Journal of
Risk and Financial Management

MDPI

Article

Self-Weighted LSE and Residual-Based QMLE of ARMA-GARCH Models †

Shiqing Ling [1,*] **and Ke Zhu** [2]

[1] Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China
[2] Department of Statistics & Actuarial Science, The University of Hong Kong, Pok Fu Lam Road, Hong Kong, China; mazhuke@hku.hk
* Correspondence: maling@ust.hk
† This article is devoted to memorize Professor Michael McAleer for his friendship and long-term support of us.

Abstract: This paper studies the self-weighted least squares estimator (SWLSE) of the ARMA model with GARCH noises. It is shown that the SWLSE is consistent and asymptotically normal when the GARCH noise does not have a finite fourth moment. Using the residuals from the estimated ARMA model, it is shown that the residual-based quasi-maximum likelihood estimator (QMLE) for the GARCH model is consistent and asymptotically normal, but if the innovations are asymmetric, it is not as efficient as that when the GARCH process is observed. Using the SWLSE and residual-based QMLE as the initial estimators, the local QMLE for ARMA-GARCH model is asymptotically normal via an one-step iteration. The importance of the proposed estimators is illustrated by simulated data and five real examples in financial markets.

Keywords: ARMA models; GARCH models; QMLE; Self-weighted LSE

Citation: Ling, Shiqing, and Ke Zhu. 2022. Self-Weighted LSE and Residual-Based QMLE of ARMA-GARCH Models. *Journal of Risk and Financial Management* 15: 90. https://doi.org/10.3390/jrfm15020090

Academic Editor: Thanasis Stengos

Received: 6 January 2022
Accepted: 16 February 2022
Published: 19 February 2022

1. Introduction

Time series models have been extensively applied in various areas and many methodologies were proposed in the literature; for example, Zhang (2003) proposed a hybrid methodology that combines both ARIMA and ANN models to improve forecasting accuracy. Since Engle (1982), the ARCH-type models have been widely used in economics and finance. In particular, the GARCH model proposed by Bollerslev (1986) has been a benchmark in the risk management. Zhang and Zhang (2020) showed that the GARCH-based option-pricing models are able to price the SPX one-month variance swap rate, that is, the CBOE Volatility Index (VIX) accurately. Setiawan et al. (2021) used the GARCH(1, 1) model to analyze stock market turmoil during COVID-19 outbreak in an emerging and developed Economy.

However, recent research showed that the usual statistical inference procedure does not work if the fourth moment of the GARCH process does not exist. To make it clear, let us consider the AR(1)-GARCH(1, 1) model

$$y_t = \phi_1 y_{t-1} + \varepsilon_t, \tag{1}$$
$$\varepsilon_t = \eta_t \sqrt{h_t} \text{ and } h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}, \tag{2}$$

where $\alpha_0 > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$, and η_t is a sequence of independent and identically distributed (i.i.d.) innovations with zero mean and unit variance. For model (1), the least squares estimator (LSE) of ϕ_1 is

$$\hat{\phi}_{LSn} \equiv \left(\frac{1}{n}\sum_{t=1}^{n} y_{t-1}^2\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} y_{t-1}y_t\right),$$

where n is the sample size. Weiss (1986) and Pantula (1989) showed that $\hat{\phi}_{LSn}$ is $\sqrt{n}$-consistent and asymptotically normal if $E\varepsilon_t^4 < \infty$. However, $E\varepsilon_t^4 = \infty$ when the tail index α of ε_t is in $(0, 4]$. In this case, Davis and Mikosch (1998) and Basrak et al. (2002) showed that ε_t has a heavy-tailed feature and its sample autocorrelation function is neither $\sqrt{n}$-consistent nor asymptotically normal. Lange (2011) showed that $\hat{\phi}_{LSn}$ is $n^{1-2/\alpha}$-consistent and converges to a stable random variable when $\alpha \in (2, 4)$. Furthermore, for the AR model with ε_t being G-GARCH$(1, 1)$ noise in He and Teräsvirta (1999), Zhang and Ling (2015) showed that

$$\frac{\sqrt{n}}{\log n}(\hat{\phi}_{LSn} - \phi_1) \longrightarrow_{\mathcal{L}} \text{Normal, if } \alpha = 4 \ (i.e.\ E\varepsilon_t^4 = \infty), \tag{3}$$

$$n^{1-\frac{2}{\alpha}}(\hat{\phi}_{LSn} - \phi_1) \longrightarrow_{\mathcal{L}} \text{Stable, if } \alpha \in (2, 4) \ (i.e.\ E\varepsilon_t^2 < \infty \text{ and } E\varepsilon_t^4 = \infty), \tag{4}$$

$$\log n(\hat{\phi}_{LSn} - \phi_1) \longrightarrow_{\mathcal{L}} \text{Stable, if } \alpha = 2 \ (i.e.\ E\varepsilon_t^2 = \infty), \tag{5}$$

$$\hat{\phi}_{LSn} - \phi_1 \longrightarrow_{\mathcal{L}} \text{Stable, if } \alpha \in (0, 2) \ (i.e.\ E\varepsilon_t^2 = \infty), \tag{6}$$

when $n \to \infty$, where $\longrightarrow_{\mathcal{L}}$ denotes the convergence in distribution. From (3)–(6), we find that the LSE not only has a slower rate of convergence but also is not asymptotically normal when $\alpha \in (0, 4)$. Thus, based on the LSE, the classical theory and methodology (e.g., t-test, Wald test, and Ljung-Box test, among others) do not work in this case. Using a simulation method, we give the regime of parameter vector (α_1, β_1) with $E\varepsilon_t^{2\iota} < \infty$ in Figure 1 when $\eta_t \sim N(0, 1)$. It can be seen that the regime of (α_1, β_1) is very small for $E\varepsilon_t^4 < \infty$ (i.e., $\alpha > 4$). In practice, the estimated value of (α_1, β_1) does not lie in this regime, usually. Thus, it is very important to study the statistical inference when $\alpha \in (0, 4]$. Zhu and Ling (2015) studied the self-weighted least absolute deviation estimator (SLADE) of the ARMA-GARCH model and showed that it is consistent and asymptotically normal when $\alpha \in (0, 4]$.

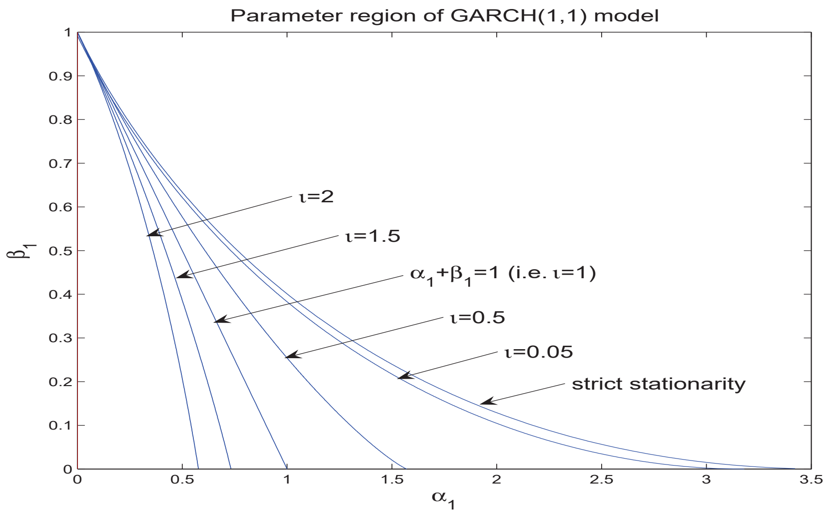

Figure 1. Parameter regime of (α_1, β_1) with $E\varepsilon_t^{2\iota} < \infty$.

This paper studies the self-weighted LSE (SWLSE) of the ARMA model with GARCH noises. It is shown that the SWLSE is consistent and asymptotically normal when the GARCH noise does not have a finite fourth moment (i.e., $\alpha \in (2, 4]$). Using the residuals from the estimated ARMA model, it is shown that the residual-based quasi-maximum likelihood estimator (QMLE) for the GARCH model is consistent and asymptotically normal, but if the innovations are asymmetric, it is not as efficient as that when the GARCH process is observed. Using the SWLSE and residual-based QMLE as the initial estimators, the local QMLE for ARMA-GARCH model is asymptotically normal via an one-step iteration.

This paper is arranged as follows. Section 2 presents the model and assumptions. Section 3 presents our main results. Section 4 presents simulation results and Section 5 gives real examples. All the proofs are deferred into the Appendix A.

2. Model and Assumptions

Assume that $\{y_t : t = 0, \pm 1, \pm 2, \cdots\}$ are generated by the ARMA-GARCH model

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \psi_i \varepsilon_{t-i} + \varepsilon_t, \tag{7}$$

$$\varepsilon_t = \eta_t \sqrt{h_t} \text{ and } h_t = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{s} \beta_i h_{t-i}, \tag{8}$$

where $\alpha_i \geq 0$ and $\beta_j \geq 0$, $i = 0, \cdots, r$, $j = 1, \cdots, s$, and η_t is defined as in (2). Denote $\gamma = (\mu, \phi_1, \cdots, \phi_p, \psi_1, \cdots, \psi_q)'$, $\delta = (\alpha_0, \alpha_1, \cdots, \alpha_r, \beta_1, \cdots, \beta_s)'$, and $\lambda = (\gamma', \delta')'$. Let γ_0, δ_0, and θ_0 be the true values of γ, δ, and θ, respectively. The parameter subspaces $\Theta_\gamma \subset R^{p+q+1}$ and $\Theta_\delta \subset R_0^{r+s+1}$ are compact, where $R = (-\infty, \infty)$ and $R_0 = [0, \infty)$. Denote $\Theta = \Theta_\gamma \times \Theta_\delta$, $m = p + q + r + s + 2$, $\alpha(z) = \sum_{i=1}^{r} \alpha_i z^i$, $\beta(z) = 1 - \sum_{i=1}^{s} \beta_i z^i$, $\phi(z) = 1 - \sum_{i=1}^{p} \phi_i z^i$, and $\psi(z) = 1 + \sum_{i=1}^{q} \psi_i z^i$. We introduce the following conditions:

Assumption 1. *θ_0 is an interior point in Θ and for each $\theta \in \Theta$, $\phi(z) \neq 0$ and $\psi(z) \neq 0$ when $|z| \leq 1$, and $\phi(z)$ and $\psi(z)$ have no common root with $\phi_p \neq 0$ or $\psi_q \neq 0$.*

Assumption 2. *$\alpha(z)$ and $\beta(z)$ have no common root, $\alpha_r + \beta_s \neq 0$, and $\sum_{i=1}^{r} \alpha_i + \sum_{j=1}^{s} \beta_j < 1$ for each $\theta \in \Theta$.*

Assumption 1 is the stationarity and invertibility condition of ARMA models, under which it follows that

$$\psi^{-1}(z) = \sum_{i=0}^{\infty} a_\psi(i) z^i \ \text{ and } \ \phi(z)\psi^{-1}(z) = \sum_{i=0}^{\infty} a_\gamma(i) z^i, \tag{9}$$

where $\sup_{\Theta_\gamma} |a_\psi(i)| = O(\rho^i)$ and $\sup_{\Theta_\gamma} |a_\gamma(i)| = O(\rho^i)$ with $\rho \in (0,1)$. Assumption 2 ensures that $\{\varepsilon_t\}$ is strictly stationary and ergodic with $E\varepsilon_t^2 < \infty$, see Ling and Li (1997) and Ling and McAleer (2002). It is also the identifiability condition for model (2) and, by Lemma 2.1 in Ling (1999), the condition $\sum_{i=1}^{s} \beta_i < 1$ is equivalent to

$$0 \leq \rho(G) < 1, \text{ where } G = \begin{pmatrix} \beta_1 & \cdots & \beta_s \\ I_{s-1} & & O \end{pmatrix}, \tag{10}$$

I_k is the $k \times k$ identity matrix, and $\rho(B)$ is the spectral radius of matrix B. Under this condition, we have

$$\beta^{-1}(z) = \sum_{i=0}^{\infty} a_\beta(i) z^i \text{ and } \alpha(z)\beta^{-1}(z) = \sum_{i=1}^{\infty} a_\delta(i) z^i, \tag{11}$$

where $\sup_{\Theta_\delta} |a_\beta(i)| = O(\rho^i)$ and $\sup_{\Theta_\delta} |a_\delta(i)| = O(\rho^i)$ with $\rho = \rho(G)$.

Given the observations $\{y_n, \cdots, y_1\}$ and initial value $Y_0 \equiv \{y_0, y_{-1}, \cdots\}$, we can write the parametric model as

$$\varepsilon_t(\gamma) = y_t - \mu - \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{i=1}^{q} \psi_i \varepsilon_{t-i}(\gamma), \tag{12}$$

$$\eta_t(\lambda) = \varepsilon_t(\gamma) / \sqrt{h_t(\lambda)} \text{ and } h_t(\lambda) = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2(\gamma) + \sum_{i=1}^{s} \beta_i h_{t-i}(\lambda). \tag{13}$$

It is easy to see that $\eta_t(\lambda_0) = \eta_t$, $\varepsilon_t(\gamma_0) = \varepsilon_t$, and $h_t(\lambda_0) = h_t$. In practice, we do not observe those y_i in Y_0 and hence they have to be replaced by some constants. This does not affect our asymptotic results, see Ling and McAleer (2003a). For simplicity, we do not study this case in details.

3. Main Results

The self-weighted estimation approach was proposed by Ling (2005) and it has been used to solve the problem on statistical inference of the heavy-tailed ARMA-GARCH model in Ling (2007) and Zhu and Ling (2011). Using a similar idea, we define the SWLSE as

$$\tilde{\gamma}_n = \arg\min_{\gamma\in\Theta_\gamma} \sum_{t=1}^{n} \frac{\varepsilon_t^2(\gamma)}{w_t},$$

where $w_t = 1 + \sum_{k=1}^{\infty} k^{-1/2-1}|y_{t-k}|$. We can state the following result:

Theorem 1. *Suppose that Assumptions 1–2 hold. Then, as $n \to \infty$,*

$$\begin{array}{ll} (i) & \tilde{\gamma}_n \longrightarrow_p \gamma_0, \\ (ii) & \sqrt{n}(\tilde{\gamma}_n - \gamma_0) \longrightarrow_{\mathcal{L}} N(0, A^{-1}BA^{-1}), \end{array}$$

where $\longrightarrow_p$ denotes the convergence in probability, $A = E(w_t^{-1}M_t)$, $B = E(w_t^{-2}h_tM_t)$, and $M_t = [\partial\varepsilon_t(\gamma_0)/\partial\gamma][\partial\varepsilon_t(\gamma_0)/\partial\gamma]'$.

The preceding result holds for any kind of ARCH-type errors only if $Eh_t < \infty$, see the proof in the Appendix A. To easily understand it, we refer to model (1) and (2) again. In this case, the information function is $E(y_{t-1}^2/w_t) \le E|y_{t-1}| < \infty$. The score function is $n^{-1/2}\sum_{t=1}^{n} y_{t-1}\varepsilon_t/w_t$ and $E(y_{t-1}\varepsilon_t/w_t)^2 \le O(1)Eh_t < \infty$, which is the condition we need for the GARCH errors. This result holds when $E\varepsilon_t^4 < \infty$, but it is not as efficient as the LSE in this case. When $E\varepsilon_t^4 = \infty$ and $E\varepsilon_t^2 < \infty$, the process y_t has a heavy tailed feature and the SWLSE has a faster rate of convergence than that of LSE. The weight function w_t can be replaced by others, see Ling (2007).

Next, we use the residual $\tilde{\varepsilon}_t \equiv \varepsilon_t(\tilde{\gamma}_n)$ from ARMA parts as the artificial observation of ε_t. The log-quasi-likelihood function based on $\tilde{\varepsilon}_t$ can be written as

$$\tilde{L}_{\delta n}(\delta) = \frac{1}{n}\sum_{t=1}^{n} \tilde{l}_t(\delta) \text{ and } \tilde{l}_t(\delta) = -\frac{1}{2}\log \tilde{h}_t(\delta) - \frac{\tilde{\varepsilon}_t^2}{2\tilde{h}_t(\delta)}, \tag{14}$$

where $\tilde{h}_t(\delta) = h_t(\lambda)|_{\gamma=\tilde{\gamma}_n}$. We define the residual-based QMLE of δ_0 as

$$\tilde{\delta}_n = \arg\max_{\delta\in\Theta_\delta} \tilde{L}_{\delta n}(\delta).$$

Denote $H_{\delta t}(\lambda) = h_t^{-2}(\lambda)[\partial h_t(\lambda)/\partial\delta][\partial h_t(\lambda)/\partial\delta']$ and $H_{\delta t}(\lambda_0)$ by $H_{\delta t}$. We now give the asymptotic properties of $\tilde{\delta}_n$ as follows.

Theorem 2. *Suppose that Assumptions 1 and 2 hold. Then, as $n \to \infty$,*

$$\begin{array}{ll} (i) & \tilde{\delta}_n \longrightarrow_p \delta_0, \ if \ E|\eta_t|^{2+\tilde{\iota}} < \infty \ for\ some\ \tilde{\iota} > 0, \\ (ii) & \sqrt{n}(\tilde{\delta}_n - \delta_0) \longrightarrow_{\mathcal{L}} N(0, (EH_{\delta t})^{-1}\Omega_\delta(EH_{\delta t})^{-1}), \ if \ E\eta_t^4 < \infty, \end{array}$$

where $\Omega_\delta = \kappa EH_{\delta t} + ED_t(A^{-1}BA^{-1})ED_t' + \kappa_3\tilde{\Omega}_\delta$, $\tilde{\Omega}_\delta = ED_tA^{-1}E(w_t^{-1}\tilde{D}_t') + E(w_t^{-1}\tilde{D}_t)A^{-1}ED_t'$, $\kappa = E\eta_t^4 - 1$, $\kappa_3 = E\eta_t^3$, $D_t = E\{h_t^{-2}[\partial h_t(\lambda_0)/\partial\delta][\partial h_t(\lambda_0)/\partial\gamma']\}$, and $\tilde{D}_t = E\{h_t^{-1/2}[\partial h_t(\lambda_0)/\partial\delta][\partial\varepsilon_t(\gamma_0)/\partial\gamma']\}$.

When η_t is symmetric and $\mu = 0$, we have $E\eta_t^3 = 0$, $ED_t = E\tilde{D}_t = 0$, and hence $\Omega_\delta = \kappa EH_{\delta t}$. When the conditional mean is zero (i.e., $y_t = \varepsilon_t$), model (7) and (8) reduces to the GARCH model. In this case, the log-quasi-likelihood function based on ε_t can be written as

$$L_{\delta n}(\delta) = \frac{1}{n}\sum_{t=1}^{n} l_t(\delta) \text{ and } l_t(\delta) = -\frac{1}{2}\log h_t(\delta) - \frac{\varepsilon_t^2}{2\,h_t(\delta)}. \tag{15}$$

Then, the global QMLE of δ_0 is defined as $\bar{\delta}_n = \arg\max_{\delta\in\Theta_\delta} L_{\delta n}(\delta)$. Berkes et al (2003) and Hall and Yao (2003) showed that $\bar{\delta}_n$ is consistent and as $n \to \infty$,

$$\sqrt{n}(\bar{\delta}_n - \delta_0) \longrightarrow_{\mathcal{L}} N(0, \kappa(EH_{\delta t})^{-1}), \text{ if } E\eta_t^4 < \infty. \tag{16}$$

From Theorem 2, we see that the efficiency of the estimated δ_0 is affected by the estimated parameters in ARMA parts unless η_t has a symmetric density and μ is known to be zero without estimation. This gives a reminder to practitioners that we need to be careful when ones use the residuals to estimate the GARCH model.

Given $\{y_n, \cdots, y_1\}$ and the initial value Y_0, we can write down the log-quasi-likelihood function of model (7) and (8) as follows:

$$L_n(\lambda) = \frac{1}{n}\sum_{t=1}^{n} l_t(\lambda) \text{ and } l_t(\lambda) = -\frac{1}{2}\log h_t(\lambda) - \frac{\varepsilon_t^2(\gamma)}{2\,h_t(\lambda)}. \tag{17}$$

Then, the global QMLE of λ_0 is defined as the maximizer of $L_n(\lambda)$ in Θ. Ling and McAleer (2003a) proved the consistency of this QMLE. But the asymptotic normality of this QMLE requires $E\varepsilon_t^4 < \infty$, see also Francq and Zakoïan (2004).

Based on $\tilde{\lambda}_n \equiv (\tilde{\gamma}_n', \tilde{\delta}_n')'$, we obtain the local QMLE through an one-step iteration

$$\hat{\lambda}_n = \tilde{\lambda}_n - \Big[\sum_{t=1}^{n} \frac{\partial^2 l_t(\tilde{\lambda}_n)}{\partial\lambda\partial\lambda'}\Big]^{-1} \sum_{t=1}^{n} \frac{\partial l_t(\tilde{\lambda}_n)}{\partial\lambda}. \tag{18}$$

As in Ling (2007), we can show that as $n \to \infty$,

$$\sqrt{n}(\hat{\lambda}_n - \lambda_0) \longrightarrow_{\mathcal{L}} N(0, \Sigma^{-1}\Omega\Sigma^{-1}),$$

where $\Sigma = E[U_t(\lambda_0)U_t'(\lambda_0)]$, $\Omega = E[U_t(\lambda_0)JU_t'(\lambda_0)]$, $J = \begin{pmatrix} 1 & \kappa_3 \\ \kappa_3 & \kappa \end{pmatrix}$, and $U_t(\lambda) = [h_t^{-1/2}\partial\varepsilon_t(\gamma)/\partial\lambda, h_t^{-1}\partial h_t(\lambda)/\partial\lambda]$. When $\eta_t \sim N(0,1)$, the local QMLE is efficient. So, Theorems 1 and 2 provide an approach to obtain an efficient estimator for the full ARMA-GARCH models under the finite second moment condition of ε_t. When η_t is not normal, the efficient and adaptive estimators can be obtained by using the results in this section and following the similar lines as in Drost et al. (1997), Drost and Klaassen (1997), Ling (2003), and Ling and McAleer (2003b).

4. Simulation Study

In this section, we assess the finite sample performance of $\tilde{\lambda}_n = (\tilde{\gamma}_n', \tilde{\delta}_n')'$ and $\hat{\lambda}_n = (\hat{\gamma}_n', \hat{\delta}_n')'$, where $\tilde{\gamma}_n$ is the SWLSE, $\tilde{\delta}_n$ is the residual-based QMLE, and $\hat{\lambda}_n$ is the local QMLE. We generate 1000 replications of sample size $n = 1000$ and 2000 from the following model

$$y_t = \phi_{10}y_{t-1} + \psi_{10}\varepsilon_{t-1} + \varepsilon_t, \tag{19}$$

$$\varepsilon_t = \eta_t\sqrt{h_t} \text{ and } h_t = \alpha_{00} + \alpha_{10}\varepsilon_{t-1}^2 + \beta_{10}h_{t-1}, \tag{20}$$

where $\gamma_0' = (\phi_{10}, \psi_{10}) = (0.4, 0.5)$, $\delta_0' = (\alpha_{00}, \alpha_{10}, \beta_{10}) = (0.1, 0.1, 0.8)$, and η_t is chosen to be the standard normal N$(0,1)$ distribution, re-scaled Laplace $L(0,1)$ distribution, or re-scaled student's $t(5)$ distribution with $E\eta_t^2 = 1$. Table 1 reports the sample bias (Bias), the sample standard deviations (SD), and the average estimated asymptotic standard deviation (AD) of $\tilde{\lambda}_n$ and $\hat{\lambda}_n$. From this table, we find that (i) each considered estimator has a small bias, and its value of SD is close to that of AD, demonstrating the validity of its asymptotic normality; (ii) $\hat{\gamma}_n$ could be slightly more efficient than $\tilde{\gamma}_n$, whereas $\hat{\delta}_n$ is as efficient as $\tilde{\delta}_n$; (iii) all estimators for $\eta_t \sim N(0,1)$ are more efficient than the corresponding ones for $\eta_t \sim L(0,1)$ or $t(5)$. All these findings are consistent with our theory in Section 3. We should mention that the QMLE of δ_0 is not reliable when the sample size n is less than 800 according to our simulation experiments and hence the results are not reported here.

Table 1. The results of $\tilde{\lambda}_n$ and $\hat{\lambda}_n$.

					$\tilde{\lambda}_n$		
η_t	n		$\tilde{\phi}_{1n}$	$\tilde{\psi}_{1n}$	$\tilde{\alpha}_{0n}$	$\tilde{\alpha}_{1n}$	$\tilde{\beta}_{1n}$
$N(0,1)$	1000	Bias	−0.0012	0.0032	0.0189	0.0012	−0.0235
		SD	0.0443	0.0423	0.0650	0.0278	0.0839
		AD	0.0424	0.0402	0.0524	0.0290	0.0726
	2000	Bias	−0.0017	0.0015	0.0083	−0.0000	−0.0103
		SD	0.0300	0.0293	0.0342	0.0204	0.0471
		AD	0.0300	0.0285	0.0332	0.0201	0.0469
					$\hat{\lambda}_n$		
			$\hat{\phi}_{1n}$	$\hat{\psi}_{1n}$	$\hat{\alpha}_{0n}$	$\hat{\alpha}_{1n}$	$\hat{\beta}_{1n}$
	1000	Bias	−0.0010	0.0022	0.0172	0.0015	−0.0210
		SD	0.0425	0.0406	0.0657	0.0282	0.0845
		AD	0.0405	0.0380	0.0526	0.0291	0.0729
	2000	Bias	−0.0016	0.0012	0.0073	−0.0000	−0.0087
		SD	0.0283	0.0274	0.0340	0.0205	0.0470
		AD	0.0286	0.0270	0.0332	0.0202	0.0469
					$\tilde{\lambda}_n$		
			$\tilde{\phi}_{1n}$	$\tilde{\psi}_{1n}$	$\tilde{\alpha}_{0n}$	$\tilde{\alpha}_{1n}$	$\tilde{\beta}_{1n}$
$L(0,1)$	1000	Bias	−0.0032	0.0035	0.0241	0.0020	−0.0304
		SD	0.0454	0.0414	0.0806	0.0381	0.1079
		AD	0.0456	0.0433	0.0639	0.0385	0.0909
	2000	Bias	−0.0001	0.0014	0.0116	0.0016	−0.0148
		SD	0.0328	0.0307	0.0426	0.0268	0.0599
		AD	0.0323	0.0307	0.0397	0.0269	0.0577
					$\hat{\lambda}_n$		
			$\hat{\phi}_{1n}$	$\hat{\psi}_{1n}$	$\hat{\alpha}_{0n}$	$\hat{\alpha}_{1n}$	$\hat{\beta}_{1n}$
	1000	Bias	−0.0027	0.0028	0.0237	0.0028	−0.0296
		SD	0.0444	0.0402	0.0918	0.0390	0.1183
		AD	0.0443	0.0416	0.0641	0.0387	0.0913
	2000	Bias	−0.0008	0.0013	0.0109	0.0019	−0.0138
		SD	0.0316	0.0296	0.0424	0.0270	0.0598
		AD	0.0313	0.0295	0.0397	0.0270	0.0578
					$\tilde{\lambda}_n$		
			$\tilde{\phi}_{1n}$	$\tilde{\psi}_{1n}$	$\tilde{\alpha}_{0n}$	$\tilde{\alpha}_{1n}$	$\tilde{\beta}_{1n}$
$t(5)$	1000	Bias	−0.0012	0.0016	0.0300	0.0046	−0.0395
		SD	0.0460	0.0445	0.0867	0.0432	0.1137
		AD	0.0454	0.0431	0.0734	0.0443	0.1038
	2000	Bias	0.0014	0.0005	0.0126	0.0025	−0.0164
		SD	0.0312	0.0305	0.0463	0.0325	0.0657
		AD	0.0323	0.0308	0.0459	0.0316	0.0666
					$\hat{\lambda}_n$		
			$\hat{\phi}_{1n}$	$\hat{\psi}_{1n}$	$\hat{\alpha}_{0n}$	$\hat{\alpha}_{1n}$	$\hat{\beta}_{1n}$
	1000	Bias	−0.0022	0.0018	0.0291	0.0054	−0.0381
		SD	0.0472	0.0448	0.0897	0.0444	0.1166
		AD	0.0443	0.0417	0.0737	0.0445	0.1042
	2000	Bias	0.0006	0.0007	0.0119	0.0030	−0.0155
		SD	0.0317	0.0296	0.0462	0.0330	0.0656
		AD	0.0315	0.0297	0.0459	0.0317	0.0667

As a comparison, we compute the classical LSE $\hat{\gamma}_{LSn} = (\hat{\phi}_{LSn}, \hat{\psi}_{LSn})'$ for γ_0 in model (19) and (20), where $\hat{\gamma}_{LSn}$ is computed in a similar way as $\tilde{\gamma}_n$ with $w_t \equiv 1$. Table 2 reports the corresponding results of $\hat{\gamma}_{LSn}$. Compared with $\tilde{\gamma}_n$ in Table 1, we find that $\hat{\gamma}_{LSn}$ is less efficient than $\tilde{\gamma}_n$ for all examined cases. This finding suggests that it seems better to fit the ARMA model by the SWLSE rather than the LSE method when the data exhibit the conditionally heteroscedastic effect.

Table 2. The results of $\hat{\gamma}_{LSn}$.

		$\eta_t \sim N(0,1)$		$\eta_t \sim L(0,1)$		$\eta_t \sim t(5)$	
n		$\hat{\phi}_{LSn}$	$\hat{\psi}_{LSn}$	$\hat{\phi}_{LSn}$	$\hat{\psi}_{LSn}$	$\hat{\phi}_{LSn}$	$\hat{\psi}_{LSn}$
1000	Bias	0.0001	0.0012	−0.0034	0.0024	−0.0033	0.0015
	SD	0.0441	0.0412	0.0482	0.0473	0.0518	0.0487
	AD	0.0437	0.0411	0.0507	0.0474	0.0525	0.0490
2000	Bias	−0.0018	0.0015	−0.0009	0.0011	−0.0008	0.0010
	SD	0.0307	0.0299	0.0350	0.0325	0.0382	0.0349
	AD	0.0311	0.0293	0.0367	0.0344	0.0382	0.0358

5. Real Examples

This section first studies the log returns ($\times 100$) of DJIA, NASDAQ, NASDAQ 100, and S&P 500 from 11 March 2015 to 10 March 2021, with a total of 1764 observations (see Figure 2). Denote each log return series by $\{y_t\}_{t=1}^{1764}$. Before fitting an AR(1)-GARCH$(1,1)$ to $\{y_t\}_{t=1}^{1764}$, we first estimate α_y, the tail index of $|y_t|$, and get the following results:

$$\begin{array}{rlrl} (\text{DJIA}) & \hat{\alpha}_y = \underset{(0.9285)}{2.3029}, & (\text{NASDAQ}) & \hat{\alpha}_y = \underset{(0.6830)}{3.2592}, \\ (\text{NASDAQ 100}) & \hat{\alpha}_y = \underset{(0.6077)}{3.6956}, & (\text{S\&P 500}) & \hat{\alpha}_y = \underset{(0.8567)}{2.5329}, \end{array}$$

where $\hat{\alpha}_y$ is the proposed estimator of α_y in Hill (2010), and the value in parentheses is the AD of $\hat{\alpha}_y$. From the above results, we can conclude that each $|y_t|$ has a finite second moment, but does not have a finite fourth moment. Hence, it is reasonable to fit four return series by using the procedure in Section 3, that is, we first obtain the SWLSE $\tilde{\gamma}_n$ and the residual-based QMLE $\tilde{\delta}_n$, and then obtain the local QMLE $\hat{\lambda}_n$. The resulting fitted models are as follows:

$$(\text{DJIA}) \left\{ \begin{array}{l} y_t = \underset{(0.0173)}{0.0859} - \underset{(0.0292)}{0.0461} y_{t-1} + \varepsilon_t, \\ h_t = \underset{(0.0109)}{0.0416} + \underset{(0.0378)}{0.2108} \varepsilon_{t-1}^2 + \underset{(0.0377)}{0.7532} h_{t-1}, \end{array} \right.$$

$$(\text{NASDAQ}) \left\{ \begin{array}{l} y_t = \underset{(0.0216)}{0.1009} - \underset{(0.0275)}{0.0663} y_{t-1} + \varepsilon_t, \\ h_t = \underset{(0.0178)}{0.0643} + \underset{(0.0335)}{0.1747} \varepsilon_{t-1}^2 + \underset{(0.0367)}{0.7826} h_{t-1}, \end{array} \right.$$

$$(\text{NASDAQ 100}) \left\{ \begin{array}{l} y_t = \underset{(0.0225)}{0.1125} - \underset{(0.0276)}{0.0654} y_{t-1} + \varepsilon_t, \\ h_t = \underset{(0.0180)}{0.0668} + \underset{(0.0325)}{0.1751} \varepsilon_{t-1}^2 + \underset{(0.0351)}{0.7855} h_{t-1}, \end{array} \right.$$

$$(\text{S\&P 500}) \left\{ \begin{array}{l} y_t = \underset{(0.0171)}{0.0910} - \underset{(0.0289)}{0.0838} y_{t-1} + \varepsilon_t, \\ h_t = \underset{(0.0117)}{0.0432} + \underset{(0.0422)}{0.2206} \varepsilon_{t-1}^2 + \underset{(0.0414)}{0.7453} h_{t-1}, \end{array} \right.$$

where all estimated parameters are the local QMLE $\hat{\lambda}_n$, and the values in parentheses are the ADs of $\hat{\lambda}_n$. From these fitted models, we can find that all estimated parameters are significantly different from zero at the level of 5%. In particular, the significant parameters in the fitted AR models imply that the U.S. stock market is not efficient during the examined period.

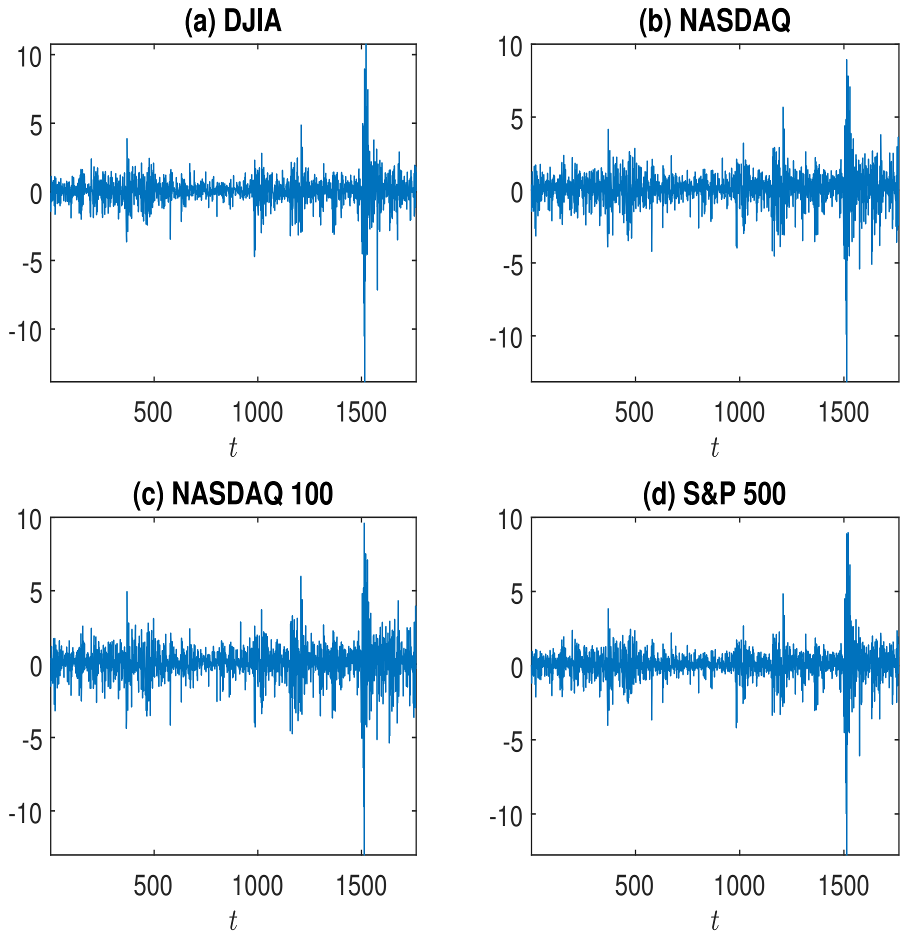

Figure 2. Log returns ($\times 100$) of DJIA, NASDAQ, NASDAQ 100, and S&P 500 from 11 March 2015 to 10 March 2021.

Next, this section considers the log returns ($\times 100$) of PHLX Oil Service Index OSX from 11 March 2015 to 10 March 2021, with a total of 1510 observations (see Figure 3). As before, we denote this log return series by $\{y_t\}_{t=1}^{1510}$, and obtain its estimate $\hat{\alpha}_y = 2.7960$ with AD $= 0.7078$. This implies that $|y_t|$ has a finite second moment, but does not have a finite fourth moment. Hence, we apply the local QMLE method to get the following fitted model for y_t:

$$
(\text{OSX}) \begin{cases} y_t = \underset{(0.0589)}{-0.0377} + \underset{(0.0307)}{0.0239} y_{t-1} + \varepsilon_t, \\ h_t = \underset{(0.0713)}{0.1329} + \underset{(0.0285)}{0.1076} \varepsilon_{t-1}^2 + \underset{(0.0304)}{0.8792} h_{t-1}. \end{cases}
$$

Unlike the fitted results for the four U.S. stock indexes above, the fitted AR coefficient for the OSX index is not significantly different from zero at the level of 5%, indicating that the oil market is efficient during the examined period.

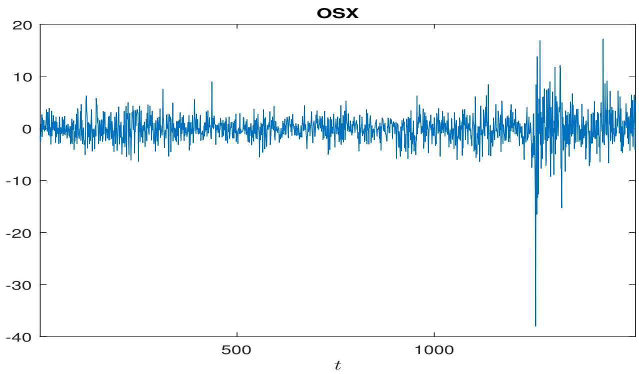

Figure 3. Log returns ($\times$100) of OSX from 11 March 2015 to 10 March 2021.

6. Concluding Remarks

This paper studied the SWLSE of the ARMA model with GARCH noises and the residual-based QMLE for the GARCH model. The consistency and asymptotic normality of SWLSE were established under a little moment condition. The importance of the proposed estimators was illustrated by simulated data and four major stock indexes and one major oil index in U.S. The ARMA-GARCH model is very important in the risk management, see He et al. (2019). In practice, ones need to build the ARMA-GARCH model from the historical data. The major contribution of our paper is to present a way to build an efficient and reliable model for this purpose. Several potential future research topics are listed as follows: first, we may extend our procedure for the hybrid methodology that combines both ARIMA and ANN models with GARCH errors as in Zhang (2003); second, we could use our procedure to analyze the energy data and build an ARMA-GARCH model for the green energy, renewable energy, and bio-energy data as discussing in An and Mikhaylov (2020); third, we may explore a linear programming or a genetic algorithm to find the QMLE of ARMA-GARCH model as presented in An et al. (2021).

Author Contributions: Conceptualization and methodology, S.L.; Data analysis, K.Z. All authors have read and agreed to the published version of the manuscript.

Funding: Ling's research was funded by Hong Kong Research Grants Commission Grants (nos. 16500117, 16303118, 16301620, and 16300621), Australian Research Council, and the NSFC. Zhu's research was funded by Hong Kong Research Grants Commission Grants (nos. 17304421, 17306818, and 17305619) and the NSFC (nos. 11690014 and 11731015).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available data sets were analyzed in this study and can be found at https://www.wsj.com/, accessed on 5 January 2022.

Acknowledgments: The authors thank the referees for careful reading and useful comments that helped to improve the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proofs

The following lemma gives two basic properties for model (7) and (8).

Lemma A1. *Suppose $\{\varepsilon_t\}$ is generated by model (8) satisfying Assumption 2. Then (i) $\{\varepsilon_t\}$ is strictly stationary and ergodic with $E\varepsilon_t^2 < \infty$, and has the following causal representation:*

$$\varepsilon_t = \eta_t \sqrt{h_t} \text{ and } h_t = \alpha_0 \Big[1 + \sum_{j=1}^{\infty} u' \prod_{i=0}^{j-1} P_{t-i} \xi_{t-j}\Big] \text{ a.s.;}$$

and (ii) there exists some $\iota \in (0,1)$ such that $E|\varepsilon_t|^{2+\iota} < \infty$ if $E|\eta_t|^{2+\tilde{\iota}} < \infty$ for some $\tilde{\iota} > 0$, where $\xi_t = (\eta_t^2, 0, \cdots, 0, 1, \cdots, 0)'_{(r+s)\times 1}$ with the first component η_t^2 and the $(r+1)$th component 1, and $u = (0, \cdots, 0, 1, \cdots, 0)'_{(r+s)\times 1}$ with the $(r+1)$th component 1, and

$$P_t = \left(\begin{array}{ccc|ccc} \alpha_1\eta_t^2 & \cdots & \alpha_r\eta_t^2 & \beta_1\eta_t^2 & \cdots & \beta_s\eta_t^2 \\ & I_{r-1} & O & & O & \\ \hline \alpha_1 & \cdots & \alpha_r & \beta_1 & \cdots & \beta_s \\ & O & & & I_{s-1} & O \end{array}\right).$$

Proof. The result in (i) is from Theorem 2.1 of Ling and Li (1997). For (ii), we first show that there exists an integer i_0 such that, for some $\tilde{\iota} \in (0,1)$,

$$E\|\prod_{k=1}^{i_0} P_{t-k}\|^{1+\tilde{\iota}_1} < 1, \tag{A1}$$

where $\|B\| = \sqrt{tr(BB')}$ for a vector or matrix B. Let $P = [\Pi, O]'_{(r+s)\times(r+s)}$ with $\Pi = (\alpha_1, \cdots, \alpha_r, \beta_1, \cdots, \beta_s)'$, C be defined as P_t with all the elements of its first row replaced by 0, and

$$P(x) = (E|\eta_t|^{2(1+x)})^{1/(1+x)}P + C.$$

Since $E|\eta_t|^{2+\tilde{\iota}} < \infty$, the spectral radius $\rho(P(x))$ is continuous in terms of x in $[0, \tilde{\iota})$. By Lemma 3.2 in Ling (1999) and Assumption 2, we know that $\rho(P(0)) = \rho(EP_t) < 1$, and there exists a constant $\tilde{\iota}_1 \in (0, \tilde{\iota})$ such that

$$\rho(P(\tilde{\iota}_1)) < \rho(EP_t) + [1 - \rho(EP_t)] < 1. \tag{A2}$$

By Corollary A.2 in Johansen (1995, p. 220) and (A2),

$$\|P^i(\tilde{\iota}_1)\| \le c[\rho(P(\tilde{\iota}_1))]^{i/2} \longrightarrow 0, \tag{A3}$$

as $i \to \infty$, where c is a constant. Let $c_j = (0, \cdots, 0, 1, 0, \cdots, 0)'_{(r+s)\times 1}$ with the jth element being 1. Since all the elements of P_t are nonnegative, it follows that

$$\|\prod_{k=1}^{i} P_t\| \le \sum_{j_1,j_2=1}^{r+s} c'_{j_1} \prod_{k=1}^{i} P_t c_{j_2}. \tag{A4}$$

By Minkowskii's inequality and (A3) and (A4), we have that

$$\begin{aligned} E\|\prod_{k=1}^{i} P_{t-k}\|^{1+\tilde{\iota}_1} &\le \Big(\sum_{j_1,j_2=1}^{r+s} \{E[c'_{j_1} \prod_{k=1}^{i} P_{t-k}c_{j_2}]^{1+\tilde{\iota}_1}\}^{1/(1+\tilde{\iota}_1)} \Big)^{1+\tilde{\iota}_1} \\ &= \Big[\sum_{j_1,j_2=1}^{r+s} \{E\Big[c'_{j_1} \prod_{k=1}^{i} (\eta_{t-k}^2 P + C)c_{j_2}\Big]^{1+\tilde{\iota}_1}\}^{1/(1+\tilde{\iota}_1)} \Big]^{1+\tilde{\iota}_1} \\ &\le \Big[\sum_{j_1,j_2=1}^{r+s} \Big(c'_{j_1} \prod_{k=1}^{i} [(E|\eta_t|^{2(1+\tilde{\iota}_1)})^{1/(1+\tilde{\iota}_1)}P + C]c_{j_2}\Big) \Big]^{1+\tilde{\iota}_1} \\ &= \Big[\sum_{j_1,j_2=1}^{r+s} c'_{j_1} P^i(\tilde{\iota}_1) c_{j_2} \Big]^{1+\tilde{\iota}_1} \longrightarrow 0, \end{aligned}$$

as $i \to \infty$. Thus, there is i_0 large enough such that (A1) holds. Using (A1) and the representation in (i), we can show that (ii) holds. This completes the proof. □

Lemma A2. *[Lemma A.1 in Ling (2007)] If Assumptions 1 and 2 hold, then there exist constants C and $\rho \in (0,1)$ such that the following holds uniformly in Θ:*

$$(i) \qquad \varepsilon_{t-1}(\gamma), \ \left\|\frac{\partial \varepsilon_t(\gamma)}{\partial \gamma}\right\|, \text{ and } \left\|\frac{\partial^2 \varepsilon_t(\gamma)}{\partial \gamma \partial \gamma'}\right\| \text{ are bounded a.s. by } \xi_{\gamma t-1},$$

$$(ii) \qquad h_t(\lambda) \text{ is bounded a.s. by } \xi^2_{\gamma t-1},$$

where $\xi_{\gamma t-1} = C(1 + \sum_{j=1}^{\infty} \rho^j |y_{t-j}|)$ with constants $\rho \in (0,1)$ and C.

Proof of Theorem. Let $L_{sn}(\gamma) = \sum_{t=1}^{n}[\varepsilon_t^2(\gamma)/w_t]/n$. First, the space Θ_γ is compact and γ_0 is an interior point in Θ_γ. Second, $L_{sn}(\gamma)$ is continuous in $\gamma \in \Theta_\gamma$ and is a measurable function of $\{y_s, s = t, t-1, \cdots\}$ for all $\gamma \in \Theta_\gamma$. Third, by Lemma A2(i),

$$E \sup_{\gamma \in \Theta_\gamma} [\varepsilon_t^2(\gamma)/w_t] \leq CE(1 + \sum_{i=0}^{\infty} \rho^i |y_{t-i}|)^2 < \infty,$$

where C is a constant. Moreover, by the ergodic theorem, $L_{sn}(\gamma) \longrightarrow_p E[\varepsilon_t^2(\gamma)/w_t]$ for each γ. Furthermore, by Theorem 3.1 in Ling and McAleer (2003a), $L_{sn}(\gamma) \longrightarrow_p E[\varepsilon_t^2(\gamma)/w_t]$ uniformly in Θ_γ. Fourth,

$$\varepsilon_t(\gamma) = \varepsilon_t - [M_t(\gamma) - M_t(\gamma_0)],$$

where $M_t(\gamma) = \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \phi_i \varepsilon_{t-i}(\gamma)$. Thus,

$$E\Big[\frac{\varepsilon_t^2(\gamma)}{w_t}\Big] = E\Big[\frac{\varepsilon_t^2(\gamma_0)}{w_t}\Big] + E\Big\{\frac{[M_t(\gamma) - M_t(\gamma_0)]^2}{w_t}\Big\} \geq E\Big[\frac{\varepsilon_t^2(\gamma_0)}{w_t}\Big],$$

where the equality holds if and only if $M_t(\gamma) = M_t(\gamma_0)$, that is, $\varepsilon_t(\gamma) = \varepsilon_t(\gamma_0)$, which holds if and only if $\gamma = \gamma_0$ under Assumption 1, that is, $E[\varepsilon_t^2(\gamma)/w_t]$ reaches its unique minimum at $\gamma = \gamma_0$. Thus, we have established all the conditions for consistency in Theorem 4.1.1 in Amemiya (1985) and hence (i) holds.

(ii) First, $\tilde{\gamma}_n$ is a consistent estimator of γ_0. Second,

$$\frac{\partial^2 L_{sn}(\gamma)}{\partial \gamma \partial \gamma'} = \frac{2}{n} \sum_{t=1}^{n} \frac{1}{w_t} \frac{\partial \varepsilon_t(\gamma)}{\partial \gamma} \frac{\partial \varepsilon_t(\gamma)}{\partial \gamma'} + \frac{2}{n} \sum_{t=1}^{n} \frac{\varepsilon_t(\gamma)}{w_t} \frac{\partial^2 \varepsilon_t(\gamma)}{\partial \gamma \partial \gamma'}$$

exists and is continuous in Θ_γ. Third, let

$$A_t(\gamma) \equiv \frac{1}{w_t} \frac{\partial \varepsilon_t(\gamma)}{\partial \gamma} \frac{\partial \varepsilon_t(\gamma)}{\partial \gamma'} + \frac{\varepsilon_t(\gamma)}{w_t} \frac{\partial^2 \varepsilon_t(\gamma)}{\partial \gamma \partial \gamma'}.$$

By Lemma A2, we can show that $E \sup_{\gamma \in \Theta_\gamma} \|A_t(\gamma)\| < \infty$. By the ergodic theorem and Theorem 3.1 in Ling and McAleer (2003a), we can show that $\partial^2 L_{sn}(\gamma)/\partial \gamma \partial \gamma'$ converges to $2EA_t(\gamma)$ uniformly in Θ_γ in probability. Since $EA_t(\gamma)$ is continuous in terms of γ, we can show that $\partial^2 L_{sn}(\gamma_n)/\partial \gamma \partial \gamma'$ converges to $2A$ in probability for any sequence γ_n, such that $\gamma_n \to \gamma_0$ in probability. Fourth,

$$\frac{\partial L_{sn}(\gamma_0)}{\partial \gamma'} = \frac{2}{n} \sum_{t=1}^{n} \frac{\varepsilon_t(\gamma_0)}{w_t} \frac{\partial \varepsilon_t(\gamma_0)}{\partial \gamma}.$$

By Lemma A2, it follows that

$$B = E\Big[\frac{\varepsilon_t^2(\gamma_0)}{w_t^2} \frac{\partial \varepsilon_t(\gamma_0)}{\partial \gamma} \frac{\partial \varepsilon_t(\gamma_0)}{\partial \gamma'}\Big] = E\Big[\frac{h_t(\lambda_0)}{w_t^2} \frac{\partial \varepsilon_t(\gamma_0)}{\partial \gamma} \frac{\partial \varepsilon_t(\gamma_0)}{\partial \gamma'}\Big] \leq C^2 Eh_t < \infty.$$

Similar to the proof of Lemma 4.2 in Ling and McAleer (2003a), we can show that A and B are positive definite. By the central limit theorem, we have that $\partial L_{sn}(\gamma_0)/\partial\gamma \longrightarrow_{\mathcal{L}} N(0, 4B)$. Thus, we have established all the conditions in Theorem 4.1.3 in Amemiya (1985), and hence $\sqrt{n}(\tilde{\gamma}_n - \gamma_0) \longrightarrow_{\mathcal{L}} N(0, A^{-1}BA^{-1})$. This completes the proof. □

The following Lemma A3(i)–(ii) is Lemma A.2 in Ling (2007) and Lemma A3(iii) is Lemma A.3(i) in Ling (2007).

Lemma A3. *If Assumptions 1 and 2 hold, then it follows that*

$$
\begin{aligned}
(i) \qquad & \sup_{\Theta}\left\|\frac{1}{h_t(\lambda)}\frac{\partial h_t(\lambda)}{\partial\delta}\right\| \le \xi_{\delta t-1}, \\
(ii) \qquad & \sup_{\Theta}\left\|\frac{1}{h_t(\lambda)}\frac{\partial^2 h_t(\lambda)}{\partial\delta\partial\delta'}\right\| \le \xi_{\delta t-1}, \\
(iii) \qquad & \sup_{\Theta}\left\|\frac{1}{\sqrt{h_t(\lambda)}}\frac{\partial h_t(\lambda)}{\partial\gamma}\right\| \le \xi_{\gamma t-1},
\end{aligned}
$$

where $\xi_{\delta t-1} = C(1 + \sum_{j=1}^{\infty}\rho^j|y_{t-j}|^{\iota_1})$ *with constants* $\rho \in (0,1)$ *and* C *for any* $\iota_1 > 0$.

To prove Theorem 2, we need to introduce another three lemmas. For their proofs, we need the condition that $E|\varepsilon_t|^{2+\tilde{\iota}_1} < \infty$ for some $\tilde{\iota}_1 > 0$. Here and in the sequel, $l_t(\delta) = l_t(\lambda)|_{\gamma=\gamma_0}$ and $h_t(\delta) = h_t(\lambda)|_{\gamma=\gamma_0}$.

Lemma A4. *If Assumptions 1 and 2 hold with* $E|\eta_t|^{2+\tilde{\iota}} < \infty$ *for some* $\tilde{\iota} > 0$, *then it follows that*

$$
\sup_{\delta\in\Theta_\delta}\left|\frac{1}{n}\sum_{t=1}^{n}[\tilde{l}_t(\delta) - l_t(\delta)]\right| = o_p(1).
$$

Proof. Since $\xi_{\gamma t}$ in Lemma A2 is strictly stationary with $E\xi_{\gamma t}^2 < \infty$, we have that $\max_{1\le t\le n}\xi_{\gamma t}/\sqrt{n} = o_p(1)$. By Taylor's expansion, Lemma A2(i), and Theorem 1(ii), it follows that

$$
\tilde{\varepsilon}_t = \varepsilon_t + (\tilde{\gamma}_n - \gamma_0)\frac{\partial\varepsilon_t(\gamma^*)}{\partial\gamma} = \varepsilon_t + o_p(1), \tag{A5}
$$

where $o_p(1)$ holds uniformly in t, and γ^* lies between γ_0 and $\tilde{\gamma}_n$. By (A5), we can readily show that

$$
\sup_{\delta\in\Theta_\delta}\left|\frac{1}{n}\sum_{t=1}^{n}\frac{\tilde{\varepsilon}_t^2 - \varepsilon_t^2}{\tilde{h}_t(\delta)}\right| = o_p(1), \tag{A6}
$$

since $\tilde{h}_t(\delta) \ge \underline{\alpha}_0$ uniformly in $\delta \in \Theta_\delta$. Note that

$$
\tilde{h}_t(\delta) = h_t(\delta) + (\tilde{\gamma}_n - \gamma_0)\frac{\partial h_t(\lambda^*)}{\partial\gamma}, \tag{A7}
$$

where $\lambda^* = (\gamma^{*'}, \delta')'$ and γ^* lies between γ_0 and $\tilde{\gamma}_n$. By Lemma A1(ii), we can show that $E(\varepsilon_t^2\xi_{\gamma t-1}^{\tilde{\iota}_1}) < \infty$ as $\tilde{\iota}_1$ is small enough. By Lemma A3(iii) and the ergodic theorem, it follows that

$$
\sup_{\lambda\in\Theta}\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^2\left\|\frac{\partial h_t(\lambda)}{\partial\gamma}\right\|^{\tilde{\iota}_1} \le \frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^2\xi_{\gamma t-1}^{\tilde{\iota}_1} = O_p(1),
$$

as $\tilde{\iota}_1$ is small enough. Thus,

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^2\Big|\frac{1}{\tilde{h}_t(\delta)}-\frac{1}{h_t(\delta)}\Big| &\le \frac{2}{\underline{\alpha}_0^{1-\tilde{\iota}_1}n}\sum_{t=1}^{n}\varepsilon_t^2\Big|\frac{1}{\tilde{h}_t(\delta)}-\frac{1}{h_t(\delta)}\Big|^{\tilde{\iota}_1} \\ &\le \frac{2}{\underline{\alpha}_0^{1+\tilde{\iota}_1}n}\sum_{t=1}^{n}\varepsilon_t^2\Big|\tilde{h}_t(\delta)-h_t(\delta)\Big|^{\tilde{\iota}_1} \\ &\le \frac{2\|\tilde{\gamma}_n-\gamma_0\|^{\tilde{\iota}_1}}{\underline{\alpha}_0^{1+\tilde{\iota}_1}n}\sum_{t=1}^{n}\varepsilon_t^2\Big\|\frac{\partial h_t(\lambda^*)}{\partial\gamma}\Big\|^{\tilde{\iota}_1} = o_p(1),\end{aligned} \tag{A8}$$

where $o_p(1)$ holds uniformly in $\delta\in\Theta_\delta$. By (A6) and (A8), it follows that

$$\sup_{\delta\in\Theta_\delta}\Big|\frac{1}{n}\sum_{t=1}^{n}\Big[\frac{\tilde{\varepsilon}_t^2}{\tilde{h}_t(\delta)}-\frac{\varepsilon_t^2}{h_t(\delta)}\Big]\Big| = o_p(1). \tag{A9}$$

Moreover, we can show that

$$\begin{aligned}&\frac{1}{n}\sum_{t=1}^{n}\Big[\log\tilde{h}_t(\delta)-\log h_t(\delta)\Big]I\{\tilde{h}_t(\delta)\ge h_t(\delta)\} \\ &=\frac{1}{n}\sum_{t=1}^{n}\Big[\log\Big[1+(\tilde{\gamma}_n-\gamma_0)\frac{1}{h_t(\delta)}\frac{\partial h_t(\lambda^*)}{\partial\gamma}\Big]\Big]I\{\tilde{h}_t(\delta)\ge h_t(\delta)\} \\ &\le\frac{1}{n\tilde{\iota}_1}\sum_{t=1}^{n}\log\Big[1+\underline{\alpha}_0^{-1}\|\tilde{\gamma}_n-\gamma_0\|\Big\|\frac{\partial h_t(\lambda^*)}{\partial\gamma}\Big\|\Big]^{\tilde{\iota}_1},\end{aligned}$$

where $\lambda^*=(\gamma^{*'},\delta')'$ and γ^* lies between γ_0 and $\tilde{\gamma}_n$. Note that there exists an $\tilde{\iota}_1$ such that $E\sup_{\lambda\in\Theta}\|\partial h_t(\lambda)/\partial\gamma\|^{\tilde{\iota}_1}<\infty$. For any $\varepsilon>0$, first taking η small enough such that $\log[1+\eta^{\tilde{\iota}_1}\underline{\alpha}_0^{-\tilde{\iota}_1}E\sup_{\lambda\in\Theta}\|\partial h_t(\lambda)/\partial\gamma\|^{\tilde{\iota}_1}]<\varepsilon^2\tilde{\iota}_1$ and then taking n large enough such that $P(\|\tilde{\gamma}_n-\gamma_0\|\ge\eta)\le\varepsilon$, it follows that

$$\begin{aligned}&P\Big(\frac{1}{n\tilde{\iota}_1}\sum_{t=1}^{n}\log\Big[1+\frac{1}{\underline{\alpha}_0}\|\tilde{\gamma}_n-\gamma_0\|\sup_{\lambda\in\Theta}\Big\|\frac{\partial h_t(\lambda)}{\partial\gamma}\Big\|\Big]^{\tilde{\iota}_1}\ge\varepsilon\Big) \\ &\le P\Big(\frac{1}{n\tilde{\iota}_1}\sum_{t=1}^{n}\log\Big[1+\frac{1}{\underline{\alpha}_0}\|\tilde{\gamma}_n-\gamma_0\|\sup_{\lambda\in\Theta}\Big\|\frac{\partial h_t(\lambda)}{\partial\gamma}\Big\|\Big]^{\tilde{\iota}_1}\ge\varepsilon,\|\tilde{\gamma}_n-\gamma_0\|\le\eta\Big) \\ &\quad+\varepsilon \\ &\le\frac{1}{n\tilde{\iota}_1\varepsilon}\sum_{t=1}^{n}E\log\Big[1+\frac{1}{\underline{\alpha}_0}\eta\sup_{\lambda\in\Theta}\Big\|\frac{\partial h_t(\lambda)}{\partial\gamma}\Big\|\Big]^{\tilde{\iota}_1}+\varepsilon \\ &=\frac{1}{\tilde{\iota}_1\varepsilon}E\log\Big[1+\frac{1}{\underline{\alpha}_0}\eta\sup_{\lambda\in\Theta}\Big\|\frac{\partial h_t(\lambda)}{\partial\gamma}\Big\|\Big]^{\tilde{\iota}_1}+\varepsilon \\ &\le\frac{1}{\tilde{\iota}_1\varepsilon}\log\Big[1+\frac{1}{\underline{\alpha}_0^{\tilde{\iota}_1}}\eta^{\tilde{\iota}_1}\sup_{\lambda\in\Theta}\Big\|\frac{\partial h_t(\lambda)}{\partial\gamma}\Big\|^{\tilde{\iota}_1}\Big]+\varepsilon\le 2\varepsilon,\end{aligned}$$

where the last second inequality holds by Jensen's inequality. Thus, as n is large enough,

$$P\Big(\sup_{\delta\in\Theta_\delta}\frac{1}{n}\sum_{t=1}^{n}[\log\tilde{h}_t(\delta)-\log h_t(\delta)]I\{\tilde{h}_t(\delta)\ge h_t(\delta)\}\ge\varepsilon\Big)\le 2\varepsilon.$$

Similarly, we can show that

$$P\Big(\sup_{\delta\in\Theta_\delta}\frac{1}{n}\sum_{t=1}^{n}[\log\tilde{h}_t(\delta)-\log h_t(\delta)]I\{\tilde{h}_t(\delta)\le h_t(\delta)\}\ge\varepsilon\Big)\le 2\varepsilon.$$

Furthermore, by (A9), the conclusion holds. This completes the proof. □

Lemma A5. *If the assumptions of Lemma A3 hold, then it follows that*

$$(i) \qquad \sup_{\delta\in\Theta_\delta}\Big\|\frac{1}{n}\sum_{t=1}^{n}\Big[\frac{\partial^2\tilde{l}_t(\delta)}{\partial\delta\partial\delta'}-\frac{\partial^2 l_t(\delta)}{\partial\delta\partial\delta'}\Big]\Big\| = o_p(1),$$

$$(ii) \qquad E\sup_{\delta\in\Theta_\delta}\Big\|\frac{\partial^2 l_t(\delta)}{\partial\delta\partial\delta'}\Big\| < \infty.$$

Proof. Denote $\tilde{V}_t(\delta) = \tilde{h}_t^{-1}(\delta)[\partial\tilde{h}_t(\delta)/\partial\delta]$ and similarly for $V_t(\delta)$. Then

$$\frac{\partial^2\tilde{l}_t(\delta)}{\partial\delta\,\partial\delta'} = -\frac{1}{2}\tilde{V}_t(\delta)\tilde{V}_t'(\delta)\frac{\tilde{\varepsilon}_t^2}{\tilde{h}_t(\delta)} + \Big[\frac{\tilde{\varepsilon}_t^2}{\tilde{h}_t(\delta)}-1\Big]\frac{\partial\tilde{V}_t(\delta)}{\partial\gamma}. \tag{A10}$$

Similarly, we can have the formula of $\partial^2 l_t(\delta)/\partial\delta\,\partial\delta'$. By (A5), we have

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\tilde{V}_t(\delta)\tilde{V}_t'(\delta)\frac{\tilde{\varepsilon}_t^2}{\tilde{h}_t(\delta)} &= \frac{1}{n}\sum_{t=1}^{n}\tilde{V}_t(\delta)\tilde{V}_t'(\delta)\frac{\varepsilon_t^2}{\tilde{h}_t(\delta)} + \frac{o_p(1)}{n}\sum_{t=1}^{n}\tilde{V}_t(\delta)\tilde{V}_t'(\delta)\frac{\varepsilon_t}{\tilde{h}_t(\delta)}\\ &+ \frac{o_p(1)}{n}\sum_{t=1}^{n}\tilde{V}_t(\delta)\tilde{V}_t'(\delta)\frac{1}{\tilde{h}_t(\delta)}.\end{aligned} \tag{A11}$$

By Lemma A3(i), $\sup_{\delta\in\Theta_\delta}\|\tilde{V}_t(\delta)\| \le \sup_\Theta\|h_t^{-1}(\lambda)[\partial h_t(\lambda)/\partial\delta]\| \le \xi_{\delta t-1}$. Furthermore, by Lemma A1, we can take ι_1 in $\xi_{\delta t-1}$ small enough such that the leading factors in the last terms are bounded uniformly in $\delta\in\Theta_\delta$. Thus, the last two terms are $o_p(1)$, and hence it follows that

$$\frac{1}{n}\sum_{t=1}^{n}\tilde{V}_t(\delta)\tilde{V}_t'(\delta)\frac{\tilde{\varepsilon}_t^2}{\tilde{h}_t(\delta)} = \frac{1}{n}\sum_{t=1}^{n}\tilde{V}_t(\delta)\tilde{V}_t'(\delta)\frac{\varepsilon_t^2}{\tilde{h}_t(\delta)} + o_p(1), \tag{A12}$$

where $o_p(1)$ holds uniformly in $\delta\in\Theta_\delta$. Moreover, by Lemma A3(i), we have

$$\begin{aligned}&\frac{1}{n}\sum_{t=1}^{n}\tilde{V}_t(\delta)\Big\|\tilde{V}_t(\delta)-V_t(\delta)\Big\|\frac{\varepsilon_t^2}{\tilde{h}_t(\delta)}\\ &\le \frac{1}{n}\sum_{t=1}^{n}\tilde{V}_t(\delta)\Big\|\tilde{V}_t(\delta)-V_t(\delta)\Big\|^{\iota}\Big[\Big\|\tilde{V}_t(\delta)\Big\|+\Big\|V_t(\delta)\Big\|\Big]^{1-\iota}\frac{\varepsilon_t^2}{\tilde{h}_t(\delta)}\\ &\le \frac{2}{n}\sum_{t=1}^{n}\xi_{\delta t-1}^{2-\iota}\Big\|\tilde{V}_t(\delta)-V_t(\delta)\Big\|^{\iota}\frac{\varepsilon_t^2}{\tilde{h}_t(\delta)}.\end{aligned} \tag{A13}$$

By Lemma A1 and taking ι and ι_1 in $\xi_{\delta t-1}$ small enough, we have

$$E\max_{1\le n<\infty}\sup_{\delta\in\Theta_\delta}\Big[\xi_{\delta t-1}^{2-\iota}\Big\|\tilde{V}_t(\delta)-V_t(\delta)\Big\|^{\iota}\frac{\varepsilon_t^2}{\tilde{h}_t(\delta)}\Big] \le CE(\xi_{\delta t-1}^2\varepsilon_t^2) < \infty,$$

where C is a constant. By the dominated convergence theorem, we can show that

$$\lim_{n\to\infty}E\sup_{\delta\in\Theta_\delta}\Big[\xi_{\delta t-1}^{2-\iota}\Big\|\tilde{V}_t(\delta)-V_t(\delta)\Big\|^{\iota}\frac{\varepsilon_t^2}{\tilde{h}_t(\delta)}\Big] = 0.$$

Thus, we can show that (A13) is $o_p(1)$ uniformly in $\delta\in\Theta_\delta$. Furthermore, by (A12),

$$\frac{1}{n}\sum_{t=1}^{n}\tilde{V}_t(\delta)\tilde{V}_t(\delta)\frac{\varepsilon_t^2}{\tilde{h}_t(\delta)} = \frac{1}{n}\sum_{t=1}^{n}\tilde{V}_t(\delta)V_t(\delta)\frac{\varepsilon_t^2}{\tilde{h}_t(\delta)} + o_p(1). \tag{A14}$$

Similarly, we can show that

$$\frac{1}{n}\sum_{t=1}^{n}\tilde{V}_t(\delta)V_t(\delta)\frac{\varepsilon_t^2}{\tilde{h}_t(\delta)}=\frac{1}{n}\sum_{t=1}^{n}V_t(\delta)V_t(\delta)\frac{\varepsilon_t^2}{\tilde{h}_t(\delta)}+o_p(1). \tag{A15}$$

Similar to (A8), we can show that

$$\frac{1}{n}\sum_{t=1}^{n}V_t(\delta)V_t(\delta)\frac{\varepsilon_t^2}{\tilde{h}_t(\delta)}=\frac{1}{n}\sum_{t=1}^{n}V_t(\delta)V_t(\delta)\frac{\varepsilon_t^2}{h_t(\delta)}+o_p(1). \tag{A16}$$

The $o_p(1)$ in (A14)–(A16) hold uniformly in $\delta\in\Theta_\delta$. By (A12) and (A14)–(A16), we have that

$$\frac{1}{n}\sum_{t=1}^{n}\tilde{V}_t(\delta)\tilde{V}_t(\delta)\frac{\tilde{\varepsilon}_t^2}{\tilde{h}_t(\delta)}=\frac{1}{n}\sum_{t=1}^{n}V_t(\delta)V_t(\delta)\frac{\varepsilon_t^2}{h_t(\delta)}+o_p(1).$$

We can show that a similar equation holds for other terms in (A10). Thus, (i) holds. By Lemmas A2 and A3, it is straightforward to show that (ii) holds. This completes the proof. □

Lemma A6. *[Lemma A.7 in Ling (2007)] If the conditions in Theorem 1 holds and $\sqrt{n}\|\lambda-\lambda_0\|\le M$, then it follows that*

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\lambda)}{\partial\lambda\partial\lambda'}=\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\lambda_0)}{\partial\lambda\partial\lambda'}+o_p(1),$$

for any fixed constant M.

Proof of Theorem. Let $\tilde{L}_n(\delta)=\sum_{t=1}^{n}\tilde{l}_t(\delta)/n$. First, the space Θ_δ is compact and δ_0 is an interior point in Θ_δ. Second, $\tilde{L}_n(\delta)$ is continuous in $\delta\in\Theta_\delta$ and is a measurable function of $\{y_s, s=t,t-1,\cdots\}$ for all $\delta\in\Theta_\delta$. Third, by Lemma A2(ii), there exist constants C and $\rho\in(0,1)$ such that

$$1\le\frac{h_t(\delta)}{\underline{\alpha}_0}\le C(1+\sum_{i=1}^{\infty}\rho^i|\varepsilon_{t-i}|)^2,$$

uniformly in $\delta\in\Theta_\delta$. By Jensen's inequality, $E\sup_{\delta\in\Theta_\delta}|\log h(\delta)|\le E\sup_{\delta\in\Theta_\delta}\log[h(\delta)/\underline{\alpha}_0]+|\log\underline{\alpha}_0|<\infty$. Thus, we can show that $E\sup_{\delta\in\Theta_\delta}|l_t(\delta)|<\infty$. By the ergodic theorem, $\sum_{t=1}^{n}l_t(\delta)/n\longrightarrow_p El_t(\delta)$ for each δ. Furthermore, by Theorem 3.1 in Ling and McAleer (2003a), $\sum_{t=1}^{n}l_t(\delta)/n\longrightarrow_p El_t(\delta)$ uniformly in Θ_δ. By Lemma A4, $\tilde{L}_n(\delta)\longrightarrow_p El_t(\delta)$ uniformly in Θ_δ. Fourth, similar to the proof of Lemma A.10 of Ling (2007), we can show that $El_t(\delta)$ reaches its unique maximum at $\delta=\delta_0$. Thus, we have established all the conditions for consistency in Theorem 4.1.1 in Amemiya (1985) and hence (i) holds.

For (ii), we first have a consistent estimator $\tilde{\delta}_n$ of δ_0. Second, $\partial^2\tilde{L}_n(\delta)/\partial\delta\partial\delta'$ exists and is continuous in Θ_δ. Third, by Lemma A5(ii), $E\sup_{\delta\in\Theta_\delta}\|\partial^2 l_t(\delta)/\partial\delta\partial\delta'\|<\infty$. By the ergodic theorem and Theorem 3.1 in Ling and McAleer (2003a), we can show that $\sum_{t=1}^{n}[\partial^2 l_t(\delta)/\partial\delta\partial\delta']/n\longrightarrow_p E[\partial^2 l_t(\delta)/\partial\delta\partial\delta']$ uniformly in Θ_δ. Since $E[\partial^2 l_t(\delta)/\partial\delta\partial\delta']$ is continuous in terms of δ, we can show that $\sum_{t=1}^{n}[\partial^2 l_t(\delta_n)/\partial\delta\partial\delta']/n\longrightarrow_p -EH_{\delta t}/4$ for any sequence δ_n, such that $\delta_n\longrightarrow_p\delta_0$. Furthermore, by Lemma A5(i), $\partial^2\tilde{L}_n(\delta_n)/\partial\delta\partial\delta'\longrightarrow_p -EH_{\delta t}/4$ for any sequence δ_n, such that $\delta_n\longrightarrow_p\delta_0$. Fourth, by Taylor's expansion, it follows that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial\tilde{l}_t(\delta_0)}{\partial\delta}=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t(\delta_0)}{\partial\delta}+\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Big[\frac{\partial^2 l_t(\lambda^*)}{\partial\delta\partial\gamma'}\Big](\tilde{\gamma}_n-\gamma_0),$$

where $\lambda^* = (\gamma^{*'}, \delta_0')'$ and γ^* lies between γ_0 and γ. By Lemma A6, we have

$$\frac{1}{n}\sum_{t=1}^{n}\Big[\frac{\partial^2 l_t(\lambda^*)}{\partial\delta\partial\gamma'}\Big] = E\Big[\frac{\partial^2 l_t(\lambda_0)}{\partial\delta\partial\gamma'}\Big] + o_p(1) = -\frac{1}{2}ED_t + o_p(1).$$

Furthermore, by Theorem 1, we can show that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial\tilde{l}_t(\delta_0)}{\partial\delta} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t(\delta_0)}{\partial\delta} + \frac{ED_tA^{-1}}{2\sqrt{n}}\sum_{t=1}^{n}\frac{\varepsilon_t(\gamma_0)}{w_t}\frac{\partial\varepsilon_t(\gamma_0)}{\partial\gamma} + o_p(1).$$

By Lemma A4, we can see that $E\|H_{\delta t}\| < \infty$ and $E\|\partial l_t(\delta_0)/\partial\delta\|^2 < \infty$. Thus, Ω_δ is finite. Similar to the proof of Lemma 4.2 in Ling and McAleer (2003a), we can show that $EH_{\delta t}$ and Ω_δ are positive definite. By the central limit theorem, we have that $n^{-1/2}\partial\tilde{L}_n(\delta_0)/\partial\delta \longrightarrow_{\mathcal{L}} N(0, \Omega_\delta/4)$. Thus, we have established all the conditions in Theorem 4.1.3 in Amemiya (1985), and hence $\sqrt{n}(\hat{\delta}_n - \delta_0) \longrightarrow_{\mathcal{L}} N(0, E^{-1}H_{\delta t}\Omega_\delta E^{-1}H_{\delta t})$. This completes the proof. □

References

An, Jaehyung, Alexey Mikhaylov, and Sang-Uk Jung. 2021. A linear programming approach for robust network revenue management in the airline industry. *Journal of Air Transport Management* 91: 101979. [CrossRef]

An, Jaehyung, and Alexey Mikhaylov. 2020. Russian energy projects in South Africa. *Journal of Energy in Southern Africa* 31: 58–64. [CrossRef]

Amemiya, Takeshi. 1985. *Advanced Econometrics*. Cambridge: Harvard University Press.

Basrak, Bojan, Richard A. Davis, and Thomas Mikosch. 2002. Regular variation of GARCH processes. *Stochastic Processes and Their Applications* 99: 95–115. [CrossRef]

Berkes, István, Lajos Horváth, and Piotr Kokoszka. 2003. GARCH processes: Structure and estimation. *Bernoulli* 9: 201–7. [CrossRef]

Bollerslev, Tim. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31: 307–27. [CrossRef]

Davis, Richard A., and Thomas Mikosch. 1998. The sample autocorrelations of heavy-tailed processs with applications to ARCH. *Annals of Statistics* 26: 2049–80. [CrossRef]

Drost, Feike C., and Chris A. J. Klaassen. 1997. Efficient estimation in semiparametric GARCH models. *Journal of Econometrics* 81: 193–221. [CrossRef]

Drost, Feike C., Chris A. J. Klaassen, and Bas J. M. Werker. 1997. Adaptive estimation in time series models. *Annals of Statistics* 25: 786–817. [CrossRef]

Engle, Robert F. 1982. Autoregressive conditional heteroskedasticity with estimates of variance of U.K. inflation. *Econometrica* 50: 987–1008. [CrossRef]

Francq, Christian, and Jean-Michel Zakoïan. 2004. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10: 605–637. [CrossRef]

Hall, Peter, and Qiwei Yao. 2003. Inference in ARCH and GARCH models. *Eonometrica* 71: 285–317. [CrossRef]

He, Changli, and Timo Teräsvirta. 1999. Properties of moments of a family of GARCH processes. *Journal of Econometrics* 92: 173–92. [CrossRef]

He, Yi, Yanxi Hou, Liang Peng, and Jiliang Sheng. 2019. Statistical inference for a relative risk measure. *Journal of Business & Economic Statistics* 37: 301–11.

Hill, Jonathan B. 2010. On tail index estimation for dependent, heterogeneous data. *Econometric Theory* 26: 1398–436. [CrossRef]

Johansen, Søren. 1995. *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford: OUP Oxford.

Lange, Theis. 2011. Tail behavior and OLS estimation in AR-GARCH models. *Statistica Sinica* 21: 1191–200. [CrossRef]

Ling, Shiqing. 1999. On the stationarity and the existence of moments of conditional heteroskedastic ARMA models. *Statistica Sinica* 9: 1119–30.

Ling, Shiqing. 2003. Adaptive estimators and tests of stationary and non-stationary short and long memory ARIMA-GARCH models. *Journal of the American Statistical Association* 98: 955–67. [CrossRef]

Ling, Shiqing. 2005. Self-weighted LAD estimation for infinite variance autoregressive models. *Journal of the Royal Statistical Society: Series B* 67: 381–93. [CrossRef]

Ling, Shiqing. 2007. Self-weighted and local quasi-maximum likelihood estimator for ARMA-GARCH/IGARCH models. *Journal of Econometrics* 140: 849–73. [CrossRef]

Ling, Shiqing, and Michael McAleer. 2002. Necessary and sufficient moment conditions for the GARCH(r,s) and asymmetric power GARCH(r,s) models. *Econometric Theory* 18: 722–29. [CrossRef]

Ling, Shiqing, and Wai Keung Li. 1997. Fractional autoregressive integrated moving-average time series with conditional heteroskedasticity. *Journal of the American Statistical Association* 92: 1184–94. [CrossRef]

Ling, Shiqing, and Michael McAleer. 2003a. Asymptotic theory for a new vector ARMA-GARCH model. *Econometric Theory* 19: 280–310. [CrossRef]

Ling, Shiqing, and Michael McAleer. 2003b. On adaptive estimation in nonstationary ARMA models with GARCH errors. *Annals of Statistics* 31: 642–74. [CrossRef]

Pantula, Sastry G. 1989. Estimation of autoregressive models with ARCH errors. *Sankhyā: The Indian Journal of Statistics, Series B* 50: 119–38.

Setiawan, Budi, Marwa Ben Abdallah, Maria Fekete-Farkas, Robert Jeyakumar Nathan, and Zoltan Zeman. 2021. GARCH (1, 1) models and analysis of stock market turmoil during COVID-19 outbreak in an emerging and developed economy. *Journal of Risk and Financial Management* 14: 576. [CrossRef]

Weiss, Andrew A. 1986. Asymptotic theory for ARCH models: Estimation and testing. *Econometrics Theory* 2: 107–31. [CrossRef]

Zhang, G. Peter. 2003. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 50: 159–75. [CrossRef]

Zhang, Rongmao, and Shiqing Ling. 2015. Asymptotic inference for AR models with heavy-tailed G-GARCH noises. *Econometric Theory* 31: 880–90. [CrossRef]

Zhang, Wenjun, and Jin E. Zhang. 2020. GARCH option pricing models and the variance risk premium. *Journal of Risk and Financial Management* 13: 51. [CrossRef]

Zhu, Ke, and Shiqing Ling. 2011. Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA-GARCH/IGARCH models. *Annals of Statistics* 39: 2131–63. [CrossRef]

Zhu, Ke, and Shiqing Ling. 2015. LADE-based inference for ARMA models with unspecified and heavy-tailed heteroscedastic noises. *Journal of the American Statistical Association* 110: 784–94. [CrossRef]