Série des Documents de Travail

n° 2020-21

Conditional asymmetry in ARCH(∞) models

Julien ROYER[1]

[1] CREST, ENSAE, Institut Polytechnique de Paris, 5 Avenue Henri Le Chatelier, 91120 Palaiseau, France ; E-mail : julien.royer@ensae.fr

Conditional asymmetry in ARCH(∞) models

Julien Royer*†[1]

[1]CREST, ENSAE, Institut Polytechnique de Paris

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 tests 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 evidences that the short memory feature of such models is not suitable for lightly traded assets.

Keywords: ARCH(∞) models, conditional asymmetry, Quasi Maximum Likelihood Estimation

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 Zakoian for their guidance and feedback.

Introduction

Since the seminal papers of Engle[18] and Bollerslev[8] in the early 1980s, GARCH models have arguably become some of the most popular models in financial econometrics. Through their various specifications, these models are able to capture some of the well established empirical facts of financial time series[1] such as the absence of sample autocorrelations of asset returns, the presence of significant sample autocorrelations of squared returns, and conditional asymmetry. While GARCH models feature some persistence, their autocovariance functions are exponentially decaying and thus do not allow autocorrelations to be far from zero at large lags. However, empirical studies have showed that some financial time series do exhibit slow decaying autocovariance functions (see for example Ding and Granger[16]). This latter empirical fact has propelled the study of memory in financial returns. The introduction of memory in volatility modelling was proposed by Robinson[38] through ARCH(∞) models. In these models, financial returns (ε_t) and volatilities write as

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

with $\omega > 0$, and α_i, $i = 1, ...,$ a sequence of positive constants. The existence of a strictly stationary and nonanticipative solution has been proved by Giraitis, Kokoszka and Leipus[25], Kazakevičius and Leipus[35], and Douc, Roueff and Soulier[17] 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 $\mu_{2s} = E|\eta_t|^{2s}$. It is noteworthy that (2) imposes the summability of the autocovariances of ε_t^2 and thus refutes the existence of a stationary solution with long memory as shown by Giraitis, Kokoszka and Leipus[25]. Indeed, a process is said to have long memory if its autocovariance function γ is not absolutely summable:

$$\sum_{k \in \mathbb{Z}} |\gamma(k)| = \infty.$$

The result of Giraitis, Kokoszka and Leipus[25] has been the source of a long academic discussion on the existence of a stationary solution with a finite fourth order moment for the FIGARCH model of Baillie, Bollerslev, and Mikkelsen[3] [2]. This question has only been recently answered by Giraitis, Surgailis and Škarnulis[28] who proved the existence of such solutions. Nevertheless, even if ARCH(∞) models are not long memory models, they allow for a very slow rate of decay of the

[1] For a review of some empirical facts of financial returns, see [12]

[2] For a review of the controversies around this issue, see for example Davidson[13]

autocorrelation function of the squared returns. Therefore, they are sometimes referred to as moderate memory models.

However, ARCH(∞) models suffer, like classic GARCH models, from an important drawback: their conditional variance only depends on the amplitude of past returns, and not their sign. It is well documented that, empirically, a volatility increase is generally stronger after a price decrease than following an increase of the same amplitude. This asymmetry was first observed by Black [7] and Christie [11] and is often referred to as *leverage effect*. Indeed, the authors suggest that this phenomenon arises due to equity-debt ratio variation prompted by a large equity price drop, thus leading to a higher risk and therefore volatility. It is however unlikely that this sole mechanism accounts for the magnitude of the observed asymmetry (see for example Schwert [41]). Another popular explanation lies in the *volatility feedback* rationale which justifies asymmetry as the result of time-varying risk premia and volatility anticipations (e.g. French, Schwert and Stambaugh [24], Campbell and Hentschel[10], and Bekaert and Wu[4]). The modelling of this empirical fact has prompted the introduction of various GARCH models, but to our best knowledge, attempts to capture both the asymmetry and the memory properties of financial time series have been scarce. A noticeable exception is the FIEGARCH model introduced by Bollerslev and Mikkelsen[9] as the long-memory extension to the Exponential-GARCH model of Nelson[37].

Although long or moderate memory models are suitable candidates to model financial time series, their use amongst practitioners has been regrettably limited. Indeed, ignoring the memory feature of financial time series may lead to spurious conclusions on widely used estimators such as the Sharpe Ratio (Ho[33]), quantile estimators (Ho[34]), or linear regression coefficients (Liu, Deo and Hurvich [36]). The aim of this paper is to buttress the use of ARCH(∞) models. Our work is organized as follows. In the first section, we introduce a new specification for ARCH(∞) models to capture the possible asymmetry and memory effect in financial returns and give a condition for the existence of a stationary solution to our model. In the second section, we focus on its statistical inference. We prove the strong consistency of the quasi maximum likelihood estimator (QMLE) and derive its asymptotic distribution. In Section 3, 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.

1 Threshold ARCH(∞) model

Following the threshold GARCH (TGARCH) model of Zakoian[43], we propose to introduce asymmetry by specifying the conditional variance as a function of the positive and negative parts of the past returns through the form of a threshold ARCH(∞) model.

Definition 1 (TARCH(∞) process)**.** *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 a TARCH(∞) process if it satisfies an equation of the form*

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \\ \sigma_t^2 = \omega + \sum\limits_{i=1}^{\infty} \alpha_i^+ \varepsilon_{t-i}^2 \mathbb{1}_{\varepsilon_{t-i} \geq 0} + \alpha_i^- \varepsilon_{t-i}^2 \mathbb{1}_{\varepsilon_{t-i} < 0} \end{cases} \tag{3}$$

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

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

$$A_s^+ = \sum_{i=1}^{\infty} \left(\alpha_i^+\right)^s \ , \ A_s^- = \sum_{i=1}^{\infty} \left(\alpha_i^-\right)^s$$

and

$$\mu_{2s}^+ = \mathbb{E}|\mathbb{1}_{\eta_t \geq 0} \ \eta_t|^{2s} \ , \ \mu_{2s}^- = \mathbb{E}|\mathbb{1}_{\eta_t < 0} \ \eta_t|^{2s}.$$

Theorem 1 (Existence of a stationary TARCH(∞) solution)**.** *If there exists $s \in (0, 1]$ such that*

$$A_s^+ \mu_{2s}^+ + A_s^- \mu_{2s}^- < 1, \tag{4}$$

there exists a unique strictly stationary and nonanticipative solution of (3) such that $\mathbb{E}|\varepsilon_t|^{2s} < \infty$. This solution is given by

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \\ \sigma_t^2 = \omega + \omega \sum\limits_{k=1}^{\infty} \sum\limits_{i_1, ..., i_k \geq 1} a_{i_1, t-i_1} ... a_{i_k, t-i_1-...-i_k} \ \eta_{t-i_1}^2 ... \eta_{t-i_1-...-i_k}^2 \end{cases} \tag{5}$$

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. *Comments on the assumption*

- *In the ARCH(∞) case, where $A_s^+ = A_s^- = A_s$, the assumption for the existence of a solution becomes $A_s(\mu_{2s}^+ + \mu_{2s}^-) < 1$ which reduces to (2) since $\mu_{2s}^+ + \mu_{2s}^- = \mu_{2s}$.*

The existence of a second order stationary solution of a TARCH(∞) can easily be ensued from Theorem 1, where the condition reduces to $A_1^+\mu_2^+ + A_1^-\mu_2^- < 1$. The existence of a fourth order stationary solution is not straightforward but drawing from the results of Giraitis and Surgailis[27], the following corollary provides a sufficient condition.

Corollary 1.1 (Existence of a fourth order stationary TARCH(∞) solution)**.** *If*

$$A_2^+\mu_4^+ + A_2^-\mu_4^- < 1,$$

the strictly stationary solution admits a fourth order moment.

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

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \\ \sigma_t^2 = \dfrac{\omega}{1-\beta} + \sum\limits_{i=1}^{\infty} \alpha\ \beta^{i-1}\varepsilon_{t-i}^2, \end{cases} \tag{6}$$

where α and β, are positive constants, $\beta < 1$, and $\omega > 0$ is obviously a particular (symmetrical) specification of (3). 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

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \\ \sigma_t^2 = \dfrac{\omega}{1-\beta} + \sum\limits_{i=1}^{\infty} \left(\alpha\ \beta^{i-1} + \gamma i^{-d-1}\right) \varepsilon_{t-i}^2 \end{cases} \tag{7}$$

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 shock effect has almost entirely disappeared after a hundred lags in the GARCH(1,1) case, while it remains clearly observable on the ARCH(∞) process.

Consider now the well known asymmetric extension to the GARCH(1,1), the GJR-GARCH(1,1) model introduced by Glosten, Jagannathan and Runkle[29]. This model is also a particular specification of a TARCH(∞) process as we can rewrite it as

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \\ \sigma_t^2 = \dfrac{\omega}{1-\beta} + \sum\limits_{i=1}^{\infty} \beta^{i-1}(\alpha^+ \mathbb{1}_{\varepsilon_{t-i}\geq 0} + \alpha^- \mathbb{1}_{\varepsilon_{t-i}<0})\varepsilon_{t-i}^2. \end{cases}$$

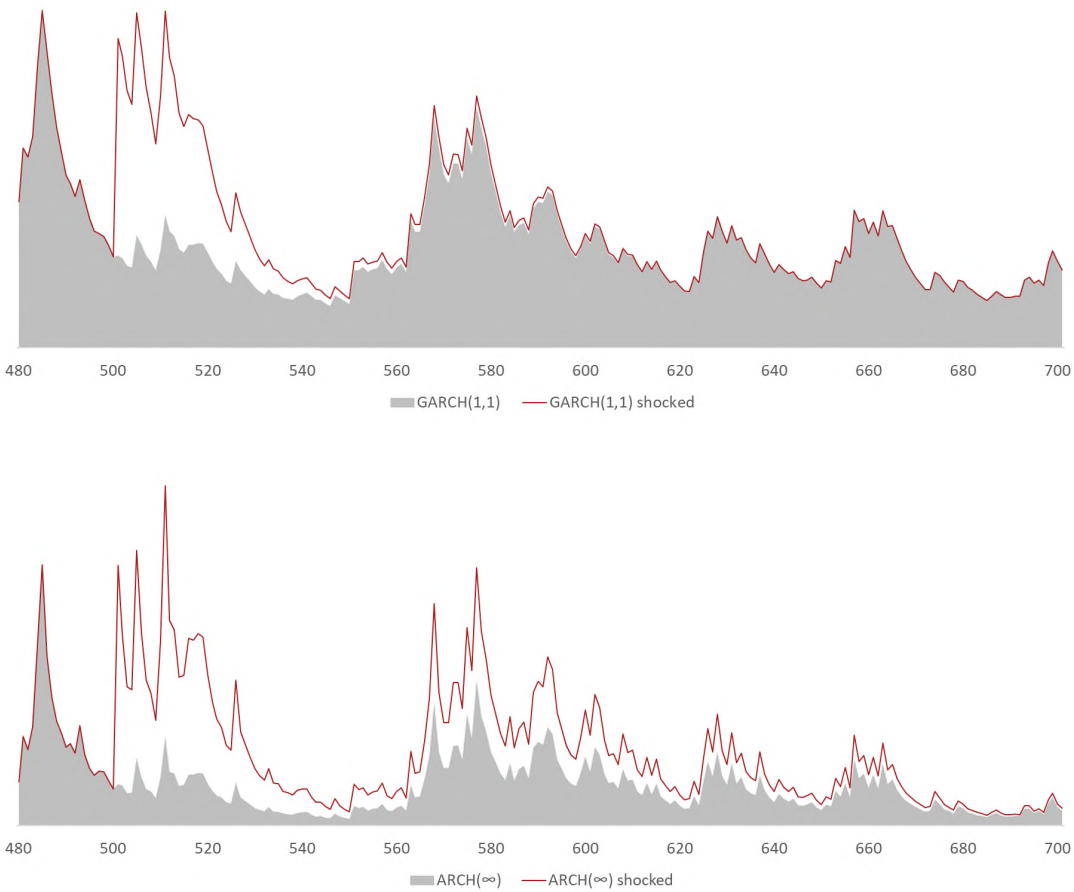


Figure 1: Effect of a shock on η_t at $t = 500$ on the conditional variance of a GARCH(1,1) process and an ARCH(∞) process with $\alpha_i = \alpha\beta^{i-1} + \gamma i^{-(d+1)}$, where $\omega = 0.01$, $\alpha = 0.15$, $\beta = 0.85$, $\gamma = 0.15$, and $d = 1$, and with $\eta_t \sim \mathcal{N}(0,1)$.

In the spirit of (7), an extension to the GJR-GARCH(1,1) model to allow for higher persistence is

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \\ \sigma_t^2 = \dfrac{\omega}{1-\beta} + \sum\limits_{i=1}^{\infty} \beta^{i-1}(\alpha^+ \mathbb{1}_{\varepsilon_{t-i}\geq 0} + \alpha^- \mathbb{1}_{\varepsilon_{t-i}<0})\varepsilon_{t-i}^2 + \gamma i^{-d-1}\varepsilon_{t-i}^2. \end{cases} \tag{8}$$

The models introduced in (7) and (8) are particularly interesting as they allow to nest GARCH-type specifications in highly persistent volatility models. It will thus allow us to verify if the classical GARCH standards are suited to model financial data that may exhibit strong persistence. The third and fifth sections will address that question.

2 Statistical inference of a Threshold ARCH(∞) 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[39], we introduce the parametric form of Model (3)

$$\begin{cases} \varepsilon_t = \sigma_t(\boldsymbol{\theta_0})\eta_t \\ \sigma_t^2(\boldsymbol{\theta_0}) = \omega_0 + \sum_{i=1}^{\infty} \alpha_i(\boldsymbol{\phi_0^+})\varepsilon_{t-i}^2 \mathbb{1}_{\varepsilon_{t-i}\geq 0} + \alpha_i(\boldsymbol{\phi_0^-})\varepsilon_{t-i}^2 \mathbb{1}_{\varepsilon_{t-i}<0} \end{cases} \tag{9}$$

where $\alpha_i(.) : \boldsymbol{\Phi} \to [0, \infty]$ are known functions and $\boldsymbol{\phi_0^+}$ and $\boldsymbol{\phi_0^-}$ are $r \times 1$ unknown vectors of parameters, and ω_0 is an unknown positive constant. We wish to estimate $\boldsymbol{\theta_0} = [\omega_0, (\boldsymbol{\phi_0^+})', (\boldsymbol{\phi_0^-})']'$ over a parameter space $\boldsymbol{\Theta}$, on the basis of n observations $\varepsilon_1, \dots, \varepsilon_n$.

The estimation of the parameters of ARCH(∞) models has been studied through two contributions. First, Giraitis and Robinson[26] proposed a Whittle estimation of $\boldsymbol{\theta_0}$ based on the observations of $\varepsilon_1^2, \dots, \varepsilon_n^2$. This method, however presents some disadvantages as discussed by the authors[3]. For their part, Robinson and Zaffaroni[39] proposed to estimate the parameter $\boldsymbol{\theta_0}$ by Quasi Maximum Likelihood following the works of Berkes, Horváth and Kokoszka[6], and Francq and Zakoïan[21] who studied the convergence and asymptotic properties of the QML estimator in the case of a GARCH(p,q) process[4]. For different assumptions, see also Hafner and Preminger[31]. In the spirit of [39], we study the QML estimator in the case of a TARCH(∞) process.

Let us rewrite the volatility of (9) as

$$\sigma_t^2(\boldsymbol{\theta_0}) = \omega_0 + \sum_{i=1}^{\infty} a_{i,t-i}(\boldsymbol{\phi_0^+}, \boldsymbol{\phi_0^-})\varepsilon_{t-i}^2 \tag{10}$$

where $a_{i,t-i}(\boldsymbol{\phi_0^+}, \boldsymbol{\phi_0^-}) = \alpha_i(\boldsymbol{\phi_0^+})\mathbb{1}_{\varepsilon_{t-i}\geq 0} + \alpha_i(\boldsymbol{\phi_0^-})\mathbb{1}_{\varepsilon_{t-i}<0}$ and remark that for all i, any t, and any $\boldsymbol{\phi}, \boldsymbol{\phi^*}$ in $\boldsymbol{\Phi}$

$$\min(\alpha_i(\boldsymbol{\phi}), \alpha_i(\boldsymbol{\phi^*})) \leq a_{i,t-i}(\boldsymbol{\phi}, \boldsymbol{\phi^*}) \leq \max(\alpha_i(\boldsymbol{\phi}), \alpha_i(\boldsymbol{\phi^*})). \tag{11}$$

We define the Quasi Maximum Likelihood estimator as follows

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

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

[4]For a review of the statistical inference results for long memory processes, see [5].

where, for any admissible value $\boldsymbol{\theta}$ of $\boldsymbol{\theta_0}$, $\tilde{\sigma}_t^2$ is defined as

$$\begin{aligned}\tilde{\sigma}_t^2(\boldsymbol{\theta}) &= \omega + \sum_{i=1}^{t-1} \alpha_i(\boldsymbol{\phi}^+)\varepsilon_{t-i}^2 \mathbb{1}_{\varepsilon_{t-i}\geq 0} + \alpha_i(\boldsymbol{\phi}^-)\varepsilon_{t-i}^2 \mathbb{1}_{\varepsilon_{t-i}<0} \\ &= \omega + \sum_{i=1}^{t-1} a_{i,t-i}(\boldsymbol{\phi}^+, \boldsymbol{\phi}^-)\varepsilon_{t-i}^2.\end{aligned} \tag{12}$$

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 $\boldsymbol{\Theta} = [\omega_L, \omega_U] \times \boldsymbol{\Phi} \times \boldsymbol{\Phi}$ 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. 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 $\alpha_k(\boldsymbol{\phi}) \neq \alpha_k(\boldsymbol{\phi}^*)$.

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

[A4] There exists a solution (ε_t) of equation (9) and for $\rho \in (\frac{1}{d+1}, 1]$, we have $\mathbb{E}|\varepsilon_t|^{2\rho} < \infty$.

[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. *Comments on the assumptions*

- *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 $\sigma_t^2(\boldsymbol{\theta})$ for any $\boldsymbol{\theta}$.*

- *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.*

- *Proposition 2 in the appendix gives a sufficent condition for Assumption **A5**.*

- *In the classic case where $\boldsymbol{\theta_0} = [\omega_0, \boldsymbol{\phi}_0']'$ with $\boldsymbol{\phi_0} = \boldsymbol{\phi_0^+} = \boldsymbol{\phi_0^-}$ our assumptions are mostly in line with the ones proposed by Robinson and Zaffaroni[39]. However, our assumptions on η_t are noticeably milder as we do not specify a particular*

restriction on the distribution as opposed to [39] in which the probability density function of η_t is assumed to be 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 $\alpha_i(\boldsymbol{\theta_0}) \leq K\alpha_j(\boldsymbol{\theta_0})$ for $i \geq j \geq 1$. Thus, our assumptions allow to nest the classical ARCH(q), or allow for some sparsity in the model coefficients.

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

Theorem 2 (Strong consistency of the Quasi Maximum Likelihood estimator)**.** *Under assumptions A1-A5, almost surely*

$$\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} \to \boldsymbol{\theta_0}, \qquad \text{as } n \to \infty$$

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

[A6] $\boldsymbol{\theta_0}$ belongs to the interior of $\boldsymbol{\Theta}$.

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

[A8] For all $i \geq 1$, $\max(\alpha_i(\boldsymbol{\phi_0^+}), \alpha_i(\boldsymbol{\phi_0^-})) \leq K\ i^{-d^*-1}$ for some $d^* > \dfrac{1}{2}$.

[A9] $\mathbb{E}|\varepsilon_t|^{2\rho} < \infty$ for some $\rho \in (\frac{4}{2d^*+3}, 1]$.

[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 $\boldsymbol{\phi}$,

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

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

(ii) There exists $i_h^+ = i_h(\boldsymbol{\phi_0^+})$ and $i_h^- = i_h(\boldsymbol{\phi_0^-})$, $h = 1, ..., r$, such that $1 \leq i_1^{+(-)} < ... < i_r^{+(-)} < \infty$ and

$$\text{rank}\left[\frac{\partial \alpha_{i_1^+}(\boldsymbol{\phi_0^+})}{\partial \boldsymbol{\phi}} ... \frac{\partial \alpha_{i_r^+}(\boldsymbol{\phi_0^+})}{\partial \boldsymbol{\phi}}\right] = \text{rank}\left[\frac{\partial \alpha_{i_1^-}(\boldsymbol{\phi_0^-})}{\partial \boldsymbol{\phi}} ... \frac{\partial \alpha_{i_r^-}(\boldsymbol{\phi_0^-})}{\partial \boldsymbol{\phi}}\right] = r.$$

[A11] For all k>0, there exists a neighborhood $V(\boldsymbol{\theta_0})$ of $\boldsymbol{\theta_0}$ such that,

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

Remarks 2.2. *Comments on the additional assumptions*

- *Assumption* ***A6*** *is required for asymptotic normality.*
- *Assumption* ***A7*** *is necessary for the existence of the variance of the score vector* $\partial l_t(\boldsymbol{\theta_0})/\partial\boldsymbol{\theta}$.
- *Assumptions* ***A8*** *and* ***A9*** *are stronger thanAssumptions* ***A3(ii)*** *and* ***A4*** *and impose a higher rate of convergence for* α_i.
- *Assumption* ***A10(i)*** *is similar to Assumption* ***A3(ii)*** *and allows the summability of the derivatives of the* α_i *functions, while Assumption* ***A10(ii)*** *ensures non singularity of the matrix* $\boldsymbol{J}$.
- *Proposition 3 in the appendix gives an example of a sufficient condition for Assumption* ***A11***.

Theorem 3 (Asymptotic normality of the Quasi Maximum Likelihood estimator)**.** *Under assumptions A1-A11,*

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta_0}) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, (\kappa_\eta - 1)\boldsymbol{J}^{-1}) \tag{13}$$

where

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

is a positive definite matrix.

3 Specification tests

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

3.1 Testing for asymmetry with a memory model

In Model (9), the symmetry hypothesis is a particular constrained representation. Testing for the significance of asymmetry can thus be achieved by testing the implied restriction on $\boldsymbol{\theta_0}$ given by

$$H_0^{\text{sym}} : \boldsymbol{R}\ \boldsymbol{\theta_0} = 0, \quad H_1^{\text{asym}} : \boldsymbol{R}\ \boldsymbol{\theta_0} \neq 0 \tag{14}$$

where $\boldsymbol{R}$ is the constraints matrix. Let c be the rank of the matrix $\boldsymbol{R}$. If we are interested in testing the full symmetry of the model, $\boldsymbol{R}$ is given by $[\mathbf{0}_{(r\times 1)}, \boldsymbol{I}_r, -\boldsymbol{I}_r]$ where $\boldsymbol{I}_r$ is the identity matrix of rank r. The hypothesis are thus

$$H_0^{\text{sym}} : \boldsymbol{\phi}_{\mathbf{0}}^{+} = \boldsymbol{\phi}_{\mathbf{0}}^{-}, \quad H_1^{\text{asym}} : \boldsymbol{\phi}_{\mathbf{0}}^{+} \neq \boldsymbol{\phi}_{\mathbf{0}}^{-}.$$

Testing for a constrained representation is highly common when testing for asymmetry in parametric models, see for example Nelson [37] or Rodríguez and Ruiz [40].

The triptych of the Wald, Rao-score, and Quasi Likelihood Ratio (LR) statistics is given by

$$\begin{aligned}
W_n^{\text{sym}} &= (\boldsymbol{R}\,\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}})' \left(\boldsymbol{R} \left(\frac{(\hat{\kappa}_\eta - 1)}{n} \hat{\boldsymbol{J}}_n^{-1} \right) \boldsymbol{R}' \right)^{-1} (\boldsymbol{R}\,\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}) \\
R_n^{\text{sym}} &= \frac{n}{\hat{\kappa}_{\eta|H_0^{\text{sym}}} - 1} \frac{\partial \tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}})}{\partial \boldsymbol{\theta}'} \hat{\boldsymbol{J}}_{n|H_0^{\text{sym}}}^{-1} \frac{\partial \tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}})}{\partial \boldsymbol{\theta}} \\
L_n^{\text{sym}} &= \frac{2n}{\hat{\kappa}_{\eta|H_0^{\text{sym}}} - 1} \left[\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}}) - \tilde{Q}_n(\tilde{\boldsymbol{\theta}}_n) \right]
\end{aligned} \tag{15}$$

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

$$\begin{aligned}
\hat{\boldsymbol{J}}_n &= \frac{1}{n} \sum_{t=1}^{n} \frac{1}{\tilde{\sigma}_t^4(\tilde{\boldsymbol{\theta}}_n)} \frac{\partial \tilde{\sigma}_t^2(\tilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} \frac{\partial \tilde{\sigma}_t^2(\tilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}'}, \\
\hat{\boldsymbol{J}}_{n|H_0^{\text{sym}}} &= \frac{1}{n} \sum_{t=1}^{n} \frac{1}{\tilde{\sigma}_t^4(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}})} \frac{\partial \tilde{\sigma}_t^2(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}})}{\partial \boldsymbol{\theta}} \frac{\partial \tilde{\sigma}_t^2(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}})}{\partial \boldsymbol{\theta}'}, \\
\hat{\kappa}_\eta &= \frac{1}{n} \sum_{t=1}^{n} \frac{\varepsilon_t^4}{\tilde{\sigma}_t^4(\tilde{\boldsymbol{\theta}}_n)} \\
\hat{\kappa}_{\eta|H_0^{\text{sym}}} &= \frac{1}{n} \sum_{t=1}^{n} \frac{\varepsilon_t^4}{\tilde{\sigma}_t^4(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}})}
\end{aligned}$$

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

Proposition 1 (Critical regions when testing for asymmetry)**.** *Under the assumptions of Theorem 3, under H_0^{sym}, the three test statistics follow a χ_c^2 distribution and the critical regions at the asymptotic level α are given by*

$$\left\{ W_n^{sym} > \chi_c^2(1-\alpha) \right\}, \left\{ R_n^{sym} > \chi_c^2(1-\alpha) \right\}, \left\{ L_n^{sym} > \chi_c^2(1-\alpha) \right\}$$

where $\chi_c^2(1-\alpha)$ is the $(1-\alpha)$-quantile of the χ^2 distribution with c degrees of freedom.

3.2 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 of 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

$$\begin{cases} \varepsilon_t = \sigma_t(\boldsymbol{\theta_0})\eta_t \\ \sigma_t^2(\boldsymbol{\theta_0}) = \dfrac{\omega_0}{1-\beta_0} + \sum\limits_{i=1}^{\infty} (\alpha_0\beta_0^{i-1} + \gamma_0 i^{-(d_0+1)})\varepsilon_{t-i}^2 \end{cases} \tag{16}$$

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

$$H_0^{\text{GARCH}} : \gamma_0 = 0, \quad H_1^{\text{ARCH}(\infty)} : \gamma_0 > 0, \tag{17}$$

which can be rewritten as $H_0^{\text{GARCH}} : \boldsymbol{R}\ \boldsymbol{\theta_0} = 0$, and $H_1^{\text{ARCH}(\infty)} : \boldsymbol{R}\ \boldsymbol{\theta_0} > 0$ with $\boldsymbol{R} = [0, 0, 0, 1]$. While this test may seem standard, it poses two major difficulties. As $\gamma_0 \geq 0$, this is a test at the boundary of the parameter space, but we only derived the distribution of the quasi maximum likelihood estimator in the interior of our parameter space. In addition, the parameter d_0 is not identified under the null hypothesis. To take into account the latter problem, one could study the supremum-tests statistics on γ_0 building upon the work of Andrews and Ploberger[2] and Hansen[32]. This is however beyond the scope of this paper. We propose to solve this issue by setting d_0 at an arbitrary value $\overline{d} > \frac{1}{2}$.

Testing problems in which the parameter is at the boundary under the null hypothesis have been widely approached in the literature. Papers dealing with this issue in the context of GARCH models adequacy include, amongst others, Andrews[1] who proposes a test for conditional homoscedasticity, and Francq and Zakoïan[22] who test the nullity of GARCH coefficients[5]. In the latter, the authors show that, under the assumption that parameters are on the boundary, the QMLE asymptotic distribution is not standard. Under additional technical assumptions, the authors derive the asymptotic distributions of the Wald, Rao-score, and LR statistics. Furthermore, they show that, when testing the nullity of only one coefficient, the Wald test has an explicit asymptotic distribution, and is locally asymptotically more powerful than the standard score test. We thus focus on the Wald statistic W_n^{GARCH}, which has the same form as W_n^{sym} in (15).

[5]See the latter for further references

In the spirit of [22], it can be shown that, under H_0^{GARCH}, $W_n^{\text{GARCH}} \xrightarrow{\mathcal{L}} \frac{1}{2}\delta_0 + \frac{1}{2}\chi_1^2$, where δ_0 is the Dirac measure at 0. Thus, the critical region of asymptotic level α is given by $\{W_n^{\text{GARCH}} > \chi_1^2(1-2\alpha)\}$.

This test can easily be extended to an asymmetric volatility model. Consider, the TARCH(∞) extension of the GJR-GARCH(1,1) presented in (8). Testing for the adequacy of the GJR-GARCH model can then be achieved by testing for $H_0^{\text{GJR}} : \gamma_0 = 0$ against $H_1^{\text{TARCH}(\infty)} : \gamma_0 > 0$. We can thus define the Wald statistic W_n^{GJR} in a similar manner, and derive its asymptotic distribution under H_0^{GJR}.

4 Simulations

In order to assess the finite sample properties of the QML estimators in the different settings studied in this paper, and to study the empirical behavior of the test statistics defined in Section 4, we carried out some Monte Carlo experiments. In the following simulations, we used Gaussian innovations ($\eta_t \sim \mathcal{N}(0,1)$).

We first wanted to assess the finite sample properties of the QML estimator in the simple case where the α_i functions are given by $\alpha_i(\boldsymbol{\phi}) = \gamma i^{-(d+1)}$. We have simulated a thousand samples of different sizes n of a symmetric ARCH(∞) process, parametrized by $\boldsymbol{\theta_0} = [\omega_0, (\boldsymbol{\phi_0})']'$, $\boldsymbol{\phi_0} = (0.4, 0.85)$. To verify that we are able to capture asymmetric behaviors, we have also simulated a TARCH(∞) with $\boldsymbol{\theta_0} = [\omega_0, (\boldsymbol{\phi_0^+})', (\boldsymbol{\phi_0^-})']'$, $\boldsymbol{\phi_0^+} = (0.4, 1.2)$, and $\boldsymbol{\phi_0^-} = (0.2, 0.75)$. On each realisation, we fitted a TARCH(∞) by QML, which gave us a thousand estimators $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}$. Table 1 presents the empirical mean and root mean squared error (RMSE) of these estimators (in brackets). We can note that the estimations results are satisfactory, although the parameters $d^{+(-)}$ may require a large sample size to be precise, which is natural as it is mostly observable at large lags.

	ARCH(∞)					TARCH(∞)				
	ω	γ^+	d^+	γ^-	d^-	ω	γ^+	d^+	γ^-	d^-
$\boldsymbol{\theta_0}$	1.00	0.40	0.85	0.40	0.85	1.00	0.40	1.20	0.20	0.75
$\tilde{\boldsymbol{\theta}}_{\mathbf{1000}}$	1.08	0.40	0.85	0.40	0.87	1.04	0.40	1.20	0.20	0.89
	(0.19)	(0.07)	(0.53)	(0.07)	(0.50)	(0.15)	(0.08)	(0.59)	(0.06)	(0.73)
$\tilde{\boldsymbol{\theta}}_{\mathbf{2000}}$	1.06	0.40	0.85	0.40	0.86	1.03	0.40	1.19	0.20	0.82
	(0.14)	(0.05)	(0.47)	(0.05)	(0.43)	(0.11)	(0.05)	(0.48)	(0.04)	(0.70)
$\tilde{\boldsymbol{\theta}}_{\mathbf{5000}}$	1.04	0.40	0.85	0.40	0.85	1.01	0.40	1.19	0.20	0.75
	(0.08)	(0.03)	(0.38)	(0.03)	(0.38)	(0.07)	(0.04)	(0.39)	(0.03)	(0.70)

Table 1: *Estimation results for 1000 simulations of size n of a symmetric ARCH(∞) process and a TARCH(∞) process with $\alpha_i(\boldsymbol{\phi}) = \gamma i^{-d-1}$*

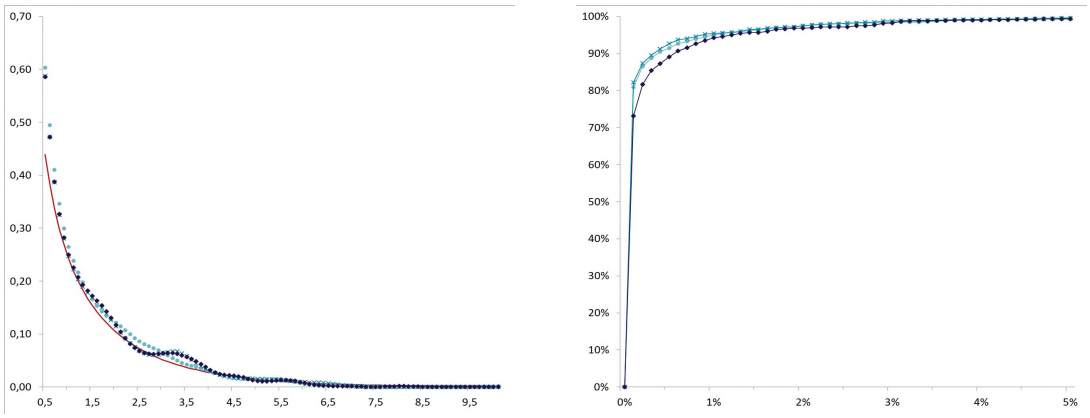
We then considered specifications that nest favored GARCH models in order to study the empirical behavior of the QMLE and the test statistics defined in the previous section. We started by simulating a thousand sample of different sizes n of a GARCH(1,1) with $(\omega_0, \alpha_0, \beta_0) = (0.2, 0.15, 0.75)$. We then simulated a thousand sample of a ARCH(∞) given by (7) with $\omega_0 = 0.2$, $\alpha_0 = 0.1$ $\beta_0 = 0.75$ and $\gamma_0 = 0.2$ which corresponds to the extension of the GARCH(1,1), and a thousand sample of a TARCH(∞) given by (8) with $\omega_0 = 0.2$, $\alpha_0^+ = 0.05$, $\alpha_0^- = 0.15$ $\beta_0 = 0.75$ and $\gamma_0 = 0.2$ which corresponds to the extension of the GJR-GARCH(1,1). On each realisation, we fitted by QML a symmetric ARCH(∞) specified as (7) and a TARCH(∞) specified as (8). This gave us sets of a thousand estimators $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}^{\text{ARCH}}$ and $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}^{\text{TARCH}}$. Empirical mean and RMSE of these estimators are reported in Table 2. Finally, in order to assess the finite sample properties of the asymptotic variance estimator, given by (13), 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 2 are quite satisfactory.

			ARCH(∞)			TARCH(∞)		
		$\boldsymbol{\theta_0}$	$\tilde{\boldsymbol{\theta}}_{\mathbf{5000}}^{\text{ARCH}}$	RMSE	$V_n^{\frac{1}{2}}$	$\tilde{\boldsymbol{\theta}}_{\mathbf{5000}}^{\text{TARCH}}$	RMSE	$V_n^{\frac{1}{2}}$
GARCH(1,1)	ω	0.20	0.20	0.03	0.03	0.20	0.03	0.03
	α	0.15	0.14	0.02	0.03	–	–	–
	α^+	–	–	–	–	0.14	0.03	0.03
	α^-	–	–	–	–	0.14	0.03	0.03
	β	0.75	0.76	0.03	0.03	0.76	0.03	0.03
	γ	–	0.02	0.03	0.04	0.02	0.03	0.04
ARCH(∞)	ω	0.20	0.21	0.04	0.03	0.21	0.04	0.03
	α	0.10	0.10	0.03	0.02	–	–	–
	α^+	–	–	–	–	0.10	0.03	0.02
	α^-	–	–	–	–	0.10	0.03	0.02
	β	0.75	0.75	0.04	0.03	0.75	0.04	0.03
	γ	0.20	0.19	0.05	0.03	0.19	0.05	0.03
TARCH(∞)	ω	0.20				0.21	0.05	0.03
	α^+	0.05				0.06	0.04	0.02
	α^-	0.15				0.16	0.05	0.03
	β	0.70				0.70	0.06	0.04
	γ	0.15				0.14	0.05	0.03

Table 2: *Estimation results for 1000 simulations of size 5000 of a symmetric ARCH(∞) process and a TARCH(∞) process*

Finally, in order to study the empirical behavior of the asymmetry test statistics defined in (15) under the null hypothesis $H_0^{\text{sym}} : \boldsymbol{\phi_0^+} = \boldsymbol{\phi_0^-}$, we performed the tests on each realisation of our ARCH(∞) simulations sample. Figure 2(a) presents kernel density estimators of the three test statistics for $n = 5000$. 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

levels 5%, are respectively 3.80%, 4.10% and 4.50%. On these simulations, the Wald statistic appears to be slightly more sensible to the Type I error. To study the empirical behavior of these statistics under H_1^{asym}, we also performed the tests on each realisation of our TARCH(∞) simulations sample. 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 5000. On these simulations, all the statistics seem to be powerful.



(a) Comparison between kernel density estimators and the χ_1^2 density on [0.5,∞) (red solid line) on 1000 simulations of a symmetric ARCH(∞) process for sample size n = 5000.

(b) Comparison of the observed powers as a function of the nominal level α, on 1000 simulations of an asymmetric TARCH(∞) process for sample size n = 5000.

Figure 2: Empirical behavior of the of the Wald (dark blue square), the Rao-score (light blue dot), and the LR (blue cross) test statistics under the null hypothesis H_0^{sym} and the alternative hypothesis H_1^{asym}.

5 Application: Are GARCH(1,1)-type models suitable for lightly traded 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) 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 May 2020[6] of 30 indices

[6]Data for the FTSE MIB and the MOEX start respectively in September 1997 and January

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 3 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 classic result in the financial literature. However, almost half of the thirty indices reject the hypothesis of a GARCH(1,1) specification at the 5% level, and a third reject the GJR-GARCH(1,1) model. Interestingly, all the indices that reject H_0^{GARCH} 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 Efficient Market Hypothesis (EMH) introduced by Fama[19]. Under the EMH, agents on the market should be able to correct any price shock almost instantaneously, inducing a very short persistence. On the contrary, in a market with fewer investors and with less liquid instruments, the integration of a shock into the prices is less efficient, resulting in higher persistence. A similar argument was used by Di Matteo et al.[14, 15]. The authors studied the value of the General Hurst Exponent to determine the maturity of a financial market. They show that the less developed a market, the stronger the persistence.

The results in Table 3 provide a compelling argument in favor of applying models with strong persistence to less developed markets. Additionally, if this persistence stems from a market inefficiency, it should also imply heterogeneity at the stock level within developed markets. Assets that are less traded should therefore exhibit stronger persistence than highly traded ones. We propose to test this hypothesis by studying the Fama and French[20] Size equity portfolios. Every year, the authors sort in ascending order of market equity all the NYSE, AMEX, and NASDAQ stocks and construct 10 decile portfolios labelled "Dec1", "Dec2", etc. Our dataset contains their daily returns from January 1975 to March 2020 and was obtained from Kenneth French's website[7]. For each portfolio, we compute the Wald statistic W_n^{GJR} to test the hypothesis that the GJR-GARCH(1,1) is well suited to model returns series. The results are presented in Figure 3. It is clear that on our whole sample, the GJR-GARCH(1,1) specification is strongly rejected for smaller Size portfolios. Moreover, we find that the portfolios composed of large stocks do not exhibit strong persistence, which confirms our finding for the US indices in Table 3.

1998.

[7]https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Index	Country	Symmetry test			GARCH test			
		W_n^{sym}	R_n^{sym}	L_n^{sym}	$\tilde{\gamma}_n^{\text{GARCH}}$	W_n^{GARCH}	$\tilde{\gamma}_n^{\text{GJR}}$	W_n^{GJR}
S&P 500	United States	**0.000**	**0.000**	**0.000**	0.000	0.500	0.000	0.500
Nasdaq 100	United States	**0.000**	**0.000**	**0.000**	0.000	0.500	0.000	0.500
S&P/TSX Composite	Canada	**0.000**	**0.000**	**0.000**	0.022	0.265	0.000	0.500
Mexico IPC	Mexico	**0.000**	**0.000**	**0.000**	0.116	**0.000**	0.037	0.072
FTSE 100	UK	**0.000**	**0.000**	**0.000**	0.037	0.093	0.000	0.500
DAX 30	Germany	**0.000**	**0.000**	**0.000**	0.000	0.500	0.000	0.500
CAC 40	France	**0.000**	**0.000**	**0.000**	0.000	0.500	0.000	0.500
Swiss Market (SMI)	Switzerland	**0.000**	**0.000**	**0.000**	0.000	0.500	0.000	0.500
AEX	Netherlands	**0.000**	**0.000**	**0.000**	0.000	0.500	0.000	0.500
FTSE MIB	Italy	**0.000**	**0.000**	**0.000**	0.000	0.500	0.000	0.500
IBEX 35	Spain	**0.000**	**0.000**	**0.000**	0.000	0.500	0.000	0.500
MOEX Russia Index	Russia	**0.002**	**0.000**	**0.000**	0.078	**0.045**	0.071	**0.049**
Wig (WIG)	Poland	**0.000**	**0.000**	**0.000**	0.053	**0.022**	0.029	0.123
Budapest BUX (BUX)	Hungary	**0.000**	**0.000**	**0.000**	0.193	**0.000**	0.146	**0.000**
Tel Aviv 125	Israel	**0.000**	**0.000**	**0.000**	0.068	**0.023**	0.000	0.500
Nikkei 225	Japan	**0.000**	**0.000**	**0.000**	0.000	0.499	0.000	0.500
KOSPI	South Korea	**0.000**	**0.000**	**0.000**	0.001	0.489	0.000	0.500
Hang Seng (HS)	Hong Kong	**0.000**	**0.000**	**0.000**	0.000	0.500	0.000	0.500
Taiwan Weighted (TAIEX)	Taiwan	**0.000**	**0.000**	**0.000**	0.009	0.379	0.000	0.500
MSCI Singapore	Singapore	**0.000**	**0.000**	**0.000**	0.057	**0.024**	0.034	0.104
Bangkok SET (BSET)	Thailand	**0.000**	**0.000**	**0.000**	0.089	**0.001**	0.056	**0.031**
PSE Composite (PSEi)	Philippines	**0.000**	**0.000**	**0.000**	0.088	0.124	0.141	**0.002**
IDX Composite	Indonesia	**0.001**	**0.001**	**0.000**	0.307	**0.000**	0.194	**0.000**
FTSE Bursa Malaysia KLCI	Malaysia	**0.000**	**0.000**	**0.000**	0.156	**0.000**	0.115	**0.000**
NIFTY 50	India	**0.000**	**0.000**	**0.000**	0.093	**0.006**	0.022	0.288
ASX All Ordinaries (AO)	Australia	**0.000**	**0.000**	**0.000**	0.067	**0.011**	0.002	0.470
MSCI New Zealand	New Zealand	**0.043**	0.072	0.068	0.128	**0.000**	0.109	**0.000**
Merval (ME)	Argentina	**0.003**	**0.000**	**0.000**	0.175	**0.000**	0.142	**0.001**
Bovespa (BO)	Brazil	**0.000**	**0.000**	**0.000**	0.034	0.139	0.000	0.500
S&P/BVL Peru General Index (IGBVL)	Peru	**0.001**	**0.000**	**0.000**	0.211	**0.000**	0.204	**0.000**

Table 3: *Asymmetry and GARCH-memory tests results on different equity indices*

However, when focusing on the 1995-onward period, the results are quite different. Indeed, we do not reject the GJR-GARCH(1,1) for the low Size portfolios, which would imply that Size stocks have become more efficient. A possible explanation could be that the search for higher returns has sparked investors' interest in small stocks, resulting in more efficient market conditions. Actually, the Size premium's existence has been disputed since the early nineties (see for example van Dijk[42]), which would support our tests results on the above mentioned time frame.

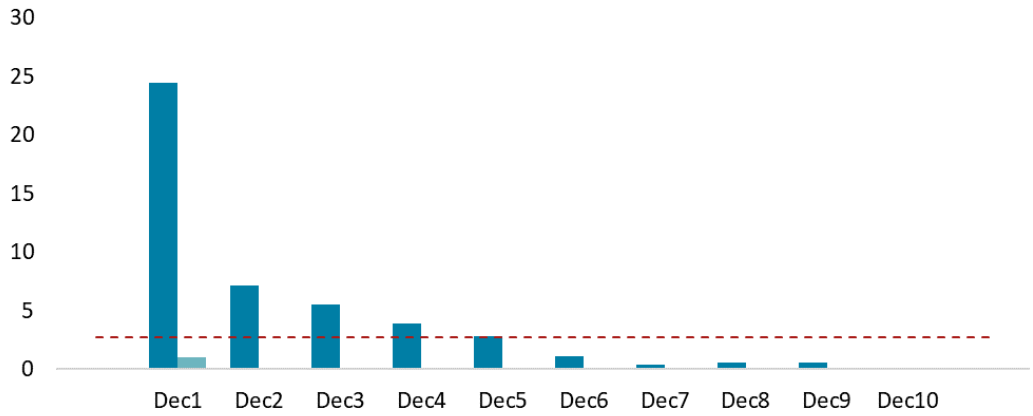


Figure 3: Wald statistics W_n^{GJR} computed on the portfolios formed on Size from 1975 (in blue) and from 1995 (in light blue). The rejection threshold of H_0^{GJR} at the 5% asymptotic level is represented by the red dashed line.

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 propose an extension of the ARCH(∞) model of Robinson[38] to account for high persistence in squared returns and conditional asymmetry. We prove the existence of a stationary solution and we derive statistical inference results, in particular, we prove the consistency and asymptotic normality of Quasi Maximum Likelihood estimator. We show that the TARCH(∞) representation nests some of the most used models in the financial industry. Those specifications allow us to derive tests procedures for conditional asymmetry, and to verify that the GARCH(1,1) memory pattern is 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 not heavily traded assets. We show that in our database, most of the emerging markets equity indices exhibit a stronger persis-

tence than the GARCH(1,1) allows. The same feature is also observable for low Size portfolios on developed markets, although it seems to be less noticeable in the most recent years. In light of these results, potential extensions include the study of the conditional risk measures behavior under strong persistence specifications. We leave this inference problem for future research.

References

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

[2] D. W. Andrews and W. Ploberger. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, pages 1383–1414, 1994.

[3] R. T. Baillie, T. Bollerslev, and H. O. Mikkelsen. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 74(1):3–30, 1996.

[4] G. Bekaert and G. Wu. Asymmetric volatility and risk in equity markets. *The review of financial studies*, 13(1):1–42, 2000.

[5] J. Beran, Y. Feng, S. Ghosh, and R. Kulik. *Long-Memory Processes.* Springer, 2016.

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

[7] F. Black. Studies of stock price volatility changes. *Proceedings from the American Statistical Association, Business and Economic Statistics Section*, pages 177–181, 1976.

[8] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

[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] J. Y. Campbell and L. Hentschel. No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of financial Economics*, 31(3):281–318, 1992.

[11] A. A. Christie. The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics*, 10(4):407–432, 1982.

[12] R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1:223–236, 2001.

[13] J. Davidson. Moment and memory properties of linear conditional heteroscedasticity models, and a new model. *Journal of Business & Economic Statistics*, 22(1):16–29, 2004.

[14] T. Di Matteo, T. Aste, and M. M. Dacorogna. Scaling behaviors in differently developed markets. *Physica A: Statistical Mechanics and its Applications*, 324(1):183–188, 2003.

[15] T. Di Matteo, T. Aste, and M. M. Dacorogna. Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development. *Journal of Banking & Finance*, 29(4):827–851, 2005.

[16] Z. Ding and C. W. Granger. Modeling volatility persistence of speculative returns: A new approach. *Journal of Econometrics*, 73(1):185–215, 1996.

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

[18] R. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

[19] E. F. Fama. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417, 1970.

[20] E. F. Fama and K. R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993.

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

[22] 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.

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

[24] K. R. French, G. W. Schwert, and R. F. Stambaugh. Expected stock returns and volatility. *Journal of financial Economics*, 19(1):3–29, 1987.

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

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

[27] L. Giraitis and D. Surgailis. ARCH-type bilinear models with double long memory. *Stochastic Processes and their Applications*, 100(1):275–300, 2002.

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

[29] 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.

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

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

[32] B. E. Hansen. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, pages 413–430, 1996.

[33] H.-C. Ho. Estimation errors of the sharpe ratio for long-memory stochastic volatility models. In *Time Series and Related Topics*, pages 165–172. Institute of Mathematical Statistics, 2006.

[34] H.-C. Ho. Sample quantile analysis for long-memory stochastic volatility models. *Journal of Econometrics*, 189(2):360–370, 2015.

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

[36] J. Liu, R. Deo, and C. Hurvich. The slow convergence of ordinary least squares estimators of α, β and portfolio weights under long-memory stochastic volatility. *Journal of Time Series Analysis*, 40(4):590–608, 2019.

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

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

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

[40] M. J. Rodríguez and E. Ruiz. Revisiting Several Popular GARCH Models with Leverage Effect: Differences and Similarities. *Journal of Financial Econometrics*, 10(4):637–668, 04 2012.

[41] G. W. Schwert. Why does stock market volatility change over time? *The journal of finance*, 44(5):1115–1153, 1989.

[42] M. A. Van Dijk. Is size dead? a review of the size effect in equity returns. *Journal of Banking & Finance*, 35(12):3263–3274, 2011.

[43] 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 TARCH(∞) solution

We develop in this section the proof of Theorem 1 and Corollary 1.1. The proof of the theorem is based on a Volterra expansion and, in this sense, follows the work of Giraitis, Kokoszka and Leipus[25], Kazakevičius and Leipus[35], and Douc, Roueff and Soulier[17].

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

$$S_t = \omega + \omega \sum_{k=1}^{\infty} \sum_{i_1,\ldots,i_k \geq 1} a_{i_1,t-i_1} \ldots a_{i_k,t-i_1-\ldots-i_k} \ \eta^2_{t-i_1} \ldots \eta^2_{t-i_1-\ldots-i_k}$$

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

$$\begin{aligned} \mathbb{E}S_t^s \leq \ & \omega^s + \omega^s \sum_{k=1}^{\infty} \sum_{i_1,\ldots,i_k \geq 1} \mathbb{E}\left(\left[(\alpha^+_{i_1})^s \ \mathbb{1}_{\eta_{t-i_1} \geq 0} + (\alpha^-_{i_1})^s \ \mathbb{1}_{\eta_{t-i_1} < 0}\right] |\eta_{t-i_1}|^{2s}\right) \ldots \\ & \mathbb{E}\left(\left[(\alpha^+_{i_k})^s \ \mathbb{1}_{\eta_{t-i_1-\ldots-i_k} \geq 0} + (\alpha^-_{i_k})^s \ \mathbb{1}_{\eta_{t-i_1-\ldots-i_k} < 0}\right] |\eta_{t-i_1-\ldots-i_k}|^{2s}\right), \end{aligned}$$

and thus

$$\mathbb{E}S_t^s \leq \omega^s \left[1 + \sum_{k=1}^{\infty} (A_s^+ \mu_{2s}^+ + A_s^- \mu_{2s}^-)^k\right] \leq \frac{\omega^s}{1 - (A_s^+ \mu_{2s}^+ + A_s^- \mu_{2s}^-)} < \infty,$$

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

$$\sum_{i=1}^{\infty} a_{i,t-i} S_{t-i} \eta^2_{t-i} = \omega \sum_{k=0}^{\infty} \sum_{i_0,\ldots,i_k \geq 1} a_{i_0,t-i_0} \ldots a_{i_k,t-i_0-\ldots-i_k} \ \eta^2_{t-i_0} \ldots \eta^2_{t-i_0-\ldots-i_k}$$

and thus we obtain the recursive equation $S_t = \omega + \sum_{i=1}^{\infty} a_{i,t-i} S_{t-i} \eta^2_{t-i}$. By setting $\varepsilon_t = \sqrt{S_t}\eta_t$, we obtain a strictly stationary and nonanticipative solution of (3) and $\mathbb{E}|\varepsilon_t|^{2s} \leq \mu_{2s}\omega^s/(1 - (A_s^+ \mu_{2s}^+ + A_s^- \mu_{2s}^-))$.

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

such that $\mathbb{E}|\varepsilon_t^*|^{2s} < \infty$. For all $q \geq 1$, by q recursive substitutions of the ε_{t-i}^{*2}, we obtain

$$\begin{aligned}
\sigma_t^2 &= \omega + \sum_{i=1}^{\infty} a_{i,t-i}\varepsilon_{t-i}^{*2} \\
&= \left\{ \omega + \omega \sum_{k=1}^{q} \sum_{i_1,\ldots,i_k \geq 1} a_{i_1,t-i_1} \ldots a_{i_k,t-i_1-\ldots-i_k} \ \eta_{t-i_1}^2 \ldots \eta_{t-i_1-\ldots-i_k}^2 \right\} \\
&+ \sum_{i_1,\ldots,i_{q+1} \geq 1} a_{i_1,t-i_1} \ldots a_{i_{q+1},t-i_1-\ldots-i_{q+1}} \ \eta_{t-i_1}^2 \ldots \eta_{t-i_1-\ldots-i_q}^2 \varepsilon_{t-i_1-\ldots-i_{q+1}}^{*2} \\
&:= \{S_{t,q}\} + R_{t,q}.
\end{aligned}$$

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

$$\mathbb{E}R_{t,q}^s \leq (A_s^+ \mu_{2s}^+ + A_s^- \mu_{2s}^-)^q \ \left(A_s^+ \mathbb{E}|\mathbb{1}_{\eta_t \geq 0} \varepsilon_t^*|^{2s} + A_s^- \mathbb{E}|\mathbb{1}_{\eta_t < 0} \varepsilon_t^*|^{2s} \right)$$

Since $A_s^+ \mu_{2s}^+ + A_s^- \mu_{2s}^- < 1$, 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 $\sigma_t^2 = S_t$ almost surely and yields $\varepsilon_t^* = \varepsilon_t$ almost surely hence concluding the proof. □

Proof of Corollary 1.1. From Theorem 1, we have that

$$\varepsilon_t^2 = \omega \ \eta_t^2 \left[1 + \sum_{k=1}^{\infty} \sum_{i_1,\ldots,i_k \geq 1} a_{i_1,t-i_1} \ldots a_{i_k,t-i_1-\ldots-i_k} \ \eta_{t-i_1}^2 \ldots \eta_{t-i_1-\ldots-i_k}^2 \right]$$

and using the c_r-inequality, we obtain

$$\left(\mathbb{E}\varepsilon_t^4\right)^{\frac{1}{2}} \leq \omega \ \left(\mathbb{E}\eta_t^4\right)^{\frac{1}{2}} \left[1 + \sum_{k=1}^{\infty} \sum_{i_1,\ldots,i_k \geq 1} \left(\mathbb{E} \left[a_{i_1,t-i_1}^2 \ldots a_{i_k,t-i_1-\ldots-i_k}^2 \ \eta_{t-i_1}^4 \ldots \eta_{t-i_1-\ldots-i_k}^4 \right] \right)^{\frac{1}{2}} \right]$$

where, for all $i \geq 1$,

$$a_{i,t-i}^2 = \left[\alpha_i^+ \mathbb{1}_{\eta_{t-i} \geq 0} + \alpha_i^- \ \mathbb{1}_{\eta_{t-i} < 0} \right]^2 = (\alpha_i^+)^2 \mathbb{1}_{\eta_{t-i} \geq 0} + (\alpha_i^-)^2 \mathbb{1}_{\eta_{t-i} < 0},$$

hence

$$\left(\mathbb{E}\varepsilon_t^4\right)^{\frac{1}{2}} \leq \frac{\omega \ \left(\mathbb{E}\eta_t^4\right)^{\frac{1}{2}}}{1 - \left(A_2^+ \mu_4^+ + A_2^- \mu_4^-\right)^{\frac{1}{2}}}$$

and thus ε_t admits a fourth moment if $\left(A_2^+ \mu_4^+ + A_2^- \mu_4^-\right)^{\frac{1}{2}} < 1$. □

A.2 Statistical inference of a Threshold ARCH(∞) 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.

Let us define the theoretical criterion

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

The theoretical QML estimator $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}$ 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 $\mathcal{I}(\boldsymbol{\phi}) = \{i \text{ such that } \alpha_i(\boldsymbol{\phi}) \neq 0\}$, and we define $\mathcal{I}_t^+$ (respectively $\mathcal{I}_t^-$) as $\mathcal{I}_t^{+(-)} = \{i \text{ such that } \varepsilon_{t-i} \geq 0 \text{ (resp. } < 0)\}$, yielding the following rewriting of (9) as

$$\sigma_t^2(\boldsymbol{\theta_0}) = \omega_0 + \sum_{i\in\mathcal{I}_t^+} \alpha_i(\boldsymbol{\phi_0^+})\varepsilon_{t-i}^2 + \sum_{j\in\mathcal{I}_t^-} \alpha_j(\boldsymbol{\phi_0^-})\varepsilon_{t-j}^2. \tag{18}$$

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

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

$$\sup_{i\in\mathcal{I}(\boldsymbol{\phi_0^+})} \sup_{\phi\in\boldsymbol{\Phi}} \frac{\alpha_i(\boldsymbol{\phi_0^+})}{\alpha_i^{1-\tau}(\boldsymbol{\phi})} \leq K \text{ and } \sup_{i\in\mathcal{I}(\boldsymbol{\phi_0^-})} \sup_{\phi\in\boldsymbol{\Phi}} \frac{\alpha_i(\boldsymbol{\phi_0^-})}{\alpha_i^{1-\tau}(\boldsymbol{\phi})} \leq K, \tag{19}$$

then

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

Proof of Proposition 2. Using the fact that for any s>0, we have $x/(1+x) \leq x^s$, Equation (18) gives

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})} \\
\leq\ & K + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{i\in\mathcal{I}_t^+\cap\mathcal{I}(\boldsymbol{\phi_0^+})} \frac{\alpha_i(\boldsymbol{\phi_0^+})\varepsilon_{t-i}^2}{\omega + \alpha_i(\boldsymbol{\phi^+})\varepsilon_{t-i}^2} + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{i\in\mathcal{I}_t^-\cap\mathcal{I}(\boldsymbol{\phi_0^-})} \frac{\alpha_i(\boldsymbol{\phi_0^-})\varepsilon_{t-i}^2}{\omega + \alpha_i(\boldsymbol{\phi^-})\varepsilon_{t-i}^2} \\
\leq\ & K + K\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{i\in\mathcal{I}_t^+\cap\mathcal{I}(\boldsymbol{\phi_0^+})} \alpha_i^{s-\tau}(\boldsymbol{\phi^+})\varepsilon_{t-i}^{2s} + K\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{i\in\mathcal{I}_t^-\cap\mathcal{I}(\boldsymbol{\phi_0^-})} \alpha_i^{s-\tau}(\boldsymbol{\phi^-})\varepsilon_{t-i}^{2s} \\
\leq\ & K + K\sum_{i=1}^{\infty} i^{-(d+1)(s-\tau)}\varepsilon_{t-i}^{2s}
\end{aligned}$$

using **A3(ii)** and (19). By taking $s = \rho$, **A4** thus yields

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

□

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) Asymptotic irrelevance of the initial values
$\lim_{n\to\infty} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |Q_n(\boldsymbol{\theta}) - \tilde{Q}_n(\boldsymbol{\theta})| = 0$

(b) Identifiability of the parameter
$(\exists t \in \mathbb{Z}$ such that $\sigma_t^2(\boldsymbol{\theta}) = \sigma_t^2(\boldsymbol{\theta_0})$ a.s.$) \Rightarrow \boldsymbol{\theta} = \boldsymbol{\theta_0}$

(c) The limit criterion is minimized at the true value
$\mathbb{E}_{\boldsymbol{\theta_0}}|l_t(\boldsymbol{\theta_0})| < \infty$, 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) Compactness of $\boldsymbol{\Theta}$ and ergodicity of $(l_t(\boldsymbol{\theta}))$
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.

This structure of the proof follows the ones for the earlier results in the classical GARCH case and is similar to the one presented in Francq and Zakoian[23]. However, the demonstrations of the intermediate results differ and, notably, the impact of the initial values requires a particular attention. The last result being similar to the GARCH case, we let the reader refer to Chapter 7 in [23] and we focus on showing the first three previous points.

(a) Asymptotic irrelevance of the initial values
Consider

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

and remark that from (9) and (12) and Assumption **A1**, we have

$$\sigma_t^2(\boldsymbol{\theta}) = \tilde{\sigma}_t^2(\boldsymbol{\theta}) + \sum_{i=t}^{\infty} a_{i,t-i}(\boldsymbol{\phi}^+, \boldsymbol{\phi}^-)\varepsilon_{t-i}^2 \geq \tilde{\sigma}_t^2(\boldsymbol{\theta}). \tag{20}$$

We denote $\chi_t = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\sigma_t^2(\boldsymbol{\theta}) - \tilde{\sigma}_t^2(\boldsymbol{\theta})|$, and we have from (20), (11), and **A3(ii)**

$$\chi_t = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{i=t}^{\infty} a_{i,t-i}(\boldsymbol{\phi}^+, \boldsymbol{\phi}^-)\varepsilon_{t-i}^2 \leq K\sum_{i=0}^{\infty} (i+t)^{-(d+1)}\varepsilon_{-i}^2,$$

whence $\mathbb{E}\chi_t^\rho \leq K\sum_{i=0}^{\infty} (i+t)^{-(d+1)\rho}\ \mathbb{E}|\varepsilon_{-i}|^{2\rho}$. From Assumption **A4**, $\mathbb{E}\,|\varepsilon_t|^{2\rho} < \infty$, with $\rho(d+1) > 1$, and since, for any $k > 1$, $\int_t^\infty x^{-k}dx = (k-1)^{-1}t^{-k+1}$, 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,

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

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

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

whence $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |B_n(\boldsymbol{\theta})| \leq \frac{K}{n}\sum_{t=1}^{n}\eta_t^2 \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\ \chi_t$. By ergodicity and independence 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 **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. Thus, we can conclude $\lim_{n\to\infty}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |Q_n(\boldsymbol{\theta}) - \tilde{Q}_n(\boldsymbol{\theta})| = 0$.

(b) Identifiability of the parameter:
Let $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, such that, for some $t \in \mathbb{Z}$, we have $\sigma_t^2(\boldsymbol{\theta}) = \sigma_t^2(\boldsymbol{\theta_0})$ almost surely. Assume $\boldsymbol{\theta} \neq \boldsymbol{\theta_0}$, and suppose that

$$\alpha_1(\boldsymbol{\phi_0^+})\mathbb{1}_{\varepsilon_{t-1}\geq 0} + \alpha_1(\boldsymbol{\phi_0^-})\mathbb{1}_{\varepsilon_{t-1}<0} \neq \alpha_1(\boldsymbol{\phi}^+)\mathbb{1}_{\varepsilon_{t-1}\geq 0} + \alpha_1(\boldsymbol{\phi}^-)\mathbb{1}_{\varepsilon_{t-1}<0}. \tag{21}$$

Then we can show that $\left[\alpha_1(\boldsymbol{\phi_0^+}) - \alpha_1(\boldsymbol{\phi}^+)\right]\ \mathbb{1}_{\eta_{t-1}\geq 0}\eta_{t-1}^2 + \left[\alpha_1(\boldsymbol{\phi_0^-}) - \alpha_1(\boldsymbol{\phi}^-)\right]\ \mathbb{1}_{\eta_{t-1}<0}\eta_{t-1}^2$ 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 yields almost surely $\alpha_1(\boldsymbol{\phi_0^+}) = \alpha_1(\boldsymbol{\phi^+})$ and $\alpha_1(\boldsymbol{\phi_0^-}) = \alpha_1(\boldsymbol{\phi^-})$ which contradicts (21). 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 $\alpha_i(\boldsymbol{\phi_0^-}) = \alpha_i(\boldsymbol{\phi^-})$. Thus, from **A3(i)**, $\boldsymbol{\phi^+} = \boldsymbol{\phi_0^+}$ and $\boldsymbol{\phi^-} = \boldsymbol{\phi_0^-}$ a.s., whence $\omega = \omega_0$ a.s., and thus $\boldsymbol{\theta} = \boldsymbol{\theta_0}$ a.s.

(c) The limit criterion is minimized at the true value:
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

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

Furthermore, we can show that it is integrable at $\boldsymbol{\theta_0}$. Using Jensen inequality, (11) and Assumption **A3(ii)**, we obtain

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

since, from **A4**, $\mathbb{E}\,|\varepsilon_t|^{2\rho} < \infty$ 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 $\boldsymbol{\theta} = \boldsymbol{\theta_0}$.

The conclusion of the proof uses a compactness argument. First note that for any neighborhood $V(\boldsymbol{\theta_0})$ of $\boldsymbol{\theta_0}$,

$$\limsup_{n\to\infty} \inf_{\boldsymbol{\theta^*} \in V(\boldsymbol{\theta_0})} \tilde{Q}_n(\boldsymbol{\theta^*}) \leq \lim_{n\to\infty} \tilde{Q}_n(\boldsymbol{\theta_0}) \leq \lim_{n\to\infty} Q_n(\boldsymbol{\theta_0}) \leq \mathbb{E}_{\boldsymbol{\theta_0}} l_1(\boldsymbol{\theta_0}).$$

The compact set $\boldsymbol{\Theta}$ is covered by the union of an arbitrary neighborhood $V(\boldsymbol{\theta_0})$ of $\boldsymbol{\theta_0}$ and the set of the neighborhoods $V(\boldsymbol{\theta})$ satisfying $\liminf_{n\to\infty} \inf_{\boldsymbol{\theta^*} \in V(\boldsymbol{\theta})} \tilde{Q}_n(\boldsymbol{\theta^*}) \geq \mathbb{E}_{\boldsymbol{\theta_0}} l_1(\boldsymbol{\theta_0})$, for $\boldsymbol{\theta} \in \boldsymbol{\Theta}/V(\boldsymbol{\theta_0})$. Thus, there exists a finite subcover of $\boldsymbol{\Theta}$ of the form $V(\boldsymbol{\theta_0})$, $V(\boldsymbol{\theta_1})$, ..., $V(\boldsymbol{\theta_k})$, whence

$$\inf_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \tilde{Q}_n(\boldsymbol{\theta}) = \min_{i=0,1,\ldots,k} \inf_{\boldsymbol{\theta^*} \in V(\boldsymbol{\theta_i}) \cap \boldsymbol{\Theta}} \tilde{Q}_n(\boldsymbol{\theta}).$$

We obtain that, for n large enough, $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}$ belongs to $V(\boldsymbol{\theta_0})$ a.s. Since this is true for any neighborhood $V(\boldsymbol{\theta_0})$, we have shown that $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} \underset{n\to\infty}{\to} \boldsymbol{\theta_0}$ a.s. □

We will now state and prove the proposition mentioned in the remark about **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*

$$\sup_{i\in\mathcal{I}(\boldsymbol{\phi_0^+})} \sup_{\phi\in V(\boldsymbol{\phi_0^+})} \frac{\alpha_i(\boldsymbol{\phi_0^+})}{\alpha_i^{1-\tau}(\boldsymbol{\phi})} \leq K \text{ and } \sup_{i\in\mathcal{I}(\boldsymbol{\phi_0^-})} \sup_{\phi\in V(\boldsymbol{\phi_0^-})} \frac{\alpha_i(\boldsymbol{\phi_0^-})}{\alpha_i^{1-\tau}(\boldsymbol{\phi})} \leq K. \tag{22}$$

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 for all $k > s$, (10) and Hölder inequality yield

$$\begin{aligned}\sigma_t^2(\boldsymbol{\theta_0}) &= \omega_0\omega^{\frac{s}{k}-1}\omega^{1-\frac{s}{k}} + \sum_{i=1}^{\infty} a_{i,t-i}(\boldsymbol{\phi_0^+},\boldsymbol{\phi_0^-}) a_{i,t-i}^{\frac{s}{k}-1}(\boldsymbol{\phi^+},\boldsymbol{\phi^-}) a_{i,t-i}^{1-\frac{s}{k}}(\boldsymbol{\phi^+},\boldsymbol{\phi^-}) \varepsilon_{t-i}^{2\frac{s}{k}} \varepsilon_{t-i}^{2-2\frac{s}{k}} \\ &\leq K\left[\omega_0^{\frac{k}{s}}\omega^{1-\frac{k}{s}} + \sum_{i=1}^{\infty} a_{i,t-i}^{\frac{k}{s}}(\boldsymbol{\phi_0^+},\boldsymbol{\phi_0^-}) a_{i,t-i}^{1-\frac{k}{s}}(\boldsymbol{\phi^+},\boldsymbol{\phi^-})\varepsilon_{t-i}^2\right]^{\frac{s}{k}} \left[\sigma_t^2(\boldsymbol{\theta})\right]^{1-\frac{s}{k}}.\end{aligned}$$

Since $\left[\sigma_t^2(\boldsymbol{\theta})\right]^{-\frac{s}{k}} \leq K$, we obtain

$$\left[\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\right]^k \leq K\left[1 + \sum_{\mathcal{I}_t^+\cap\mathcal{I}(\boldsymbol{\phi_0^+})} \frac{\alpha_i^k(\boldsymbol{\phi_0^+})}{\alpha_i^k(\boldsymbol{\phi^+})}\alpha_i^s(\boldsymbol{\phi^+})\varepsilon_{t-i}^{2s} + \sum_{\mathcal{I}_t^-\cap\mathcal{I}(\boldsymbol{\phi_0^-})} \frac{\alpha_i^k(\boldsymbol{\phi_0^-})}{\alpha_i^k(\boldsymbol{\phi^-})}\alpha_i^s(\boldsymbol{\phi^-})\varepsilon_{t-i}^{2s}\right]$$

whence, from (22) and Assumptions **A3(ii)** and **A4**, by taking $s = \rho$, there exists a neighborhood such that

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

Indeed, from the arbitrariness of τ in (22), we can find a τ such that $(d+1)(\rho - k\tau) > 1$. □

Before developing the proof of Theorem 3, it is useful to state the following lemma.

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

$$\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial^k\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}...\partial\theta_{i_k}}\right|^p < \infty.$$

Proof of Lemma 1. From (18) and **A10(i)**, we have, for all $j_1, j_2 \in \{1, ..., r\}$,

$$\frac{\partial \sigma_t^2}{\partial \theta_1} = \frac{\partial \sigma_t^2}{\partial \omega} = 1, \ \frac{\partial \sigma_t^2}{\partial \theta_{1+j_1}} = \sum_{i \in \mathcal{I}_t^+} \frac{\partial \alpha_i}{\partial \phi_{j_1}^+} \varepsilon_{t-i}^2 \ \text{ and } \ \frac{\partial \sigma_t^2}{\partial \theta_{1+r+j_2}} = \sum_{i \in \mathcal{I}_t^-} \frac{\partial \alpha_i}{\partial \phi_{j_2}^-} \varepsilon_{t-i}^2 \tag{23}$$

whence $\dfrac{\partial^2 \sigma_t^2}{\partial \omega \partial \theta_{i_1}} = 0$ and $\dfrac{\partial^2 \sigma_t^2}{\partial \phi_{j_1}^+ \partial \omega} = \dfrac{\partial^2 \sigma_t^2}{\partial \phi_{j_1}^+ \partial \phi_{j_2}^-} = \dfrac{\partial^2 \sigma_t^2}{\partial \phi_{j_2}^- \partial \omega} = 0$. It is thus sufficient to show that for all $j_{1,h}, j_{2,h} \in \{1, ..., r\}$, $h = 1, ..., k$, $k \leq 3$, we have

$$\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \frac{1}{\sigma_t^2(\boldsymbol{\theta})} \frac{\partial^k \sigma_t^2(\boldsymbol{\theta})}{\partial \phi_{j_{1,1}}^+ ... \partial \phi_{j_{1,k}}^+} \right|^p < \infty \text{ and } \mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \frac{1}{\sigma_t^2(\boldsymbol{\theta})} \frac{\partial^k \sigma_t^2(\boldsymbol{\theta})}{\partial \phi_{j_{2,1}}^- ... \partial \phi_{j_{2,k}}^-} \right|^p < \infty.$$

From (23) we have

$$\left| \frac{\partial^k \sigma_t^2(\boldsymbol{\theta})}{\partial \phi_{j_{1,1}}^+ ... \partial \phi_{j_{1,k}}^+} \right| \leq \sum_{i \in \mathcal{I}_t^+} \left| \frac{\partial^k \alpha_i(\boldsymbol{\phi}^+)}{\partial \phi_{j_{1,1}}^+ ... \partial \phi_{j_{1,k}}^+} \right| \left[\varepsilon_{t-i}^2\right]^{\frac{\rho}{p}} \left[\varepsilon_{t-i}^2\right]^{1-\frac{\rho}{p}} \left[\alpha_i(\boldsymbol{\phi}^+)\right]^{1-\frac{\rho}{p}} \left[\alpha_i(\boldsymbol{\phi}^+)\right]^{\frac{\rho}{p}-1}$$

and from the Hölder inequality, **A3(ii)** and **A10(i)**, we obtain, for any $p > \rho$,

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

for all $\xi > 0$. We may choose ξ such that $(d+1)(\rho - p\xi) > 1$ and thus we have $\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \frac{1}{\sigma_t^2(\boldsymbol{\theta})} \frac{\partial^k \sigma_t^2(\boldsymbol{\theta})}{\partial \phi_{j_{1,1}}^+ ... \partial \phi_{j_{1,k}}^+} \right|^p < \infty$ and similarly $\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \frac{1}{\sigma_t^2(\boldsymbol{\theta})} \frac{\partial^k \sigma_t^2(\boldsymbol{\theta})}{\partial \phi_{j_{2,1}}^- ... \partial \phi_{j_{2,k}}^-} \right|^p < \infty$ allowing us to conclude.

□

We can now develop the proof of Theorem 3 on the asymptotic normality of the QMLE.

Proof of Theorem 3. From Theorem 2, we have that $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}$ converges to $\boldsymbol{\theta_0}$ which, from **A6**, belongs in 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 $\boldsymbol{\theta}_{\boldsymbol{ij}}^*$ are between $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}$ and $\boldsymbol{\theta_0}$.

We will show the result by proving that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial \tilde{l}_t}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta_0}) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, (\kappa_\eta - 1)\boldsymbol{J})$$

and that

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\tilde{l}_t}{\partial\theta_i\partial\theta_j}(\boldsymbol{\theta}^*_{\boldsymbol{ij}}) \to J(i,j) \text{ in probability.}$$

To do so, we will show the following intermediate results:

(a) Integrability of the derivatives of the criterion at $\boldsymbol{\theta_0}$
$\mathbb{E}_{\boldsymbol{\theta_0}}\left\|\frac{\partial l_t(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}}\frac{\partial l_t(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta'}}\right\| < \infty$, $\mathbb{E}_{\boldsymbol{\theta_0}}\left\|\frac{\partial^2 l_t(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta'}}\right\| < \infty$

(b) Invertibility of $\boldsymbol{J}$ and connection with the variance of the criterion derivative
$\boldsymbol{J}$ is invertible and $\mathbb{V}_{\boldsymbol{\theta_0}}\left[\frac{\partial l_t(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}}\right] = (\kappa_\eta - 1)\boldsymbol{J}$

(c) Uniform integrability of the third-order derivatives of the criterion
There exists a neighborhood $V(\boldsymbol{\theta_0})$ of $\boldsymbol{\theta_0}$ such that, for all $k_1, k_2, k_3 \in \{1, ..., 2r+1\}$,

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

(d) Asymptotic decrease of the effect of the initial values
$\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{\partial l_t(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}} - \frac{\partial\tilde{l}_t(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}}\right)\right\|$ and $\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})}\left\|\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta'}} - \frac{\partial^2\tilde{l}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta'}}\right)\right\|$ tend to 0 in probability as n tends to infinity.

(e) Central Limit Theorem for martingale increments
$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta_0}) \xrightarrow{\mathcal{L}} \mathcal{N}(\boldsymbol{0}, (\kappa_\eta - 1)\boldsymbol{J})$

(f) Use of a second Taylor expansion and of the ergodic theorem
$\sum_{t=1}^{n}\frac{\partial^2 l_t}{\partial\theta_i\partial\theta_j}(\boldsymbol{\theta}^*_{\boldsymbol{ij}}) \to J(i,j)$ in probability

The last two results being similar to the GARCH case, we let the reader refer to chapter 7 in [23] and we focus on showing the first four previous points.

(a) Integrability of the derivatives of the criterion at $\boldsymbol{\theta_0}$

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

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

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

$$\mathbb{E}_{\boldsymbol{\theta_0}} \left\| \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta_0}) \right\| < \infty, \ \mathbb{E}_{\boldsymbol{\theta_0}} \left\| \frac{1}{\sigma_t^2} \frac{\partial^2 \sigma_t^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta'}}(\boldsymbol{\theta_0}) \right\| < \infty, \ \mathbb{E}_{\boldsymbol{\theta_0}} \left\| \frac{1}{\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}} \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta'}}(\boldsymbol{\theta_0}) \right\| < \infty.$$

The first and the second inequalities directly follow from (18) and Lemma 1, while the boundedness at $\boldsymbol{\theta_0}$ of $\dfrac{1}{\sigma_t^2}\dfrac{\partial \sigma_t^2}{\partial \omega}$, $\dfrac{1}{\sigma_t^2}\dfrac{\partial \sigma_t^2}{\partial \phi_{j1}^+}$ and $\dfrac{1}{\sigma_t^2}\dfrac{\partial \sigma_t^2}{\partial \phi_{j_2}^-}$ entail the last inequality.

(b) Invertibility of $\boldsymbol{J}$ and connection with the variance of the criterion derivative

Since 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, we have

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

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

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

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

$$\lambda_0 + \sum_{i=1}^{\infty} \left[\sum_{j=1}^{r} \lambda_j^+ \frac{\partial \alpha_i(\boldsymbol{\phi_0^+})}{\partial \phi_j^+} \mathbb{1}_{\varepsilon_{t-i} \geq 0} + \sum_{k=1}^{r} \lambda_k^- \frac{\partial \alpha_i(\boldsymbol{\phi_0^-})}{\partial \phi_k^-} \mathbb{1}_{\varepsilon_{t-i} < 0} \right] \varepsilon_{t-i}^2 = 0.$$

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

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

whence $\eta_{t-1}^2 \in \mathcal{F}_{t-2}$ and thus, by independence, $\sum_{j=1}^{r} \lambda_j^+ \dfrac{\partial \alpha_1(\boldsymbol{\phi_0^+})}{\partial \phi_j^+} \mathbb{1}_{\eta_{t-1} \geq 0} \eta_{t-1}^2$ is constant a.s. However, from **A2**, η_1 takes at least two positive values, which

implies $\boldsymbol{\lambda}^{+\prime}\dfrac{\partial\alpha_1(\boldsymbol{\phi}_{\mathbf{0}}^{+})}{\partial\boldsymbol{\phi}^{+}} = 0$ a.s. Iterating this argument we obtain that for all $i_h = i_h(\boldsymbol{\phi}_{\mathbf{0}}^{+}), i_h = 1, ..., r$, we have $\boldsymbol{\lambda}^{+\prime}\dfrac{\partial\alpha_{i_h}(\boldsymbol{\phi}_{\mathbf{0}}^{+})}{\partial\boldsymbol{\phi}^{+}} = 0$ and thus from **A10(ii)** we must have $\boldsymbol{\lambda}^{+} = \mathbf{0}$. Similarly, we obtain $\boldsymbol{\lambda}^{-} = \mathbf{0}$. This implies $\lambda_0 = 0$ and contradicts the singularity of $\boldsymbol{J}$.

(c) Uniform integrability of the third-order derivatives of the criterion
We have

$$\begin{aligned}\frac{\partial^3 l_t(\boldsymbol{\theta})}{\partial\theta_{i_1}\partial\theta_{i_2}\partial\theta_{i_3}} &= \left[1-\frac{\varepsilon_t^2}{\sigma_t^2}\right]\left[\frac{1}{\sigma_t^2}\frac{\partial^3\sigma_t^2}{\partial\theta_{i_1}\partial\theta_{i_2}\partial\theta_{i_3}}\right]\\ &+\left[2\frac{\varepsilon_t^2}{\sigma_t^2}-1\right]\left[\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_{i_1}}\right]\left[\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta_{i_2}\partial\theta_{i_3}}\right]\\ &+\left[2\frac{\varepsilon_t^2}{\sigma_t^2}-1\right]\left[\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_{i_2}}\right]\left[\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta_{i_1}\partial\theta_{i_3}}\right]\\ &+\left[2\frac{\varepsilon_t^2}{\sigma_t^2}-1\right]\left[\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_{i_3}}\right]\left[\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta_{i_1}\partial\theta_{i_2}}\right]\\ &+\left[2-6\frac{\varepsilon_t^2}{\sigma_t^2}\right]\left[\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_{i_1}}\right]\left[\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_{i_2}}\right]\left[\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_{i_3}}\right](\boldsymbol{\theta}).\end{aligned}$$

From Assumptions **A7** and **A11**, and the triangle inequality, there exists a neighborhood $V(\boldsymbol{\theta}_{\mathbf{0}})$ of $\boldsymbol{\theta}_{\mathbf{0}}$ such that, $\left\|\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_{\mathbf{0}})}\dfrac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})}\right\|_2 = \sqrt{\kappa_\eta}\left\|\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_{\mathbf{0}})}\dfrac{\sigma_t^2(\boldsymbol{\theta}_{\mathbf{0}})}{\sigma_t^2(\boldsymbol{\theta})}\right\|_2 < \infty$.
Using Lemma 1, the Cauchy-Schwartz inequality, and the Hölder inequality, we have for all $i_1, i_2, i_3 \in \{1, ..., 2r+1\}$

$$\mathbb{E}_{\boldsymbol{\theta}_{\mathbf{0}}}\sup_{\boldsymbol{\theta}\in V_\tau(\boldsymbol{\theta}_{\mathbf{0}})}\left|\left[1-\frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})}\right]\left[\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial^3\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}\partial\theta_{i_2}\partial\theta_{i_3}}\right]\right| < \infty,$$

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{\theta}_{\mathbf{0}}}\sup_{\boldsymbol{\theta}\in V_\tau(\boldsymbol{\theta}_{\mathbf{0}})}\left|\left[2\frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})}-1\right]\left[\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right]\left[\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial^2\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_2}\partial\theta_{i_3}}\right]\right|\\ &\leq\left\|\sup_{\boldsymbol{\theta}\in V_\tau(\boldsymbol{\theta}_{\mathbf{0}})}\left|2\frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})}-1\right|\right\|_2\left\|\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right|\right\|_4\left\|\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial^2\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_2}\partial\theta_{i_3}}\right|\right\|_4\\ &<\infty,\end{aligned}$$

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{\theta}_{\mathbf{0}}}\sup_{\boldsymbol{\theta}\in V_\tau(\boldsymbol{\theta}_{\mathbf{0}})}\left|\left[2-6\frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})}\right]\left[\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right]\left[\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_2}}\right]\left[\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_3}}\right]\right|\\ &\leq\left\|\sup_{\boldsymbol{\theta}\in V_\tau(\boldsymbol{\theta}_{\mathbf{0}})}\left|2-6\frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})}\right|\right\|_2\max_h\left\|\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_h}}\right|\right\|_6^3 < \infty,\end{aligned}$$

which concludes the proof.

(d) Asymptotic decrease of the effect of the initial values

Let us first consider the derivatives of $\tilde{\sigma}_t^2$. Similarly to (18), we can rewrite (12) as

$$\tilde{\sigma}_t^2(\boldsymbol{\theta_0}) = \omega_0 + \sum_{i\in\tilde{\mathcal{I}}_t^+} \alpha_i(\boldsymbol{\phi_0^+})\varepsilon_{t-i}^2 + \sum_{j\in\tilde{\mathcal{I}}_t^-} \alpha_j(\boldsymbol{\phi_0^-})\varepsilon_{t-j}^2 \tag{24}$$

where we denote by $\tilde{\mathcal{I}}_t^+$ (respectively $\tilde{\mathcal{I}}_t^-$) the sets

$$\tilde{\mathcal{I}}_t^{+(-)} = \{i < t \text{ such that } \varepsilon_{t-i} \geq 0(< 0)\}.$$

Remark now that, from **A3(ii)** and **A11**, on a neighborhood $V(\boldsymbol{\theta_0})$ of $\boldsymbol{\theta_0}$, we have

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})} \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2(\boldsymbol{\theta})} &= \eta_t^2 \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})} \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left[1 + \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})} \frac{\sigma_t^2(\boldsymbol{\theta}) - \tilde{\sigma}_t^2(\boldsymbol{\theta})}{\tilde{\sigma}_t^2(\boldsymbol{\theta})}\right] \\
&\leq K\eta_t^2 \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})} \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left[1 + \sum_{i=0}^{\infty} i^{-(d+1)}\varepsilon_{-i}^2\right] \\
&\leq K\eta_t^2 \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})} \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}
\end{aligned} \tag{25}$$

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.

In addition, we can write

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

and from the Markov inequality we have

$$\begin{aligned}
&\mathbb{P}\left[\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[\frac{\partial l_t(\boldsymbol{\theta_0})}{\partial\theta_k} - \frac{\partial\tilde{l}_t(\boldsymbol{\theta_0})}{\partial\theta_k}\right]\right| > \varepsilon\right] \\
&\leq \frac{1}{\varepsilon}\left[\mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} A_t(\boldsymbol{\theta_0})\right| + \mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} B_t(\boldsymbol{\theta_0})\right| + \mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} C_t(\boldsymbol{\theta_0})\right|\right].
\end{aligned} \tag{26}$$

First consider $\mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} A_t(\boldsymbol{\theta_0})\right|$. From (20), we have

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

whence, using the independence of η_t^2 with σ_t^2 and its derivatives at $\boldsymbol{\theta_0}$, (11), **A2** and **A8**, $\mathbb{E}_{\boldsymbol{\theta_0}}|A_t(\boldsymbol{\theta_0})|^\rho \leq K\mathbb{E}_{\boldsymbol{\theta_0}}\left(\left[\sum_{i=t}^{\infty} i^{-(d^*+1)}\varepsilon_{t-i}^2\right]^\rho\left|\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_k}(\boldsymbol{\theta_0})\right|^\rho\right)$. Since $\rho<1$ from Assumption **A9**, there exists some $\xi>0$ such that $\rho(1+\xi)\leq 1$. Hence, from Hölder inequality, along with Lemma 1, we obtain

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta_0}}|A_t(\boldsymbol{\theta_0})|^\rho &\leq K\left(\mathbb{E}_{\boldsymbol{\theta_0}}\left[\sum_{i=t}^{\infty} i^{-(d^*+1)}\varepsilon_{t-i}^2\right]^{\rho(1+\xi)}\right)^{\frac{1}{1+\xi}}\left(\mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_k}(\boldsymbol{\theta_0})\right|^{\rho\frac{\xi+1}{\xi}}\right)^{\frac{\xi}{1+\xi}}\\
&\leq K\sum_{i=0}^{\infty}(t+i)^{-(d^*+1)\rho}\left(\mathbb{E}_{\boldsymbol{\theta_0}}|\varepsilon_{-i}|^{2\rho(1+\xi)}\right)^{\frac{1}{1+\xi}}\\
&\leq Kt^{-(d^*+1)\rho+1},
\end{aligned}$$

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

Consider now $\mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} B_t(\boldsymbol{\theta_0})\right|$. We have, from (25)

$$\begin{aligned}
|B_t(\boldsymbol{\theta_0})| &= \left|\left[1-\frac{\varepsilon_t^2}{\tilde{\sigma}_t^2}\right]\left[\frac{1}{\tilde{\sigma}_t^2}-\frac{1}{\sigma_t^2}\right]\left[\frac{\partial\sigma_t^2}{\partial\theta_k}\right]\right|(\boldsymbol{\theta_0})\\
&\leq K\eta_t^2\left[\frac{\sigma_t^2(\boldsymbol{\theta_0})-\tilde{\sigma}_t^2(\boldsymbol{\theta_0})}{\tilde{\sigma}_t^2(\boldsymbol{\theta_0})}\right]\left|\left[\frac{1}{\sigma_t^2(\boldsymbol{\theta_0})}\frac{\partial\sigma_t^2(\boldsymbol{\theta_0})}{\partial\theta_k}\right]\right|,
\end{aligned}$$

and thus $\mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}B_t(\boldsymbol{\theta_0})\right|^\rho \underset{n\to\infty}{\to} 0$ from the same previous arguments. Using Markov inequality, we can conclude $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}|B_t(\boldsymbol{\theta_0})|$ tends to 0 in probability.

Finally consider $\mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} C_t(\boldsymbol{\theta_0})\right|$. From (23) and (24), and from Assumptions

A10(i) and **A3(ii)**, we have for all $\xi > 0$,

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

and thus $\mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}C_t(\boldsymbol{\theta_0})\right|^{\rho} \leq Kn^{-\frac{1}{2}\rho}\sum_{t=1}^{n}t^{-(d^*+1)\rho(1-\xi)+1} \leq Kn^{-(d^*+1)\rho(1-\xi)+2} \underset{n\to\infty}{\to}$ 0 since, from **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 (26) 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 $\boldsymbol{\theta_0}$.

We denote $\chi_t = \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})}|\sigma_t^2(\boldsymbol{\theta})-\tilde{\sigma}_t^2(\boldsymbol{\theta})|$, and we have from (20), (11), and **A3(ii)**, $\chi_t = \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})}\sum_{i=t}^{\infty}a_{i,t-i}(\boldsymbol{\phi}^+,\boldsymbol{\phi}^-)\varepsilon_{t-i}^2 \leq K\sum_{i=0}^{\infty}(i+t)^{-(d+1)}\varepsilon_{-i}^2$, whence $\mathbb{E}\chi_t^{\rho} \leq Kt^{-(d+1)\rho+1}$ since, from **A4**, $\mathbb{E}\left|\varepsilon_t\right|^{2\rho}<\infty$. This shows that χ_t has a finite moment of order ρ and thus is finite a.s. Furthermore, since $\rho(d+1)>1$, the dominated convergence theorem entails $\lim_{t\to\infty}\chi_t = 0$ a.s.

Let us now denote

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

where $V(\boldsymbol{\theta_0})$ is a neighborhood of $\boldsymbol{\theta_0}$ and $i_1, i_2 \in \{1,\ldots,2r+1\}$. From (23) and

(24) we easily obtain $\chi_t^{(1)} = 0$, and from **A10(i)** and **A3(ii)**, we have for all $\xi > 0$,

$$\begin{aligned}
\chi_t^{(i_1)} &\leq \begin{cases} \sum_{i=t}^{\infty} \max_{j\in\{1,\dots,r\}} \sup_{\boldsymbol{\phi}\in V(\boldsymbol{\phi}_0^+)} \left|\frac{\partial \alpha_i(\boldsymbol{\phi})}{\partial \phi_j^+}\right| \varepsilon_{t-i}^2 \text{ if } 1 < i_1 \leq r+1 \\ \sum_{i=t}^{\infty} \max_{j\in\{1,\dots,r\}} \sup_{\boldsymbol{\phi}\in V(\boldsymbol{\phi}_0^-)} \left|\frac{\partial \alpha_i(\boldsymbol{\phi})}{\partial \phi_j^-}\right| \varepsilon_{t-i}^2 \text{ if } r+1 < i_1 \leq 2r+1 \end{cases} \\
&\leq K \sum_{i=0}^{\infty} (i+t)^{-(d+1)(1-\xi)} \varepsilon_{-i}^2,
\end{aligned}$$

whence $\mathbb{E}\left(\chi_t^{(i_1)}\right)^{\rho} \leq K t^{-(d+1)\rho(1-\xi)+1}$ since, from **A4**, $\mathbb{E}\left|\varepsilon_t\right|^{2\rho} < \infty$. This shows that for any i_1, $\chi_t^{(i_1)}$ has a finite moment of order ρ and thus is finite a.s. 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 $\lim_{t\to\infty} \chi_t^{(i_1)} = 0$ a.s. The same arguments yield $\lim_{t\to\infty} \chi_t^{(i_1,i_2)} = 0$ a.s. 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{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^2(\boldsymbol{\theta})} \frac{\partial^2 \sigma_t^2(\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^2(\boldsymbol{\theta})} \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}} \right] \left[\frac{1}{\sigma_t^2(\boldsymbol{\theta})} \frac{\partial \sigma_t^2(\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^2(\boldsymbol{\theta})} \frac{\partial \tilde{\sigma}_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}} \right] \left[\frac{1}{\sigma_t^2(\boldsymbol{\theta})} \frac{\partial \sigma_t^2(\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^2(\boldsymbol{\theta})} \frac{\partial \tilde{\sigma}_t^2(\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^2(\boldsymbol{\theta})} \frac{\partial \sigma_t^2(\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 1, for all $p > 0$, $i_1, i_2 = 1, ..., 2r+1$,

$$\mathbb{E}_{\boldsymbol{\theta}_0} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{1}{\tilde{\sigma}_t^2(\boldsymbol{\theta})} \frac{\partial \tilde{\sigma}_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}} \right|^p < \infty \quad \text{and} \quad \mathbb{E}_{\boldsymbol{\theta}_0} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{1}{\tilde{\sigma}_t^2(\boldsymbol{\theta})} \frac{\partial^2 \tilde{\sigma}_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}\partial\theta_{i_2}} \right|^p < \infty. \tag{27}$$

Then, from independence of η_t^2 with σ_t^2 and its derivatives, Assumption **A11**,

Lemma 1, (25), and (27) we have, using Hölder inequality, for all i_1, i_2,

$$\begin{aligned}
&\mathbb{E}\left[\eta_t^2 \sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\left[\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial^2\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}\partial\theta_{i_2}}\right]\right|\right] < \infty \\
&\mathbb{E}\left[\eta_t^2 \sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left[\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right]\left[\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_2}}\right]\right|\right] < \infty \\
&\mathbb{E}\left[\eta_t^2 \sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left[\frac{1}{\tilde{\sigma}_t^2(\boldsymbol{\theta})}\frac{\partial\tilde{\sigma}_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right]\left[\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_2}}\right]\right|\right] < \infty \qquad (28)\\
&\mathbb{E}\left[\eta_t^2 \sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left[\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right]\right|\right] < \infty \\
&\mathbb{E}\left[\eta_t^2 \sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left[\frac{1}{\tilde{\sigma}_t^2(\boldsymbol{\theta})}\frac{\partial\tilde{\sigma}_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right]\right|\right] < \infty.
\end{aligned}$$

Since χ_t, $\chi_t^{(i_1)}$, and $\chi_t^{(i_1,i_2)}$ tends to 0 almost surely as t tends to infinity, and (28), Toeplit lemma entails that

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

almost surely, which concludes the proof.

Using Slutsky lemma along with the previous intermediate results allows us to conclude that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial\tilde{l}_t}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta_0}) \xrightarrow{\mathcal{L}} \mathcal{N}(0,(\kappa_\eta - 1)\boldsymbol{J})$$

and that

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\tilde{l}_t}{\partial\theta_i\partial\theta_j}(\boldsymbol{\theta^*_{ij}}) \to J(i,j) \text{ in probability}$$

which ends the proof.

□

A.3 Specification tests

Proof of Proposition 1. This is a standard result for testing linear constraints. See for example Chapter 17 of Gouriéroux and Monfort[30] for proofs of the asymptotic distributions. □

Supplementary content - Conditional asymmetry in ARCH(∞) models

Julien Royer*[1]

[1]CREST, ENSAE, Institut Polytechnique de Paris

Appendix B Detailed proofs and technical results

B.1 Existence of a stationary TARCH(∞) solution

We develop in this section the proof of Theorem 1 and Corollary 1.1.

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

$$\begin{aligned} S_t &= \omega + \omega \sum_{k=1}^{\infty} \sum_{i_1,\ldots,i_k \geq 1} a_{i_1,t-i_1} \ldots a_{i_k,t-i_1-\ldots-i_k} \ \eta^2_{t-i_1} \ldots \eta^2_{t-i_1-\ldots-i_k} \\ &= \omega + \omega \sum_{k=1}^{\infty} \sum_{i_1,\ldots,i_k \geq 1} \left[\alpha^+_{i_1} \ \mathbb{1}_{\eta_{t-i_1} \geq 0} + \alpha^-_{i_1} \ \mathbb{1}_{\eta_{t-i_1} < 0}\right] \eta^2_{t-i_1} \ldots \\ &\qquad \left[\alpha^+_{i_k} \ \mathbb{1}_{\eta_{t-i_1-\ldots-i_k} \geq 0} + \alpha^-_{i_k} \ \mathbb{1}_{\eta_{t-i_1-\ldots-i_k} < 0}\right] \eta^2_{t-i_1-\ldots-i_k} \end{aligned}$$

defined in $[0, +\infty]$. Since $s \in (0, 1]$, we have

$$\begin{aligned} S^s_t &\leq \omega^s + \omega^s \sum_{k=1}^{\infty} \sum_{i_1,\ldots,i_k \geq 1} \left[\alpha^+_{i_1} \ \mathbb{1}_{\eta_{t-i_1} \geq 0} + \alpha^-_{i_1} \ \mathbb{1}_{\eta_{t-i_1} < 0}\right]^s \eta^{2s}_{t-i_1} \ldots \\ &\qquad \left[\alpha^+_{i_k} \ \mathbb{1}_{\eta_{t-i_1-\ldots-i_k} \geq 0} + \alpha^-_{i_k} \ \mathbb{1}_{\eta_{t-i_1-\ldots-i_k} < 0}\right]^s \eta^{2s}_{t-i_1-\ldots-i_k} \\ &\leq \omega^s + \omega^s \sum_{k=1}^{\infty} \sum_{i_1,\ldots,i_k \geq 1} \left[(\alpha^+_{i_1})^s \ \mathbb{1}_{\eta_{t-i_1} \geq 0} + (\alpha^-_{i_1})^s \ \mathbb{1}_{\eta_{t-i_1} < 0}\right] \eta^{2s}_{t-i_1} \ldots \\ &\qquad \left[(\alpha^+_{i_k})^s \ \mathbb{1}_{\eta_{t-i_1-\ldots-i_k} \geq 0} + (\alpha^-_{i_k})^s \ \mathbb{1}_{\eta_{t-i_1-\ldots-i_k} < 0}\right] \eta^{2s}_{t-i_1-\ldots-i_k} \end{aligned}$$

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

and from the independence of (η_t), it follows that

$$\begin{aligned}\mathbb{E}S_t^s \leq \quad & \omega^s + \omega^s \sum_{k=1}^{\infty} \sum_{i_1,\ldots,i_k \geq 1} \mathbb{E}\left(\left[(\alpha_{i_1}^+)^s \ \mathbb{1}_{\eta_{t-i_1}\geq 0} + (\alpha_{i_1}^-)^s \ \mathbb{1}_{\eta_{t-i_1}<0}\right] |\eta_{t-i_1}|^{2s}\right) \ldots \\ & \mathbb{E}\left(\left[(\alpha_{i_k}^+)^s \ \mathbb{1}_{\eta_{t-i_1-\ldots-i_k}\geq 0} + (\alpha_{i_k}^-)^s \ \mathbb{1}_{\eta_{t-i_1-\ldots-i_k}<0}\right] |\eta_{t-i_1-\ldots-i_k}|^{2s}\right),\end{aligned}$$

and thus

$$\begin{aligned}\mathbb{E}S_t^s \quad & \leq \omega^s \left[1 + \sum_{k=1}^{\infty} (A_s^+ \mu_{2s}^+ + A_s^- \mu_{2s}^-)^k\right] \\ & \leq \frac{\omega^s}{1 - (A_s^+ \mu_{2s}^+ + A_s^- \mu_{2s}^-)} < \infty,\end{aligned}$$

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

$$\begin{aligned}&\sum_{i=1}^{\infty} a_{i,t-i} S_{t-i} \eta_{t-i}^2 \\ &= \omega \sum_{i_0=1}^{\infty} a_{i_0,t-i_0} \eta_{t-i_0}^2 + \\ &\omega \sum_{i_0=1}^{\infty} a_{i_0,t-i_0} \eta_{t-i_0}^2 \sum_{k=1}^{\infty} \sum_{i_1,\ldots,i_k \geq 1} a_{i_1,t-i_0-i_1} \ldots a_{i_k,t-i_0-\ldots-i_k} \ \eta_{t-i_0-i_1}^2 \ldots \eta_{t-i_0-i_1-\ldots-i_k}^2 \\ &= \omega \sum_{k=0}^{\infty} \sum_{i_0,\ldots,i_k \geq 1} a_{i_0,t-i_0} \ldots a_{i_k,t-i_0-\ldots-i_k} \ \eta_{t-i_0}^2 \ldots \eta_{t-i_0-\ldots-i_k}^2\end{aligned}$$

and thus we obtain the recursive equation

$$S_t = \omega + \sum_{i=1}^{\infty} a_{i,t-i} S_{t-i} \eta_{t-i}^2.$$

By setting $\varepsilon_t = \sqrt{S_t}\eta_t$, we obtain a strictly stationary and nonanticipative solution of (3) and $\mathbb{E}|\varepsilon_t|^{2s} \leq \mu_{2s}\omega^s/(1 - (A_s^+ \mu_{2s}^+ + A_s^- \mu_{2s}^-))$.

Now denote by (ε_t^*) any strictly stationary and nonanticipative solution of (3) such that $\mathbb{E}|\varepsilon_t^*|^{2s} < \infty$. For all $q \geq 1$, by q recursive substitutions of the ε_{t-i}^{*2}, we obtain

$$\begin{aligned}\sigma_t^2 \quad & = \omega + \sum_{i=1}^{\infty} a_{i,t-i} \varepsilon_{t-i}^{*2} \\ & = \left\{\omega + \omega \sum_{k=1}^{q} \sum_{i_1,\ldots,i_k \geq 1} a_{i_1,t-i_1} \ldots a_{i_k,t-i_1-\ldots-i_k} \ \eta_{t-i_1}^2 \ldots \eta_{t-i_1-\ldots-i_k}^2\right\} \\ & + \sum_{i_1,\ldots,i_{q+1} \geq 1} a_{i_1,t-i_1} \ldots a_{i_{q+1},t-i_1-\ldots-i_{q+1}} \ \eta_{t-i_1}^2 \ldots \eta_{t-i_1-\ldots-i_q}^2 \varepsilon_{t-i_1-\ldots-i_{q+1}}^{*2} \\ & := \{S_{t,q}\} + R_{t,q}.\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E}R_{t,q}^s &\leq \sum_{i_1,\dots,i_{q+1}\geq 1} \mathbb{E}\left(\left[(\alpha_{i_1}^+)^s \ \mathbb{1}_{\eta_{t-i_1}\geq 0} + (\alpha_{i_1}^-)^s \ \mathbb{1}_{\eta_{t-i_1}<0}\right] |\eta_{t-i_1}|^{2s}\right) \dots \\
&\quad \mathbb{E}\left(\left[(\alpha_{i_{q+1}}^+)^s \ \mathbb{1}_{\eta_{t-i_1-\dots-i_{q+1}}\geq 0} + (\alpha_{i_{q+1}}^-)^s \ \mathbb{1}_{\eta_{t-i_1-\dots-i_{q+1}}<0}\right] |\varepsilon^*_{t-i_1-\dots-i_{q+1}}|^{2s}\right) \\
&\leq (A_s^+\mu_{2s}^+ + A_s^-\mu_{2s}^-)^q \ \left(A_s^+\mathbb{E}|\mathbb{1}_{\eta_t\geq 0}\varepsilon_t^*|^{2s} + A_s^-\mathbb{E}|\mathbb{1}_{\eta_t<0}\varepsilon_t^*|^{2s}\right)
\end{aligned}$$

Since $A_s^+\mu_{2s}^+ + A_s^-\mu_{2s}^- < 1$, 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 $\sigma_t^2 = S_t$ almost surely and yields $\varepsilon_t^* = \varepsilon_t$ almost surely hence concluding the proof. □

Proof of Corollary 1.1. From Theorem 1, we have that

$$\varepsilon_t^2 = \omega \ \eta_t^2 \left[1 + \sum_{k=1}^{\infty} \sum_{i_1,\dots,i_k\geq 1} a_{i_1,t-i_1}\dots a_{i_k,t-i_1-\dots-i_k} \ \eta_{t-i_1}^2\dots\eta_{t-i_1-\dots-i_k}^2\right]$$

and using the c_r-inequality, we obtain

$$\left(\mathbb{E}\varepsilon_t^4\right)^{\frac{1}{2}} \leq \omega \ \left(\mathbb{E}\eta_t^4\right)^{\frac{1}{2}} \left[1 + \sum_{k=1}^{\infty} \sum_{i_1,\dots,i_k\geq 1} \left(\mathbb{E}\left[a_{i_1,t-i_1}^2\dots a_{i_k,t-i_1-\dots-i_k}^2 \ \eta_{t-i_1}^4\dots\eta_{t-i_1-\dots-i_k}^4\right]\right)^{\frac{1}{2}}\right]$$

where, for all $i \geq 1$,

$$a_{i,t-i}^2 = \left[\alpha_i^+\mathbb{1}_{\eta_{t-i}\geq 0} + \alpha_i^- \ \mathbb{1}_{\eta_{t-i}<0}\right]^2 = (\alpha_i^+)^2\mathbb{1}_{\eta_{t-i}\geq 0} + (\alpha_i^-)^2\mathbb{1}_{\eta_{t-i}<0},$$

hence

$$\left(\mathbb{E}\varepsilon_t^4\right)^{\frac{1}{2}} \leq \frac{\omega \ \left(\mathbb{E}\eta_t^4\right)^{\frac{1}{2}}}{1 - \left(A_2^+\mu_4^+ + A_2^-\mu_4^-\right)^{\frac{1}{2}}}$$

and thus ε_t admits a fourth moment if $\left(A_2^+\mu_4^+ + A_2^-\mu_4^-\right)^{\frac{1}{2}} < 1$. □

B.2 Statistical inference of a Threshold ARCH(∞) 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.

Let us define the theoretical criterion

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

The theoretical QML estimator $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}$ 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 $\mathcal{I}(\boldsymbol{\phi}) = \{i \text{ such that } \alpha_i(\boldsymbol{\phi}) \neq 0\}$, and we define $\mathcal{I}_t^+$ (respectively $\mathcal{I}_t^-$) as $\mathcal{I}_t^{+(-)} = \{i \text{ such that } \varepsilon_{t-i} \geq 0 \text{ (resp. } < 0)\}$, yielding the following rewriting of (9) as

$$\sigma_t^2(\boldsymbol{\theta_0}) = \omega_0 + \sum_{i \in \mathcal{I}_t^+} \alpha_i(\boldsymbol{\phi_0^+})\varepsilon_{t-i}^2 + \sum_{j \in \mathcal{I}_t^-} \alpha_j(\boldsymbol{\phi_0^-})\varepsilon_{t-j}^2. \tag{B.1}$$

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

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

$$\sup_{i \in \mathcal{I}(\boldsymbol{\phi_0^+})} \sup_{\phi \in \boldsymbol{\Phi}} \frac{\alpha_i(\boldsymbol{\phi_0^+})}{\alpha_i^{1-\tau}(\boldsymbol{\phi})} \leq K \text{ and } \sup_{i \in \mathcal{I}(\boldsymbol{\phi_0^-})} \sup_{\phi \in \boldsymbol{\Phi}} \frac{\alpha_i(\boldsymbol{\phi_0^-})}{\alpha_i^{1-\tau}(\boldsymbol{\phi})} \leq K, \tag{B.2}$$

then

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

Proof of Proposition 2. Using the fact that for any s>0, we have $x/(1+x) \leq x^s$, Equation (B.1) gives

$$\begin{aligned}
&\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})} \\
=&\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \frac{\omega_0 + \sum_{i \in \mathcal{I}_t^+} \alpha_i(\boldsymbol{\phi_0^+})\varepsilon_{t-i}^2 + \sum_{j \in \mathcal{I}_t^-} \alpha_j(\boldsymbol{\phi_0^-})\varepsilon_{t-j}^2}{\omega + \sum_{i' \in \mathcal{I}_t^+} \alpha_{i'}(\boldsymbol{\phi^+})\varepsilon_{t-i'}^2 + \sum_{j' \in \mathcal{I}_t^-} \alpha_{j'}(\boldsymbol{\phi^-})\varepsilon_{t-j'}^2} \\
\leq& K + \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \sum_{i \in \mathcal{I}_t^+ \cap \mathcal{I}(\boldsymbol{\phi_0^+})} \frac{\alpha_i(\boldsymbol{\phi_0^+})\varepsilon_{t-i}^2}{\omega + \alpha_i(\boldsymbol{\phi^+})\varepsilon_{t-i}^2} + \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \sum_{i \in \mathcal{I}_t^- \cap \mathcal{I}(\boldsymbol{\phi_0^-})} \frac{\alpha_i(\boldsymbol{\phi_0^-})\varepsilon_{t-i}^2}{\omega + \alpha_i(\boldsymbol{\phi^-})\varepsilon_{t-i}^2} \\
\leq& K + \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \omega^{-s} \sum_{i \in \mathcal{I}_t^+ \cap \mathcal{I}(\boldsymbol{\phi_0^+})} \frac{\alpha_i(\boldsymbol{\phi_0^+})}{\alpha_i(\boldsymbol{\phi^+})} \alpha_i^s(\boldsymbol{\phi^+})\varepsilon_{t-i}^{2s} \\
&+ \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \omega^{-s} \sum_{i \in \mathcal{I}_t^- \cap \mathcal{I}(\boldsymbol{\phi_0^-})} \frac{\alpha_i(\boldsymbol{\phi_0^-})}{\alpha_i(\boldsymbol{\phi^-})} \alpha_i^s(\boldsymbol{\phi^-})\varepsilon_{t-i}^{2s} \\
\leq& K + \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \omega^{-s} \sum_{i \in \mathcal{I}_t^+ \cap \mathcal{I}(\boldsymbol{\phi_0^+})} \frac{\alpha_i(\boldsymbol{\phi_0^+})}{\alpha_i^{1-\tau}(\boldsymbol{\phi^+})} \alpha_i^{s-\tau}(\boldsymbol{\phi^+})\varepsilon_{t-i}^{2s}
\end{aligned}$$

$$
\begin{aligned}
&+\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \omega^{-s} \sum_{i\in\mathcal{I}_t^-\cap\mathcal{I}(\boldsymbol{\phi_0^-})} \frac{\alpha_i(\boldsymbol{\phi_0^-})}{\alpha_i^{1-\tau}(\boldsymbol{\phi}^-)}\alpha_i^{s-\tau}(\boldsymbol{\phi}^-)\varepsilon_{t-i}^{2s} \\
\leq & K + K\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{i\in\mathcal{I}_t^+\cap\mathcal{I}(\boldsymbol{\phi_0^+})} \alpha_i^{s-\tau}(\boldsymbol{\phi}^+)\varepsilon_{t-i}^{2s} + K\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{i\in\mathcal{I}_t^-\cap\mathcal{I}(\boldsymbol{\phi_0^-})} \alpha_i^{s-\tau}(\boldsymbol{\phi}^-)\varepsilon_{t-i}^{2s} \\
\leq & K + K\sum_{i=1}^{\infty} i^{-(d+1)(s-\tau)}\varepsilon_{t-i}^{2s}
\end{aligned}
$$

using Assumptions **A3(ii)** and (B.2). This yields

$$
\mathbb{E}_{\boldsymbol{\theta_0}}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})} \leq K + \omega^{-s}K'\sum_{i=1}^{\infty} i^{-(d+1)(s-\tau)}\mathbb{E}_{\boldsymbol{\theta_0}}|\varepsilon_{t-i}|^{2s}.
$$

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

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

□

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

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

(b) Identifiability of the parameter
($\exists t\in\mathbb{Z}$ such that $\sigma_t^2(\boldsymbol{\theta}) = \sigma_t^2(\boldsymbol{\theta_0})$ a.s.) $\Rightarrow \boldsymbol{\theta} = \boldsymbol{\theta_0}$

(c) The limit criterion is minimized at the true value
$\mathbb{E}_{\boldsymbol{\theta_0}}|l_t(\boldsymbol{\theta_0})| < \infty$, 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) Compactness of $\boldsymbol{\Theta}$ and ergodicity of $(l_t(\boldsymbol{\theta}))$
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.

This structure of the proof follows the ones for the earlier results in the classical GARCH case and is similar to the one presented in Francq and Zakoian[2]. However, the demonstrations of the intermediate results differ and, noticeably, the

impact of the initial values requires a particular attention.

In the following, we detail the demonstration of the four previous points:
(a) Asymptotic irrelevance of the initial values
Consider

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

and remark that from (9), (12), and **A1**, we have

$$\begin{aligned} \sigma_t^2(\boldsymbol{\theta}) &= \omega + \sum_{i=1}^{\infty} a_{i,t-i}(\boldsymbol{\phi}^+, \boldsymbol{\phi}^-)\varepsilon_{t-i}^2 \\ &= \omega + \sum_{i=1}^{t-1} a_{i,t-i}(\boldsymbol{\phi}^+, \boldsymbol{\phi}^-)\varepsilon_{t-i}^2 + \sum_{i=t}^{\infty} a_{i,t-i}(\boldsymbol{\phi}^+, \boldsymbol{\phi}^-)\varepsilon_{t-i}^2 \\ &= \tilde{\sigma}_t^2(\boldsymbol{\theta}) + \sum_{i=t}^{\infty} a_{i,t-i}(\boldsymbol{\phi}^+, \boldsymbol{\phi}^-)\varepsilon_{t-i}^2 \\ &\geq \tilde{\sigma}_t^2(\boldsymbol{\theta}). \end{aligned} \tag{B.3}$$

We denote $\chi_t = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\sigma_t^2(\boldsymbol{\theta}) - \tilde{\sigma}_t^2(\boldsymbol{\theta})|$, and we have from (B.3), (11), and **A3(ii)**

$$\begin{aligned} \chi_t &= \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{i=t}^{\infty} a_{i,t-i}(\boldsymbol{\phi}^+, \boldsymbol{\phi}^-)\varepsilon_{t-i}^2 \\ &\leq \sum_{i=t}^{\infty} \sup_{\boldsymbol{\phi}\in\boldsymbol{\Phi}} \alpha_i(\boldsymbol{\phi})\varepsilon_{t-i}^2 \\ &\leq K\sum_{i=t}^{\infty} i^{-(d+1)}\varepsilon_{t-i}^2 \\ &\leq K\sum_{i=0}^{\infty} (i+t)^{-(d+1)}\varepsilon_{-i}^2, \end{aligned}$$

whence

$$\mathbb{E}\chi_t^{\rho} \leq K\sum_{i=0}^{\infty} (i+t)^{-(d+1)\rho}\, \mathbb{E}|\varepsilon_{-i}|^{2\rho}.$$

Since from **A4**, $\mathbb{E}\,|\varepsilon_t|^{2\rho} < \infty$, with $\rho(d+1) > 1$, and since for any $k > 1$ we have

$$\int_t^{\infty} x^{-k}dx = \left[-\frac{x^{-k+1}}{k-1} \right]_t^{\infty} = \frac{t^{-k+1}}{k-1},$$

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,

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

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

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |A_n(\boldsymbol{\theta})| \leq \frac{K}{n} \sum_{t=1}^{n} \chi_t \tag{B.4}$$

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

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

whence

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

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 **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.

We can conclude

$$\lim_{n\to\infty} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |Q_n(\boldsymbol{\theta}) - \tilde{Q}_n(\boldsymbol{\theta})| \leq \lim_{n\to\infty} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |A_n(\boldsymbol{\theta})| + \lim_{n\to\infty} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |B_n(\boldsymbol{\theta})| = 0.$$

(b) Identifiability of the parameter:
Let $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, such that, for some $t \in \mathbb{Z}$, we have $\sigma_t^2(\boldsymbol{\theta}) = \sigma_t^2(\boldsymbol{\theta_0})$ almost surely. Assume $\boldsymbol{\theta} \neq \boldsymbol{\theta_0}$, and suppose that

$$\alpha_1(\boldsymbol{\phi_0^+})\mathbb{1}_{\varepsilon_{t-1}\geq 0} + \alpha_1(\boldsymbol{\phi_0^-})\mathbb{1}_{\varepsilon_{t-1}<0} \neq \alpha_1(\boldsymbol{\phi^+})\mathbb{1}_{\varepsilon_{t-1}\geq 0} + \alpha_1(\boldsymbol{\phi^-})\mathbb{1}_{\varepsilon_{t-1}<0}. \tag{B.5}$$

Then we have

$$\begin{aligned}
&\omega_0 + \sum_{i=1}^{\infty} \alpha_i(\boldsymbol{\phi_0^+})\varepsilon_{t-i}^2\mathbb{1}_{\varepsilon_{t-i}\geq 0} + \alpha_i(\boldsymbol{\phi_0^-})\varepsilon_{t-i}^2\mathbb{1}_{\varepsilon_{t-i}<0} \\
&\qquad = \omega + \sum_{i=1}^{\infty} \alpha_i(\boldsymbol{\phi^+})\varepsilon_{t-i}^2\mathbb{1}_{\varepsilon_{t-i}\geq 0} + \alpha_i(\boldsymbol{\phi^-})\varepsilon_{t-i}^2\mathbb{1}_{\varepsilon_{t-i}<0} \\
&\Leftrightarrow \left(\left[\alpha_1(\boldsymbol{\phi_0^+}) - \alpha_1(\boldsymbol{\phi^+})\right]\mathbb{1}_{\eta_{t-1}\geq 0} + \left[\alpha_1(\boldsymbol{\phi_0^-}) - \alpha_1(\boldsymbol{\phi^-})\right]\mathbb{1}_{\eta_{t-1}>0}\right)\sigma_{t-1}^2(\boldsymbol{\theta_0})\eta_{t-1}^2 \\
&\qquad = \omega - \omega_0 + \sum_{i=2}^{\infty} \left(\left[\alpha_i(\boldsymbol{\phi^+}) - \alpha_i(\boldsymbol{\phi_0^+})\right]\mathbb{1}_{\eta_{t-i}\geq 0}\right. \\
&\qquad\quad \left. + \left[\alpha_i(\boldsymbol{\phi^-}) - \alpha_i(\boldsymbol{\phi_0^-})\right]\mathbb{1}_{\eta_{t-i}<0}\right)\sigma_{t-i}^2(\boldsymbol{\theta_0})\eta_{t-i}^2.
\end{aligned}$$

Whence $\left(\left[\alpha_1(\boldsymbol{\phi_0^+}) - \alpha_1(\boldsymbol{\phi^+})\right]\mathbb{1}_{\eta_{t-1}\geq 0} + \left[\alpha_1(\boldsymbol{\phi_0^-}) - \alpha_1(\boldsymbol{\phi^-})\right]\mathbb{1}_{\eta_{t-1}<0}\right)\eta_{t-1}^2$ belongs to $\mathcal{F}\left(\eta_{t-2}^2, \eta_{t-3}^2, ...\right)$ and thus, by independence, is almost surely constant, which yields

$$\begin{cases} \left[\alpha_1(\boldsymbol{\phi_0^+}) - \alpha_1(\boldsymbol{\phi^+})\right]\mathbb{1}_{\eta_{t-1}\geq 0}\ \eta_{t-1}^2 \text{ is constant almost surely} \\ \left[\alpha_1(\boldsymbol{\phi_0^-}) - \alpha_1(\boldsymbol{\phi^-})\right]\mathbb{1}_{\eta_{t-1}<0}\ \eta_{t-1}^2 \text{ is constant almost surely} \end{cases}. \tag{B.6}$$

Since, from **A2**, η_1 takes at least two positive (respectively negative) values, (B.6) implies almost surely

$$\begin{cases} \alpha_1(\boldsymbol{\phi_0^+}) = \alpha_1(\boldsymbol{\phi^+}) \\ \alpha_1(\boldsymbol{\phi_0^-}) = \alpha_1(\boldsymbol{\phi^-}) \end{cases}$$

which contradicts (B.5). Recursively, we obtain that $\sigma_t^2(\boldsymbol{\theta}) = \sigma_t^2(\boldsymbol{\theta_0})$ implies that, for all i,

$$\begin{cases} \alpha_i(\boldsymbol{\phi_0^+}) = \alpha_i(\boldsymbol{\phi^+}) \\ \alpha_i(\boldsymbol{\phi_0^-}) = \alpha_i(\boldsymbol{\phi^-}) \end{cases}$$

and thus, from **A3(i)**, $\boldsymbol{\phi^+} = \boldsymbol{\phi_0^+}$ and $\boldsymbol{\phi^-} = \boldsymbol{\phi_0^-}$ almost surely, whence $\omega = \omega_0$ almost surely, and thus $\boldsymbol{\theta} = \boldsymbol{\theta_0}$ almost surely.
(c) The limit criterion is minimized at the true value:
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

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta_0}}\left[l_t^-(\boldsymbol{\theta})\right] &= \mathbb{E}_{\boldsymbol{\theta_0}} \max[0, -l_t(\boldsymbol{\theta})] \\
&= \mathbb{E}_{\boldsymbol{\theta_0}} \max\left[0, -\log\sigma_t^2(\boldsymbol{\theta}) - \frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})}\right] \\
&\leq \mathbb{E}_{\boldsymbol{\theta_0}} \max\left[0, -\log\sigma_t^2(\boldsymbol{\theta})\right] \\
&\leq \mathbb{E}_{\boldsymbol{\theta_0}} \max\left[0, -\log\omega\right] \\
&< \infty.
\end{aligned}.$$

Furthermore, we can show that it is integrable at $\boldsymbol{\theta_0}$. Using Jensen inequality, (11), and **A3(ii)**, we obtain

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

since, from **A4**, $\mathbb{E}\left|\varepsilon_t\right|^{2\rho} < \infty$ 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 $\boldsymbol{\theta} = \boldsymbol{\theta_0}$.

(d) Compactness of $\boldsymbol{\Theta}$ and ergodicity of $(l_t(\boldsymbol{\theta}))$

For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, and any positive integer k, let $V_k(\boldsymbol{\theta})$ be the open ball of center $\boldsymbol{\theta}$ and radius $1/k$. Because of the asymptotic irrelevance of the initial values, we have

$$\begin{aligned}
\liminf_{n\to\infty}\inf_{\boldsymbol{\theta}^*\in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}}\tilde{Q}_n(\boldsymbol{\theta}^*) &\geq \liminf_{n\to\infty}\inf_{\boldsymbol{\theta}^*\in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} Q_n(\boldsymbol{\theta}^*) \\
&\qquad -\limsup_{n\to\infty}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|Q_n(\boldsymbol{\theta}) - \tilde{Q}_n(\boldsymbol{\theta})| \\
&\geq \liminf_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\inf_{\boldsymbol{\theta}^*\in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} l_t(\boldsymbol{\theta}^*).
\end{aligned}$$

From the uniform ergodic theorem, the sequences $(l_t(\boldsymbol{\theta}^*))$ and $\left(\inf_{\boldsymbol{\theta}^*\in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} l_t(\boldsymbol{\theta}^*)\right)$ are ergodic and strictly stationary, and admit an expectation in $\mathbb{R}\cup\{\infty\}$. Using

the ergodic theorem for non-integrable processes [1], we have

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \inf_{\boldsymbol{\theta}^*\in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} l_t(\boldsymbol{\theta}^*) = \mathbb{E}_{\boldsymbol{\theta_0}} \inf_{\boldsymbol{\theta}^*\in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} l_1(\boldsymbol{\theta}^*).$$

By Beppo-Levi theorem, $\mathbb{E}_{\boldsymbol{\theta_0}} \inf_{\boldsymbol{\theta}^*\in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} l_1(\boldsymbol{\theta}^*)$ increases to $\mathbb{E}_{\boldsymbol{\theta_0}} l_1(\boldsymbol{\theta})$ as $k\to\infty$. The limit criterion being minimized at the true value $\boldsymbol{\theta_0}$, we obtain

$$\liminf_{n\to\infty} \inf_{\boldsymbol{\theta}^*\in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} \tilde{Q}_n(\boldsymbol{\theta}^*) > \mathbb{E}_{\boldsymbol{\theta_0}} l_1(\boldsymbol{\theta_0}).$$

The conclusion of the proof uses a compactness argument. First note that for any neighborhood $V(\boldsymbol{\theta_0})$ of $\boldsymbol{\theta_0}$,

$$\begin{aligned}\limsup_{n\to\infty} \inf_{\boldsymbol{\theta}^*\in V(\boldsymbol{\theta_0})} \tilde{Q}_n(\boldsymbol{\theta}^*) &\le \lim_{n\to\infty} \tilde{Q}_n(\boldsymbol{\theta_0}) \\ &\le \lim_{n\to\infty} Q_n(\boldsymbol{\theta_0}) \\ &\le \mathbb{E}_{\boldsymbol{\theta_0}} l_1(\boldsymbol{\theta_0}).\end{aligned}$$

The compact set $\boldsymbol{\Theta}$ is covered by the union of an arbitrary neighborhood $V(\boldsymbol{\theta_0})$ of $\boldsymbol{\theta_0}$ and the set of the neighborhoods $V(\boldsymbol{\theta})$ satisfying $\liminf_{n\to\infty} \inf_{\boldsymbol{\theta}^*\in V(\boldsymbol{\theta})} \tilde{Q}_n(\boldsymbol{\theta}^*) \ge \mathbb{E}_{\boldsymbol{\theta_0}} l_1(\boldsymbol{\theta_0})$ for $\boldsymbol{\theta}\in\boldsymbol{\Theta}/V(\boldsymbol{\theta_0})$. Thus, there exists a finite subcover of $\boldsymbol{\Theta}$ of the form $V(\boldsymbol{\theta_0})$, $V(\boldsymbol{\theta_1})$, ..., $V(\boldsymbol{\theta_k})$, whence

$$\inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \tilde{Q}_n(\boldsymbol{\theta}) = \min_{i=0,1,\dots,k} \inf_{\boldsymbol{\theta}^*\in V(\boldsymbol{\theta_i})\cap\boldsymbol{\Theta}} \tilde{Q}_n(\boldsymbol{\theta}).$$

We obtain that, for n large enough, $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}$ belongs to $V(\boldsymbol{\theta_0})$ almost surely. Since this is true for any neighborhood $V(\boldsymbol{\theta_0})$, we have shown that, almost surely,

$$\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} \underset{n\to\infty}{\to} \boldsymbol{\theta_0}.$$

□

We will know 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*

$$\sup_{i\in\mathcal{I}(\boldsymbol{\phi_0^+})} \sup_{\phi\in V(\boldsymbol{\phi_0^+})} \frac{\alpha_i(\boldsymbol{\phi_0^+})}{\alpha_i^{1-\tau}(\boldsymbol{\phi})} \le K \text{ and } \sup_{i\in\mathcal{I}(\boldsymbol{\phi_0^-})} \sup_{\phi\in V(\boldsymbol{\phi_0^-})} \frac{\alpha_i(\boldsymbol{\phi_0^-})}{\alpha_i^{1-\tau}(\boldsymbol{\phi})} \le K. \tag{B.7}$$

[1] If (X_t) is an ergodic and strictly stationary process and if $\mathbb{E}X_1$ exists in $\mathbb{R}\cup\{+\infty\}$ then

$$\frac{1}{n}\sum_{t=1}^{n} X_t \underset{n\to\infty}{\to} \mathbb{E}X_1 \quad a.s.$$

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 for all $k > s$, (10) and Hölder inequality yield

$$\begin{aligned}
&\sigma_t^2(\boldsymbol{\theta_0}) \\
&= \omega_0 + \sum_{i=1}^{\infty} a_{i,t-i}(\boldsymbol{\phi_0^+}, \boldsymbol{\phi_0^-})\varepsilon_{t-i}^2 \\
&= \omega_0 \omega^{\frac{s}{k}-1}\omega^{1-\frac{s}{k}} + \sum_{i=1}^{\infty} a_{i,t-i}(\boldsymbol{\phi_0^+}, \boldsymbol{\phi_0^-}) a_{i,t-i}^{\frac{s}{k}-1}(\boldsymbol{\phi^+}, \boldsymbol{\phi^-}) a_{i,t-i}^{1-\frac{s}{k}}(\boldsymbol{\phi^+}, \boldsymbol{\phi^-}) \varepsilon_{t-i}^{2\frac{s}{k}} \varepsilon_{t-i}^{2-2\frac{s}{k}} \\
&\leq K \left[\omega_0^{\frac{k}{s}}\omega^{1-\frac{k}{s}} + \sum_{i=1}^{\infty} a_{i,t-i}^{\frac{k}{s}}(\boldsymbol{\phi_0^+}, \boldsymbol{\phi_0^-}) a_{i,t-i}^{1-\frac{k}{s}}(\boldsymbol{\phi^+}, \boldsymbol{\phi^-})\varepsilon_{t-i}^2\right]^{\frac{s}{k}} \left[\sigma_t^2(\boldsymbol{\theta})\right]^{1-\frac{s}{k}}.
\end{aligned}$$

Since $\left[\sigma_t^2(\boldsymbol{\theta})\right]^{-\frac{s}{k}} \leq K$, we obtain

$$\begin{aligned}
&\left[\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\right]^k \\
&\leq K \left[1 + \sum_{i=1}^{\infty} a_{i,t-i}^k(\boldsymbol{\phi_0^+}, \boldsymbol{\phi_0^-}) a_{i,t-i}^{s-k}(\boldsymbol{\phi^+}, \boldsymbol{\phi^-})\varepsilon_{t-i}^{2s}\right] \\
&\leq K \left[1 + \sum_{i\in\mathcal{I}_t^+\cap\mathcal{I}(\boldsymbol{\phi_0^+})} \frac{\alpha_i^k(\boldsymbol{\phi_0^+})}{\alpha_i^k(\boldsymbol{\phi^+})}\alpha_i^s(\boldsymbol{\phi^+})\varepsilon_{t-i}^{2s} + \sum_{i\in\mathcal{I}_t^-\cap\mathcal{I}(\boldsymbol{\phi_0^-})} \frac{\alpha_i^k(\boldsymbol{\phi_0^-})}{\alpha_i^k(\boldsymbol{\phi^-})}\alpha_i^s(\boldsymbol{\phi^-})\varepsilon_{t-i}^{2s}\right],
\end{aligned}$$

whence, from (B.7) and **A3(ii)**, there exists a neighborhood such that

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})} \left[\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\right]^k \\
&\leq K \left[1 + \sum_{i\in\mathcal{I}_t^+\cap\mathcal{I}(\boldsymbol{\phi_0^+})} \sup_{\boldsymbol{\phi^+}\in V(\boldsymbol{\phi_0^+})} \alpha_i^{s-k\tau}(\boldsymbol{\phi^+})\varepsilon_{t-i}^{2s} + \sum_{i\in\mathcal{I}_t^-\cap\mathcal{I}(\boldsymbol{\phi_0^-})} \sup_{\boldsymbol{\phi^-}\in V(\boldsymbol{\phi_0^-})} \alpha_i^{s-k\tau}(\boldsymbol{\phi^-})\varepsilon_{t-i}^{2s}\right] \\
&\leq K \left[1 + \sum_{i=1}^{\infty} i^{-(d+1)(s-k\tau)}\varepsilon_{t-i}^{2s}\right]
\end{aligned}$$

and thus, by taking $s = \rho$, there exists a neighborhood such that

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

from **A4**, and since from the arbitrariness of τ in (B.7), we can find a τ such that $(d+1)(\rho - k\tau) > 1$. □

Before developing the proof of Theorem 3, it is useful to state the following lemma.

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

$$\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \frac{1}{\sigma_t^2(\boldsymbol{\theta})} \frac{\partial^k \sigma_t^2(\boldsymbol{\theta})}{\partial \theta_{i_1} ... \partial \theta_{i_k}} \right|^p < \infty.$$

Proof of Lemma 1. From (B.1) and Assumption **A10(i)**, we have, for all $j_1, j_2 \in \{1, ..., r\}$,

$$\begin{aligned} \frac{\partial \sigma_t^2}{\partial \theta_1} &= \frac{\partial \sigma_t^2}{\partial \omega} = 1 \\ \frac{\partial \sigma_t^2}{\partial \theta_{1+j_1}} &= \frac{\partial \sigma_t^2}{\partial \phi_{j_1}^+} = \sum_{i \in \mathcal{I}_t^+} \frac{\partial \alpha_i}{\partial \phi_{j_1}^+} \varepsilon_{t-i}^2 \\ \frac{\partial \sigma_t^2}{\partial \theta_{1+r+j_2}} &= \frac{\partial \sigma_t^2}{\partial \phi_{j_2}^-} = \sum_{i \in \mathcal{I}_t^-} \frac{\partial \alpha_i}{\partial \phi_{j_2}^-} \varepsilon_{t-i}^2 \end{aligned} \tag{B.8}$$

whence $\dfrac{\partial^2 \sigma_t^2}{\partial \omega \partial \theta_{i_1}} = 0$ and $\dfrac{\partial^2 \sigma_t^2}{\partial \phi_{j_1}^+ \partial \omega} = \dfrac{\partial^2 \sigma_t^2}{\partial \phi_{j_1}^+ \partial \phi_{j_2}^-} = \dfrac{\partial^2 \sigma_t^2}{\partial \phi_{j_2}^- \partial \omega} = 0$. It is thus sufficient to show that for all $j_{1,h}, j_{2,h} \in \{1, ..., r\}$, $h = 1, ..., k$, $k \leq 3$, we have

$$\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \frac{1}{\sigma_t^2(\boldsymbol{\theta})} \frac{\partial^k \sigma_t^2(\boldsymbol{\theta})}{\partial \phi_{j_{1,1}}^+ ... \partial \phi_{j_{1,k}}^+} \right|^p < \infty \text{ and } \mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left| \frac{1}{\sigma_t^2(\boldsymbol{\theta})} \frac{\partial^k \sigma_t^2(\boldsymbol{\theta})}{\partial \phi_{j_{2,1}}^- ... \partial \phi_{j_{2,k}}^-} \right|^p < \infty.$$

From (B.8) we have

$$\begin{aligned} & \left| \frac{\partial^k \sigma_t^2(\boldsymbol{\theta})}{\partial \phi_{j_{1,1}}^+ ... \partial \phi_{j_{1,k}}^+} \right| \\ \leq & \sum_{i \in \mathcal{I}_t^+} \left| \frac{\partial^k \alpha_i(\boldsymbol{\phi}^+)}{\partial \phi_{j_{1,1}}^+ ... \partial \phi_{j_{1,k}}^+} \right| \varepsilon_{t-i}^2 \\ \leq & \sum_{i \in \mathcal{I}_t^+} \left| \frac{\partial^k \alpha_i(\boldsymbol{\phi}^+)}{\partial \phi_{j_{1,1}}^+ ... \partial \phi_{j_{1,k}}^+} \right| \left[\varepsilon_{t-i}^2\right]^{\frac{\rho}{p}} \left[\varepsilon_{t-i}^2\right]^{1-\frac{\rho}{p}} \left[\alpha_i(\boldsymbol{\phi}^+)\right]^{1-\frac{\rho}{p}} \left[\alpha_i(\boldsymbol{\phi}^+)\right]^{\frac{\rho}{p}-1} \end{aligned}$$

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

$$\begin{aligned} & \left| \frac{\partial^k \sigma_t^2(\boldsymbol{\theta})}{\partial \phi_{j_{1,1}}^+ ... \partial \phi_{j_{1,k}}^+} \right| \\ \leq & \left[\sum_{i \in \mathcal{I}_t^+} \left| \frac{\partial^k \alpha_i(\boldsymbol{\phi}^+)}{\partial \phi_{j_{1,1}}^+ ... \partial \phi_{j_{1,k}}^+} \right|^{\frac{p}{\rho}} \alpha_i^{1-\frac{p}{\rho}}(\boldsymbol{\phi}^+) \varepsilon_{t-i}^2 \right]^{\frac{\rho}{p}} \left[\sum_{i \in \mathcal{I}_t^+} \alpha_i(\boldsymbol{\phi}^+) \varepsilon_{t-i}^2 \right]^{1-\frac{\rho}{p}} \\ \leq & \left[\sum_{i \in \mathcal{I}_t^+} \left| \frac{\partial^k \alpha_i(\boldsymbol{\phi}^+)}{\partial \phi_{j_{1,1}}^+ ... \partial \phi_{j_{1,k}}^+} \right|^{\frac{p}{\rho}} \alpha_i^{1-\frac{p}{\rho}}(\boldsymbol{\phi}^+) \varepsilon_{t-i}^2 \right]^{\frac{\rho}{p}} \left[\sigma_t^2(\boldsymbol{\theta})\right]^{1-\frac{\rho}{p}}, \end{aligned}$$

whence

$$\left|\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial^k\sigma_t^2(\boldsymbol{\theta})}{\partial\phi^+_{j_{1,1}}...\partial\phi^+_{j_{1,k}}}\right|^p \leq \sum_{i\in\mathcal{I}_t^+}\left|\frac{\partial^k\alpha_i(\boldsymbol{\phi}^+)}{\partial\phi^+_{j_{1,1}}...\partial\phi^+_{j_{1,k}}}\right|^p \alpha_i^{\rho-p}(\boldsymbol{\phi}^+)|\varepsilon_{t-i}|^{2\rho}\left[\sigma_t^2(\boldsymbol{\theta})\right]^{-\rho}$$
$$\leq K\sum_{i\in\mathcal{I}_t^+}\left|\frac{\partial^k\alpha_i(\boldsymbol{\phi}^+)}{\partial\phi^+_{j_{1,1}}...\partial\phi^+_{j_{1,k}}}\right|^p \alpha_i^{\rho-p}(\boldsymbol{\phi}^+)|\varepsilon_{t-i}|^{2\rho}$$

and from Assumptions **A3(ii)** and **A10(i)**

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

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 $\mathbb{E}_{\boldsymbol{\theta_0}}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\dfrac{1}{\sigma_t^2(\boldsymbol{\theta})}\dfrac{\partial^k\sigma_t^2(\boldsymbol{\theta})}{\partial\phi^+_{j_{1,1}}...\partial\phi^+_{j_{1,k}}}\right|^p < \infty$ and similarly we can show $\mathbb{E}_{\boldsymbol{\theta_0}}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\dfrac{1}{\sigma_t^2(\boldsymbol{\theta})}\dfrac{\partial^k\sigma_t^2(\boldsymbol{\theta})}{\partial\phi^-_{j_{2,1}}...\partial\phi^-_{j_{2,k}}}\right|^p < \infty$ whence $\mathbb{E}_{\boldsymbol{\theta_0}}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\dfrac{1}{\sigma_t^2(\boldsymbol{\theta})}\dfrac{\partial^k\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}...\partial\theta_{i_k}}\right|^p < \infty$

□

We can now develop the proof of Theorem 3 on the asymptotic normality of the QMLE.

Proof of Theorem 3. From Theorem 2, we have that $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}$ converges to $\boldsymbol{\theta_0}$ which, from **A6**, belongs in 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

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

where the $\boldsymbol{\theta^*_{ij}}$ are between $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}$ and $\boldsymbol{\theta_0}$.

We will show the result by proving that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n\frac{\partial\tilde{l}_t}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta_0}) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0},(\kappa_\eta - 1)\boldsymbol{J})$$

and that

$$\frac{1}{n}\sum_{t=1}^{n} \frac{\partial^2 \tilde{l}_t}{\partial\theta_i \partial\theta_j}(\boldsymbol{\theta}^*_{\boldsymbol{ij}}) \to J(i,j) \text{ in probability.}$$

To do so, we will show the following intermediate results:

(a) Integrability of the derivatives of the criterion at $\boldsymbol{\theta_0}$
$\mathbb{E}_{\boldsymbol{\theta_0}} \left\| \frac{\partial l_t(\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta}} \frac{\partial l_t(\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta'}} \right\| < \infty$, $\mathbb{E}_{\boldsymbol{\theta_0}} \left\| \frac{\partial^2 l_t(\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta'}} \right\| < \infty$

(b) Invertibility of $\boldsymbol{J}$ and connection with the variance of the criterion derivative
$\boldsymbol{J}$ is invertible and $\mathbb{V}_{\boldsymbol{\theta_0}} \left[\frac{\partial l_t(\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta}} \right] = (\kappa_\eta - 1)\boldsymbol{J}$

(c) Uniform integrability of the third-order derivatives of the criterion
There exists a neighborhood $V(\boldsymbol{\theta_0})$ of $\boldsymbol{\theta_0}$ such that, for all $k_1, k_2, k_3 \in \{1, ..., 2r+1\}$,

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

(d) Asymptotic decrease of the effect of the initial values
$\left\| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\frac{\partial l_t(\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta}} - \frac{\partial \tilde{l}_t(\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta}} \right) \right\|$ and $\sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta_0})} \left\| \frac{1}{n} \sum_{t=1}^{n} \left(\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta'}} - \frac{\partial^2 \tilde{l}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta'}} \right) \right\|$ tend to 0 in probability as n tends to infinity.

(e) Central Limit Theorem for martingale increments
$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial l_t}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta_0}) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, (\kappa_\eta - 1)\boldsymbol{J})$

(f) Use of a second Taylor expansion and of the ergodic theorem
$\sum_{t=1}^{n} \frac{\partial^2 l_t}{\partial\theta_i \partial\theta_j}(\boldsymbol{\theta}^*_{\boldsymbol{ij}}) \to J(i,j)$ in probability

In the following, we detail the demonstration of the six previous points:

(a) Integrability of the derivatives of the criterion at $\boldsymbol{\theta_0}$

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

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

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

$$\mathbb{E}_{\boldsymbol{\theta_0}}\left\|\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta_0})\right\| < \infty,\ \mathbb{E}_{\boldsymbol{\theta_0}}\left\|\frac{1}{\sigma_t^2}\frac{\partial^2 \sigma_t^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta'}}(\boldsymbol{\theta_0})\right\| < \infty,\ \mathbb{E}_{\boldsymbol{\theta_0}}\left\|\frac{1}{\sigma_t^4}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta'}}(\boldsymbol{\theta_0})\right\| < \infty.$$

From (B.1) and Lemma 1, we have that for any $j_1, j_2 \in \{1, ..., r\}$

$$\frac{\partial \sigma_t^2}{\partial \omega}(\boldsymbol{\theta_0}) = 1,\ \left|\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \phi_{j_1}^+}\right|(\boldsymbol{\theta_0}) < \infty \text{ and } \left|\frac{1}{\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \phi_{j_2}^-}\right|(\boldsymbol{\theta_0}) < \infty$$

which proves the first inequality, and

$$\begin{gathered}
\frac{\partial^2 \sigma_t^2}{\partial \omega \partial \phi_{j_1}^+}(\boldsymbol{\theta_0}) = \frac{\partial^2 \sigma_t^2}{\partial \omega \partial \phi_{j_2}^-}(\boldsymbol{\theta_0}) = \frac{\partial^2 \sigma_t^2}{\partial \phi_{j_1}^+ \partial \phi_{j_2}^-}(\boldsymbol{\theta_0}) = 0, \\
\left|\frac{1}{\sigma_t^2}\frac{\partial^2 \sigma_t^2}{\partial \phi_{j_1}^+ \partial \phi_{j_2}^+}\right|(\boldsymbol{\theta_0}) < \infty \text{ and } \left|\frac{1}{\sigma_t^2}\frac{\partial^2 \sigma_t^2}{\partial \phi_{j_1}^- \partial \phi_{j_2}^-}\right|(\boldsymbol{\theta_0}) < \infty
\end{gathered}$$

which proves the second inequality.

Since $\dfrac{1}{\sigma_t^2}\dfrac{\partial \sigma_t^2}{\partial \omega}$, $\dfrac{1}{\sigma_t^2}\dfrac{\partial \sigma_t^2}{\partial \phi_{j1}^+}$ and $\dfrac{1}{\sigma_t^2}\dfrac{\partial \sigma_t^2}{\partial \phi_{j_2}^-}$ are bounded at $\boldsymbol{\theta_0}$, we can conclude that

$$\mathbb{E}_{\boldsymbol{\theta_0}}\left\|\frac{1}{\sigma_t^4}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}}\frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta'}}(\boldsymbol{\theta_0})\right\| < \infty$$

which finishes the proof.

(b) Invertibility of J and connection with the variance of the criterion derivative

Since 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, we have

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

because $\mathbb{E}_{\boldsymbol{\theta_0}}\eta_t^2 = 1$ 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 **A7** we can write

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

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

$$\begin{aligned}&\boldsymbol{\lambda}'\boldsymbol{J}\boldsymbol{\lambda} = 0\\ \Leftrightarrow\ &\mathbb{E}_{\boldsymbol{\theta_0}}\left[\frac{1}{\sigma_t^4(\boldsymbol{\theta_0})}\left(\boldsymbol{\lambda'}\frac{\partial\sigma_t^2(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}}\right)^2\right] = 0\\ \Leftrightarrow\ &\boldsymbol{\lambda}'\left[\frac{\partial\sigma_t^2(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}}\right] = 0\\ \Leftrightarrow\ &\lambda_0 + \sum_{i=1}^{\infty}\left[\sum_{j=1}^{r}\lambda_j^+\frac{\partial\alpha_i(\boldsymbol{\phi_0^+})}{\partial\phi_j^+}\mathbb{1}_{\varepsilon_{t-i}\geq 0} + \sum_{k=1}^{r}\lambda_k^-\frac{\partial\alpha_i(\boldsymbol{\phi_0^-})}{\partial\phi_k^-}\mathbb{1}_{\varepsilon_{t-i}<0}\right]\varepsilon_{t-i}^2 = 0.\end{aligned}$$

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

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

whence $\eta_{t-1}^2 \in \mathcal{F}\left(\eta_{t-2}^2, ...\right)$ and thus, by independence,

$$\begin{cases}\sum_{j=1}^{r}\lambda_j^+\frac{\partial\alpha_1(\boldsymbol{\phi_0^+})}{\partial\phi_j^+}\mathbb{1}_{\eta_{t-1}\geq 0}\eta_{t-1}^2 \text{ is constant almost surely}\\ \sum_{k=1}^{r}\lambda_k^-\frac{\partial\alpha_1(\boldsymbol{\phi_0^-})}{\partial\phi_k^-}\mathbb{1}_{\eta_{t-1}<0}\eta_{t-1}^2 \text{ is constant almost surely}\end{cases} \tag{B.9}$$

However, since, from **A2**, η_1 takes at least two positive (respectively negative) values, (B.9) implies almost surely

$$\begin{cases}\boldsymbol{\lambda}^{+\prime}\ \frac{\partial\alpha_1(\boldsymbol{\phi_0^+})}{\partial\boldsymbol{\phi}^+} = 0\\ \boldsymbol{\lambda}^{-\prime}\ \frac{\partial\alpha_1(\boldsymbol{\phi_0^-})}{\partial\boldsymbol{\phi}^-} = 0.\end{cases}$$

Iterating this argument we obtain that for all $i_h = i_h(\boldsymbol{\phi}_{\mathbf{0}}^+), i_h = 1, ..., r$, we have $\boldsymbol{\lambda}^{+'} \dfrac{\partial \alpha_{i_h}(\boldsymbol{\phi}_{\mathbf{0}}^+)}{\partial \boldsymbol{\phi}^+} = 0$ and thus from **A10(ii)** we must have $\boldsymbol{\lambda}^+ = \mathbf{0}$. Similarly, we obtain $\boldsymbol{\lambda}^- = \mathbf{0}$. This implies $\lambda_0 = 0$ and contradicts the singularity of $\boldsymbol{J}$.

(c) Uniform integrability of the third-order derivatives of the criterion

We have

$$\begin{aligned}\frac{\partial^3 l_t(\boldsymbol{\theta})}{\partial\theta_{i_1}\partial\theta_{i_2}\partial\theta_{i_3}} &= \left[1-\frac{\varepsilon_t^2}{\sigma_t^2}\right]\left[\frac{1}{\sigma_t^2}\frac{\partial^3\sigma_t^2}{\partial\theta_{i_1}\partial\theta_{i_2}\partial\theta_{i_3}}\right] \\ &+\left[2\frac{\varepsilon_t^2}{\sigma_t^2}-1\right]\left[\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_{i_1}}\right]\left[\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta_{i_2}\partial\theta_{i_3}}\right] \\ &+\left[2\frac{\varepsilon_t^2}{\sigma_t^2}-1\right]\left[\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_{i_2}}\right]\left[\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta_{i_1}\partial\theta_{i_3}}\right] \\ &+\left[2\frac{\varepsilon_t^2}{\sigma_t^2}-1\right]\left[\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_{i_3}}\right]\left[\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta_{i_1}\partial\theta_{i_2}}\right] \\ &+\left[2-6\frac{\varepsilon_t^2}{\sigma_t^2}\right]\left[\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_{i_1}}\right]\left[\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_{i_2}}\right]\left[\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_{i_3}}\right](\boldsymbol{\theta}).\end{aligned}$$

By **A11**, there exists a neighborhood $V(\boldsymbol{\theta}_{\mathbf{0}})$ of $\boldsymbol{\theta}_{\mathbf{0}}$ such that,

$$\mathbb{E}_{\boldsymbol{\theta}_{\mathbf{0}}} \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_{\mathbf{0}})} \left[\frac{\sigma_t^2(\boldsymbol{\theta}_{\mathbf{0}})}{\sigma_t^2(\boldsymbol{\theta})}\right]^2 < \infty, \tag{B.10}$$

and the triangle inequality gives

$$\left\|\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_{\mathbf{0}})}\frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})}\right\|_2 = \sqrt{\kappa_\eta}\left\|\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_{\mathbf{0}})}\frac{\sigma_t^2(\boldsymbol{\theta}_{\mathbf{0}})}{\sigma_t^2(\boldsymbol{\theta})}\right\|_2 < \infty$$

by **A7**.

Using Lemma 1, we have for all $i_1, i_2, i_3 \in \{1, ..., 2r+1\}$

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{\theta}_{\mathbf{0}}}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right|^p < \infty, \\ &\mathbb{E}_{\boldsymbol{\theta}_{\mathbf{0}}}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial^2\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}\partial\theta_{i_2}}\right|^p < \infty, \\ &\mathbb{E}_{\boldsymbol{\theta}_{\mathbf{0}}}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial^3\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}\partial\theta_{i_2}\partial\theta_{i_3}}\right|^p < \infty,\end{aligned} \tag{B.11}$$

and we thus obtain, using the Cauchy-Schwartz inequality and the Hölder inequal-

ity,

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

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

and

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta}\in V_\tau(\boldsymbol{\theta_0})} \left| \left[2-6\frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})}\right] \left[\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right] \left[\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_2}}\right] \left[\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_3}}\right]\right| \\
&\quad \leq \left\| \sup_{\boldsymbol{\theta}\in V_\tau(\boldsymbol{\theta_0})} \left|2-6\frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})}\right| \right\|_2 \max_h \left\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_h}}\right| \right\|_6^3 \\
&\quad < \infty,
\end{aligned}$$

which concludes the proof.

(d) Asymptotic decrease of the effect of the initial values

Let us first consider the derivatives of $\tilde{\sigma}_t^2$. Similarly to (B.1), we can rewrite (12) as

$$\tilde{\sigma}_t^2(\boldsymbol{\theta_0}) = \omega_0 + \sum_{i\in\tilde{\mathcal{I}}_t^+} \alpha_i(\boldsymbol{\phi_0^+})\varepsilon_{t-i}^2 + \sum_{j\in\tilde{\mathcal{I}}_t^-} \alpha_j(\boldsymbol{\phi_0^-})\varepsilon_{t-j}^2 \tag{B.12}$$

where we denote by $\tilde{\mathcal{I}}_t^+$ (respectively $\tilde{\mathcal{I}}_t^-$) the sets

$$\tilde{\mathcal{I}}_t^{+(-)} = \{i < t \text{ such that } \varepsilon_{t-i} \geq 0 (< 0)\}.$$

From (B.12) we have for any $j_{1,h}, j_{2,h} \in \{1, ..., r\}$, $h \in \{1, 2\}$

$$\begin{aligned}
&\frac{\partial\tilde{\sigma}_t^2(\boldsymbol{\theta})}{\partial\omega} = 1\ ,\ \ \frac{\partial\tilde{\sigma}_t^2(\boldsymbol{\theta})}{\partial\phi_{j_{1,1}}^+} = \sum_{i\in\tilde{\mathcal{I}}_t^+} \frac{\partial\alpha_i(\boldsymbol{\phi^+})}{\partial\phi_{j_{1,1}}^+}\varepsilon_{t-i}^2\ ,\ \ \frac{\partial\tilde{\sigma}_t^2(\boldsymbol{\theta})}{\partial\phi_{j_{2,1}}^-} = \sum_{i\in\tilde{\mathcal{I}}_t^-} \frac{\partial\alpha_i(\boldsymbol{\phi^-})}{\partial\phi_{j_{2,1}}^-}\varepsilon_{t-i}^2, \\
&\frac{\partial^2\tilde{\sigma}_t^2(\boldsymbol{\theta})}{\partial\omega\partial\phi_{j_{1,1}}^+} = \frac{\partial^2\tilde{\sigma}_t^2(\boldsymbol{\theta})}{\partial\omega\partial\phi_{j_{2,1}}^-} = \frac{\partial^2\tilde{\sigma}_t^2(\boldsymbol{\theta})}{\partial\phi_{j_{1,1}}^+\partial\phi_{j_{2,1}}^-} = 0\ , \\
&\text{and} \\
&\frac{\partial^2\tilde{\sigma}_t^2(\boldsymbol{\theta})}{\partial\phi_{j_{1,1}}^+\partial\phi_{j_{1,2}}^+} = \sum_{i\in\tilde{\mathcal{I}}_t^+} \frac{\partial^2\alpha_i(\boldsymbol{\phi^+})}{\partial\phi_{j_{1,1}}^+\partial\phi_{j_{1,2}}^+}\varepsilon_{t-i}^2\ ,\ \ \frac{\partial^2\tilde{\sigma}_t^2(\boldsymbol{\theta})}{\partial\phi_{j_{2,1}}^-\partial\phi_{j_{2,2}}^-} = \sum_{i\in\tilde{\mathcal{I}}_t^-} \frac{\partial^2\alpha_i(\boldsymbol{\phi^-})}{\partial\phi_{j_{2,1}}^-\partial\phi_{j_{2,2}}^-}\varepsilon_{t-i}^2.
\end{aligned} \tag{B.13}$$

Remark now that, from **A3(ii)** and **A11**, on a neighborhood $V(\boldsymbol{\theta_0})$ of $\boldsymbol{\theta_0}$, we have

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})} \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2(\boldsymbol{\theta})} &= \eta_t^2 \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})} \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})} \left[1 + \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})} \frac{\sigma_t^2(\boldsymbol{\theta}) - \tilde{\sigma}_t^2(\boldsymbol{\theta})}{\tilde{\sigma}_t^2(\boldsymbol{\theta})}\right] \\
&\leq K\eta_t^2 \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})} \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})} \left[1 + \sum_{i=0}^{\infty}(t+i)^{-(d+1)}\varepsilon_{-i}^2\right] \\
&\leq K\eta_t^2 \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})} \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})} \left[1 + \sum_{i=0}^{\infty}i^{-(d+1)}\varepsilon_{-i}^2\right] \\
&\leq K\eta_t^2 \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})} \frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}
\end{aligned} \tag{B.14}$$

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

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

therefore we can write

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

From the Markov inequality we have

$$\begin{aligned}
&\mathbb{P}\left[\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[\frac{\partial l_t(\boldsymbol{\theta_0})}{\partial \theta_k} - \frac{\partial \tilde{l}_t(\boldsymbol{\theta_0})}{\partial \theta_k}\right]\right| > \varepsilon\right] \\
&\quad \leq \frac{1}{\varepsilon}\,\mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[\frac{\partial l_t(\boldsymbol{\theta_0})}{\partial \theta_k} - \frac{\partial \tilde{l}_t(\boldsymbol{\theta_0})}{\partial \theta_k}\right]\right| \\
&\quad \leq \frac{1}{\varepsilon}\left[\mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} A_t(\boldsymbol{\theta_0})\right| + \mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} B_t(\boldsymbol{\theta_0})\right| + \mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} C_t(\boldsymbol{\theta_0})\right|\right]
\end{aligned} \tag{B.15}$$

First consider $\mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} A_t(\boldsymbol{\theta_0})\right|$. From (B.3), we have

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

whence, using the independence of η_t^2 with σ_t^2 and its derivatives at $\boldsymbol{\theta_0}$, (11), **A2** and **A8**,

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta_0}}|A_t(\boldsymbol{\theta_0})|^\rho &\leq K\mathbb{E}_{\boldsymbol{\theta_0}}|\eta_t|^{2\rho}\mathbb{E}_{\boldsymbol{\theta_0}}\left(\left[\sum_{i=t}^{\infty} a_{i,t-i}(\boldsymbol{\phi_0^+},\boldsymbol{\phi_0^-})\varepsilon_{t-i}^2\right]^\rho\left|\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_k}(\boldsymbol{\theta_0})\right|^\rho\right) \\
&\leq K\mathbb{E}_{\boldsymbol{\theta_0}}\left(\left[\sum_{i=t}^{\infty} i^{-(d^*+1)}\varepsilon_{t-i}^2\right]^\rho\left|\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_k}(\boldsymbol{\theta_0})\right|^\rho\right).
\end{aligned}$$

Since $\rho<1$ from **A9**, there exists some $\xi>0$ such that $\rho(1+\xi)\leq 1$. Hence, from Hölder inequality, along with Lemma 1, we obtain

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta_0}}|A_t(\boldsymbol{\theta_0})|^\rho &\leq K\left(\mathbb{E}_{\boldsymbol{\theta_0}}\left[\sum_{i=t}^{\infty} i^{-(d^*+1)}\varepsilon_{t-i}^2\right]^{\rho(1+\xi)}\right)^{\frac{1}{1+\xi}}\left(\mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_k}(\boldsymbol{\theta_0})\right|^{\rho\frac{\xi+1}{\xi}}\right)^{\frac{\xi}{1+\xi}} \\
&\leq K\left(\sum_{i=t}^{\infty} i^{-(d^*+1)\rho(1+\xi)}\mathbb{E}_{\boldsymbol{\theta_0}}|\varepsilon_{t-i}|^{2\rho(1+\xi)}\right)^{\frac{1}{1+\xi}} \\
&\leq K\sum_{i=t}^{\infty} i^{-(d^*+1)\rho}\left(\mathbb{E}_{\boldsymbol{\theta_0}}|\varepsilon_{t-i}|^{2\rho(1+\xi)}\right)^{\frac{1}{1+\xi}} \\
&\leq K\sum_{i=0}^{\infty} (t+i)^{-(d^*+1)\rho}\left(\mathbb{E}_{\boldsymbol{\theta_0}}|\varepsilon_{-i}|^{2\rho(1+\xi)}\right)^{\frac{1}{1+\xi}} \\
&\leq Kt^{-(d^*+1)\rho+1},
\end{aligned}$$

and thus

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

since from **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.

Consider now $\mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} B_t(\boldsymbol{\theta_0})\right|$. We have, from (B.14)

$$\begin{aligned}|B_t(\boldsymbol{\theta_0})| &= \left|\left[1-\frac{\varepsilon_t^2}{\tilde{\sigma}_t^2}\right]\left[\frac{1}{\tilde{\sigma}_t^2}-\frac{1}{\sigma_t^2}\right]\left[\frac{\partial\sigma_t^2}{\partial\theta_k}\right]\right|(\boldsymbol{\theta_0})\\ &\leq K\eta_t^2\left[\frac{\sigma_t^2(\boldsymbol{\theta_0})-\tilde{\sigma}_t^2(\boldsymbol{\theta_0})}{\tilde{\sigma}_t^2(\boldsymbol{\theta_0})}\right]\left|\left[\frac{1}{\sigma_t^2(\boldsymbol{\theta_0})}\frac{\partial\sigma_t^2(\boldsymbol{\theta_0})}{\partial\theta_k}\right]\right|,\end{aligned}$$

and thus

$$\mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}B_t(\boldsymbol{\theta_0})\right|^{\rho} \underset{n\to\infty}{\to} 0$$

from the same previous arguments. Using Markov inequality, we can conclude $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}|B_t(\boldsymbol{\theta_0})|$ tends to 0 in probability.

Finally consider $\mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} C_t(\boldsymbol{\theta_0})\right|$. From (B.8) and (B.13), and from **A10(i)** and **A3(ii)**, we have for all $\xi>0$,

$$\begin{aligned}|C_t(\boldsymbol{\theta_0})| &= \left|\left[1-\frac{\varepsilon_t^2}{\tilde{\sigma}_t^2}\right]\left[\frac{1}{\tilde{\sigma}_t^2}\right]\left[\frac{\partial\sigma_t^2}{\partial\theta_k}-\frac{\partial\tilde{\sigma}_t^2}{\partial\theta_k}\right]\right|(\boldsymbol{\theta_0})\\ &\leq K\eta_t^2\left|\frac{\partial\sigma_t^2(\boldsymbol{\theta_0})}{\partial\theta_k}-\frac{\partial\tilde{\sigma}_t^2(\boldsymbol{\theta_0})}{\partial\theta_k}\right|\\ &\leq K\eta_t^2\sum_{i=t}^{\infty}\max\left(\alpha_i(\boldsymbol{\phi_0^+}),\alpha_i(\boldsymbol{\phi_0^-})\right)^{1-\xi}\varepsilon_{t-i}^2\\ &\leq K\eta_t^2\sum_{i=t}^{\infty}i^{-(d+1)(1-\xi)}\varepsilon_{t-i}^2\\ &\leq K\eta_t^2\sum_{i=0}^{\infty}(t+i)^{-(d+1)(1-\xi)}\varepsilon_{-i}^2,\end{aligned}$$

and thus

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

since, from **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 (B.15) 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 $\boldsymbol{\theta_0}$.
We denote $\chi_t = \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta_0})} |\sigma_t^2(\boldsymbol{\theta}) - \tilde{\sigma}_t^2(\boldsymbol{\theta})|$, and we have from (B.3), (11), and **A3(ii)**

$$\begin{aligned}
\chi_t &= \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta_0})} \sum_{i=t}^{\infty} a_{i,t-i}(\boldsymbol{\phi}^+, \boldsymbol{\phi}^-)\varepsilon_{t-i}^2 \\
&\leq \sum_{i=t}^{\infty} \sup_{\boldsymbol{\phi} \in \boldsymbol{\Phi}} \alpha_i(\boldsymbol{\phi})\varepsilon_{t-i}^2 \\
&\leq K \sum_{i=t}^{\infty} i^{-(d+1)}\varepsilon_{t-i}^2 \\
&\leq K \sum_{i=0}^{\infty} (i+t)^{-(d+1)}\varepsilon_{-i}^2,
\end{aligned}$$

whence

$$\begin{aligned}
\mathbb{E}\chi_t^{\rho} &\leq K \sum_{i=0}^{\infty} (i+t)^{-(d+1)\rho}\ \mathbb{E}|\varepsilon_{-i}|^{2\rho} \\
&\leq K t^{-(d+1)\rho+1}
\end{aligned}$$

since, from **A4**, $\mathbb{E}\,|\varepsilon_t|^{2\rho} < \infty$. 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

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

and

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

where $V(\boldsymbol{\theta_0})$ is a neighborhood of $\boldsymbol{\theta_0}$ and $i_1, i_2 \in \{1, \ldots, 2r+1\}$.

From (B.8) and (B.13) we easily obtain $\chi_t^{(1)} = 0$, and from **A10(i)** and **A3(ii)**, we have for all $\xi > 0$,

$$\chi_t^{(i_1)} \leq \begin{cases} \sum_{i=t}^{\infty} \max_{j\in\{1,\ldots,r\}} \sup_{\boldsymbol{\phi} \in V(\boldsymbol{\phi_0^+})} \left| \frac{\partial \alpha_i(\boldsymbol{\phi})}{\partial \phi_j^+} \right| \varepsilon_{t-i}^2 \text{ if } 1 < i_1 \leq r+1 \\ \sum_{i=t}^{\infty} \max_{j\in\{1,\ldots,r\}} \sup_{\boldsymbol{\phi} \in V(\boldsymbol{\phi_0^-})} \left| \frac{\partial \alpha_i(\boldsymbol{\phi})}{\partial \phi_j^-} \right| \varepsilon_{t-i}^2 \text{ if } r+1 < i_1 \leq 2r+1 \end{cases}$$

whence

$$\begin{aligned}\chi_t^{(i_1)} &\leq \sum_{i=t}^{\infty} \sup_{\boldsymbol{\phi}\in\boldsymbol{\Phi}} \alpha_i(\boldsymbol{\phi})^{(1-\xi)}\varepsilon_{t-i}^2 \\ &\leq K\sum_{i=t}^{\infty} i^{-(d+1)(1-\xi)}\varepsilon_{t-i}^2 \\ &\leq K\sum_{i=0}^{\infty} (i+t)^{-(d+1)(1-\xi)}\varepsilon_{-i}^2.\end{aligned}$$

Finally, we have

$$\begin{aligned}\mathbb{E}\left(\chi_t^{(i_1)}\right)^{\rho} &\leq \sum_{i=0}^{\infty} (i+t)^{-(d+1)\rho(1-\xi)}\mathbb{E}|\varepsilon_{-i}|^{2\rho} \\ &\leq Kt^{-(d+1)\rho(1-\xi)+1}\end{aligned}$$

since, from **A4**, $\mathbb{E}\left|\varepsilon_t\right|^{2\rho} < \infty$. 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}\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^2}\frac{\partial^2\sigma_t^2}{\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^2}-\frac{1}{\tilde{\sigma}_t^2}\right)\frac{\partial^2\sigma_t^2}{\partial\theta_{i_1}\partial\theta_{i_2}} + \frac{1}{\tilde{\sigma}_t^2}\left(\frac{\partial^2\sigma_t^2}{\partial\theta_{i_1}\partial\theta_{i_2}} - \frac{\partial^2\tilde{\sigma}_t^2}{\partial\theta_{i_1}\partial\theta_{i_2}}\right)\right] \\ &+\left[2\frac{\varepsilon_t^2}{\sigma_t^2} - 2\frac{\varepsilon_t^2}{\tilde{\sigma}_t^2}\right]\left[\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_{i_1}}\right]\left[\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_{i_2}}\right] \\ &+\left[2\frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} - 1\right]\left[\left(\frac{1}{\sigma_t^2}-\frac{1}{\tilde{\sigma}_t^2}\right)\frac{\partial\sigma_t^2}{\partial\theta_{i_1}} + \frac{1}{\tilde{\sigma}_t^2}\left(\frac{\partial\sigma_t^2}{\partial\theta_{i_1}} - \frac{\partial\tilde{\sigma}_t^2}{\partial\theta_{i_1}}\right)\right]\left[\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_{i_2}}\right] \\ &\left.+\left[2\frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} - 1\right]\left[\left(\frac{1}{\sigma_t^2}-\frac{1}{\tilde{\sigma}_t^2}\right)\frac{\partial\sigma_t^2}{\partial\theta_{i_2}} + \frac{1}{\tilde{\sigma}_t^2}\left(\frac{\partial\sigma_t^2}{\partial\theta_{i_2}} - \frac{\partial\tilde{\sigma}_t^2}{\partial\theta_{i_2}}\right)\right]\left[\frac{1}{\tilde{\sigma}_t^2}\frac{\partial\tilde{\sigma}_t^2}{\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^2(\boldsymbol{\theta})}\frac{\partial^2\sigma_t^2(\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^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right]\left[\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\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^2(\boldsymbol{\theta})}\frac{\partial\tilde{\sigma}_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right]\left[\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\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^2(\boldsymbol{\theta})}\frac{\partial\tilde{\sigma}_t^2(\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^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\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 1, for all $p>0$, $i_1,i_2=1,...,2r+1$,

$$
\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta_0}}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{\tilde{\sigma}_t^2(\boldsymbol{\theta})}\frac{\partial\tilde{\sigma}_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right|^p<\infty \\
&\mathbb{E}_{\boldsymbol{\theta_0}}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{\tilde{\sigma}_t^2(\boldsymbol{\theta})}\frac{\partial^2\tilde{\sigma}_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}\partial\theta_{i_2}}\right|^p<\infty.
\end{aligned}
\tag{B.16}
$$

Then, from independence of η_t^2 with σ_t^2 and its derivatives, **A11**, Lemma 1, (B.14), and (B.16) we have, using Hölder inequality, for all i_1,i_2,

$$
\begin{aligned}
&\mathbb{E}\left[\eta_t^2\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\left[\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial^2\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}\partial\theta_{i_2}}\right]\right|\right]<\infty \\
&\mathbb{E}\left[\eta_t^2\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left[\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right]\left[\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_2}}\right]\right|\right]<\infty \\
&\mathbb{E}\left[\eta_t^2\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left[\frac{1}{\tilde{\sigma}_t^2(\boldsymbol{\theta})}\frac{\partial\tilde{\sigma}_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right]\left[\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_2}}\right]\right|\right]<\infty \\
&\mathbb{E}\left[\eta_t^2\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left[\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right]\right|\right]<\infty \\
&\mathbb{E}\left[\eta_t^2\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left[\frac{1}{\tilde{\sigma}_t^2(\boldsymbol{\theta})}\frac{\partial\tilde{\sigma}_t^2(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right]\right|\right]<\infty.
\end{aligned}
\tag{B.17}
$$

Since χ_t, $\chi_t^{(i_1)}$, and $\chi_t^{(i_1,i_2)}$ tends to 0 almost surely as t tends to infinity, and (B.17), Toeplit lemma entails that

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

almost surely, which concludes the proof.

(e) Central Limit Theorem for martingale increments
Using the fact that $\sigma_t^2(\boldsymbol{\theta_0})$ 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

$$\mathbb{E}_{\boldsymbol{\theta_0}}\left[\frac{\partial l_t(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}}\ |\varepsilon_u, u<t\right] = \frac{1}{\sigma_t^4(\boldsymbol{\theta_0})}\left[\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta_0})\right]\ \mathbb{E}_{\boldsymbol{\theta_0}}\left[\ \sigma_t^2(\boldsymbol{\theta_0}) - \varepsilon_t^2|\varepsilon_u, u<t\right] = 0$$

and we have shown that $\mathbb{V}_{\boldsymbol{\theta_0}}\left[\frac{\partial l_t}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta_0})\right]$ if 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 $\boldsymbol{\lambda} \in \mathbb{R}^{2r+1}$, the sequence $\left\{\boldsymbol{\lambda}'\ \frac{\partial l_t(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}}, \varepsilon_t\right\}_t$ is a square integrable ergodic stationary martingale difference. The Cramer-Wold theorem and the following corollary entail

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta_0}) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, (\kappa_\eta - 1)\boldsymbol{J})$$

*Corollary (**Billingsley, 1961**)*[1]: if $(v_t, \mathcal{F}_t)_t$ is a stationary and ergodic sequence of square integrable martingale increments such that $\sigma_v^2 = \mathbb{V}(v_t) \neq 0$ then

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} v_t \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_v^2)$$

(f) Use of a second Taylor expansion and of the ergodic theorem
Consider the Taylor expansion of the criterion at $\boldsymbol{\theta_0}$, we have for all i and j,

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

where $\boldsymbol{\tilde{\theta}_{ij}}$ is between $\boldsymbol{\theta^*_{ij}}$ and $\boldsymbol{\theta_0}$. The almost sure convergence of $\boldsymbol{\tilde{\theta}_{ij}}$ to $\boldsymbol{\theta_0}$, the ergodic theorem and the uniform integrability of the third-order derivatives of the criterion 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 $\|\boldsymbol{\theta}^*_{\boldsymbol{ij}}-\boldsymbol{\theta_0}\|\to 0$ almost surely, we have for all $\varepsilon>0$,

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

and by the ergodic theorem,

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

Using Slutsky lemma along with the previous intermediate results allows us to conclude that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial\tilde{l}_t}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta_0})\overset{\mathcal{L}}{\to}\mathcal{N}(0,(\kappa_\eta-1)\boldsymbol{J})$$

and that

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\tilde{l}_t}{\partial\theta_i\partial\theta_j}(\boldsymbol{\theta}^*_{\boldsymbol{ij}})\to J(i,j)\text{ in probability}$$

which ends the proof.

□

B.3 Specification tests

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

Proof of Proposition 1. We begin by studying the asymptotic distribution of the Wald statistic under the null H_0^{sym}. From 13 and Slutsky lemma, we obtain

$$\sqrt{n}\boldsymbol{R}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}-\boldsymbol{\theta_0})\overset{\mathcal{L}}{\to}\mathcal{N}(\boldsymbol{0},\boldsymbol{R}(\kappa_\eta-1)\boldsymbol{J}^{-1}\boldsymbol{R}') \tag{B.18}$$

and from the quadratic form, we thus have

$$\begin{aligned}&\sqrt{n}(\boldsymbol{R}\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}})'\left[\boldsymbol{R}(\kappa_\eta-1)J^{-1}\boldsymbol{R}')\right]^{-1}\sqrt{n}(\boldsymbol{R}\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}})\overset{\mathcal{L}}{\to}\chi^2_c \\ \Leftrightarrow\quad &(\boldsymbol{R}\,\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}})'\left(\boldsymbol{R}\left(\frac{(\hat{\kappa}_\eta-1)}{n}\hat{\boldsymbol{J}}_n^{-1}\right)\boldsymbol{R}'\right)^{-1}(\boldsymbol{R}\,\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}})=W_n^{\text{sym}}\overset{\mathcal{L}}{\to}\chi^2_c\end{aligned}$$

under $H_0^{\text{sym}} : \boldsymbol{R\theta_0} = 0$. Thus, the critical region of the Wald test at the asymptotic level α is $\{W_n^{\text{sym}} > \chi_c^2(1-\alpha)\}$.

To study the Rao-score statistic, we first introduce the Lagrangian function associated with the likelihood optimization problem constrained by H_0^{sym}, $\tilde{Q}_n(\boldsymbol{\theta}) + (\boldsymbol{R\theta})'\boldsymbol{\lambda}$. The first-order condition is then

$$\frac{\partial \tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}})}{\partial \boldsymbol{\theta}} + \boldsymbol{R}'\tilde{\boldsymbol{\lambda}}_n = 0 \tag{B.19}$$

with $\tilde{\boldsymbol{\lambda}}_n$ the Lagrange multipliers vector.

In the following, we use the notation $a \overset{o_P(1)}{=} b$ meaning $a = b + o_P(1)$. Under H_0^{sym}, we have

$$\sqrt{n}\boldsymbol{R}\tilde{\boldsymbol{\theta}}_n = \boldsymbol{R}\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0})$$

and

$$0 = \sqrt{n}\boldsymbol{R}\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}} = \boldsymbol{R}\sqrt{n}(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}} - \boldsymbol{\theta_0})$$

since $\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}}$ is the constrained estimator. By subtraction, we thus obtain

$$\sqrt{n}\boldsymbol{R}\tilde{\boldsymbol{\theta}}_n = \boldsymbol{R}\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}}). \tag{B.20}$$

Using Taylor expansions, we can also notice that

$$0 = \sqrt{n}\frac{\partial \tilde{Q}_n(\tilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} \overset{o_P(1)}{=} \sqrt{n}\frac{\partial \tilde{Q}_n(\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta}} + \sqrt{n}\boldsymbol{J}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}) \tag{B.21}$$

and

$$\sqrt{n}\frac{\partial \tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}})}{\partial \boldsymbol{\theta}} \overset{o_P(1)}{=} \sqrt{n}\frac{\partial \tilde{Q}_n(\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta}} + \sqrt{n}\boldsymbol{J}(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}} - \boldsymbol{\theta_0})$$

which yields by subtraction

$$\sqrt{n}\frac{\partial \tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}})}{\partial \boldsymbol{\theta}} \overset{o_P(1)}{=} -\sqrt{n}\boldsymbol{J}(\tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}})$$

hence

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}}) \overset{o_P(1)}{=} -\sqrt{n}\boldsymbol{J}^{-1}\frac{\partial \tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}})}{\partial \boldsymbol{\theta}}. \tag{B.22}$$

From (B.19), (B.20) and (B.22), we thus obtain

$$\sqrt{n}\boldsymbol{R}\tilde{\boldsymbol{\theta}}_n \overset{o_P(1)}{=} \boldsymbol{R}\boldsymbol{J}^{-1}\boldsymbol{R}'\sqrt{n}\tilde{\boldsymbol{\lambda}}_n$$

which yields

$$\sqrt{n}\tilde{\boldsymbol{\lambda}}_n \overset{o_P(1)}{=} \left[\boldsymbol{R}\boldsymbol{J}^{-1}\boldsymbol{R}'\right]^{-1}\sqrt{n}\boldsymbol{R}\tilde{\boldsymbol{\theta}}_n$$

hence from (B.18), under H_0^{sym},

$$\sqrt{\frac{n}{\kappa_\eta - 1}}\tilde{\boldsymbol{\lambda}}_n \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \left[\boldsymbol{R}\boldsymbol{J}^{-1}\boldsymbol{R}'\right]^{-1}\right)$$

as $n \to \infty$. Taking the quadratic form, we obtain under H_0^{sym},

$$\frac{n}{\hat{\kappa}_{\eta|H_0^{\text{sym}}} - 1}(\tilde{\boldsymbol{\lambda}}_n'\boldsymbol{R})\hat{\boldsymbol{J}}_{n|H_0^{\text{sym}}}^{-1}(\boldsymbol{R}'\tilde{\boldsymbol{\lambda}}_n) \xrightarrow{\mathcal{L}} \chi_c^2$$

and (B.19) yields

$$\frac{n}{\hat{\kappa}_{\eta|H_0^{\text{sym}}} - 1}\frac{\partial\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}})}{\partial\boldsymbol{\theta}'}\hat{\boldsymbol{J}}_{n|H_0^{\text{sym}}}^{-1}\frac{\partial\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}})}{\partial\boldsymbol{\theta}} = R_n^{\text{sym}} \xrightarrow{\mathcal{L}} \chi_c^2.$$

It follows that the critical region of the Rao-score test at the asymptotic level α is $\{R_n^{\text{sym}} > \chi_c^2(1-\alpha)\}$.

We finally focus on the Quasi Likelihood Ratio statistic. Using Taylor expansions, we get

$$\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_n) \overset{o_P(1)}{=} \tilde{Q}_n(\boldsymbol{\theta_0}) + \frac{\partial\tilde{Q}_n(\boldsymbol{\theta_0})}{\partial\theta'}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}) + \frac{1}{2}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0})'J(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0})$$

and

$$\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}}) \overset{o_P(1)}{=} \tilde{Q}_n(\boldsymbol{\theta_0}) + \frac{\partial\tilde{Q}_n(\boldsymbol{\theta_0})}{\partial\theta'}(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}} - \boldsymbol{\theta_0}) + \frac{1}{2}(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}} - \boldsymbol{\theta_0})'J(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}} - \boldsymbol{\theta_0}),$$

hence, by subtraction,

$$\begin{aligned}(\hat{\kappa}_{\eta|H_0^{\text{sym}}} - 1)L_n^{\text{sym}} \overset{o_P(1)}{=} \; & 2n\frac{\partial\tilde{Q}_n(\boldsymbol{\theta_0})}{\partial\theta'}(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}} - \tilde{\boldsymbol{\theta}}_n) + n(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}} - \boldsymbol{\theta_0})'J(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}} - \boldsymbol{\theta_0}) \\ & -n(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0})'J(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}),\end{aligned}$$

and, from (B.21),

$$\begin{aligned}(\hat{\kappa}_{\eta|H_0^{\text{sym}}} - 1)L_n^{\text{sym}} \overset{o_P(1)}{=} \; & 2n(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0})'J(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}} - \tilde{\boldsymbol{\theta}}_n) \\ & +n(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}} - \boldsymbol{\theta_0})'J(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}} - \boldsymbol{\theta_0}) - n(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0})'J(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}) \\ \overset{o_P(1)}{=} \; & n(\tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}})'J(\tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}}).\end{aligned}$$

From (B.22), it follows that under H_0^{sym},

$$L_n^{\text{sym}} \overset{o_P(1)}{=} \frac{2n}{\hat{\kappa}_{\eta|H_0^{\text{sym}}} - 1}\frac{\partial\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}})}{\partial\boldsymbol{\theta}'}\hat{\boldsymbol{J}}_{n|H_0^{\text{sym}}}^{-1}\frac{\partial\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0^{\text{sym}}})}{\partial\boldsymbol{\theta}} = R_n^{\text{sym}} \xrightarrow{\mathcal{L}} \chi_c^2$$

as $n \to \infty$. Hence, the critical region of the Quasi Likelihood Ratio test at the asymptotic level α is $\{L_n^{\text{sym}} > \chi_c^2(1-\alpha)\}$ □

References

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

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