Testing the existence of moments and estimating the tail index of augmented garch processes

Francq, Christian and Zakoian, Jean-Michel

2021

Testing the Existence of Moments and Estimating the Tail Index of Augmented GARCH Processes

Christian Francq
CREST and University of Lille
and
Jean-Michel Zakoian
CREST and University of Lille

November 5, 2021

Abstract

We investigate the problem of testing finiteness of moments for a class of semi-parametric augmented GARCH models encompassing most commonly used specifications. The existence of positive-power moments of the strictly stationary solution is characterized through the Moment Generating Function (MGF) of the model, defined as the MGF of the logarithm of the random autoregressive coefficient in the volatility dynamics. We establish the asymptotic distribution of the empirical MGF, from which tests of moments are deduced. Alternative tests relying on the 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, showing that semi-parametric estimation of the MME offers an interesting alternative to Hill's nonparametric estimator of the tail index.

Keywords: APARCH model, Bahadur slopes, Hill's estimator, Local asymptotic power, Maximal moment exponent, Moment generating function

1 Introduction

Volatility of financial returns certainly constitutes the most important concept in decision making based on risk analysis, portfolio management or asset pricing. For this reason, a plethora of models has emerged during the last four decades. Among them, GARCH-type formulations continue to attract the greatest attention, in particular due to their simplicity of use, flexibility and their seemingly infinite capacity of extensions.

By construction, GARCH models are based on specifications of the *conditional* variance but, indirectly, the volatility dynamics constrains the shape of the marginal distribution of the returns process, in particular through the *unconditional* moments. For most classes of GARCH models, moments do not exist at any order and their existence is not a simple consequence of the model coefficients, but also depends intricately (not only through the moments) of the innovations distribution. Necessary and sufficient conditions for the existence of moments of GARCH processes are well-known, at least for the standard GARCH formulation (e.g. Ling and McAleer (2002)), but little attention has been devoted to testing these conditions. 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 asymptotic properties of the Quasi-Maximum Likelihood (QML) estimators of GARCH models hold without any extra moment assumption, many applications rest on finite unconditional moments. Moreover, the existence of moments for financial returns is *per se* an interesting issue, which regularly gives rise to controversial views in the empirical finance literature.

The present paper proposes new methods for testing the existence of moments for a general class of GARCH-type processes. A first step in this direction has been taken in Francq and Zakoïan (2021a) who proposed tests of the existence of even-order moments for the standard GARCH model. In this set up, the problem essentially 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 (2019) for a bootstrap-based approach). However, this approach cannot be extended to other GARCH formulations for which the moment conditions are not so explicit. Moreover, it does not allow to handle non even-order moments, in particular non-integer power moments.

1.1 Augmented GARCH

We consider the class of *augmented* GARCH processes (see e.g. Aue et al. (2006)), defined as

$$\left\{ \begin{array}{rcl} \epsilon_t & = & \sigma_t \eta_t, \\ \sigma_t^\delta & = & \omega(\eta_{t-1}) + a(\eta_{t-1})\sigma_{t-1}^\delta \end{array} \right. \tag{1}$$

where $(\eta_t)_{t\geq 0}$ is a sequence of independent and identically distributed (iid) random variables with zero mean and unit variance, δ is a positive constant, and the functions $\omega(\cdot)$ and $a(\cdot)$ satisfy $\omega : \mathbb{R} \to [\underline{\omega}, +\infty)$ and $a : \mathbb{R} \to [\underline{a}, +\infty)$, for some $\underline{\omega} > 0$ and $\underline{a} \geq 0$. This class, introduced by He and Teräsvirta (1999), encompasses most GARCH-type models introduced in the literature.

1.2 Two characterizations of the existence of moments

Under appropriate conditions, the model admits a strictly stationary solution (ϵ_t) which has a moment of order $u\delta$, for $u > 0$, if and only if $E|\eta_t|^{u\delta} < \infty$ and $E(\sigma_t^{u\delta}) < \infty$. The latter condition can be formulated as follows (see Ling and McAleer (2002) and Aue et al. (2006))

$$E(\sigma_t^{u\delta}) < \infty \qquad \Leftrightarrow \qquad E[a^u(\eta_1)] < 1 \quad \text{and} \quad E[\omega^u(\eta_1)] < \infty. \tag{2}$$

The behaviour of the function $u \mapsto E[a^u(\eta_1)]$, which will be called throughout *Moment Generating Function* (MGF) of Model (1) is thus crucial for the existence of moments. In general, the MGF cannot be expressed as a function of moments of η_t, making the approach developed in Francq and Zakoïan (2021a) inapplicable in this framework.

Under mild conditions discussed below, there exists a unique $u_0 > 0$ such that $E[a^{u_0}(\eta_1)] = 1$ and the moment condition can be written

$$E(\sigma_t^{u\delta}) < \infty \qquad \Leftrightarrow \qquad u < u_0. \tag{3}$$

Following the terminology of Berkes et al. (2003), who proposed an estimator of the coefficient for standard GARCH(1,1) models, the coefficient u_0 will be referred to as the *Maximal Moment Exponent* (MME). Under mild additional assumptions, this coefficient will be related to the *tail index* of the distribution of ϵ_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 MGF and MME. Relying on a semi-parametric version of Model (1), in which the functions a and ω depend on a finite-dimensional parameter $\boldsymbol{\theta}_0$ but the distribution of η_t is left unspecified, we will provide conditions for the consistency and asymptotic normality of the empirical MGF and MME

$$S_n^{(u)} = \frac{1}{n}\sum_{t=1}^{n} a^u(\hat{\eta}_t; \widehat{\boldsymbol{\theta}}_n), \qquad \hat{u}_n = \sup\{u > 0; \;\; S_n^{(u)} \leq 1\}, \tag{4}$$

where $\widehat{\boldsymbol{\theta}}_n$ denotes any consistent estimator of $\boldsymbol{\theta}_0$, and $\hat{\eta}_t, t = 1, \dots, n$ denote the residuals.

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

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

where $\hat{\upsilon}_u^2$ and $\widehat{w}_{\hat{u}_n}^2$ denote consistent estimators of the asymptotic variances of $S_n^{(u)}$ and $\hat{u}_n$, respectively, tests of the moment condition $E(\sigma_t^{u\delta}) < \infty$ are defined by the rejection regions

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

where Φ is the $\mathcal{N}(0,1)$ cumulative distribution function. Assuming that η_t has a known density f, or a parametric density $f(\cdot; \boldsymbol{\nu})$, parametric versions $V_n^{(u)}$ and $W_n^{(u)}$ of the statistic U will also be introduced.

1.4 Contributions of the paper

For the semi-parametric version of Model (1), we study the aforementioned tests for the existence of moments. The model being semi-parametric we will not rely on 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, some of our results being particularized for the QML and ML methods.

Our main contributions are as follows:

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

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

c) we propose new tests of the moment condition;

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

e) we provide power comparisons of the semi-parametric and parametric tests under local alternatives or using the Bahadur approach.

1.5 Organisation of the paper

In Section 2, we develop the asymptotic theory for the empirical MGF. Section 3 derives the test based on the MGF, while Section 4 derives the test based on the MME. Comparisons based on local alternatives are studied in Section 5. The case where the power δ is unknown is studied in Section 6. An empirical illustration is displayed in Section 7. Section 8 concludes. Finally, in appendix we present the proofs of our results, additional properties and Monte-Carlo experiments.

2 Estimating the MGF of the augmented GARCH

Consider a semi-parametric version of Model (1) defined by the equations

$$\begin{cases} \epsilon_t &= \sigma_t(\boldsymbol{\theta}_0)\eta_t, \\ \sigma_t^\delta(\boldsymbol{\theta}_0) &= \omega(\eta_{t-1};\boldsymbol{\theta}_0) + a(\eta_{t-1};\boldsymbol{\theta}_0)\sigma_{t-1}^\delta(\boldsymbol{\theta}_0) \end{cases} \tag{5}$$

where $\delta > 0$ is given (see Section 6 for an extension) and $\boldsymbol{\theta}_0 \in \mathbb{R}^d$ is a vector of parameters. Let $\boldsymbol{\theta}$ denote a generic value of the parameter, which is assumed to belong to a compact parameter set $\Theta \subset \mathbb{R}^d$. Assume that, for any $\boldsymbol{\theta} \in \Theta$, the functions $\omega(\cdot;\boldsymbol{\theta})$ and $a(\cdot;\boldsymbol{\theta})$ satisfy $\omega(\cdot;\boldsymbol{\theta}) : \mathbb{R} \to [\underline{\omega}, +\infty)$ and $a(\cdot;\boldsymbol{\theta}) : \mathbb{R} \to [\underline{a}, +\infty)$.

The second equation in (5) has the form of a stochastic recurrence equation (SRE) which enables to study its probability properties. Let (ϵ_t) denote the strictly stationary, non-anticipative[1] and ergodic solution of Model (5) (under Assumption **A1** in Appendix A). Given

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

observations $\epsilon_1, \dots, \epsilon_n$, and arbitrary initial values $\tilde{\epsilon}_0$ and $\tilde{\sigma}_0 > 0$ we define, for $t = 1, \dots, n$ and any $\boldsymbol{\theta}$ belonging to Θ,

$$\tilde{\sigma}_t^\delta(\boldsymbol{\theta}) = \omega\left(\frac{\epsilon_{t-1}}{\tilde{\sigma}_{t-1}(\boldsymbol{\theta})}; \boldsymbol{\theta}\right) + a\left(\frac{\epsilon_{t-1}}{\tilde{\sigma}_{t-1}(\boldsymbol{\theta})}; \boldsymbol{\theta}\right)\tilde{\sigma}_{t-1}^\delta(\boldsymbol{\theta})$$

where $\tilde{\sigma}_0(\boldsymbol{\theta}) = \tilde{\sigma}_0$ and $\epsilon_0 = \tilde{\epsilon}_0$. The above SRE raises the question of the *invertibility* of the model, which holds only if $\tilde{\sigma}_t^\delta(\boldsymbol{\theta})$ does not depend asymptotically on the initialization (see Straumann and Mikosch (2006), Blasques et al. (2018)). Under condition **A3** below, the sequence $(\tilde{\sigma}_t^\delta(\boldsymbol{\theta}))_{t\geq 0}$ can be approximated by a stationary ergodic process $(\sigma_t^\delta(\boldsymbol{\theta}))$ solution of the SRE

$$\sigma_t^\delta(\boldsymbol{\theta}) \quad = \quad \omega\left(\frac{\epsilon_{t-1}}{\sigma_{t-1}(\boldsymbol{\theta})}; \boldsymbol{\theta}\right) + a\left(\frac{\epsilon_{t-1}}{\sigma_{t-1}(\boldsymbol{\theta})}; \boldsymbol{\theta}\right)\sigma_{t-1}^\delta(\boldsymbol{\theta}), \qquad t \in \mathbb{Z}. \tag{6}$$

Lemma 1 in appendix provides conditions for the existence of a strictly stationary solution to the previous SRE. Assume that for some $s > 0$, $E[a^s(\eta_1; \boldsymbol{\theta}_0)] < \infty$. For $0 < u \leq s$, consider the estimator $S_n^{(u)}$ defined in (4) of the MGF $S_\infty^{(u)} := E[a^u(\eta_1; \boldsymbol{\theta}_0)]$ where $\widehat{\boldsymbol{\theta}}_n$ denotes any strongly consistent estimator of $\boldsymbol{\theta}_0 \in \Theta$, $\hat{\eta}_t = \epsilon_t/\hat{\sigma}_t$, with $\hat{\sigma}_t = \tilde{\sigma}_t(\widehat{\boldsymbol{\theta}}_n)$. To simplify the presentation, precise assumptions, labelled **A1-A10** are relegated to Appendix A. In particular, a moment assumption on $a(\eta_t, \boldsymbol{\theta}_0)$ is required. This assumption is in general much weaker than the corresponding assumption for the observed process. In some models, the moment assumption on $a(\eta_t, \boldsymbol{\theta}_0)$ is inocuous (as in the Beta-t-GARCH of Harvey (2013) and Creal et al. (2013) where the variables $a(\eta_t, \boldsymbol{\theta}_0)$ are bounded). In general, this assumption can be assessed using the filtered variables $a(\tilde{\eta}_t, \widehat{\boldsymbol{\theta}}_n)$ and by applying the nonparametric approach of Hill (2015).

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

Theorem 1 *Under* **A1**-**A6** *and* **A7**(u) *with* $0 < u \leq s/2$, *we have*

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

where $\boldsymbol{\Sigma} = E(\boldsymbol{\Delta}_t\boldsymbol{\Upsilon}\boldsymbol{\Delta}_t')$, $\psi_u = Var[a^u(\eta_1; \boldsymbol{\theta}_0)]$, $\boldsymbol{\xi}_u = \boldsymbol{\Lambda}E[\boldsymbol{V}(\eta_t)a^u(\eta_t; \boldsymbol{\theta}_0)]$, $\boldsymbol{g}_u = E\left(\boldsymbol{g}_{u,t}\right)$ *where* $\boldsymbol{g}_{u,t} = \left[\frac{\partial}{\partial\boldsymbol{\theta}}a^u\{\eta_t(\boldsymbol{\theta}); \boldsymbol{\theta}\}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$. *Moreover* $v_u^2 > 0$ *whenever* $Var\{a^u(\eta_t; \boldsymbol{\theta}_0) + \boldsymbol{g}_u'\boldsymbol{\Delta}_{t-1}\boldsymbol{V}(\eta_t)\} > 0$. [2]

The asymptotic variance of the empirical MGF has a more explicit form in the case of the GARCH(1,1) ($\delta = 2$) for two important estimation methods: the Gaussian QML and the ML.

Corollary 1 (GARCH(1,1)) *For the standard GARCH(1,1), under the assumptions of Theorem 1, letting* $M_{x,y} = E[\eta_t^{2x}(\alpha_0\eta_t^2 + \beta_0)^y]$, $x, y \in \mathbb{R}$, *and*

$$\boldsymbol{\Omega} = E\left(\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right), \qquad \boldsymbol{J} = E\left(\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\right), \tag{8}$$

[2] A sufficient condition is the positive-definiteness of the covariance matrix of the vector $\{a^u(\eta_t; \boldsymbol{\theta}_0), \boldsymbol{V}'(\eta_t)\}$.

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*

$$\upsilon_u^2 \;=\; c_\eta u^2\left\{\boldsymbol{m}_u'\boldsymbol{J}^{-1}\boldsymbol{m}_u - \alpha_0^2 M_{1,u-1}^2\right\} + M_{0,2u} - M_{0,u}^2, \tag{9}$$

where $c_\eta = \kappa_4 - 1$ *with* $\kappa_4 = E\eta_t^4$ *for the QMLE, and* $c_\eta = 4/\iota_f$ *for the MLE, where* $\iota_f = \int\left\{1 + yf'(y)/f(y)\right\}^2 f(y)d\mu(y)$ *is the Fisher information for scale.*[3]

An example of model of the form (1) is the APARCH (Asymmetric Power ARCH) of Ding et al. (1993) defined by $\omega(\eta) = \omega$, $a(\eta) = \alpha_+|\eta|^\delta \mathbb{1}_{\eta>0} + \alpha_-|\eta|^\delta \mathbb{1}_{\eta<0} + \beta$. For APARCH estimated by QML, the assumptions of Theorem 1 can be considerably reduced.

Corollary 2 (APARCH model) *Under the following assumptions: i)* $P(\eta_t > 0) \in (0,1)$, *the support of the distribution of* η_t *contains at least three points, and* $E(|\eta_t|^{s\delta}) < \infty$ *with* $s\delta \geq 4$; *ii)* $\boldsymbol{\Theta} \subset [\underline{\omega}, \infty) \times (0,\infty)^2 \times [0,1)$ *is compact and* $\boldsymbol{\theta}_0 \in \overset{\circ}{\boldsymbol{\Theta}}$, *iii)* $E\log a(\eta_1, \boldsymbol{\theta}_0) < 0$, *the conclusions of Theorem 1 hold for the QML estimator and* $u \leq s/2$.

3 Testing the existence of moments of given order using the MGF

For $u > 0$, consider the $u\delta$-th order moments testing problems

$$\boldsymbol{H}_{0,u}: \quad E(|\epsilon_t|^{u\delta}) < \infty \quad \text{against} \quad \boldsymbol{H}_{1,u}: \quad E(|\epsilon_t|^{u\delta}) = \infty, \tag{10}$$

and

$$\boldsymbol{H}_{0,u}^*: \quad E(|\epsilon_t|^{u\delta}) = \infty \quad \text{against} \quad \boldsymbol{H}_{1,u}^*: \quad E(|\epsilon_t|^{u\delta}) < \infty. \tag{11}$$

Note that by (2), under the conditions

$$E|\eta_1|^{u\delta} < \infty, \quad E[\omega^u(\eta_1)] < \infty, \tag{12}$$

the testing problem can be equivalently written as

$$\boldsymbol{H}_{0,u}: \quad E\{a^u(\eta_t)\} < 1 \quad \text{against} \quad \boldsymbol{H}_{1,u}: \quad E\{a^u(\eta_t)\} \geq 1, \tag{13}$$

and similarly for $\boldsymbol{H}_{0,u}^*$. Let the test statistic based on the empirical MGF

$$T_n^{(u)} = \frac{\sqrt{n}\left\{S_n^{(u)} - 1\right\}}{\hat{\upsilon}_u}, \quad \text{where} \quad \hat{\upsilon}_u^2 = \hat{\boldsymbol{g}}_u'\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{g}}_u + \hat{\psi}_u + 2\hat{\boldsymbol{g}}_u'\hat{\xi}_u,$$

provided $\hat{\upsilon}_u^2 > 0$, with

$$\hat{\boldsymbol{g}}_u = \frac{1}{n}\sum_{t=1}^n \frac{\partial}{\partial\boldsymbol{\theta}} a^u\left(\frac{\epsilon_t}{\tilde{\sigma}_t}(\widehat{\boldsymbol{\theta}}_n); \widehat{\boldsymbol{\theta}}_n\right), \quad \hat{\psi}_u = \frac{1}{n}\sum_{t=1}^n a^{2u}\left(\frac{\epsilon_t}{\tilde{\sigma}_t}(\widehat{\boldsymbol{\theta}}_n); \widehat{\boldsymbol{\theta}}_n\right) - \left\{\frac{1}{n}\sum_{t=1}^n a^u\left(\frac{\epsilon_t}{\tilde{\sigma}_t}(\widehat{\boldsymbol{\theta}}_n); \widehat{\boldsymbol{\theta}}_n\right)\right\}^2$$

and $\hat{\xi}_u$ and $\hat{\boldsymbol{\Sigma}}$ strongly consistent estimators of ξ_u and $\boldsymbol{\Sigma}$.

[3] assuming that η_t has a density f with respect to some σ-finite measure μ. Conditions for the existence of ι_f are provided in Assumptions **B1-B2** of Appendix A.

Proposition 1 *Under the assumptions of Theorem 1 with $\upsilon_u > 0$ and under (12), a test of $\boldsymbol{H}_{0,u}$ [resp. $\boldsymbol{H}^*_{0,u}$] at the asymptotic level $\underline{\alpha} \in (0,1)$ is defined by the rejection region*

$$C_T^{(u)} = \{T_n^{(u)} > \Phi^{-1}(1-\underline{\alpha})\}, \qquad [resp. \;\; \{T_n^{(u)} < \Phi^{-1}(\underline{\alpha})\}]. \tag{14}$$

This result provides an extension of a test studied by Francq and Zakoian (2021a) in the case where u is even and (ϵ_t) follows a standard GARCH. In this framework, the moment condition is an explicit function of $\boldsymbol{\theta}_0$ and moments of η_t. The test statistic is thus computed differently, but is equivalent to $T_n^{(u)}$, as the next example illustrates.

Example 1 (2nd-order stationarity testing ($u = 1$) in standard GARCH ($\delta = 2$)) *We have $a(\eta, \boldsymbol{\theta}) = \alpha\eta^2 + \beta$. When the model is estimated by Gaussian QML we have, by Corollary 1, $\upsilon_1^2 = (\kappa_4 - 1)\boldsymbol{e}_0'\boldsymbol{J}^{-1}\boldsymbol{e}_0 + (\alpha_0 + \beta_0)^2 - 1$, where $\boldsymbol{e}_0' = (0,1,1)$. Thus under $\mathbf{H}_{0,1}$,*

$$S_n^{(1)} \;\; = \;\; \frac{1}{n}\sum_{t=1}^{n}(\hat{\alpha}_n\hat{\eta}_t^2 + \hat{\beta}_n) = \hat{\alpha}_n + \hat{\beta}_n + o_P(1), \quad \upsilon_1^2 = (\kappa_4 - 1)\boldsymbol{e}_0'\boldsymbol{J}^{-1}\boldsymbol{e}_0.$$

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

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

4 Estimating the MME and alternative tests

In the next proposition, we gather existing results on the existence of a finite MME.

Proposition 2 *Suppose $\gamma = E\log a(\eta_1) < 0$.*

i) If $P[a(\eta_1) \le 1] = 1$: for all $u > 0$, $E[a^u(\eta_1)] < 1$, and $E(\sigma_t^{u\delta}) < \infty$ if $E[\omega^u(\eta_1)] < \infty$.

ii) If $P[a(\eta_1) \le 1] < 1$ and $1 \le E[a^s(\eta_1)] < \infty$ for some $s > 0$: there exists a unique $u_0 > 0$ such that $E[a^{u_0}(\eta_1)] = 1$.
Moreover, if $E[a^u(\eta_1)] < 1$ and $E[a^v(\eta_1)] > 1$ for $0 < u < v$ then $u_0 \in (u,v)$. In addition, if $E[\omega^{u_0}(\eta_1)] < \infty$, then $E(\sigma_t^{u\delta}) < \infty$ for all $u < u_0$, and $E(\sigma_t^{u\delta}) = \infty$ for $u \ge u_0$.

If ii) holds, the law of $\log a(\eta_1)$ is nonarithmetic, 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 $x \to \infty$.

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

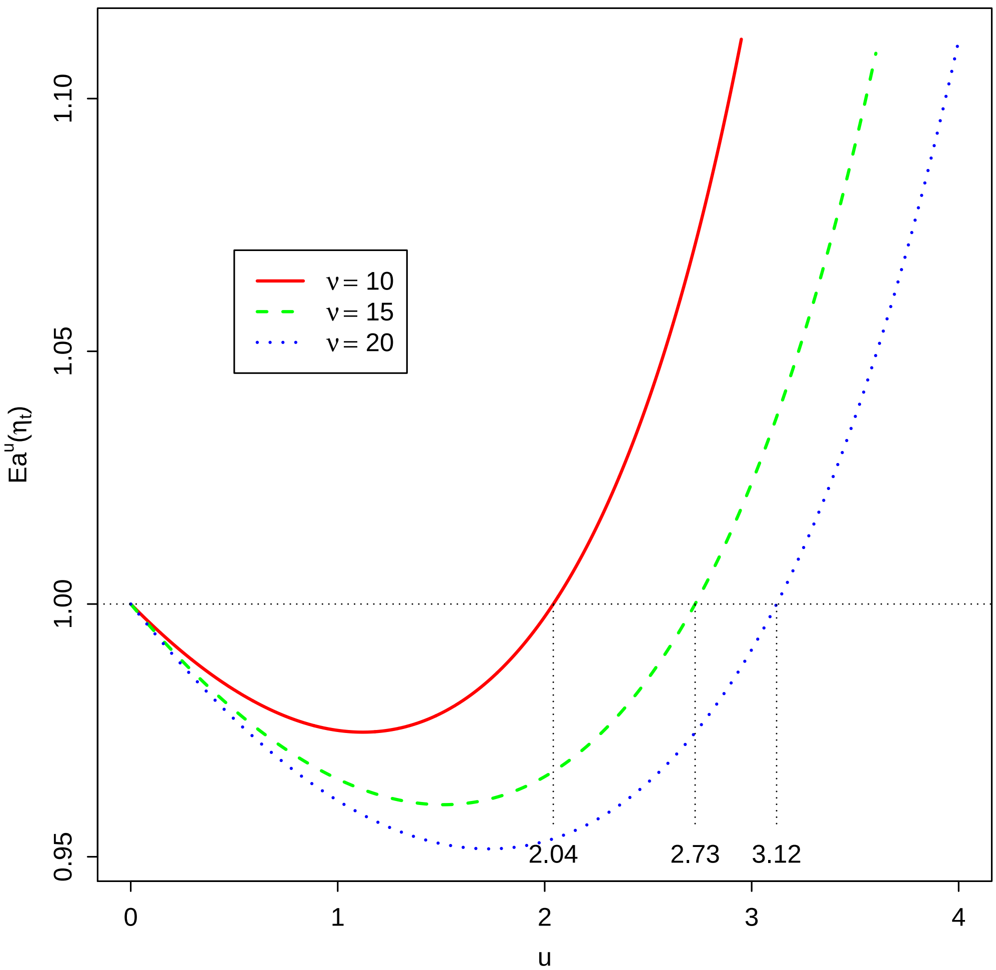

Figure 1: MGF for the standard GARCH(1,1) model with $\alpha_0 = 0.10, \beta_0 = 0.85$ and for Student errors with ν degrees of freedom. Values of the MME u_0 are displayed over the horizontal axis.

Remark 2 *The tail properties in this proposition–established in the case of standard GARCH by Mikosch and Stărică (2000) and for augmented GARCH by Zhang and Ling (2015)–show that, under mild additional assumptions, the coefficient δu_0 is also the tail index of the augmented GARCH process. Conditions for the existence of a tail index for general SRE were derived by Basrak et al. (2002), and Kesten (1973) characterized this coefficient as the solution of an equation taking the form $E[a^{u_0}(\eta_1)] = 1$ in the case of augmented GARCH(1,1) processes.*

Proposition 2 is illustrated in Figure 1 for Student distributions with $\nu = 10, 15$ and 20.

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

4.1 Semi-parametric estimation of the MME

The following result is the sample counterpart of Proposition 2.

Proposition 3 *Suppose* $\gamma_n := \frac{1}{n}\sum_{t=1}^{n} \log a(\hat{\eta}_t; \widehat{\boldsymbol{\theta}}_n) < 0$.

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

Conversely, if $a(\hat{\eta}_t; \widehat{\boldsymbol{\theta}}_n) > 1$ *for at least one* $1 \leq t \leq n$, *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 $\hat{u}_n = \sup\{u > 0;\ \ S_n^{(u)} \leq 1\}$, we have $\hat{u}_n = \infty$ when $a(\hat{\eta}_t; \widehat{\boldsymbol{\theta}}_n) \leq 1$ for all $1 \leq t \leq n$, and $\hat{u}_n = u_n$ (of Proposition 3) in the opposite case. We will show the strong consistency of $\hat{u}_n$.

Theorem 2 *Suppose* $\gamma = E\{a(\eta_t)\} < 0$, *with* $a(\eta) = a(\eta; \boldsymbol{\theta}_0)$. *Under* **A1-A4, A6** *and* $E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{\partial}{\partial \boldsymbol{\theta}} \log a(\epsilon_t/\sigma_t(\boldsymbol{\theta}); \boldsymbol{\theta}) \right\| < \infty$, *we have* $\gamma_n \to \gamma$, *a.s. Moreover, if*

i) $P[a(\eta_1) \leq 1] = 1$, *then* $\hat{u}_n \to \infty$, *a.s.*

ii) $P[a(\eta_1) > 1] > 0$, *and* $1 < E[a^s(\eta_1)] < \infty$ *for some* $s > 0$, *then* $\hat{u}_n \to u_0$, *a.s., where* $u_0 > 0$ *is such that* $E[a^{u_0}(\eta_1)] = 1$.

In order to obtain the asymptotic distribution of $\hat{u}_n$, we will now show the following 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 on the space $\mathcal{C}$ equipped with uniform distance.

Theorem 3 *If* **A1-A6** *and* **A7**(u_2) *hold, for* $[u_1, u_2] \subset (0, s/2)$

$$\sqrt{n}\left\{S_n^{(u)} - S_\infty^{(u)}\right\} \stackrel{\mathcal{C}[u_1,u_2]}{\Longrightarrow} \Gamma(u) \tag{15}$$

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

Let $D_\infty^{(u)} = E[a^u(\eta_1; \boldsymbol{\theta}_0) \log\{a(\eta_1; \boldsymbol{\theta}_0)\}]$ the first-order derivative of the MGF $u \to S_\infty^{(u)}$, which is well-defined for $u < s$ under **A1**. Note that $D_\infty^{(u_0)}$ is positive (in view of the convexity of the MGF). The asymptotic distribution of the MME was derived in the standard GARCH case by Mikosch and Stărică (2000) and Berkes et al. (2003), for Double AR(1) models by Chan et al. (2013), and for both models using a least absolute deviation estimator by Zhang et al. (2019). For the augmented GARCH, we have the following result.

Theorem 4 *Under the assumptions of Theorem 3, if Assumption ii) of Theorem 2 holds, with* $u_0 \in (0, u_2)$, *we have*

$$\sqrt{n}(\hat{u}_n - u_0) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, w_{u_0}^2 := \{D_\infty^{(u_0)}\}^{-2} \upsilon_{u_0}^2\right).$$

This result allows to build asymptotic confidence intervals (CI) for the MME u_0 and also, by Proposition 2, for the tail index of the distribution of ϵ_t. Hill's estimator of the tail index has been studied for time series models under different dependence assumptions (as for instance in Drees (2000) or Resnick and Stărică (1998)). However, this estimator crucially depends on the choice of the fraction of sample on which it is computed (see for instance Figure 1 in Zhu and Ling (2011)). Moreover, Baek, Pipiras, Wendt and Abry (2009) showed that the Hill estimator is extremely biased for estimating the tail index of ARCH-type models. Even for iid 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[4] (see below experiments in the numerical section using Student distributions). The derivation of CI for the tail index using Hill's estimator is even more challenging. By Theorem 4 one can estimate the tail index of an augmented GARCH at a parametric rate, instead of resorting to extreme value statistics. A similar situation occurs for the estimation of the density of a GARCH(1,1) since, by exploiting the dynamic structure of the model, Delaigle et al. (2016) managed to provide a root-n consistent estimator. Trapani (2016) 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.

Now consider testing (10) for a given $u > 0$. Note that the null assumption can be equivalently written $\boldsymbol{H}_{0,u}: \quad u < u_0$. Let the test statistic,

$$U_n^{(u)} = \frac{\sqrt{n}\,\{u - \hat{u}_n\}}{\widehat{w}_{\hat{u}_n}}, \quad \text{where} \quad \widehat{w}_u^2 = \left\{ \frac{1}{n} \sum_{t=1}^{n} a^u(\hat{\eta}_t; \widehat{\boldsymbol{\theta}}_n) \log\{a(\hat{\eta}_t; \widehat{\boldsymbol{\theta}}_n)\} \right\}^{-2} \hat{v}_u^2.$$

Proposition 4 *Under the assumptions of Theorem 4 with $w_{u_0}^2 > 0$, and (12), a test of $\boldsymbol{H}_{0,u}$ [resp. $\boldsymbol{H}_{0,u}^*$] at the asymptotic level $\underline{\alpha} \in (0,1)$ is defined by the rejection region*

$$C_U^{(u)} = \{U_n^{(u)} > \Phi^{-1}(1-\underline{\alpha})\}, \qquad [\textit{resp.} \quad \{U_n^{(u)} < \Phi^{-1}(\underline{\alpha})\}], \tag{16}$$

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

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

4.2 Purely parametric estimators of the MME

In this section, we assume that η_t has a density f which is positive everywhere, with third-order derivatives and satisfying some regularity assumptions displayed in Appendix A. These regularity conditions are satisfied for numerous distributions, including the Gaussian distribution, and entail the existence of the Fisher information for scale ι_f introduced in Corollary 1.

[4] According to Drees et al. (2000), "One would have to be paranormal to discern with confidence the true value from the Hill plot."

4.2.1 When the errors density is known

When the density f of η_t is known, under the assumption ii) of Theorem 2, given $\boldsymbol{\theta}$ the maximal moment exponent $u_0 = u_{0,f}(\boldsymbol{\theta})$ can be obtained by solving the implicit equation

$$\int a^{u_0}(x;\boldsymbol{\theta})f(x)dx = 1.$$

Under **B3** this solution satisfies, by the implicit function theorem,

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

For the classical GARCH(1,1), we have $\boldsymbol{r}_{u_0} = u_0\boldsymbol{m}_{u_0}$ where $\boldsymbol{m}_{u_0}$ is defined in Corollary 1.

Let $\hat{u}_{n,f} = u_{0,f}(\widehat{\boldsymbol{\theta}}_{n,ML})$ where $\widehat{\boldsymbol{\theta}}_{n,ML}$ is the MLE of $\boldsymbol{\theta}_0$, that is, the estimator of u_0 obtained by solving

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

Note that $\hat{u}_{n,f}$ is the ML estimator of u_0 (by the functional invariance of the ML estimator) which is not the case of $\hat{u}_n$ (even when $\widehat{\boldsymbol{\theta}}_n$ is the ML estimator of $\boldsymbol{\theta}_0$).

Under regularity assumptions (derived by Berkes and Horváth (2004) in the case of the standard GARCH(p,q) model), the MLE of $\boldsymbol{\theta}_0$ satisfies an expansion displayed in **B4** (see Appendix A). Let the test statistic

$$V_n^{(u)} = \frac{\sqrt{n}(u - \hat{u}_{n,f})}{\hat{\sigma}_f},$$

where $\hat{\sigma}_f$ is a consistent estimator of $\sigma_f = \left(\frac{4}{\iota_f}\frac{\partial u_0}{\partial \boldsymbol{\theta}'}\boldsymbol{J}^{-1}\frac{\partial u_0}{\partial \boldsymbol{\theta}}\right)^{1/2} = \frac{1}{D_\infty^{(u_0)}}\left\{\frac{4}{\iota_f}\boldsymbol{r}'_{u_0}\boldsymbol{J}^{-1}\boldsymbol{r}_{u_0}\right\}^{1/2}$.

Proposition 5 *Let Assumption ii) of Theorem 2,* (12)*, Assumptions* **B1-B4** *hold, and let* $\boldsymbol{r}_{u_0} \neq \mathbf{0}$. *Then, a test of* $\boldsymbol{H}_{0,u}$ *[resp.* $\boldsymbol{H}_{0,u}^*$*] at the asymptotic level* $\underline{\alpha} \in (0,1)$ *is defined by the rejection region*

$$C_V^{(u)} = \{V_n^{(u)} > \Phi^{-1}(1-\underline{\alpha})\}, \qquad [\textit{resp.} \;\; \{V_n^{(u)} < \Phi^{-1}(\underline{\alpha})\}], \tag{17}$$

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

4.2.2 When the error density is parametrized

In practical situations, assuming that the density f of η_t is known is not realistic. Alternatively, the density can be supposed to be known up to some finite parameter: $f(\cdot) = f(\cdot, \boldsymbol{\nu}_0)$ 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 $u_0 = u_{0,f}(\boldsymbol{\varphi})$ of

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

Under **B5**, we have

$$\frac{\partial u_{0,f}(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\theta}} = \frac{-1}{D_\infty^{(u_0)}}\boldsymbol{r}_{u_0},\ \frac{\partial u_{0,f}(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\nu}} = \frac{-1}{D_\infty^{(u_0)}}\boldsymbol{s}_{u_0},\ \boldsymbol{s}_{u_0} := E\left(a^{u_0}(\eta_t;\boldsymbol{\theta}_0)\frac{1}{f(\eta_t;\boldsymbol{\nu}_0)}\frac{\partial f(\eta_t;\boldsymbol{\nu}_0)}{\partial \boldsymbol{\nu}}\right).$$

Let $\widehat{u}_{n,\hat{f}} = u_{0,f}(\widehat{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\nu}}_n)$ where $(\widehat{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\nu}}_n)$ is the MLE of $\boldsymbol{\varphi}_0$, which can be obtained by solving

$$\int a^{\widehat{u}_{n,\hat{f}}}(x;\widehat{\boldsymbol{\theta}}_n)f(x,\widehat{\boldsymbol{\nu}}_n)dx = 1. \tag{18}$$

The asymptotic properties of the ML estimator of $\boldsymbol{\varphi}_0$ were established by Straumann (Chapter 6, 2005). For the sake of brevity, we defer to this reference for precise assumptions under which such properties rely. We assume that the MLE satisfies the expansion in **B6**. Let the test statistic $W_n^{(u)} = \frac{\sqrt{n}(u-\hat{u}_{n,\hat{f}})}{\hat{\varsigma}_f}$ where $\hat{\varsigma}_f$ is a consistent estimator of

$$\varsigma_f = \left\{\left(\frac{\partial u_0}{\partial \boldsymbol{\theta}'}, \frac{\partial u_0}{\partial \boldsymbol{\nu}'}\right)\mathfrak{J}^{-1}\left(\frac{\partial u_0}{\partial \boldsymbol{\theta}'}, \frac{\partial u_0}{\partial \boldsymbol{\nu}'}\right)'\right\}^{1/2} = \frac{1}{D_\infty^{(u_0)}}\left\{\left(\boldsymbol{r}'_{u_0}, \boldsymbol{s}'_{u_0}\right)\mathfrak{J}^{-1}\left(\boldsymbol{r}'_{u_0}, \boldsymbol{s}'_{u_0}\right)'\right\}^{1/2}.$$

Proposition 6 *Let Assumption ii) of Theorem 2,* (12)*, Assumptions* **B1-B2** *with* $f(\cdot)$ *replaced by* $f(\cdot;\boldsymbol{\nu}_0)$*,* **B5-B6** *hold, and let* $(\boldsymbol{r}'_{u_0}, \boldsymbol{s}'_{u_0}) \neq \mathbf{0}$*. Then, a test of* $\boldsymbol{H}_{0,u}$ *[resp.* $\boldsymbol{H}^*_{0,u}$*] at the asymptotic level* $\underline{\alpha} \in (0,1)$ *is defined by the rejection region*

$$C_W^{(u)} = \{W_n^{(u)} > \Phi^{-1}(1-\underline{\alpha})\}, \qquad [\mathit{resp.}\ \ \{W_n^{(u)} < \Phi^{-1}(\underline{\alpha})\}], \tag{19}$$

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

5 Asymptotic power comparisons

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 the Le Cam theory, they are also equivalent under local alternatives. In this section, we will compare these tests with the parametric ones and also provide non-local comparisons.

5.1 Asymptotic power under local alternatives

Conditional on ϵ_0 and σ_0, the density of the observations $(\epsilon_1, \ldots, \epsilon_n)$ satisfying (5) is given by $L_{n,f}(\boldsymbol{\theta}_0) = \prod_{t=1}^n \sigma_t^{-1}(\boldsymbol{\theta}_0)f\left\{\sigma_t^{-1}(\boldsymbol{\theta}_0)\epsilon_t\right\}$.

Around $\boldsymbol{\theta}_0 \in \overset{\circ}{\Theta}$, let 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 by $P_{n,\boldsymbol{\tau}}$ (resp. P_0) the distribution of the observations when the parameter is $\boldsymbol{\theta}_n$ (resp. $\boldsymbol{\theta}_0$). Under the assumptions ii) of Theorem 2, for given f and $\boldsymbol{\theta}_0$, there exists a unique $u_0 := u_0(\boldsymbol{\theta}_0, f)$ 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 (1997) showed that for standard GARCH, the log-likelihood ratio $\Lambda_{n,f}(\boldsymbol{\theta}_n, \boldsymbol{\theta}_0) = \log L_{n,f}(\boldsymbol{\theta}_n)/L_{n,f}(\boldsymbol{\theta}_0)$ satisfies the LAN property

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

where

$$\boldsymbol{\Delta}_{n,f}(\boldsymbol{\theta}_0) = \frac{-1}{\sqrt{n}}\sum_{t=1}^{n} g_1(\eta_t)\frac{1}{\sigma_t}\frac{\partial\sigma_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\mathfrak{I}}_f\right), \quad \boldsymbol{\mathfrak{I}}_f = \iota_f E\left(\frac{1}{\sigma_t^2}\frac{\partial\sigma_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\right)$$

under $P_{\boldsymbol{\theta}_0}$ as $n \to \infty$. Note that the so-called central sequence $\boldsymbol{\Delta}_{n,f}(\boldsymbol{\theta}_0)$ is conditional on the initial values. It is shown in Drost et al. (1997) and Ling and McAleer (2003) that (22) continues to hold in more general frameworks. Lee and Taniguchi (2005) 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 (LAP) of our tests.

Proposition 7 *Under Assumptions* **B1-B2**, (22) *and the assumptions of Propositions 1 and 4, respectively, the LAP of the tests of* $\boldsymbol{H}_{0,u_0}$ *defined in (14) and (16) are given by*

$$\lim_{n\to\infty} P_{n,\boldsymbol{\tau}}\left(C_T^{(u_0)}\right) = \lim_{n\to\infty} P_{n,\boldsymbol{\tau}}\left(C_U^{(u_0)}\right) = \Phi\left\{c_{f,u_0}(\theta_0) - \Phi^{-1}(1-\underline{\alpha})\right\} \tag{23}$$

where, using $g_1(y) = 1 + y\frac{f'}{f}(y)$,

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

For instance, in the standard GARCH(1,1) model estimated by QML or by ML, computations reported in appendix show that, with obvious notation,

$$c_{f,u_0}^{QML}(\boldsymbol{\theta}_0) = \frac{u_0}{\upsilon_{u_0,QML}}\boldsymbol{\tau}'\boldsymbol{m}_{u_0} \leq c_{f,u_0}^{ML}(\boldsymbol{\theta}_0) = \frac{u_0}{\upsilon_{u_0,ML}}\boldsymbol{\tau}'\boldsymbol{m}_{u_0}, \tag{24}$$

where the denominators are displayed in (9). It can be noted that the tests are locally asymptotically unbiased (i.e. $c_{f,u_0} > 0$) whenever $\tau_3/\tau_2 \geq \beta_0/\alpha_0$.

In the usual 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 give conditions for the tests T and U to be optimal in this sense.

Proposition 8 *Assume that $u_0(\boldsymbol{\theta}_0 + \frac{\varepsilon \boldsymbol{e}}{\sqrt{n}}, f) < u_0(\boldsymbol{\theta}_0, f)$ for n large enough and any $\varepsilon > 0$. Then, under the assumptions of Proposition 7, any asymptotically locally UMPU test for testing $\boldsymbol{H}_{0,u}$: $u_0(\boldsymbol{\theta}_0, f) > u$ against $\boldsymbol{H}_{1,n,u}$: $u(\boldsymbol{\theta}_0 + \frac{\varepsilon \boldsymbol{e}}{\sqrt{n}}, f) \leq u$ has asymptotic power bounded by*

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

For the standard GARCH(1,1) 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 η_t has the form

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

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

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

Proposition 9 *Under the assumptions of Propositions 5 and 7, the LAP of the test of $\boldsymbol{H}_{0,u_0}$ defined in (17) is given by*

$$\lim_{n\to\infty} P_{n,\boldsymbol{\tau}}\left(C_V^{(u_0)}\right) = \Phi\left\{d_{f,u_0}(\theta_0) - \Phi^{-1}(1 - \underline{\alpha})\right\} \quad \text{where} \quad d_{f,u_0}(\boldsymbol{\theta}_0) = \frac{\boldsymbol{r}_{u_0}'\boldsymbol{\tau}}{\sqrt{\frac{4}{\iota_f}\boldsymbol{r}_{u_0}'\boldsymbol{J}^{-1}\boldsymbol{r}_{u_0}}}. \tag{27}$$

Under the assumptions of Proposition 8, the test $C_V^{(u_0)}$ is optimal if the vectors $\boldsymbol{r}_{u_0}$ and $\boldsymbol{e}$ are collinear.

Next, we turn to the case of Section 4.2.2 where the errors density is parametrized 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 by $P_{n,\boldsymbol{\tau}}$ (resp. P_0) the distribution of the 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 $\Lambda_n(\boldsymbol{\varphi}_0 + \boldsymbol{\tau}/\sqrt{n}, \boldsymbol{\varphi}_0) = \log L_{n,f}(\boldsymbol{\varphi}_n)/L_{n,f}(\boldsymbol{\varphi}_0)$.

The LAN property (22) holds when the density f can be treated as an infinite-dimensional nuisance parameter. In Francq and Zakoïan (2021b), we show 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

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

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

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

The next result provides the LAP of the test W.

Proposition 10 *Under Assumptions* **B1-B2**, (29)-(30)*, and the assumptions of Proposition 6, the LAP of the test of $\boldsymbol{H}_{0,u_0}$ defined in (19) is given by*

$$\lim_{n\to\infty} P_{n,\boldsymbol{\tau}}\left(C_W^{(u_0)}\right) = \Phi\left\{e_{f,u_0}(\theta_0) - \Phi^{-1}(1-\underline{\alpha})\right\}, \quad e_{f,u_0}(\boldsymbol{\theta}_0) = \frac{\boldsymbol{r}'_{u_0}\boldsymbol{\tau}_1 + \boldsymbol{s}'_{u_0}\boldsymbol{\tau}_2}{\sqrt{\left(\boldsymbol{r}'_{u_0}, \boldsymbol{s}'_{u_0}\right)\mathfrak{J}^{-1}\left(\boldsymbol{r}'_{u_0}, \boldsymbol{s}'_{u_0}\right)'}}.$$

Under the assumptions of Proposition 8, the test $C_W^{(u_0)}$ is optimal if the vectors $\left(\boldsymbol{r}'_{u_0}, \boldsymbol{s}'_{u_0}\right)$ and $\boldsymbol{\tau}'$ are colinear.

Propositions 7,9 and 10 (with $\boldsymbol{\tau}_2 = 0$) are illustrated in Figure 2 for Student distributions with $\nu = 5, 20, 30$ and ∞. In the GARCH(1,1) case the LAPs of the tests T, U, V and W depend on $\boldsymbol{\tau}$ through $\boldsymbol{m}'_{u_0}\boldsymbol{\tau}$, which is thus reported in 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 test W.

5.2 Comparisons based on Bahadur slopes

To be able to distinguish the tests T and U, we turn to the Bahadur approach. We will also compare them with the tests V and W requiring 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. 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 $n\to\infty$,

$$T_n^{(u)} = \frac{\sqrt{n}\left\{S_n^{(u)} - 1\right\}}{\hat{\upsilon}_u} \sim \frac{\sqrt{n}\left\{S_\infty^{(u)} - 1\right\}}{\upsilon_u}, \quad U_n^{(u)} = \frac{\sqrt{n}\left\{u - \hat{u}_n\right\}}{\widehat{w}_{\hat{u}_n}} \sim \frac{\sqrt{n}\left\{u - u_0\right\}}{w_{u_0}},$$
$$V_n^{(u)} = \frac{\sqrt{n}(u - \hat{u}_{n,\hat{f}})}{\hat{\sigma}_f} \sim \frac{\sqrt{n}(u-u_0)}{\sigma_f}, \qquad W_n^{(u)} = \frac{\sqrt{n}(u - \hat{u}_{0,\hat{f}})}{\hat{\varsigma}_f} \sim \frac{\sqrt{n}(u-u_0)}{\varsigma_f}.$$

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

$$c_T(u) = \frac{\left\{S_\infty^{(u)} - 1\right\}^2}{\upsilon_u^2}, \; c_U(u) = \frac{\{u-u_0\}^2}{w_{u_0}^2}, \; c_V(u) = \frac{\{u-u_0\}^2}{\sigma_f^2} \quad \text{and} \quad c_W(u) = \frac{\{u-u_0\}^2}{\varsigma_f^2}.$$

In the Bahadur sense, the test $T_n^{(u)}$ is more efficient than $U_n^{(u)}$ if and only if

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

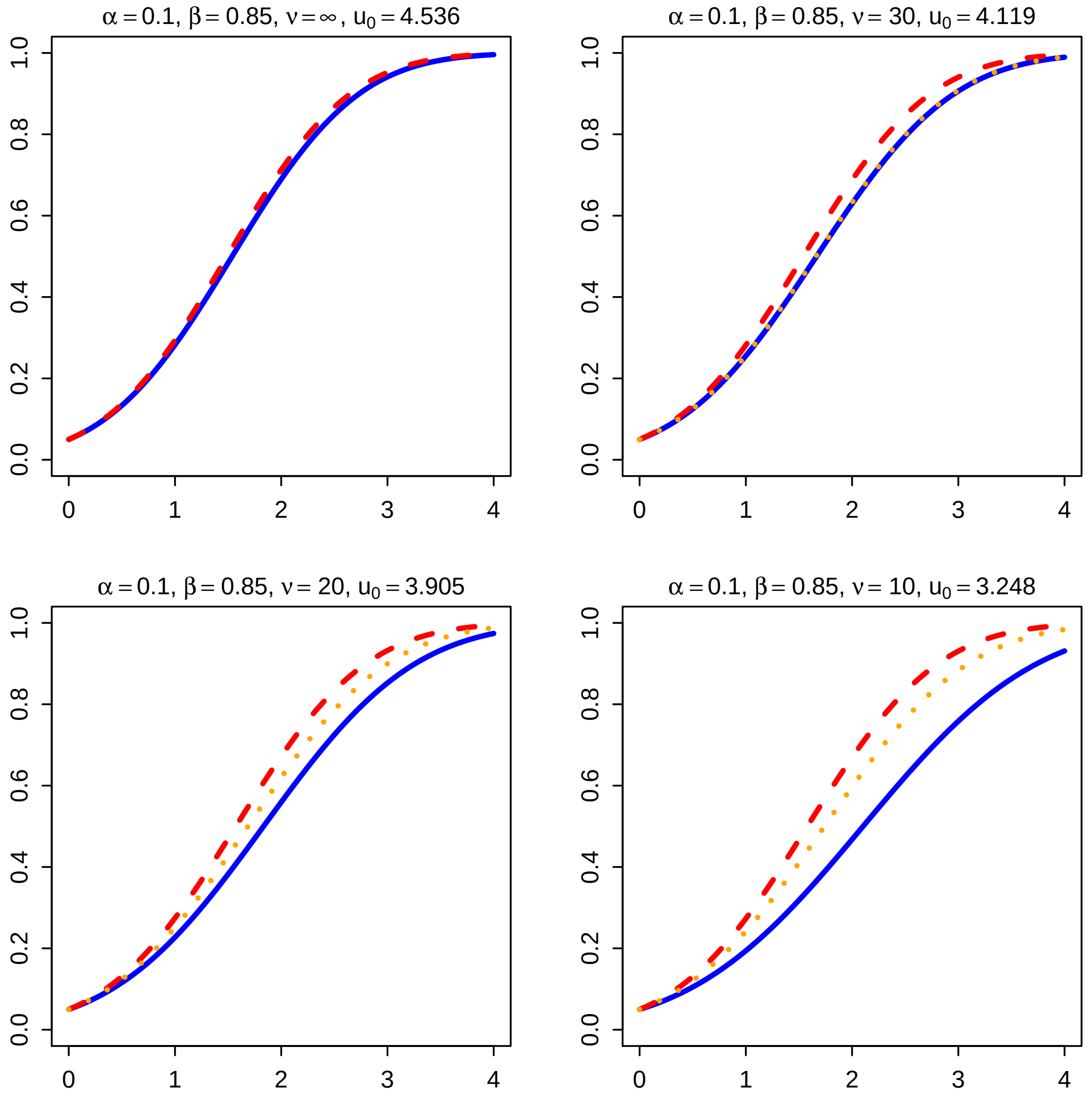

Figure 2: LAPs of the test T and U (blue line) based on the Gaussian QML, the test V (dotted red line), and the test W (dotted orange line) as functions of $\boldsymbol{m}_{u_0}'\boldsymbol{\tau}$, for the standard GARCH(1,1) model with $\alpha_0 = 0.10, \beta_0 = 0.85$ and for Student errors with ν degrees of freedom.

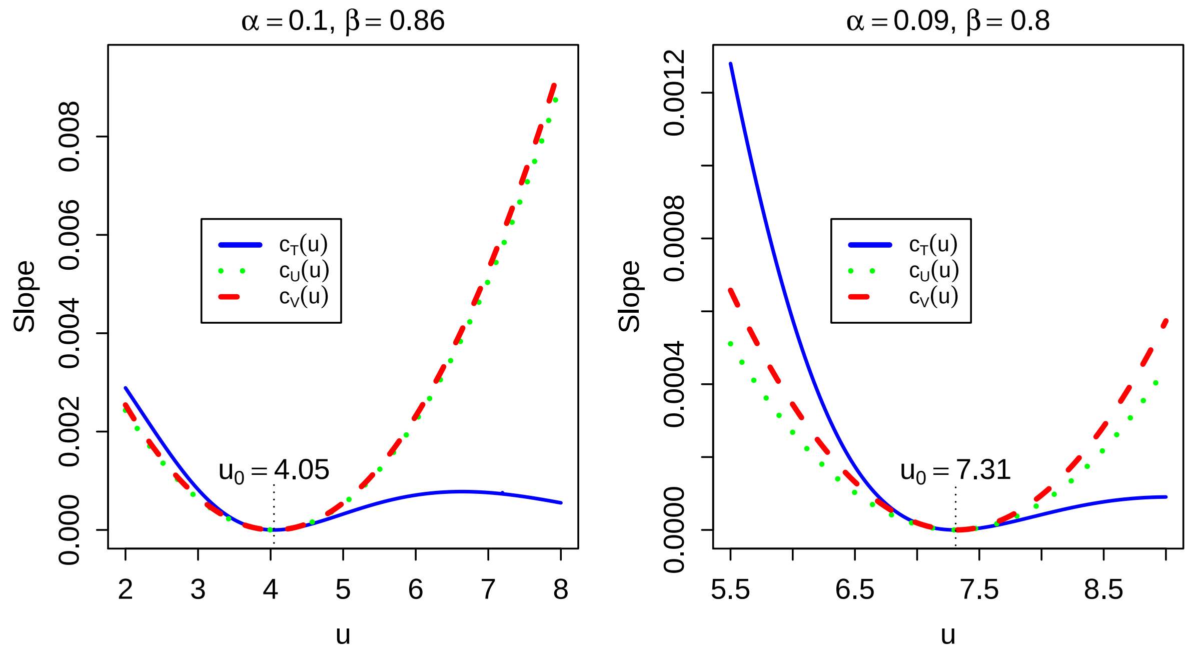

Figure 3: Asymptotic slopes of the tests T, U and V for Gaussian errors and the standard GARCH(1,1) models.

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

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

Note that the latter condition does not depend on u, i.e. on the alternative.

Examples of asymptotic slopes in the standard GARCH(1,1) case with Gaussian errors are displayed in Figure 3. It is clear from these graphs that, for the alternative $\boldsymbol{H}_{1,u}: \quad u > u_0$, the test U based on the MME is more efficient than the 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 are in favor of the test T. The test V has always better power 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 the 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 the test T can be satisfied while the condition $u_0 < s/2$, required for the validity of the test U, might be violated. Similar graphs for Student errors are reported in appendix.

Monte-Carlo experiments displayed in appendix illustrate the lack of power of the test T, compared to its competitors, in agreement with Figure 3.

6 Selecting δ

In practice, estimating $\delta > 0$ is very challenging. Even if the asymptotic normality of the joint QML estimator of δ and $\boldsymbol{\theta}_0$ has been established, the value of δ can be extremely difficult to identify in finite sample (see Hamadeh and Zakoïan (2011)). The quasi-likelihood being very flat in the direction of δ, estimating this coefficient may entail considerable numerical difficulties and result in poor accuracy. For this reason, instead of treating δ as a real-valued parameter, practitioners tend to select δ from a finite set of values corresponding to well-known models such as the standard or GJR-GARCH ($\delta = 2$) or the T-GARCH ($\delta = 1$). To reflect the existence of several candidates for δ, assume that the true value δ_0 belongs to a finite set,

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

For the sake of illustration, we focus on the APARCH model and the QML estimator.

Write the vector of parameters as $\boldsymbol{\vartheta} = (\delta, \boldsymbol{\theta}')'$ and assume $\boldsymbol{\vartheta} \in \mathcal{D} \times \boldsymbol{\Theta}$ where $\boldsymbol{\Theta}$ is a compact subset of $(0, \infty) \times [0, \infty)^3 \times [0, 1)$. The true parameters value is denoted by $\boldsymbol{\vartheta}_0 = (\delta_0, \boldsymbol{\theta}_0')'$. In order to define the QMLE of $\boldsymbol{\vartheta}$, we define recursively $\tilde{\sigma}_t$, for $t \geq 1$, by

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

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

$$\widehat{\boldsymbol{\vartheta}}_n^{QML} = \arg\min_{\boldsymbol{\vartheta} \in \mathcal{D} \times \boldsymbol{\Theta}} \tilde{\mathbf{l}}_n(\boldsymbol{\vartheta}), \qquad \tilde{\mathbf{l}}_n(\boldsymbol{\vartheta}) = n^{-1} \sum_{t=1}^n \tilde{\ell}_t, \qquad \tilde{\ell}_t = \tilde{\ell}_t(\boldsymbol{\vartheta}) = \frac{\epsilon_t^2}{\tilde{\sigma}_t^2} + \ln \tilde{\sigma}_t^2.$$

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

Proposition 11 *Under the following assumptions: i) η_t has a positive density on some neighborhood of zero, $E(|\eta_t|^{s\delta_0}) < \infty$ with $s\delta_0 \geq 4$; ii) $\boldsymbol{\theta}_0 \in \overset{\circ}{\boldsymbol{\Theta}}$, 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 conclusions of Theorems 1 and 3 hold. If η_t has a positive density over the real line, the conclusion of Theorem 4 holds.*

As a consequence, the tests of the previous sections can be applied without modification for this model.

7 Empirical application

Davis and Mikosch (2009) 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 (2021a) 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, they fitted a standard GARCH(1,1) model and, using $T_n^{(u)}$ for testing 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) we are able to test finiteness of any *real-order* moment, and (iii) our analysis is not restricted to the standard GARCH model, there is hope to improve the results obtained in Francq and Zakoïan (2021a).

We thus reconsidered the same series and estimated APARCH(1,1) models, by 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 ν degrees of freedom for the iid 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 are given into brackets. The model estimated by Student-ML is

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

where $\mathrm{St}(\nu)$ denotes the standardized Student with ν degrees of freedom. Note that the volatilities estimated by QML and ML are almost the same. Results displayed in appendix show that the QMLE and MLE residuals do not show any sign of dependence and that the distribution of the residuals is better represented by the Student than by the 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, which is $3 + 6/(\nu - 4) = 3.848$, than the Gausian kurtosis). Table 1 shows that the tests based on $U_n^{(u)}$ and $W_n^{(u)}$ give similar results, and 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 the 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 the asymptotic confidence level 95%, an estimated CI 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 empirical MGF $S_n^{(u)}$ is drawn in red in Figure 5. This curve crosses the horizontal line $y = 1$ at $\widehat{u}_0 = 7.9$, the estimated value of u_0 based on $U_n^{(u)}$. To have an idea of the variability of this estimator without relying on the asymptotic theory, we simulated APARCH(1,1) models with parameter $\widehat{\boldsymbol{\theta}}_n$–the QMLE computed on the Total series–and a noise whose distribution is that of the QML residuals. The MGF computed on the first 20 replications of this bootstrap simulation are plotted in Figure 5. Using 10000 bootstrap replications, an empirical 95% CI for u_0 is $[5.7, 9.8]$, which is quite similar to the estimates obtained from the asymptotic theory. The two estimation methods based on $U_n^{(u)}$ and $W_n^{(u)}$ therefore give a similar estimated tail index but, as expected, the fully parametric method based on $W_n^{(u)}$ provides a tighter CI. We thus find strong evidence for

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

the existence of finite moments of order 5 or 6, which makes it possible to validate certain statistical procedures, such as the construction of confidence intervals for the prediction of the squared returns over a long horizon. By contrast, as can be seen in Figure 4, the usual Hill estimator does not seem to be informative on the value of the tail index, both on the Total series (left graph) and on a simulation of the model for which we know that $u_0 = 7.8$ is maximum moment order. Note that Figure 4 is in perfect agreement with Figures 2 and 3 of Baek et al. (2009).

From this study we draw the conclusions that: 1) the tests proposed here are much more effective than the Hill estimator to assess the value of the tail index of a GARCH-type model; 2) estimating the maximum moment order is a difficult problem (since the CI remain large, even in a fully parametric framework); 3) at least for the Total series, moments seem to exist at an order much larger than two, which leads to relativize the overly pessimistic statement quoted at the beginning of this section.

8 Conclusion

In this paper we introduced statistics for testing the existence of moments of given order based on the MGF and MME of augmented GARCH processes. The tests are amenable to different parametric or semi-parametric estimators of the model parameter. We provided local and non-local asymptotic comparisons of the tests and illutrated their usefulness on a real financial series. Estimation of the MME offers an interesting alternative to Hill's estimator of the tail index which is often non informative in practice.

A potential extension of our tests would concern higher-order GARCH volatilities of the form

$$\sigma_t^\delta = \omega(\eta_{t-1}, \dots, \eta_{t-k}) + a(\eta_{t-1}, \dots, \eta_{t-k})\sigma_{t-k}^\delta$$

where $k \geq 1$. This extension is left for future investigations.

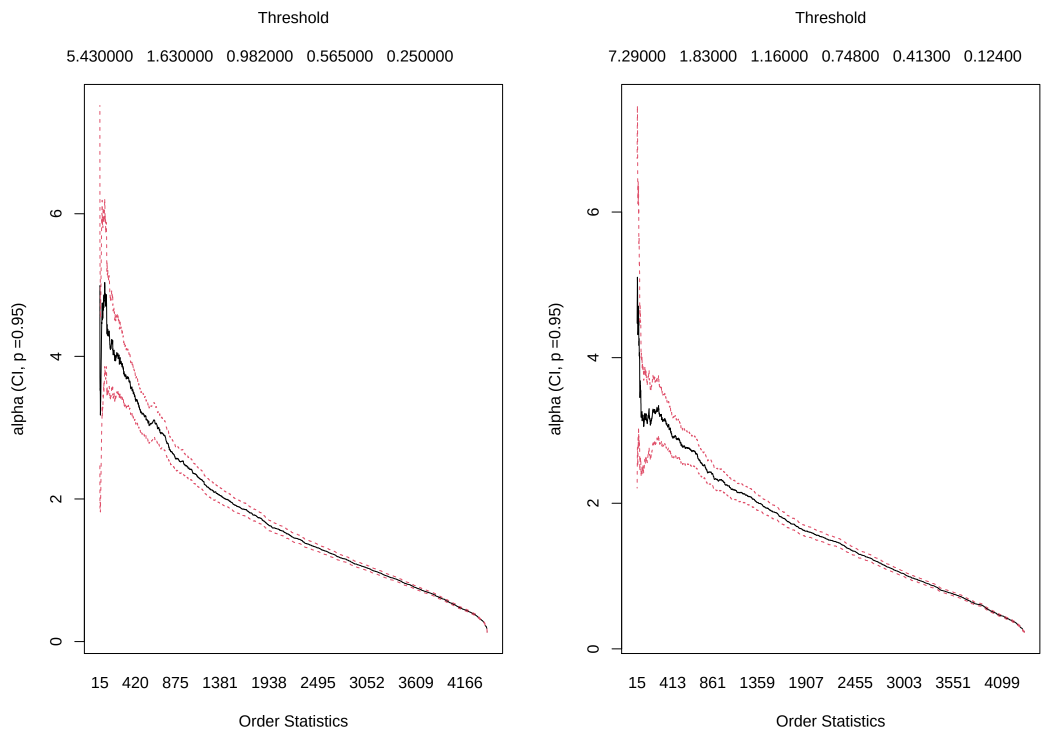

Figure 4: Hill plots of the absolute value of the Total return series (right graph) and of a simulation of an APARCH model with tail index 7.8 (left graph).

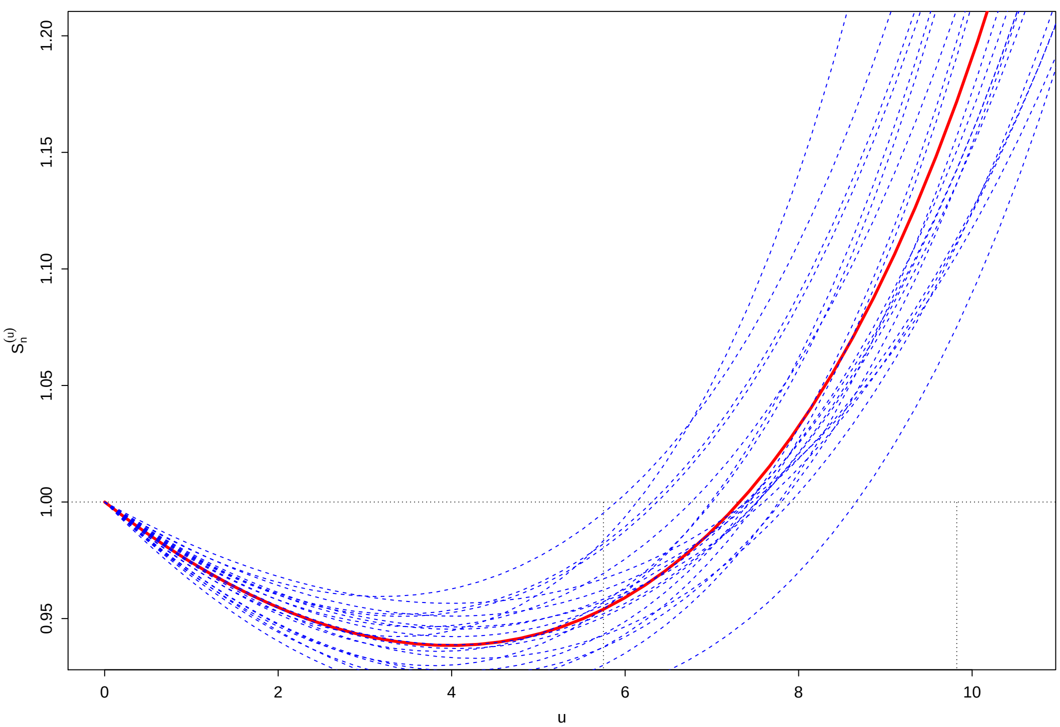

Figure 5: Empirical MGF for the APARCH(1,1) model fitted on the Total return series (red full line), MGF of 20 bootstrap replications (blue dotted line), and 95% bootstrap interval (delimited by vertical dotted lines) over 10000 bootstrap replications.

Appendix

A Assumptions

The next assumptions are used in the semi-parametric framework of Sections 2, 3 and 4.1.

A1: $E[\omega^{\varsigma}(\eta_1, \boldsymbol{\theta}_0)] < \infty$, $E \log a(\eta_1, \boldsymbol{\theta}_0) < 0$ and $E[a^s(\eta_1, \boldsymbol{\theta}_0)] < \infty$ for some $\varsigma > 0$ and $s > 0$.

A2: For any $\boldsymbol{\theta} \in \Theta$, there exists $z_0 > \underline{\omega}$ such that

$$E \log^+ \omega \left(\frac{\epsilon_t}{z_0^{1/\delta}}; \boldsymbol{\theta} \right) + \log^+ a \left(\frac{\epsilon_t}{z_0^{1/\delta}}; \boldsymbol{\theta} \right) < \infty, \; E \log \sup_{z \geq \underline{\omega}} \left| \frac{\partial}{\partial z} \left\{ \omega \left(\frac{\epsilon_t}{z^{1/\delta}}; \boldsymbol{\theta} \right) + a \left(\frac{\epsilon_t}{z^{1/\delta}}; \boldsymbol{\theta} \right) z \right\} \right| < 0.$$

A3: The $\mathcal{F}_{t-1}$-measurable function $\boldsymbol{\theta} \to (\sigma_t(\boldsymbol{\theta}), \tilde{\sigma}_t(\boldsymbol{\theta}))$ is a.s. twice continuously differentiable. Moreover, $\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |\sigma_t(\boldsymbol{\theta}) - \tilde{\sigma}_t(\boldsymbol{\theta})| + \left| \frac{\partial \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial \tilde{\sigma}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right| \leq K_t \rho^t$ where $K_t \in \mathcal{F}_{t-1}$ and $\sup_t E(K_t^r) < \infty$ for some $r > 0$.

A4: 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 $E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \|\partial \sigma_t(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}\|^r < \infty$.

A5: $\boldsymbol{\theta}_0$ belongs to the interior $\overset{\circ}{\Theta}$ of Θ 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, $\boldsymbol{V} : \mathbb{R} \mapsto \mathbb{R}^k$ 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 $E\boldsymbol{V}(\eta_t) = 0$, $\operatorname{var}\{\boldsymbol{V}(\eta_t)\} = \boldsymbol{\Upsilon}$ is nonsingular and $E\boldsymbol{\Delta}_t = \boldsymbol{\Lambda}$ is full row rank.

A6: For almost all ϵ, 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 $(\epsilon, \sigma, \boldsymbol{\theta}) \in \mathbb{R} \times [\underline{\omega}, +\infty) \times V(\boldsymbol{\theta}_0)$,

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

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

A7(u)**:** There exist $p, q > 0$ such that $\frac{1}{p} + \frac{2}{q} = 1$ and

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

Assumption **A1** ensures the existence of a strictly stationary and ergodic solution (ϵ_t) to Model (5), while **A2** ensures the existence of strictly stationary and ergodic solution to the SRE (6) by Lemma 1. Assumption **A3** is introduced to control the effect of the initial values on the statistics under study as the sample size increases. **A5** is a mild assumption which is fulfilled by commonly used estimators of volatility parameters, as illustrated in Corollary 1. The other assumptions reduce considerably for particular specifications of the MGF, see for instance Corollary 2.

The next assumptions are used for the fully-parametric framework of Section 4.2.

B1: $\lim_{|y|\to\infty} yf(y) = 0$ and $\lim_{|y|\to\infty} y^2 f'(y) = 0$.

B2: For some positive constants K and ς,

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

B3: The function $\boldsymbol{\theta} \mapsto \int a^{u_0}(x;\boldsymbol{\theta})f(x)dx$ is continuously differentiable under the integral sign.

B4: Letting $g_1(y) = 1 + y\frac{f'}{f}(y)$ we have

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

B5: The functions $\boldsymbol{\theta} \mapsto \int a^{u_0}(x;\boldsymbol{\theta})f(x,\boldsymbol{\nu})dx$, $\quad \boldsymbol{\nu} \mapsto \int a^{u_0}(x;\boldsymbol{\theta})f(x,\boldsymbol{\nu})dx$ are continuously differentiable under the integral signs.

B6: We have

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

where

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

B Examples of augmented GARCH models

C Conditions for the existence of a strictly stationary solution to the SRE (6)

Lemma 1 *Let (X_t) be a stationary and ergodic process. Suppose that for some differentiable functions $\omega : \mathbb{R} \to [\underline{\omega}, +\infty)$ and $a : \mathbb{R} \to [\underline{a}, +\infty)$, where $\underline{\omega} > 0$ and $\underline{a} \ge 0$, and for $\delta > 0$,*

$$E\log^+\omega\left(\frac{X_t}{z_0^{1/\delta}}\right) + E\log^+ a\left(\frac{X_t}{z_0^{1/\delta}}\right) < \infty, \quad E\log\sup_{z\ge\underline{\omega}}\left|\frac{\partial}{\partial z}\left\{\omega\left(\frac{X_t}{z^{1/\delta}}\right) + a\left(\frac{X_t}{z^{1/\delta}}\right)z\right\}\right| < 0,$$

Table 2: Examples of models satisfying (5) (with $x^+ = \max(x, 0), x^- = \max(-x, 0)$)

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] $\sigma_t = \omega + \alpha|\epsilon_{t-1}| + \beta\sigma_{t-1}$

[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] $\sigma_t^\delta = \omega + \alpha(|\epsilon_{t-1}| - \xi\epsilon_{t-1})^\delta + \beta\sigma_{t-1}^\delta$

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

for some $z_0 \geq \underline{\omega}$. Then there exists a stationary and ergodic solution (Z_t), $Z_t \in [\underline{\omega}, \infty)$, to the SRE

$$Z_t = \omega\left(\frac{X_{t-1}}{Z_{t-1}^{1/\delta}}\right) + a\left(\frac{X_{t-1}}{Z_{t-1}^{1/\delta}}\right)Z_{t-1}, \qquad t \in \mathbb{Z}.$$

D Proofs

D.1 Proof of Lemma 1

We provide a direct proof of this result, which could be obtained as a consequence of the much more general (but non-explicit) result of (Straumann and Mikosch (2006), Theorem 2.8). For all $n \in \mathbb{N}, n > 0$ let

$$Z_{t,n} = \omega\left(\frac{X_{t-1}}{Z_{t-1,n-1}^{1/\delta}}\right) + a\left(\frac{X_{t-1}}{Z_{t-1,n-1}^{1/\delta}}\right)Z_{t-1,n-1} := \varphi(X_{t-1}, Z_{t-1,n-1}), \qquad t \in \mathbb{Z},$$

where $Z_{t,0} = z_0$. For fixed n, the sequence $(Z_{t,n})_t$ is stationary and ergodic. By the mean-value theorem,

$$\sup_{z \neq \tilde{z}, z \wedge \tilde{z} \geq \underline{\omega}} \left|\frac{\varphi(X_t, z) - \varphi(X_t, \tilde{z})}{z - \tilde{z}}\right| \leq \Lambda_t := \sup_{z \in [\underline{\omega}, \infty)} \left|\frac{\partial\varphi(X_t, z)}{\partial z}\right|.$$

It follows that

$$|Z_{t,n} - Z_{t,n-1}| \leq \Lambda_{t-1}|Z_{t-1,n-1} - Z_{t-1,n-2}| \leq \Lambda_{t-1}\Lambda_{t-2}\ldots\Lambda_{t-n+1}|\varphi(X_{t-n}, z_0) - z_0|.$$

Thus, for $n < m$,

$$|Z_{t,m} - Z_{t,n}| \leq \sum_{k=0}^{m-n-1} |Z_{t,m-k} - Z_{t,m-k-1}|$$

$$\leq \sum_{k=0}^{m-n-1} \Lambda_{t-1} \ldots \Lambda_{t-m+k+1} |\varphi(X_{t-m+k}, z_0) - z_0|$$

$$\leq \sum_{j=n+1}^{\infty} \Lambda_{t-1} \ldots \Lambda_{t-j+1} |\varphi(X_{t-j}, z_0) - z_0| \to 0 \quad \text{a.s. as} \quad n \to \infty.$$

The latter convergence follows from the Cauchy rule applied to the infinite sum, using $E \log^+ \lambda_t < \infty$ and $E \log \lambda_t < 0$. We have shown that, a.s., $(Z_{t,n})_{n\in\mathbb{N}}$ is a Cauchy sequence. The conclusion follows by standard arguments.

D.2 Proof of Theorem 1

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

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

where $\tilde{\eta}_t(\theta) = \epsilon_t \tilde{\sigma}_t^{-1}(\boldsymbol{\theta})$.

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

$$\begin{aligned} \sqrt{n}\left\{S_n^{(u)} - S_\infty^{(u)}\right\} &= \sqrt{n}\left\{\tilde{S}_n^{(u)}(\widehat{\boldsymbol{\theta}}_n) - S_n^{(u)}(\widehat{\boldsymbol{\theta}}_n)\right\} + \left[\frac{\partial S_n^{(u)}(\boldsymbol{\theta}_n^*)}{\partial \boldsymbol{\theta}'} - \frac{\partial S_n^{(u)}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right] \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\ &+ \frac{\partial S_n^{(u)}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + \sqrt{n}\left\{S_n^{(u)}(\boldsymbol{\theta}_0) - S_\infty^{(u)}\right\}, \end{aligned} \tag{31}$$

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

Write $a(\frac{\epsilon}{\sigma}; \boldsymbol{\theta}) = b(\epsilon, \sigma; \boldsymbol{\theta})$ where $b : \mathbb{R} \times \mathbb{R}^+ \times \boldsymbol{\Theta} \mapsto \mathbb{R}^+$. Under **A6**, for such function b or $\log b$, ∇_σ (resp. $\nabla_{\boldsymbol{\theta}}$) denotes the partial derivative with respect to σ (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 σ (resp. σ and $\boldsymbol{\theta}$).[5] With this notation, we can write for instance

$$\frac{\partial}{\partial \boldsymbol{\theta}} a\left(\eta_t(\boldsymbol{\theta}); \boldsymbol{\theta}\right) = \frac{\partial}{\partial \boldsymbol{\theta}} b\left(\eta_t, \sigma_t(\boldsymbol{\theta}); \boldsymbol{\theta}\right) = \nabla_{\boldsymbol{\theta}} b\left(\epsilon_t, \sigma_t(\boldsymbol{\theta}); \boldsymbol{\theta}\right) + \nabla_\sigma b\left(\epsilon_t, \sigma_t(\boldsymbol{\theta}); \boldsymbol{\theta}\right) \frac{\partial}{\partial \boldsymbol{\theta}} \sigma_t(\boldsymbol{\theta}).$$

The proof of the theorem will be a consequence of the following lemmas whose proofs are provided below.

Lemma 2 *If the conditions of Theorem 1 are satisfied, then*

$$\sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} n \left| \tilde{S}_n^{(u)}(\boldsymbol{\theta}) - S_n^{(u)}(\boldsymbol{\theta}) \right| = O(1) \quad a.s. \tag{32}$$

[5] 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) case, 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 $\nabla_{\boldsymbol{\theta}} \log b(\epsilon, \sigma; \boldsymbol{\theta}) = \left(0, \frac{1}{\alpha}\right)'$.

Lemma 3 *If the conditions of Theorem 1 are satisfied, then*

$$\frac{\partial S_n^{(u)}(\boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta}} - \frac{\partial S_n^{(u)}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \to 0,$$

a.s. as $n \to \infty$, *for any sequence* $(\boldsymbol{\theta}_n)$ *such that* $\boldsymbol{\theta}_n \to \boldsymbol{\theta}_0$ *a.s.*

Lemma 4 *If the conditions of Theorem 1 are satisfied, then*

$$\frac{\partial S_n^{(u)}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \to E\left[\frac{\partial}{\partial \boldsymbol{\theta}} a^u\{\eta_t(\boldsymbol{\theta}); \boldsymbol{\theta}\}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \quad a.s. \text{ as } n \to \infty.$$

Now we complete the proof of Theorem 1. In view of (31), the strong consistency of $\widehat{\boldsymbol{\theta}}_n$ and the previous lemmas, we have

$$\sqrt{n}\left\{S_n^{(u)} - S_\infty^{(u)}\right\} = \sqrt{n}\left\{S_n^{(u)}(\boldsymbol{\theta}_0) - S_\infty^{(u)}\right\} + \boldsymbol{g}_u'\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_P(1). \tag{33}$$

The asymptotic distribution in (7) follows from **A5** and the CLT for stationary second-order martingale differences of (6).

D.3 Proof of lemma 2

A Taylor expansion shows that for $\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)$

$$\begin{aligned} a^u\{\tilde{\eta}_t(\boldsymbol{\theta}); \boldsymbol{\theta}\} - a^u\{\eta_t(\boldsymbol{\theta}); \boldsymbol{\theta}\} &= b^u\{\epsilon_t, \tilde{\sigma}_t(\boldsymbol{\theta}); \boldsymbol{\theta}\} - b^u\{\epsilon_t, \sigma_t(\boldsymbol{\theta}); \boldsymbol{\theta}\} \\ &= u b^u\{\epsilon_t, \sigma_t^*; \boldsymbol{\theta}\} \nabla_\sigma \log b(\epsilon_t, \sigma_t^*; \boldsymbol{\theta})\{\tilde{\sigma}_t(\boldsymbol{\theta}) - \sigma_t(\boldsymbol{\theta})\} \end{aligned} \tag{34}$$

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

Then, using **A3**, **A6** and the c_r inequality, we deduce

$$|a^u\{\tilde{\eta}_t(\boldsymbol{\theta}); \boldsymbol{\theta}\} - a^u\{\eta_t(\boldsymbol{\theta}); \boldsymbol{\theta}\}| \leq u 2^u C^{u+1}\left\{\left(\frac{|\epsilon_t|}{\sigma_t^*}\right)^{\tau(u+1)} + 1\right\} K_t \rho^t.$$

The r.h.s. of the above inequality is bounded by a variable of the form $X_t\rho^t$ where X_t admits a small moment, uniformly in t, using **A3-A4** and noting that

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

Thus

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

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

D.4 Proof of lemma 3

We have

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

From Hölder inequality and $\mathbf{A7}(u)$

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

By a Taylor expansion of $\frac{\partial S_n^{(u)}(\boldsymbol{\theta}_n)}{\partial\boldsymbol{\theta}}$ around $\boldsymbol{\theta}_0$, the conclusion follows.

D.5 Proof of lemma 4

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 Hölder inequality, $\mathbf{A7}(u)$ and the ergodic theorem.

D.6 Proof of Corollary 1

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 (2013b)), we obtain $\boldsymbol{g}_u'\boldsymbol{\xi}_u = 0$ and the formula for υ_u^2 follows.

ii) If the model is estimated by ML we have

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

where $g_1(y) = 1 + y\frac{f'}{f}(y)$. Noting that

$$Ea(\eta_t;\theta)g_1(\eta_t) = \alpha + \beta + \int(\alpha x^2 + \beta)xf'(x)dx = \alpha + \beta - \int(3\alpha x^2 + \beta)f(x)dx = -2\alpha,$$

we have, using $\boldsymbol{\Omega}'\boldsymbol{J}^{-1}\boldsymbol{\Omega} = 1$ (see Remark 3 in Francq and Zakoïan (2013b)) and $\boldsymbol{J}^{-1}\boldsymbol{\Omega} = (\omega_0, \alpha_0, 0)'$,

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

and

$$\boldsymbol{g}_u'\boldsymbol{\Sigma}\boldsymbol{g}_u = \frac{4u^2}{\iota_f}\left\{\boldsymbol{m}_u - \alpha_0 M_{1,u-1}\boldsymbol{\Omega}\right\}'\boldsymbol{J}^{-1}\left\{\boldsymbol{m}_u - \alpha_0 M_{1,u-1}\boldsymbol{\Omega}\right\} = \frac{4u^2}{\iota_f}\left\{\boldsymbol{m}_u'\boldsymbol{J}^{-1}\boldsymbol{m}_u - \alpha_0^2 M_{1,u-1}^2\right\}.$$

Thus

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

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

$$\boldsymbol{m}_u'\boldsymbol{J}^{-1}\boldsymbol{m}_u - \alpha_0^2 M_{1,u-1}^2 = \boldsymbol{m}_u'\boldsymbol{J}^{-1}\boldsymbol{m}_u - (\boldsymbol{m}_u'\boldsymbol{J}^{-1}\boldsymbol{\Omega})^2 \geq \boldsymbol{m}_u'\boldsymbol{J}^{-1}\boldsymbol{m}_u - (\boldsymbol{m}_u'\boldsymbol{J}^{-1}\boldsymbol{m}_u)(\boldsymbol{\Omega}'\boldsymbol{J}^{-1}\boldsymbol{\Omega}) = 0.$$

D.7 Proof of Corollary 2

The consistency and asymptotic normality of the QMLE were established by Hamadeh and Zakoïan (2011). The fact that Assumptions **A1** and **A3-A5** hold true can be found in the proof of their Theorems 2.1 and 2.2. In particular, Assumption **A5** holds with

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

where $\boldsymbol{\Omega}_\delta = E(\boldsymbol{D}_t)$, $\boldsymbol{J}_\delta = E(\boldsymbol{D}_t\boldsymbol{D}_t')$ with $\boldsymbol{D}_t = \boldsymbol{D}_t(\boldsymbol{\theta}_0)$ and $\boldsymbol{D}_t(\boldsymbol{\theta}) = \sigma_t^{-\delta}(\boldsymbol{\theta})\partial\sigma_t^\delta(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. Noting that the strictly stationary solution admits a small-order moment, and that the derivative in **A2** is equal to β, this assumption is obviously satisfied. Hamadeh and Zakoian showed 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). Noting that $b(\epsilon, \sigma, \boldsymbol{\theta}) = \frac{|\epsilon|^\delta}{\sigma^\delta}(\alpha_+\mathbb{1}_{\epsilon>0} + \alpha_-\mathbb{1}_{\epsilon<0}) + \beta$ the last two conditions of $\mathbf{A7}(u)$ are thus satisfied for any $q > 0$, while the first condition is satisfied for p close enough to 1 since $u < s$. Assumption **A6** is satisfied for $\tau = \delta$.

D.8 Proof of Proposition 1

Noting that $\hat{\boldsymbol{g}}_u = \frac{\partial \tilde{S}_n^{(u)}(\widehat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}}$, Lemmas 3, 4 and 5 below, together with the consistency of $\widehat{\boldsymbol{\theta}}_n$, entail that $\hat{\boldsymbol{g}}_u$ is a consistent estimator of $\boldsymbol{g}_u$.

Lemma 5 *If the conditions of Theorem 1 are satisfied, then*

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

The estimator of ψ_u can be handled similarly. Using also the consistency of $\hat{\xi}_u$ and $\hat{\boldsymbol{\Sigma}}$, it follows that $\hat{v}_u$ is a consistent estimator of v_u. Now we have

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

which tends to $\underline{\alpha}$ as $n \to \infty$ by Theorem 1.

D.9 Proof of lemma 5

We have

$$\begin{aligned}
& \frac{\partial S_n^{(u)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial \tilde{S}_n^{(u)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\
= \quad & \frac{1}{n}\sum_{t=1}^n u\left\{b^u(\epsilon_t, \sigma_t(\boldsymbol{\theta}); \boldsymbol{\theta}) - b^u(\epsilon_t, \tilde{\sigma}_t(\boldsymbol{\theta}); \boldsymbol{\theta})\right\}\left\{\nabla_\sigma \log b(\epsilon_t, \sigma_t(\boldsymbol{\theta}); \boldsymbol{\theta})\frac{\partial \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \nabla_{\boldsymbol{\theta}} \log b(\epsilon_t, \sigma_t(\boldsymbol{\theta}); \boldsymbol{\theta})\right\} \\
& + \frac{1}{n}\sum_{t=1}^n u b^u(\epsilon_t, \tilde{\sigma}_t(\boldsymbol{\theta}); \boldsymbol{\theta})\left\{\nabla_\sigma \log b(\epsilon_t, \sigma_t(\boldsymbol{\theta}); \boldsymbol{\theta}) - \nabla_\sigma \log b(\epsilon_t, \tilde{\sigma}_t(\boldsymbol{\theta}); \boldsymbol{\theta})\right\}\frac{\partial \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\
& + \frac{1}{n}\sum_{t=1}^n u b^u(\epsilon_t, \tilde{\sigma}_t(\boldsymbol{\theta}); \boldsymbol{\theta})\nabla_\sigma \log b(\epsilon_t, \tilde{\sigma}_t(\boldsymbol{\theta}); \boldsymbol{\theta})\left(\frac{\partial \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial \tilde{\sigma}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) \\
& + \frac{1}{n}\sum_{t=1}^n u b^u(\epsilon_t, \tilde{\sigma}_t(\boldsymbol{\theta}); \boldsymbol{\theta})\left\{\nabla_{\boldsymbol{\theta}} \log b(\epsilon_t, \sigma_t(\boldsymbol{\theta}); \boldsymbol{\theta}) - \nabla_{\boldsymbol{\theta}} \log b(\epsilon_t, \tilde{\sigma}_t(\boldsymbol{\theta}); \boldsymbol{\theta})\right\} \\
:= \quad & \Delta_{1n}(\boldsymbol{\theta}) + \Delta_{2n}(\boldsymbol{\theta}) + \Delta_{3n}(\boldsymbol{\theta}) + \Delta_{4n}(\boldsymbol{\theta}).
\end{aligned}$$

First consider $\Delta_{1n}(\boldsymbol{\theta})$. By the proof of Lemma 2 we have

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

where X_t admits a small moment. By **A4** and **A6**, the other summands involved in Δ_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 Δ_{2n}. Another Taylor expansion yields

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

where σ_t^* is between $\tilde{\sigma}_t(\boldsymbol{\theta})$ and $\sigma_t(\boldsymbol{\theta})$. The same arguments show that $\sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} |\Delta_{2n}(\boldsymbol{\theta})| \to 0$, in probability as $n \to \infty$. The last two terms can be handled similarly.

D.10 Proof of Proposition 2

The proof is given for convenience, but very similar results have been established in the articles already cited.

The i) is obvious in view of (2): the condition $P[a(\eta_1) \leq 1] = 1$ entails $E[a^u(\eta_1)] \leq 1$ and the inequality is strict because $\gamma < 0$.
Now suppose $P[a(\eta_1) \leq 1] < 1$ and let $\epsilon > 0$ such that $P[a(\eta_1) > 1 + \epsilon] > 0$. Then $S_\infty^{(u)} = E[a^u(\eta_1)] > (1+\epsilon)^u P[a(\eta_1) > 1 + \epsilon] \to \infty$ as $u \to \infty$. For any $\eta > 0$, the function $u \mapsto a^u(\eta)$ is convex. Thus $u \mapsto E[a^u(\eta_1)]$ is convex on $(0, s]$. We consider two cases: a) when $P[a(\eta_1) = 0] = p > 0$ we have $S_\infty^{(0^+)} = 1 - p < S_\infty^{(0)} = 1$. In view of the convexity and the fact $S_\infty^{(s)} \geq 1$, the conclusion follows; b) when $P[a(\eta_1) = 0] = 0$, the right derivative of $u \mapsto S_\infty^{(u)}$ in a neighborhood of 0 is negative. Thus there exists $0 < s_0 < s$ such that the function $u \mapsto E[a^u(\eta_1)]$ decreases over $(0, s_0)$ and increases over $(s_0, s]$. Since $S_\infty^{(s)} \geq 1$, it follows that there is the unique $u > 0$ such that $E[a^u(\eta_1)] = 1$. Finally, by (2), moments of σ_t do not exist at any order: $E(\sigma_t^{u\delta}) < \infty$ for all $u < u_0$ such that $E[\omega^u(\eta_1)] < \infty$, and $E(\sigma_t^{u\delta}) = \infty$ for $u > u_0$. The proof of ii) follows.

The tail result on σ_t is established using Theorem 4.1 in (21). The tail result on ϵ_t follows by the arguments given by Mikosch and Stărică (2000) in the proof of their Theorem 2.1.

D.11 Proof of Proposition 3

We apply Proposition 2 substituting the empirical distribution on $\{a(\hat{\eta}_t; \widehat{\boldsymbol{\theta}}_n) : t = 1, \ldots, n\}$ for the theoretical distribution of $a(\eta_1)$. The condition on the existence of $s > 0$ vanishes because moments exist at any order for the empirical distribution.

D.12 Proof of Theorem 2

Similar to (34) we have

$$\log b(\epsilon_t, \sigma_t(\boldsymbol{\theta}); \boldsymbol{\theta}) - \log b(\epsilon_t, \tilde{\sigma}_t(\boldsymbol{\theta}); \boldsymbol{\theta}) \quad = \quad \nabla_\sigma \log b(\epsilon_t, \sigma_t^*; \boldsymbol{\theta})\{\tilde{\sigma}_t(\boldsymbol{\theta}) - \sigma_t(\boldsymbol{\theta})\}, \qquad (35)$$

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

$$\log b(\epsilon_t, \sigma_t(\boldsymbol{\theta}); \boldsymbol{\theta}) - \log b(\epsilon_t, \sigma_t(\boldsymbol{\theta}_0); \boldsymbol{\theta}_0) \quad = \quad \frac{\partial}{\partial \boldsymbol{\theta}'} \log b(\epsilon_t, \sigma_t(\boldsymbol{\theta}^*); \boldsymbol{\theta}^*)(\boldsymbol{\theta} - \boldsymbol{\theta}_0), \qquad (36)$$

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

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

The a.s. convergence of γ_n follows by the ergodic theorem.

By the same arguments, and those of the proofs of Lemmas 2, 3 and 4,

$$S_n^{(u)} \to S_\infty^{(u)} \quad a.s., \quad \text{for any } u \text{ such that} \quad S_\infty^{(u)} < \infty. \tag{37}$$

Now, we turn to case i). We have $S_\infty^{(u)} < 1$ by Proposition 2, thus $S_n^{(u)} < 1$ for n large enough by (37). It follows, by Proposition 3, that $\hat{u}_n > u$ for all u and n large enough.

Turning to case ii), we note that, for $\varepsilon \in (0, \max\{u_0, s - u_o\})$,

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

thus the consistency of $\hat{u}_n$.

D.13 Proof of Theorem 3

Let $\Gamma_n(u) = \sqrt{n}\left\{S_n^{(u)} - S_\infty^{(u)}\right\}$ and, in view of (33), let

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

We will show that

$$\Gamma_n^0 \overset{\mathcal{C}[u_1,u_2]}{\Longrightarrow} \Gamma, \qquad \sup_{u\in(u_1,u_2)} |\Gamma_n(u) - \Gamma_n^0(u)| = o_P(1). \tag{38}$$

By the Cramér-Wold device, and by arguments used in the proof of Theorem 1, it can be established that the finite-dimensional distributions of Γ_n^0 converge to those of Γ. By showing that

$$\text{the sequence} \quad \{\Gamma_n^0(u_1)\} \quad \text{is tight} \tag{39}$$

and, for some constant $K > 0$,

$$E[\Gamma_n^0(u) - \Gamma_n^0(v)]^2 \le K(u-v)^2, \tag{40}$$

the tightness of the sequence $\{\Gamma_n^0\}$ will be established, according to Theorem 12.3 of Billingsley (1968). The weak convergence in (38) will follow from Theorem 8.1 of Billingsley (1968).

The convergence in distribution of $\{\Gamma_n^0(u_1)\}$ entails (39).

We have

$$\Gamma_n^0(u) - \Gamma_n^0(v) \quad = \quad \frac{1}{\sqrt{n}}\sum_{t=1}^{n} a^u\{\eta_t; \boldsymbol{\theta}_0\} - E[a^u\{\eta_t; \boldsymbol{\theta}_0\}] - a^v\{\eta_t; \boldsymbol{\theta}_0\} + E[a^v\{\eta_t; \boldsymbol{\theta}_0\}]$$

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

Note that

$$
\begin{aligned}
E\Delta_{n,1}^2(u,v) &= \operatorname{Var}\{a^u(\eta_t)-a^v(\eta_t)\}\\
&\leq (u-v)^2E\left(\{a^{2u_1}(\eta_t)+a^{2u_2}(\eta_t)\}\{\log a(\eta_t)\}^2\right)\leq K(u-v)^2.
\end{aligned}
$$

Now

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

where the u_i^* belong to (u,v) and the existence of the expectation follows from $\mathbf{A7}(u_2)$. Moreover $\operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n\boldsymbol{\Delta}_{t-1}\boldsymbol{V}(\eta_t)\right)=\boldsymbol{\Sigma}$. It follows that $E\Delta_{n,2}^2(u,v)\leq K(u-v)^2$. This completes the proof of the weak convergence in (38).

We will now show the convergence in probability in (38). We have, for $\boldsymbol{\theta}^*$ between $\widehat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$,

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

We have, by **A3**, with $\sigma_t^*(\boldsymbol{\theta})$ between $\tilde{\sigma}_t(\boldsymbol{\theta})$ and $\sigma_t(\boldsymbol{\theta})$

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

uniformly in u, noting that, by **A6**, the supremum admits a small-order moment. The second term, $R_{n,2}(u)$, can be handled by a Taylor expansion of $\frac{\partial}{\partial\boldsymbol{\theta}}b^u\{\epsilon_t,\sigma_t(\boldsymbol{\theta}^*);\boldsymbol{\theta}^*\}$ around $\boldsymbol{\theta}_0$. The third term, $R_{n,3}(u)$, is an $o_P(1)$ uniformly in u by **A5** and using the fact that $\sup_{u\in(u_1,u_2)}\|\boldsymbol{g}_u\|<\infty$.

D.14 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 $\Gamma_n(\hat{u}_n) \overset{\mathcal{L}}{\to} \Gamma(u_0)$. Indeed, for $\varsigma, \varsigma' > 0$,

$$\begin{aligned} P[\Gamma_n(\hat{u}_n) > x] &\leq P[\Gamma_n(\hat{u}_n) > x, |\hat{u}_n - u_0| \leq \varsigma] + P[|\hat{u}_n - u_0| > \varsigma] \\ &\leq P[\sup_{|u-u_0|\leq\varsigma} \Gamma_n(u) > x] + P[|\hat{u}_n - u_0| > \varsigma] \\ &\leq P[\Gamma_n(u_0) > x - \varsigma'] + P[\sup_{|u-u_0|\leq\varsigma} |\Gamma_n(u) - \Gamma_n(u_0)| > \varsigma'] + P[|\hat{u}_n - u_0| > \varsigma]. \end{aligned}$$

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

D.15 Proof of Proposition 4

The arguments are the same as in the proof of Proposition 1, using the asymptotic normality of $\sqrt{n}(\hat{u}_n - u_0)$ established in Theorem 4.

D.16 Proof of Proposition 5

By the delta method we have,

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

In view of **B4** and the consistency of $\hat{\sigma}_f$, we deduce

$$V_n^{(u_0)} = \frac{\sqrt{n}(u_0 - \hat{u}_{n,f})}{\hat{\sigma}_f} = \frac{-2}{\sigma_f \iota_f \sqrt{n}} \sum_{t=1}^{n} \frac{\partial u_0}{\partial \boldsymbol{\theta}'} \boldsymbol{J}^{-1} \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}} g_1(\eta_t) + o_P(1), \tag{41}$$

from which the conclusion follows.

D.17 Proof of Proposition 6

The proof is similar to that of Proposition 5, relying on **B6** and the Taylor expansion

$$\begin{aligned} \frac{\sqrt{n}(u_0 - \hat{u}_{0,\hat{f}})}{\hat{\varsigma}_f} &= \frac{-1}{\varsigma_f}\left(\frac{\partial u_0}{\partial \boldsymbol{\theta}'}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + \frac{\partial u_0}{\partial \boldsymbol{\nu}'}\sqrt{n}(\widehat{\boldsymbol{\nu}}_n - \boldsymbol{\nu}_0)\right) + o_P(1) \\ &= \frac{-1}{\varsigma_f}\begin{bmatrix} \frac{\partial u_0}{\partial \boldsymbol{\theta}'} & \frac{\partial u_0}{\partial \boldsymbol{\nu}'} \end{bmatrix} \mathfrak{J}^{-1} \begin{pmatrix} \frac{1}{2\sqrt{n}} \sum_{t=1}^n \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}} g_1(\eta_t) \\ \frac{-1}{\sqrt{n}} \sum_{t=1}^n \frac{1}{f(\eta_t;\boldsymbol{\nu}_0)} \frac{\partial f(\eta_t;\boldsymbol{\nu}_0)}{\partial \boldsymbol{\nu}} \end{pmatrix} + o_P(1). \end{aligned} \tag{42}$$

D.18 Proof of Proposition 7 and inequality (24)

In the proof of Theorem 1 we have seen that

$$T_n^{(u)} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{a^u(\eta_t) - 1}{\upsilon_u} + \boldsymbol{g}_u' \frac{1}{\upsilon_u \sqrt{n}} \sum_{t=1}^{n} \boldsymbol{\Delta}_{t-1} \boldsymbol{V}(\eta_t) + o_P(1), \tag{43}$$

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

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

Le Cam's third lemma (see *e.g.* van der Vaart, 1998, page 90) shows that

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

The conclusion of Proposition 7 easily follows for the two tests using (20).

With the notations used in the proof of Corollary 1, for the standard GARCH(1,1) 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

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

Moreover,

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

Thus, in the standard GARCH(1,1) case,

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

where the formulas for υ_{u_0} are displayed in (9) for the ML and QML estimators.

D.19 Proof of Proposition 8

Relation (22) implies that

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

which is the distribution of the log-likelihood ratio in the statistical model $\mathcal{N}\left(\boldsymbol{\tau},\mathfrak{J}_f^{-1}\right)$ 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 $\{L_{n,f}(\boldsymbol{\theta}_0+\boldsymbol{\tau}/\sqrt{n}),\boldsymbol{\tau}\in\mathcal{T}\}$ converge to the gaussian experiment $\left\{\mathcal{N}\left(\boldsymbol{\tau},\mathfrak{J}_f^{-1}\right),\boldsymbol{\tau}\in\mathcal{T}\right\}$.

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

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

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

For $\boldsymbol{\tau}=\varepsilon\boldsymbol{e}_0$, $\varepsilon>0$, in the case of the standard GARCH(1,1),

$$\begin{aligned} c_{f,1}^{QML}(\boldsymbol{\theta}_0) =& \frac{\varepsilon\boldsymbol{e}_0'\boldsymbol{e}_0}{\sqrt{(\kappa_4-1)e_0'\boldsymbol{J}^{-1}e_0}} \leq c_{f,1}^{ML}(\boldsymbol{\theta}_0) = \frac{\varepsilon\boldsymbol{e}_0'\boldsymbol{e}_0}{\sqrt{\frac{4}{\iota_f}e_0'\boldsymbol{J}^{-1}e_0+\alpha_0^2\left(\kappa_4-1-\frac{4}{\iota_f}\right)}} \\ & \leq c_\varepsilon=\varepsilon\frac{\iota_f^{1/2}\boldsymbol{e}_0'\boldsymbol{e}_0}{2\sqrt{\boldsymbol{e}_0'\boldsymbol{J}^{-1}\boldsymbol{e}_0}}, \end{aligned}$$

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 η_t has the form (26) (see Francq and Zakoïan (2013a), Proposition 5.5)).

D.20 Proof of Proposition 9

By the arguments of the proof of Proposition 7, using (41), we obtain

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

D.21 Proof of Proposition 10

Follows by the arguments of the proof of Proposition 7, using (42) and the LAN property (29)-(30).

D.22 Proof of Proposition 11

The strong consistency of $\widehat{\boldsymbol{\vartheta}}_n^{QML}$ follows from Theorem 3.1 in Hamadeh and Zakoïan (2011). 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 Theorems 1 and 3 are satisfied for n large enough when δ is replaced by $\widehat{\delta}_n^{QML}$. If η_t has a positive density over the real line, Assumption ii) of Theorem 2 holds and the conclusion follows.

E Complement to Section 5.2

Examples of asymptotic slopes in the standard GARCH(1,1) case with Student errors are displayed in Figure 6.

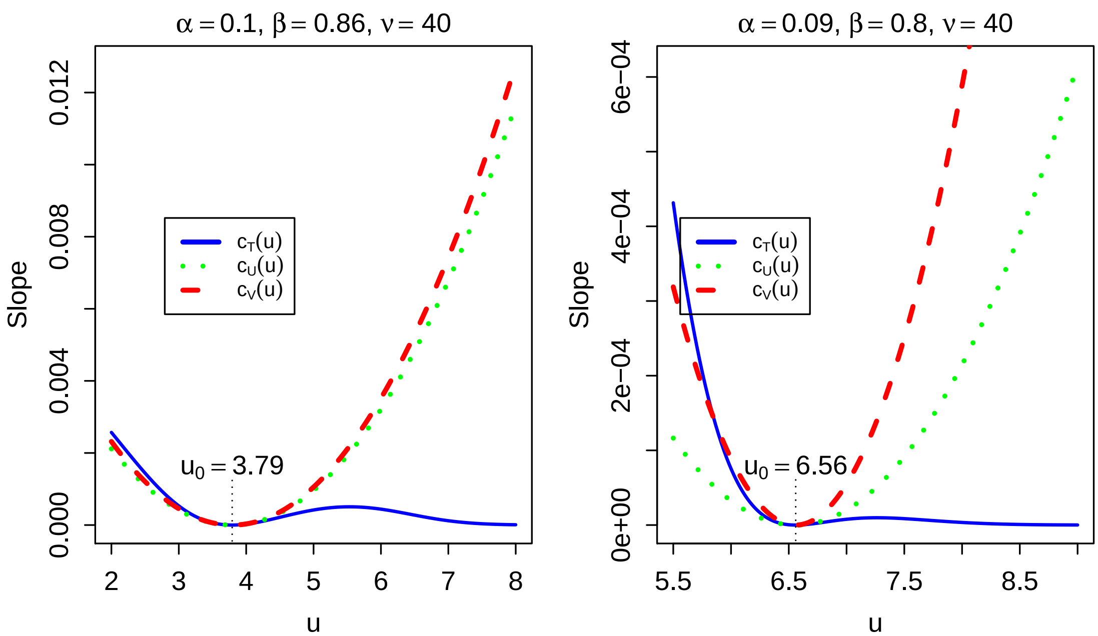

Figure 6: Asymptotic slopes of the tests T, U and V for Student errors ($\nu = 40$) and the standard GARCH(1,1) models.

F Monte Carlo experiments

We first made 10,000 simulations of a standard GARCH(1,1) 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 output is the lack of power of the test T, compared to its competitors, in agreement with Figure 3. Even for large sample sizes, the test T is too conservative but the levels of the tests U and V at the boundary of the null are correct. As expected the test V is slightly more powerful than the test U. The CI based on the statistics $\widehat{u}_n$ and $\widehat{u}_{n,f}$ (lines $U_n^{(u)}$ and $V_n^{(u)}$) are similar and, as expected, slightly tighter with the fully parametric method (that based on $\widehat{u}_{n,f}$ with f Gaussian). Note that the coverage probabilities are excellent (*i.e.* very close to nominal level $1-\alpha$) when $n = 4000$ or $n = 8000$. Results displayed in appendix concern the test of $\boldsymbol{H}_{0,u}^*$, for the same experiments. In agreement with Figure 3 these results are more favorable to the test T, even if the level is poorly controlled.

Next, we consider the class of the Beta-t-GARCH introduced in Harvey (2013) and Creal et al. (2013), 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 degree of freedom ν. This model is of the form (1) with $\delta = 2$, $\omega(\eta) = \omega$ and $a(\eta) = \beta + \frac{\alpha(\nu+1)\eta^2}{(\nu-2)+\eta^2}$. Results obtained for simulations of a Beta-t-GARCH model lead to similar conclusions (see the appendix).

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

n	α	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

The results reported in Table 4 concern the test of $\boldsymbol{H}_{0,u}^*$, for the same experiments as in Table 3.

The results reported in Tables 5 and 6 are obtained for simulations of the Beta-t-GARCH model:

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

with errors density

$$f(y) = \frac{1}{\sqrt{(\nu-2)\pi}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{y^2}{\nu-2}\right)^{-\frac{\nu+1}{2}},$$

with $\nu > 2$ and $\theta = (\omega, \alpha, \beta, \nu)'$ which belongs to the parameter space Θ, a subset of $(\underline{\omega}, \infty)^2 \times [0,1) \times (2, \infty)$ for some $\underline{\omega} > 0$. For the chosen parameters we have $u_0 = 3.5$. 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. The conclusions are similar to those drawn for Tables 3 and 4.

G A complement to Section 7

The QMLE and MLE residuals of the Total return series do not show any sign of dependence (on Figure 7, 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.

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 Processes and their Applications* **99** 95–116.

[4] 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.

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

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

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

n	α	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: As 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	α	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: As Table 5, but for the null $\boldsymbol{H}^*_{0,u}$

n	α	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

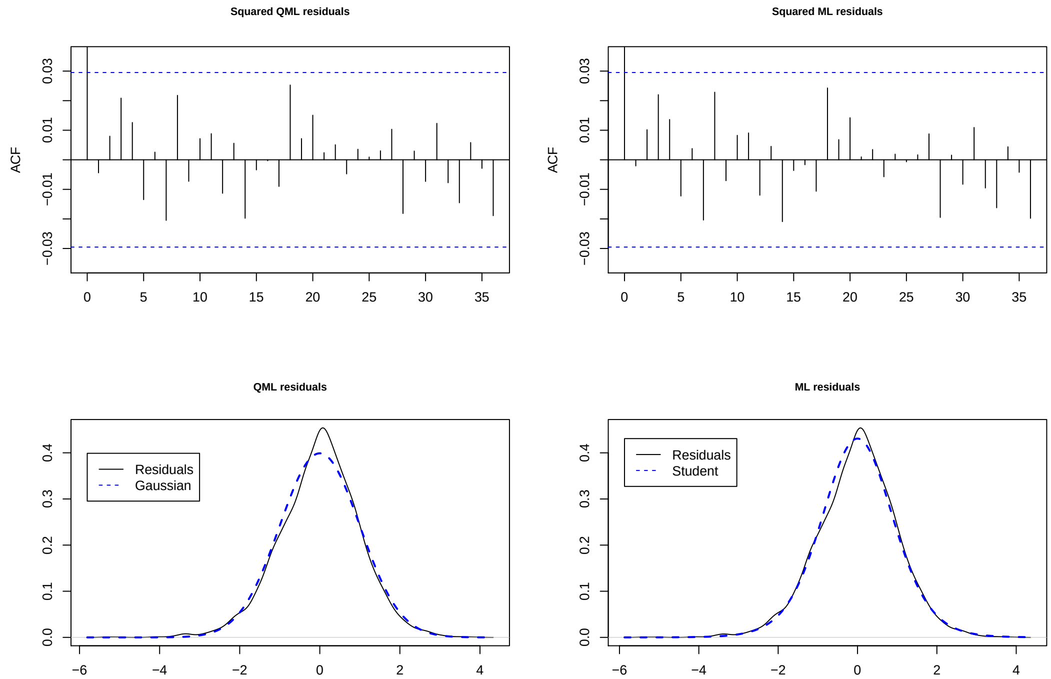

Figure 7: Autocorrelations of the squares of the QML and ML residuals, and empirical distributions of the QML and ML residuals, after fitting an APARCH on the Total return series.

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

[8] 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.

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

[10] 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.

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

[12] 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.

[13] Drees, H. (2000) Weighted approximations of tail processes for β-mixing random variables. *Annals of Applied Probability* **10** 1274–1301.

[14] Drees, H., Resnick, S., and L. de Haan (2000) How to make a Hill plot. *The Annals of Statistics* **28** 254–274.

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

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

[17] Francq, C. and J-M. Zakoïan (2013a) Inference in nonstationary asymmetric GARCH models. *The Annals of Statistics* **41** 1970–1998.

[18] Francq, C. and J-M. Zakoïan (2013b) Optimal predictions of powers of conditionally heteroskedastic processes. *Journal of the Royal Statistical Society - Series B* **75** 345–367.

[19] Francq, C. and J-M. Zakoïan (2021a) Testing the existence of moments for GARCH processes. *Journal of Econometrics*, forthcoming.

[20] Francq, C. and J-M. Zakoïan (2021b) LAN of GARCH-type and score-driven time-series models. *Discussion paper.*

[21] Goldie, C.M. (1991) Implicit renewal theory and tails of solutions of random equations. *The Annals of Applied Probability* **1** 126–166.

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

[23] Harvey, A.(2013) *Dynamic models for volatility and heavy tails.* Cambridge University Press.

[24] He, C. and T. Teräsvirta (1999) Properties of moments of a family of GARCH processes. *Journal of Econometrics* **92** 173–192.

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

[26] Hill, J.B. (2015) Tail index estimation for a filtered dependent time series. *Statistica Sinica* **25**, 609–629.

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

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

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

[30] Ling, S. and M. McAleer (2003) Adaptative estimation in nonstationary ARMA models with GARCH errors. *The Annals of Statistics* **31** 642–674.

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

[32] Resnick, S.I. and C. Stărică (1998) Tail index estimation for dependent data. *Annals of Applied Probability* **8** 1156–1183.

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

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

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

[36] 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 Reviews* **38** 151–169.

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

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