coefficients and contractions. *SIAM J. Control and Optim.* **31** 942–959.

[13] Brandt, A. (1986) The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Adv. Appl. Probability* **18** 211–220.

[14] Chan, N.H., Li, D., Peng, L. and R. Zhang (2013). Tail index of an AR(1) model with ARCH(1) errors. *Econometric Theory* **29** 920–940.

[15] Creal, D., Koopman, S.J. and A. Lucas (2013). Generalized autoregressive score models with applications. *J. Appl. Econometrics* **28** 777–795.

[16] Davis, R. and T. Mikosch (2009). Extreme value theory for GARCH processes. In T. Andersen, R. Davis, J.-P. Kreiss, and T. Mikosch (Eds.), *Handbook of Financial Time Series*, 187–200. New York: Springer.

[17] Davis, R. and S. Resnick (1986) Limit theory for the sample covariance and correlation functions of moving averages. *Ann. Statist.* **14** 533–558.

[18] Delaigle, A., Meister, A. and J. Rombouts (2016). Root-T consistent density estimation in GARCH models. *J. Econometrics* **192** 55–63.

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

[20] Drees, H., Resnick, S. and L. de Haan (2000). How to make a Hill plot. *Ann. Statist.* **28** 254–274.

[21] Drost, F.C. and C.A.J. Klaassen (1997). Efficient estimation in semiparametric GARCH models. *J. Econometrics* **81** 193–221.

[22] Drost, F.C., Klaassen, C.A.J. and B.J.M. Werker (1997). Adaptive estimation in time-series models. *Ann. Statist.* **25** 786–817.

[23] Embrechts, P., Klüppelberg, C. and T. Mikosch (1997). *Modelling extremal events: for insurance and finance.* Springer, New-York.

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

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

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

[27] Francq, C. and J-M. Zakoïan (2009). A tour in the asymptotic theory of GARCH estimation. In *Handb. of Financial Time Series*, pp 85–111. Berlin, Heidelberg: Springer.

[28] Francq, C. and J-M. Zakoïan (2013a) Inference in nonstationary asymmetric GARCH models. *Ann. Statist.* **41** 1970–1998.

[29] Francq, C. and J-M. Zakoïan (2013b) Optimal predictions of powers of conditionally heteroskedastic processes. *J. Roy. Statist. Soc. - Series B* **75** 345–367.

[30] Francq, C. and J.M. Zakoïan (2019). *GARCH Models: Structure, Statistical Inference and Financial Applications.* John Wiley, Second edition.

[31] Francq, C. and J-M. Zakoïan (2022). Testing the existence of moments for GARCH processes. *J. Econometrics* **227** 47–64.

[32] Francq, C. and J-M. Zakoïan (2023). Local asymptotic normality of general conditionally heteroskedastic and score-driven time-series models. *Econometric Theory* **39** 1067–1092.

[33] Glosten, L.R., Jaganathan, R. and D. Runkle (1993). On the relation between the expected values and the volatility of the nominal excess return on stocks. *J. Finance* **48** 1779–1801.

[34] Goldie, C.M. (1991). Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probability* **1** 126–166.

[35] Hamadeh, T.. and J-M. Zakoïan (2011). Asymptotic properties of LS and QML estimators for a class of nonlinear GARCH Processes. *J. Statist. Plann. Inference* **141** 488–507.

[36] Harvey, A.(2013) *Dynamic Models for Volatility and Heavy Tails.* Cambridge University Press.

[37] Heinemann, A. (2019). A bootstrap test for the existence of moments for GARCH processes. *Preprint arXiv:1902.01808v3.*

[38] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3** 1163–1174.

[39] Hörmann, S. (2008). Augmented GARCH sequences: dependence structure and asymptotics. *Bernoulli* **14** 543–561.

[40] Kesten, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131** 207–248.

[41] Lee, S. and M. Taniguchi (2005). Asymptotic theory for ARCH-SM models: LAN and residual empirical processes. *Statist. Sinica* **15** 215–234.

[42] Ling, S. and M. McAleer (2002). Stationarity and the existence of moments of a family of GARCH processes. *J. Econometrics* **106** 109–117.

[43] Ling, S. and M. McAleer (2003). Adaptative estimation in nonstationary ARMA models with GARCH errors. *Ann. Statist.* **31** 642–674.

[44] Mikosch, T. and C. Stărică (2000). Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *Ann. Statist.* **28** 1427–1451.

[45] Ng, W.L. and C.Y. Yau (2018) Test for the existence of finite moments via bootstrap. *Nonparametr. Stat.* 30, 28–48.

[46] Straumann, D. (2005). *Estimation in Conditionally Heteroscedastic Time Series Models.* Lecture Notes in Statistics, Springer Berlin Heidelberg.

[47] Straumann, D. and T. Mikosch (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. *Ann. Statist.* **34** 2449–2495.

[48] Taylor, S. J. (1994) Modeling stochastic volatility: A review and comparative study. *Math. Finance* **4**, 183–204.

[49] Trapani, L. (2016). Testing for (in)finite moments. *J. Econometrics* **191** 57–68.

[50] van der Vaart, A.W. (1998). *Asymptotic statistics.* Cambridge University Press, United Kingdom.

[51] Zakoïan, J-M. (1994). Threshold heteroskedastic models. *J. Econom. Dynam. Control* **18**,931–955.

[52] Zhang, R., Li, C. and L. Peng (2019). Inference for the tail index of a GARCH(1,1) model and an AR(1) model with ARCH(1) errors. *Econometric Rev.* **38** 151–169.

[53] Zhang, R. and S. Ling (2015). Asymptotic inference for AR models with heavy-tailed G-GARCH noises. *Econometric Theory* **31** 880–890.

[54] Zhu, K. and S. Ling (2011). Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA-GARCH/IGARCH models. *Ann. Statist.* **39** 2131–2163.