Supplementary content - Conditional asymmetry in Power ARCH(∞) models

Julien Royer*[1]

[1]CREST, ENSAE, Institut Polytechnique de Paris

Appendix B GARCH(1,1)-type test when the parameter d is considered fixed

B.1 Asymptotics for the Wald statistic

Non-identification of the parameter d under H_0^{GARCH} makes the derivation of the Wald asymptotic distribution infeasible. In the main text of the paper, we proposed to estimate this distribution by bootstrap. Another solution is to consider that the unidentified parameter is known and set to a value $\overline{d} > \frac{1}{2}$. Under this assumption, the identification problem disappears and we can use Theorem 4 to obtain the asymptotic distribution of the Wald statistic.

Proposition 4. *Under the assumptions of Theorem 4, under H_0^{GARCH},*

$$W_n^{GARCH}(\overline{d}) \xrightarrow{\mathcal{L}} \frac{1}{2}\Delta_0 + \frac{1}{2}\chi_1^2,$$

where Δ_0 is the Dirac measure at 0. Thus, the critical region of asymptotic level ν is given by $\{W_n^{GARCH} > \chi_1^2(1-2\nu)\}$.

This test can easily be extended to an asymmetric volatility model with a different δ. Consider, the APARCH(∞) specification presented in (8). Testing for the adequacy of the GJR-GARCH model ($\delta = 2$) or the TGARCH ($\delta = 1$) can then be achieved by testing for $H_0 : \gamma_0 = 0$ against $H_1^{\text{APARCH}(\infty)} : \gamma_0 > 0$. We can thus define the Wald statistic in a similar manner and derive its asymptotic distribution under the null.

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

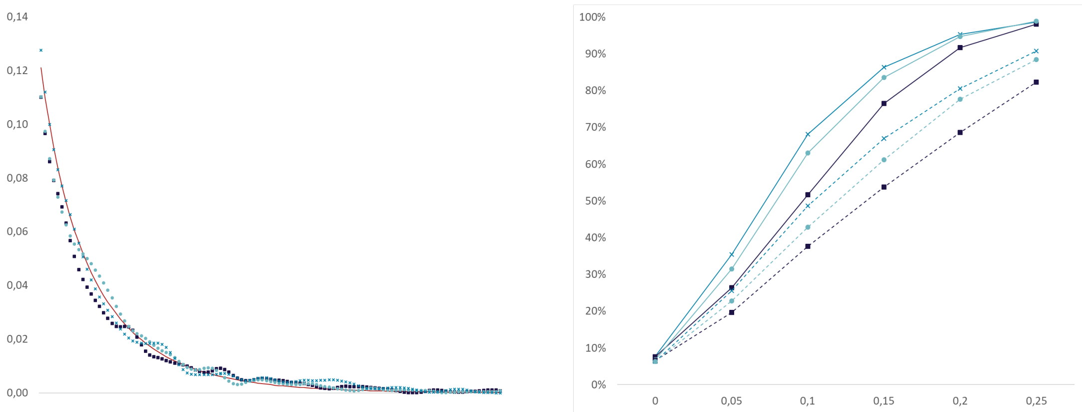
B.2 Finite sample properties of the statistic

We propose to study the empirical behavior of the test statistics defined in Section B.1. In the following simulations, we use Gaussian innovations ($\eta_t \sim \mathcal{N}(0,1)$).

We first study the empirical level of the statistic. Figure 4a presents kernel estimators for $W_n(\overline{d})$ when testing for a GARCH(1,1), a GJR-GARCH(1,1) and a TGARCH(1,1) against an APARCH(∞) model of form (8) with $\delta = 2$ and 1 respectively under H_0. The parameters used for the simulations are $\boldsymbol{\theta_0} = (0.2, 0.15, 0.75)$ for the GARCH model and $\boldsymbol{\theta_0} = (0.2, 0.05, 0.2, 0.75)$ for both the GJR-GARCH and the TGARCH. All kernels estimators are obtained from 1000 replications and are close to their theoretical asymptotic distributions. The relative rejection frequency of the test statistics, at the asymptotic levels 5%, are respectively 5.70%, 5.10% and 5.70% which is not significantly different from 5%.

We now turn to the empirical properties of the statistics under $H_1 : \gamma_0 > 0$. We first study the empirical power of the statistics under the assumption that parameters $\overline{d}$ and δ are well specified. We consider the parametric form of Model (8) with $\boldsymbol{\theta_0} = (\omega_0, \alpha_0^+, \alpha_0^-, \beta_0, \gamma_0) = (0.25, 0.05, 0.15, 0.7, \gamma_0)$ where γ_0 ranges from 0 to 0.25, $\overline{d} = 1$, and either $\delta = 2$ or $\delta = 1$ which corresponds to testing for a GJR-GARCH(1,1) or a TGARCH(1,1). In addition we consider an ARCH(∞) specified as Model (7) where $\boldsymbol{\theta_0} = (\omega_0, \alpha_0, \beta_0, \gamma_0) = (0.25, 0.1, 0.7, \gamma_0)$ where γ_0 ranges from 0 to 0.25 and $\overline{d} = 1$ which allows us to test for a GARCH(1,1). Figure 4b compares the observed powers of the three tests, that is, the relative frequency of rejection of the null hypothesis $H_0 : \gamma_0 = 0$ on the 1000 independent realizations of length $n = 2500$ and $n = 5000$, as a function of γ_0. On these simulations, we see that the three test statistics seem powerful even for low values of γ_0.

We then consider the empirical power of the test statistics when either $\overline{d}$ or δ is assumed known but is misspecified. We simulate 1000 replications of size 5000 of the parametric form of Model (8) with $\boldsymbol{\theta_0} = (\omega_0, \alpha_0^+, \alpha_0^-, \beta_0, \gamma_0, d_0) = (0.25, 0.05, 0.1, 0.7, 0.25, 1.0)$ and $\delta = 2$. We then conduct the GJR-GARCH(1,1) test assuming $\gamma = 2$ is well specified but $\overline{d}$ is misspecified. Table 4a presents the observed power of the test at different asymptotic level ν and for different values of $\overline{d}$. Even when $\overline{d}$ is far from its true value, the empirical power remains high. In addition we conduct the test assuming, this time, that $\overline{d} = 1.0$ is well specified but the power δ is misspecified. This is equivalent to testing that a short memory APARCH(1,1) with power δ is suited to model a persistent TARCH(∞). Table 4b presents the observed power of the test at different asymptotic level ν and for different values of δ. Again, even when δ is misspecified, the empirical power remains high, meaning our test statistic appears robust to misspecifications in $\overline{d}$ and δ.



(a) $\chi^2_1/2$ *density (red solid line) and kernel density estimators when testing for a GARCH (dark blue square), a GJR-GARCH (blue cross) and a TGARCH(light blue dot) on 1000 simulations with* $n = 5000$.

(b) *Observed powers of the GARCH test(dark blue square), GJR-GARCH test (light blue dot), and TGARCH test (blue cross) has a function of* γ_0, *on 1000 simulations with* $n = 2500$ *(dashed line) and* $n = 5000$ *(solid line).*

Figure 4 – *Empirical behavior of the Wald statistics when d is fixed.*

Appendix C Additional Monte Carlo experiments

We assess the finite sample properties of the QML estimator in the case where $\delta = 2$ and $\alpha_i^{+(-)}(\boldsymbol{\theta}) = \gamma^{+(-)} i^{-(d^{+(-)}+1)}$ with $\boldsymbol{\theta} = (\omega, \gamma^+, d^+, \gamma^-, d^-)$, which corresponds to an hyperbolic decay. We have simulated a thousand samples of different sizes n for $\boldsymbol{\theta_0} = (1.0, 0.40, 0.85, 0.40, 0.85)$, which generates a standard ARCH(∞) process, as well as $\boldsymbol{\theta_0} = (1.0, 0.40, 1.2, 0.20, 0.75)$, which generates a TARCH(∞) process. On each realisation, we fitted a TARCH(∞) by QML, which gave us a thousand estimators $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}$. Table 5 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.

Appendix D Additional application

D.1 Are GARCH(1,1)-type specifications suitable to model peripheral stocks?

The results in Table 2 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

		ν		
		1%	5%	10%
$\overline{d}$	0.55	70.9%	89.3%	94.6%
	1.0	91.3%	96.6%	98.3%
	1.5	92.7%	96.8%	98.4%
	2.0	92.3%	96.6%	97.9%
	2.5	91.2%	96.3%	97.6%
	5.0	89.6%	95.7%	97.3%
	7.5	89.5%	95.5%	97.3%
	10.0	89.5%	95.5%	97.3%

(a) Observed power of the GJR-GARCH(1,1) test when $d = 1.0$ and $\overline{d}$ is misspecified.

		ν		
		1%	5%	10%
δ	0.5	82.1%	91.1%	93.7%
	1.0	86.9%	94.1%	96.4%
	1.5	89.4%	95.7%	97.9%
	2.0	91.3%	96.6%	98.3%
	2.5	90.5%	97.0%	98.4%
	3.0	85.8%	95.5%	97.9%

(b) Observed power of the APARCH(1,1) test when δ is misspecified.

Table 4 – Observed power of the GARCH(1,1)-type test under $H_1 : \gamma_0 > 0$ for an asymptotic level ν of 1%, 5%, and 10%, when the model is misspecified.

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

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[4] 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[1]. 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 5. 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 2.

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

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 less peripheral. 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[8]), which would support our tests results on the above mentioned time frame.

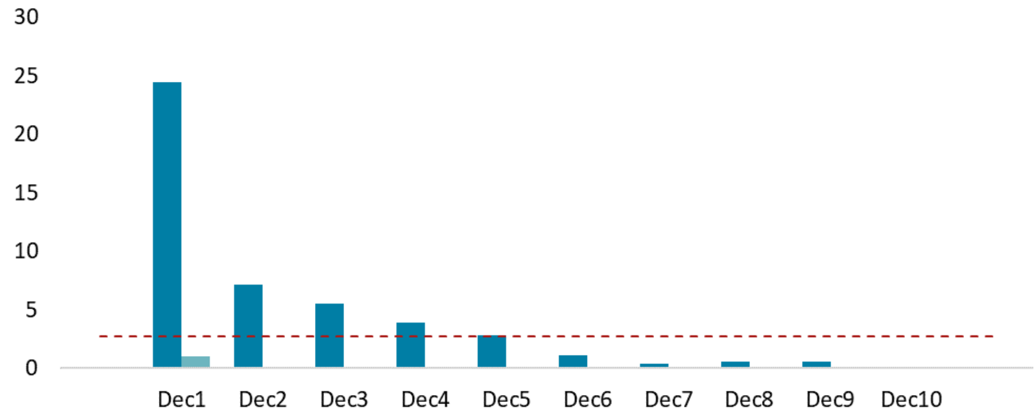


Figure 5 – Wald statistics W_n^{GJR} computed on the portfolios formed on Size from 1975 (in blue) and from 1995 (in light blue). The rejection threshold of H_0^{GJR} at the 5% asymptotic level is represented by the red dashed line.

Appendix E Detailed proofs and technical results

This appendix provides the proofs and technical results of the paper in a detailed manner. In particular, proofs of Theorem 4 and Propositions 1 and 4 were left out of the paper for the sake of brevity and are developed hereafter.

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

We develop in this section the proof of Theorem 1. The proof is based on a Volterra expansion and, in this sense, follows the work of Giraitis, Kokoszka and Leipus[6], Kazakevičius and Leipus[7], and Douc, Roueff and Soulier[3].

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

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

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

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

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

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

and thus

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

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

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

and thus we obtain the recursive equation

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

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

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

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

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

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

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

In addition, Theorem 36.4 in Billingsley[2] entails the ergodicity of the stationary solution, hence concluding the proof. □

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

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

Let us define the theoretical criterion

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

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

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

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

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

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

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

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

and using the fact that for any $s>0$, we have $x/(1+x)\leq x^s$, we obtain

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

using assumptions **A3(ii)** and (E.2). This yields

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

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

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

If $\delta\geq 2$, Jensen inequality yields $\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{\sigma_t^{2}(\boldsymbol{\theta_0})}{\sigma_t^{2}(\boldsymbol{\theta})}<\infty$, and if $\delta<2$, Minkowski inequality gives **A4**, we thus obtain

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

from assumption **A3(ii)**, which concludes the proof.

□

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

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

Proof of Lemma 1. Consider

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

and remark that $\sigma_t^2(\boldsymbol{\theta}) \geq \tilde{\sigma}_t^2(\boldsymbol{\theta})$, since we have

$$
\sigma_t^\delta(\boldsymbol{\theta}) = \tilde{\sigma}_t^\delta(\boldsymbol{\theta}) + \sum_{i=t}^{\infty} a_{i,t-i}(\boldsymbol{\phi}) |\varepsilon_{t-i}|^\delta \geq \tilde{\sigma}_t^\delta(\boldsymbol{\theta}). \tag{E.3}
$$

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

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

whence

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

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

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

we obtain

$$
\mathbb{E}\chi_t^\rho \leq K t^{-(d+1)\rho+1}.
$$

This shows that χ_t has a finite moment of order ρ and thus is finite almost surely. Furthermore, since $\rho(d+1) > 1$, the dominated convergence theorem entails $\lim_{t\to\infty} \chi_t = 0$ almost surely.

Then, we have

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

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

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

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

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

whence

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

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

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

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

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

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

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\eta_t^2\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\frac{\sigma_t^2(\boldsymbol{\theta})}{\tilde{\sigma}_t^2(\boldsymbol{\theta})} &= \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\eta_t^2\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left[\frac{\sigma_t^{\delta}(\boldsymbol{\theta})}{\tilde{\sigma}_t^{\delta}(\boldsymbol{\theta})}\right]^{2/\delta}\\
&\leq K\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\eta_t^2\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left[1+\sum_{i=t}^{\infty}i^{-d-1}|\varepsilon_{t-i}|^{\delta}\right]^{2/\delta}\\
&\leq K\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\eta_t^2\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left[1+\sum_{i=0}^{\infty}i^{-d-1}|\varepsilon_{-i}|^{\delta}\right]^{2/\delta}\\
&\leq K\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\eta_t^2\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}
\end{aligned} \tag{E.5}$$

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

Thus, we have

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

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

We can conclude

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

□

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 $\sigma_t^{\delta}(\boldsymbol{\theta}) = \sigma_t^{\delta}(\boldsymbol{\theta_0})$ 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.

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

- **(a) Asymptotic irrelevance of the initial values**
  This is directly obtained from Lemma 1.

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

$$\alpha_1^+(\boldsymbol{\phi_0})\mathbb{1}_{\varepsilon_{t-1}\geq 0} + \alpha_1^-(\boldsymbol{\phi_0})\mathbb{1}_{\varepsilon_{t-1}<0} \neq \alpha_1^+(\boldsymbol{\phi})\mathbb{1}_{\varepsilon_{t-1}\geq 0} + \alpha_1^-(\boldsymbol{\phi})\mathbb{1}_{\varepsilon_{t-1}<0}. \tag{E.6}$$

  Then we have

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

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

$$\begin{cases} \left[\alpha_1^+(\boldsymbol{\phi_0}) - \alpha_1^+(\boldsymbol{\phi})\right] \mathbb{1}_{\eta_{t-1}\geq 0} \; |\eta_{t-1}|^\delta \text{ is constant almost surely} \\ \left[\alpha_1^-(\boldsymbol{\phi_0}) - \alpha_1^-(\boldsymbol{\phi})\right] \mathbb{1}_{\eta_{t-1}<0} \; |\eta_{t-1}|^\delta \text{ is constant almost surely} \end{cases}. \tag{E.7}$$

  Since from assumption **A2** η_1 takes at least two positive (respectively negative) values, (E.7) implies almost surely $\alpha_1^+(\boldsymbol{\phi_0}) = \alpha_1^+(\boldsymbol{\phi})$ and $\alpha_1^-(\boldsymbol{\phi_0}) = \alpha_1^-(\boldsymbol{\phi})$, which contradicts (E.6). Recursively, we obtain that $\sigma_t^2(\boldsymbol{\theta}) = \sigma_t^2(\boldsymbol{\theta_0})$ implies that, for all i, $\alpha_i^+(\boldsymbol{\phi_0}) = \alpha_i^+(\boldsymbol{\phi})$ and $\alpha_i^-(\boldsymbol{\phi_0}) = \alpha_i^-(\boldsymbol{\phi})$ and thus, from assumption **A3(i)**, $\boldsymbol{\phi}^+ = \boldsymbol{\phi}_0^+$ and $\boldsymbol{\phi}^- = \boldsymbol{\phi}_0^-$ 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

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

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

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

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

In addition, we have

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

since, for any $x > 0$, $\log x \leq x - 1$.
We can conclude by noticing that $\mathbb{E}_{\boldsymbol{\theta_0}}[l_t(\boldsymbol{\theta})] = \mathbb{E}_{\boldsymbol{\theta_0}}[l_t(\boldsymbol{\theta_0})]$ if and only if $\dfrac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})} = 1$ almost surely, and thus, by identifiability of the parameter, if and only if $\boldsymbol{\theta} = \boldsymbol{\theta_0}$.

- **(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

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

The sequences $(l_t(\boldsymbol{\theta^*}))$ and $\left(\inf_{\boldsymbol{\theta^*}\in V_k(\boldsymbol{\theta})\cap\boldsymbol{\Theta}} l_t(\boldsymbol{\theta^*})\right)$ being measurable functions of ε_t and its past values, they are ergodic and strictly stationary, and admit an expectation in $\mathbb{R}\cup\{\infty\}$. Using the ergodic theorem for non-integrable processes [2], 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 $\boldsymbol{\theta_0}$,

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

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 $V(\boldsymbol{\theta})$, $\boldsymbol{\theta}\in\boldsymbol{\Theta}/V(\boldsymbol{\theta_0})$, satisfying $\liminf_{n\to\infty}\inf_{\boldsymbol{\theta^*}\in V(\boldsymbol{\theta})}\tilde{Q}_n(\boldsymbol{\theta^*}) \geq \mathbb{E}_{\boldsymbol{\theta_0}} l_1(\boldsymbol{\theta_0})$. 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, $\boldsymbol{\tilde{\theta}_n}$ 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,

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

□

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

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

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*

$$\sup_{i\in\mathcal{I}^+(\boldsymbol{\phi_0})}\sup_{\phi\in V(\boldsymbol{\phi_0})}\frac{\alpha_i^+(\boldsymbol{\phi_0})}{(\alpha_i^+)^{1-\tau}(\boldsymbol{\phi})}\leq K \text{ and } \sup_{i\in\mathcal{I}^-(\boldsymbol{\phi_0})}\sup_{\phi\in V(\boldsymbol{\phi_0})}\frac{\alpha_i^-(\boldsymbol{\phi_0})}{(\alpha_i^-)^{1-\tau}(\boldsymbol{\phi})}\leq K. \quad \text{(E.8)}$$

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

$$\begin{aligned}
&\sigma_t^\delta(\boldsymbol{\theta_0})\\
&=\omega_0+\sum_{i=1}^{\infty}a_{i,t-i}(\boldsymbol{\phi_0})|\varepsilon_{t-i}|^\delta\\
&=\omega_0\omega^{\frac{s}{k}-1}\omega^{1-\frac{s}{k}}+\sum_{i=1}^{\infty}a_{i,t-i}(\boldsymbol{\phi_0})a_{i,t-i}^{\frac{s}{k}-1}(\boldsymbol{\phi})a_{i,t-i}^{1-\frac{s}{k}}(\boldsymbol{\phi})|\varepsilon_{t-i}|^{\delta\frac{s}{k}}|\varepsilon_{t-i}|^{\delta-\delta\frac{s}{k}}\\
&=\sum_{i=0}^{\infty}x_iy_i
\end{aligned}$$

with

$$\begin{aligned}
x_0&=\omega_0\omega^{\frac{s}{k}-1} && \text{and} \quad y_0=\omega^{1-\frac{s}{k}}\\
x_i&=a_{i,t-i}(\boldsymbol{\phi_0})a_{i,t-i}^{\frac{s}{k}-1}(\boldsymbol{\phi})|\varepsilon_{t-i}|^{\delta\frac{s}{k}} && \quad\quad\; y_i=a_{i,t-i}^{1-\frac{s}{k}}(\boldsymbol{\phi})|\varepsilon_{t-i}|^{\delta-\delta\frac{s}{k}}
\end{aligned}$$

for all $i>0$. Since x_i and y_i are positive for all $i\geq 0$, from Hölder inequality, we have

$$\sigma_t^\delta(\boldsymbol{\theta_0})\leq\left(\sum_{i=0}^{\infty}x_i^{\frac{k}{s}}\right)^{\frac{s}{k}}\left(\sum_{i=0}^{\infty}y_i^{\frac{1}{1-s/k}}\right)^{1-\frac{s}{k}}.$$

Replacing x_i and y_i by their expression, we obtain

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

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

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

whence, from (E.8) and assumption **A3(ii)**, there exists a neighborhood such that

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

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

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

from assumption **A4**, and since from the arbitrariness of τ in (E.8), we can find a τ such that $(d+1)(\rho-k\tau)>1$. □

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

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

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

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

$$\frac{\partial\sigma_t^\delta}{\partial\theta_1}=\frac{\partial\sigma_t^\delta}{\partial\omega}=1\quad\text{and}\quad\frac{\partial\sigma_t^\delta}{\partial\theta_{1+j_1}}=\frac{\partial\sigma_t^\delta}{\partial\phi_{j_1}}=\sum_{i\in\mathcal{I}_t^+}\frac{\partial\alpha_i^+}{\partial\phi_{j_1}}|\varepsilon_{t-i}|^\delta+\sum_{i\in\mathcal{I}_t^-}\frac{\partial\alpha_i^-}{\partial\phi_{j_1}}|\varepsilon_{t-i}|^\delta. \quad \text{(E.9)}$$

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

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

From (E.9), and assumptions **A3(ii)** and **A10(i)** we have

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

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

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

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

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

and thus

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

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

□

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

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

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

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

$$\begin{aligned}
\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \frac{2}{\delta}\left[1 - \frac{\varepsilon_t^2}{\sigma_t^2}\right]\left[\frac{1}{\sigma_t^\delta}\frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\theta}}\right] \\
\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta'}} &= \frac{2}{\delta}\left[1 - \frac{\varepsilon_t^2}{\sigma_t^2}\right]\left[\frac{1}{\sigma_t^\delta}\frac{\partial^2 \sigma_t^\delta}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta'}}\right] + \frac{2}{\delta}\left[\frac{\delta+2}{\delta}\frac{\varepsilon_t^2}{\sigma_t^2} - 1\right]\left[\frac{1}{\sigma_t^\delta}\frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\theta}}\right]\left[\frac{1}{\sigma_t^\delta}\frac{\partial \sigma_t^\delta}{\partial \boldsymbol{\theta'}}\right].
\end{aligned} \tag{E.10}$$

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

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

From (E.1) and Lemma 2, we have that for any $j_1, j_2 \in \{1, ..., r\}$

$$\frac{\partial \sigma_t^\delta}{\partial \omega}(\boldsymbol{\theta_0}) = 1 \quad \text{and} \quad \left|\frac{1}{\sigma_t^\delta}\frac{\partial \sigma_t^\delta}{\partial \phi_{j_1}}\right|(\boldsymbol{\theta_0}) < \infty$$

which proves the first inequality, and

$$\frac{\partial^2 \sigma_t^\delta}{\partial \omega \partial \phi_{j_1}}(\boldsymbol{\theta_0}) = 0 \quad \text{and} \quad \left|\frac{1}{\sigma_t^\delta}\frac{\partial^2 \sigma_t^\delta}{\partial \phi_{j_1}\partial \phi_{j_2}}\right|(\boldsymbol{\theta_0}) < \infty$$

which proves the second inequality.

Since $\frac{1}{\sigma_t^\delta}\frac{\partial \sigma_t^\delta}{\partial \omega}$ and $\frac{1}{\sigma_t^\delta}\frac{\partial \sigma_t^\delta}{\partial \phi_{j_1}}$ are bounded at $\boldsymbol{\theta_0}$, we can conclude that

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

which finishes the proof. □

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

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

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

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

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

because $\mathbb{E}_{\boldsymbol{\theta_0}}\eta_t^2 = 1$ from assumptions **A2**.

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

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

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

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

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

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

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

$$\begin{cases}
\sum_{j=1}^{r}\lambda_j\frac{\partial\alpha_1^+(\boldsymbol{\phi_0})}{\partial\phi_j}\mathbb{1}_{\eta_{t-1}\geq 0}|\eta_{t-1}|^\delta \text{ is constant almost surely} \\
\sum_{k=1}^{r}\lambda_k\frac{\partial\alpha_1^-(\boldsymbol{\phi_0})}{\partial\phi_k}\mathbb{1}_{\eta_{t-1}<0}|\eta_{t-1}|^\delta \text{ is constant almost surely}
\end{cases} \tag{E.11}$$

However, since, from assumption **A2**, η_1 takes at least two positive (respectively negative) values, (E.11) implies almost surely

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

Iterating this argument we obtain that for all $i_h^{+(-)} = i_h^{+(-)}(\boldsymbol{\phi_0}), i_h^{+(-)} = 1, ..., r$, we have $\boldsymbol{\lambda}' \dfrac{\partial \alpha_{i_h}^{+(-)}(\boldsymbol{\phi_0})}{\partial \boldsymbol{\phi}} = 0$ and thus from assumption **A10(ii)** we must have $\boldsymbol{\lambda} = \mathbf{0}$. This implies $\lambda_0 = 0$ and contradicts the singularity of $\boldsymbol{J}$. □

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

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

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

Proof. From (E.10), we have

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

In addition, we have

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

By assumption **A11**, there exists a neighborhood $V(\boldsymbol{\theta_0})$ of $\boldsymbol{\theta_0}$ such that,

$$\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta_0})} \left[\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})} \right]^2 < \infty, \tag{E.12}$$

and the triangle inequality gives

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

by assumption **A7**.

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

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1}} \right|^p < \infty, \\
&\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial^2 \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1} \partial \theta_{i_2}} \right|^p < \infty, \\
&\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial^3 \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1} \partial \theta_{i_2} \partial \theta_{i_3}} \right|^p < \infty,
\end{aligned} \tag{E.13}$$

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

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta}\in V_\tau(\boldsymbol{\theta_0})} \left| \left[1 - \frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})} \right] \left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial^3 \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1} \partial \theta_{i_2} \partial \theta_{i_3}} \right] \right| < \infty, \\
&\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta}\in V_\tau(\boldsymbol{\theta_0})} \left| \left[\frac{\delta+2}{2} \frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})} - 1 \right] \left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1}} \right] \left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial^2 \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_2} \partial \theta_{i_3}} \right] \right| \\
&\quad \leq \left\| \sup_{\boldsymbol{\theta}\in V_\tau(\boldsymbol{\theta_0})} \left| \frac{\delta+2}{2} \frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})} - 1 \right| \right\|_2 \left\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1}} \right| \right\|_4 \left\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial^2 \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_2} \partial \theta_{i_3}} \right| \right\|_4 \\
&\quad < \infty, \\
&\text{and} \\
&\mathbb{E}_{\boldsymbol{\theta_0}} \sup_{\boldsymbol{\theta}\in V_\tau(\boldsymbol{\theta_0})} \left| \left[1 - \frac{\delta^2+3\delta+2}{\delta^2} \frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})} \right] \left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_1}} \right] \left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_2}} \right] \left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_3}} \right] \right| \\
&\quad \leq \left\| \sup_{\boldsymbol{\theta}\in V_\tau(\boldsymbol{\theta_0})} \left| 1 - \frac{\delta^2+3\delta+2}{\delta^2} \frac{\varepsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})} \right| \right\|_2 \max_h \left\| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{1}{\sigma_t^\delta(\boldsymbol{\theta})} \frac{\partial \sigma_t^\delta(\boldsymbol{\theta})}{\partial \theta_{i_h}} \right| \right\|_6^3 \\
&\quad < \infty,
\end{aligned}$$

which concludes the proof. □

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

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

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

Proof. First, remark that, from assumption **A3(ii)** and **A11**, on a neighborhood

$V(\boldsymbol{\theta_0})$ of $\boldsymbol{\theta_0}$, we have similarly to (E.5)

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

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

We have

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

therefore we can write

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

From the Markov inequality we have

$$
\begin{aligned}
&\mathbb{P}\left[\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[\frac{\partial l_t(\boldsymbol{\theta_0})}{\partial \theta_k} - \frac{\partial \tilde{l}_t(\boldsymbol{\theta_0})}{\partial \theta_k}\right]\right| > \varepsilon\right] \\
&\quad \leq \frac{1}{\varepsilon}\,\mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[\frac{\partial l_t(\boldsymbol{\theta_0})}{\partial \theta_k} - \frac{\partial \tilde{l}_t(\boldsymbol{\theta_0})}{\partial \theta_k}\right]\right| \\
&\quad \leq \frac{1}{\varepsilon}\frac{2}{\delta}\left[\mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} A_t(\boldsymbol{\theta_0})\right| + \mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} B_t(\boldsymbol{\theta_0})\right| + \mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} C_t(\boldsymbol{\theta_0})\right|\right] \qquad \text{(E.15)}
\end{aligned}
$$

First consider $\mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} A_t(\boldsymbol{\theta_0})\right|$. From (E.3) and (E.14), we have

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

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

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

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 2, we obtain

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta_0}}|A_t(\boldsymbol{\theta_0})|^\rho &\leq K\left(\mathbb{E}_{\boldsymbol{\theta_0}}\left[\sum_{i=t}^{\infty} i^{-(d^*+1)}|\varepsilon_{t-i}|^\delta\right]^{\rho(1+\xi)}\right)^{\frac{1}{1+\xi}}\left(\mathbb{E}_{\boldsymbol{\theta_0}}\left|\frac{1}{\sigma_t^\delta}\frac{\partial\sigma_t^\delta}{\partial\theta_k}(\boldsymbol{\theta_0})\right|^{\rho\frac{\xi+1}{\xi}}\right)^{\frac{\xi}{1+\xi}} \\
&\leq K\left(\sum_{i=t}^{\infty} i^{-(d^*+1)\rho(1+\xi)}\mathbb{E}_{\boldsymbol{\theta_0}}|\varepsilon_{t-i}|^{\delta\rho(1+\xi)}\right)^{\frac{1}{1+\xi}} \\
&\leq K\sum_{i=t}^{\infty} i^{-(d^*+1)\rho}\left(\mathbb{E}_{\boldsymbol{\theta_0}}|\varepsilon_{t-i}|^{\delta\rho(1+\xi)}\right)^{\frac{1}{1+\xi}} \\
&\leq K\sum_{i=0}^{\infty}(t+i)^{-(d^*+1)\rho}\left(\mathbb{E}_{\boldsymbol{\theta_0}}|\varepsilon_{-i}|^{\delta\rho(1+\xi)}\right)^{\frac{1}{1+\xi}} \\
&\leq Kt^{-(d^*+1)\rho+1},
\end{aligned} \tag{E.16}$$

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 assumption **A9** we have $(d^* + \frac{3}{2})\rho - 2 > 0$. Using Markov inequality, we can conclude $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}|A_t(\boldsymbol{\theta_0})|$ tends to 0 in probability.

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 (E.14) and similarly to (E.5),

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

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 (E.9), and from assumptions **A10(i)** and **A3(ii)**, we have for all $\xi > 0$,

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

and thus

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

since, from assumption **A8** and **A9**, there exists a ξ such that $(d^*+1)\rho(1-\xi) - 2 > 0$. Using Markov inequality, we can conclude $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} |C_t(\boldsymbol{\theta_0})|$ tends to 0 in probability.

Hence (E.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 $\chi_t=\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta_0})}|\sigma_t^\delta(\boldsymbol{\theta})-\tilde{\sigma}_t^\delta(\boldsymbol{\theta})|$, and we have from (E.3) and assumption **A3(ii)**

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

whence

$$\begin{aligned}\mathbb{E}\chi_t^\rho &\leq K\sum_{i=0}^{\infty}(i+t)^{-(d+1)\rho}\ \mathbb{E}|\varepsilon_{-i}|^{\delta\rho}\\ &\leq Kt^{-(d+1)\rho+1}\end{aligned}\tag{E.17}$$

since, from assumption **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 almost surely. Furthermore, since $\rho(d+1)>1$, the dominated convergence theorem entails $\lim_{t\to\infty}\chi_t=0$ almost surely.

Let us now denote

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

and

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

where $V(\boldsymbol{\theta_0})$ is a neighborhood of $\boldsymbol{\theta_0}$ and $i_1,i_2\in\{1,\dots,r\}$. From (E.9) we easily obtain $\chi_t^{(1)}=0$, and from assumptions **A10(i)** and **A3(ii)**, we have for all $\xi>0$,

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

whence

$$\mathbb{E}\left(\chi_t^{(i_1)}\right)^\rho \begin{array}{l} \leq \sum_{i=0}^{\infty} (i+t)^{-(d+1)\rho(1-\xi)} \mathbb{E}|\varepsilon_{-i}|^{\delta\rho} \\ \leq Kt^{-(d+1)\rho(1-\xi)+1} \end{array} \tag{E.18}$$

since, from assumption **A4**, $\mathbb{E}\,|\varepsilon_t|^{\delta\rho} < \infty$. 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} \frac{2}{\delta} \sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})} \left| \left[\frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} - \frac{\varepsilon_t^2}{\sigma_t^2} \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial^2 \sigma_t^\delta}{\partial\theta_{i_1}\partial\theta_{i_2}} \right] \right. \\
&+ \left[1 - \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} \right] \left[\left(\frac{1}{\sigma_t^\delta} - \frac{1}{\tilde{\sigma}_t^\delta} \right) \frac{\partial^2 \sigma_t^\delta}{\partial\theta_{i_1}\partial\theta_{i_2}} + \frac{1}{\tilde{\sigma}_t^\delta} \left(\frac{\partial^2 \sigma_t^\delta}{\partial\theta_{i_1}\partial\theta_{i_2}} - \frac{\partial^2 \tilde{\sigma}_t^\delta}{\partial\theta_{i_1}\partial\theta_{i_2}} \right) \right] \\
&+ \left[\frac{2+\delta}{\delta} \frac{\varepsilon_t^2}{\sigma_t^2} - \frac{2+\delta}{\delta} \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial\theta_{i_1}} \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial\theta_{i_2}} \right] \\
&+ \left[\frac{2+\delta}{\delta} \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} - 1 \right] \left[\left(\frac{1}{\sigma_t^\delta} - \frac{1}{\tilde{\sigma}_t^\delta} \right) \frac{\partial \sigma_t^\delta}{\partial\theta_{i_1}} + \frac{1}{\tilde{\sigma}_t^\delta} \left(\frac{\partial \sigma_t^\delta}{\partial\theta_{i_1}} - \frac{\partial \tilde{\sigma}_t^\delta}{\partial\theta_{i_1}} \right) \right] \left[\frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial\theta_{i_2}} \right] \\
&\left. + \left[\frac{2+\delta}{\delta} \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} - 1 \right] \left[\left(\frac{1}{\sigma_t^\delta} - \frac{1}{\tilde{\sigma}_t^\delta} \right) \frac{\partial \sigma_t^\delta}{\partial\theta_{i_2}} + \frac{1}{\tilde{\sigma}_t^\delta} \left(\frac{\partial \sigma_t^\delta}{\partial\theta_{i_2}} - \frac{\partial \tilde{\sigma}_t^\delta}{\partial\theta_{i_2}} \right) \right] \left[\frac{1}{\tilde{\sigma}_t^\delta} \frac{\partial \tilde{\sigma}_t^\delta}{\partial\theta_{i_1}} \right] \right| (\boldsymbol{\theta}),
\end{aligned}$$

which yields

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

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

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta_0}}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{\tilde{\sigma}_t^\delta(\boldsymbol{\theta})}\frac{\partial\tilde{\sigma}_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right|^p<\infty \\
&\mathbb{E}_{\boldsymbol{\theta_0}}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{\tilde{\sigma}_t^\delta(\boldsymbol{\theta})}\frac{\partial^2\tilde{\sigma}_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}\partial\theta_{i_2}}\right|^p<\infty.
\end{aligned} \tag{E.19}$$

Then, from independence of η_t^2 with σ_t^δ and its derivatives, assumption **A11**, Lemma 2, (E.14), and (E.19) we have, using Hölder inequality, for all i_1,i_2,

$$\begin{aligned}
&\mathbb{E}\left[\eta_t^2\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\left[\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})}\frac{\partial^2\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}\partial\theta_{i_2}}\right]\right|\right]<\infty \\
&\mathbb{E}\left[\eta_t^2\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})}\frac{\partial\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right]\left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})}\frac{\partial\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_2}}\right]\right|\right]<\infty \\
&\mathbb{E}\left[\eta_t^2\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left[\frac{1}{\tilde{\sigma}_t^\delta(\boldsymbol{\theta})}\frac{\partial\tilde{\sigma}_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right]\left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})}\frac{\partial\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_2}}\right]\right|\right]<\infty \\
&\mathbb{E}\left[\eta_t^2\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left[\frac{1}{\sigma_t^\delta(\boldsymbol{\theta})}\frac{\partial\sigma_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right]\right|\right]<\infty \\
&\mathbb{E}\left[\eta_t^2\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}\left|\frac{\sigma_t^2(\boldsymbol{\theta_0})}{\sigma_t^2(\boldsymbol{\theta})}\left[\frac{1}{\tilde{\sigma}_t^\delta(\boldsymbol{\theta})}\frac{\partial\tilde{\sigma}_t^\delta(\boldsymbol{\theta})}{\partial\theta_{i_1}}\right]\right|\right]<\infty.
\end{aligned} \tag{E.20}$$

Since χ_t, $\chi_t^{(i_1)}$, and $\chi_t^{(i_1,i_2)}$ tend to 0 almost surely as t tends to infinity, and (E.20), Toeplitz lemma combined with Markov inequality entail

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

in probability, which concludes the proof. □

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

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

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

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

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

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

and we have from Lemma 4 that $\mathbb{V}_{\boldsymbol{\theta_0}}\left[\dfrac{\partial l_t}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta_0})\right]$ is finite. In view of the invertibility of $\boldsymbol{J}$ and the assumptions on the distribution of η_t (which entails $0<\kappa_\eta-1<\infty$), this covariance matrix is non-degenerate. It follows that, for all $\boldsymbol{\lambda}\in\mathbb{R}^{r+1}$, the sequence $\left\{\boldsymbol{\lambda}'\dfrac{\partial l_t(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}},\varepsilon_t\right\}_t$ is a square integrable ergodic stationary martingale difference. The Cramer-Wold theorem and the following corollary entail

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

*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

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

The ergodic theorem entails $\boldsymbol{J}_n\to\boldsymbol{J}$ as $n\to\infty$ almost surely. Finally, the conclusion follows from Slutsky lemma.

□

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

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

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

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \frac{\partial \tilde{l}_t}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta_0}) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, (\kappa_\eta - 1)\boldsymbol{J}) \tag{E.21}$$

and that

$$\frac{1}{n}\sum_{t=1}^{n} \frac{\partial^2 \tilde{l}_t}{\partial \theta_i \partial \theta_j}(\boldsymbol{\theta}^*_{\boldsymbol{ij}}) \to J(i,j) \text{ in probability.} \tag{E.22}$$

Using lemmas 3, 4, 6, and 7 along with Slutsky lemma directly yields (E.21).

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

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

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

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

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

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

and by the ergodic theorem,

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

Using Slutsky lemma along with the previous lemmas allows us to obtain (E.22) which ends the proof. □

We now turn to the proof of Theorem 4. This proof is very similar to the one established by Francq and Zakoïan[5] in the GARCH(p,q) case when some coefficients are equal to zeros. In the following, we use the notation $a \overset{o_P(1)}{=} b$ meaning $a = b + o_P(1)$.

Proof of Theorem 4. From the proof of Theorem 3, we have

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta_0}) \overset{o_P(1)}{=} \boldsymbol{Z}_n = -\boldsymbol{J}_n^{-1}\sqrt{n}\frac{\partial Q_n(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}}. \tag{E.23}$$

when $\boldsymbol{\theta_0}$ belongs to the interior of Θ. This relation does not hold when $\boldsymbol{\theta_0} \in \partial\boldsymbol{\Theta}$ since then at least one element of the vector $(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta_0})$ is a positive random variable. We will show that, in the general case, for all $\boldsymbol{\theta_0} \in \boldsymbol{\Theta}$, we have

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta_0}) \overset{o_P(1)}{=} \boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}} = \arg\inf_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}}[\boldsymbol{\lambda} - \boldsymbol{Z}_n]'\boldsymbol{J}_n[\boldsymbol{\lambda} - \boldsymbol{Z}_n]. \tag{E.24}$$

Note than when $\boldsymbol{\theta_0}$ belongs to the interior of $\boldsymbol{\Theta}$, we have $\boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}} = \boldsymbol{Z}_n$ and (E.24) reduces to (E.23). $\boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}}$ can be seen as the orthogonal projection of $\boldsymbol{Z}_n$ on Λ for the inner product $\langle \boldsymbol{x}, \boldsymbol{y}\rangle_{\boldsymbol{J}_n} = \boldsymbol{x}'\boldsymbol{J}_n\boldsymbol{y}$. This projection can be approximated by

$$\sqrt{n}(\boldsymbol{\theta}_{\boldsymbol{J}_n}(\boldsymbol{Z}_n) - \boldsymbol{\theta_0}) \quad \text{with} \quad \boldsymbol{\theta}_{\boldsymbol{J}_n}(\boldsymbol{Z}_n) = \arg\inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\|\boldsymbol{Z}_n - \sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta_0})\|_{\boldsymbol{J}_n}$$

which is the the projection of $\boldsymbol{Z}_n$ on the space $\sqrt{n}(\boldsymbol{\Theta} - \boldsymbol{\theta_0})$ which increases to $\boldsymbol{\Lambda}$. Using a Taylor expansion for a function with right partial derivatives, we have for all $\boldsymbol{\theta}$ and $\boldsymbol{\theta_0} \in \boldsymbol{\Theta}$,

$$\begin{aligned}
\tilde{Q}_n(\boldsymbol{\theta}) &= \tilde{Q}_n(\boldsymbol{\theta_0}) + \frac{\partial\tilde{Q}_n(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta'}}(\boldsymbol{\theta} - \boldsymbol{\theta_0}) + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta_0})'\frac{\partial^2\tilde{Q}_n(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta'}}(\boldsymbol{\theta} - \boldsymbol{\theta_0}) + R_n(\boldsymbol{\theta}) \\
&= \tilde{Q}_n(\boldsymbol{\theta_0}) + \frac{1}{2n}\boldsymbol{Z}_n'\boldsymbol{J}_n\sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta_0}) - \frac{1}{2n}\sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta_0})'\boldsymbol{J}_n\boldsymbol{Z}_n \\
&\quad + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta_0})'\boldsymbol{J}_n(\boldsymbol{\theta} - \boldsymbol{\theta_0}) + R_n(\boldsymbol{\theta}) + R_n^*(\boldsymbol{\theta}) \\
&= \tilde{Q}_n(\boldsymbol{\theta_0}) + \frac{1}{2n}\|\boldsymbol{Z}_n - \sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta_0})\|^2_{\boldsymbol{J}_n} - \frac{1}{2n}\boldsymbol{Z}_n'\boldsymbol{J}_n\boldsymbol{Z}_n + R_n(\boldsymbol{\theta}) + R_n^*(\boldsymbol{\theta})
\end{aligned} \tag{E.25}$$

where $R_n(\boldsymbol{\theta})$ and $R_n^*(\boldsymbol{\theta})$ are remainder terms. To conclude the proof, we will prove the following intermediate results. For all $\boldsymbol{\theta_0} \in \boldsymbol{\Theta}$

(a) $\sqrt{n}(\boldsymbol{\theta}_{\boldsymbol{J}_n}(\boldsymbol{Z}_n) - \boldsymbol{\theta_0}) = O_P(1)$

(b) $\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}) = O_P(1)$

(c) For any sequence $(\boldsymbol{\theta}_n)$ such that $\sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta_0}) = O_P(1)$, $R_n(\boldsymbol{\theta}_n) = o_P(n^{-1})$ and $R_n^*(\boldsymbol{\theta}_n) = o_P(n^{-1})$

(d) $\|\boldsymbol{Z}_n - \sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0})\|^2_{\boldsymbol{J}_n} \overset{o_P(1)}{=} \|\boldsymbol{Z}_n - \boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}}\|^2_{\boldsymbol{J}_n}$

(e) $\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}) \overset{o_P(1)}{=} \boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}}$

(f) $\boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}} \overset{\mathcal{L}}{\to} \boldsymbol{\lambda}^{\boldsymbol{\Lambda}}$

Lemma 4 ensures that for n sufficiently large, $\|.\|_{\boldsymbol{J}_n}$ almost surely defines a norm. Using the triangular inequality and the fact that $\boldsymbol{\theta}_{\boldsymbol{J}_n}(\boldsymbol{Z}_n)$ minimizes $\|\boldsymbol{Z}_n - \sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta_0})\|_{\boldsymbol{J}_n}$ over $\boldsymbol{\Theta}$, we have

$$\|\sqrt{n}(\boldsymbol{\theta}_{\boldsymbol{J}_n}(\boldsymbol{Z}_n) - \boldsymbol{\theta_0})\|_{\boldsymbol{J}_n} \le \|\boldsymbol{Z}_n - \sqrt{n}(\boldsymbol{\theta}_{\boldsymbol{J}_n}(\boldsymbol{Z}_n) - \boldsymbol{\theta_0})\|_{\boldsymbol{J}_n} + \|\boldsymbol{Z}_n\|_{\boldsymbol{J}_n} \le \|\boldsymbol{Z}_n\|_{\boldsymbol{J}_n} + \|\boldsymbol{Z}_n\|_{\boldsymbol{J}_n}$$

and from Lemma 7, we have $\|\boldsymbol{Z}_n\|_{\boldsymbol{J}_n} + \|\boldsymbol{Z}_n\|_{\boldsymbol{J}_n} = O_P(1)$ which concludes the proof of (a).

By the Taylor expansion

$$\tilde{Q}_n(\boldsymbol{\theta}) = \tilde{Q}_n(\boldsymbol{\theta_0}) + \frac{\partial \tilde{Q}_n(\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta} - \boldsymbol{\theta_0}) + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta_0})' \frac{\partial^2 \tilde{Q}_n(\boldsymbol{\theta}^*_{\boldsymbol{ij}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}(\boldsymbol{\theta} - \boldsymbol{\theta_0})$$

with $\boldsymbol{\theta}^*_{\boldsymbol{ij}}$ between $\boldsymbol{\theta_0}$ and $\boldsymbol{\theta}$, we have

$$R_n(\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta_0})' \left[\frac{\partial^2 \tilde{Q}_n(\boldsymbol{\theta}^*_{\boldsymbol{ij}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \frac{\partial^2 \tilde{Q}_n(\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] (\boldsymbol{\theta} - \boldsymbol{\theta_0}). \tag{E.26}$$

Theorem 2, and Lemmas 5 and 6 ensure $\left[\frac{\partial^2 \tilde{Q}_n(\boldsymbol{\theta}^*_{\boldsymbol{ij}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \frac{\partial^2 \tilde{Q}_n(\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] \to 0$ as n tends to infinity when $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}$, and thus $R_n(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}) = o_P(\|\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta_0}\|^2_{\boldsymbol{J}_n})$.

In addition,

$$R_n^*(\boldsymbol{\theta}) = \left[\frac{\partial \tilde{Q}_n(\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta}} - \frac{\partial Q_n(\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta}} \right] (\boldsymbol{\theta} - \boldsymbol{\theta_0}) + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta_0})' \left[\frac{\partial^2 \tilde{Q}_n(\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \boldsymbol{J}_n \right] (\boldsymbol{\theta} - \boldsymbol{\theta_0}). \tag{E.27}$$

and from Lemma 6, we have $R_n^*(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}) = o_P(n^{-1/2}\|\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}_{\mathbf{0}}\|_{\boldsymbol{J}_n}) + o_P(\|\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}_{\mathbf{0}}\|^2_{\boldsymbol{J}_n})$.

Since $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}$ minimizes $\tilde{Q}_n$ over $\boldsymbol{\Theta}$, equation (E.25) yields

$$\begin{aligned}\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}) - \tilde{Q}_n(\boldsymbol{\theta}_{\mathbf{0}}) &= \frac{1}{2n}\Big[\|\boldsymbol{Z}_n - \sqrt{n}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}_{\mathbf{0}})\|^2_{\boldsymbol{J}_n} - \|\boldsymbol{Z}_n\|^2_{\boldsymbol{J}_n} \\ &\quad + o_P(\|n^{1/2}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}_{\mathbf{0}})\|_{\boldsymbol{J}_n}) + o_P(\|n^{1/2}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}_{\mathbf{0}})\|^2_{\boldsymbol{J}_n})\Big] \leq 0\end{aligned}$$

and thus

$$\|\boldsymbol{Z}_n - \sqrt{n}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}_{\mathbf{0}})\|^2_{\boldsymbol{J}_n} \leq \Big[\|\boldsymbol{Z}_n\|_{\boldsymbol{J}_n} + o_P(\|n^{1/2}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}_{\mathbf{0}})\Big]^2.$$

By the triangular inequality, we obtain

$$\begin{aligned}\|\sqrt{n}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}_{\mathbf{0}})\|_{\boldsymbol{J}_n} &\leq \|\sqrt{n}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}_{\mathbf{0}}) - \boldsymbol{Z}_n\|_{\boldsymbol{J}_n} + \|\boldsymbol{Z}_n\|_{\boldsymbol{J}_n} \\ &\leq 2\|\boldsymbol{Z}_n\|_{\boldsymbol{J}_n} + o_P(\|n^{1/2}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}_{\mathbf{0}}\|_{\boldsymbol{J}_n})\end{aligned}$$

whence $\|\sqrt{n}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}_{\mathbf{0}})\|_{\boldsymbol{J}_n}[1 + o_P(1)] \leq 2\|\boldsymbol{Z}_n\|_{\boldsymbol{J}_n} = O_P(1)$, which proves (b).

From lemmas 5 and 6, equation (E.26) entail $R_n(\boldsymbol{\theta}_{\boldsymbol{n}}) = o_P(\|\boldsymbol{\theta}_{\boldsymbol{n}} - \boldsymbol{\theta}_{\mathbf{0}}\|^2) = o_P(n^{-1})$, which proves the first part of (c), while equation (E.27) similarly yields $R_n^*(\boldsymbol{\theta}_{\boldsymbol{n}}) = o_P(n^{-1/2}\|\boldsymbol{\theta}_{\boldsymbol{n}} - \boldsymbol{\theta}_{\mathbf{0}}\|_{\boldsymbol{J}_n}) + o_P(\|\boldsymbol{\theta}_{\boldsymbol{n}} - \boldsymbol{\theta}_{\mathbf{0}}\|^2_{\boldsymbol{J}_n}) = o_P(n^{-1})$ which concludes the proof of (c).

By (a)-(c) and (E.25), we have

$$\begin{aligned}0 &\leq \|\boldsymbol{Z}_n - \sqrt{n}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}_{\mathbf{0}})\|^2_{\boldsymbol{J}_n} - \|\boldsymbol{Z}_n - \sqrt{n}(\boldsymbol{\theta}_{\boldsymbol{J}_n}(\boldsymbol{Z}_n) - \boldsymbol{\theta}_{\mathbf{0}})\|^2_{\boldsymbol{J}_n} \\ &= 2n[\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}) - \tilde{Q}_n(\boldsymbol{\theta}_{\boldsymbol{J}_n}(\boldsymbol{Z}_n))] - 2n[(R_n + R_n^*)(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}) - (R_n + R_n^*)(\boldsymbol{\theta}_{\boldsymbol{J}_n})] \\ &\leq -2n[(R_n + R_n^*)(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}) - (R_n + R_n^*)(\boldsymbol{\theta}_{\boldsymbol{J}_n})] = o_P(1)\end{aligned}$$

since $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}$ minimizes $\tilde{Q}_n$ and $\boldsymbol{\theta}_{\boldsymbol{J}_n}$ minimizes $\|\boldsymbol{Z}_n - \sqrt{n}(\boldsymbol{\theta}_{\boldsymbol{J}_n}(\boldsymbol{Z}_n) - \boldsymbol{\theta}_{\mathbf{0}})\|_{\boldsymbol{J}_n}$. Since for n sufficiently large, we have $\sqrt{n}(\boldsymbol{\theta}_{\boldsymbol{J}_n} - \boldsymbol{\theta}_{\mathbf{0}}) = \boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}}$, (d) holds.

The vector $\boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}}$ being the projection of $\boldsymbol{Z}_n$ on the convex set Λ for the scalar product $\langle \boldsymbol{x}, \boldsymbol{y}\rangle_{\boldsymbol{J}_n}$, it is characterized by $\boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}} \in \Lambda$, $\langle \boldsymbol{Z}_n \boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}}, \boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}} - \boldsymbol{\lambda}\rangle$ for all $\lambda \in \Lambda$ (see for example Lemma 1.1 in Zarantonello[9]). Thus we obtain

$$\begin{aligned}\|\sqrt{n}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}_{\mathbf{0}}) - \boldsymbol{Z}_n\|^2_{\boldsymbol{J}_n} &= \|\sqrt{n}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}_{\mathbf{0}}) - \boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}}\|^2_{\boldsymbol{J}_n} + \|\boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}} - \boldsymbol{Z}_n\|^2_{\boldsymbol{J}_n} \\ &\quad + 2\langle\sqrt{n}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}_{\mathbf{0}}) - \boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}}, \boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}} - \boldsymbol{Z}_n\rangle_{\boldsymbol{J}_n} \\ &\geq \|\sqrt{n}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}_{\mathbf{0}}) - \boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}}\|^2_{\boldsymbol{J}_n} + \|\boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}} - \boldsymbol{Z}_n\|^2_{\boldsymbol{J}_n}\end{aligned}$$

whence, by (d),

$$\|\sqrt{n}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}_{\mathbf{0}}) - \boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}}\|^2_{\boldsymbol{J}_n} \leq \|\boldsymbol{Z}_n - \sqrt{n}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}_{\mathbf{0}})\|^2_{\boldsymbol{J}_n} - \|\boldsymbol{Z}_n - \boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}}\|^2_{\boldsymbol{J}_n} = o_P(1)$$

which proves (e).

Finally, Lemma 7 entails $(\boldsymbol{J}_n, \boldsymbol{Z}_n) \to (\boldsymbol{J}, \boldsymbol{Z})$. In addition, $\boldsymbol{\lambda}_n^{\boldsymbol{\Lambda}} = f(\boldsymbol{J}_n, \boldsymbol{Z}_n)$ and $\boldsymbol{\lambda}^{\boldsymbol{\Lambda}} = f(\boldsymbol{J}, \boldsymbol{Z})$ where f is a continuous function, except on the set of the points $(\boldsymbol{J}_n, \boldsymbol{Z}_n)$ such that $\boldsymbol{J}_n$ is singular, which is a set of $P_{(\boldsymbol{J},\boldsymbol{Z})}$-probability zero. Thus, the continuous mapping theorem entails (f).

The proof of Theorem 4 directly follows from (e) and (f). □

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

$$\frac{\sigma_{\delta,t}(\boldsymbol{\theta})}{\sigma_{\delta_0,t}(\boldsymbol{\theta_0})} = 1 \text{ almost surely} \Rightarrow \delta = \delta_0. \tag{E.28}$$

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

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

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

$$\left[\frac{\sigma_{\delta,t}(\boldsymbol{\theta})}{\sigma_{\delta_0,t}(\boldsymbol{\theta_0})}\right]^{\delta_0} = g_{\Psi^+}(|\eta_{t-1}|^{\delta_0}) \mathbb{1}_{\eta_{t-1} \geq 0} + g_{\Psi^-}(|\eta_{t-1}|^{\delta_0}) \mathbb{1}_{\eta_{t-1} < 0} \tag{E.29}$$

where $\Psi^{+(-)} = (\omega_{\delta_0,t-2}(\boldsymbol{\theta_0}), \omega_{\delta,t-2}(\boldsymbol{\theta}), \delta/\delta_0, \alpha_1^{+(-)}(\boldsymbol{\phi_0}) \sigma_{\delta_0,t-1}^{\delta_0}(\boldsymbol{\theta_0}), \alpha_1^{+(-)}(\boldsymbol{\phi}) \sigma_{\delta_0,t-1}^{\delta}(\boldsymbol{\theta_0}))$. Thus $\sigma_{\delta,t}(\boldsymbol{\theta}) = \sigma_{\delta_0,t}(\boldsymbol{\theta_0})$ implies i) $\delta = \delta_0$ or ii) $\alpha_1^+(\boldsymbol{\phi}) = \alpha_1^+(\boldsymbol{\phi_0}) = 0$ and $\alpha_1^-(\boldsymbol{\phi}) = \alpha_1^-(\boldsymbol{\phi_0}) = 0$. In the latter case, (E.29) holds by replacing η_{t-1} by η_{t-2}. Iterating the arguments, under A2', the first equality in (E.28) entails either i) $\delta = \delta_0$ or ii) $\alpha_i^+(\boldsymbol{\phi}) = \alpha_i^+(\boldsymbol{\phi_0}) = 0$ and $\alpha_i^-(\boldsymbol{\phi}) = \alpha_i^-(\boldsymbol{\phi_0}) = 0$ for all $i \geq 1$. The latter is precluded by Assumption A3(i), thus we have shown (E.28), which concludes the proof using Theorem 2 and arguments of its proof. □

E.3 Specification tests

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

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

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

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

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

$$\sqrt{n}\|\boldsymbol{r}_m(\boldsymbol{\theta_0}) - \tilde{\boldsymbol{r}}_m(\boldsymbol{\theta_0})\| = o_P(1) \text{ and } \sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})} \left\| \frac{\partial \boldsymbol{r}_m(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial \tilde{\boldsymbol{r}}_m(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| = o_P(1). \quad \text{(E.30)}$$

We have

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

Similarly to (E.16), we have

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\theta_0}} |A_t(\boldsymbol{\theta_0})|^\rho &\le K\mathbb{E}_{\boldsymbol{\theta_0}} \left| \frac{\sigma_t^2(\boldsymbol{\theta_0}) - \tilde{\sigma}_t^2(\boldsymbol{\theta_0})}{\tilde{\sigma}_t^2(\boldsymbol{\theta_0})} \right|^\rho \\ &\le K t^{-(d^*+1)\rho+1} \end{aligned}$$

and thus

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

since from assumption **A9** we have $(d^* + \frac{3}{2})\rho - 2 > 0$. Using Markov inequality, we can conclude $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} |A_t(\boldsymbol{\theta_0})|$ tends to 0 in probability. Similar arguments yield that $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} |B_t(\boldsymbol{\theta_0})|$ tends to 0 in probability, which proves the first part of (E.30).

In addition, we have

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

whence, for all $k \in \{1, \ldots, r+1\}$, using similar notations as in (E.17) and (E.18),

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

Then similarly to (E.20), from independence of η_t^2 with σ_t^δ and its derivatives, assumption **A11**, Lemma 2, (E.14), and (E.19) we have, using Hölder inequality, and Toeplitz lemma,

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

since χ_t and $\chi_t^{(k)}$ tend to 0 almost surely as t tends to infinity.

In a like manner, we obtain that

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

Using Markov inequality, we thus obtain that $n^{-1}\sum_{t=1}^n\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}|\partial A_t(\boldsymbol{\theta})/\partial\theta_k|\to 0$ in probability as n tends to infinity. Similar arguments yield the convergence of the term $n^{-1}\sum_{t=1}^n\sup_{\boldsymbol{\theta}\in\mathcal{V}(\boldsymbol{\theta_0})}|\partial B_t(\boldsymbol{\theta})/\partial\theta_k|$ and thus we have shown the second part of (E.30).

Using a Taylor expansion of $\tilde{r}_h$ at $\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}$ for $h=1,\dots,m$ along with (E.30) yields

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

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

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

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

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

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

$$\sqrt{n}\tilde{\boldsymbol{r}}_m(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}})\overset{o_P(1)}{=}\sqrt{n}\boldsymbol{r}_m(\boldsymbol{\theta_0})+\boldsymbol{C}_m\sqrt{n}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}-\boldsymbol{\theta_0}).\tag{E.31}$$

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

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

thus the central limit theorem applied to the martingale difference

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

shows that

$$\begin{aligned}\sqrt{n}\begin{pmatrix}\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}}-\boldsymbol{\theta_0}\\ \boldsymbol{r}_m(\boldsymbol{\theta_0})\end{pmatrix} &\overset{o_P(1)}{=} \frac{1}{\sqrt{n}}\sum_{t=1}^n s_t\begin{pmatrix}\boldsymbol{J}^{-1}\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}}\\ \boldsymbol{s}_{t-1:t-m}\end{pmatrix}\\ &\overset{\mathcal{L}}{\to}\mathcal{N}\left(\boldsymbol{0},\begin{bmatrix}(\kappa_\eta-1)\boldsymbol{J}^{-1} & (\kappa_\eta-1)\boldsymbol{J}^{-1}\boldsymbol{C}'_m\\ (\kappa_\eta-1)\boldsymbol{C}_m\boldsymbol{J}^{-1} & (\kappa_\eta-1)^2\boldsymbol{I}_m\end{bmatrix}\right).\end{aligned} \tag{E.32}$$

From (E.31) and (E.32), we obtain

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

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

$$(\kappa_\eta-1)\boldsymbol{I}_m-\boldsymbol{C}_m\boldsymbol{J}^{-1}\boldsymbol{C}'_m=\mathbb{E}_{\boldsymbol{\theta_0}}\boldsymbol{V}\boldsymbol{V}',\text{ with }\boldsymbol{V}=\boldsymbol{s}_{-1:-m}+\boldsymbol{C}_m\boldsymbol{J}^{-1}\frac{2}{\delta}\frac{1}{\sigma_0^2}\frac{\partial\sigma_0^2(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}}.$$

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

$$\boldsymbol{\lambda}'\boldsymbol{s}_{-1:-m}+\boldsymbol{\mu}'\frac{1}{\sigma_0^2}\frac{\partial\sigma_0^2(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}}\text{ a.s., where }\boldsymbol{\mu}=\frac{2}{\delta}\boldsymbol{\lambda}'\boldsymbol{C}_m\boldsymbol{J}^{-1}. \tag{E.33}$$

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

$$\begin{aligned}\boldsymbol{\mu}'\frac{\partial\sigma_0^2(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}} &= \mu_1+\sum_{i=1}^{\infty}\left[\sum_{j=2}^{r+1}\mu_j\frac{\partial\alpha_i^+(\boldsymbol{\phi_0})}{\partial\phi_j}\mathbb{1}_{\eta_{-i}\ge 0}+\sum_{k=2}^{r+1}\mu_k\frac{\partial\alpha_i^-(\boldsymbol{\phi_0})}{\partial\phi_k}\mathbb{1}_{\eta_{-i}<0}\right]\sigma_{-i}^\delta(\boldsymbol{\theta_0})|\eta_{-i}|^\delta\\ &= \mu_1+\left[\sum_{j=2}^{r+1}\mu_j\frac{\partial\alpha_1^+(\boldsymbol{\phi_0})}{\partial\phi_j}|\eta_{-1}^+|^\delta+\sum_{k=2}^{r+1}\mu_k\frac{\partial\alpha_1^-(\boldsymbol{\phi_0})}{\partial\phi_k}|\eta_{-1}^-|^\delta\right]\sigma_{-1}^\delta(\boldsymbol{\theta_0})+R_{-2}\end{aligned}$$

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

$$\begin{aligned}\sigma_0^\delta(\boldsymbol{\theta_0})\boldsymbol{\lambda}'\boldsymbol{s}_{-1:-m} &= (\omega_0+a_{1,t-1}(\boldsymbol{\phi_0})|\eta_{-1}|^\delta\sigma_{-1}^\delta(\boldsymbol{\theta_0})+R_{-2})(\lambda_1\eta_{-1}^2+R_{-2})\\ &= \lambda_1\sigma_{-1}^\delta(\boldsymbol{\theta_0})a_{1,t-1}(\boldsymbol{\phi_0})|\eta_{-1}|^{\delta+2}+(\omega_0\lambda_1+R_{-2})\eta_{-1}^2+R_{-2}.\end{aligned}$$

Thus (E.33) entails almost surely

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

and

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

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

Proof of Proposition 1. We begin by studying the asymptotic distribution of statistics under the assumptions that the parameters are in the interior of the parameter space.

From (12) and Slutsky lemma, we obtain

$$\sqrt{n}(\boldsymbol{R}\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{k} - \boldsymbol{R}\boldsymbol{\theta_0} + \boldsymbol{k}) = \sqrt{n}\boldsymbol{R}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta_0}) \xrightarrow{\mathcal{L}} \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}(\kappa_\eta - 1)\boldsymbol{J}^{-1}\boldsymbol{R}') \quad \text{(E.34)}$$

and from the quadratic form, we thus have

$$\sqrt{n}(\boldsymbol{R}\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{k})' \left[\boldsymbol{R}(\kappa_\eta - 1)J^{-1}\boldsymbol{R}')\right]^{-1} \sqrt{n}(\boldsymbol{R}\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{k}) \xrightarrow{\mathcal{L}} \chi_c^2$$
$$\Leftrightarrow (\boldsymbol{R}\,\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{k})' \left(\boldsymbol{R}\left(\frac{(\hat{\kappa}_\eta - 1)}{n}\hat{\boldsymbol{J}}_n^{-1}\right)\boldsymbol{R}'\right)^{-1} (\boldsymbol{R}\,\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{k}) = W_n \xrightarrow{\mathcal{L}} \chi_c^2$$

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

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

$$\frac{\partial \tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0})}{\partial \boldsymbol{\theta}} + \boldsymbol{R}'\tilde{\boldsymbol{\lambda}}_n = 0 \quad \text{(E.35)}$$

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

Under H_0, we have

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

and

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

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

$$\sqrt{n}(\boldsymbol{R}\tilde{\boldsymbol{\theta}}_n - \boldsymbol{k}) = \boldsymbol{R}\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_{n|H_0}). \tag{E.36}$$

Using Taylor expansions, we can also notice that

$$0 = \sqrt{n}\frac{\partial \tilde{Q}_n(\tilde{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} \overset{o_P(1)}{=} \sqrt{n}\frac{\partial \tilde{Q}_n(\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta}} + \sqrt{n}\boldsymbol{J}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}) \tag{E.37}$$

and

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

which yields by subtraction

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

hence

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_{n|H_0}) \overset{o_P(1)}{=} -\sqrt{n}\boldsymbol{J}^{-1}\frac{\partial \tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0})}{\partial \boldsymbol{\theta}}. \tag{E.38}$$

From (E.35), (E.36) and (E.38), we thus obtain

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

which yields

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

hence from (E.34), under H_0,

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

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

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

and (E.35) yields

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

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

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

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

hence, by subtraction,

$$\begin{aligned}(\hat{\kappa}_{\eta|H_0}-1)L_n \overset{o_P(1)}{=}\ & 2n\frac{\partial \tilde{Q}_n(\boldsymbol{\theta_0})}{\partial \theta'}(\tilde{\boldsymbol{\theta}}_{n|H_0} - \tilde{\boldsymbol{\theta}}_n) + n(\tilde{\boldsymbol{\theta}}_{n|H_0} - \boldsymbol{\theta_0})' J(\tilde{\boldsymbol{\theta}}_{n|H_0} - \boldsymbol{\theta_0}) \\ & - n(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0})' J(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta_0}),\end{aligned} \tag{E.39}$$

and, from (E.37),

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

From (E.38), it follows that under H_0,

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

as $n \to \infty$. Hence, the critical region of the Quasi Likelihood Ratio test at the asymptotic level α is $\{L_n > \chi^2_c(1-\alpha)\}$. This concludes the proof of Proposition 1(i).

We now turn to the second part of the proposition, when the parameter is allowed to be on the boundary of the parameter space. For the ease of notation, we assume that $\underline{\phi_i} = 0$ for all $i = 1, \dots, r$. This is often the case in contitional volatility models as boundary conditions are necessary to ensure positivity of the conditional variance. In addition, without loss of generality, we consider testing that the last d_2 coefficients of $\boldsymbol{\theta_0}$ are on the boundary. We thus split $\boldsymbol{\theta_0}$ into two components $\boldsymbol{\theta_0} = (\boldsymbol{\theta_0}^{(1)}, \boldsymbol{\theta_0}^{(2)})'$, where $\boldsymbol{\theta_0}^{(i)} \in \mathbb{R}^{d_i}$, $d_1 + d_2 = 1 + r$. The null hypothesis is thus $H_0 : \boldsymbol{\theta_0}^{(2)} = \boldsymbol{R}\boldsymbol{\theta_0} = \mathbf{0}_{d_2\times 1}$ with $\boldsymbol{R} = (\mathbf{0}_{d_2\times d_1}, \boldsymbol{I}_{d_2})$. Let $H : \boldsymbol{\theta_0}^{(1)} = \overline{\boldsymbol{R}}\boldsymbol{\theta_0} > \mathbf{0}_{d_1\times 1}$ with $\boldsymbol{R} = (\boldsymbol{I}_{d_1}, \mathbf{0}_{d_1\times d_2})$ denote the maintained assumption.

From (13) and a direct application of the continuous mapping theorem, we have that, under H_0,

$$\sqrt{n}(\boldsymbol{R}\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{k} - \boldsymbol{R}\boldsymbol{\theta_0} + \boldsymbol{k}) = \sqrt{n}\boldsymbol{R}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta_0}) \xrightarrow{\mathcal{L}} \boldsymbol{R}\boldsymbol{\lambda}^{\boldsymbol{\Lambda}}$$

which yields

$$W_n \xrightarrow{\mathcal{L}} \boldsymbol{\lambda}^{\Lambda\prime}\boldsymbol{R}'[(\kappa_\eta - 1)\boldsymbol{R}\boldsymbol{J}^{-1}\boldsymbol{R}']^{-1}\boldsymbol{\lambda}^{\Lambda\prime}\boldsymbol{R}.$$

We now turn to the Rao-score statistic. Since $\tilde{\boldsymbol{\theta}}^{(1)}_{n|H_0}$ is a consistent estimator of $\boldsymbol{\theta_0}^{(1)}$, we have $\tilde{\boldsymbol{\theta}}^{(1)}_{n|H_0} > 0$ for n large enough. Therefore $\partial\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0})/\partial\theta_i = 0$ for $i = 1, \ldots, d_1$, or equivalently

$$\frac{\partial\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0})}{\partial\boldsymbol{\theta}} = \boldsymbol{R'}\frac{\partial\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0})}{\partial\boldsymbol{\theta}^{(2)}}. \tag{E.40}$$

A Taylor expansion yields

$$\sqrt{n}\frac{\partial\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0})}{\partial\boldsymbol{\theta}} \overset{o_P(1)}{=} \sqrt{n}\frac{\partial\tilde{Q}_n(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}} + \boldsymbol{J}(\tilde{\boldsymbol{\theta}}_{n|H_0} - \boldsymbol{\theta_0}). \tag{E.41}$$

The last d_2 components of this vector relation give

$$\sqrt{n}\frac{\partial\tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0})}{\partial\boldsymbol{\theta}^{(2)}} \overset{o_P(1)}{=} \sqrt{n}\frac{\partial\tilde{Q}_n(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}^{(2)}} + \boldsymbol{R}\boldsymbol{J}(\tilde{\boldsymbol{\theta}}_{n|H_0} - \boldsymbol{\theta_0}) \tag{E.42}$$

while the first d_1 components give

$$\mathbf{0} \overset{o_P(1)}{=} \sqrt{n}\frac{\partial\tilde{Q}_n(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}^{(1)}} + \sqrt{n}\overline{\boldsymbol{R}}\boldsymbol{J}\overline{\boldsymbol{R}}'(\tilde{\boldsymbol{\theta}}^{(1)}_{n|H_0} - \boldsymbol{\theta_0}^{(1)}) \tag{E.43}$$

using

$$\overline{\boldsymbol{R}}'(\tilde{\boldsymbol{\theta}}^{(1)}_{n|H_0} - \boldsymbol{\theta_0}^{(1)}) = \tilde{\boldsymbol{\theta}}_{n|H_0} - \boldsymbol{\theta_0}. \tag{E.44}$$

In view of (E.43), we have

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}^{(1)}_{n|H_0} - \boldsymbol{\theta_0}^{(1)}) \overset{o_P(1)}{=} (\overline{\boldsymbol{R}}\tilde{\boldsymbol{J}}_{n|H_0}\overline{\boldsymbol{R}}')^{-1}\sqrt{n}\frac{\partial\tilde{Q}_n(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}^{(1)}}. \tag{E.45}$$

From (E.40), (E.42), (E.44) and (E.45), we obtain

$$
\begin{aligned}
R_n &= \frac{n}{\tilde{\kappa}_{n|H_0}-1}\frac{\partial \tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0})}{\partial \boldsymbol{\theta}^{(2)\prime}}\boldsymbol{R}\tilde{\boldsymbol{J}}^{-1}_{n|H_0}\boldsymbol{R}'\frac{\partial \tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0})}{\partial \boldsymbol{\theta}^{(2)}} \\
&\overset{o_P(1)}{=} \frac{n}{\kappa_\eta-1}\left\|\frac{\partial \tilde{Q}_n(\tilde{\boldsymbol{\theta}}_{n|H_0})}{\partial \boldsymbol{\theta}^{(2)}}\right\|^2_{\boldsymbol{R}\boldsymbol{J}^{-1}\boldsymbol{R}'} \\
&\overset{o_P(1)}{=} \frac{n}{\kappa_\eta-1}\left\|\frac{\partial \tilde{Q}_n(\boldsymbol{\theta}_{\mathbf{0}})}{\partial \boldsymbol{\theta}^{(2)}}+\boldsymbol{R}\boldsymbol{J}\overline{\boldsymbol{R}}'(\tilde{\boldsymbol{\theta}}^{(1)}_{n|H_0}-\boldsymbol{\theta}^{(1)}_{\mathbf{0}})\right\|^2_{\boldsymbol{R}\boldsymbol{J}^{-1}\boldsymbol{R}'} \\
&\overset{o_P(1)}{=} \frac{n}{\kappa_\eta-1}\left\|\frac{\partial \tilde{Q}_n(\boldsymbol{\theta}_{\mathbf{0}})}{\partial \boldsymbol{\theta}^{(2)}}+\boldsymbol{R}\boldsymbol{J}\overline{\boldsymbol{R}}'(\overline{\boldsymbol{R}}\boldsymbol{J}\overline{\boldsymbol{R}}')^{-1}\frac{\partial \tilde{Q}_n(\boldsymbol{\theta}_{\mathbf{0}})}{\partial \boldsymbol{\theta}^{(1)}}\right\|^2_{\boldsymbol{R}\boldsymbol{J}^{-1}\boldsymbol{R}'}.
\end{aligned}
$$

Now recall that under H_0,

$$
\begin{pmatrix}\boldsymbol{W}_1\\ \boldsymbol{W}_2\end{pmatrix} := \sqrt{\frac{n}{\kappa_\eta-1}}\begin{pmatrix}\frac{\partial Q_n(\boldsymbol{\theta}_{\mathbf{0}})}{\partial \boldsymbol{\theta}^{(1)}}\\ \frac{\partial Q_n(\boldsymbol{\theta}_{\mathbf{0}})}{\partial \boldsymbol{\theta}^{(2)}}\end{pmatrix} \overset{\mathcal{L}}{\to} \mathcal{N}\left(\mathbf{0}, \boldsymbol{J}=\begin{pmatrix}\boldsymbol{J}_{11} & \boldsymbol{J}_{12}\\ \boldsymbol{J}_{21} & \boldsymbol{J}_{22}\end{pmatrix}\right). \tag{E.46}
$$

As $\boldsymbol{R}\boldsymbol{J}^{-1}\boldsymbol{R}' = (\boldsymbol{J}_{22}-\boldsymbol{J}_{21}\boldsymbol{J}_{11}^{-1}\boldsymbol{J}_{12})^{-1}$, the asymptotic distribution of R_n is that of $(\boldsymbol{W}_2-\boldsymbol{J}_{21}\boldsymbol{J}_{11}^{-1}\boldsymbol{W}_1)'(\boldsymbol{J}_{22}-\boldsymbol{J}_{21}\boldsymbol{J}_{11}^{-1}\boldsymbol{J}_{12})^{-1})(\boldsymbol{W}_2-\boldsymbol{J}_{21}\boldsymbol{J}_{11}^{-1}\boldsymbol{W}_1)$ under H_0, which follows a $\chi^2_{d_2}$ since $\boldsymbol{W}_2-\boldsymbol{J}_{21}\boldsymbol{J}_{11}^{-1}\boldsymbol{W}_1 \sim \mathcal{N}(0, \boldsymbol{J}_{22}-\boldsymbol{J}_{21}\boldsymbol{J}_{11}^{-1}\boldsymbol{J}_{12})$.

Finally, we turn to L_n. From (E.39), we have

$$
\begin{aligned}
\frac{\hat{\kappa}_{\eta|H_0}-1}{2}L_n \overset{o_P(1)}{=} & -n\left[\frac{1}{2}\frac{\partial \tilde{Q}_n(\boldsymbol{\theta}_{\mathbf{0}})}{\partial \boldsymbol{\theta}^{(1)\prime}}(\overline{\boldsymbol{R}}\boldsymbol{J}\overline{\boldsymbol{R}}^{-1})\frac{\partial \tilde{Q}_n(\boldsymbol{\theta}_{\mathbf{0}})}{\partial \boldsymbol{\theta}^{(1)}}\right. \\
&\left. +\frac{\partial \tilde{Q}_n(\boldsymbol{\theta}_{\mathbf{0}})}{\partial \boldsymbol{\theta}'}(\tilde{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_{\mathbf{0}})+\frac{1}{2}(\tilde{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_{\mathbf{0}})'J(\tilde{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_{\mathbf{0}})\right].
\end{aligned}
$$

Under H_0, by showing $\sqrt{n}\begin{pmatrix}\frac{\partial \tilde{Q}_n(\boldsymbol{\theta}_{\mathbf{0}})}{\partial \boldsymbol{\theta}}\\ \tilde{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_{\mathbf{0}}\end{pmatrix} \overset{\mathcal{L}}{\to} \begin{pmatrix}-\boldsymbol{J}\boldsymbol{Z}\\ \boldsymbol{\lambda}^{\boldsymbol{\Lambda}}\end{pmatrix}$, it can be seen that the asymptotic distribution of L_n is the law of

$$
L = \frac{1}{2}\boldsymbol{Z}'\boldsymbol{J}'\overline{\boldsymbol{R}}'\boldsymbol{J}_{11}^{-1}\overline{\boldsymbol{R}}\boldsymbol{J}\boldsymbol{Z}+\boldsymbol{Z}'\boldsymbol{J}'\boldsymbol{\lambda}^{\boldsymbol{\Lambda}\prime}-\frac{1}{2}\boldsymbol{\lambda}^{\boldsymbol{\Lambda}\prime}\boldsymbol{J}\boldsymbol{\lambda}^{\boldsymbol{\Lambda}}.
$$

Since,

$$
\boldsymbol{J}'\overline{\boldsymbol{R}}'\boldsymbol{J}_{11}^{-1}\overline{\boldsymbol{R}}\boldsymbol{J} = \boldsymbol{J}-(\kappa_\eta-1)\boldsymbol{\Omega} \quad \text{with } (\kappa_\eta-1)\boldsymbol{\Omega} = \begin{pmatrix}0 & 0\\ 0 & \boldsymbol{J}_{22}-\boldsymbol{J}_{21}\boldsymbol{J}_{11}^{-1}\boldsymbol{J}_{12}\end{pmatrix},
$$

we have

$$
\begin{aligned}
L &= -\frac{1}{2}\boldsymbol{Z}'\boldsymbol{J}\boldsymbol{Z}+\frac{1}{2}\boldsymbol{Z}'(\kappa_\eta-1)\boldsymbol{\Omega}\boldsymbol{Z}+\boldsymbol{Z}'\boldsymbol{J}'\boldsymbol{\lambda}^{\boldsymbol{\Lambda}}-\frac{1}{2}\boldsymbol{\lambda}^{\boldsymbol{\Lambda}\prime}\boldsymbol{J}\boldsymbol{\lambda}^{\boldsymbol{\Lambda}} \\
&= -\frac{1}{2}(\boldsymbol{\lambda}^{\boldsymbol{\Lambda}}-\boldsymbol{Z})'\boldsymbol{J}(\boldsymbol{\lambda}^{\boldsymbol{\Lambda}}-\boldsymbol{Z})+\frac{\kappa_\eta-1}{2}\boldsymbol{Z}'\boldsymbol{\Omega}\boldsymbol{Z}
\end{aligned}
$$

which concludes the proof. □

Proof of Proposition 4. First, note that, under H_0^{GARCH}, the continuous mapping theorem and Theorem 4 entail

$$W_n^{\text{GARCH}} \xrightarrow{\mathcal{L}} \boldsymbol{\lambda}^{\boldsymbol{\Lambda}\prime}\boldsymbol{\Omega}\boldsymbol{\lambda}^{\boldsymbol{\Lambda}}$$

with $\boldsymbol{\Omega} = \boldsymbol{R}'[(\kappa_\eta - 1)\boldsymbol{R}\boldsymbol{J}^{-1}\boldsymbol{R}']\boldsymbol{R}$. In addition, since we are in the case where only one coefficient is at the boundary of the parameter space, $\boldsymbol{\lambda}^{\boldsymbol{\Lambda}}$ has a trivial form. We have $\boldsymbol{\Lambda} = \mathbb{R}^{r-1} \times [0, \infty)$ and $\boldsymbol{R} = (0, \dots, 0, 1)$. We thus get

$$\boldsymbol{\lambda}^{\boldsymbol{\Lambda}\prime} = \boldsymbol{Z}\mathbb{1}_{Z_r \geq 0} + \boldsymbol{P}\boldsymbol{Z}\mathbb{1}_{Z_r < 0}, \text{ with } \boldsymbol{P} = \boldsymbol{I}_r - \boldsymbol{J}^{-1}\boldsymbol{R}'[\boldsymbol{R}\boldsymbol{J}^{-1}\boldsymbol{R}']^{-1}\boldsymbol{R}$$

where Z_d denotes the last component of vector $\boldsymbol{Z}$, and thus it follows that

$$\boldsymbol{\lambda}^{\boldsymbol{\Lambda}\prime} = \boldsymbol{Z} - [Z_r\mathbb{1}_{Z_r<0}]\boldsymbol{c}$$

where vector $\boldsymbol{c} = \mathbb{E}[Z_r\boldsymbol{Z}]/\mathbb{V}[Z_d]$ is the last column of $\boldsymbol{J}^{-1}$ divided by its (r, r)-component. We thus obtain

$$W_n^{\text{GARCH}} \xrightarrow{\mathcal{L}} \frac{(\boldsymbol{R}\boldsymbol{\lambda}^{\boldsymbol{\Lambda}})^2}{\boldsymbol{R}[(\kappa_\eta - 1)\boldsymbol{J}^{-1}]\boldsymbol{R}'} = \frac{(\boldsymbol{R}\boldsymbol{\lambda}^{\boldsymbol{\Lambda}})^2}{\mathbb{V}[Z_r]} = U^2\mathbb{1}_{U\geq 0} \sim \frac{1}{2}\Delta_0 + \frac{1}{2}\chi_1^2$$

where $U \sim \mathcal{N}(0, 1)$. □

References

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

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

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

[4] 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.

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

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

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

[8] 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.

[9] E. H. Zarantonello. Projections on convex sets in Hilbert space and spectral theory. In *Contributions to nonlinear functional analysis*, pages 237–424. Elsevier, 1971.