Conditional asymmetry in Power ARCH(∞) models

Julien Royer[*†1]

[1]CREST, ENSAE, Institut Polytechnique de Paris

July 2021

Abstract

We consider an extension of ARCH(∞) models to account for conditional asymmetry in the presence of high persistence. After stating existence and stationarity conditions, this paper develops the statistical inference of such models and proves the consistency and asymptotic distribution of a Quasi Maximum Likelihood estimator. Some particular specifications are studied and we introduce a Portmanteau test of goodness-of-fit. In addition, test procedures for asymmetry and GARCH validity are derived. Finally, we present an application on a set of equity indices to reexamine the preeminence of GARCH(1,1) specifications. We find strong evidence that the short memory feature of such models is not suitable for peripheral assets.

Keywords: Quasi Maximum Likelihood Estimation, Moderate memory, Testing parameters on the boundary, Recursive design bootstrap

JEL classification: C22, C51, C58

[*]5 Avenue Henri Le Chatelier, 91120 Palaiseau, France; E-mail: julien.royer@ensae.fr

[†]The author thanks Christian Francq and Jean-Michel Zakoïan for their guidance and feedback as well as participants of the CIRM Meeting on "New Results on Time Series and their Statistical Applications", and the ICEEE 2021 meeting for their comments.

Introduction

Despite their tremendous success in the financial and econometric literature, standard GARCH models are inappropriate for capturing strong volatility persistence. In practice, autocorrelations of squared returns often decay slowly, a property hardly compatible with the exponential decrease of squared GARCH processes. This motivated the introduction of ARCH(∞) by Robinson[36], providing series with longer memory than the classical GARCH specifications. In these models, financial returns (ε_t) and volatilities write as

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t, & (\eta_t) \text{ iid } (0,1) \\ \sigma_t^2 = \omega + \sum\limits_{i=1}^{\infty} \alpha_i \varepsilon_{t-i}^2 \end{cases} \tag{1}$$

with $\omega > 0$, and α_i, $i = 1, ...,$ a sequence of nonnegative constants. The existence of a strictly stationary and nonanticipative solution has been proved by Giraitis, Kokoszka and Leipus[21], Kazakevičius and Leipus[31], and Douc, Roueff and Soulier[14] under the condition

$$A_s \mu_{2s} < 1 \tag{2}$$

for some $s \in (0, 1]$, where $A_s = \sum\limits_{i=1}^{\infty} \alpha_i^s$ and $\mu_{2s} = E|\eta_t|^{2s}$. Condition (2) entails summability of the autocovariances of ε_t^2 and thus is not compatible with the usual concept of long memory (see Giraitis, Kokoszka and Leipus[21] and Zaffaroni[38]). However, this condition is compatible with a slow decay of the autocorrelation function of the squares and is sometimes referred to as moderate memory models[1].

Another well-documented empirical fact concerning stock returns is the leverage effect, namely the higher impact on the current volatility of past price decreases rather than increases of the same magnitude. To the best of our knowledge, attempts to capture both the asymmetry and the memory properties of financial time series have been scarce. A noticeable exception is the fractionally integrated EGARCH model introduced by Bollerslev and Mikkelsen[9]. The estimation of such models has been particularly studied by Zaffaroni[39].

Although long or moderate memory models are suitable candidates to model financial time series, their use amongst practitioners has been regrettably limited. The aim of this paper is to buttress the use of ARCH(∞) models. Our work is organized as follows. In Section 1, we introduce a new specification for ARCH(∞) models aiming at capturing

[1] The question of existence of a stationary solution with a finite fourth order moment for the FI-GARCH model has given rise to a long academic discussion, until the paper by Giraitis, Surgailis and Škarnulis[23] in which the existence of such solution was established.

the possible asymmetry and memory effect in financial returns. We establish a condition for the existence of a stationary solution. In Section 2, we focus on statistical inference. We prove the strong consistency of the quasi maximum likelihood estimator (QMLE) and derive its asymptotic distribution, allowing the parameter to belong to the frontier of the parameter space. In Section 3, we focus on hypothesis testing. We establish the asymptotic distribution of a Portmanteau statistic to test for the goodness-of-fit of our model. In addition, we design procedures to test for asymmetry and the adequacy of GARCH(1,1)-type specifications. Monte Carlo experiments are conducted in Section 4. Section 5 presents an application on a wide set of equity indices to reexamine the preeminence of GARCH(1,1)-type models. Finally, Section 6 concludes. Proofs and technical results are relegated to an appendix. Additional simulations and applications are available in a supplementary file.

1 Asymmetric Power ARCH(∞) model

Modeling asymmetry has led to the introduction of numerous specifications of the conditional volatility process. Among them, a popular and very general class of models is the Asymmetric Power GARCH (APARCH(p,q)) of Ding, Granger and Engle[13], defined by

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \\ \sigma_t^\delta = \omega + \sum_{i=1}^{q} \alpha_i^+ |\varepsilon_{t-i}|^\delta \mathbb{1}_{\varepsilon_{t-i}\geq 0} + \alpha_i^- |\varepsilon_{t-i}|^\delta \mathbb{1}_{\varepsilon_{t-i}<0} + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\delta \end{cases}$$

where $\omega > 0$, the coefficients are nonnegative constants, and δ is a positive constant. We propose an ARCH(∞) extension of this model defined as follows.

Definition 1. *Let (η_t) be an iid sequence of random variables such that $\mathbb{E}\eta_0 = 0$ and $\mathbb{E}\eta_0^2 = 1$. Then, (ε_t) is called an APARCH(∞) process if it satisfies an equation of the form*

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \\ \sigma_t^\delta = \omega + \sum_{i=1}^{\infty} \alpha_i^+ |\varepsilon_{t-i}|^\delta \mathbb{1}_{\varepsilon_{t-i}\geq 0} + \alpha_i^- |\varepsilon_{t-i}|^\delta \mathbb{1}_{\varepsilon_{t-i}<0} \end{cases} \tag{3}$$

with $\omega > 0$, $\delta > 0$, and where α_i^+ and α_i^-, $i = 1, \ldots$, are sequences of nonnegative constants.

Note that this specification is very general and includes standard ARCH(∞) as well as the Threshold-ARCH(∞) model which corresponds to $\delta = 2$. TARCH(∞) models were first considered by Bardet and Wintenberger[2] as a particular example of a more general causal process.

The following theorem gives a condition for the existence of a strictly stationary and nonanticipative solution to an APARCH(∞) model defined by (3). For any $s > 0$, let

$$A_s^+ = \sum_{i=1}^{\infty} \left(\alpha_i^+\right)^s \ , \ A_s^- = \sum_{i=1}^{\infty} \left(\alpha_i^-\right)^s \quad \text{and} \quad \mu_{\delta s}^+ = \mathbb{E}|\mathbb{1}_{\eta_t \geq 0} \ \eta_t|^{\delta s} \ , \ \mu_{\delta s}^- = \mathbb{E}|\mathbb{1}_{\eta_t < 0} \ \eta_t|^{\delta s}.$$

Theorem 1. *If there exists $s \in (0,1]$ such that*

$$A_s^+ \mu_{\delta s}^+ + A_s^- \mu_{\delta s}^- < 1, \tag{4}$$

there exists a unique, strictly stationary, ergodic, and nonanticipative solution of (3) such that $\mathbb{E}|\varepsilon_t|^{\delta s} < \infty$. This solution is given by

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \\ \sigma_t^\delta = \omega + \omega \sum_{k=1}^{\infty} \sum_{i_1,\dots,i_k \geq 1} a_{i_1,t-i_1} \dots a_{i_k,t-i_1-\dots-i_k} \ |\eta_{t-i_1}|^\delta \dots |\eta_{t-i_1-\dots-i_k}|^\delta \end{cases} \tag{5}$$

with $a_{i,t-j} = \alpha_i^+ \ \mathbb{1}_{\eta_{t-j} \geq 0} + \alpha_i^- \ \mathbb{1}_{\eta_{t-j} < 0}$.

Remarks 1.1.

– *In the ARCH(∞) case, where $\delta = 2$ and $A_s^+ = A_s^- = A_s$, Assumption (4) reduces to (2) since $\mu_{2s}^+ + \mu_{2s}^- = \mu_{2s}$. For the TARCH($\infty$), Bardet and Wintenberger[2] establish the sufficient second order stationarity condition $\sum_{i=1}^{\infty} \max(\alpha_i^+, \alpha_i^-) < 1$ which is stronger than (4) since $A_1^+ \mu_2^+ + A_1^- \mu_2^- \leq \sum_{i=1}^{\infty} \max(\alpha_i^+, \alpha_i^-) \mu_2$.*

It is worth noticing that the process introduced in (3) nests some widely used models in the financial industry. For example, the ARCH(∞) representation of the classical GARCH(1,1) process

$$\varepsilon_t = \sigma_t \eta_t, \quad \sigma_t^2 = \frac{\omega}{1-\beta} + \sum_{i=1}^{\infty} \alpha \ \beta^{i-1} \varepsilon_{t-i}^2, \tag{6}$$

where α and β, are positive constants, $\beta < 1$, and $\omega > 0$ is obviously a particular (symmetrical) specification of (3) with $\delta = 2$. Of course, this specification has short memory as the ARCH(∞) coefficients decay exponentially to zero. A more persistent specification of (3) based on the GARCH(1,1) model (6) is

$$\varepsilon_t = \sigma_t \eta_t, \quad \sigma_t^2 = \frac{\omega}{1-\beta} + \sum_{i=1}^{\infty} \left(\alpha \ \beta^{i-1} + \gamma i^{-d-1}\right) \varepsilon_{t-i}^2 \tag{7}$$

with $\gamma > 0$ and $d > 0$, where the coefficients have an hyperbolic decay. Figure 1 presents the effect of a shock on the conditional variance of a GARCH(1,1) and on an ARCH(∞) process specified as (7) for the same simulation of the iid process. It is seen that the shock at $t = 500$ is less persistent for a GARCH(1,1) process than for the ARCH(∞) one. Even if the β used in this illustration is fairly high (0.85), the effect of the shock has almost entirely disappeared after a hundred lags in the GARCH(1,1) case, while it remains clearly observable on the ARCH(∞) process.

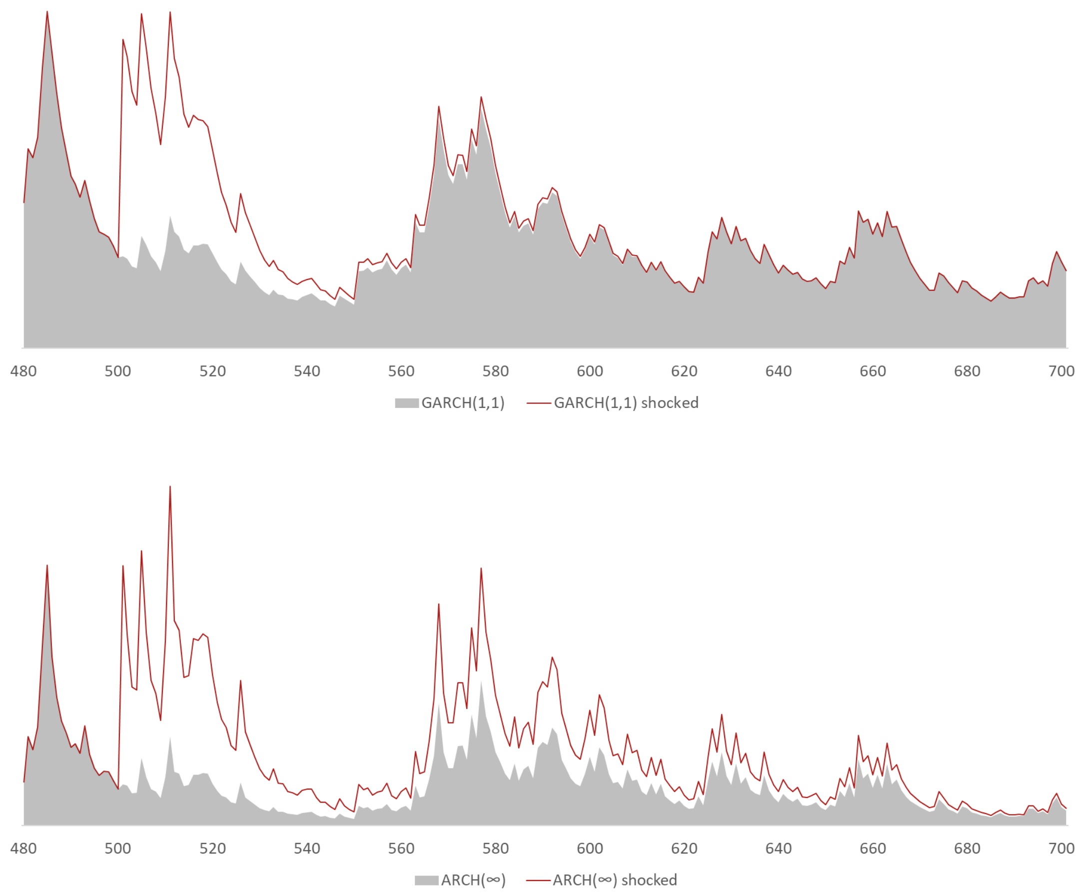


Figure 1: Effect of a shock on η_t at $t = 500$ on the conditional variance of a GARCH(1,1) process and an ARCH(∞) process with $\alpha_i = \alpha\beta^{i-1} + \gamma i^{-(d+1)}$, where $\omega = 0.01$, $\alpha = 0.1$, $\beta = 0.85$, $\gamma = 0.15$, and $d = 1$, and with $\eta_t \sim \mathcal{N}(0, 1)$.

Some well known asymmetric extensions to the GARCH(1,1) are also particular specifications of model (3). Consider the following APARCH(∞) specification

$$\varepsilon_t = \sigma_t \eta_t, \quad \sigma_t^\delta = \frac{\omega}{1-\beta} + \sum_{i=1}^{\infty} \beta^{i-1} (\alpha^+ \mathbb{1}_{\varepsilon_{t-i} \geq 0} + \alpha^- \mathbb{1}_{\varepsilon_{t-i} < 0}) |\varepsilon_{t-i}|^\delta$$

which is the rewriting of an APARCH(1,1) as an APARCH(∞). The GJR-GARCH(1,1) model introduced by Glosten, Jagannathan and Runkle[24] is obtained when $\delta = 2$, and the Threshold GARCH (TGARCH) model of Zakoïan[40] is obtained when $\delta = 1$. In the spirit of (7), an extension to the APARCH(1,1) model to allow for higher persistence is then

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \\ \sigma_t^\delta = \dfrac{\omega}{1-\beta} + \sum\limits_{i=1}^{\infty} \beta^{i-1} (\alpha^+ \mathbb{1}_{\varepsilon_{t-i} \geq 0} + \alpha^- \mathbb{1}_{\varepsilon_{t-i} < 0}) |\varepsilon_{t-i}|^\delta + \gamma i^{-d-1} |\varepsilon_{t-i}|^\delta. \end{cases} \tag{8}$$

The models introduced in (7) and (8) are particularly interesting as they allow to nest GARCH-type specifications in highly persistent volatility models. They will be used throughout the paper to illustrate the assumptions required to establish asymptotic results.

2 Statistical inference of an APARCH(∞) process

Direct estimation of the models defined in (1) and (3) is not feasible without constraining the infinite sequence of coefficients and requires considering a parametrization. Building upon Robinson and Zaffaroni[37], we introduce the parametric form of Model (3)

$$\begin{cases} \varepsilon_t = \sigma_t(\boldsymbol{\theta_0}) \eta_t \\ \sigma_t^\delta(\boldsymbol{\theta_0}) = \omega_0 + \sum\limits_{i=1}^{\infty} \alpha_i^+(\boldsymbol{\phi_0}) |\varepsilon_{t-i}|^\delta \mathbb{1}_{\varepsilon_{t-i} \geq 0} + \alpha_i^-(\boldsymbol{\phi_0}) |\varepsilon_{t-i}|^\delta \mathbb{1}_{\varepsilon_{t-i} < 0} \end{cases} \tag{9}$$

where $\alpha_i^+(.), \alpha_i^-(.) : \boldsymbol{\Phi} \to [0, \infty]$ are known functions, $\boldsymbol{\phi_0}$ is a $r \times 1$ unknown vector of parameters, ω_0 is an unknown positive constant, and $\delta > 0$ is a known parameter. We wish to estimate $\boldsymbol{\theta_0} = (\omega_0, \boldsymbol{\phi_0}')'$ over a parameter space $\boldsymbol{\Theta}$, on the basis of n observations $\varepsilon_1, \dots, \varepsilon_n$. For example, the parametric form of Model (8) would then be $\alpha_i^{+(-)}(\boldsymbol{\phi_0}) = \alpha_0^{+(-)} \beta_0 + \gamma_0 i^{-d_0-1}$ with $\boldsymbol{\phi_0} = (\alpha_0^+, \alpha_0^-, \beta_0, \gamma_0, d_0)$.

Following the works of Berkes, Horváth and Kokoszka[4], and Francq and Zakoïan[16] for the GARCH(p,q) process, asymptotic properties of the QMLE for APARCH(p, q) models have been established by Hamadeh and Zakoïan[28], and extended by Francq and Thieu[15]. In [28], the authors show that the empirical estimation of the power parameter δ, although theoretically possible, is difficult to achieve. Following Francq and Thieu[15], we therefore consider that the parameter δ is fixed and known. In general, this parameter is fixed to 1 (TGARCH) or 2 (GJR-GARCH) by practitioners. A comment on how to choose this parameter is however provided at the end of this section.

Estimation of the parameters of ARCH(∞) models has been first studied by Giraitis and Robinson[22] who proposed a Whittle estimation of $\boldsymbol{\theta_0}$. However, this method presents some drawbacks as discussed by the authors[2]. Linton and Mammen[34] studied semi-parametric estimation of a ARCH(∞) model without parametric specification of the effect of past returns on the conditional variance, but their method requires the existence of a fourth moment for ε_t. Robinson and Zaffaroni[37] proposed to estimate the parameter $\boldsymbol{\theta_0}$ by QML under milder assumptions on the observed process. For different assumptions, see also Hafner and Preminger[27]. Finally, Bardet and Wintenberger[2] studied the QMLE for ARCH(∞) and TARCH(∞) processes under mild assumptions but at the cost of imposing higher moments on ε_t. In the spirit of [37], we study the QMLE in the case of an APARCH(∞) process.

Let us rewrite the volatility in (9) as

$$\sigma_t^\delta(\boldsymbol{\theta_0}) = \omega_0 + \sum_{i=1}^{\infty} a_{i,t-i}(\boldsymbol{\phi_0})|\varepsilon_{t-i}|^\delta \tag{10}$$

where $a_{i,t-i}(\boldsymbol{\phi}) = \alpha_i^+(\boldsymbol{\phi})\mathbb{1}_{\varepsilon_{t-i}\geq 0} + \alpha_i^-(\boldsymbol{\phi})\mathbb{1}_{\varepsilon_{t-i}<0}$ and note that for all i, any t, and any $\boldsymbol{\phi}$ in $\boldsymbol{\Phi}$, $a_{i,t-i}(\boldsymbol{\phi}) \leq \max(\alpha_i^+(\boldsymbol{\phi}), \alpha_i^-(\boldsymbol{\phi}))$. We define the QMLE as

$$\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\text{Argmin}}\ \tilde{Q}_n(\boldsymbol{\theta}),\ \tilde{Q}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\tilde{l}_t(\boldsymbol{\theta}),\ \tilde{l}_t(\boldsymbol{\theta}) = \log\tilde{\sigma}_t^2(\boldsymbol{\theta}) + \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2(\boldsymbol{\theta})}$$

where, for any admissible value $\boldsymbol{\theta}$ of $\boldsymbol{\theta_0}$, $\tilde{\sigma}_t^\delta$ is defined as ω for $t = 1$ and for $t > 1$

$$\tilde{\sigma}_t^\delta(\boldsymbol{\theta}) = \omega + \sum_{i=1}^{t-1} \alpha_i^+(\boldsymbol{\phi})|\varepsilon_{t-i}|^\delta\mathbb{1}_{\varepsilon_{t-i}\geq 0} + \alpha_i^-(\boldsymbol{\phi})|\varepsilon_{t-i}|^\delta\mathbb{1}_{\varepsilon_{t-i}<0} = \omega + \sum_{i=1}^{t-1} a_{i,t-i}(\boldsymbol{\phi})|\varepsilon_{t-i}|^\delta. \tag{11}$$

To show strong consistency, the following assumptions are used, and we denote from now on by K a generic positive constant.

[A1] The parameter space is of the form $\boldsymbol{\Theta} = [\omega_L, \omega_U] \times \boldsymbol{\Phi}$ where $0 < \omega_L < \omega_U < \infty$, and $\boldsymbol{\Phi} \subset \mathbb{R}^r$ is a compact space.

[A2] The η_t are iid with $\mathbb{E}\eta_0 = 0$, $\mathbb{E}\eta_0^2 = 1$ and the distribution of the positive (resp.

[2]For example, the existence of a fourth-moment of ε_t is required for consistency and an eighth-moment for asymptotic normality.

negative) part of (η_t) is non-degenerate.

[A3] (i) For any $\boldsymbol{\phi}$ and $\boldsymbol{\phi}^* \in \boldsymbol{\Phi}$ such that $\boldsymbol{\phi} \neq \boldsymbol{\phi}^*$, there exists $k \geq 1$ such that $\alpha_k^+(\boldsymbol{\phi}) \neq \alpha_k^+(\boldsymbol{\phi}^*)$ and $\alpha_k^-(\boldsymbol{\phi}) \neq \alpha_k^-(\boldsymbol{\phi}^*)$.

(ii) For all $i \geq 1$, $\sup_{\boldsymbol{\phi}\in\boldsymbol{\Phi}} \max(\alpha_i^+(\boldsymbol{\phi}), \alpha_i^-(\boldsymbol{\phi})) \leq K\, i^{-d-1}$ for some $d > 0$.

[A4] There exists a solution (ε_t) of equation (9) such that $\mathbb{E}|\varepsilon_t|^{(2\wedge\delta)\rho} < \infty$ for $\rho > \frac{1}{d+1}$.

[A5] $\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \dfrac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})} < \infty$.

Remarks 2.1.

- *The compactness assumption* ***A1*** *is standard for QML estimation. Assumptions* ***A2*** *and* ***A3(i)*** *are needed for identifiability. The former is slightly stronger than needed in the ARCH(∞) case where only the distribution of (η_t^2) needs to be non-degenerate. Assumption* ***A3(ii)*** *along with Assumption* ***A4*** *entail the existence of $\sigma_t^\delta(\boldsymbol{\theta})$ for any $\boldsymbol{\theta}$. Note that the d in assumptions* ***A3(ii)*** *and* ***A4*** *is the same and that these assumptions may be stronger than (4). Nevertheless, Assumption* ***A4*** *is quite mild as, for a large value of d, it would only imply the existence of a small moment. For example, it is the case for the GARCH(1,1) model where the α_i are exponentially decaying. Note that a sufficient condition for Assumption* ***A5*** *is of course $\mathbb{E}\varepsilon_t^2 < \infty$. Proposition 2 in the appendix gives a different sufficient condition for* ***A5*** *to hold without additional moment condition for ε_t.*

- *In the classical ARCH(∞) case where $\delta = 2$ and α_i^+ and α_i^- are equal, though our assumptions are mostly in line with the ones proposed by Robinson and Zaffaroni[37], they are noticeably milder concerning the distribution of η_t. Indeed, as opposed to [37], we do not specify that the density of η_t is well-behaved near 0. Furthermore, our assumptions on α_i are also milder as we allow our coefficients to be equal to 0 and do not impose $\alpha_i(\boldsymbol{\phi_0}) \leq K\alpha_j(\boldsymbol{\phi_0})$ for $i \geq j \geq 1$. Note that Robinson and Zaffaroni considered a slightly more general model $y_t = \mu + \varepsilon_t$ allowing for a drift. Our model could similarly be extended to take into account this parameter but for the sake of clarity we assume that this drift parameter is known and equal to 0.*

Notice that Model (8) where $\alpha_i^{+(-)}(\boldsymbol{\phi_0}) = \alpha_0^{+(-)}\beta_0^{i-1} + \gamma_0 i^{-(d_0+1)}$ and $\boldsymbol{\Phi} \subset (0,\infty)^5$ satisfies the proposed assumptions. In particular, Assumption **A3(i)** is satisfied if $\alpha_0^{+(-)}$, β_0, and γ_0 are positive, which ensures Assumption **A5** using Proposition 2.

The following result states the strong consistency of $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}$.

Theorem 2. *Under assumptions **A1-A5**, almost surely*

$$\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} \to \boldsymbol{\theta}_{\mathbf{0}}, \qquad \text{as } n \to \infty.$$

To show the asymptotic normality, the following additional assumptions are considered.

[A6] $\boldsymbol{\theta}_{\mathbf{0}}$ belongs to the interior of $\boldsymbol{\Theta}$.

[A7] $\kappa_\eta = \mathbb{E}\eta_0^4 < \infty$.

[A8] For all $i \geq 1$, $\max(\alpha_i^+(\boldsymbol{\phi}_{\mathbf{0}}), \alpha_i^-(\boldsymbol{\phi}_{\mathbf{0}})) \leq K\ i^{-d^*-1}$ for some $d^* > \dfrac{1}{2}$.

[A9] $\mathbb{E}|\varepsilon_t|^{(2\wedge\delta)\rho} < \infty$ for some $\rho > \frac{4}{2d^*+3}$.

[A10] (i) For all j, α_j has continuous kth derivative on $\boldsymbol{\Phi}$, $k \leq 3$, such that, denoting ϕ_i the ith element of $\boldsymbol{\phi}$,

$$\left|\frac{\partial^k \alpha_j^+(\boldsymbol{\phi})}{\partial\phi_{i_1}...\partial\phi_{i_k}}\right| \leq K(\alpha_j^+)^{1-\xi}(\boldsymbol{\phi}) \quad \text{and} \quad \left|\frac{\partial^k \alpha_j^-(\boldsymbol{\phi})}{\partial\phi_{i_1}...\partial\phi_{i_k}}\right| \leq K(\alpha_j^-)^{1-\xi}(\boldsymbol{\phi})$$

for all $\xi > 0$ and all $i_h = 1, ..., r$, $h = 1, ..., k$.

(ii) There exists $i_h^+ = i_h^+(\boldsymbol{\phi}_{\mathbf{0}})$ and $i_h^- = i_h^-(\boldsymbol{\phi}_{\mathbf{0}})$, $h = 1, ..., r$, such that $1 \leq i_1^{+(-)} < ... < i_r^{+(-)} < \infty$ and

$$\operatorname{rank}\left[\frac{\partial\alpha_{i_1^+}^+(\boldsymbol{\phi}_{\mathbf{0}})}{\partial\boldsymbol{\phi}} ... \frac{\partial\alpha_{i_r^+}^+(\boldsymbol{\phi}_{\mathbf{0}})}{\partial\boldsymbol{\phi}}\right] = \operatorname{rank}\left[\frac{\partial\alpha_{i_1^-}^-(\boldsymbol{\phi}_{\mathbf{0}})}{\partial\boldsymbol{\phi}} ... \frac{\partial\alpha_{i_r^-}^-(\boldsymbol{\phi}_{\mathbf{0}})}{\partial\boldsymbol{\phi}}\right] = r.$$

[A11] For all k>0, there exists a neighborhood $V(\boldsymbol{\theta}_{\mathbf{0}})$ of $\boldsymbol{\theta}_{\mathbf{0}}$ such that,

$$\mathbb{E}_{\boldsymbol{\theta}_{\mathbf{0}}} \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_{\mathbf{0}})} \left[\frac{\sigma_t^\delta(\boldsymbol{\theta}_{\mathbf{0}})}{\sigma_t^\delta(\boldsymbol{\theta})}\right]^k < \infty.$$

Remarks 2.2.

- *Assumption **A6** is required for asymptotic normality. Assumption **A7** is necessary for the existence of the variance of the score vector $\partial l_t(\boldsymbol{\theta}_{\mathbf{0}})/\partial\boldsymbol{\theta}$. Assumptions **A8** and **A9** are stronger than Assumptions **A3(ii)** and **A4** and impose a higher rate of convergence for $\alpha_i^{+(-)}$. Assumption **A10(i)** is similar to Assumption **A3(ii)** and allows the summability of the derivatives of the $\alpha_i^{+(-)}$ functions, while Assumption **A10(ii)** ensures non singularity of the matrix $\boldsymbol{J}$. The particular rates of convergence of the $\alpha_i^{+(-)}$ functions and their derivatives imposed in Assumptions **A3(ii)**, **A8** and **A10(i)** are crucial to show the asymptotic irrelevance of the initial values and the* *integrability of the derivatives in a neighborhood of $\boldsymbol{\theta_0}$. Proposition 3 in the appendix gives an example of a sufficient condition for **A11**.*

Note again that Model (8) satisfies the set of additional assumptions if $d_0 > 0.5$. In particular, Assumption **A10** is satisfied and Assumption **A11** holds from Proposition 3.

Theorem 3. *Under assumptions A1-A11,*

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta_0}) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, (\kappa_\eta - 1)\boldsymbol{J}^{-1}) \tag{12}$$

where

$$\boldsymbol{J} = \frac{4}{\delta^2}\mathbb{E}_{\boldsymbol{\theta_0}}\left[\frac{1}{\sigma_t^{2\delta}(\boldsymbol{\theta_0})}\frac{\partial\sigma_t^\delta(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^\delta(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta'}}\right]$$

is a positive definite matrix.

Next, we consider an extension allowing the true parameter value to lie on the boundary of $\boldsymbol{\Theta}$, relaxing Assumption **A6**. Let $\boldsymbol{\Phi} = [\underline{\phi_1}, \overline{\phi_1}] \times \cdots \times [\underline{\phi_r}, \overline{\phi_r}]$, define $\partial\boldsymbol{\Phi} = \{\boldsymbol{\phi_0} \in \Phi : \phi_{0,i} = \underline{\phi_i}$ for some $i > 0\}$ and let $\boldsymbol{\phi_0}(\varepsilon)$ the vector obtained by replacing $\phi_{0,i}$ by $\underline{\phi_i} + \varepsilon$ for all i such that $\phi_{0,i} = \underline{\phi_i}$. Similarly to Francq and Zakoïan[17] for the case of GARCH(p, q) models, we make the following assumption to prevent $\boldsymbol{\phi_0}$ from reaching the upper bound of $\boldsymbol{\Phi}$.

[A6'] There exists $\varepsilon > 0$ such that $\boldsymbol{\theta_0}(\varepsilon) = [\omega_0, \boldsymbol{\phi_0}(\varepsilon)']'$ belongs to the interior of $\boldsymbol{\Theta}$.

The following theorem establishes the asymptotic distribution of $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}$ when $\boldsymbol{\theta_0}$ may be on the boundary.

Theorem 4. *Under the assumptions of Theorem 3 where A6 is replaced by A6',*

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta_0}) \xrightarrow{\mathcal{L}} \boldsymbol{\lambda}^\Lambda := \arg\inf_{\boldsymbol{\lambda}\in\Lambda}[\boldsymbol{\lambda} - \boldsymbol{Z}]'\boldsymbol{J}[\boldsymbol{\lambda} - \boldsymbol{Z}] \tag{13}$$

with $\boldsymbol{Z} \sim \mathcal{N}(\mathbf{0}, (\kappa_\eta - 1)\boldsymbol{J}^{-1})$, $\Lambda = \Lambda(\boldsymbol{\theta_0}) = \Lambda_1 \times \cdots \times \Lambda_{r+1}$, where $\Lambda_1 = \mathbb{R}$ and for $i = 2, \ldots, r + 1$, $\Lambda_i = \mathbb{R}$ if $\phi_{0,i} \neq \underline{\phi_i}$ and $\Lambda_i = [\underline{\phi_i}, \infty)$ otherwise.

Remarks 2.3.

- *We emphasize that the asymptotic distribution of the QMLE is obtained without any additional assumption on the moments of ε_t. Similarly, Francq and Zakoïan[17] establish the asymptotic distribution of the QMLE without assuming additional moment assumption by ensuring that assumptions **A5** and **A11** are satisfied.*

- *The asymptotic distribution in (13) is the orthogonal projection of a normal vector distribution onto a convex cone, see [17] for a practical derivation of this limiting distribution.*

Asymptotic results for $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}$ have been obtained under the assumption that δ was known. Although the choice of the power parameter may have little influence on the fitted volatility (see Francq and Thieu[15] and references therein), a practitioner might be unsure of which model to select. As the number of unknown parameters in $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}$ is the same for different choices of δ, it seems natural to select the model with the highest quasi likelihood. The following proposition justifies this approach.

Let us denote by δ_0 the true value of the power parameter and replace $\sigma_t^{\delta}(\boldsymbol{\theta_0})$ in (9) and $\tilde{\sigma}_t^{\delta}(\boldsymbol{\theta_0})$ in (11) by $\sigma_{\delta_0,t}^{\delta_0}(\boldsymbol{\theta_0})$ and $\tilde{\sigma}_{\delta_0,t}^{\delta}(\boldsymbol{\theta_0})$ to underline that they actually depend on the value of δ_0. Consider a set $\boldsymbol{\mathcal{D}}$ of k candidates for δ_0 (e.g. $\delta_0 = 1$ for the APARCH(∞) extension of the TGARCH, or $\delta_0 = 2$ for a TARCH(∞)) such as

$$\delta_0 \in \boldsymbol{\mathcal{D}} = \{\delta_1, \dots, \delta_k\}, \quad \delta_i > 0, \quad i = 1, \dots, k \tag{14}$$

and let

$$(\tilde{\delta}_n, \tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}) = \underset{\delta \in \boldsymbol{\mathcal{D}}, \boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{Argmin}} \tilde{Q}_{\delta,n}(\boldsymbol{\theta}), \;\; \tilde{Q}_{\delta,n}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} \tilde{l}_{\delta,t}(\boldsymbol{\theta}), \;\; \tilde{l}_{\delta,t}(\boldsymbol{\theta}) = \log \tilde{\sigma}_{\delta,t}^2(\boldsymbol{\theta}) + \frac{\varepsilon_t^2}{\tilde{\sigma}_{\delta,t}^2(\boldsymbol{\theta})}. \tag{15}$$

We need to slightly strengthen the assumption on the distribution of η_t.

[A2'] The η_t are iid with $\mathbb{E}\eta_0 = 0$, $\mathbb{E}\eta_0^2 = 1$ and the positive (resp. negative) part of (η_t) takes more than 3 values.

Theorem 5. *Under the assumptions of Theorem 2, Assumption A2', and (14), almost surely $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} \to \boldsymbol{\theta_0}$ as $n \to \infty$ and $\tilde{\delta}_n = \delta_0$ for n large enough. Moreover, under the corresponding additional assumptions, the asymptotic distribution of $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}$ is still given by Theorem 3 or Theorem 4.*

Note that in Model (8), Assumption **A3(i)** do not allow for the parameters β_0 or γ_0 to be equal to zeros for identification reasons. A particular discussion on how to test $\gamma_0 = 0$ is proposed in Section 3.3.

3 Specification tests

The presence of asymmetry and memory in financial time series has been well documented. However, in order to select the most parsimonious model, it is critical to test their statistical significance and the adequacy of the chosen model. This section introduces simple test procedures for goodness-of-fit, asymmetry and strong (non-exponentially decaying) memory.

3.1 Portmanteau goodness-of-fit test for APARCH(∞) models

Since their introduction by Box and Pierce[10] tests based on residuals autocorrelations, the so-called Portmanteau tests, have become widely used in econometrics. To test the adequacy of conditional volatility models, Li and Mak[33] proposed to use Portmanteau tests based on squared residuals autocorrelations. Asymptotic properties of these tests have been established by Berkes, Horváth and Kokoszka[3] for standard GARCH(p,q) models and by Carbon and Francq[11] in the APARCH(p,q) case. To our best knowledge, these results having not yet been extended to the ARCH(∞) literature, this section aims at filling that gap. One should note that other kinds of goodness-of-fit tests exist. In particular, Hidalgo and Zaffaroni[30] propose a goodness-of-fit test based on the estimated spectral distribution function. However, contrary to Portmanteau tests, their statistic has a nonstandard asymptotic distribution and requires bootstrap procedures to compute critical values.

Let us consider the null hypothesis H_0^{GoF} that the process (ε_t) satisfies model (3). We define the autocovariances of the squared residuals by

$$\hat{r}_h = n^{-1} \sum_{t=h+1}^{n} (\hat{\eta}_t^2 - 1)(\hat{\eta}_{t-h}^2 - 1) \ , \ \text{with } \hat{\eta}_t^2 = \varepsilon_t^2 / \tilde{\sigma}_t^2(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}),$$

and let $\hat{\boldsymbol{r}}_m = (\hat{r}_1, \dots, \hat{r}_m)$ for any $1 \leq m \leq n$, and $\hat{\boldsymbol{C}}_m$ the $m \times (r+1)$ matrix whose elements (h,k) are given by

$$\hat{\boldsymbol{C}}_m(h,k) = -\frac{2}{\delta n} \sum_{t=h+1}^{n} (\hat{\eta}_{t-h}^2 - 1) \frac{1}{\tilde{\sigma}_t^\delta(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}})} \frac{\partial \tilde{\sigma}_t^\delta(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}})}{\partial \theta_k}.$$

In addition, let

$$\hat{\boldsymbol{J}}_n = \frac{4}{\delta^2} \frac{1}{n} \sum_{t=1}^{n} \frac{1}{\tilde{\sigma}_t^{2\delta}(\tilde{\boldsymbol{\theta}}_n)} \frac{\partial \tilde{\sigma}_t^\delta(\tilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} \frac{\partial \tilde{\sigma}_t^\delta(\tilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}'} \quad \text{and} \quad \hat{\kappa}_\eta = \frac{1}{n} \sum_{t=1}^{n} \frac{\varepsilon_t^4}{\tilde{\sigma}_t^4(\tilde{\boldsymbol{\theta}}_n)}$$

be consistent estimators of $\boldsymbol{J}$ and κ_η respectively (from arguments in the proofs of Theorem 3, Lemma 1, and the ergodic theorem).

The following theorem establishes the asymptotic distribution of the Portmanteau test statistic.

Theorem 6. *Under H_0^{GoF}, under the assumptions of Theorem 3 and assumption **A2'**,*

$$n \hat{\boldsymbol{r}}_m' \hat{\boldsymbol{D}}^{-1} \hat{\boldsymbol{r}}_m \xrightarrow{\mathcal{L}} \chi_m^2,$$

with $\hat{\boldsymbol{D}} = (\hat{\kappa}_\eta - 1)^2 \boldsymbol{I}_m - (\hat{\kappa}_\eta - 1)\hat{\boldsymbol{C}}_m \hat{\boldsymbol{J}}_n^{-1} \hat{\boldsymbol{C}}'_m$.

The adequacy of the APARCH(∞) model (3) is then rejected at the asymptotic level ν when $n\hat{\boldsymbol{r}}'_m \hat{\boldsymbol{D}}^{-1} \hat{\boldsymbol{r}}_m > \chi^2_m(1-\nu)$ where $\chi^2_m(1-\nu)$ is the $(1-\nu)$-quantile of the χ^2 distribution with m degrees of freedom.

3.2 Testing for linear constraints on the parameters

We are now interested in testing for a general hypothesis of the form

$$H_0 : \boldsymbol{R}\, \boldsymbol{\theta_0} = \boldsymbol{k}, \quad H_1 : \boldsymbol{R}\, \boldsymbol{\theta_0} \neq \boldsymbol{k} \tag{16}$$

where $\boldsymbol{R}$ is the constraints matrix and $\boldsymbol{k}$ is a constant vector. Let c be the rank of the matrix $\boldsymbol{R}$. The triptych of the Wald, Rao-score, and Quasi Likelihood Ratio (LR) statistics to test (16) is given by

$$\begin{aligned} W_n &= (\boldsymbol{R}\, \tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{k})' \left(\boldsymbol{R} \left(\frac{(\hat{\kappa}_\eta - 1)}{n} \hat{\boldsymbol{J}}_n^{-1} \right) \boldsymbol{R}' \right)^{-1} (\boldsymbol{R}\, \tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{k}) \\ R_n &= \frac{n}{\hat{\kappa}_{\eta|H_0} - 1} \frac{\partial \tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0})}{\partial \boldsymbol{\theta}'} \hat{\boldsymbol{J}}^{-1}_{n|H_0} \frac{\partial \tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0})}{\partial \boldsymbol{\theta}} \\ L_n &= \frac{2n}{\hat{\kappa}_{\eta|H_0} - 1} \left[\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0}) - \tilde{Q}_n(\tilde{\boldsymbol{\theta}}_n) \right] \end{aligned} \tag{17}$$

where $\tilde{\boldsymbol{\theta}}_{n|H_0}$ is the QMLE restricted by H_0 and

$$\hat{\boldsymbol{J}}_{n|H_0} = \frac{4}{\delta^2} \frac{1}{n} \sum_{t=1}^n \frac{1}{\tilde{\sigma}_t^{2\delta}(\tilde{\boldsymbol{\theta}}_{n|H_0})} \frac{\partial \tilde{\sigma}_t^\delta(\tilde{\boldsymbol{\theta}}_{n|H_0})}{\partial \boldsymbol{\theta}} \frac{\partial \tilde{\sigma}_t^\delta(\tilde{\boldsymbol{\theta}}_{n|H_0})}{\partial \boldsymbol{\theta}'}, \quad \text{and} \quad \hat{\kappa}_{\eta|H_0} = \frac{1}{n} \sum_{t=1}^n \frac{\varepsilon_t^4}{\tilde{\sigma}_t^4(\tilde{\boldsymbol{\theta}}_{n|H_0})}$$

are consistent estimators of $\boldsymbol{J}$ and κ_η respectively, under H_0.

Proposition 1. *Under* $H_0 : \boldsymbol{R}\, \boldsymbol{\theta_0} = \boldsymbol{k}$,

(i) under the assumptions of Theorem 3,

$$W_n \xrightarrow{\mathcal{L}} \chi^2_c, \quad R_n \xrightarrow{\mathcal{L}} \chi^2_c, \quad \textit{and } L_n \xrightarrow{\mathcal{L}} \chi^2_c,$$

(ii) under the assumptions of Theorem 4,

$$\begin{aligned} &W_n \xrightarrow{\mathcal{L}} \boldsymbol{\lambda}^{\Lambda\prime} \boldsymbol{R}'[(\kappa_\eta - 1)\boldsymbol{R}\boldsymbol{J}^{-1}\boldsymbol{R}']^{-1} \boldsymbol{\lambda}^{\Lambda\prime} \boldsymbol{R}, \quad R_n \xrightarrow{\mathcal{L}} \chi^2_c, \quad \textit{and} \\ &L_n \xrightarrow{\mathcal{L}} -\frac{1}{2}(\boldsymbol{\lambda}^\Lambda - \boldsymbol{Z})'\boldsymbol{J}(\boldsymbol{\lambda}^\Lambda - \boldsymbol{Z}) + \frac{1}{2}\boldsymbol{Z}'\boldsymbol{R}'[\boldsymbol{R}\boldsymbol{J}^{-1}\boldsymbol{R}']^{-1}\boldsymbol{R}\boldsymbol{Z}. \end{aligned}$$

Note that in Model (9), the symmetry hypothesis is generally a particular constrained representation. Testing for the significance of asymmetry can thus be achieved by testing an implied restriction on $\boldsymbol{\theta_0}$. For example, if we consider the parametric version of specification (8) obtained by setting $\boldsymbol{\phi_0} = (\alpha_0^+, \alpha_0^-, \beta_0, \gamma_0, d_0)$ and $\alpha_i^{+(-)}(\boldsymbol{\phi_0}) = \beta_0^{i-1}\alpha_0^{+(-)} + \gamma_0 i^{-(d_0+1)}$, the symmetry hypothesis is given by

$$H_0^{\mathrm{sym}} : \alpha_0^+ = \alpha_0^-, \quad H_1^{\mathrm{asym}} : \alpha_0^+ \neq \alpha_0^-$$

which is a particular form of (16). Testing for a constrained representation is highly common when testing for asymmetry in parametric models, see for example Nelson [35].

3.3 Testing for GARCH(1,1) specifications

Despite the development of multiple extensions, the GARCH(1,1) model remains preeminent in the financial industry and literature. Although this model admits an ARCH(∞) representation, it imposes an exponential decay on its coefficients. We propose to study the validity of a GARCH(1,1) representation by allowing these coefficients to decay in a slower manner. In order to do so, consider the following ARCH(∞) parametrization

$$\begin{cases} \varepsilon_t = \sigma_t(\boldsymbol{\theta_0})\eta_t \\ \sigma_t^2(\boldsymbol{\theta_0}) = \dfrac{\omega_0}{1-\beta_0} + \sum\limits_{i=1}^{\infty} (\alpha_0 \beta_0^{i-1} + \gamma_0 i^{-(d_0+1)})\varepsilon_{t-i}^2 \end{cases} \tag{18}$$

with $\alpha_0 > 0$, $\beta_0 > 0$, $\gamma_0 \geq 0$, and $d_0 > 0$. Testing the validity of a GARCH(1,1) representation can then be achieved by testing

$$H_0^{\mathrm{GARCH}} : \gamma_0 = 0, \quad H_1^{\mathrm{ARCH}(\infty)} : \gamma_0 > 0, \tag{19}$$

which can be rewritten as $H_0^{\mathrm{GARCH}} : \boldsymbol{R}\ \boldsymbol{\theta_0} = 0$, and $H_1^{\mathrm{ARCH}(\infty)} : \boldsymbol{R}\ \boldsymbol{\theta_0} > 0$ with $\boldsymbol{R} = (0,0,0,1,0)$. While this test may seem standard, it poses a major difficulty. Indeed, the parameter d_0 is not identified under the null hypothesis, thus we cannot directly use Proposition 1(ii) to obtain the asymptotic distribution of the test statistics when the parameter is on the boundary. A simple solution could be to assume that the parameter d_0 is known and fixed at a value $\overline{d} > 0.5$. Under this assumption, the Wald statistic distribution is a mixture of a χ_1^2 and a Dirac measure at 0, both with weight 1/2 (see Proposition 5 in the supplementary file). In addition, Francq and Zakoïan[18] show that, when testing the nullity of only one coefficient, the Wald test is locally asymptotically more powerful than the standard score test. Although setting the unidentified under the null parameter at an arbitrary value facilitates the derivation of the asymptotic distribution, choosing a value $\overline{d}$ that is far from d_0 may lead to spurious results[3]. Asymptotic results, when the presence of a coefficient on the boundary of the parameter space involves the non identification of a second parameter, have been established by Andrews[1]. However, the limiting distributions in such case are highly non-standard.

[3]See the supplementary file for a study of the empirical power of the test when $\overline{d}$ is misspecified.

Instead, we propose to use a residual-based bootstrap procedure to approximate the statistic asymptotic distribution. Using the terminology of Beutner, Heinemann and Smeekes[6], we propose the following recursive design bootstrap procedure for testing (19) on a sample of n observations $\varepsilon_1, \dots, \varepsilon_n$. Let us denote by $\tilde{\boldsymbol{\theta}}_n^c = (\tilde{\omega}^c, \tilde{\alpha}^c, \tilde{\beta}^c)'$ the QMLE of a GARCH(1,1) model and let $\tilde{\sigma}_t^c(\tilde{\boldsymbol{\theta}}_n^c)$ the estimated volatility process.

1. On the observations, compute $\tilde{\boldsymbol{\theta}}_n^c$ and the standardized residuals $\tilde{\eta}_t^c = \hat{\eta}_t^c / s_n^c$ where $\hat{\eta}_t^c = \varepsilon_t / \tilde{\sigma}_t^c(\tilde{\boldsymbol{\theta}}_n^c)$ and $(s_n^c)^2 = n^{-1} \sum_{t=1}^n (\hat{\eta}_t^c)^2$. Denote by F_n^* the empirical distribution of $\tilde{\eta}_t^c$.

2. Simulate a trajectory of length n of a GARCH(1,1) with parameter $\tilde{\boldsymbol{\theta}}_n^c$ and where the innovations $\eta_t^* \overset{\text{iid}}{\sim} F_n^*$. On this simulation, compute the unconstrained estimator $\tilde{\boldsymbol{\theta}}_n^*$ of an ARCH(∞) and compute the statistic $W_n^{\text{GARCH}*}$.

3. On the observations, compute the unconstrained estimator $\tilde{\boldsymbol{\theta}}_n$ of an ARCH(∞) and compute the statistic W_n.

4. Repeat B times Step 2 and denote by $W_n^{\text{GARCH}*1} \dots W_n^{\text{GARCH}*B}$ the obtained bootstrap test statistics. Approximate the p-value of the test H_0^{GARCH} against $H_1^{\text{ARCH}(\infty)}$ by $B^{-1} \sum_{b=1}^B \mathbb{1}_{W_n^{\text{GARCH}*b} \geq W_n^{\text{GARCH}}}$.

This test can easily be extended to an asymmetric volatility model with a different δ_0. Consider, the APARCH(∞) specification presented in (8). Testing for the adequacy of the GJR-GARCH model ($\delta_0 = 2$) or the TGARCH ($\delta_0 = 1$) can then be achieved by testing for $H_0 : \gamma_0 = 0$ against $H_1^{\text{APARCH}(\infty)} : \gamma_0 > 0$. By changing the constrained model, in steps 1 and 2, and the unconstrained model in step 3, we obtain the corresponding test procedures. In the case of ARCH(∞) models, the validity of this approach has been established by Hidalgo and Zaffaroni[30]. Note that the proposed bootstrap might be invalid when other parameters than γ_0 are on the boundary. However, the procedure can be modified to account for such problems using the recent technique introduced by Cavaliere et al[12].

4 Simulations

In order to assess the finite sample properties of the QMLE in the different settings studied in this paper and to study the empirical behavior of the test statistics defined in Section 3, we carry out some Monte Carlo experiments. In the following simulations, we use Gaussian innovations ($\eta_t \sim \mathcal{N}(0,1)$).

We focus on specifications (7) and (8) that nest several favored volatility models. We want to estimate $\boldsymbol{\theta}_\mathbf{0} = (\omega_0, \alpha_0^+, \alpha_0^-, \beta_0, \gamma_0, d_0)$. We start by simulating a thousand sample of size $n = 5000$ of different specifications including symmetric models (i.e. with

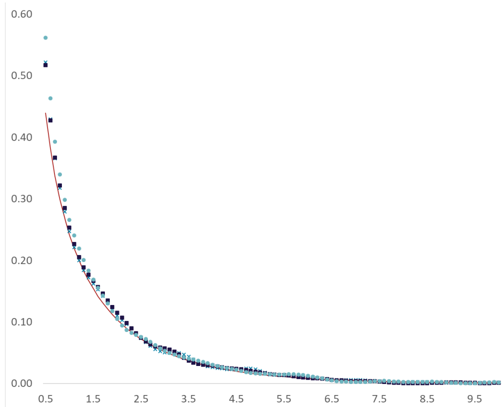
$\alpha_0^+ = \alpha_0^-$) either for $\delta_0 = 1$ or $\delta_0 = 2$. Empirical mean and RMSE of the obtained QMLE are reported in Table 1 as well as the empirical mean of $\tilde{\delta}_n$ obtained from Proposition 1 when the candidates for δ_0 range from 0.5 to 3 with a 0.25 step. In order to assess the finite sample properties of the asymptotic variance estimator, given by (12), we can compare $V_n^{1/2} = \text{diag}[(\hat{\kappa}_\eta - 1)\hat{\boldsymbol{J}}_n^{-1}]^{1/2}/\sqrt{n}$ to the RMSE. On that matter, the results in Table 1 are quite satisfactory. Note that $d_0 = 1$ allows to easily derive sufficient stationary conditions for model (8) as the Riemann sum $\sum_{i=1}^{\infty} i^{-2} = \pi^2/6$ and thus, for $\delta = 2$, Theorem 1 entails the existence of a second order stationary solution if $\max(\alpha_0^+, \alpha_0^-)/(1-\beta_0) + \gamma_0\pi^2/6 < 1$, which is verified for the $\boldsymbol{\theta_0}$ reported in Table 1.

		$\delta_0 = 2$				$\delta_0 = 1$			
	$\boldsymbol{\theta_0}$	$\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}$	Bias	RMSE	$V_n^{\frac{1}{2}}$	$\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}$	Bias	RMSE	$V_n^{\frac{1}{2}}$
ω	0.20	0.209	0.009	0.053	0.029	0.206	0.006	0.046	0.024
α^+	0.05	0.057	0.007	0.038	0.021	0.055	0.005	0.038	0.017
α^-	0.15	0.158	0.008	0.045	0.026	0.155	0.005	0.043	0.019
β	0.70	0.688	-0.012	0.067	0.035	0.691	-0.009	0.060	0.029
γ	0.15	0.140	-0.010	0.049	0.029	0.142	-0.008	0.048	0.022
d	1.0	0.983	-0.017	0.383	0.157	0.980	-0.020	0.388	0.115
				$\tilde{\delta}_n = 2.001$				$\tilde{\delta}_n = 0.988$	
ω	0.20	0.211	0.011	0.057	0.029	0.213	0.013	0.064	0.019
α^+	0.10	0.108	0.008	0.038	0.021	0.108	0.008	0.040	0.013
α^-	0.10	0.108	0.008	0.037	0.021	0.109	0.009	0.039	0.013
β	0.75	0.739	-0.011	0.053	0.027	0.736	-0.014	0.064	0.017
γ	0.20	0.189	-0.011	0.049	0.029	0.189	-0.011	0.050	0.019
d	1.0	0.989	-0.011	0.372	0.129	0.973	-0.027	0.398	0.090
				$\tilde{\delta}_n = 2.005$				$\tilde{\delta}_n = 0.997$	

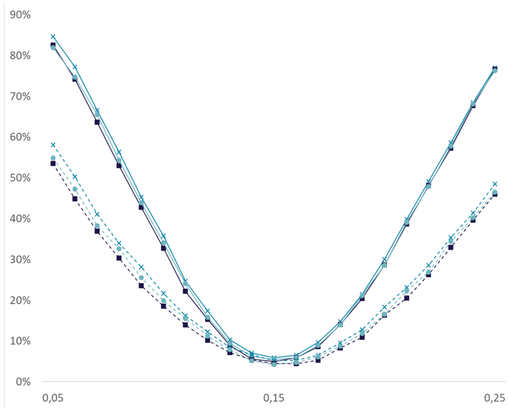
Table 1: *Estimation results for 1000 simulations of size 5000 of an APARCH(∞) process defined as (8) with different specifications and for $\delta_0 = 1$ and 2.*

We then turn to the asymptotic properties of the tests statistics introduced in Section 3. Note that in model (8), the null hypothesis $H_0^{\text{sym}} : \alpha_0^+ = \alpha_0^-$, is a linear constraint on $\boldsymbol{\theta_0}$ with $\boldsymbol{R} = (0, 1, -1, 0, 0, 0)$. We denote by W_n^{sym}, R_n^{sym} and L_n^{sym} the Wald, Rao, and Quasi-Likelihood Ratio test statistics derived from (17). Figure 2(a) presents kernel density estimators of the three test statistics for $n = 5000$ under H_0^{sym} obtained with 5000 replications for $\boldsymbol{\theta_0} = (0.2, 0.1, 0.1, 0.75, 0.2, 1)$. All kernel estimators are close to the asymptotic distribution χ_1^2. In addition, the relative rejection frequency of the Wald, Rao-score, and LR test statistics, at the asymptotic level 5%, are respectively 5.38%, 5.70% and 5.78%, while when using 5000 independent replications, the empirical level should belong to the confidence interval $[4.40\%, 5.60\%]$, hence the Wald statistic seems to better control the error of first kind. To study the empirical behavior of these statistics under H_1^{asym}, we also performed the tests on each realization of a TARCH(∞) simulations sample when $\boldsymbol{\theta_0} = (0.2, \alpha_0^+, 0.15, 0.5, 0.25, 1)$ and α_0^+ ranges from 0.05 to

0.25. Figure 2(b) compares the observed powers of the three tests, that is, the relative frequency of rejection of the null hypothesis of symmetry on the 1000 independent realizations of length $n = 2500$ and $n = 5000$, as a function of α_0^+. On these simulations, we see that the three test statistics seem powerful but may require a large number of observations to capture a weak asymmetry.



(a) *Comparison between kernel density estimators and the χ_1^2 density on [0.5,∞) (red solid line) on 5000 simulations of a symmetric ARCH(∞) process for sample size $n = 5000$.*

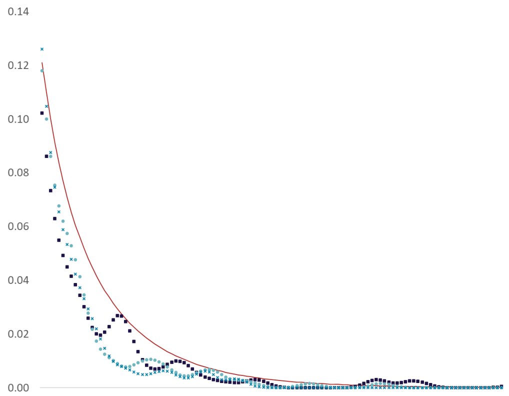


(b) *Observed powers as a function of α_0^+ when $\alpha_0^- = 0.15$, on 1000 simulations with $n = 2500$ (dashed line) and $n = 5000$ (solid line).*

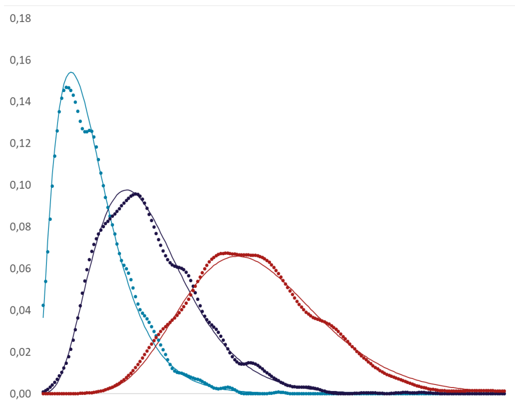
Figure 2: *Empirical behavior of the Wald (dark blue square), the Rao-score (light blue dot), and the LR (blue cross) test statistics.*

Figure 3(a) presents kernel estimators for the Wald test statistics defined in Section 3.2 when testing for a GARCH(1,1), a GJR-GARCH(1,1) and a TGARCH(1,1) against an APARCH(∞) model of form (8) with $\delta_0 = 2$ and 1 respectively under H_0. The statistics have been obtained by adapting the "Warp-Speed" bootstrap techniques introduced by Giacomini, Politis and White [20] to reduce the computational burden of the bootstrap procedure. The parameters used for the simulations are $\boldsymbol{\theta_0} = (0.2, 0.15, 0.75)$ for the GARCH model and $\boldsymbol{\theta_0} = (0.2, 0.05, 0.2, 0.75)$ for both the GJR-GARCH and the TGARCH. All kernels estimators are obtained from 1000 replications. We clearly see that the estimated distributions are different from the theoretical asymptotic distributions when there is no identification issue. The relative rejection frequency of the test statistics, at the asymptotic levels 5%, are respectively 3.70%, 4.60% and 5.50%. We then repeat the experience under H_1 with $\boldsymbol{\theta_0}$ similar to the top part of Table 1. The obtained empirical power of the three test statistics, at the asymptotic levels 5%, are respectively 95.5%, 71.4% and 69.3%. It thus appears that on these realizations, the GARCH(1-1) test has a better power but seems to have a lower control of the error of first kind.

Finally, Figure 3(b) presents the empirical kernels of the Portmanteau statistic for the goodness-of-fit test presented in Section 3.1. The kernels are obtained from 1000 simulations of an APARCH(∞) with $\delta = 1$ and $\boldsymbol{\theta_0} = (0.2, 0.05, 0.15, 0.7, 0.15, 1.0)$, for m =5, 10 and 20 lags. All are close to the theoretical asymptotic distributions. The relative rejection frequency of the test statistics, at the asymptotic levels 5%, are 5.0%, 5.8%, 3.2% and 5.9% for 5, 10, 20 and 50 lags respectively.



(a) $\chi_1^2/2$ *density (red solid line) and kernel density estimators when testing for a GARCH (dark blue square), a GJR-GARCH (blue cross) and a TGARCH(light blue dot) on 1000 simulations with* $n = 5000$.



(b) *Kernel density estimators (dots) and asymptotic distributions (solid line) of the Portmanteau test statistic for* $m = 5$ *(light blue),* $m = 10$ *(dark blue) and* $m = 20$ *(red) on 1000 simulations of size* $n = 5000$.

Figure 3: *Kernel density estimators for the GARCH-type test statistics and for the goodness-of-fit test under their respective null hypothesis.*

Additional simulations results are presented in the supplementary file.

5 Application: Are GARCH(1,1)-type models suitable for peripheral markets?

Despite the development of numerous extensions, short memory models, and in particular GARCH(1,1) specifications, remain the preferred choice for most academics and practitioners when studying volatility. However, the weak persistence they impose might be too restrictive to accurately model some financial time series. We propose to test the GARCH(1,1), TGARCH(1,1), and GJR-GARCH(1,1) specifications on a broad set of equity indices to verify whether their preeminence is justified.

Our dataset contains daily returns from January 1995 to December 2020[4] of 30 indices in their local currency, from five regions with the following breakdown: 4 in North America (S&P500, Nasdaq, TSX, Mexico IPC), 11 in Europe (FTSE, DAX, CAC, SMI, AEX, FTSE MIB, IBEX, MOEX, WIG, BUX, TA-125), 10 in Asia (Nikkei, KOSPI, Hang Seng, TAIEX, MSCI Singapore, BSET, PSEi, IDX, KLCI, NIFTY), 2 in Oceania (ASX AO, MSCI New Zealand), and 3 in South America (Merval, Bovespa, IGBVL).

Table 2 presents the p-values of the statistics for the symmetry test, and the GARCH-type tests presented in Section 3. The vast majority of indices reject the symmetry assumption, which is a classical result in the financial literature. However, almost half of the thirty indices reject the hypothesis of a GARCH(1,1) specification, and eight reject the GJR-GARCH(1,1) or the TGARCH(1,1) model at the 5% level. Interestingly, all the indices that reject the hypothesis of short memory are from emerging markets. This suggests that the level of development of a financial market has implications on the persistence patterns exhibited by its assets. A possible explanation stems from the difficulty to invest in peripheral markets with fewer investors and with less liquid instruments, which leads to a slower integration of shocks and ultimately higher persistence.

In addition, we propose to study the ability of our model to forecast tail risk measures. We study six competing models, corresponding to the last columns of Table 2. The first column is the standard GARCH(1,1) process, the second is an ARCH(∞) model that nests the GARCH(1,1) similarly to equation (7), the third column corresponds to the GJR-GARCH(1,1), the fourth to the TARCH(∞) model (8) with $\delta = 2$, the fifth column corresponds to the TGARCH(1,1) and finally the last column corresponds to the APARCH(∞) model (8) with $\delta = 1$. For each specification, we fit the model on the sample from 1995 to 2017, and compute daily one-day ahead forecasts for the 95%-Value-at-Risk using the residuals obtained from each models

$$\mathrm{VaR}_{95\%} = \hat{\sigma}_{t+1}^{\delta} F_{\hat{\eta}_t}^{-1}(0.05) \text{ with } \hat{\eta}_t = \varepsilon_t / \hat{\sigma}_t$$

where $F_{\hat{\eta}_t}$ is a non parametric estimator of the distribution of the residuals. We thus obtain approximately 750 forecasts for each index ranging from January 2018 to December 2020. The last six columns of Table 2 give the frequency of violation of the VaR forecasts in each model for every index. Aside from the Bovespa index, all the competing models have a frequency of violations that is not statistically different from 5% using Kupiec's test[32] at the 95% confidence level. From this perspective, it thus seems difficult to choose between the studied specifications. We therefore propose to use the Model Confidence Set (MCS) procedure of Hansen, Lunde, and Nason[29] to select

[4] Data for the FTSE MIB and the MOEX start respectively in September 1997 and January 1998.

the best predictive models amongst our competitors. The idea of the MCS procedure is to sequentially eliminate competitors until the set of remaining models does not reject the hypothesis of equal predictive ability. To perform this series of tests, we used the R package developed by Bernardi and Catania[5] and used the asymmetric VaR loss function of González-Rivera, Lee, and Mishra[25] to compute the losses associated with the VaR forecasts. The results of the MCS procedures are also presented in Table 2. For each index, the models included in the Superior Set of Models (SSM) at the 80%-confidence level are marked with a star. A notable result is that for more than a third of the studied indices, the GARCH(1,1) and ARCH(∞) models are excluded from the SSM. This is a clear argument for the use of asymmetric models, even if the impact of asymmetry is less obvious on tail measures than on volatility. Surprisingly, however, asymmetric short memory models are often included in the SSM even if the hypothesis of nullity of γ_0 is rejected for some indices. Finally, the APARCH(∞) model with $\delta = 1$ is always included in the SSM, which seems to validate the pertinence of our model.

6 Concluding remarks

Although econometric models allowing for a strong persistence of the volatility of financial returns have been introduced in the academic literature for a long time, short memory models are still preferred by most practitioners. In this paper, we proposed an extension of the ARCH(∞) model of Robinson[36] to account for high persistence in power-transformed returns and conditional asymmetry. We proved the existence of a stationary solution and derived statistical inference results. In particular, we proved the consistency and asymptotic normality of QMLE. We showed that the APARCH(∞) representation nests some of the most used models in the financial industry and introduced a Portmanteau type goodness-of-fit test to verify the adequacy of such models. We derived test procedures for conditional asymmetry and to verify that GARCH(1,1)-type memory patterns are sufficient to model financial returns. In this regard, the results of the application on real data provide a remarkable argument for the use of moderate memory models when studying peripheral assets. We showed that in our database, most of the emerging markets equity indices exhibit a stronger persistence than the standard APARCH(1,1) allows for. The study of conditional Value-at-Risk measures seems to validate the pertinence of the proposed extensions. Although it would be of interest to derive asymptotic results for such quantities under stronger persistence, we leave this problem for future research.

Index	Country	Symmetry tests			GARCH-type tests			Frequency of rejections of the $\mathrm{VaR}_{95\%}$ (in %)					
		W_n^{sym}	R_n^{sym}	L_n^{sym}	W_n^{GARCH}	W_n^{GJR}	W_n^{TGARCH}	GARCH	ARCH(∞)	GJR	TARCH(∞)	TGARCH	APARCH(∞)
S&P 500	USA	**0.000**	**0.000**	**0.000**	0.958	0.991	0.996	5.04	5.04	4.24*	4.38	5.17*	5.17*
Nasdaq 100	USA	**0.000**	**0.000**	**0.000**	0.858	0.996	0.993	5.17*	5.17*	4.77*	4.77*	5.44*	5.44*
S&P/TSX	Canada	**0.000**	**0.000**	**0.000**	0.392	0.614	0.955	5.59*	5.59*	4.79*	4.79*	4.92*	4.92*
Mexico IPC	Mexico	**0.000**	**0.000**	**0.000**	**0.000**	0.156	0.066	5.60*	4.93*	5.20*	5.20*	4.93*	4.93*
FTSE 100	UK	**0.000**	**0.000**	**0.000**	0.204	0.875	0.890	5.15*	5.15*	5.15*	5.15*	4.88*	4.88*
DAX 30	Germany	**0.000**	**0.000**	**0.000**	0.966	0.995	0.999	5.56	5.56	4.90*	4.90*	5.30*	5.30*
CAC 40	France	**0.000**	**0.000**	**0.000**	0.943	1.000	0.997	5.75	5.88	4.71*	4.71*	5.36*	5.36*
SMI	Switzerland	**0.000**	**0.000**	**0.000**	0.930	0.993	0.991	4.67	4.67	5.07*	5.07*	5.34*	5.34*
AEX	Netherlands	**0.000**	**0.000**	**0.000**	0.960	0.995	0.998	5.75	5.75	5.10*	5.10*	5.36*	5.36*
FTSE MIB	Italy	**0.000**	**0.000**	**0.000**	0.947	0.993	0.996	5.80	5.80	5.01*	5.01	5.54*	5.54*
IBEX 35	Spain	**0.000**	**0.000**	**0.000**	0.953	0.992	0.998	5.49	5.49	4.71*	4.71*	4.84*	4.84*
MOEX	Russia	**0.003**	**0.000**	**0.000**	0.068	0.066	0.146	3.72*	3.72*	3.72*	3.59	3.86*	3.86*
WIG	Poland	**0.000**	**0.000**	**0.000**	**0.034**	0.160	0.158	5.50	5.63*	5.36*	5.23*	5.50*	5.50*
BUX	Hungary	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	3.92*	4.59*	4.19*	4.05*	4.05*	4.59*
TA125	Israel	**0.000**	**0.000**	**0.000**	**0.031**	0.942	0.972	5.31	5.18	5.31	5.31	5.18*	5.18*
Nikkei 225	Japan	**0.000**	**0.000**	**0.000**	0.625	0.978	0.987	4.81	4.81	4.54	4.54	4.95*	4.95*
KOSPI	South Korea	**0.000**	**0.000**	**0.000**	0.402	0.937	0.957	5.03*	5.03*	5.30	5.30	5.16*	5.16*
Hang Seng	Hong Kong	**0.000**	**0.000**	**0.000**	0.967	0.995	0.998	6.23*	6.23*	5.28*	5.28*	5.42*	5.42*
TAIEX	Taiwan	**0.000**	**0.000**	**0.000**	0.530	0.986	0.997	4.52	4.52	4.52	4.52	4.79*	4.79*
MSCI Singapore	Singapore	**0.000**	**0.000**	**0.000**	**0.035**	0.199	0.259	4.66*	5.06*	4.93*	5.06*	4.93*	4.93*
BSET	Thailand	**0.000**	**0.000**	**0.000**	**0.003**	**0.049**	0.110	6.02	6.29*	5.61*	5.75*	5.20*	5.20*
PSEi	Philippines	**0.001**	**0.000**	**0.000**	0.112	**0.004**	**0.021**	5.36*	5.36*	5.50*	5.36*	4.81*	4.95*
IDX	Indonesia	**0.001**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	5.81*	5.67*	4.84*	4.98*	4.98	4.98*
FTSE KLCI	Malaysia	**0.000**	**0.000**	**0.000**	**0.000**	**0.002**	**0.000**	6.55*	6.28*	6.14*	6.28*	5.87*	6.14*
NIFTY 50	India	**0.000**	**0.000**	**0.000**	**0.020**	0.440	0.656	5.02*	5.02*	4.75*	4.75*	4.61*	4.75*
ASX AO	Australia	**0.000**	**0.000**	**0.000**	**0.019**	0.643	0.778	5.93*	5.53*	5.93*	6.06*	6.32*	6.32*
MSCI NZ	New Zealand	**0.032**	0.054	**0.047**	**0.000**	**0.000**	**0.000**	5.05*	5.45*	4.92*	5.32*	5.19*	5.85*
Merval	Argentina	**0.005**	**0.000**	**0.000**	**0.000**	**0.000**	**0.002**	6.74	6.88	6.74*	6.74*	7.02*	6.74*
Bovespa	Brazil	**0.000**	**0.000**	**0.000**	0.187	0.994	0.989	3.64*	3.64*	3.64*	3.64*	3.78*	3.78*
S&P/BVL	Peru	**0.002**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	4.51*	5.17*	4.77*	4.77*	4.91*	5.31*

Table 2: *Symmetry tests, GARCH-memory tests, and validation of VaR forecasts on different equity indices*

References

[1] D. W. Andrews. Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*, 69(3):683–734, 2001.

[2] J.-M. Bardet and O. Wintenberger. Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes. *The Annals of Statistics*, 37(5B):2730–2759, 2009.

[3] I. Berkes, L. Horváth, and P. Kokoszka. Asymptotics for GARCH squared residual correlations. *Econometric Theory*, pages 515–540, 2003.

[4] I. Berkes, L. Horváth, and P. Kokoszka. GARCH processes: Structure and estimation. *Bernoulli*, 9(2):201–227, 2003.

[5] M. Bernardi and L. Catania. The model confidence set package for R. *International Journal of Computational Economics and Econometrics*, 8(2):144–158, 2018.

[6] E. Beutner, A. Heinemann, and S. Smeekes. A residual bootstrap for conditional value-at-risk. *arXiv preprint arXiv:1808.09125*, 2018.

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

[8] P. Billingsley. *Probability and measure.* John Wiley & Sons, 3rd edition, 1995.

[9] T. Bollerslev and H. Ole Mikkelsen. Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*, 73(1):151–184, 1996.

[10] G. E. Box and D. A. Pierce. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65(332):1509–1526, 1970.

[11] M. Carbon and C. Francq. Portmanteau goodness-of-fit test for asymmetric power GARCH models. *Austrian Journal of Statistics*, 40(1&2):55–64, 2011.

[12] G. Cavaliere, H. B. Nielsen, R. S. Pedersen, and A. Rahbek. Bootstrap inference on the boundary of the parameter space, with application to conditional volatility models. *Journal of Econometrics*, 2020.

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

[14] R. Douc, F. Roueff, and P. Soulier. On the existence of some ARCH(∞) processes. *Stochastic Processes and their Applications*, 118(5):755–761, 2008.

[15] C. Francq and L. Q. Thieu. QML inference for volatility models with covariates. *Econometric Theory*, 35(1):37–72, 2019.

[16] C. Francq and J.-M. Zakoïan. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli*, 10(4):605–637, 2004.

[17] C. Francq and J.-M. Zakoïan. Quasi-maximum likelihood estimation in garch processes when some coefficients are equal to zero. *Stochastic Processes and their Applications*, 117(9):1265–1284, 2007.

[18] C. Francq and J.-M. Zakoïan. Testing the nullity of GARCH coefficients: correction of the standard tests and relative efficiency comparisons. *Journal of the American Statistical Association*, 104(485):313–324, 2009.

[19] C. Francq and J.-M. Zakoïan. *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons, 2nd edition, 2019.

[20] R. Giacomini, D. N. Politis, and H. White. A warp-speed method for conducting monte carlo experiments involving bootstrap estimators. *Econometric Theory*, 29(3):567–589, 2013.

[21] L. Giraitis, P. Kokoszka, and R. Leipus. Stationary ARCH models: Dependence structure and central limit theorem. *Econometric Theory*, 16(1):3–22, 2000.

[22] L. Giraitis and P. M. Robinson. Whittle estimation of ARCH models. *Econometric Theory*, 17(3):608–631, 2001.

[23] L. Giraitis, D. Surgailis, and A. Škarnulis. Stationary integrated ARCH(∞) and AR(∞) processes with finite variance. *Econometric Theory*, 34(6):1159–1179, 2018.

[24] L. R. Glosten, R. Jagannathan, and D. E. Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801, 1993.

[25] G. González-Rivera, T.-H. Lee, and S. Mishra. Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of forecasting*, 20(4):629–645, 2004.

[26] C. Gouriéroux and A. Monfort. *Statistics and Econometric Models*, volume 2 of *Themes in Modern Econometrics*. Cambridge University Press, 1995.

[27] C. M. Hafner and A. Preminger. On asymptotic theory for ARCH (∞) models. *Journal of Time Series Analysis*, 38(6):865–879, 2017.

[28] T. Hamadeh and J.-M. Zakoïan. Asymptotic properties of LS and QML estimators for a class of nonlinear GARCH processes. *Journal of Statistical Planning and Inference*, 141(1):488–507, 2011.

[29] P. R. Hansen, A. Lunde, and J. M. Nason. The model confidence set. *Econometrica*, 79(2):453–497, 2011.

[30] J. Hidalgo and P. Zaffaroni. A goodness-of-fit test for ARCH(∞) models. *Journal of Econometrics*, 141(2):973–1013, 2007.

[31] V. Kazakevičius and R. Leipus. On stationarity in the ARCH(∞) model. *Econometric Theory*, 18(1):1–16, 2002.

[32] P. Kupiec. Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, 3(2), 1995.

[33] W. K. Li and T. Mak. On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. *Journal of Time Series Analysis*, 15(6):627–636, 1994.

[34] O. Linton and E. Mammen. Estimating semiparametric ARCH(∞) models by kernel smoothing methods. *Econometrica*, 73(3):771–836, 2005.

[35] D. B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370, 1991.

[36] P. M. Robinson. Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics*, 47(1):67–84, 1991.

[37] P. M. Robinson and P. Zaffaroni. Pseudo-maximum likelihood estimation of ARCH(∞) models. *The Annals of Statistics*, 34(3):1049–1074, 2006.

[38] P. Zaffaroni. Stationarity and memory of ARCH(∞) models. *Econometric Theory*, pages 147–160, 2004.

[39] P. Zaffaroni. Whittle estimation of EGARCH and other exponential volatility models. *Journal of Econometrics*, 151(2):190–200, 2009.

[40] J.-M. Zakoïan. Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18(5):931–955, 1994.

Appendix A Proofs and technical results

This appendix provides the proofs and technical results in a condensed manner. A more detailed version is available in a supplement to this paper.

A.1 Existence of a stationary APARCH(∞) solution

We develop in this section the proof of Theorem 1. The proof is based on a Volterra expansion and, in this sense, follows the work of Giraitis, Kokoszka and Leipus[21], Kazakevičius and Leipus[31], and Douc, Roueff and Soulier[14].

Proof of Theorem 1. First, let us remark that $\sigma_t > 0$ which implies for any t, $\mathbb{1}_{\varepsilon_t \geq 0} = \mathbb{1}_{\eta_t \geq 0}$, and consider the random variable

$$S_t = \omega + \omega \sum_{k=1}^{\infty} \sum_{i_1,\dots,i_k \geq 1} a_{i_1,t-i_1} \dots a_{i_k,t-i_1-\dots-i_k} \, |\eta_{t-i_1}|^{\delta} \dots |\eta_{t-i_1-\dots-i_k}|^{\delta}$$

defined in $[0,+\infty]$. From the independence of (η_t) and since $s \in (0,1]$, we have

$$\begin{aligned} \mathbb{E}S_t^s \leq \; & \omega^s + \omega^s \sum_{k=1}^{\infty} \sum_{i_1,\dots,i_k \geq 1} \mathbb{E}\left(\left[(\alpha_{i_1}^+)^s \, \mathbb{1}_{\eta_{t-i_1} \geq 0} + (\alpha_{i_1}^-)^s \, \mathbb{1}_{\eta_{t-i_1} < 0} \right] |\eta_{t-i_1}|^{\delta s} \right) \dots \\ & \mathbb{E}\left(\left[(\alpha_{i_k}^+)^s \, \mathbb{1}_{\eta_{t-i_1-\dots-i_k} \geq 0} + (\alpha_{i_k}^-)^s \, \mathbb{1}_{\eta_{t-i_1-\dots-i_k} < 0} \right] |\eta_{t-i_1-\dots-i_k}|^{\delta s} \right), \end{aligned}$$

and thus

$$\mathbb{E}S_t^s \leq \omega^s \left[1 + \sum_{k=1}^{\infty} (A_s^+ \mu_{\delta s}^+ + A_s^- \mu_{\delta s}^-)^k \right] \leq \frac{\omega^s}{1 - (A_s^+ \mu_{\delta s}^+ + A_s^- \mu_{\delta s}^-)} < \infty,$$

whence S_t is finite almost surely. In addition, we have

$$\sum_{i=1}^{\infty} a_{i,t-i} S_{t-i} |\eta_{t-i}|^{\delta} = \omega \sum_{k=0}^{\infty} \sum_{i_0,\dots,i_k \geq 1} a_{i_0,t-i_0} \dots a_{i_k,t-i_0-\dots-i_k} \, |\eta_{t-i_0}|^{\delta} \dots |\eta_{t-i_0-\dots-i_k}|^{\delta}$$

and thus we obtain the recursive equation $S_t = \omega + \sum_{i=1}^{\infty} a_{i,t-i} S_{t-i} |\eta_{t-i}|^{\delta}$. By setting $\varepsilon_t = S_t^{1/\delta} \eta_t$, we obtain a strictly stationary and nonanticipative solution of (3) and $\mathbb{E}|\varepsilon_t|^{\delta s} \leq \mu_{\delta s} \omega^s / (1 - (A_s^+ \mu_{\delta s}^+ + A_s^- \mu_{\delta s}^-))$. In addition, Theorem 36.4 in Billingsley[8] entails the ergodicity of the stationary solution.

Now denote by (ε_t^*) any strictly stationary and nonanticipative solution of (3) such that

$\mathbb{E}|\varepsilon_t^*|^{\delta s} < \infty$. For all $q \geq 1$, by q recursive substitutions of the $\varepsilon_{t-i}^{*\delta}$, we obtain

$$\begin{aligned}
\sigma_t^\delta &= \omega + \sum_{i=1}^{\infty} a_{i,t-i}|\varepsilon_{t-i}|^{*\delta} \\
&= \left\{ \omega + \omega \sum_{k=1}^{q} \sum_{i_1,\dots,i_k \geq 1} a_{i_1,t-i_1} \dots a_{i_k,t-i_1-\dots-i_k} \; |\eta_{t-i_1}|^\delta \dots |\eta_{t-i_1-\dots-i_k}|^\delta \right\} \\
&+ \sum_{i_1,\dots,i_{q+1} \geq 1} a_{i_1,t-i_1} \dots a_{i_{q+1},t-i_1-\dots-i_{q+1}} \; |\eta_{t-i_1}|^\delta \dots |\eta_{t-i_1-\dots-i_q}|^\delta |\varepsilon_{t-i_1-\dots-i_{q+1}}|^{*\delta} \\
&:= \{S_{t,q}\} + R_{t,q}.
\end{aligned}$$

Since (ε_t^*) is nonanticipative, it is independent of $\eta_{t'}$ for any $t' > t$. Hence, since $s \in (0,1]$,

$$\mathbb{E}R_{t,q}^s \leq (A_s^+ \mu_{\delta s}^+ + A_s^- \mu_{\delta s}^-)^q \; \left(A_s^+ \mathbb{E}|\mathbb{1}_{\eta_t \geq 0} \varepsilon_t^*|^{\delta s} + A_s^- \mathbb{E}|\mathbb{1}_{\eta_t < 0} \varepsilon_t^*|^{\delta s} \right)$$

Since $A_s^+ \mu_{\delta s}^+ + A_s^- \mu_{\delta s}^- < 1$, we have $\sum_{q \geq 1} \mathbb{E}R_{t,q}^s < \infty$, whence $R_{t,q}$ tends to 0 almost surely as $q \to \infty$. Furthermore, $S_{t,q}$ tends to S_t almost surely as $q \to \infty$, which implies $\sigma_t^\delta = S_t$ almost surely and yields $\varepsilon_t^* = \varepsilon_t$ almost surely, hence concluding the proof. □

A.2 Statistical inference of an APARCH(∞) process

We develop in this section the proofs of the main results of Section 2 on consistency and asymptotic normality of the QMLE in our model. Note that in the following proofs, it will not be restrictive to assume $\rho < 1$.

Let us define the theoretical criterion

$$Q_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} l_t(\boldsymbol{\theta}), \;\; l_t(\boldsymbol{\theta}) = \log \sigma_t^2(\boldsymbol{\theta}) + \frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})}, \;\; \hat{\boldsymbol{\theta}}_{\boldsymbol{n}} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{Argmin}} \; Q_n(\boldsymbol{\theta}).$$

The theoretical QMLE $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}$ is infeasible, and we will thus study the feasible estimator $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}$, which is conditional to initial values. We will show that the choice of the initial values is unimportant for the asymptotic properties of the QMLE.

In the following, we denote $\mathcal{I}^+(\boldsymbol{\phi})$ (respectively $\mathcal{I}^-(\boldsymbol{\phi})$) the sets $\{i$ such that $\alpha_i^{+(-)}(\boldsymbol{\phi}) \neq 0\}$, and we define $\mathcal{I}_t^+$ (respectively $\mathcal{I}_t^-$) as $\mathcal{I}_t^{+(-)} = \{i$ such that $\varepsilon_{t-i} \geq 0$ (resp. $< 0)\}$, yielding the following rewriting of (9) as

$$\sigma_t^\delta(\boldsymbol{\theta_0}) = \omega_0 + \sum_{i \in \mathcal{I}_t^+} \alpha_i^+(\boldsymbol{\phi_0})|\varepsilon_{t-i}|^\delta + \sum_{j \in \mathcal{I}_t^-} \alpha_j^-(\boldsymbol{\phi_0})|\varepsilon_{t-j}|^\delta. \tag{20}$$

We first state and prove the property mentioned in the remark about assumption **A5**.

Proposition 2. *Under assumptions A1-A4, if there exists* $0 < \tau < \rho - (d+1)^{-1}$ *such that*

$$\sup_{i \in \mathcal{I}^+(\boldsymbol{\phi_0})} \sup_{\phi \in \boldsymbol{\Phi}} \frac{\alpha_i^+(\boldsymbol{\phi_0})}{(\alpha_i^+)^{1-\tau}(\boldsymbol{\phi})} \leq K \text{ and } \sup_{i \in \mathcal{I}^-(\boldsymbol{\phi_0})} \sup_{\phi \in \boldsymbol{\Phi}} \frac{\alpha_i^-(\boldsymbol{\phi_0})}{(\alpha_i^-)^{1-\tau}(\boldsymbol{\phi})} \leq K, \tag{21}$$

then

$$\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})} < \infty.$$

Condition (21) limits the set of eligible functions $\alpha_i^{+(-)}$ that allows the integrability of $\varepsilon_t^2/\sigma_t^2(\boldsymbol{\theta})$ without additional moment assumptions on ε_t. In particular, functions such that $\inf_{\phi \in \boldsymbol{\Phi}} \alpha_i^{+(-)}(\boldsymbol{\phi}) = 0$ when $\alpha_i^{+(-)}(\boldsymbol{\phi_0}) \neq 0$ are excluded. Note however that the latter case does not exclude every $\alpha_i^{+(-)}$ that can be equal to 0. For example, $\alpha_i(\boldsymbol{\phi}) = [(1 + \sin(i\pi/4))/2]\gamma i^{-(d+1)}$ where $\boldsymbol{\phi} = (\gamma, d)$ and $\boldsymbol{\Phi} \subset (0, \infty)^2$ is periodically equal to 0 and still verifies (21). The requirement for higher moment for ε_t is also required for the estimation of ARCH(q) models or GARCH models when coefficients are equal to 0 (see for example Francq and Zakoïan[17] or Cavaliere et al[12]).

Proof of Proposition 2. Let us first note that if $c > 0$ and for all i in a set I, $a_i \geq 0$ and $b_i \geq 0$ then $\frac{\sum_{i \in I} a_i}{c + \sum_{j \in I} b_j} \leq \sum_{i \in I} \frac{a_i}{c + b_i}$. Since $\omega_L > 0$ and for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ we have $\alpha_i^{+(-)}(\boldsymbol{\theta}) \geq 0$, using the previous elementary inequality and the fact that for any s>0, we have $x/(1+x) \leq x^s$, equation (20) gives

$$\begin{aligned}
&\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \frac{\sigma_t^\delta(\boldsymbol{\theta_0})}{\sigma_t^\delta(\boldsymbol{\theta})} \\
\leq\ & \frac{\omega_0}{\omega} + \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \sum_{i \in \mathcal{I}_t^+ \cap \mathcal{I}^+(\boldsymbol{\phi_0})} \frac{\alpha_i^+(\boldsymbol{\phi_0})|\varepsilon_{t-i}|^\delta}{\omega + \alpha_i^+(\boldsymbol{\phi})|\varepsilon_{t-i}|^\delta} + \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \sum_{i \in \mathcal{I}_t^- \cap \mathcal{I}^-(\boldsymbol{\phi_0})} \frac{\alpha_i^-(\boldsymbol{\phi_0})|\varepsilon_{t-i}|^\delta}{\omega + \alpha_i^-(\boldsymbol{\phi})|\varepsilon_{t-i}|^\delta} \\
\leq\ & K + K \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \sum_{i \in \mathcal{I}_t^+ \cap \mathcal{I}^+(\boldsymbol{\phi_0})} (\alpha_i^+)^{s-\tau}(\boldsymbol{\phi})|\varepsilon_{t-i}|^{\delta s} + K \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \sum_{i \in \mathcal{I}_t^- \cap \mathcal{I}^-(\boldsymbol{\phi_0})} (\alpha_i^-)^{s-\tau}(\boldsymbol{\phi})|\varepsilon_{t-i}|^{\delta s} \\
\leq\ & K + K \sum_{i=1}^{\infty} i^{-(d+1)(s-\tau)}|\varepsilon_{t-i}|^{\delta s}
\end{aligned}$$

using assumptions **A3(ii)** and (21). This yields

$$\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \frac{\sigma_t^\delta(\boldsymbol{\theta_0})}{\sigma_t^\delta(\boldsymbol{\theta})} \leq K + \omega^{-s} K' \sum_{i=1}^{\infty} i^{-(d+1)(s-\tau)} \mathbb{E}_{\boldsymbol{\theta_0}} |\varepsilon_{t-i}|^{\delta s}.$$

By taking $s = \rho$ we have that $(d+1)(s-\tau) > 1$ by assumption **A4**, we thus obtain

$$\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \frac{\sigma_t^\delta(\boldsymbol{\theta_0})}{\sigma_t^\delta(\boldsymbol{\theta})} < \infty.$$

If $\delta < 2$, using Minkowski inequality and assumption **A4**, we obtain

$$\left[\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left(\frac{\sigma_t^\delta(\boldsymbol{\theta_0})}{\sigma_t^\delta(\boldsymbol{\theta})}\right)^{2/\delta}\right]^{\delta/2} \leq K + K\sum_{i=1}^{\infty} i^{-(d+1)(\rho-\tau)} \left[\mathbb{E}_{\boldsymbol{\theta_0}}|\varepsilon_{t-i}|^{2\rho}\right]^{\delta/2} < \infty$$

from assumption **A3(ii)**, and if $\delta \geq 2$, Jensen inequality yields $\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})} < \infty$, which concludes the proof.

□

The following lemma shows the asymptotic irrelevance of the initial values on Q_n.

Lemma 1. *Under assumptions A1-A5,* $\lim_{n\to\infty} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |Q_n(\boldsymbol{\theta}) - \tilde{Q}_n(\boldsymbol{\theta})| = 0$.

Proof of Lemma 1. Consider

$$Q_n(\boldsymbol{\theta}) - \tilde{Q}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} \log \frac{\sigma_t^2(\boldsymbol{\theta})}{\tilde{\sigma}_t^2(\boldsymbol{\theta})} + \frac{1}{n}\sum_{t=1}^{n} \varepsilon_t^2 \left(\frac{1}{\sigma_t^2(\boldsymbol{\theta})} - \frac{1}{\tilde{\sigma}_t^2(\boldsymbol{\theta})}\right) := A_n(\boldsymbol{\theta}) + B_n(\boldsymbol{\theta}).$$

and remark that $\sigma_t^2(\boldsymbol{\theta}) \geq \tilde{\sigma}_t^2(\boldsymbol{\theta})$, since we have $\sigma_t^\delta(\boldsymbol{\theta}) = \tilde{\sigma}_t^\delta(\boldsymbol{\theta}) + \sum_{i=t}^{\infty} a_{i,t-i}(\boldsymbol{\phi})|\varepsilon_{t-i}|^\delta$. We denote $\chi_t = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\sigma_t^\delta(\boldsymbol{\theta}) - \tilde{\sigma}_t^\delta(\boldsymbol{\theta})|$, and we have from assumption **A3(ii)**

$$\chi_t = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{i=t}^{\infty} a_{i,t-i}(\boldsymbol{\phi})|\varepsilon_{t-i}|^\delta \leq K\sum_{i=0}^{\infty} (i+t)^{-(d+1)}|\varepsilon_{-i}|^\delta,$$

whence $\mathbb{E}\chi_t^\rho \leq K\sum_{i=0}^{\infty} (i+t)^{-(d+1)\rho}\ \mathbb{E}|\varepsilon_{-i}|^{\delta\rho}$. Since from assumption **A4**, $\mathbb{E}\,|\varepsilon_t|^{\delta\rho} < \infty$, with $\rho(d+1) > 1$, and since for any $k > 1$ we have $\int_t^\infty x^{-k}dx = (k-1)^{-1}t^{-k+1}$, we obtain $\mathbb{E}\chi_t^\rho \leq Kt^{-(d+1)\rho+1}$. This shows that χ_t has a finite moment of order ρ and thus is finite almost surely. Furthermore, since $\rho(d+1) > 1$, the dominated convergence theorem entails $\lim_{t\to\infty} \chi_t = 0$ almost surely. Then, we have

$$|A_n(\boldsymbol{\theta})| = \frac{2}{\delta n}\sum_{t=1}^{n} \log\left[1 + \frac{\sigma_t^\delta(\boldsymbol{\theta}) - \tilde{\sigma}_t^\delta(\boldsymbol{\theta})}{\tilde{\sigma}_t^2(\boldsymbol{\theta})}\right] \leq \frac{K}{n}\sum_{t=1}^{n} \sigma_t^\delta(\boldsymbol{\theta}) - \tilde{\sigma}_t^\delta(\boldsymbol{\theta})$$

since $\log(1+x) \leq x$ for $x \geq 0$ and, for all t, $\tilde{\sigma}_t^2(\boldsymbol{\theta}) \geq \omega$. Therefore, we obtain $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |A_n(\boldsymbol{\theta})| \leq Kn^{-1}\sum_{t=1}^{n} \chi_t$ and from Cesaro mean convergence theorem, we obtain $\lim_{n\to\infty} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |A_n(\boldsymbol{\theta})| = 0$ almost surely.

Consider now

$$|B_n(\boldsymbol{\theta})| \leq \frac{K}{n}\sum_{t=1}^{n}\eta_t^2 \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})} \frac{\max[\sigma_t^{2-\delta}(\boldsymbol{\theta}), \tilde{\sigma}_t^{2-\delta}(\boldsymbol{\theta})]}{\tilde{\sigma}_t^2(\boldsymbol{\theta})} \left[\sigma_t^{\delta}(\boldsymbol{\theta}) - \tilde{\sigma}_t^{\delta}(\boldsymbol{\theta})\right]$$

whence

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |B_n(\boldsymbol{\theta})| \quad \leq \frac{K}{n}\sum_{t=1}^{n}\eta_t^2 \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})} \frac{\max[\sigma_t^{2-\delta}(\boldsymbol{\theta}), \tilde{\sigma}_t^{2-\delta}(\boldsymbol{\theta})]}{\tilde{\sigma}_t^2(\boldsymbol{\theta})}\chi_t.$$

If $\delta \geq 2$, $\sigma_t^{2-\delta}(\boldsymbol{\theta}) \leq \tilde{\sigma}_t^{2-\delta}(\boldsymbol{\theta})$ and since $\tilde{\sigma}_t^{-\delta}(\boldsymbol{\theta}) \leq \omega^{-\delta} < \infty$ from **A1**, we have

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |B_n(\boldsymbol{\theta})| \quad \leq \frac{K}{n}\sum_{t=1}^{n}\eta_t^2 \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}[\tilde{\sigma}_t^2(\boldsymbol{\theta})]^{-\delta}\chi_t \leq \frac{K}{n}\sum_{t=1}^{n}\eta_t^2 \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\chi_t.$$

If $\delta < 2$, $\sigma_t^{2-\delta}(\boldsymbol{\theta}) \geq \tilde{\sigma}_t^{2-\delta}(\boldsymbol{\theta})$ and we have

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |B_n(\boldsymbol{\theta})| \leq \frac{K}{n}\sum_{t=1}^{n}\eta_t^2 \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})} \frac{\sigma_t^2(\boldsymbol{\theta})}{\tilde{\sigma}_t^2(\boldsymbol{\theta})}\chi_t.$$

From assumptions **A3(ii)** and **A4**, we have

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \eta_t^2 \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})} \frac{\sigma_t^2(\boldsymbol{\theta})}{\tilde{\sigma}_t^2(\boldsymbol{\theta})} &= \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \eta_t^2 \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})} \left[\frac{\sigma_t^{\delta}(\boldsymbol{\theta})}{\tilde{\sigma}_t^{\delta}(\boldsymbol{\theta})}\right]^{2/\delta} \\
&\leq K \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \eta_t^2 \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})} \left[1 + \sum_{i=0}^{\infty} i^{-d-1}|\varepsilon_{-i}|^{\delta}\right]^{2/\delta} \\
&\leq K \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \eta_t^2 \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}
\end{aligned} \tag{22}$$

where K is finite almost surely and does not depend on t since $\sum_{i=0}^{\infty} i^{-(d+1)}|\varepsilon_{-i}|^{\delta}$ admits a moment of order ρ and thus is finite almost surely.

Thus, we have

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |B_n(\boldsymbol{\theta})| \leq \frac{K}{n}\sum_{t=1}^{n}\eta_t^2 \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\chi_t.$$

By ergodicity and independance of η_t^2 with σ_t^2, we have that $\frac{1}{n}\sum_{t=1}^{n}\eta_t^2 \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}$ tends to $\mathbb{E}\eta_t^2 \ \mathbb{E} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}$ almost surely as n tends to infinity. Since $\chi_t \to 0$ almost surely and $\mathbb{E}\eta_t^2 \ \mathbb{E} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})} < \infty$ by assumptions **A2** and **A5**, from Toeplitz lemma we obtain $\lim_{n\to\infty} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |B_n(\boldsymbol{\theta})| = 0$ almost surely, which concludes the proof. □

***Proof of Theorem** 2.* The proof of the strong consistency of the QMLE is achieved by proving the four following intermediate results and using a compactness argument:

(a) $\lim_{n\to\infty} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |Q_n(\boldsymbol{\theta}) - \tilde{Q}_n(\boldsymbol{\theta})| = 0$

(b) $(\exists t \in \mathbb{Z}$ such that $\sigma_t^\delta(\boldsymbol{\theta}) = \sigma_t^\delta(\boldsymbol{\theta_0})$ a.s.$)\ \Rightarrow \boldsymbol{\theta} = \boldsymbol{\theta_0}$

(c) $\mathbb{E}_{\boldsymbol{\theta_0}}|l_t(\boldsymbol{\theta_0})| < \infty$, and if $\boldsymbol{\theta} \neq \boldsymbol{\theta_0}$, $\mathbb{E}_{\boldsymbol{\theta_0}} l_t(\boldsymbol{\theta}) > \mathbb{E}_{\boldsymbol{\theta_0}} l_t(\boldsymbol{\theta_0})$

(d) For any $\boldsymbol{\theta} \neq \boldsymbol{\theta_0}$, there exists a neighborhood $V(\boldsymbol{\theta})$ such that $\liminf_{n\to\infty} \inf_{\boldsymbol{\theta}^*\in V(\boldsymbol{\theta})} \tilde{Q}_n(\boldsymbol{\theta}^*) > \mathbb{E}_{\boldsymbol{\theta_0}} l_1(\boldsymbol{\theta_0})$ a.s.

(a) is directly obtained from Lemma 1.

Now let $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, such that, for some $t \in \mathbb{Z}$, we have $\sigma_t^\delta(\boldsymbol{\theta}) = \sigma_t^\delta(\boldsymbol{\theta_0})$ almost surely. Assume $\boldsymbol{\theta} \neq \boldsymbol{\theta_0}$, and suppose that

$$\alpha_1^+(\boldsymbol{\phi_0})\mathbb{1}_{\varepsilon_{t-1}\geq 0} + \alpha_1^-(\boldsymbol{\phi_0})\mathbb{1}_{\varepsilon_{t-1}<0} \neq \alpha_1^+(\boldsymbol{\phi})\mathbb{1}_{\varepsilon_{t-1}\geq 0} + \alpha_1^-(\boldsymbol{\phi})\mathbb{1}_{\varepsilon_{t-1}<0}. \tag{23}$$

Then $\left(\left[\alpha_1^+(\boldsymbol{\phi_0}) - \alpha_1^+(\boldsymbol{\phi})\right]\mathbb{1}_{\eta_{t-1}\geq 0} + \left[\alpha_1^-(\boldsymbol{\phi_0}) - \alpha_1^-(\boldsymbol{\phi})\right]\mathbb{1}_{\eta_{t-1}<0}\right)|\eta_{t-1}|^\delta$ belongs to the σ-field $\mathcal{F}_{t-2}$ generated by $\{\eta_s : s \leq t-2\}$ and thus, by independence, is almost surely constant. Since, from **A2**, η_1 takes at least two positive (respectively negative) values, this implies almost surely $\alpha_1^+(\boldsymbol{\phi_0}) = \alpha_1^+(\boldsymbol{\phi})$ and $\alpha_1^-(\boldsymbol{\phi_0}) = \alpha_1^-(\boldsymbol{\phi})$, which contradicts (23). Recursively, we obtain that $\sigma_t^2(\boldsymbol{\theta}) = \sigma_t^2(\boldsymbol{\theta_0})$ implies that, for all i, $\alpha_i^+(\boldsymbol{\phi_0}) = \alpha_i^+(\boldsymbol{\phi})$ and $\alpha_i^-(\boldsymbol{\phi_0}) = \alpha_i^-(\boldsymbol{\phi})$ and thus, from assumption **A3(i)**, $\boldsymbol{\phi}^+ = \boldsymbol{\phi_0}^+$ and $\boldsymbol{\phi}^- = \boldsymbol{\phi_0}^-$ a.s., whence $\omega = \omega_0$ a.s., and thus $\boldsymbol{\theta} = \boldsymbol{\theta_0}$ almost surely, which proves **(b)**.

We now turn to **(c)**. First, notice that, even if the limit criterion may not be integrable at some point of $\boldsymbol{\Theta}$, it is well defined in $\mathbb{R}\cup\{+\infty\}$. Indeed

$$\mathbb{E}_{\boldsymbol{\theta_0}}\left[l_t^-(\boldsymbol{\theta})\right] = \mathbb{E}_{\boldsymbol{\theta_0}} \max[0; -l_t(\boldsymbol{\theta})] \leq \mathbb{E}_{\boldsymbol{\theta_0}} \max\left[0; -\log\sigma_t^2(\boldsymbol{\theta})\right] < \infty.$$

Furthermore, we can show that it is integrable at $\boldsymbol{\theta_0}$. Using Jensen inequality and assumption **A3(ii)**, we obtain

$$\mathbb{E}_{\boldsymbol{\theta_0}}\left[l_t(\boldsymbol{\theta_0})\right] = 1 + \mathbb{E}_{\boldsymbol{\theta_0}} \log\sigma_t^2(\boldsymbol{\theta_0}) \leq 1 + \frac{2}{\delta\rho}\log\left(\omega^\rho + K\sum_{i=1}^\infty i^{-(d+1)\rho}\mathbb{E}_{\boldsymbol{\theta_0}}|\varepsilon_{t-i}|^{\delta\rho}\right) < \infty$$

since, from assumption **A4**, $\mathbb{E}\,|\varepsilon_t|^{\delta\rho} < \infty$ and $\rho(d+1) > 1$. Thus, $\mathbb{E}_{\boldsymbol{\theta_0}}|l_t(\boldsymbol{\theta_0})|$ is well defined in $\mathbb{R}$. In addition, we have

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\theta_0}}\left[l_t(\boldsymbol{\theta})\right]-\mathbb{E}_{\boldsymbol{\theta_0}}\left[l_t(\boldsymbol{\theta_0})\right] &= \mathbb{E}_{\boldsymbol{\theta_0}}\left[\log\frac{\sigma_t^2(\boldsymbol{\theta})}{\sigma_t^2(\boldsymbol{\theta_0})}\right]+\mathbb{E}_{\boldsymbol{\theta_0}}\left[\frac{\sigma_t^2(\boldsymbol{\theta_0})\eta_t^2}{\sigma_t^2(\boldsymbol{\theta})}-\eta_t^2\right]\\ &\geq -\log\left[\mathbb{E}_{\boldsymbol{\theta_0}}\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\right]+\mathbb{E}_{\boldsymbol{\theta_0}}\left[\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\right]-1\geq 0\end{aligned}$$

since, for any $x>0$, $\log x\leq x-1$. We can conclude by noticing that $\mathbb{E}_{\boldsymbol{\theta_0}}[l_t(\boldsymbol{\theta})]=\mathbb{E}_{\boldsymbol{\theta_0}}[l_t(\boldsymbol{\theta_0})]$ if and only if $\dfrac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}=1$ almost surely, and thus, by identifiability of the parameter, if and only if $\boldsymbol{\theta}=\boldsymbol{\theta_0}$.
Finally, the proof of **(d)** is similar to the one presented in Francq and Zakoïan[19] and is left out of this paper the sake of brevity.
The conclusion of the proof follows from the previous four intermediate results and a compactness argument similar to [19]. □

We now state and prove the property mentioned in the remark about assumption **A11**.

Proposition 3. *Under assumptions A1-A4, if for all $\tau>0$, there exists a neighborhood $V(\boldsymbol{\theta_0})$ of $\boldsymbol{\theta_0}$ such that*

$$\sup_{i\in\mathcal{I}^+(\boldsymbol{\phi_0})}\ \sup_{\phi\in V(\boldsymbol{\phi_0})}\frac{\alpha_i^+(\boldsymbol{\phi_0})}{(\alpha_i^+)^{1-\tau}(\boldsymbol{\phi})}\leq K \text{ and } \sup_{i\in\mathcal{I}^-(\boldsymbol{\phi_0})}\ \sup_{\phi\in V(\boldsymbol{\phi_0})}\frac{\alpha_i^-(\boldsymbol{\phi_0})}{(\alpha_i^-)^{1-\tau}(\boldsymbol{\phi})}\leq K. \tag{24}$$

then, for all k>0, there exists some neighborhood $V(\boldsymbol{\theta_0})$ of $\boldsymbol{\theta_0}$ such that

$$\mathbb{E}_{\boldsymbol{\theta_0}}\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})}\left[\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\right]^k<\infty.$$

Proof of Proposition 3. For all $s\in(0,1]$ and $k>s$, (10) and Hölder inequality yield

$$\begin{aligned}\sigma_t^\delta(\boldsymbol{\theta_0}) &= \omega_0\omega^{\frac{s}{k}-1}\omega^{1-\frac{s}{k}}+\sum_{i=1}^{\infty}a_{i,t-i}(\boldsymbol{\phi_0})a_{i,t-i}^{\frac{s}{k}-1}(\boldsymbol{\phi})a_{i,t-i}^{1-\frac{s}{k}}(\boldsymbol{\phi})|\varepsilon_{t-i}|^{\delta\frac{s}{k}}|\varepsilon_{t-i}|^{\delta-\delta\frac{s}{k}}\\ &\leq\left(\sum_{i=0}^{\infty}x_i^{\frac{k}{s}}\right)^{\frac{s}{k}}\left(\sum_{i=0}^{\infty}y_i^{\frac{1}{1-s/k}}\right)^{1-\frac{s}{k}}\\ &\leq K\left[\omega_0^{\frac{k}{s}}\omega^{1-\frac{k}{s}}+\sum_{i=1}^{\infty}a_{i,t-i}^{\frac{k}{s}}(\boldsymbol{\phi_0})a_{i,t-i}^{1-\frac{k}{s}}(\boldsymbol{\phi})|\varepsilon_{t-i}|^{\delta}\right]^{\frac{s}{k}}\left[\sigma_t^\delta(\boldsymbol{\theta})\right]^{1-\frac{s}{k}}.\end{aligned}$$

with $x_0=\omega_0\omega^{\frac{s}{k}-1}$, $x_i=a_{i,t-i}(\boldsymbol{\phi_0})a_{i,t-i}^{\frac{s}{k}-1}(\boldsymbol{\phi})|\varepsilon_{t-i}|^{\delta\frac{s}{k}}$, $y_0=\omega^{1-\frac{s}{k}}$, $y_i=a_{i,t-i}^{1-\frac{s}{k}}(\boldsymbol{\phi})|\varepsilon_{t-i}|^{\delta-\delta\frac{s}{k}}$.
Since $\left[\sigma_t^\delta(\boldsymbol{\theta})\right]^{-\frac{s}{k}}\leq K$, we obtain $\left[\sigma_t^\delta(\boldsymbol{\theta_0})/\sigma_t^\delta(\boldsymbol{\theta})\right]^k\leq$

$$K\left[1+\sum_{\mathcal{I}_t^+\cap\mathcal{I}^+(\boldsymbol{\phi_0})}\frac{(\alpha_i^+)^k(\boldsymbol{\phi_0})}{(\alpha_i^+)^k(\boldsymbol{\phi})}(\alpha_i^+)^s(\boldsymbol{\phi})|\varepsilon_{t-i}|^{\delta s}+\sum_{\mathcal{I}_t^-\cap\mathcal{I}^-(\boldsymbol{\phi_0})}\frac{(\alpha_i^-)^k(\boldsymbol{\phi_0})}{(\alpha_i^-)^k(\boldsymbol{\phi})}(\alpha_i^-)^s(\boldsymbol{\phi})|\varepsilon_{t-i}|^{\delta s}\right],$$

whence, from (24) and assumptions **A3(ii)** and **A4**, by taking $s = \rho$, there exists a neighborhood such that

$$\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta_0})} \left[\frac{\sigma_t^\delta(\boldsymbol{\theta_0})}{\sigma_t^\delta(\boldsymbol{\theta})} \right]^k \leq K \left[1 + \sum_{i=1}^{\infty} i^{-(d+1)(\rho - k\tau)} \mathbb{E}_{\boldsymbol{\theta_0}} |\varepsilon_{t-i}|^{\delta\rho} \right] < \infty.$$

Indeed, from the arbitrariness of τ, we can find a τ such that $(d+1)(\rho - k\tau) > 1$. $\square$

Before developing the proofs of Theorems 3 and 4, it is useful to state the following lemmas. Note that the function $l_t(\boldsymbol{\theta})$ may be non-defined in a neighborhood of $\boldsymbol{\theta_0}$ when $\boldsymbol{\theta_0} \in \partial\Theta$ since the volatility process $\sigma_t^\delta(\boldsymbol{\theta})$ can take negative values. For ease of notation, we denote by $\partial\sigma_t^\delta(\boldsymbol{\theta_0})/\partial\boldsymbol{\theta}$ the vector of partial derivatives $(\partial\sigma_t^\delta(\boldsymbol{\theta_0})/\partial\theta_i)_{i=1,\ldots,r+1}$ where the j-th derivatives is replaced by the right derivative when $\phi_{0,j} = \underline{\phi_j}$. The same convention is applied to the derivatives of l_t, Q_t, $\tilde{\sigma}_t^\delta$, $\tilde{l}_t$, and $\tilde{Q}_t$.

Lemma 2. *Under assumptions A1-A10, for all* $i_h = 1, ..., r+1$, $h = 1, ..., k$, $k \leq 3$, *and for all* $p > 0$, *we have*

$$\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial^k \sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1} \ldots \partial\theta_{i_k}} \right|^p < \infty.$$

***Proof of Lemma** 2.* From (20) and assumption **A10(i)**, we have, for all $j_1 \in \{1, ..., r\}$,

$$\frac{\partial\sigma_t^\delta}{\partial\theta_1} = \frac{\partial\sigma_t^\delta}{\partial\omega} = 1 \quad \text{and} \quad \frac{\partial\sigma_t^\delta}{\partial\theta_{1+j_1}} = \frac{\partial\sigma_t^\delta}{\partial\phi_{j_1}} = \sum_{i \in \mathcal{I}_t^+} \frac{\partial\alpha_i^+}{\partial\phi_{j_1}} |\varepsilon_{t-i}|^\delta + \sum_{i \in \mathcal{I}_t^-} \frac{\partial\alpha_i^-}{\partial\phi_{j_1}} |\varepsilon_{t-i}|^\delta. \tag{25}$$

It is thus sufficient to show that for all $j_h \in \{1, ..., r\}$, $h = 1, ..., k$, $k \leq 3$, we have

$$\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial^k \sigma_t^\delta(\boldsymbol{\theta})}{\partial\phi_{j_1} \ldots \partial\phi_{j_k}} \right|^p < \infty$$

From (25), and assumptions **A3(ii)** and **A10(i)** we have

$$\begin{aligned}
\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \frac{\partial^k \sigma_t^\delta(\boldsymbol{\theta})}{\partial\phi_{j_1} \ldots \partial\phi_{j_k}} \right| &\leq \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \sum_{i \in \mathcal{I}_t^+} \left| \frac{\partial^k \alpha_i^+(\boldsymbol{\phi})}{\partial\phi_{j_1} \ldots \partial\phi_{j_k}} \right| |\varepsilon_{t-i}|^\delta + \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \sum_{i \in \mathcal{I}_t^-} \left| \frac{\partial^k \alpha_i^-(\boldsymbol{\phi})}{\partial\phi_{j_1} \ldots \partial\phi_{j_k}} \right| |\varepsilon_{t-i}|^\delta \\
&\leq K \sum_{i \in \mathcal{I}_t^+} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} (\alpha_i^+)^{(1-\xi)}(\boldsymbol{\theta}) |\varepsilon_{t-i}|^\delta + K \sum_{i \in \mathcal{I}_t^-} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} (\alpha_i^-)^{(1-\xi)}(\boldsymbol{\theta}) |\varepsilon_{t-i}|^\delta \\
&\leq K \sum_{i=1}^{\infty} i^{-(d+1)(1-\xi)} |\varepsilon_{t-i}|^\delta
\end{aligned}$$

and from the Hölder inequality we obtain, for all $p > \rho$

$$\begin{aligned} & \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \frac{\partial^k \sigma_t^\delta(\boldsymbol{\theta})}{\partial \phi_{j_1} ... \partial \phi_{j_k}} \right| \\ \leq \quad & K \left[\sum_{i=1}^{\infty} \left[i^{-(d+1)(1-\xi)} \right]^{\frac{p}{\rho}} a_{i,t-i}^{1-\frac{p}{\rho}}(\boldsymbol{\phi}) |\varepsilon_{t-i}|^\delta \right]^{\frac{\rho}{p}} \left[\sum_{i=1}^{\infty} a_{i,t-i}(\boldsymbol{\phi}) |\varepsilon_{t-i}| \delta \right]^{1-\frac{\rho}{p}} \\ \leq \quad & K \left[\sum_{i=1}^{\infty} \left[i^{-(d+1)(1-\xi)} \right]^{\frac{p}{\rho}} a_{i,t-i}^{1-\frac{p}{\rho}}(\boldsymbol{\phi}) |\varepsilon_{t-i}|^\delta \right]^{\frac{\rho}{p}} \left[\sigma_t^\delta(\boldsymbol{\theta}) \right]^{1-\frac{\rho}{p}}, \end{aligned}$$

whence, from assumptions **A3(ii)** and **A10(i)**,

$$\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \frac{1}{\sigma_t^2(\boldsymbol{\theta})} \frac{\partial^k \sigma_t^2(\boldsymbol{\theta})}{\partial \phi_{j_{1,1}}^+ ... \partial \phi_{j_{1,k}}^+} \right|^p \leq K \sum_{i=1}^{\infty} i^{-(d+1)(\rho - p\xi)} \mathbb{E}_{\boldsymbol{\theta_0}} |\varepsilon_{t-i}|^{\delta\rho}$$

for all $\xi > 0$. Since $\rho > \dfrac{1}{d+1}$, we may choose ξ such that $(d+1)(\rho - p\xi) > 1$ and thus we have $\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \dfrac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \dfrac{\partial^k \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1} ... \partial \theta_{i_k}} \right|^p < \infty$.

□

The following lemma shows the integrability of the criterion derivatives at $\boldsymbol{\theta_0}$.

Lemma 3. *Under the assumptions of Theorem 3 or Theorem 4,*

$$\mathbb{E}_{\boldsymbol{\theta_0}} \left\| \frac{\partial l_t(\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta}} \frac{\partial l_t(\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta}'} \right\| < \infty \quad \textit{and} \quad \mathbb{E}_{\boldsymbol{\theta_0}} \left\| \frac{\partial^2 l_t(\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\| < \infty$$

Proof. We have $l_t(\boldsymbol{\theta}) = \log \sigma_t^2(\boldsymbol{\theta}) + \dfrac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})}$, thus we obtain

$$\begin{aligned} \frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \frac{2}{\delta} \left[1 - \frac{\varepsilon_t^2}{\sigma_t^2} \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\theta}} \right] \\ \frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} &= \frac{2}{\delta} \left[1 - \frac{\varepsilon_t^2}{\sigma_t^2} \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial^2 \sigma_t^\delta}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] + \frac{2}{\delta} \left[\frac{\delta + 2}{\delta} \frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\theta}} \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\theta}'} \right]. \end{aligned} \tag{26}$$

Note that at $\boldsymbol{\theta_0}$, $\dfrac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta_0})} = \eta_t^2$ is independent of σ_t^2 and its derivatives. It thus suffices to show

$$\mathbb{E}_{\boldsymbol{\theta_0}} \left\| \frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta_0}) \right\| < \infty, \ \mathbb{E}_{\boldsymbol{\theta_0}} \left\| \frac{1}{\sigma_t^\delta} \frac{\partial^2 \sigma_t^\delta}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}(\boldsymbol{\theta_0}) \right\| < \infty \text{ and } \mathbb{E}_{\boldsymbol{\theta_0}} \left\| \frac{1}{\sigma_t^{2\delta}} \frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\theta}} \frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta_0}) \right\| < \infty.$$

The first two inequalities directly follow from Lemma 2. Since for all j, $\sigma_t^{-\delta} \partial \sigma_t^\delta / \partial \theta_j$ is bounded at $\boldsymbol{\theta_0}$, we obtain the last inequality, which concludes the proof. □

The following lemma shows the non-singularity of J and how it connects with the variance of the criterion derivatives.

Lemma 4. *Under the assumptions of Theorem 3 or Theorem 4,*

$$\boldsymbol{J} \text{ is invertible and } \mathbb{V}_{\boldsymbol{\theta_0}}\left[\frac{\partial l_t(\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta}}\right] = (\kappa_\eta - 1)\boldsymbol{J}$$

Proof. Since at $\boldsymbol{\theta_0}$, $\varepsilon_t^2/\sigma_t^2(\boldsymbol{\theta_0}) = \eta_t^2$ is independent of σ_t^2 and its derivatives, we have

$$\mathbb{E}_{\boldsymbol{\theta_0}}\left[\frac{\partial l_t}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta_0})\right] = \frac{2}{\delta}\mathbb{E}_{\boldsymbol{\theta_0}}[1-\eta_t^2]\mathbb{E}_{\boldsymbol{\theta_0}}\left[\frac{1}{\sigma_t^\delta}\frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta_0})\right] = 0$$

from **A2**. Moreover, in view of integrability of the derivatives of the criterion at $\boldsymbol{\theta_0}$, $\boldsymbol{J} = \mathbb{E}_{\boldsymbol{\theta_0}}\left[\frac{\partial^2 l_t(\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right]$ exists, and from assumption **A7** we can write

$$\mathbb{V}_{\boldsymbol{\theta_0}}\left[\frac{\partial l_t}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta_0})\right] = \frac{4}{\delta^2}\mathbb{E}_{\boldsymbol{\theta_0}}[(1-\eta_t^2)^2]\mathbb{E}_{\boldsymbol{\theta_0}}\left[\frac{1}{\sigma_t^\delta}\frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\theta}}\frac{1}{\sigma_t^\delta}\frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta_0})\right] = (\kappa_\eta - 1)\boldsymbol{J}.$$

Assume now that $\boldsymbol{J}$ is singular, then there exists a non-zero vector $\boldsymbol{\Lambda} = [\lambda_0, \boldsymbol{\lambda}']'$, with $\boldsymbol{\lambda} \in \mathbb{R}^r$, such that almost surely $\boldsymbol{\Lambda}'\boldsymbol{J}\boldsymbol{\Lambda} = 0$, which is equivalent to

$$\lambda_0 + \sum_{i=1}^{\infty}\left[\sum_{j=1}^{r}\lambda_j\frac{\partial \alpha_i^+(\boldsymbol{\phi_0})}{\partial \phi_j}\mathbb{1}_{\varepsilon_{t-i}\geq 0} + \sum_{k=1}^{r}\lambda_k\frac{\partial \alpha_i^-(\boldsymbol{\phi_0})}{\partial \phi_k}\mathbb{1}_{\varepsilon_{t-i}< 0}\right]|\varepsilon_{t-i}|^\delta = 0.$$

Now, assume $\sum_{j=1}^{r}\lambda_j\frac{\partial \alpha_1^+(\boldsymbol{\phi_0})}{\partial \phi_j}\mathbb{1}_{\varepsilon_{t-1}\geq 0} + \sum_{k=1}^{r}\lambda_k\frac{\partial \alpha_1^-(\boldsymbol{\phi_0})}{\partial \phi_k}\mathbb{1}_{\varepsilon_{t-1}< 0} \neq 0$, then it follows

$$\begin{aligned}&\left[\sum_{j=1}^{r}\lambda_j\frac{\partial \alpha_1^+(\boldsymbol{\phi_0})}{\partial \phi_j}\mathbb{1}_{\eta_{t-1}\geq 0} + \sum_{k=1}^{r}\lambda_k\frac{\partial \alpha_1^-(\boldsymbol{\phi_0})}{\partial \phi_k}\mathbb{1}_{\eta_{t-1}< 0}\right]|\eta_{t-1}|^\delta\sigma_{t-1}^\delta(\boldsymbol{\theta_0})\\ &= -\lambda_0 - \sum_{i=2}^{\infty}\left[\sum_{j=1}^{r}\lambda_j\frac{\partial \alpha_i^+(\boldsymbol{\phi_0})}{\partial \phi_j}\mathbb{1}_{\eta_{t-i}\geq 0} + \sum_{k=1}^{r}\lambda_k\frac{\partial \alpha_i^-(\boldsymbol{\phi_0})}{\partial \phi_k}\mathbb{1}_{\eta_{t-i}< 0}\right]|\eta_{t-i}|^\delta\sigma_{t-i}^\delta(\boldsymbol{\theta_0})\end{aligned}$$

whence $\eta_{t-1}^\delta \in \mathcal{F}\left(\eta_{t-2}^\delta, \ldots\right)$ and thus, by independence, $\sum_{j=1}^{r}\lambda_j\frac{\partial \alpha_1^+(\boldsymbol{\phi_0})}{\partial \phi_j}\mathbb{1}_{\eta_{t-1}\geq 0}|\eta_{t-1}|^\delta$ is constant almost surely and thus $\boldsymbol{\lambda}'\ \frac{\partial \alpha_1^+(\boldsymbol{\phi_0})}{\partial \boldsymbol{\phi}} = 0$ almost surely since, from assumption **A2**, η_1 takes at least two positive values. Iterating this argument for $\alpha_i^{+(-)}$, we obtain that for all $i_h^{+(-)} = i_h^{+(-)}(\boldsymbol{\phi_0}), i_h^{+(-)} = 1, \ldots, r$, we have $\boldsymbol{\lambda}'\ \frac{\partial \alpha_{i_h}^{+(-)}(\boldsymbol{\phi_0})}{\partial \boldsymbol{\phi}} = 0$ and thus from assumption **A10(ii)** we must have $\boldsymbol{\lambda} = \mathbf{0}$. This implies $\lambda_0 = 0$ and contradicts the singularity of $\boldsymbol{J}$. $\square$

The following lemma shows the uniform integrability of the second and third order of the criterion derivatives.

Lemma 5. *Under the assumptions of Theorem 3 or Theorem 4, for any $\varepsilon > 0$, there exists a neighborhood $\mathcal{V}(\boldsymbol{\theta_0})$ of $\boldsymbol{\theta_0}$ such that for all $k_1, k_2, k_3 \in \{1, \dots, r+1\}$,*

$$\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta_0})} \left| \frac{\partial^2 l_t(\boldsymbol{\theta_0})}{\partial\theta_{k_1}\partial\theta_{k_2}} \right| < \infty \text{ and } \mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta_0})} \left| \frac{\partial^3 l_t(\boldsymbol{\theta_0})}{\partial\theta_{k_1}\partial\theta_{k_2}\partial\theta_{k_3}} \right| < \infty \quad a.s.$$

Proof. We have

$$\begin{aligned}
\frac{\partial^3 l_t(\boldsymbol{\theta})}{\partial\theta_{i_1}\partial\theta_{i_2}\partial\theta_{i_3}} = \frac{2}{\delta} \Bigg\{ & \left[1 - \frac{\varepsilon_t^2}{\sigma_t^2}\right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial^3 \sigma_t^\delta}{\partial\theta_{i_1}\partial\theta_{i_2}\partial\theta_{i_3}}\right] \\
& + \left[\frac{\delta+2}{2}\frac{\varepsilon_t^2}{\sigma_t^2} - 1\right] \left[\frac{1}{\sigma_t^\delta}\frac{\partial\sigma_t^\delta}{\partial\theta_{i_1}}\right] \left[\frac{1}{\sigma_t^\delta}\frac{\partial^2\sigma_t^\delta}{\partial\theta_{i_2}\partial\theta_{i_3}}\right] \\
& + \left[\frac{\delta+2}{2}\frac{\varepsilon_t^2}{\sigma_t^2} - 1\right] \left[\frac{1}{\sigma_t^\delta}\frac{\partial\sigma_t^\delta}{\partial\theta_{i_2}}\right] \left[\frac{1}{\sigma_t^\delta}\frac{\partial^2\sigma_t^\delta}{\partial\theta_{i_1}\partial\theta_{i_3}}\right] \\
& + \left[\frac{\delta+2}{2}\frac{\varepsilon_t^2}{\sigma_t^2} - 1\right] \left[\frac{1}{\sigma_t^\delta}\frac{\partial\sigma_t^\delta}{\partial\theta_{i_3}}\right] \left[\frac{1}{\sigma_t^\delta}\frac{\partial^2\sigma_t^\delta}{\partial\theta_{i_1}\partial\theta_{i_2}}\right] \\
& + 2\left[1 - \frac{\delta^2+3\delta+2}{\delta^2}\frac{\varepsilon_t^2}{\sigma_t^2}\right] \left[\frac{1}{\sigma_t^\delta}\frac{\partial\sigma_t^\delta}{\partial\theta_{i_1}}\right] \left[\frac{1}{\sigma_t^\delta}\frac{\partial\sigma_t^\delta}{\partial\theta_{i_2}}\right] \left[\frac{1}{\sigma_t^\delta}\frac{\partial\sigma_t^\delta}{\partial\theta_{i_3}}\right] \Bigg\} (\boldsymbol{\theta}).
\end{aligned}$$

From assumptions **A7** and **A11**, and the triangular inequality, there exists a neighborhood $V(\boldsymbol{\theta_0})$ of $\boldsymbol{\theta_0}$ such that,

$$\left\| \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})} \frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})} \right\|_2 = \sqrt{\kappa_\eta} \left\| \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})} \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})} \right\|_2 < \infty.$$

Using Lemma 2, the Cauchy-Schwartz inequality and the Hölder inequality, we have for all $i_1, i_2, i_3 \in \{1, ..., r+1\}$

$$\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta}\in V_\tau(\boldsymbol{\theta_0})} \left| \left[1 - \frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})}\right] \left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial^3\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}\partial\theta_{i_2}\partial\theta_{i_3}}\right] \right| < \infty,$$

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta}\in V_\tau(\boldsymbol{\theta_0})} \left| \left[\frac{\delta+2}{2}\frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})} - 1\right] \left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})}\frac{\partial\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right] \left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})}\frac{\partial^2\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_2}\partial\theta_{i_3}}\right] \right| \\
&\leq \left\| \sup_{\boldsymbol{\theta}\in V_\tau(\boldsymbol{\theta_0})} \left| \frac{\delta+2}{2}\frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})} - 1 \right| \right\|_2 \left\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{1}{\sigma_t^\delta(\boldsymbol{\theta})}\frac{\partial\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}} \right| \right\|_4 \left\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{1}{\sigma_t^\delta(\boldsymbol{\theta})}\frac{\partial^2\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_2}\partial\theta_{i_3}} \right| \right\|_4 \\
&< \infty,
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta}\in V_\tau(\boldsymbol{\theta_0})} \left| \left[1 - \frac{\delta^2+3\delta+2}{\delta^2}\frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})}\right] \left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})}\frac{\partial\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right] \left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})}\frac{\partial\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_2}}\right] \left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})}\frac{\partial\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_3}}\right] \right| \\
&\leq \left\| \sup_{\boldsymbol{\theta}\in V_\tau(\boldsymbol{\theta_0})} \left| 1 - \frac{\delta^2+3\delta+2}{\delta^2}\frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})} \right| \right\|_2 \max_h \left\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{1}{\sigma_t^\delta(\boldsymbol{\theta})}\frac{\partial\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_h}} \right| \right\|_6^3 \\
&< \infty,
\end{aligned}$$

which concludes the proof. □

The following lemma shows the asymptotic irrelevance of the initial values on the derivatives of the criterion.

Lemma 6. *Under the assumptions of Theorem 3 or Theorem 4,*

$$\left\| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\frac{\partial l_t(\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta}} - \frac{\partial \tilde{l}_t(\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta}} \right) \right\| \xrightarrow{P} 0 \quad \textit{and} \quad \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta_0})} \left\| \frac{1}{n} \sum_{t=1}^{n} \left(\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta'}} - \frac{\partial^2 \tilde{l}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta'}} \right) \right\| \xrightarrow{P} 0$$

Proof. First, remark that, from assumption **A3(ii)** and **A11**, on a neighborhood $V(\boldsymbol{\theta_0})$ of $\boldsymbol{\theta_0}$, we have similarly to (22)

$$\sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta_0})} \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2(\boldsymbol{\theta})} = \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta_0})} \eta_t^2 \, \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})} \, \left[\frac{\sigma_t^\delta(\boldsymbol{\theta})}{\tilde{\sigma}_t^\delta(\boldsymbol{\theta})} \right]^{2/\delta} \leq K \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta_0})} \eta_t^2 \, \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})} \tag{27}$$

where K is finite almost surely and does not depend on t since $\sum_{i=0}^{\infty} i^{-(d+1)} \varepsilon_{-i}^2$ admits a moment of order ρ and thus is finite almost surely.

We have

$$\frac{\partial \tilde{l}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{2}{\delta} \left[1 - \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} \right] \left[\frac{1}{\tilde{\sigma}_t^\delta} \frac{\partial \tilde{\sigma}_t^\delta}{\partial \boldsymbol{\theta}} \right] (\boldsymbol{\theta}) = \frac{2}{\delta} \left[1 - \eta_t^2 \frac{\sigma_t^2}{\tilde{\sigma}_t^2} \right] \left[\frac{1}{\tilde{\sigma}_t^2} \frac{\partial \tilde{\sigma}_t^2}{\partial \boldsymbol{\theta}} \right] (\boldsymbol{\theta}),$$

therefore we can write

$$\begin{aligned} \left| \frac{\partial l_t(\boldsymbol{\theta_0})}{\partial \theta_k} - \frac{\partial \tilde{l}_t(\boldsymbol{\theta_0})}{\partial \theta_k} \right| &= \frac{2}{\delta} \Bigg| \left[\frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} - \frac{\varepsilon_t^2}{\sigma_t^2} \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta_k} \right] + \left[1 - \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} \right] \left[\frac{1}{\tilde{\sigma}_t^\delta} - \frac{1}{\sigma_t^\delta} \right] \left[\frac{\partial \sigma_t^\delta}{\partial \theta_k} \right] \\ &\quad + \left[1 - \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} \right] \left[\frac{1}{\tilde{\sigma}_t^\delta} \right] \left[\frac{\partial \sigma_t^\delta}{\partial \theta_k} - \frac{\partial \tilde{\sigma}_t^\delta}{\partial \theta_k} \right] \Bigg| (\boldsymbol{\theta_0}) \\ &= \frac{2}{\delta} \left| A_t + B_t + C_t \right| (\boldsymbol{\theta_0}) \end{aligned}$$

From the Markov inequality we have

$$\begin{aligned} &\mathbb{P} \left[\left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[\frac{\partial l_t(\boldsymbol{\theta_0})}{\partial \theta_k} - \frac{\partial \tilde{l}_t(\boldsymbol{\theta_0})}{\partial \theta_k} \right] \right| > \varepsilon \right] \\ &\quad \leq \frac{1}{\varepsilon} \, \mathbb{E}_{\boldsymbol{\theta_0}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[\frac{\partial l_t(\boldsymbol{\theta_0})}{\partial \theta_k} - \frac{\partial \tilde{l}_t(\boldsymbol{\theta_0})}{\partial \theta_k} \right] \right| \\ &\quad \leq \frac{1}{\varepsilon} \frac{2}{\delta} \left[\mathbb{E}_{\boldsymbol{\theta_0}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} A_t(\boldsymbol{\theta_0}) \right| + \mathbb{E}_{\boldsymbol{\theta_0}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} B_t(\boldsymbol{\theta_0}) \right| + \mathbb{E}_{\boldsymbol{\theta_0}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} C_t(\boldsymbol{\theta_0}) \right| \right] \end{aligned} \tag{28}$$

From (27), we have

$$|A_t(\boldsymbol{\theta_0})| = \left| \left[\frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} - \frac{\varepsilon_t^2}{\sigma_t^2} \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta_k} \right] \right| (\boldsymbol{\theta_0}) \leq K \eta_t^2 \left[\sum_{i=t}^{\infty} a_{i,t-i}(\boldsymbol{\phi_0}) |\varepsilon_{t-i}|^\delta \right] \left| \frac{1}{\sigma_t^\delta(\boldsymbol{\theta_0})} \frac{\partial \sigma_t^\delta}{\partial \theta_k}(\boldsymbol{\theta_0}) \right|.$$

Using the independence of η_t^2 with σ_t^δ and its derivatives at $\boldsymbol{\theta_0}$, **A2**, **A8** and **A9**, Hölder inequality, and Lemma 2,

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta_0}} |A_t(\boldsymbol{\theta_0})|^\rho &\leq K \left(\mathbb{E}_{\boldsymbol{\theta_0}} \left[\sum_{i=t}^{\infty} i^{-(d^*+1)} |\varepsilon_{t-i}|^\delta \right]^{\rho(1+\xi)} \right)^{\frac{1}{1+\xi}} \left(\mathbb{E}_{\boldsymbol{\theta_0}} \left| \frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial \theta_k}(\boldsymbol{\theta_0}) \right|^{\rho \frac{\xi+1}{\xi}} \right)^{\frac{\xi}{1+\xi}} \\
&\leq K \sum_{i=0}^{\infty} (t+i)^{-(d^*+1)\rho} \left(\mathbb{E}_{\boldsymbol{\theta_0}} |\varepsilon_{-i}|^{\delta\rho(1+\xi)} \right)^{\frac{1}{1+\xi}} \\
&\leq K t^{-(d^*+1)\rho+1},
\end{aligned} \tag{29}$$

for some $\xi > 0$ such that $\rho(1+\xi) \leq 1$ and thus

$$\mathbb{E}_{\boldsymbol{\theta_0}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} A_t(\boldsymbol{\theta_0}) \right|^\rho \leq K n^{-(d^*+\frac{3}{2})\rho+2} \underset{n\to\infty}{\to} 0$$

since from assumption **A9** we have $(d^* + \frac{3}{2})\rho - 2 > 0$. Using Markov inequality, we can conclude $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} |A_t(\boldsymbol{\theta_0})|$ tends to 0 in probability. Similar arguments yield $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} |B_t(\boldsymbol{\theta_0})|$ tends to 0 in probability. In addition, from (25), and from assumptions **A10(i)** and **A3(ii)**, we have for all $\xi > 0$,

$$|C_t(\boldsymbol{\theta_0})| = \left| \left[1 - \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} \right] \left[\frac{1}{\tilde{\sigma}_t^\delta} \right] \left[\frac{\partial \sigma_t^\delta}{\partial \theta_k} - \frac{\partial \tilde{\sigma}_t^\delta}{\partial \theta_k} \right] \right| (\boldsymbol{\theta_0}) \leq K \eta_t^2 \sum_{i=0}^{\infty} (t+i)^{-(d+1)(1-\xi)} |\varepsilon_{-i}|^\delta,$$

and thus

$$\mathbb{E}_{\boldsymbol{\theta_0}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} C_t(\boldsymbol{\theta_0}) \right|^\rho \leq n^{-\frac{1}{2}\rho} \sum_{t=1}^{n} \mathbb{E}_{\boldsymbol{\theta_0}} |C_t(\boldsymbol{\theta_0})|^\rho \leq K n^{-(d^*+1)\rho(1-\xi)+2} \underset{n\to\infty}{\to} 0$$

since, from assumption **A8** and **A9**, there exists a ξ such that $(d^*+1)\rho(1-\xi) - 2 > 0$. Using Markov inequality, we can conclude $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} |C_t(\boldsymbol{\theta_0})|$ tends to 0 in probability. Hence (28) yields $\mathbb{P}\left[\left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[\frac{\partial l_t(\boldsymbol{\theta_0})}{\partial \theta_k} - \frac{\partial \tilde{l}_t(\boldsymbol{\theta_0})}{\partial \theta_k} \right] \right| > \varepsilon \right] \to 0$ for all $\varepsilon > 0$ which concludes the proof of the first inequality.

Now consider the asymptotic impact of the initial values on the second-order derivatives of the criterion in a neighborhood of $\boldsymbol{\theta_0}$. Let us denote

$$\chi_t := \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})} |\sigma_t^\delta(\boldsymbol{\theta}) - \tilde{\sigma}_t^\delta(\boldsymbol{\theta})| = \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})} \sum_{i=t}^{\infty} a_{i,t-i}(\boldsymbol{\phi})|\varepsilon_{t-i}|^\delta \leq K\sum_{i=0}^{\infty}(i+t)^{-(d+1)}|\varepsilon_{-i}|^\delta \tag{30}$$

from assumption **A3(ii)** , whence $\mathbb{E}\chi_t^\rho \leq Kt^{-(d+1)\rho+1}$ from **A4**. This shows that χ_t has a finite moment of order ρ and thus is finite almost surely. Furthermore, since $\rho(d+1) > 1$, the dominated convergence theorem entails $\lim_{t\to\infty} \chi_t = 0$ almost surely. Let us now denote

$$\chi_t^{(i_1)} = \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})} \left| \frac{\partial\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}} - \frac{\partial\tilde{\sigma}_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}} \right| \text{ and } \chi_t^{(i_1,i_2)} = \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})} \left| \frac{\partial^2\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}\partial\theta_{i_2}} - \frac{\partial^2\tilde{\sigma}_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}\partial\theta_{i_2}} \right| \tag{31}$$

where $V(\boldsymbol{\theta_0})$ is a neighborhood of $\boldsymbol{\theta_0}$ and $i_1, i_2 \in \{1,\ldots,r\}$. From assumptions **A10(i)** and **A3(ii)**, we have for all $\xi > 0$,

$$\chi_t^{(1+i_1)} \leq \sum_{i=t}^{\infty} \sup_{\boldsymbol{\phi}\in V(\boldsymbol{\phi_0})} \max\left(\left|\frac{\partial\alpha_i^+(\boldsymbol{\phi})}{\partial\phi_{i_1}}\right|, \left|\frac{\partial\alpha_i^-(\boldsymbol{\phi})}{\partial\phi_{i_1}}\right| \right) |\varepsilon_{t-i}|^\delta \leq K\sum_{i=0}^{\infty}(i+t)^{-(d+1)(1-\xi)}|\varepsilon_{-i}|^\delta,$$

whence $\mathbb{E}\left(\chi_t^{(i_1)}\right)^\rho \leq Kt^{-(d+1)\rho(1-\xi)+1}$ from **A4**. This shows that for any i_1, $\chi_t^{(i_1)}$ has a finite moment of order ρ and thus is finite almost surely. Furthermore, since $\rho(d+1) > 1$, we can find a $\xi > 0$ such that $\rho(d+1)(1-\xi) > 1$, and thus the dominated convergence theorem entails $\lim_{t\to\infty} \chi_t^{(i_1)} = 0$ almost surely. The same arguments yield $\lim_{t\to\infty} \chi_t^{(i_1,i_2)} = 0$ almost surely for any i_1, i_2.

Consider now

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})} \left| \frac{1}{n}\sum_{t=1}^{n} \left[\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\theta_{i_1}\partial\theta_{i_2}} - \frac{\partial^2 \tilde{l}_t(\boldsymbol{\theta})}{\partial\theta_{i_1}\partial\theta_{i_2}} \right] \right| \\
\leq\ & \frac{1}{n}\sum_{t=1}^{n} \frac{2}{\delta} \sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})} \left| \left[\frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} - \frac{\varepsilon_t^2}{\sigma_t^2} \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial^2\sigma_t^\delta}{\partial\theta_{i_1}\partial\theta_{i_2}} \right] \right. \\
&+ \left[1 - \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2}\right] \left[\left(\frac{1}{\sigma_t^\delta} - \frac{1}{\tilde{\sigma}_t^\delta}\right) \frac{\partial^2\sigma_t^\delta}{\partial\theta_{i_1}\partial\theta_{i_2}} + \frac{1}{\tilde{\sigma}_t^\delta} \left(\frac{\partial^2\sigma_t^\delta}{\partial\theta_{i_1}\partial\theta_{i_2}} - \frac{\partial^2\tilde{\sigma}_t^\delta}{\partial\theta_{i_1}\partial\theta_{i_2}} \right) \right] \\
&+ \left[\frac{2+\delta}{\delta}\frac{\varepsilon_t^2}{\sigma_t^2} - \frac{2+\delta}{\delta}\frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} \right] \left[\frac{1}{\sigma_t^\delta}\frac{\partial\sigma_t^\delta}{\partial\theta_{i_1}} \right] \left[\frac{1}{\sigma_t^\delta}\frac{\partial\sigma_t^\delta}{\partial\theta_{i_2}} \right] \\
&+ \left[\frac{2+\delta}{\delta}\frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} - 1 \right] \left[\left(\frac{1}{\sigma_t^\delta} - \frac{1}{\tilde{\sigma}_t^\delta}\right) \frac{\partial\sigma_t^\delta}{\partial\theta_{i_1}} + \frac{1}{\tilde{\sigma}_t^\delta}\left(\frac{\partial\sigma_t^\delta}{\partial\theta_{i_1}} - \frac{\partial\tilde{\sigma}_t^\delta}{\partial\theta_{i_1}} \right) \right] \left[\frac{1}{\sigma_t^\delta}\frac{\partial\sigma_t^\delta}{\partial\theta_{i_2}} \right] \\
&+ \left. \left[\frac{2+\delta}{\delta}\frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} - 1 \right] \left[\left(\frac{1}{\sigma_t^\delta} - \frac{1}{\tilde{\sigma}_t^\delta}\right) \frac{\partial\sigma_t^\delta}{\partial\theta_{i_2}} + \frac{1}{\tilde{\sigma}_t^\delta}\left(\frac{\partial\sigma_t^\delta}{\partial\theta_{i_2}} - \frac{\partial\tilde{\sigma}_t^\delta}{\partial\theta_{i_2}} \right) \right] \left[\frac{1}{\tilde{\sigma}_t^\delta}\frac{\partial\tilde{\sigma}_t^\delta}{\partial\theta_{i_1}} \right] \right| (\boldsymbol{\theta}),
\end{aligned}$$

which yields

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\frac{1}{n}\sum_{t=1}^{n}\left[\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\theta_{i_1}\partial\theta_{i_2}}-\frac{\partial^2\tilde{l}_t(\boldsymbol{\theta})}{\partial\theta_{i_1}\partial\theta_{i_2}}\right]\right|\\
&\leq\frac{K}{n}\sum_{t=1}^{n}\eta_t^2\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\left[\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})}\frac{\partial^2\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}\partial\theta_{i_2}}\right]\right|\ \chi_t\\
&\quad+\frac{K}{n}\sum_{t=1}^{n}\eta_t^2\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})}\frac{\partial\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right]\left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})}\frac{\partial\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_2}}\right]\right|\ \chi_t\\
&\quad+\frac{K}{n}\sum_{t=1}^{n}\eta_t^2\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left[\frac{1}{\tilde{\sigma}_t^\delta(\boldsymbol{\theta})}\frac{\partial\tilde{\sigma}_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right]\left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})}\frac{\partial\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_2}}\right]\right|\ \chi_t\\
&\quad+\frac{K}{n}\sum_{t=1}^{n}\eta_t^2\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left[\frac{1}{\tilde{\sigma}_t^\delta(\boldsymbol{\theta})}\frac{\partial\tilde{\sigma}_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right]\right|\ \chi_t^{(i_2)}\\
&\quad+\frac{K}{n}\sum_{t=1}^{n}\eta_t^2\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})}\frac{\partial\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right]\right|\ \chi_t^{(i_1)}\\
&\quad+\frac{K}{n}\sum_{t=1}^{n}\eta_t^2\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\right|\ \chi_t^{(i_1,i_2)}.
\end{aligned}$$

We can first notice that, from the same arguments used to show Lemma 2, for all $p>0$, $i_1,i_2=1,...,r+1$,

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta_0}}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{\tilde{\sigma}_t^\delta(\boldsymbol{\theta})}\frac{\partial\tilde{\sigma}_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right|^p<\infty\\
&\mathbb{E}_{\boldsymbol{\theta_0}}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{\tilde{\sigma}_t^\delta(\boldsymbol{\theta})}\frac{\partial^2\tilde{\sigma}_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}\partial\theta_{i_2}}\right|^p<\infty.
\end{aligned}\tag{32}$$

Then, from independence of η_t^2 with σ_t^δ and its derivatives, assumption **A11**, Lemma 2, (27), and (32) we have, using Hölder inequality, for all i_1,i_2,

$$\begin{aligned}
&\mathbb{E}\left[\eta_t^2\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\left[\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})}\frac{\partial^2\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}\partial\theta_{i_2}}\right]\right|\right]<\infty\\
&\mathbb{E}\left[\eta_t^2\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})}\frac{\partial\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right]\left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})}\frac{\partial\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_2}}\right]\right|\right]<\infty\\
&\mathbb{E}\left[\eta_t^2\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left[\frac{1}{\tilde{\sigma}_t^\delta(\boldsymbol{\theta})}\frac{\partial\tilde{\sigma}_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right]\left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})}\frac{\partial\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_2}}\right]\right|\right]<\infty\\
&\mathbb{E}\left[\eta_t^2\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})}\frac{\partial\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right]\right|\right]<\infty\\
&\mathbb{E}\left[\eta_t^2\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left[\frac{1}{\tilde{\sigma}_t^\delta(\boldsymbol{\theta})}\frac{\partial\tilde{\sigma}_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right]\right|\right]<\infty.
\end{aligned}\tag{33}$$

Since χ_t, $\chi_t^{(i_1)}$, and $\chi_t^{(i_1,i_2)}$ tend to 0 almost surely as t tends to infinity, and (33), Toeplitz lemma combined with Markov inequality entail

$$\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\frac{1}{n}\sum_{t=1}^{n}\left[\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\theta_{i_1}\partial\theta_{i_2}}-\frac{\partial^2 \tilde{l}_t(\boldsymbol{\theta})}{\partial\theta_{i_1}\partial\theta_{i_2}}\right]\right|\underset{n\to\infty}{\to}0$$

in probability, which concludes the proof. □

Finally, the following lemma shows the asymptotic normality of the normalized score.

Lemma 7. *Under the assumptions of Theorem 3 or Theorem 4,*

$$\boldsymbol{Z}_n=-\boldsymbol{J}_n^{-1}\sqrt{n}\frac{\partial Q_n(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}}\xrightarrow{\mathcal{L}}\boldsymbol{Z}\ ,\ \text{with }\boldsymbol{Z}\sim\mathcal{N}(\mathbf{0},(\kappa_\eta-1)\boldsymbol{J})$$

where $\boldsymbol{J}_n^{-1}=\dfrac{\partial^2 Q_n(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}\boldsymbol{\theta}'}$ *is an almost surely positive definite matrix for n sufficiently large.*

Proof. Using the fact that $\sigma_t^\delta(\boldsymbol{\theta_0})$ and its derivatives belong to the σ-field generated by $\{\varepsilon_{t-i},i\geq 0\}$, and the fact that $\mathbb{E}_{\boldsymbol{\theta_0}}\left[\varepsilon_t^2|\varepsilon_u,u<t\right]=\sigma_t^2(\boldsymbol{\theta_0})$, we have

$$\mathbb{E}_{\boldsymbol{\theta_0}}\left[\frac{\partial l_t(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}}\ |\varepsilon_u,u<t\right]=\frac{1}{\sigma_t^\delta(\boldsymbol{\theta_0})}\left[\frac{\partial\sigma_t^\delta}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta_0})\right]\ \mathbb{E}_{\boldsymbol{\theta_0}}\left[\ \sigma_t^2(\boldsymbol{\theta_0})-\varepsilon_t^2|\varepsilon_u,u<t\right]=0$$

and we have from Lemma 4 that $\mathbb{V}_{\boldsymbol{\theta_0}}\left[\dfrac{\partial l_t}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta_0})\right]$ is finite. In view of the invertibility of $\boldsymbol{J}$ and the assumptions on the distribution of η_t (which entails $0<\kappa_\eta-1<\infty$), this covariance matrix is non-degenerate. It follows that, for all $\boldsymbol{\lambda}\in\mathbb{R}^{r+1}$, the sequence $\left\{\boldsymbol{\lambda}'\dfrac{\partial l_t(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}},\varepsilon_t\right\}_t$ is a square integrable ergodic stationary martingale difference. The Cramer-Wold theorem and the central limit theorem for square-integrable martingale difference of Billingsley[7] entail $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta_0})\xrightarrow{\mathcal{L}}\mathcal{N}(\mathbf{0},(\kappa_\eta-1)\boldsymbol{J})$. The ergodic theorem entails $\boldsymbol{J}_n\to\boldsymbol{J}$ as $n\to\infty$ almost surely and thus the conclusion follows from Slutsky lemma.

□

W can now develop the proof of Theorem 3.

Proof of Theorem 3. From Theorem 2, we have that $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}$ converges to $\boldsymbol{\theta_0}$ which, from assumption **A6**, belongs to the interior of $\boldsymbol{\Theta}$, whence the derivative of the criterion is equal to zero at $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}$. It follows that, by a standard Taylor expansion at $\boldsymbol{\theta_0}$, we have

$$0=\frac{\partial\tilde{Q}_n}{\partial\boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}})=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial\tilde{l}_t}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta_0})+\left[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\tilde{l}_t}{\partial\theta_i\partial\theta_j}(\boldsymbol{\theta}^*_{\boldsymbol{ij}})\right]\sqrt{n}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}-\boldsymbol{\theta_0})$$

where the $\boldsymbol{\theta^*_{ij}}$ are between $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}$ and $\boldsymbol{\theta_0}$. We will show the result by proving that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial \tilde{l}_t}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta_0}) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0},(\kappa_\eta-1)\boldsymbol{J})\text{and}\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 \tilde{l}_t}{\partial\theta_i\partial\theta_j}(\boldsymbol{\theta^*_{ij}}) \to J(i,j) \text{ in probability.} \quad (34)$$

Using lemmas 3, 4, 6, and 7 along with Slutsky lemma directly yields the first part of (34).

Consider now a second Taylor expansion of the criterion at $\boldsymbol{\theta_0}$. We have for all i and j,

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t}{\partial\theta_i\partial\theta_j}(\boldsymbol{\theta^*_{ij}}) = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t}{\partial\theta_i\partial\theta_j}(\boldsymbol{\theta_0}) + \frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t}{\partial\boldsymbol{\theta'}}\left[\frac{\partial^2 l_t}{\partial\theta_i\partial\theta_j}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{ij}})\right](\boldsymbol{\theta^*_{ij}}-\boldsymbol{\theta_0})$$

where $\tilde{\boldsymbol{\theta}}_{\boldsymbol{ij}}$ is between $\boldsymbol{\theta^*_{ij}}$ and $\boldsymbol{\theta_0}$. The almost sure convergence of $\tilde{\boldsymbol{\theta}}_{\boldsymbol{ij}}$ to $\boldsymbol{\theta_0}$, the ergodic theorem and the uniform integrability of the third-order derivatives of the criterion (from Lemma 5) imply that almost surely

$$\begin{aligned}\limsup_{n\to\infty}\left\|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t}{\partial\boldsymbol{\theta'}}\left[\frac{\partial^2 l_t}{\partial\theta_i\partial\theta_j}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{ij}})\right]\right\| &\leq \limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left\|\frac{\partial l_t}{\partial\boldsymbol{\theta'}}\left[\frac{\partial^2 l_t}{\partial\theta_i\partial\theta_j}(\boldsymbol{\theta})\right]\right\| \\ &\leq \mathbb{E}_{\boldsymbol{\theta_0}}\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left\|\frac{\partial l_t}{\partial\boldsymbol{\theta'}}\left[\frac{\partial^2 l_t}{\partial\theta_i\partial\theta_j}(\boldsymbol{\theta})\right]\right\| \\ &< \infty\end{aligned}$$

Since $\|\boldsymbol{\theta^*_{ij}}-\boldsymbol{\theta_0}\| \to 0$ almost surely, we have for all $\varepsilon > 0$,

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t}{\partial\boldsymbol{\theta'}}\left[\frac{\partial^2 l_t}{\partial\theta_i\partial\theta_j}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{ij}})\right](\boldsymbol{\theta^*_{ij}}-\boldsymbol{\theta_0})\right| \leq \varepsilon\right] = 1$$

and by the ergodic theorem,

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t}{\partial\theta_i\partial\theta_j}(\boldsymbol{\theta_0}) \xrightarrow{\mathbb{P}} J(i,j).$$

Using Slutsky lemma along with the previous lemmas allows us to obtain the last part of (34) which ends the proof. □

The proof of Theorem 4 is exactly similar to the one developed by Francq and Zakoïan[17] using the previously established lemmas and is left out of this paper for the sake of brevity. The detailed proof is available in the supplementary file.

Proof of Theorem 5. It suffices to show that $\tilde{\delta}_n = \delta_0$ for n large enough, the other results being easily obtained from the proofs of Theorems 2, 3 and 4. We first show that

$$\frac{\sigma_{\delta,t}(\boldsymbol{\theta})}{\sigma_{\delta_0,t}(\boldsymbol{\theta_0})} = 1 \text{ almost surely} \Rightarrow \delta = \delta_0. \tag{35}$$

We have, denoting $\eta_t^{+(-)} = \eta_t \mathbb{1}_{\eta_t \geq (<) 0}$,

$$\begin{aligned} \sigma_{\delta,t}^{\delta}(\boldsymbol{\theta}) &= \omega + \sum_{i=1}^{\infty} \alpha_i^{+}(\boldsymbol{\phi}) \sigma_{\delta_0,t-i}^{\delta}(\boldsymbol{\theta_0}) |\eta_{t-i}^{+}|^{\delta} + \alpha_i^{-}(\boldsymbol{\phi}) \sigma_{\delta_0,t-i}^{\delta}(\boldsymbol{\theta_0}) |\eta_{t-i}^{-}|^{\delta} \\ &= \omega_{\delta,t-2}(\boldsymbol{\theta}) + \alpha_1^{+}(\boldsymbol{\phi}) \sigma_{\delta_0,t-1}^{\delta}(\boldsymbol{\theta_0}) |\eta_{t-1}^{+}|^{\delta} + \alpha_1^{-}(\boldsymbol{\phi}) \sigma_{\delta_0,t-1}^{\delta}(\boldsymbol{\theta_0}) |\eta_{t-1}^{-}|^{\delta} \end{aligned}$$

where $\omega_{\delta,t-2}(\boldsymbol{\theta}) = \omega + \sum_{i=2}^{\infty} \alpha_i^{+}(\boldsymbol{\phi}) \sigma_{\delta_0,t-i}^{\delta}(\boldsymbol{\theta_0}) |\eta_{t-i}^{+}|^{\delta} + \alpha_i^{-}(\boldsymbol{\phi}) \sigma_{\delta_0,t-i}^{\delta}(\boldsymbol{\theta_0}) |\eta_{t-i}^{-}|^{\delta}$ is measurable with respect to $\mathcal{F}_{t-2}$. Let $\Psi = (a, b, r, c, d) \in (0, \infty)^3 \times [0, \infty)^2$ and let the function $g_\Psi : [0, \infty) \to (0; \infty)$ defined by $g_\Psi(x) = (a + bx)^{-1}(c + dx^r)^{1/r}$. We have $g'_\Psi(x) = 0$ if and only if $adx^{r-1} = bc$, whence $g_\Psi(x) = 1$ cannot have more than two solutions, except if i) $r = 1, a = c, b = d$, or ii) $b = d = 0$ and $c = a^r$. Conditionally on $\mathcal{F}_{t-1}$ we have

$$\left[\frac{\sigma_{\delta,t}(\boldsymbol{\theta})}{\sigma_{\delta_0,t}(\boldsymbol{\theta_0})}\right]^{\delta_0} = g_{\Psi^+}(|\eta_{t-1}|^{\delta_0}) \mathbb{1}_{\eta_{t-1} \geq 0} + g_{\Psi^-}(|\eta_{t-1}|^{\delta_0}) \mathbb{1}_{\eta_{t-1} < 0} \tag{36}$$

where $\Psi^{+(-)} = (\omega_{\delta_0,t-2}(\boldsymbol{\theta_0}), \omega_{\delta,t-2}(\boldsymbol{\theta}), \delta/\delta_0, \alpha_1^{+(-)}(\boldsymbol{\phi_0}) \sigma_{\delta_0,t-1}^{\delta_0}(\boldsymbol{\theta_0}), \alpha_1^{+(-)}(\boldsymbol{\phi}) \sigma_{\delta_0,t-1}^{\delta}(\boldsymbol{\theta_0}))$. Thus $\sigma_{\delta,t}(\boldsymbol{\theta}) = \sigma_{\delta_0,t}(\boldsymbol{\theta_0})$ implies i) $\delta = \delta_0$ or ii) $\alpha_1^{+}(\boldsymbol{\phi}) = \alpha_1^{+}(\boldsymbol{\phi_0}) = 0$ and $\alpha_1^{-}(\boldsymbol{\phi}) = \alpha_1^{-}(\boldsymbol{\phi_0}) = 0$. In the latter case, (36) holds by replacing η_{t-1} by η_{t-2}. Iterating the arguments, under A2', the first equality in (35) entails either i) $\delta = \delta_0$ or ii) $\alpha_i^{+}(\boldsymbol{\phi}) = \alpha_i^{+}(\boldsymbol{\phi_0}) = 0$ and $\alpha_i^{-}(\boldsymbol{\phi}) = \alpha_i^{-}(\boldsymbol{\phi_0}) = 0$ for all $i \geq 1$. The latter is precluded by Assumption A3(i), thus we have shown (35), which concludes the proof using Theorem 2 and arguments of its proof. □

A.3 Specification tests

We develop in this section the proofs of the results of Section 3.

Proof of Theorem 6. Let us define for $0 < h < n$

$$r_h = n^{-1} \sum_{t=h+1}^{n} s_t s_{t-h} \ , \text{ with } s_t = \eta_t^2 - 1,$$

and let $\boldsymbol{r}_m = (r_1, \dots, r_m)'$ for any $1 \leq m \leq n$. Let $s_t(\boldsymbol{\theta})$ (respectively $\tilde{s}_t(\boldsymbol{\theta})$) be the random variable obtained by replacing η_t by $\eta_t(\boldsymbol{\theta}) = \varepsilon_t / \sigma_t(\boldsymbol{\theta})$ (respectively $\tilde{\eta}_t(\boldsymbol{\theta}) = \varepsilon_t / \tilde{\sigma}_t(\boldsymbol{\theta})$). Let $r_h(\boldsymbol{\theta})$ and $\tilde{r}_h(\boldsymbol{\theta})$ be defined with the same convention.

We first prove the asymptotic irrelevance of the initial values on $\boldsymbol{r}_m$

$$\sqrt{n}\|\boldsymbol{r}_m(\boldsymbol{\theta_0})-\tilde{\boldsymbol{r}}_m(\boldsymbol{\theta_0})\|=o_P(1) \text{ and } \sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left\|\frac{\partial\boldsymbol{r}_m(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}-\frac{\partial\tilde{\boldsymbol{r}}_m(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|=o_P(1). \tag{37}$$

We have

$$\begin{aligned} s_t(\boldsymbol{\theta})s_{t-h}(\boldsymbol{\theta})-\tilde{s}_t(\boldsymbol{\theta})\tilde{s}_{t-h}(\boldsymbol{\theta}) \quad &= (s_t(\boldsymbol{\theta})-\tilde{s}_t(\boldsymbol{\theta}))s_{t-h}(\boldsymbol{\theta})+(s_{t-h}(\boldsymbol{\theta})-\tilde{s}_{t-h}(\boldsymbol{\theta}))\tilde{s}_t(\boldsymbol{\theta}) \\ &:= A_t(\boldsymbol{\theta})+B_t(\boldsymbol{\theta}) \end{aligned}$$

Similarly to (29), we have

$$\mathbb{E}_{\boldsymbol{\theta_0}}\left|A_t(\boldsymbol{\theta_0})\right|^\rho\leq K\mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{\sigma_t^2(\boldsymbol{\theta_0})-\tilde{\sigma}_t^2(\boldsymbol{\theta_0})}{\tilde{\sigma}_t^2(\boldsymbol{\theta_0})}\right|^\rho\leq Kt^{-(d^*+1)\rho+1}$$

and thus

$$\mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}A_t(\boldsymbol{\theta_0})\right|^\rho\leq Kn^{-(d^*+\frac{3}{2})\rho+2}\underset{n\to\infty}{\to}0$$

since from assumption **A9** we have $(d^*+\frac{3}{2})\rho-2>0$. Using Markov inequality, we can conclude $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}|A_t(\boldsymbol{\theta_0})|$ tends to 0 in probability. Similar arguments yield that $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}|B_t(\boldsymbol{\theta_0})|$ tends to 0 in probability, which proves the first part of (37). In addition, we have

$$\frac{\partial s_t}{\partial\boldsymbol{\theta}}-\frac{\partial\tilde{s}_t}{\partial\boldsymbol{\theta}}=\frac{-2}{\delta}\left[\left[\frac{\varepsilon_t^2}{\sigma_t^2}-\frac{\varepsilon_t^2}{\tilde{\sigma}_t^2}\right]\left[\frac{1}{\sigma_t^\delta}\frac{\partial\sigma_t^\delta}{\partial\boldsymbol{\theta}}\right]+\frac{\varepsilon_t^2}{\tilde{\sigma}_t^2}\left[\frac{1}{\sigma_t^\delta}-\frac{1}{\tilde{\sigma}_t^\delta}\right]\frac{\partial\sigma_t^\delta}{\partial\boldsymbol{\theta}}+\frac{\varepsilon_t^2}{\tilde{\sigma}_t^2}\frac{1}{\tilde{\sigma}_t^\delta}\left[\frac{\partial\sigma_t^\delta}{\partial\boldsymbol{\theta}}-\frac{\partial\tilde{\sigma}_t^\delta}{\partial\boldsymbol{\theta}}\right]\right]$$

whence, for all $k\in\{1,\dots,r+1\}$, using similar notations as in (30) and (31),

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})}\left|\frac{\partial s_t}{\partial\theta_k}-\frac{\partial\tilde{s}_t}{\partial\theta_k}\right|s_{t-h}\leq\;& K\eta_{t-h}^2\eta_t^2\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})}\frac{\sigma_{t-h}^2(\boldsymbol{\theta_0})}{\sigma_{t-h}^2(\boldsymbol{\theta})}\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left|\frac{1}{\sigma_t^\delta}\frac{\partial\sigma_t^\delta}{\partial\theta_k}\right|\chi_t \\ &+K\eta_{t-h}^2\eta_t^2\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})}\frac{\sigma_{t-h}^2(\boldsymbol{\theta_0})}{\sigma_{t-h}^2(\boldsymbol{\theta})}\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\tilde{\sigma}_t^2(\boldsymbol{\theta})}\left|\frac{1}{\sigma_t^\delta}\frac{\partial\sigma_t^\delta}{\partial\theta_k}\right|\chi_t \\ &+K\eta_{t-h}^2\eta_t^2\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})}\frac{\sigma_{t-h}^2(\boldsymbol{\theta_0})}{\sigma_{t-h}^2(\boldsymbol{\theta})}\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\tilde{\sigma}_t^2(\boldsymbol{\theta})}\chi_t^{(k)}.\end{aligned}$$

Then similarly to (33), from independence of η_t^2 with σ_t^δ and its derivatives, assumption **A11**, Lemma 2, (27), and (32) we have, using Hölder inequality, and Toeplitz lemma,

$$\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial s_t}{\partial\theta_k}-\frac{\partial\tilde{s}_t}{\partial\theta_k}\right)s_{t-h}\right|\underset{n\to\infty}{\to}0$$

since χ_t and $\chi_t^{(k)}$ tend to 0 almost surely as t tends to infinity. In a like manner, we obtain that

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial s_{t-h}}{\partial\theta_k}(s_t-\tilde{s}_t)\right|\leq\mathbb{E}\left[\frac{K}{n}\sum_{t=1}^{n}\eta_{t-h}^2\eta_t^2\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\frac{\sigma_{t-h}^2(\boldsymbol{\theta_0})}{\sigma_{t-h}^2(\boldsymbol{\theta})}\left|\frac{1}{\sigma_t^\delta}\frac{\partial\sigma_t^\delta}{\partial\theta_k}\right|\chi_t\right]\underset{n\to\infty}{\to}0.$$

Using Markov inequality, we thus obtain that $n^{-1}\sum_{t=1}^n\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}|\partial A_t(\boldsymbol{\theta})/\partial\theta_k|\to 0$ in probability as n tends to infinity. Similar arguments yield the convergence of the term $n^{-1}\sum_{t=1}^n\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}|\partial B_t(\boldsymbol{\theta})/\partial\theta_k|$ and thus we have shown the second part of (37).

Using a Taylor expansion of $\tilde{r}_h$ at $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}$ for $h=1,\dots,m$ along with (37) yields

$$\sqrt{n}\tilde{r}_h(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}})=\sqrt{n}\tilde{r}_h(\boldsymbol{\theta_0})+\frac{\partial\tilde{r}_h(\boldsymbol{\theta}_n^*)}{\partial\boldsymbol{\theta}}\sqrt{n}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}-\boldsymbol{\theta_0})\overset{o_P(1)}{=}\sqrt{n}r_h(\boldsymbol{\theta_0})+\frac{\partial r_h(\boldsymbol{\theta}_n^*)}{\partial\boldsymbol{\theta}}\sqrt{n}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}-\boldsymbol{\theta_0})$$

for some $\boldsymbol{\theta}_n^*$ between $\boldsymbol{\theta_0}$ and $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}$. In addition, assumption **A11** and Lemma 2 entail that there exists a neighborhood $\mathcal{V}(\boldsymbol{\theta_0})$ of $\boldsymbol{\theta_0}$ such that for all $i,j\in\{1,\dots,r+1\}$

$$\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{\partial^2 s_t(\boldsymbol{\theta})s_{t-h}(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right|<\infty.$$

Using a second Taylor expansion, the ergodic theorem, and Theorem 2, we thus obtain for all $0<h<n$

$$\frac{\partial r_h(\boldsymbol{\theta}_n^*)}{\partial\boldsymbol{\theta}}\to\mathbb{E}_{\boldsymbol{\theta_0}}\left[\frac{\partial s_t s_{t-h}(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}}\right]=-\frac{2}{\delta}\mathbb{E}_{\boldsymbol{\theta_0}}\left[s_{t-h}(\boldsymbol{\theta_0})\frac{1}{\sigma_t^\delta(\boldsymbol{\theta_0})}\frac{\partial\sigma_t^\delta(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}}\right]$$

since $\mathbb{E}_{\boldsymbol{\theta_0}}[s_t\partial s_{t-h}(\boldsymbol{\theta_0})/\partial\boldsymbol{\theta}]=0$ and thus we have

$$\sqrt{n}\tilde{\boldsymbol{r}}_m(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}})\overset{o_P(1)}{=}\sqrt{n}\boldsymbol{r}_m(\boldsymbol{\theta_0})+\boldsymbol{C}_m\sqrt{n}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}-\boldsymbol{\theta_0}).\tag{38}$$

We now derive the asymptotic distribution of $\sqrt{n}(\boldsymbol{r}_m(\boldsymbol{\theta_0}),\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}-\boldsymbol{\theta_0})$. Let us denote $\boldsymbol{s}_{t-1:t-m}=(s_{t-1},\dots,s_{t-m})'$ and remark that $\boldsymbol{r}_m(\boldsymbol{\theta_0})\overset{o_P(1)}{=}n^{-1}\sum_{t=1}^n s_t\boldsymbol{s}_{t-1:t-m}$. From the proof of Theorem 3, we have

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}-\boldsymbol{\theta_0})\overset{o_P(1)}{=}\boldsymbol{J}^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\eta_t^2-1)\frac{1}{\sigma_t^2(\boldsymbol{\theta_0})}\frac{\partial\sigma_t^2(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}},$$

thus the central limit theorem applied to the martingale difference

$$\left\{\left(s_t\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2(\boldsymbol{\theta_0})}{\partial\theta'},s_t\boldsymbol{s}_{t-1:t-m}'\right)';\mathcal{F}(\eta_u,u\leq t)\right\}$$

shows that

$$\sqrt{n}\begin{pmatrix} \tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}_{\boldsymbol{0}} \\ \boldsymbol{r}_m(\boldsymbol{\theta}_{\boldsymbol{0}}) \end{pmatrix} \overset{o_P(1)}{=} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} s_t \begin{pmatrix} \boldsymbol{J}^{-1}\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2(\boldsymbol{\theta}_{\boldsymbol{0}})}{\partial \boldsymbol{\theta}} \\ \boldsymbol{s}_{t-1:t-m} \end{pmatrix}$$
$$\xrightarrow{\mathcal{L}} \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix} (\kappa_\eta - 1)\boldsymbol{J}^{-1} & (\kappa_\eta - 1)\boldsymbol{J}^{-1}\boldsymbol{C}_m' \\ (\kappa_\eta - 1)\boldsymbol{C}_m\boldsymbol{J}^{-1} & (\kappa_\eta - 1)^2\boldsymbol{I}_m \end{bmatrix}\right). \tag{39}$$

From (38) and (39), we obtain

$$\sqrt{n}\tilde{\boldsymbol{r}}_m(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}) \xrightarrow{\mathcal{L}} \mathcal{N}(\boldsymbol{0}, \boldsymbol{D}) \text{ , with } \boldsymbol{D} = (\kappa_\eta - 1)^2\boldsymbol{I}_m - (\kappa_\eta - 1)\boldsymbol{C}_m\boldsymbol{J}^{-1}\boldsymbol{C}_m'$$

and we can show that $\hat{\boldsymbol{D}} \to \boldsymbol{D}$ almost surely as $n \to \infty$. Finally, we show that $\boldsymbol{D}$ is invertible. From assumption **A2**, the law of η_t^2 is non-degenerated hence $\kappa_\eta > 1$ and it suffices to show the non singularity of

$$(\kappa_\eta - 1)\boldsymbol{I}_m - \boldsymbol{C}_m\boldsymbol{J}^{-1}\boldsymbol{C}_m' = \mathbb{E}_{\boldsymbol{\theta}_{\boldsymbol{0}}}\boldsymbol{V}\boldsymbol{V}', \text{ with } \boldsymbol{V} = \boldsymbol{s}_{-1:-m} + \boldsymbol{C}_m\boldsymbol{J}^{-1}\frac{2}{\delta}\frac{1}{\sigma_0^2}\frac{\partial \sigma_0^2(\boldsymbol{\theta}_{\boldsymbol{0}})}{\partial \boldsymbol{\theta}}.$$

If this matrix is singular, then there exists $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)'$ such that $\boldsymbol{\lambda} \neq 0$ and

$$\boldsymbol{\lambda}'\boldsymbol{s}_{-1:-m} + \boldsymbol{\mu}'\frac{1}{\sigma_0^2}\frac{\partial \sigma_0^2(\boldsymbol{\theta}_{\boldsymbol{0}})}{\partial \boldsymbol{\theta}} \text{ a.s., where } \boldsymbol{\mu} = \frac{2}{\delta}\boldsymbol{\lambda}'\boldsymbol{C}_m\boldsymbol{J}^{-1}. \tag{40}$$

If $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{r+1}) = 0$, then $\boldsymbol{\lambda}'\boldsymbol{s}_{-1:-m} = 0$ almost surely, and thus there exists $j \in \{1, \ldots, m\}$ such that $s_{-j} \in \mathcal{F}(s_t, t \neq -j)$, which is impossible since s_t are independent and non-degenerated, and thus we have $\boldsymbol{\mu} \neq 0$. Denoting by R_t any random variable measurable with respect to $\mathcal{F}(\eta_u, u \leq t)$, we have

$$\begin{aligned}
\boldsymbol{\mu}'\frac{\partial \sigma_0^2(\boldsymbol{\theta}_{\boldsymbol{0}})}{\partial \boldsymbol{\theta}} &= \mu_1 + \sum_{i=1}^{\infty}\left[\sum_{j=2}^{r+1}\mu_j\frac{\partial \alpha_i^+(\boldsymbol{\phi}_{\boldsymbol{0}})}{\partial \phi_j}\mathbb{1}_{\eta_{-i}\geq 0} + \sum_{k=2}^{r+1}\mu_k\frac{\partial \alpha_i^-(\boldsymbol{\phi}_{\boldsymbol{0}})}{\partial \phi_k}\mathbb{1}_{\eta_{-i}<0}\right]\sigma_{-i}^\delta(\boldsymbol{\theta}_{\boldsymbol{0}})|\eta_{-i}|^\delta \\
&= \mu_1 + \left[\sum_{j=2}^{r+1}\mu_j\frac{\partial \alpha_1^+(\boldsymbol{\phi}_{\boldsymbol{0}})}{\partial \phi_j}|\eta_{-1}^+|^\delta + \sum_{k=2}^{r+1}\mu_k\frac{\partial \alpha_1^-(\boldsymbol{\phi}_{\boldsymbol{0}})}{\partial \phi_k}|\eta_{-1}^-|^\delta\right]\sigma_{-1}^\delta(\boldsymbol{\theta}_{\boldsymbol{0}}) + R_{-2}
\end{aligned}$$

where $\eta_t^{+(-)} = \eta_t\mathbb{1}_{\eta_t \geq (<) 0}$. In addition, we have

$$\begin{aligned}
\sigma_0^\delta(\boldsymbol{\theta}_{\boldsymbol{0}})\boldsymbol{\lambda}'\boldsymbol{s}_{-1:-m} &= (\omega_0 + a_{1,t-1}(\boldsymbol{\phi}_{\boldsymbol{0}})|\eta_{-1}|^\delta\sigma_{-1}^\delta(\boldsymbol{\theta}_{\boldsymbol{0}}) + R_{-2})(\lambda_1\eta_{-1}^2 + R_{-2}) \\
&= \lambda_1\sigma_{-1}^\delta(\boldsymbol{\theta}_{\boldsymbol{0}})a_{1,t-1}(\boldsymbol{\phi}_{\boldsymbol{0}})|\eta_{-1}|^{\delta+2} + (\omega_0\lambda_1 + R_{-2})\eta_{-1}^2 + R_{-2}.
\end{aligned}$$

Thus (40) entails almost surely

$$0 = \lambda_1 \sigma_{-1}^{\delta}(\boldsymbol{\theta_0})\alpha_1^{+(-)}(\boldsymbol{\phi_0})|\eta_{-1}^{+(-)}|^{\delta+2} + \left[\sum_{j=2}^{r+1} \mu_j \frac{\partial \alpha_1^{+(-)}(\boldsymbol{\phi_0})}{\partial \phi_j}\right] \sigma_{-1}^{\delta}(\boldsymbol{\theta_0})|\eta_{-1}^{+(-)}|^{\delta}$$
$$+(\omega_0\lambda_1 + R_{-2})|\eta_{-1}^{+(-)}|^2 + R_{-2}.$$

Since an equation of the form $a|x|^{\delta+2} + b|x|^{\delta} + c|x|^2 + d = 0$ cannot have more than three positive or more than three negative roots, except if all the coefficients are equal to 0, assumption **A2'** implies $\sum_{j=2}^{r+1} \mu_j \frac{\partial \alpha_1^{+}(\boldsymbol{\phi_0})}{\partial \phi_j} = 0$ and $\sum_{j=2}^{r+1} \mu_j \frac{\partial \alpha_1^{-}(\boldsymbol{\phi_0})}{\partial \phi_j} = 0$ almost surely. Iterating this argument, we obtain for all $i_h^{+}(\boldsymbol{\phi_0})$ and $i_h^{-}(\boldsymbol{\phi_0})$ a similar result, and thus from assumption **A10(ii)**, we must have $\boldsymbol{\mu} = 0$ which is impossible and thus contradicts the singularity of $\boldsymbol{D}$, concluding the proof.

□

Proof of Proposition 1. The first part of the proposition is a standard result for testing linear constraints. See for example Chapter 17 of Gouriéroux and Monfort[26] for proofs of the asymptotic distributions. The second part directly follows from Theorem 2 in Francq and Zakoïan[18]. □