
Volume 140 Issue 2 October 2007 ISSN 0304-4076
ELSEVIER
JOURNAL OF Econometrics
Editors:
TAKESHI AMEMIYA
A. RONALD GALLANT
JOHN F. GEWEKE
CHENG HSIAO
PETER ROBINSON
ARNOLD ZELLNER
Associate Editors:
YACINE AÏT-SAHALIA
BADI H. BALTAGI
MICHAEL BRANDT
MARCUS J. CHAMBERS
SONGNIAN CHEN
VICTOR CHERNOZHUKOV
MANFRED DEISTLER
MIGUEL A. DELGADO
JEAN-MARIE DUFOUR
SYLVIA FRUHWIRTH-SCHNATTER
ERIC GHYSELS
JOHN C. HAM
JAVIER HIDALGO
HAN HONG
YONGMIAO HONG
BO E. HONORÉ
MAXWELL L. KING
YUICHI KITAMURA
GARY KOOP
CHUNG-MING KUAN
NAOTO KUNITOMO
KAJAL LAHIRI
QI LI
TONG LI
OLIVER LINTON
ROSA L. MATZKIN
FRANZ C. PALM
DALE J. POIRIER
BENEDIKT PÖTSCHER
INGMAR PRUCHA
PETER C. REISS
ERIC RENAULT
FRANK SCHORFHEIDE
ROBIN SICKLES
FALLAW SOWELL
GERARD VAN DEN BERG
HERMAN VAN DIJK
QUANG H. VUONG
EDWARD VYTLACIL
TOM WANSBEEK
TAO ZHA
CONTENTS
L.-f. Lee, Identification and estimation of econometric models with group interactions, contextual factors and fixed effects . . . 333
A. Daouia and L. Simar, Nonparametric efficiency analysis: A multivariate conditional quantile approach . . . 375
S. Chib, Analysis of treatment response data without the joint distribution of potential outcomes . . . 401
J. Reif, Asymptotic behaviour of regression pre-test estimators with minimal Bayes risk . . . 413
Y. Omori, S. Chib, N. Shephard and J. Nakajima, Stochastic volatility with leverage: Fast and efficient likelihood inference . . . 425
J. Hualde and P.M. Robinson, Root-n-consistent estimation of weak fractional cointegration . . . 450
V.G. Duggal, C. Saltzman and L.R. Klein, Infrastructure and productivity: An extension to private infrastructure and IT productivity . . . 485
J.M. Carro, Estimating dynamic panel data discrete choice models with fixed effects . . . 503
CONTENTS CONTINUED ON BACK COVER

Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 140 (2007) 849–873
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models

Shiqing Ling*

Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong

Available online 1 September 2006

Abstract

The limit distribution of the quasi-maximum likelihood estimator (QMLE) for parameters in the ARMA-GARCH model remains an open problem when the process has infinite 4th moment. We propose a self-weighted QMLE and show that it is consistent and asymptotically normal under only a fractional moment condition. Based on this estimator, the asymptotic normality of the local QMLE is established for the ARMA model with GARCH (finite variance) and IGARCH errors. Using the self-weighted and the local QMLEs, we construct Wald statistics for testing linear restrictions on the parameters, and their limiting distributions are given. In addition, we show that the tail index of the IGARCH process is always 2, which is independently of interest.

JEL classification: C12; C22

Keywords: Asymptotic normality; ARMA-GARCH model; GARCH model; Quasi-maximum likelihood estimation; Self-weighted estimation

1. Introduction

(G)ARCH-type models have been extensively used in economics and finance since Engle (1982) and Bollerslev (1986). Weiss (1986) first presented the asymptotic theory for the ARMA-ARCH model assuming finite fourth moment (see also Tsay, 1987). The ARCH (1) model is defined as $\varepsilon_t = \eta_t\sqrt{h_t}$ and $h_t = \alpha_0 + \alpha\varepsilon_{t-1}^2$, where $\alpha_0 > 0$, $\alpha > 0$ and $\eta_t \sim$ i.i.d. N(0, 1). It is well known that $\{\varepsilon_t\}$ is strictly stationary when $\alpha \in (0, 3.5620\cdots)$, and ε_t has a

*Tel.: +852 23587459; fax: +852 23591016.
E-mail address: maling@ust.hk.

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2006.07.016

finite second moment when $\alpha \in (0, 1)$, a finite fourth moment when $\alpha \in (0, 0.57\cdots)$ and a finite eighth moment when $\alpha \in (0, 0.3\cdots)$. It is clear that the moment conditions of ε_t link directly to the restriction on the parameters and that $E\varepsilon_t^4 < \infty$ is a very strong condition. The problem on the asymptotic theory for (G)ARCH-type models under weak moment conditions has attracted a lot of attention in econometrics and statistics.

For the GARCH(1,1) model, including the case when $E\varepsilon_t^2 = \infty$, Lee and Hansen (1994) and Lumsdaine (1996) showed that the quasi-maximum likelihood estimator (QMLE) of the parameters is consistent and asymptotically normal. For the GARCH(r, s) model as given below in Section 2, Berkes et al. (2003) proved strong consistency and asymptotic normality of the QMLE. The same results were proved by Hall and Yao (2003) and Francq and Zakoïan (2004). But Hall and Yao (2003) assumed that $E\varepsilon_t^2 < \infty$ and discussed the case when $E\eta_t^4 = \infty$. As far as we know, the weakest condition is presented by Francq and Zakoïan (2004). Consistency of the QMLE was also discussed by Jeantheau (1998) and Ling and McAleer (2003a). We also refer to recent works by Peng and Yao (2003) and Berkes and Horvath (2004) in this area. Basically, the asymptotic theory of the GARCH model is known.

However, we know less about the asymptotic theory of the ARMA-GARCH model. When $E\varepsilon_t^4 < \infty$, the consistency and asymptotic normality of its local QMLE were given by Ling and Li (1997), while the strong consistency and asymptotic normality of its global QMLE were proved by Francq and Zakoïan (2004). It is challenging to establish the limiting distribution of the QMLE when $E\varepsilon_t^4 = \infty$. Li et al. (2002) identified this as an open problem. This difficulty was also recognized by Francq and Zakoïan (2004). To understand this, we begin with the AR(1) model: $y_t = \phi y_{t-1} + \varepsilon_t$ with ε_t in (2.2) and let $\hat{\phi}_n$ be the LSE of ϕ. Then, $\sqrt{n}(\hat{\phi}_n - \phi) = (\sum_{t=1}^n y_{t-1}^2/n)^{-1}(\sum_{t=1}^n y_{t-1}\varepsilon_t/\sqrt{n})$. The central limit theorem requires $E(y_{t-1}\varepsilon_t)^2 = E(y_{t-1}^2 h_t) < \infty$ and hence the condition $E\varepsilon_t^4 < \infty$ should be minimal for asymptotic normality of $\hat{\phi}_n$. This is similar to the AR model with i.i.d. errors. The minimal condition for asymptotic normality of the LSE is $Ey_t^2 < \infty$. When $Ey_t^2 = \infty$, the asymptotic distributions of existing estimators, such as LSE, LAD and M-estimators, do not have a closed form (see Davis et al., 1992). Ling, 2005 introduced a self-weighted LAD estimator. The purpose of the weighting in Ling (2005) is to downweight the covariance matrix such that asymptotic normality can be recovered. For the ARCH model, a similar weighted LAD and L^p-estimator are used by Horvath and Liese (2004) and Chan and Peng (2005).

This paper first proposes a self-weighted QMLE for the ARMA-GARCH model. The basic idea is to use a weight to control the quasi-information matrix as in (3.4) below. We show that the self-weighted QMLE is consistent and asymptotically normal under only a fractional moment condition, i.e., $E|\varepsilon_t|^\iota < \infty$ for some $\iota > 0$. For the QMLE, $h_t(\theta)$ itself is one sort of weight in the quasi-log-likelihood function (3.3). Ling (2004) shows that the QMLE is asymptotically normal even if $E\varepsilon_t^2 = \infty$ for a double AR model and, hence, it is reasonable to expect that asymptotic normality of the QMLE holds when $E\varepsilon_t^4 = \infty$. We next establish asymptotic normality of the local QMLE for ARMA-GARCH (finite variance) and -IGARCH models. Based on the two estimators, Wald statistics are constructed for testing linear restrictions on the parameters. When $E\varepsilon_t^2 = \infty$, our result is entirely different from that in Mikosch et al. (1995) for the infinite-variance ARMA with i.i.d errors.

This paper is organized as follows. Section 2 presents the model and assumptions. Section 3 studies the self-weighted QMLE. Section 4 studies the local QMLE. Concluding remarks are offered in Section 5. All proofs are given in the Appendix.

2. Model and assumptions

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

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \psi_i \varepsilon_{t-i} + \varepsilon_t, \tag{2.1}$$

$$\varepsilon_t = \eta_t \sqrt{h_t} \quad \text{and} \quad h_t = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{s} \beta_i h_{t-i}, \tag{2.2}$$

where $\alpha_0 > 0$, $\alpha_i \geqslant 0$ $(i = 1, \ldots, r)$, $\beta_j \geqslant 0$ $(j = 1, \ldots, s)$, and η_t is a sequence of i.i.d. random variables with zero mean and variance 1. Denote $\gamma = (\mu, \phi_1, \ldots, \phi_p, \psi_1, \ldots, \psi_q)'$, $\delta = (\alpha_0, \alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_s)'$, and $\theta = (\gamma', \delta')'$. 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 $R_0 = [0, \infty)$. Let $\Theta = \Theta_\gamma \times \Theta_\delta$ and $m = p + q + r + s + 2$ and θ_0 be the true value of θ. Denote $\alpha(z) = \sum_{i=1}^{r} \alpha_i z^i$, $\beta(z) = 1 - \sum_{i=1}^{s} \beta_i z^i$, $\phi(z) = 1 - \sum_{i=1}^{p} \phi_i z^i$ and $\psi(z) = 1 + \sum_{i=1}^{q} \psi_i z^i$. We introduce the following conditions:

Assumption 2.1. θ_0 is an interior point in Θ and for each $\theta \in \Theta$, $\phi(z) \neq 0$ and $\psi(z) \neq 0$ when $|z| \leqslant 1$, and $\phi(z)$ and $\psi(z)$ have no common root with $\phi_p \neq 0$ or $\psi_q \neq 0$.

Assumption 2.2. $\alpha(z)$ and $\beta(z)$ have no common root, $\alpha(1) \neq 0$, $\alpha_r + \beta_s \neq 0$, and $\sum_{i=1}^{s} \beta_i < 1$ for each $\theta \in \Theta$.

Assumption 2.3. η_t^2 has a nondegenerate distribution with $E\eta_t^2 = 1$.

Assumption 2.4. $E|\varepsilon_t|^{2\iota} < \infty$ for some $\iota > 0$.

Assumption 2.1 implies the stationarity, invertibility and identifiability of model (2.1), under which it follows that

$$\psi^{-1}(z) = \sum_{i=0}^{\infty} a_\psi(i) z^i \quad \text{and} \quad \phi(z)\psi^{-1}(z) = \sum_{i=0}^{\infty} a_\gamma(i) z^i, \tag{2.3}$$

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 2.2 is the identifiability condition for model (2.2) and, by Lemma 2.1 in Ling (1999), the condition $\sum_{i=1}^{s} \beta_i < 1$ is equivalent to

$$0 \leqslant \rho(G) < 1 \quad \text{where } G = \begin{pmatrix} \beta_1 & \cdots & \beta_s \\ I_{s-1} & & O \end{pmatrix}, \tag{2.4}$$

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

$$\beta^{-1}(z) = \sum_{i=0}^{\infty} a_\beta(i) z^i \quad \text{and} \quad \alpha(z)\beta^{-1}(z) = \sum_{i=1}^{\infty} a_\delta(i) z^i, \tag{2.5}$$

where $\sup_{\Theta_\delta} a_\beta(i) = O(\rho^i)$, $\sup_{\Theta_\delta} a_\delta(i) = O(\rho^i)$ and $\rho = \rho(G)$. Assumption 2.3 is necessary to ensure that h_t is not almost surely (a.s.) a constant. When $\iota = 1$, the necessary and sufficient condition for Assumption 2.4 is

$$\sum_{i=1}^{r}\alpha_{0i}+\sum_{i=1}^{s}\beta_{0i}<1, \tag{2.6}$$

under which the GARCH model (2.2) has a finite variance. Furthermore, we give the necessary and sufficient condition for Assumption 2.4 when $\iota\in(0,1]$ in the following theorem.

Theorem 2.1. *Let* $\iota\in(0,1]$ *and* $\tilde{\beta}=\min\{\alpha_{0i},\beta_{0j}:i=0,1,\ldots,r,j=1,\ldots,s\}$. *Suppose* $\{\varepsilon_t\}$ *is generated by model* (2.2). (i) *If*

$$\text{there exists an integer } i_0 \text{ such that } \mathrm{E}\left\|\prod_{k=0}^{i_0-1}A_k\right\|^{\iota}<1, \tag{2.7}$$

then $\{\varepsilon_t\}$ *is strictly stationary and ergodic with* $\mathrm{E}|\varepsilon_t|^{2\iota}<\infty$; (ii) *If* $\tilde{\beta}>0$ *and* $\{\varepsilon_t\}$ *is strictly stationary with* $\mathrm{E}|\varepsilon_t|^{2\iota}<\infty$, *then* (2.7) *holds*; (iii) *if* $\tilde{\beta}>0$,

$$\sum_{i=1}^{r}\alpha_{0i}+\sum_{j=1}^{s}\beta_{0j}=1, \tag{2.8}$$

and η_t *has a positive density on* R *such that* $\mathrm{E}|\eta_t|^{\tau}<\infty$ *for all* $\tau<\tau_0$ *and* $\mathrm{E}|\eta_t|^{\tau_0}=\infty$ *for some* $\tau_0\in(0,\infty]$, *then* $\lim_{x\to\infty}x^2P(|\varepsilon_t|>x)$ *exists and is positive, where*

$$A_t=\begin{pmatrix}\alpha_{01}\eta_t^2 & \cdots & \alpha_{0r}\eta_t^2 & \beta_{01}\eta_t^2 & \cdots & \beta_{0s}\eta_t^2\\ & I_{r-1} & O & & O & \\ \alpha_{01} & \cdots & \alpha_{0r} & \beta_{01} & \cdots & \beta_{0s}\\ & O & & & I_{s-1} & O\end{pmatrix},$$

and $\|B\|=\sqrt{\mathrm{tr}(BB')}$ for a vector or matrix B.

We can show that (2.6) and (2.7) are equivalent when $\iota=1$. Under (2.8), model (2.2) is the IGARCH model with an infinite variance. Theorem 2.1 (iii) implies that the tail index of the IGARCH(r,s) process is always 2. When $r=s=1$, this tail index was also obtained by Basrak et al. (2002).

3. Self-weighted QMLE for ARMA-GARCH

Given the observations $\{y_n,\ldots,y_1\}$ and the initial values $\{y_0,y_{-1},y_{-2},\ldots\}$ which are generated by models (2.1)–(2.2), we can write the parametric model as

$$\varepsilon_t(\gamma)=y_t-\mu-\sum_{i=1}^{p}\phi_i y_{t-i}-\sum_{i=1}^{q}\psi_i\varepsilon_{t-i}(\gamma), \tag{3.1}$$

$$\eta_t(\theta)=\varepsilon_t(\gamma)/\sqrt{h_t(\theta)}\quad\text{and}\quad h_t(\theta)=\alpha_0+\sum_{i=1}^{r}\alpha_i\varepsilon_{t-i}^2(\gamma)+\sum_{i=1}^{s}\beta_i h_{t-i}(\theta). \tag{3.2}$$

Here, $\eta_t(\theta_0) = \eta_t$, $\varepsilon_t(\gamma_0) = \varepsilon_t$ and $h_t(\theta_0) = h_t$. The log-quasi-likelihood function based on $\{\varepsilon_t(\gamma) : t = 1, \ldots, n\}$ is

$$L_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} l_t(\theta) \quad \text{and} \quad l_t(\theta) = -\frac{1}{2}\log h_t(\theta) - \frac{\varepsilon_t^2(\gamma)}{2h_t(\theta)}. \tag{3.3}$$

The QMLE of θ_0 is defined as the maximizer of $L_n(\theta)$ on Θ. The quasi-score function and the quasi-information matrix are given in the Appendix. Eq. (A.18) in the Appendix shows that

$$\sup_{\Theta}\left\|\frac{\partial^2 l_t(\theta)}{\partial\gamma\,\partial\gamma'}\right\| \leqslant C\xi_{\rho t-1}^4(1+\eta_t^2), \tag{3.4}$$

where $\xi_{\rho t} = 1 + \sum_{i=0}^{\infty}\rho^i|y_{t-i}|$, and $\rho \in (0,1)$ and C are some constants not depending on t. This is why we generally need $E\varepsilon_t^4 < \infty$ for asymptotic normality of the QMLE. It is possible to obtain an estimator such that it is asymptotically normal if we can downweight $\xi_{\rho t-1}^4$. Thus, we introduce the weighted log-quasi-likelihood function:

$$L_{sn}(\theta) = \frac{1}{n}\sum_{t=1}^{n} w_t l_t(\theta), \tag{3.5}$$

where $l_t(\theta)$ is defined as in (3.3) and w_t satisfies the following assumption:

Assumption 3.1. $w_t = w(y_{t-1}, y_{t-2}, \ldots)$ and w is a measurable, positive and bounded function on R^{Z_0} with $E(w_t\xi_{\rho t-1}^4) < \infty$ and $Z_0 = \{0, 1, 2, \ldots\}$ for any $\rho \in (0,1)$.

In practice, we do not have the initial values y_i when $i \leqslant 0$ and hence they have to be replaced by some constants. Denote $\varepsilon_t(\gamma)$, $h_t(\theta)$ and w_t as $\tilde{\varepsilon}_t(\gamma)$, $\tilde{h}_t(\theta)$ and $\tilde{w}_t$, respectively, when y_i is a constant not depending on parameters when $i \leqslant 0$. The weighted log-quasi-likelihood function (3.5) is modified as follows:

$$\tilde{L}_{sn}(\theta) = \frac{1}{n}\sum_{t=1}^{n} \tilde{w}_t\tilde{l}_t(\theta) \quad \text{and} \quad \tilde{l}_t(\theta) = -\frac{1}{2}\log\tilde{h}_t(\theta) - \frac{\tilde{\varepsilon}_t^2(\gamma)}{\tilde{h}_t(\theta)}. \tag{3.6}$$

The following assumption makes the initial value y_i ignorable when $i \leqslant 0$:

Assumption 3.2. $E|w_t - \tilde{w}_t|^{\iota_0/4} = O(t^{-2})$, where $\iota_0 = \min\{\iota, 1\}$.

Since the weight w_t is determined by $\{y_t\}$ itself, the maximizer of $\tilde{L}_{sn}(\theta)$ on Θ is called the self-weighted QMLE, denoted as $\hat{\theta}_{sn}$. Let $\to_p$ and $\to_{\mathscr{L}}$, respectively, denote convergence in probability and in distribution as $n \to \infty$. Our result for $\hat{\theta}_{sn}$ is as follows.

Theorem 3.1. *Suppose that Assumptions* 2.1–2.4 *and* 3.1–3.2 *hold. Then,*

(i) $\hat{\theta}_{sn} \longrightarrow_p \theta_0$,

(ii) $\sqrt{n}(\hat{\theta}_{sn} - \theta_0) \longrightarrow_{\mathscr{L}} N(0, \Sigma_0^{-1}\Omega_0\Sigma_0^{-1})$ *if* $E\eta_t^4 < \infty$ *and* $J > 0$,

where $\Sigma_0 = E[w_t U_t(\theta_0)U_t'(\theta_0)]$, $\Omega_0 = E[w_t^2 U_t(\theta_0) J U_t'(\theta_0)]$, $J = \begin{pmatrix} 1 & \kappa_3 \\ \kappa_3 & \kappa \end{pmatrix}$, $\kappa_3 = E\eta_t^3/\sqrt{2}$, $\kappa = (E\eta_t^4 - 1)/2$ *and* $U_t(\theta) = [h_t^{-1/2}\partial\varepsilon_t(\gamma)/\partial\theta, (\sqrt{2}h_t)^{-1}\partial h_t(\theta)/\partial\theta]$.

It is readily shown that $J > 0$ if and only if $P(\eta_t^2 - c\eta_t - 1 = 0) < 1$ for any $c \in R$. A simple condition for this is that η_t has a positive density on some interval. Since θ_0 is an interior point, it excludes the ARCH as a special case of model (2.2). However, our framework allows us to deal with an ARCH model when it has been correctly identified. From its proof, Theorem 3.1 holds for the ARMA-ARCH model by removing the components corresponding to β_j's, and it holds similarly for the AR-ARCH and ARCH models.

The matrices Σ_0 and Ω_0 can be consistently estimated by their sample averages, i.e.,

$$\hat{\Sigma}_{0n} = \frac{1}{n}\sum_{t=1}^{n} \tilde{w}_t \hat{U}_{st} \hat{U}'_{st} \quad \text{and} \quad \hat{\Omega}_{0n} = \frac{1}{n}\sum_{t=1}^{n} \tilde{w}_t^2 \hat{U}_{st} \hat{J}_{sn} \hat{U}'_{st}, \tag{3.7}$$

where $\hat{U}_{st} = \tilde{U}_t(\hat{\theta}_{sn})$ and $\hat{J}_{sn}$ is defined as J with κ_3 and κ replaced by

$$\hat{\kappa}_{3sn} = \frac{1}{\sqrt{2}\hat{w}_1}\sum_{t=1}^{n} \tilde{w}_t \left[\frac{\tilde{\varepsilon}_t(\hat{\theta}_{sn})}{\sqrt{\tilde{h}_t(\hat{\theta}_{sn})}}\right]^3 \quad \text{and} \quad \hat{\kappa}_{sn} = \frac{1}{2\hat{w}_1}\sum_{t=1}^{n} \tilde{w}_t \frac{\tilde{\varepsilon}_t^4(\hat{\theta}_{sn})}{\tilde{h}_t^2(\hat{\theta}_{sn})} - \frac{1}{2},$$

respectively, and $\hat{w}_1 = \sum_{t=1}^{n} \tilde{w}_t$. Based on these, we can undertake statistical inference for the ARMA-GARCH model, such as the goodness-of-fit test in Li and Mak (1994). Here, we only consider the Wald test statistic, denoted by W_{sn}, for the p_1 linear hypothesis of the form: $\mathrm{H}_0 : \Gamma\theta = \theta_{10}$, in the usual notation. Using a similar method as for (A.17) and (A.19), we can show that $\hat{\Sigma}_{0n} = \Sigma_0 + o_p(1)$ and $\hat{\Omega}_{0n} = \Omega_0 + o_p(1)$. Thus, by Theorem 3.1, we have a corollary as follows.

Corollary 3.1. *If Assumptions* 2.1–2.4 *and* 3.1–3.2 *hold*, $J > 0$ *and* $\mathrm{E}\eta_t^4 < \infty$, *then it follows that*

$$W_{sn} = n(\Gamma\hat{\theta}_{sn} - \theta_{10})'(\Gamma\hat{\Sigma}_{1n}^{-1}\hat{\Omega}_{1n}\hat{\Sigma}_{1n}^{-1}\Gamma')^{-1}(\Gamma\hat{\theta}_{sn} - \theta_{10}) \longrightarrow_{\mathscr{L}} \chi^2_{p_1},$$

under H_0, *where* $\hat{\Sigma}_{0n}$ *and* $\hat{\Omega}_{0n}$ *are defined as in* (3.7).

We should mention that W_{sn} cannot be used for testing if the coefficients in the GARCH part are zero since θ_0 is an interior point in Θ under H_0. To use the result in this section, we need to select a weight w_t. Obviously, there are a lot of weights that satisfy Assumptions 3.1–3.2. When $\iota = \frac{1}{2}$ (i.e., $\mathrm{E}|\varepsilon_t| < \infty$), as in Ling (2005), one natural weight is

$$w_t = \left(\max\left\{1, C^{-1}\sum_{k=1}^{\infty} \frac{1}{k^9}|y_{t-k}| I\{|y_{t-k}| > C\}\right\}\right)^{-4}, \tag{3.8}$$

for some $C > 0$. This weight satisfies Assumptions 3.1–3.2 and has a connection to Huber's robust estimator for the regression model. It downweights the quasi-information matrices with large points (in absolute value) such that the magnitudes of its elements are not larger than C^4, but takes full advantage of all matrices without these points. When $q = s = 0$ (AR-ARCH model), for any $\iota > 0$, the weight can be selected as

$$w_t = \left(\max\left\{1, C^{-1}\sum_{k=1}^{p+r} \frac{1}{k^9}|y_{t-k}| I\{|y_{t-k}| > C\}\right\}\right)^{-4}. \tag{3.9}$$

When $\iota \in \left(0, \frac{1}{2}\right)$ and $q > 0$ or $s > 0$, the weight w_t may not be well defined and needs to be modified as follows:

$$w_t = \left(\max\left\{1, C^{-1}\sum_{k=1}^{\infty} \frac{1}{k^{1+8/\iota}}|y_{t-k}| I\{|y_{t-k}| > C\}\right\}\right)^{-4}. \tag{3.10}$$

In this case, we need to determine ι such that Assumption 2.4 holds. This is not an easy task, but there are two ways to do this. One way is to use a simulation method to verify the condition (2.7) for a possible ι. This is the same as the method used for verifying the stationarity condition of model (2.2) suggested by Bougerol and Picard (1992) and Basrak et al. (2002). But this method needs some information on the distribution of η_t. The test statistic in Koul and Ling (2006) can be used to test possible distributions. Another way is to use the Hill estimator to estimate the tail index of $\{y_t\}$ and its estimator may provide some useful guidelines for the choice of ι. The constant ι can be any value less than the tail index of $\{y_t\}$.

In general, it is difficult to compare the efficiency of estimators based on different weights. However, when $\mathrm{E}\|U_t(\theta_0)\|^2<\infty$, $\Sigma_0^{-1}\Omega_0\Sigma_0^{-1}$, based on the weights in (3.8)–(3.10), converges to the asymptotic covariance matrix of the QMLE in Section 4 as $C\to\infty$. When $\mathrm{E}\|U_t(\theta_0)\|^2=\infty$ (see the example on the AR-ARCH model in Francq and Zakoïan, 2004), for the weight in (3.9) or (3.10), we have

$$\|\Sigma_0^{-1}\Omega_0\Sigma_0^{-1}\|\leqslant\lambda\|\Sigma_0^{-1}\|\to 0 \quad \text{as } C\to\infty,$$

for some $\lambda>0$. That is, the asymptotic variance of the self-weighted QMLE can be as small as we want only if C is large enough. For the self-weighted LAD estimator for the infinite-variance AR model with i.i.d. errors in Ling (2005), simulation results show that it works well when C is the 90% or 95%-quantile of data $\{y_1,\ldots,y_n\}$.

4. Local QMLE for ARMA-GARCH/IGARCH

When studying the quasi-information matrix in detail, we find that $\mathrm{E}H_{\gamma t}(\theta)<\infty$ only if $\mathrm{E}\varepsilon_t^2<\infty$, where $H_{\gamma t}(\theta)=h_t^{-2}(\theta)[\partial h_t(\theta)/\partial\gamma][\partial h_t(\theta)/\partial\gamma']$ (see the proof in the Appendix). Another second-moment condition is required by the factor $\varepsilon_t^2(\gamma)/h_t(\theta)$ in (A.3). Note that $\varepsilon_t^2(\gamma_0)/h_t(\theta_0)=\eta_t^2$ is independent of $\mathscr{F}_{t-1}$. If we restrict the estimator on the subspace $\Theta_n=\{\theta:\|\theta-\theta_0\|\leqslant M/\sqrt{n}\}$ for any fixed $M>0$, $\varepsilon_t^2(\gamma)/h_t(\theta)$ is expected to be sufficiently close to η_t^2. Thus, $\mathrm{E}\varepsilon_t^2<\infty$ may be sufficient for asymptotic normality of the local QMLE. In addition, since the tail index of the IGARCH process is 2, $h_t(\theta)$ may be able to reduce a little bit of the required moment of ε_t in $H_{\gamma t}(\theta)$ such that the asymptotic normality of the local QMLE holds for the ARMA-IGARCH model. Thus, this section focuses on the local QMLE.

By using $\hat\theta_{sn}$ in Theorem 3.1 as an initial estimator of θ_0, we obtain the local MLE through the following one-step iteration:

$$\hat\theta_n=\hat\theta_{sn}-\left[\sum_{t=1}^{n}\frac{\partial^2\tilde l_t(\hat\theta_{sn})}{\partial\theta\,\partial\theta'}\right]^{-1}\sum_{t=1}^{n}\frac{\partial\tilde l_t(\hat\theta_{sn})}{\partial\theta}. \tag{4.1}$$

For this local QMLE, we have the following result:

Theorem 4.1. *Suppose that Assumptions* 2.1–2.3 *and* 3.1–3.2 *hold and that* (2.6) *or the condition of* Theorem 2.1 (iii) *is satisfied. If* $\mathrm{E}\eta_t^4<\infty$, $J>0$ *and* $\hat\theta_n$ *is obtained through* (4.1), *then*

$$\sqrt{n}(\hat\theta_n-\theta_0)\longrightarrow_{\mathscr{L}}\mathrm{N}(0,\Sigma^{-1}\Omega\Sigma^{-1}),$$

where $\Sigma=\mathrm{E}[U_t(\theta_0)U_t'(\theta_0)]$ *and* $\Omega=\mathrm{E}[U_t(\theta_0)JU_t'(\theta_0)]$.

For the ARMA-IGARCH model, the condition $\beta_{01} \neq 0$ is critical in the proof (see Lemmas A.5–A.6) and hence Theorem 4.1 cannot be applied to the ARMA-IARCH model. However, when $\mathrm{E}\varepsilon_t^2 < \infty$, we can take $\tilde{\imath} = 0$ in Lemma A.5 such that all the proofs follow. Thus, Theorem 4.1 in this case holds for ARMA-ARCH, GARCH and ARCH models only if we remove the reductant parameters and the corresponding components in the covariance matrix.

In general, it is not easy to compare the efficiency of the QMLE and the self-weighted QMLE in Theorem 3.1. However, when $J = \mathrm{diag}\{1, 1\}$, i.e., η_t has the same moments as those of N(0, 1) up to fourth-order, we can show that the QMLE is more efficient than the self-weighted QMLE. In fact, in this case,

$$\Sigma = \Omega = \mathrm{E}(X_{1t}X'_{1t} + X_{2t}X'_{2t}),$$
$$\Sigma_0 = \mathrm{E}(w_t X_{1t}X'_{1t} + w_t X_{2t}X'_{2t}) \quad \text{and} \quad \Omega_0 = \mathrm{E}(w_t^2 X_{1t}X'_{1t} + w_t^2 X_{2t}X'_{2t}),$$

where $X_{1t} = h_t^{-1/2}\partial\varepsilon_t(\gamma)/\partial\theta$ and $X_{2t} = (2h_t)^{-1}\partial h_t(\theta)/\partial\theta$. Let b and c be two any m-dimensional constant vectors. Then,

$$\begin{aligned} c'\Sigma_0 bb'\Sigma_0 c &= \{\mathrm{E}[(c'X_{1t}w_t)(X'_{1t}b) + (c'X_{2t}w_t)(X'_{2t}b)]\}^2 \\ &\leqslant \left\{ \sqrt{\mathrm{E}(c'X_{1t}w_t)^2\mathrm{E}(X'_{1t}b)^2} + \sqrt{\mathrm{E}(c'X_{2t}w_t)^2\mathrm{E}(X'_{2t}b)^2} \right\}^2 \\ &\leqslant [\mathrm{E}(c'X_{1t}w_t)^2 + \mathrm{E}(c'X_{2t}w_t)^2][\mathrm{E}(X'_{1t}b)^2 + \mathrm{E}(X'_{2t}b)^2] \\ &= c'\Omega_0 cb'\Sigma b = c'\Omega_0(b'\Sigma b)c. \end{aligned}$$

Thus, $\Omega_0 b'\Sigma b - \Sigma_0 bb'\Sigma_0 \geqslant 0$ (a positive semi-definite matrix) and hence $b'\Sigma_0\Omega_0^{-1}\Sigma_0 b = \mathrm{tr}(\Omega_0^{-1/2}\Sigma_0 bb'\Sigma_0\Omega_0^{-1/2}) \leqslant \mathrm{tr}(b'\Sigma b) = b'\Sigma b$. Thus, we have $\Sigma_0^{-1}\Omega_0\Sigma_0^{-1} \geqslant \Sigma^{-1}$.

For the pure (G)ARCH model (2.2), the asymptotic covariance matrices of the QMLE $\hat{\theta}_n$ and the self-weighted QMLE $\hat{\theta}_{sn}$ are

$$\Sigma^{-1}\Omega\Sigma^{-1} = \kappa\mathrm{E}^{-1}(X_{2t}X'_{2t}),$$
$$\Sigma_0^{-1}\Omega_0\Sigma_0^{-1} = \kappa\mathrm{E}^{-1}(w_t X_{2t}X'_{2t})\mathrm{E}(w_t^2 X_{2t}X'_{2t})\mathrm{E}^{-1}(w_t X_{2t}X'_{2t}),$$

respectively. Using the same method as for the previous case, we can show that $\Sigma_0^{-1}\Omega_0\Sigma_0^{-1} \geqslant \Sigma^{-1}\Omega\Sigma^{-1}$. Thus, the QMLE is always more efficient than the self-weighted QMLE. For the pure ARCH model, the asymptotic covariance of the weighted L^2-estimator in Horvath and Liese (2004) is $\Sigma_1^{-1}\Omega_1\Sigma_1^{-1}$, where $\Sigma_1 = \mathrm{E}(w_t Z_{t-1}Z'_{t-1})$ and $\Omega_1 = \kappa\mathrm{E}(w_t^2 h_t^2 Z_{t-1}Z'_{t-1})$ with $Z_t = (1, \varepsilon_t^2, \ldots, \varepsilon_{t-r}^2)'$. Note that $X_{2t} = Z_t/h_t$ in this special case. In general, the QMLE is more efficient than the weighted L^2-estimator. In fact, for any $m \times 1$ nonzero constant vectors b and c,

$$\begin{aligned} c'\Sigma_1 bb'\Sigma_1 c &= \left\{ \mathrm{E}\left[(w_t h_t c' Z_{t-1})\left(\frac{1}{h_t} b' Z_{t-1}\right)\right]\right\}^2 \\ &\leqslant \mathrm{E}(w_t h_t c' Z_{t-1})^2 \mathrm{E}\left(\frac{1}{h_t} b' Z_{t-1}\right)^2 = c'\Omega_1 cb'\Sigma b/\kappa. \end{aligned}$$

Thus, $\Sigma_1 bb'\Sigma_1 \leqslant \Omega_1 b'\Sigma b/\kappa$ and hence $b'\Sigma_1\Omega_1^{-1}\Sigma_1 b = \mathrm{tr}(\Omega_1^{-1/2}\Sigma_1 bb'\Sigma_1\Omega_1^{-1/2}) \leqslant b'\Sigma b/\kappa$. Thus, $\Sigma_1^{-1}\Omega_1\Sigma_1^{-1} \geqslant \kappa\Sigma^{-1}$. Based on the weight (3.9), $\Sigma_0^{-1}\Omega_0\Sigma_0^{-1} \to \kappa\Sigma^{-1}$ as $C \to \infty$, while $\Sigma_1^{-1}\Omega_1\Sigma_1^{-1} \to \infty$ if $\mathrm{E}\|Z_t\|^2 < \infty$ but $\mathrm{E}\|Z_t\|^4 = \infty$ as $C \to \infty$. Using $\|Z_{t-1}\|/h_t \leqslant$ a constant a.s., we can show that the self-weighted QMLE based on the weight (3.9) with a large C is more efficient than the weighted-L^2 estimator based on the weight $1/(1+\|Z_{t-1}\|^2)$suggested by Horvath and Liese (2004).

The covariance matrices, Σ and Ω, can be estimated by

$$\hat{\Sigma}_n = \frac{1}{n}\sum_{t=1}^{n} \hat{U}_t\hat{U}'_t \quad \text{and} \quad \hat{\Omega}_n = \frac{1}{n}\sum_{t=1}^{n} \hat{U}_t\hat{J}_n\hat{U}'_t, \tag{4.2}$$

where $\hat{U}_t = \tilde{U}_t(\hat{\theta}_n)$ and $\hat{J}_n$ is defined as J with κ_3 and κ replaced by $\hat{\kappa}_{3n}$ and $\hat{\kappa}_n$,

$$\hat{\kappa}_{3n} = \frac{1}{\sqrt{2}n}\sum_{t=1}^{n}\left[\frac{\tilde{\varepsilon}_t(\hat{\theta}_n)}{\sqrt{\tilde{h}_t(\hat{\theta}_n)}}\right]^3 \quad \text{and} \quad \hat{\kappa}_n = \frac{1}{2n}\sum_{t=1}^{n}\frac{\tilde{\varepsilon}_t^4(\hat{\theta}_n)}{\tilde{h}_t^2(\hat{\theta}_n)} - \frac{1}{2}.$$

Using Lemmas A.5–A.6, Lemmas 6 as for Lemma 7, we can show that $\hat{J}_n = J + o_p(1)$, $\hat{\Sigma}_n = \Sigma + o_p(1)$ and $\hat{\Omega}_n = \Omega + o_p(1)$. Thus, by Theorem 4.1, we have the following corollary for testing the null hypothesis H_0, which is defined as in Corollary 3.1:

Corollary 4.1. *Under the assumption of Theorem 4.1, it follows that*

$$W_n = n(\Gamma\hat{\theta}_n - \theta_{10})'\left(\Gamma\hat{\Sigma}_n^{-1}\hat{\Omega}_n\hat{\Sigma}_n^{-1}\Gamma'\right)^{-1}(\Gamma\hat{\theta}_n - \theta_{10}) \longrightarrow_{\mathscr{L}} \chi^2_{p_1},$$

under H_0, *where* $\hat{\Sigma}_n$ *and* $\hat{\Omega}_n$ *are defined as in* (4.2).

5. Concluding remarks

Given a data set, different estimators may give different results in practice. To see if the QMLE should be used, it will be helpful to estimate the tail index of the data. If it is greater than 4, then the global QMLE can be used. If it is in [2, 4], a two-step estimator should be considered, i.e., first obtaining a self-weighted QMLE and then using it to obtain the QMLE via a one-step iteration. When $\eta_t \sim N(0, 1)$, the QMLE is the MLE and hence it is efficient. Theorem 3 with Theorem 3.1 provide an approach to obtain an efficient estimator for the ARMA-GARCH (finite variance)/IGARCH models. Such an approach is novel and has never appeared in the literature before. To make sure if $\eta_t \sim N(0, 1)$, we can perform a test by using the statistic in Koul and Ling (2006). If there is strong evidence that η_t is not normal, we can further perform the adaptive estimator along the lines as in Drost et al. (1997) and Ling and McAleer (2003b), subject to some regular conditions on the density of η_t.

When the tail index is in (0, 2), we should consider only the self-weighted QMLE. But the results will depend on the choice of weights or the constant C in (3.8)–(3.10). As the Co-Editor has noted, we should acknowledge that using other weights may produce different finite sample results and a different asymptotic variance. We do not have a theory to support the choice of the weight or C yet. Which result is more reliable should depend on the features of the data and their source. It remains a difficult problem to set a sense for comparing different estimators and to select the C or other weights such that the estimator is optimal under this sense.

In summary, this paper has proposed a self-weighted QMLE for parameters in the ARMA-GARCH model and showed that it is consistent and asymptotically normal under only a fractional moment condition of GARCH errors. Asymptotic normality of the local QMLE has been established for ARMA model with GARCH and IGARCH errors.

Wald statistics have been investigated for testing linear restrictions on the parameters in the model. In all results, we assume that $E\eta_t^4 < \infty$. This assumption can be relaxed by using the LAD and the self-weighted method in this paper. The self-weighted principle can be applied to other estimators, such as the M- and the quantile-estimators, to other ARCH-type models, such as threshold AR-ARCH models, and to multivariate time series models.

Acknowledgments

The author greatly thanks Co-Editor Peter M. Robinson, the Associate Editor, two referees, Professors V.A. Unkefer and Hao Yu for their invaluable comments and thanks Professors C. Francq and L. Horvath for sending their reprints during the revision of this paper and the Hong Kong RGC (CERG #HKUST6273/03H, #HKUST6428/06H and #HKUST602205/05P) for financial support. This paper is the revised version of Ling (2003), which included the self-weighted LSE for the ARMA-GARCH model and was completed when the author visited the Department of Statistics, Stanford University. The author thanks the department for its hospitality. This early version was presented in the 2004 International Symposium on Forecasting, Sydney, and the 2004 International Chinese Statistical Association, Singapore.

Appendix A. Proofs

Proof of Theorem 2.1. By (2.7) and Jensen's inequality, we can show that $E\ln\|\prod_{i=0}^{i_0-1} A_i\| < 0$. Using the same method as in Bougerol and Picard (1992), we can show that the following representation holds:

$$\varepsilon_t = \eta_t\sqrt{h_t} \quad \text{and} \quad h_t = \alpha_{00}\left[1 + \sum_{j=1}^{\infty} u'_{r+1}\prod_{i=0}^{j-1} A_{t-i}\zeta_{t-j}\right] \quad \text{a.s.,}$$

and hence $\{\varepsilon_t\}$ is strictly stationary and ergodic, $\zeta_t = (\eta_t^2, 0, \ldots, 0, 1, \ldots, 0)'_{(r+s)\times 1}$ with the first component η_t^2 and the $(r+1)$th component 1, and $u_i = (0, \ldots, 0, 1, \ldots, 0)'_{(r+s)\times 1}$ with the ith component 1.

Let $B_t = \zeta_t + \sum_{j=1}^{i_0-1}\prod_{i=0}^{j-1} A_{t-i}\zeta_{t-j}$ and $\tilde{A}_t = \prod_{i=0}^{i_0-1} A_{t-i}$. We rewrite h_t as

$$h_t = \alpha_{00}\left[u'_{r+1}B_t + \sum_{k=1}^{\infty} u'_{r+1}\prod_{i_1=0}^{k-1}\tilde{A}_{t-i_0 i_1}B_{t-ki_0}\right], \tag{A.1}$$

where $u'_{r+1}\zeta_t = 1$ is used. By (A.1), it follows that

$$Eh_t^{\iota} \leqslant O(1) + O(1)\sum_{k=1}^{\infty}(E\|\tilde{A}_t\|^{\iota})^k < \infty.$$

Thus, $E|\varepsilon_t|^{2\iota} < \infty$ and hence (i) holds.

Note that model (2.2) has only one strictly stationary solution with the representation (A.1) (see Bougerol and Picard, 1992). Denote $h_{Jt} = \alpha_{00}(1 + \sum_{j=1}^{J} u'_{r+1}\prod_{i=0}^{j-1} A_{t-i}\xi_{t-j})$. Then $(h_{Jt} - h_{J-1,t})^{\iota} \to 0$ a.s. when $J \to \infty$. $\{h_{Jt}^{\iota}\}$ is an increasing sequence in terms of J and $E\sup_J h_{Jt}^{\iota} \leqslant Eh_t^{\iota} < \infty$. Thus, by the dominated convergence theorem, we have $E(u'_{r+1}\prod_{i=0}^{J-1} A_{t-i}\zeta_{t-j})^{\iota} = E(h_{Jt} - h_{J-1,t})^{\iota}/\alpha_{00}^{\iota} \to 0$ as $J \to \infty$. Since $\tilde{\beta} > 0$, there exist J_1 and J_2 such that all the elements of $d_{1t} \equiv (u'_{r+1}\prod_{i=0}^{J_1-1} A_{t-i})'$ and $d_{2t} \equiv \prod_{i=0}^{J_2-1} A_{t-i}\zeta_{t-j}$are a.s.

positive, and $0<\mathrm{E}d_{1it}^{\iota}<\infty$ and $0<\mathrm{E}d_{2it}^{\iota}<\infty$, where d_{jit} is the ith element of d_{jt} as $j=1,2$. Note that

$$\mathrm{E}\left(u_{r+1}'\prod_{i=0}^{J_1+i_0+J_2-1}A_{t-i}\zeta_{t-j}\right)^{\iota}=\mathrm{E}\left[d_{1t}'\left(\prod_{i=J_1}^{J_1+i_0-1}A_{t-i}\right)d_{2,J_1+i_0}\right]^{\iota}\to 0,$$

as $i_0\to\infty$. Let a_{ijt} be the (i,j)-element of $\prod_{i=J_1}^{J_1+i_0-1}A_{t-i}$. From the preceding equation, we know that $\mathrm{E}(d_{1it}a_{ijt}d_{2j,J_1+i_0})^{\iota}\to 0$ as $i_0\to\infty$. Since d_{1it}, a_{ijt}, and d_{2j,J_1+i_0} are independent, we know that $\mathrm{E}a_{ijt}^{\iota}\to 0$ as $i_0\to\infty$. Thus, there exists i_0 such that $\mathrm{E}\|\prod_{i=J_1}^{J_1+i_0-1}A_{t-i}\|^{\iota}$ $=\mathrm{E}\|\prod_{i=0}^{i_0-1}A_{t-i}\|^{\iota}<1$, i.e., (ii) holds.

(iii) Since (2.6) is the necessary and sufficient condition for $\mathrm{E}\varepsilon_t^2<\infty$, (2.8) implies $\mathrm{E}\varepsilon_t^2=\infty$. By (ii) of this theorem, $\mathrm{E}\|\prod_{k=0}^{i_0-1}A_k\|\geqslant 1$ for any $i_0\geqslant 1$. Thus,

$$\ln\mathrm{E}\|\prod_{k=1}^{n}A_k\|\geqslant 0,$$

for all $n\geqslant 1$. Note that $u_i'\prod_{k=1}^{n}A_ku_j$ is the (i,j)th element of $\prod_{k=1}^{n}A_k$. We have

$$\begin{aligned}\mathrm{E}\left\|\prod_{k=1}^{n}A_k\right\|&=\mathrm{E}\left[\sum_{i=1}^{r+s}\sum_{j=1}^{r+s}\left(u_i'\prod_{k=1}^{n}A_ku_j\right)^2\right]^{1/2}\\&\leqslant\mathrm{E}\left[\sum_{i=1}^{r+s}\sum_{j=1}^{r+s}\left(u_i'\prod_{k=1}^{n}A_ku_j\right)\right]=\sum_{i=1}^{r+s}\sum_{j=1}^{r+s}(u_i'A^nu_j)\leqslant(r+s)^2,\end{aligned}$$

where $A=\mathrm{E}A_t$. By the previous two inequalities, it follows that

$$\lim_{n\to\infty}\frac{1}{n}\ln\,\mathrm{E}\|A_1\ldots A_n\|=0.$$

By Theorems 2.4 and 3.1 (B) in Basrak et al. (2002), the conclusion (iii) holds. This completes the proof. □

We next give some basic formulas as follows:

$$\begin{aligned}\frac{\partial\,\varepsilon_t(\gamma)}{\partial\mu}&=-\psi^{-1}(1),\\ \frac{\partial\,\varepsilon_t(\gamma)}{\partial\,\phi_j}&=-\psi^{-1}(B)y_{t-j},\quad 1\leqslant j\leqslant p,\\ \frac{\partial\,\varepsilon_t(\gamma)}{\partial\psi_j}&=-\psi^{-1}(B)\varepsilon_{t-j}(\theta),\quad 1\leqslant j\leqslant q,\\ h_t(\theta)&=\frac{\alpha_0}{\beta(1)}+\beta^{-1}(B)\alpha(B)\varepsilon_t^2(\gamma),\\ \frac{\partial\,h_t(\theta)}{\partial\delta}&=\beta^{-1}(B)\tilde{\varepsilon}_t(\theta),\\ \frac{\partial h_t(\theta)}{\partial\gamma}&=2\beta^{-1}(B)\alpha(B)\left[\varepsilon_t(\gamma)\frac{\partial\varepsilon_t(\gamma)}{\partial\gamma}\right],\end{aligned}$$

where $\tilde{\varepsilon}_t(\theta)=[1,\varepsilon_{t-1}^2(\gamma),\ldots,\varepsilon_{t-r}^2(\gamma),h_{t-1}(\theta),\ldots,h_{t-s}(\theta)]'$ and B is the back-shift operator. Similarly, we can write down the formula for $\partial^2\varepsilon_t(\gamma)/\partial\gamma_{i_1}\partial\gamma_{i_2}$ and $\partial^2h_t(\theta)\,/\partial\theta_{j_1}\partial\theta_{j_2}$, where $i_1,i_2=1,\ldots,p+q+1$ and $j_1,j_2=1,\ldots,m$. We further give the quasi-score function and the quasi-information matrix as follows.

$$\frac{\partial l_t(\theta)}{\partial \theta} = -\frac{\varepsilon_t(\gamma)}{h_t(\theta)}\frac{\partial \varepsilon_t(\gamma)}{\partial \theta} + \frac{1}{2h_t(\theta)}\left(\frac{\varepsilon_t^2(\gamma)}{h_t(\theta)} - 1\right)\frac{\partial h_t(\theta)}{\partial \theta}, \tag{A.2}$$

$$\begin{aligned}\frac{\partial^2 l_t(\theta)}{\partial \gamma\,\partial \gamma'} = & -\frac{1}{h_t(\theta)}\frac{\partial\,\varepsilon_t(\gamma)}{\partial \gamma}\frac{\partial\,\varepsilon_t(\gamma)}{\partial \gamma'} - \frac{1}{2h_t^2(\theta)}\frac{\partial h_t(\theta)}{\partial \gamma}\frac{\partial h_t(\theta)}{\partial \gamma'}\frac{\varepsilon_t^2(\gamma)}{h_t(\theta)} \\ & + \frac{1}{2}\left[\frac{\varepsilon_t^2(\gamma)}{h_t(\theta)} - 1\right]\left[-\frac{1}{h_t^2(\theta)}\frac{\partial h_t(\theta)}{\partial \gamma}\frac{\partial h_t(\theta)}{\partial \gamma'} + \frac{1}{h_t(\theta)}\frac{\partial^2 h_t(\theta)}{\partial \gamma\,\partial \gamma'}\right] \\ & - \frac{2\varepsilon_t(\gamma)}{h_t(\theta)}\frac{\partial\,\varepsilon_t(\gamma)}{\partial \gamma}\frac{\partial\,h_t(\theta)}{\partial \gamma'} - \frac{\varepsilon_t(\gamma)}{h_t(\theta)}\frac{\partial^2\,\varepsilon_t(\gamma)}{\partial \gamma\,\partial \gamma'},\end{aligned} \tag{A.3}$$

$$\begin{aligned}\frac{\partial^2 l_t(\theta)}{\partial \delta\,\partial \delta'} = & -\frac{1}{2h_t^2(\theta)}\frac{\partial h_t(\theta)}{\partial \delta}\frac{\partial h_t(\theta)}{\partial \delta'}\frac{\varepsilon_t^2(\gamma)}{h_t(\theta)} \\ & + \frac{1}{2}\left[\frac{\varepsilon_t^2(\gamma)}{h_t(\theta)} - 1\right]\left[-\frac{1}{h_t^2(\theta)}\frac{\partial h_t(\theta)}{\partial \delta}\frac{\partial h_t(\theta)}{\partial \delta'} + \frac{1}{h_t(\theta)}\frac{\partial^2 h_t(\theta)}{\partial \delta\,\partial \delta'}\right].\end{aligned} \tag{A.4}$$

Similarly, we can write down $\partial^2 l_t(\theta)/(\partial\gamma\,\partial\delta')$. We now give three basic Lemmas. They are commonly used in the proof.

Lemma A.1. *Let $\xi_{\rho t}$ be defined as in* (3.4). *If Assumptions* 2.1–2.2 *hold, then there exist constants C and $\rho \in (0,1)$ such that the following holds uniformly in Θ*:

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

(ii) $h_t(\theta)$ *is bounded a.s. by* $C\xi_{\rho t-1}^2$.

Proof. It directly comes from (2.3) and (2.5). This completes the proof. □

Lemma A.2. *Let $\xi_{\rho t}$ be defined as in* (3.4). *Under Assumptions* 2.1–2.2, *there exists a neighborhood Θ_0 of θ_0 and a constant $\rho \in (0,1)$ such that*

(i) $\sup_{\Theta_0}\left\|\frac{1}{h_t(\theta)}\frac{\partial h_t(\theta)}{\partial \delta}\right\| \leqslant C\xi_{\rho t-1}^{\iota_1}$,

(ii) $\sup_{\Theta_0}\left\|\frac{1}{h_t(\theta)}\frac{\partial^2 h_t(\theta)}{\partial \delta\,\partial \delta'}\right\| \leqslant C\xi_{\rho t-1}^{\iota_1}$,

for any $\iota_1 \in (0,1)$, where C is a constant independent of ι_1 and t.

Proof. Let G be defined as in (2.4). It is not difficult to show that

$$h_t(\theta) = C_\theta + \sum_{i=1}^{r}\sum_{j=1}^{\infty}\alpha_i u' G^j u \varepsilon_{t-i-j}^2(\gamma), \tag{A.5}$$

where $C_\theta = \alpha_0/(1-\sum_{i=1}^{s}\beta_s)$ and $u' = (1,0,\ldots,0)_{s\times 1}$. Similarly, we can show that

$$\begin{aligned}\frac{\partial h_t(\theta)}{\partial \beta_k} &= \sum_{j=1}^{\infty} u' G^j u h_{t-k-j}(\theta)\\ &= C_1 + \sum_{i=1}^{r}\sum_{j=1}^{\infty}\sum_{j_1=1}^{\infty} \alpha_i u' G^j u u' G^{j_1} u \varepsilon^2_{t-k-i-j-j_1}(\gamma), \end{aligned} \tag{A.6}$$

where C_1 is some constant. Note that $x(1+x)^{-1} < x^{\iota_1}$ as $x>0$ for any $\iota_1 \in (0,1)$. We have, for any $\iota_1 \in (0,1)$,

$$\begin{aligned} C(k,i,j,j_1) \equiv \max_{\Theta}&\left[\frac{\alpha_i u' G^{k+j+j_1} u \varepsilon^2_{t-k-i-j-j_1}(\gamma)}{C_\theta + \alpha_i u' G^{k+j+j_1} u \varepsilon^2_{t-k-i-j-j_1}(\gamma)}\right]\\ &\leqslant \max_{\Theta}[\alpha_i u' G^{k+j+j_1} u \varepsilon^2_{t-k-i-j-j_1}(\gamma)/C_\theta]^{\iota_1}\\ &\leqslant C_2 \rho^{j+j_1} |\varepsilon_{t-k-i-j-j_1}(\gamma)|^{\iota_1}, \end{aligned} \tag{A.7}$$

where C_2 is a constant. Since θ_0 is an interior point in Θ, there exists a neighborhood Θ_0 of θ_0 such that $\underline{\beta} = \min\{\beta_1 : \theta \in \Theta_0\} > 0$. Furthermore, because each element of G is nonnegative, it is easy to see that, for any constant vector c with all elements being nonnegative, $c'uu'c \leqslant c'Gc/\underline{\beta}$, and hence $u'G^j uu'G^{j_1}u \leqslant u'G^j G G^{j_1} u/\underline{\beta} = u'G^{j+j_1+1}u/\underline{\beta}$. Again, since all the elements of G are nonnegative, it follows that

$$\begin{aligned} u'G^{k+j+j_1}u = u'GG^{k+j+j_1-1}u &= (\beta_1,\ldots,\beta_s)G^{k+j+j_1-1}u\\ &\geqslant \underline{\beta}\, u'G^{k+j+j_1-1}u\\ &\geqslant \cdots \geqslant \underline{\beta}^{k-1} u'G^{j+j_1+1}u \geqslant \underline{\beta}^k u'G^j uu'G^{j_1}u. \end{aligned} \tag{A.8}$$

Thus, by (A.5) and (A.8), we have

$$\begin{aligned} &\max_{\Theta_0}\left[\frac{\alpha_i u'G^j uu'G^{j_1}u\varepsilon^2_{t-k-i-j-j_1}(\gamma)}{h_t(\theta)}\right]\\ &\leqslant \max_{\Theta_0}\left[\frac{\alpha_i u'G^j uu'G^{j_1}u\varepsilon^2_{t-k-i-j-j_1}(\gamma)}{C_\theta + \alpha_i u'G^{k+j+j_1}u\varepsilon^2_{t-k-i-j-j_1}(\theta)}\right] \leqslant \frac{1}{\underline{\beta}^k} C(k,i,j,j_1). \end{aligned} \tag{A.9}$$

By (A.6), (A.7) and (A.9), for any $\iota_1 \in (0,1)$, it follows that

$$\sup_{\Theta_0}\left|\frac{1}{h_t(\theta)}\frac{\partial h_t(\theta)}{\partial \beta_k}\right| \leqslant \frac{C_4}{\underline{\beta}^k}\left(1 + \sum_{j=1}^{\infty}\rho^j|\varepsilon_{t-k-j}(\gamma)|^{\iota_1}\right), \quad k=1,\ldots,s, \tag{A.10}$$

where $\rho \in (0,1)$ and $C>0$ are some constants. Similarly, we can show that

$$\sup_{\Theta_0}\left|\frac{1}{h_t(\theta)}\frac{\partial h_t(\theta)}{\partial \alpha_k}\right| \leqslant C\left(1 + \sum_{j=1}^{\infty}\rho^j\left|\varepsilon_{t-k-j}(\gamma)\right|^{\iota_1}\right), \quad k=1,\ldots,r. \tag{A.11}$$

Let ρ be the maximizer of the ρ in (A.11) and Lemma A.1(i). By Lemma A.1 (i),

$$\begin{aligned}\sum_{j=1}^{\infty}\rho^j|\varepsilon_{t-k-j}(\gamma)|^{\iota_1} &\leqslant C\sum_{j=1}^{\infty}\rho^j|\xi_{\rho t-k-j}|^{\iota_1}\\ &\leqslant C\left(1+\sum_{j=1}^{\infty}\rho^j|\xi_{\rho t-k-j}|\right)^{\iota_1}\left[\sum_{j=1}^{\infty}\rho^j\frac{|\xi_{\rho t-k-j}|^{\iota_1}}{(1+\sum_{j=1}^{\infty}\rho^j|\xi_{\rho t-k-j}|)^{\iota_1}}\right]\\ &\leqslant C\left(1+\sum_{j=1}^{\infty}\rho^j|\xi_{\rho t-k-j}|\right)^{\iota_1}\sum_{j=1}^{\infty}\rho^{(1-\iota_1)j}\\ &\leqslant C\left(1+\sum_{j=1}^{\infty}\rho^j\sum_{l=0}^{\infty}\rho^l|y_{t-k-j-l}|\right)^{\iota_1}\leqslant C\xi_{\tilde\rho t-k-1}^{\iota_1},\end{aligned}\tag{A.12}$$

for some $\tilde\rho\in(0,1)$, where the constant C may be different in different inequalities. Thus, using (A.10)–(A.12), we can claim that (i) holds. Similarly, we can show that (ii) holds. This completes the proof. $\square$

Lemma A.3. *Let $\xi_{\rho t}$ be defined as in* (3.4). *Under Assumptions* 2.1–2.2, *there exist constants C and $\rho\in(0,1)$ such that*

$$\text{(i)}\quad \sup_{\Theta}\left\|\frac{1}{\sqrt{h_t(\theta)}}\frac{\partial h_t(\theta)}{\partial\gamma}\right\|\leqslant C\xi_{\rho t-1},$$

$$\text{(ii)}\quad \sup_{\Theta}\left\|\frac{1}{\sqrt{h_t(\theta)}}\frac{\partial^2 h_t(\theta)}{\partial\gamma\,\partial\gamma'}\right\|\leqslant C\xi_{\rho t-1},$$

$$\text{(iii)}\quad \sup_{\Theta}\left\|\frac{1}{\sqrt{h_t(\theta)}}\frac{\partial^2 h_t(\theta)}{\partial\delta\,\partial\gamma'}\right\|\leqslant C\xi_{\rho t-1}.$$

Proof. By (2.5), we have the following expansions:

$$h_t(\theta)=\alpha_0\beta^{-1}(1)+\sum_{i=1}^{\infty}a_\delta(i)\varepsilon_{t-i}^2(\gamma),$$

$$\frac{\partial h_t(\theta)}{\partial\gamma}=2\sum_{i=1}^{\infty}a_\delta(i)\left[\varepsilon_{t-i}(\gamma)\frac{\partial\varepsilon_{t-i}(\gamma)}{\partial\gamma}\right].$$

Thus, $h_t(\theta)\geqslant a_\delta(i)\varepsilon_{t-i}^2(\gamma)$. By Lemma A.1(i), it follows that

$$\begin{aligned}\sup_{\Theta}\left\|\frac{1}{\sqrt{h_t(\theta)}}\frac{\partial h_t(\theta)}{\partial\gamma}\right\| &= 2\sup_{\Theta}\left\|\sum_{i=1}^{\infty}\frac{a_\delta(i)\varepsilon_{t-i}(\gamma)}{\sqrt{h_t(\theta)}}\frac{\partial\varepsilon_{t-i}(\gamma)}{\partial\gamma}\right\|\\ &\leqslant 2\sup_{\Theta}\left\|\sum_{i=1}^{\infty}\sqrt{a_\delta(i)}\frac{\partial\varepsilon_{t-i}(\gamma)}{\partial\gamma}\right\|\leqslant\xi_{\rho t-1},\end{aligned}$$

i.e., (i) holds. Similarly, we can show that (ii)–(iii) hold. This completes the proof. $\square$

Lemma A.4. *Under Assumptions* 2.1–2.2, *for any* $\imath_1 \in (0,1)$, *it follows that*

$$\text{(i)}\quad \sup_{\Theta} |l_t(\theta) - \tilde{l}_t(\theta)| \leqslant \mathrm{O}(\rho^t) R_t^{1+\imath_1},$$

$$\text{(ii)}\quad \sup_{\Theta} \left\| \frac{\partial l_t(\theta)}{\partial \theta} - \frac{\partial \tilde{l}_t(\theta)}{\partial \theta} \right\| \leqslant \mathrm{O}(\rho^t) R_t^{2+\imath_1},$$

$$\text{(iii)}\quad \sup_{\Theta} \left\| \frac{\partial^2 l_t(\theta)}{\partial \theta\, \partial \theta'} - \frac{\partial^2 \tilde{l}_t(\theta)}{\partial \theta\, \partial \theta'} \right\| \leqslant \mathrm{O}(\rho^t) R_t^{3+\imath_1},$$

where $R_t = 1 + \sum_{i=0}^{t} \xi_{\rho t-i}^2$ *and* $\xi_{\rho t}$ *is defined as in* (3.4) *for some* $\rho \in (0,1)$.

Proof. It is straightforward to show that there exists a $\rho \in (0,1)$ such that

$$\sup_{\Theta_\gamma} |\varepsilon_t(\gamma) - \tilde{\varepsilon}_t(\gamma)| = \mathrm{O}(\rho^t)\xi_{\rho 0},$$

$$\sup_{\Theta_\gamma} |\varepsilon_t^2(\gamma) - \tilde{\varepsilon}_t^2(\gamma)| = \mathrm{O}(\rho^t)\xi_{\rho t}\xi_{\rho 0},$$

$$\begin{aligned}\sup_{\Theta} |h_t(\theta) - \tilde{h}_t(\theta)| &\leqslant \mathrm{O}(1) \sum_{i=1}^{t-1} \rho^i \sup_{\Theta_\gamma} |\varepsilon_{t-i}^2(\gamma) - \tilde{\varepsilon}_{t-i}^2(\gamma)| + \mathrm{O}(\rho^t)\xi_{\rho 0}^2 \\ &\leqslant \mathrm{O}(\rho^t)\xi_{\rho 0} \sum_{i=1}^{t-1} \xi_{\rho t-i} + \mathrm{O}(\rho^t)\xi_{\rho 0}^2 \leqslant \mathrm{O}(\rho^t) R_t,\end{aligned}$$

$$\begin{aligned}\sup_{\Theta} \left| \frac{1}{h_t^k(\theta)} - \frac{1}{\tilde{h}_t^k(\theta)} \right| &\leqslant \mathrm{O}(1) \sup_{\Theta} \left| \frac{1}{h_t^k(\theta)} - \frac{1}{\tilde{h}_t^k(\theta)} \right|^{\imath_1} \\ &\leqslant \mathrm{O}(1) \sup_{\Theta} |h_t(\theta) - \tilde{h}_t(\theta)|^{\imath_1} = \mathrm{O}(\rho^t) R_t^{\imath_1},\end{aligned}$$

for any $\imath_1 \in (0,1)$ and $k \geqslant 1$. Note that $\sup_{\Theta} |\tilde{\varepsilon}_t^2(\gamma)| \leqslant \mathrm{O}(1)\xi_{\rho t}^2$. There is a constant $C > 0$ such that

$$\begin{aligned}|l_t(\theta) - \tilde{l}_t(\theta)| &\leqslant \log[1 + C|h_t(\theta) - \tilde{h}_t(\theta)|] \\ &\quad + \frac{1}{h_t} |\varepsilon_t^2(\gamma) - \tilde{\varepsilon}_t^2(\gamma)| + \tilde{\varepsilon}_t^2(\gamma) \left| \frac{1}{h_t} - \frac{1}{\tilde{h}_t} \right|.\end{aligned}$$

By the preceding inequalities, we can show that (i) holds. Furthermore, we have

$$\sup_{\Theta_\gamma} \left\| \frac{\partial \varepsilon_t(\gamma)}{\partial \gamma} - \frac{\partial \tilde{\varepsilon}_t(\gamma)}{\partial \gamma} \right\| = \mathrm{O}(\rho^t)\xi_{\rho 0},$$

$$\begin{aligned}\sup_{\Theta} \left\| \frac{\partial h_t(\theta)}{\partial \gamma} - \frac{\partial \tilde{h}_t(\theta)}{\partial \gamma} \right\| &\leqslant \mathrm{O}(1) \sum_{i=1}^{t-1} \rho^i \sup_{\Theta_\gamma} \left\| \varepsilon_{t-i}(\gamma) \frac{\partial \varepsilon_{t-i}(\gamma)}{\partial \gamma} \right. \\ &\qquad \left. - \tilde{\varepsilon}_{t-i}(\gamma) \frac{\partial \tilde{\varepsilon}_{t-i}(\gamma)}{\partial \gamma} \right\| + \mathrm{O}(\rho^t)\xi_{\rho 0}^2 \\ &\leqslant \mathrm{O}(\rho^t)\xi_{\rho 0} \sum_{i=1}^{t-1} (\xi_{\rho t-i} + \xi_{\rho t-i-1}) + \mathrm{O}(\rho^t)\xi_{\rho 0}^2 \leqslant \mathrm{O}(\rho^t) R_t,\end{aligned}$$

$$\sup_{\Theta}\left\|\frac{\partial h_t(\theta)}{\partial\delta}-\frac{\partial\tilde{h}_t(\theta)}{\partial\delta}\right\|\leqslant O(1)\sum_{i=1}^{t-1}\rho^i\sup_{\Theta_\gamma}\left|\varepsilon_{t-i}^2(\gamma)-\tilde{\varepsilon}_{t-i}^2(\gamma)\right|+O(\rho^t)\xi_{\rho 0}^2$$
$$\leqslant O(\rho^t)\xi_{\rho 0}\sum_{i=1}^{t-1}\xi_{\rho t-i}+O(\rho^t)\xi_{\rho 0}^2\leqslant O(\rho^t)R_t.$$

Using these inequalities, we can show that (ii)–(iii) hold. This completes the proof. □

Proof of Theorem 3.1(i). First, the space Θ is compact and θ_0 is an interior point in Θ. Second, $\tilde{L}_{sn}(\theta)$ is continuous in $\theta\in\Theta$ and is a measurable function of $\{y_i,\ i=t,t-1,\ldots\}$ for all $\theta\in\Theta$.

Third, by Lemma A.1, it follows that

$$\sup_{\Theta}|\varepsilon_t(\gamma)|\leqslant|\varepsilon_t(\gamma_0)|+\sup_{\Theta}\|\gamma-\gamma_0\|\sup_{\Theta}\left\|\frac{\partial\varepsilon_t(\gamma)}{\partial\gamma}\right\|$$
$$\leqslant|\eta_t|\sqrt{h_t(\theta_0)}+O(1)\xi_{\rho t-1}\leqslant O(1)|\eta_t|(1+\xi_{\rho t-1}),\tag{A.13}$$

$$1\leqslant\sup_{\Theta}\frac{h_t(\theta)}{\underline{\alpha}_0}\leqslant O(1)\xi_{\rho t-1}^2,\tag{A.14}$$

for some $\rho\in(0,1)$, where $\underline{\alpha}_0=\inf_\Theta\{\alpha_0:\theta\in\Theta\}$. By Assumption 3.1 and (A.13)–(A.14), $E\sup_\Theta[w_t\varepsilon_t^2(\gamma)/h_t(\theta)]<\infty$. Since w is a bounded function, by Jensen's inequality, Assumption 2.4 and (A.14), $E\sup_\Theta|w_t\log h_t(\theta)|\leqslant O(1)E\sup_\Theta|\log h_t^{\iota/2}(\theta)|\leqslant O(1)[\log E\sup_\Theta(h_t(\theta)/\underline{\alpha}_0)^{\iota/2}+|\log\underline{\alpha}_0|]<\infty$. Thus, we can claim that $E\sup_\Theta|w_t l_t(\theta)|<\infty$. By the ergodic theorem, $L_{sn}(\theta)\to E[w_t l_t(\theta)]$ a.s. for each $\theta\in\Theta$. Furthermore, by Theorem 3.1 in Ling and McAleer (2003a), it follows that

$$\sup_{\Theta}|L_{sn}(\theta)-E[w_t l_t(\theta)]|\to_p 0.\tag{A.15}$$

It is straightforward to show that $\sup_\Theta[|l_t(\theta)|+|\tilde{l}_t(\theta)|]\leqslant O(1)\xi_{\rho t}^2$. By Lemma A.4 (i),

$$\sup_{\Theta}|w_t l_t(\theta)-\tilde{w}_t\tilde{l}_t(\theta)|\leqslant w_t\sup_{\Theta}|l_t(\theta)-\tilde{l}_t(\theta)|+|w_t-\tilde{w}_t|\sup_{\Theta}\tilde{l}_t(\theta)|$$
$$\leqslant O(1)\rho^t R_t^{1+\iota_1}+O(1)|w_t-\tilde{w}_t|\xi_{\rho t}^2,$$
$$E(|w_t-\tilde{w}_t|\xi_{\rho t}^2)^{\iota_0/8}\leqslant\{E|w_t-\tilde{w}_t|^{\iota_0/4}E\xi_{\rho t}^{\iota_0/2}\}^{1/2}=O(t^{-1}).$$

By the preceding inequality, and Assumptions 2.4 and 3.2, we can show that

$$\frac{1}{n}\sum_{t=1}^{n}\sup_{\Theta}|w_t l_t(\theta)-\tilde{w}_t\tilde{l}_t(\theta)|=o_p(1).\tag{A.16}$$

By (A.15)–(A.16), it follows that

$$\sup_{\Theta}|\tilde{L}_{sn}(\theta)-E[w_t l_t(\theta)]|=o_p(1).\tag{A.17}$$

Fourth, by (2.3), it follows that

$$\varepsilon_t(\gamma) - \varepsilon_t(\gamma_0) = \sum_{i=1}^{\infty}[a_\gamma(i) - a_{\gamma_0}(i)]y_{t-i}$$
$$= (\gamma - \gamma_0)' \sum_{i=1}^{\infty} a_{\gamma^*}(i)y_{t-i} = (\gamma - \gamma_0)' \frac{\partial \varepsilon_t(\gamma^*)}{\partial \gamma},$$

where γ^* is between θ and θ_0. Thus,

$$\mathrm{E}[w_t l_t(\theta)] = -Ew_t \log h_t(\theta) - \mathrm{E}\frac{w_t[\varepsilon_t(\gamma_0) + (\gamma-\gamma_0)'\partial\varepsilon_t(\theta^*)/\partial\theta]^2}{h_t(\theta)}$$
$$= \left\{-\mathrm{E}w_t \log h_t(\theta) - \mathrm{E}\frac{w_t h_t(\theta_0)}{h_t(\theta)}\right\}$$
$$- (\gamma-\gamma_0)'\mathrm{E}\left[\frac{w_t}{h_t(\theta)}\frac{\partial\varepsilon_t(\theta^*)}{\partial\gamma}\frac{\partial\varepsilon_t(\theta^*)}{\partial\gamma'}\right](\gamma-\gamma_0) \equiv L_1(\theta) + L_2(\theta).$$

$L_2(\theta)$ obtains its maximum at zero if and only if $\gamma = \gamma_0$, and

$$L_1(\theta) = \mathrm{E}w_t\left[\log\frac{h_t(\theta_0)}{h_t(\theta)} - \frac{h_t(\theta_0)}{h_t(\theta)}\right] - \mathrm{E}w_t \log h_t(\theta_0).$$

Note that, for any $M > 0$, $g(M) \equiv \log M - M \leqslant -1$ with equality only if $M = 1$. Let $M_t = h_t(\theta_0)/h_t(\theta)$. When $M_t = 1$ a.s., we have $\mathrm{E}[w_t g(M_t)] = -\mathrm{E}w_t$. If $P(M_t = 1) \neq 1$, then $P(g(M_t) < g(1)) \neq 0$, so that $\mathrm{E}[w_t g(M_t)] < \mathrm{E}[w_t g(1)] = -\mathrm{E}w_t$. Thus, $L_1(\theta)$ reaches at its maximum $-1 - \mathrm{E}[w_t \log h_t(\theta_0)]$, and this occurs if and only if $h_t(\theta) = h_t(\theta_0)$. Since $\max_\Theta \mathrm{E}[w_t l_t(\theta)] \leqslant \max_\Theta L_1(\theta) + \max_\Theta L_2(\theta)$, $\max_\Theta \mathrm{E}[w_t l_t(\theta)] = -1 - \mathrm{E}[w_t \log h_t(\theta_0)]$ if and only if $\max_\Theta L_2(\theta) = 0$ and $\max_\Theta L_1(\theta) = -1 - \mathrm{E}[w_t \log h_t(\theta_0)]$, which occurs if and only if $\gamma = \gamma_0$ and $h_t(\theta) = h_t(\theta_0)$ a.s. (see e.g., Francq and Zakoïan, 2004). Since $h_t(\theta)|_{\gamma=\gamma_0} = h_t(\theta_0)$ a.s. if and only if $\delta = \delta_0$, we can claim that $L_1(\theta)$ reaches its maximum $-1 - \mathrm{E}\log h_t(\theta_0)$ if and only if $\theta = \theta_0$. Thus, $\mathrm{E}[w_t l_t(\theta)]$ is uniquely maximized at θ_0.

Thus, we have established all the conditions for consistency in Theorem 4.1.1 in Amemiya (1985) and hence (i) holds. This completes the proof. □

Proof of Theorem 3.1(ii). First, $\hat{\theta}_{sn} \to_p \theta_0$ as $n \to \infty$. Second, $\partial^2 l_t(\theta)/\partial\theta\,\partial\theta'$ exists and is continuous in Θ. Third, by (A.3), (A.13) and Lemmas A.1 and A.3,

$$\sup_\Theta\left\|\frac{\partial^2 l_t(\theta)}{\partial\gamma\,\partial\gamma'}\right\| \leqslant O(1)\sup_\Theta\left\{\left\|\frac{\partial\,\varepsilon_t(\gamma)}{\partial\gamma}\right\|^2 + \left\|\frac{1}{\sqrt{h_t(\theta)}}\frac{\partial h_t(\theta)}{\partial\gamma}\right\|^2 \varepsilon_t^2(\gamma)\right.$$
$$+ \left(\varepsilon_t^2(\gamma)+1\right)\left(\left\|\frac{1}{\sqrt{h_t(\theta)}}\frac{\partial h_t(\theta)}{\partial\gamma}\right\|^2 + \left\|\frac{1}{\sqrt{h_t(\theta)}}\frac{\partial^2 h_t(\theta)}{\partial\gamma\,\partial\gamma'}\right\|\right)$$
$$\left. + |\varepsilon_t(\gamma)|\left\|\frac{\partial\,\varepsilon_t(\gamma)}{\partial\gamma}\right\|\left\|\frac{1}{\sqrt{h_t(\theta)}}\frac{\partial\,h_t(\theta)}{\partial\gamma'}\right\| + |\varepsilon_t(\gamma)|\left\|\frac{\partial^2\,\varepsilon_t(\gamma)}{\partial\gamma\,\partial\gamma'}\right\|\right\}$$

$$\leqslant O(1)\Big\{\xi_{\rho t-1}^2+\xi_{\rho t-1}^4(1+|\eta_t|)^2+[\xi_{\rho t-1}^2(1+|\eta_t|)^2+1]\times(\xi_{\rho t-1}^2+\xi_{\rho t-1})+\xi_{\rho t-1}(1+|\eta_t|)(\xi_{\rho t-1}^2+\xi_{\rho t-1})]\Big\}\leqslant O(1)\xi_{\rho t-1}^4(1+\eta_t^2), \tag{A.18}$$

where O(1) holds uniformly in t. By (A.4), (A.13) and Lemma A.2, there exists a neighborhood Θ_0 of θ_0 such that

$$\sup_{\Theta_0}\left\|\frac{\partial^2 l_t(\theta)}{\partial\delta\,\partial\delta'}\right\|\leqslant O(1)\sup_{\Theta_0}\left\{\left\|\frac{1}{h_t(\theta)}\frac{\partial h_t(\theta)}{\partial\delta}\right\|^2\varepsilon_t^2(\gamma)+(\varepsilon_t^2(\gamma)+1)\left[\left\|\frac{1}{h_t(\theta)}\frac{\partial h_t(\theta)}{\partial\delta}\right\|^2+\left\|\frac{1}{h_t(\theta)}\frac{\partial^2 h_t(\theta)}{\partial\delta\,\partial\delta'}\right\|\right]\right\}$$
$$\leqslant O(1)\{\xi_{\rho t-1}^4(1+|\eta_t|)^2+[\xi_{\rho t-1}^2(1+|\eta_t|)^2+1](\xi_{\rho t-1}+\xi_{\rho t-1}^2)\},$$

where O(1) holds uniformly in t. Now, by Assumption 3.1, it is readily seen that $E\sup_{\Theta}\|w_t\partial^2 l_t(\theta)/\partial\gamma\,\partial\gamma'\|<\infty$ and $E\sup_{\Theta_0}\|w_t\partial^2 l_t(\theta)/\partial\delta\,\partial\delta'\|<\infty$. Similarly, we can show that $E\sup_{\Theta_0}\|w_t\partial^2 l_t(\theta)/\partial\gamma\,\partial\delta'\|<\infty$. Thus, we can claim that

$$E\sup_{\Theta_0}\left\|w_t\frac{\partial^2 l_t(\theta)}{\partial\theta\,\partial\theta'}\right\|<\infty.$$

By the ergodic theorem and Theorem 3.1 in Ling and McAleer (2003a), we can show that $\partial^2 L_{sn}(\theta)/\partial\theta\,\partial\theta'$ converges to $E[w_t\partial^2 l_t(\theta)\partial\theta\,\partial\theta']$ uniformly in Θ_0 in probability. Similar to (A.16), using Lemma A.4(iii), we can show that $\sup_{\Theta_0}\|\partial^2 L_{sn}(\theta_n)/\partial\theta\,\partial\theta'-\partial^2\tilde{L}_{sn}(\theta_n)/\partial\theta\,\partial\theta'\|=o_p(1)$. Since $E[w_t\partial^2 l_t(\theta)\partial\theta\,\partial\theta']$ is continuous in terms of θ, for any sequence θ_n such that $\theta_n\to\theta_0$ in probability, we can show that

$$\frac{\partial^2\tilde{L}_{sn}(\theta_n)}{\partial\theta\partial\theta'}=-\frac{1}{2}\Sigma_0+o_p(1). \tag{A.19}$$

Fourth, for the previous neighborhood Θ_0, by (A.4), (A.13) and Lemmas A.1–A.3,

$$\sup_{\Theta_0}\left\|\frac{\partial l_t(\theta)}{\partial\theta}\right\|\leqslant O(1)\sup_{\Theta_0}\left\{|\varepsilon_t(\gamma)|\left\|\frac{\partial\varepsilon_t(\gamma)}{\partial\theta}\right\|+\left\|\frac{1}{h_t(\theta)}\frac{\partial h_t(\theta)}{\partial\theta}\right\|[\varepsilon_t^2(\gamma)+1]\right\}$$
$$\leqslant O(1)\Big\{(1+|\eta_t|)\xi_{\rho t-1}^2+(\xi_{\rho t-1}+\xi_{\rho t-1})[(1+|\eta_t|)^2\xi_{\rho t-1}^2+1]\Big\}$$
$$\leqslant O(1)\xi_{\rho t-1}^3(1+\eta_t^2),$$

where O(1) holds uniformly in t. Thus,

$$\Omega_0=4E\left[w_t^2\frac{\partial l_t(\theta_0)}{\partial\theta}\frac{\partial l_t(\theta_0)}{\partial\theta'}\right]<\infty.$$

Similar to the proof of Lemma 4.2 in Ling and McAleer (2003a), we can show that Σ_0 and Ω_0 are positive definite (see also Francq and Zakoïan, 2004). By the central limiting theorem, we have $\partial L_{sn}(\theta_0)/\partial\theta\to_{\mathscr{L}}N(0,\Omega_0/4)$. Similar to (A.16), using Lemma A.4(ii), we can show that $\sqrt{n}\|\partial L_{sn}(\theta_0)/\partial\theta-\partial\tilde{L}_{sn}(\theta_0)/\partial\theta\|=o_p(1)$ and hence $\partial\tilde{L}_{sn}(\theta_0)/\partial\theta\to_{\mathscr{L}}N(0,\Omega_0/4)$. Thus, we have established all the conditions in Theorem 4.1.3 in Amemiya (1985) and hence $\sqrt{n}(\hat{\theta}_{sn}-\theta_0)\to_{\mathscr{L}}N(0,\Sigma_0^{-1}\Omega_0\Sigma_0^{-1})$. This completes the proof. □

The following lemma is a key result for the ARMA-IGARCH model.

Lemma A.5. *Let* $\xi_{\rho t}$ *be defined as in* (3.4) *and* $\xi_{0\varrho t} = 1 + \sum_{i=0}^{\infty} \varrho^i |\varepsilon_{t-i}|$. *If Assumptions* 2.1–2.2 *hold, then for any* $\rho, \varrho \in (0,1)$, *there exist constants* $\varrho_1, \tilde{\imath} \in (0,1)$ *and* C *not depending on* t *such that*

(i) $\xi_{\rho t} \leqslant C\xi_{0\varrho_1 t}$ a.s.,

(ii) $\dfrac{\xi_{0\varrho t-1}}{\sqrt{h_t(\theta_0)}} \leqslant C\xi_{0\varrho_1 t-1}^{1-\tilde{\imath}}$ a.s..

Proof. By Assumption 2.1, y_t has the following expansion:

$$y_t = \mathrm{O}(1)\sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i},$$

where $\rho \in (0,1)$. By this, it is straightforward to show that (i) holds. Let G_0 be defined as G in (2.4) with $\beta_i = \beta_{0i}$. Since all $\beta_{0i} > 0$, it is not difficult to see that $u' G_0^j u \geqslant \beta_{01}^j$, where u is defined as in (A.5). Let $\tilde{\imath} \in (0,1)$ such that $\beta_{01}^{\tilde{\imath}} > \varrho$. Note that $\beta_{01}^{\iota_1 j/2} < 1$. Then, by (A.5), it follows that

$$\begin{aligned}
\frac{\xi_{0\varrho t-1}}{\sqrt{h_t(\theta_0)}} &\leqslant \mathrm{O}(1)\left[1 + \sum_{i=0}^{\infty}\left(\frac{\varrho}{\beta_{01}^{\tilde{\imath}}}\right)^i |\varepsilon_{t-i}|^{1-2\tilde{\imath}}\left(\frac{\beta_{01}^i \varepsilon_{t-i}^2}{C_{\theta_0} + \beta_{01}^i \varepsilon_{t-i}^2}\right)^{\tilde{\imath}}\right] \\
&\leqslant \mathrm{O}(1)\left[1 + \sum_{i=0}^{\infty}\left(\frac{\varrho}{\beta_{01}^{\tilde{\imath}}}\right)^i |\varepsilon_{t-i}|^{1-2\tilde{\imath}}(\beta_{01}^j \varepsilon_{t-i}^2)^{\tilde{\imath}/2}\right] \\
&\leqslant \mathrm{O}(1)\left[1 + \sum_{i=0}^{\infty}\left(\frac{\varrho}{\beta_{01}^{\tilde{\imath}}}\right)^i |\varepsilon_{t-i}|^{1-\tilde{\imath}}\right] \leqslant \xi_{0\varrho_1 t-1}^{1-\tilde{\imath}},
\end{aligned}$$

for some $\varrho_1 \in (0,1)$, where the last step holds using the same method as for (A.12). Thus, (ii) holds. This completes the proof. □

Lemma A.6. *If the assumptions of Theorem* 4.1 *hold and* $\sqrt{n}\|\theta_n - \theta_0\| \leqslant M$, *then it follows that*

(i) $\varepsilon_t(\gamma_n) = \varepsilon_t(\gamma_0) + \mathrm{o_p}(1)\sqrt{h_t(\theta_0)}$,

(ii) $h_t(\theta_n) = h_t(\theta_0) + \mathrm{o_p}(1)h_t(\theta_0)$,

where $\mathrm{o_p}(1)$ *holds uniformly in* $t = 1, \ldots, n$.

Proof. First, it is straightforward to show that

$$\varepsilon_t(\gamma_n) = \varepsilon_t(\gamma_0) + \mathrm{O}(n^{-1/2})\xi_{\rho t-1}.$$

Furthermore, by Lemma A.5(i), we have a constant $\varrho \in (0,1)$ such that

$$\varepsilon_t(\gamma_n) = \varepsilon_t(\gamma_0) + \mathrm{O}(n^{-1/2})\xi_{0\varrho t-1}. \tag{A.20}$$

By Theorem 2.1 (iii), for any $\tilde{\imath}, \varrho \in (0,1)$, we have $\mathrm{E}(\xi_{0\varrho t}^{1-\tilde{\imath}})^2 < \infty$. Thus,

$$\max_{1 \leqslant t \leqslant n} \xi_{0\varrho t-1}^{1-\tilde{\imath}}/\sqrt{n} = \mathrm{o_p}(1), \tag{A.21}$$

for any $\tilde{\imath}, \varrho \in (0,1)$. Thus, by (A.20) and Lemma A.5 (ii), we can see that (i) holds.

Define G_n and G_0 as G in (2.4) with $\theta=\theta_n$ and θ_0, respectively. By (A.5),

$$h_t(\theta_n)-h_t(\theta_0)=\left[C_{\theta_n}-C_{\theta_0}+\sum_{i=1}^{r}\sum_{j=1}^{\infty}(\alpha_{ni}u'G_n^ju-\alpha_{0i}u'G_0^ju)\varepsilon^2_{t-i-j}(\gamma_0)\right]$$
$$+\sum_{i=1}^{r}\sum_{j=1}^{\infty}\alpha_{ni}u'G_n^ju\left[\varepsilon^2_{t-i-j}(\gamma_n)-\varepsilon^2_{t-i-j}(\gamma_0)\right]$$
$$\equiv B_{1n}+B_{2n}, \tag{A.22}$$

where u is defined as in (A.5). There is a constant $c>0$ such that $G_0(1-c/\sqrt{n})\leqslant G_n\leqslant G_0(1+c/\sqrt{n})$, where "$B\leqslant C$" for matrices $B=(b_{ij})$ and $C=(c_{ij})$ means that $b_{ij}\leqslant c_{ij}$ for all i and j. Thus, $|u'G\ n^ju-u'G_0^ju)|\leqslant\max\{|(1-c/\sqrt{n})^j-1|,|(1+c/\sqrt{n})^j-1|\}u'G_0^ju\leqslant O(j/\sqrt{n})(1+c/\sqrt{n})^ju'G_0^ju$. Furthermore, since $C_{\theta_n}-C_{\theta_0}=O(n^{-1/2})$ and α_{0i} and α_{ni} are bounded, it follows that

$$B_{1n}\leqslant O\left(\frac{1}{\sqrt{n}}\right)+O\left(\frac{1}{\sqrt{n}}\right)\sum_{i=1}^{r}\sum_{j=1}^{\infty}u'G_n^ju\varepsilon^2_{t-i-j}(\gamma_0)$$
$$+\sum_{i=1}^{r}\sum_{j=1}^{\infty}\alpha_{0i}\left|u'G_n^ju-u'G_0^ju)\right|\varepsilon^2_{t-i-j}(\gamma_0)$$
$$\leqslant O\left(\frac{1}{\sqrt{n}}\right)+O\left(\frac{1}{\sqrt{n}}\right)\sum_{i=1}^{r}\sum_{j=1}^{\infty}\left(1+\frac{c}{\sqrt{n}}\right)^j ju'G_0^ju\varepsilon^2_{t-i-j}.$$

Using (A.5) and (A.7), as n is large enough, we have

$$j\left(1+\frac{c}{\sqrt{n}}\right)^j\frac{\alpha_{0i}u'G_0^ju\varepsilon^2_{t-i-j}}{h_t(\theta_0)}\leqslant j\left(1+\frac{c}{\sqrt{n}}\right)^j\frac{\alpha_{0i}u'G_0^ju\varepsilon^2_{t-i-j}}{C_{\theta_0}+\alpha_{0i}u'G_0^ju\varepsilon^2_{t-i-j}}$$
$$\leqslant j\left(1+\frac{c}{\sqrt{n}}\right)^j(\alpha_{0i}u'G_0^ju\varepsilon^2_{t-i-j})^{\iota_1}\leqslant O(\rho^j)|\varepsilon_{t-i-j}|^{2\iota_1},$$

for some $\rho\in(0,1)$ and any $\iota_1\in(0,1)$. Thus, as for (A.12), we can show that

$$B_{1n}\leqslant O\left(\frac{1}{\sqrt{n}}\right)+h_t(\theta_0)O\left(\frac{1}{\sqrt{n}}\right)\sum_{i=1}^{r}\sum_{j=1}^{\infty}\rho^j|\varepsilon_{t-i-j}|^{2\iota_1}$$
$$\leqslant O\left(\frac{1}{\sqrt{n}}\right)\left[1+h_t(\theta_0)\sum_{i=1}^{r}\xi^{2\iota_1}_{0\varrho t-i-j}\right]=h_t(\theta_0)o_p(1), \tag{A.23}$$

by (A.21), where $\iota_1\in(0,\frac{1}{2})$. By (A.5), $h_t(\theta_0)/\alpha_{0i}\geqslant u'G_0^ju\varepsilon^2_{t-i-j}$. By (A.20),

$$B_{2n}=O(1)\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{r}\sum_{j=1}^{\infty}u'G_n^ju|\varepsilon_{t-i-j}|\xi_{0\varrho t-i-j}+\frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{\infty}u'G_n^ju\xi^2_{0\varrho t-i-j}\right]$$
$$\leqslant O(1)\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{r}\sum_{j=1}^{\infty}\left(1+\frac{c}{\sqrt{n}}\right)^ju'G_0^ju|\varepsilon_{t-i-j}|\xi_{0\varrho t-i-j}+\frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{\infty}\rho^j\xi^2_{0\varrho t-i-j}\right]$$

$$\leqslant \mathrm{O}(1)\left[\frac{\sqrt{h_t(\theta_0)}}{\sqrt{n}}\sum_{i=1}^{r}\sum_{j=1}^{\infty}\left(1+\frac{c}{\sqrt{n}}\right)^j (u'G_0^j u)^{1/2}\xi_{0\varrho t-i-j}+\frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{\infty}\rho^j\xi^2_{0\varrho t-i-j}\right]$$

$$\leqslant \mathrm{O}\left(\frac{1}{\sqrt{n}}\right)\sqrt{h_t(\theta_0)}\sum_{i=1}^{r}\sum_{j=1}^{\infty}\rho^j\xi_{0\varrho t-i-j}+\mathrm{O}\left(\frac{1}{n}\right)\left(\sum_{i=1}^{r}\sum_{j=1}^{\infty}\rho^j\xi_{0\varrho t-i-j}\right)^2,$$

for some $\rho, \varrho \in (0,1)$. Reordering $\sum_{i=1}^{r}\sum_{j=1}^{\infty}\sum_{k=0}^{\infty}\rho^j\varrho^k|\varepsilon_{t-i-j-k}|$, we can show that there exists $\tilde{\varrho} \in (0,1)$ such that

$$B_{2n} = \mathrm{O}(1)\left[\frac{\sqrt{h_t(\theta_0)}}{\sqrt{n}}\xi_{0\tilde{\varrho}t-1}+\frac{1}{n}\xi^2_{0\tilde{\varrho}t-1}\right] = \mathrm{o_p}(1)h_t(\theta_0), \tag{A.24}$$

by (A.21) and Lemma A.5 (ii), where $\mathrm{o_p}(1)$ holds uniformly in $t = 1,\ldots,n$. By (A.22)–(A.24), (ii) holds. This completes the proof. □

Lemma A.7. *If the assumptions of Theorem 3 hold and $\sqrt{n}\|\theta_n - \theta_0\| \leqslant M$, then it follows that*

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta_n)}{\partial\theta\partial\theta'} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta_0)}{\partial\theta\,\partial\theta'} + \mathrm{o_p}(1) \quad \textit{for any fixed constant } M.$$

Proof. We first show that

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta_n)}{\partial\gamma\,\partial\gamma'} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta_0)}{\partial\gamma\,\partial\gamma'} + \mathrm{o_p}(1). \tag{A.25}$$

$\partial^2 l_t(\theta)/\partial\gamma\,\partial\gamma'$ in (A.3) includes five terms. We only provide the proof of the following equation, while other terms in (A.3) can be proved either easily or similarly.

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial h_t(\theta_n)}{\partial\gamma}\frac{\partial h_t(\theta_n)}{\partial\gamma'}\frac{\varepsilon_t^2(\gamma_n)}{h_t^3(\theta_n)} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial h_t(\theta_0)}{\partial\gamma}\frac{\partial h_t(\theta_0)}{\partial\gamma'}\frac{\varepsilon_t^2(\gamma_0)}{h_t^3(\theta_0)} + \mathrm{o_p}(1). \tag{A.26}$$

By Lemma A.6(ii), we have

$$\left|\frac{1}{\sqrt{h_t(\theta_n)}} - \frac{1}{\sqrt{h_t(\theta_0)}}\right| = \left|\frac{h_t(\theta_n) - h_t(\theta_0)}{\sqrt{h_t(\theta_n)h_t(\theta_0)}(\sqrt{h_t(\theta_n)} + \sqrt{h_t(\theta_0)})}\right|$$
$$= \frac{\mathrm{o_p}(1)}{\sqrt{h_t(\theta_0)[1+\mathrm{o_p}(1)]}[\sqrt{1+\mathrm{o_p}(1)}+1]} = \frac{\mathrm{o_p}(1)}{\sqrt{h_t(\theta_0)}},$$

where $\mathrm{o_p}(1)$ holds uniformly in $t = 1,\ldots,n$. By Lemma A.6,

$$\frac{|\varepsilon_t(\gamma_n) - \varepsilon_t(\gamma_0)|}{\sqrt{h_t(\theta_n)}} = \frac{\mathrm{o_p}(1)}{\sqrt{1+\mathrm{o_p}(1)}} = \mathrm{o_p}(1),$$

where $\mathrm{o_p}(1)$ holds uniformly in $t = 1,\ldots,n$. By the preceding two inequalities,

$$\left|\frac{\varepsilon_t(\gamma_n)}{\sqrt{h_t(\theta_n)}} - \frac{\varepsilon_t(\gamma_0)}{\sqrt{h_t(\theta_0)}}\right| \leqslant \frac{|\varepsilon_t(\gamma_n) - \varepsilon_t(\gamma_0)|}{\sqrt{h_t(\theta_n)}} + |\varepsilon_t(\gamma_0)|\left|\frac{1}{\sqrt{h_t(\theta_n)}} - \frac{1}{\sqrt{h_t(\theta_0)}}\right|$$
$$\leqslant \mathrm{o_p}(1) + \frac{|\varepsilon_t(\gamma_0)|\mathrm{o_p}(1)}{\sqrt{h_t(\theta_0)}} = \mathrm{o_p}(1) + |\eta_t|\mathrm{o_p}(1),$$

where $o_p(1)$ holds uniformly in $t = 1, \ldots, n$. Thus,

$$\left|\frac{\varepsilon_t^2(\gamma_n)}{h_t(\theta_n)} - \frac{\varepsilon_t^2(\gamma_0)}{h_t(\theta_0)}\right| \leqslant 2\left|\frac{\varepsilon_t(\gamma_0)}{\sqrt{h_t(\theta_0)}}\right|\left|\frac{\varepsilon_t(\gamma_n)}{\sqrt{h_t(\theta_n)}} - \frac{\varepsilon_t(\gamma_0)}{\sqrt{h_t(\theta_0)}}\right| + \left|\frac{\varepsilon_t(\gamma_n)}{\sqrt{h_t(\theta_n)}} - \frac{\varepsilon_t(\gamma_0)}{\sqrt{h_t(\theta_0)}}\right|^2$$
$$= o_p(1) + \eta_t^2 o_p(1). \tag{A.27}$$

By Lemmas A.3, A.5(ii) and A.6(ii), there exists a neighborhood Θ_0 of θ_0 such that

$$\sup_{\Theta_0}\left\|\frac{\partial h_t(\theta)}{\partial \gamma}\frac{\partial h_t(\theta)}{\partial \gamma'}\frac{1}{h_t(\theta)}\right\|\frac{1}{h_t(\theta_n)} \leqslant \frac{O(1)\xi_{\rho,t-1}^2}{h_t(\theta_0)[1+o_p(1)]} = O_p(1)\xi_{0\varrho,t-1}^{2-\tilde{\imath}}, \tag{A.28}$$

where $O_p(1)$ holds uniformly in $t = 1, \ldots, n$. By (A.27)–(A.28), it follows that

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial h_t(\theta_n)}{\partial \gamma}\frac{\partial h_t(\theta_n)}{\partial \gamma'}\frac{1}{h_t^2(\theta_n)}\left|\frac{\varepsilon_t^2(\gamma_n)}{h_t(\theta_n)} - \frac{\varepsilon_t^2(\gamma_0)}{h_t(\theta_0)}\right|$$
$$\leqslant \frac{o_p(1)}{n}\sum_{t=1}^{n}\left(\xi_{0\varrho,t-1}^{2-\tilde{\imath}} + \xi_{0\varrho,t-1}^{2-\tilde{\imath}}\eta_t^2\right) = o_p(1),$$

since $E\xi_{0\varrho,t-1}^{2-\tilde{\imath}} < \infty$. Thus, by Lemma A.6(ii) and (A.28), it is not hard to see that

$$\frac{1}{n}\sum_{t=1}^{n}\left[\frac{\partial h_t(\theta_n)}{\partial \gamma}\frac{\partial h_t(\theta_n)}{\partial \gamma'}\frac{\eta_t^2}{h_t^2(\theta_n)} - \frac{\partial h_t(\theta_n)}{\partial \gamma}\frac{\partial h_t(\theta_n)}{\partial \gamma'}\frac{\eta_t^2}{h_t(\theta_n)h_t(\theta_0)}\right]$$
$$= \frac{1}{n}\sum_{t=1}^{n}\left\{\frac{\partial h_t(\theta_n)}{\partial \gamma}\frac{\partial h_t(\theta_n)}{\partial \gamma'}\frac{\eta_t^2}{h_t(\theta_n)}\left[\frac{1}{h_t(\theta_n)} - \frac{1}{h_t(\theta_0)}\right]\right\}$$
$$= \frac{o_p(1)}{n}\sum_{t=1}^{n}\left[\frac{\partial h_t(\theta_n)}{\partial \gamma}\frac{\partial h_t(\theta_n)}{\partial \gamma'}\frac{\eta_t^2}{h_t^2(\theta_n)}\right] = o_p(1)\frac{1}{n}\sum_{t=1}^{n}\xi_{0\varrho,t-1}^{2-\tilde{\imath}}\eta_t^2 = o_p(1),$$

since $\sum_{t=1}^{n}\xi_{0\varrho,t-1}^{2-\tilde{\imath}}\eta_t^2/n = O_p(1)$, where $o_p(1)$ holds uniformly in $t = 1, \ldots, n$. By the two preceding inequalities,

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial h_t(\theta_n)}{\partial \gamma}\frac{\partial h_t(\theta_n)}{\partial \gamma'}\frac{\varepsilon_t^2(\gamma_n)}{h_t^3(\theta_n)} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial h_t(\theta_n)}{\partial \gamma}\frac{\partial h_t(\theta_n)}{\partial \gamma'}\frac{\eta_t^2}{h_t^2(\theta_n)} + o_p(1)$$
$$= \frac{1}{n}\sum_{t=1}^{n}\frac{\partial h_t(\theta_n)}{\partial \gamma}\frac{\partial h_t(\theta_n)}{\partial \gamma'}\frac{\eta_t^2}{h_t(\theta_n)h_t(\theta_0)} + o_p(1).$$

By Lemmas A.3 and A.5, $E\sup_{\Theta_0}\{\|h_t^{-1/2}(\theta)\partial h_t(\theta)/\partial\gamma\|^2\eta_t^2/h_t(\theta_0)\} \leqslant O(1)E\xi_{0\varrho,t-1}^{2-\tilde{\imath}} < \infty$. Furthermore, by the preceding equation, the dominated convergence theorem and Markov's theorem, we can show that (A.26) holds.

We next show that

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta_n)}{\partial \delta\,\partial \delta'} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta_0)}{\partial \delta\,\partial \delta'} + o_p(1). \tag{A.29}$$

There are two terms in $\partial^2 l_t(\theta)/\partial\delta\,\partial\delta'$, see (A.4). We only provide the proof of the following equation, while another term can be similarly proved.

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial h_t(\theta_n)}{\partial\delta}\frac{\partial h_t(\theta_n)}{\partial\delta'}\frac{\varepsilon_t^2(\gamma_n)}{h_t^3(\theta_n)}=\frac{1}{n}\sum_{t=1}^{n}\frac{\partial h_t(\theta_0)}{\partial\delta}\frac{\partial h_t(\theta_0)}{\partial\delta'}\frac{\varepsilon_t^2(\gamma_0)}{h_t^3(\theta_0)}+\mathrm{o_p}(1). \tag{A.30}$$

By Lemma A.2, it follows that

$$\left\|\frac{\partial h_t(\theta_n)}{\partial\delta}\frac{\partial h_t(\theta_n)}{\partial\delta'}\frac{1}{h_t^2(\theta_n)}\right\|\leqslant \mathrm{O}(1)\xi_{\rho t-1}^{\iota_1}. \tag{A.31}$$

By (A.27) and (A.31), we have

$$\frac{\partial h_t(\theta_n)}{\partial\delta}\frac{\partial h_t(\theta_n)}{\partial\delta'}\frac{\varepsilon_t^2(\gamma_n)}{h_t^3(\theta_n)}=\frac{\partial h_t(\theta_n)}{\partial\delta}\frac{\partial h_t(\theta_n)}{\partial\delta'}\frac{\varepsilon_t^2(\gamma_0)}{h_t^2(\theta_n)h_t(\theta_0)}+\mathrm{o_p}(1)\xi_{\rho t-1}^{\iota_1}(1+\eta_t^2).$$

By Lemma A.2, $\mathrm{E}\,(\sup_{\Theta_0}\|h_t^{-1}(\theta)\partial h_t(\theta)/\partial\delta\|^2\eta_t^2)<\infty$. Furthermore, by the preceding two equations, the dominated convergence theorem and Markov's theorem, we can show that (A.30) holds. Similarly, we can show that $n^{-1}\sum_{t=1}^{n}\partial^2 l_t(\theta_n)/\partial\delta\,\partial\gamma' = n^{-1}\sum_{t=1}^{n}\partial^2 l_t(\theta_0)/\partial\delta\,\partial\gamma'$ $+\mathrm{o_p}(1)$. This completes the proof. □

Proof of Theorem 4.1. We first show that

$$\mathrm{E}\left\|\frac{\partial^2 l_t(\theta_0)}{\partial\theta\,\partial\theta'}\right\|<\infty \quad\text{and}\quad \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_t(\theta_0)}{\partial\theta\,\partial\theta'}=-\frac{1}{2}\Sigma+\mathrm{o_p}(1). \tag{A.32}$$

By Lemma A.5(ii) and Lemma A.3, we have

$$\mathrm{E}\left\|\frac{\partial h_t(\theta_0)}{\partial\gamma}\frac{\partial h_t(\theta_0)}{\partial\gamma'}\frac{\varepsilon_t^2(\gamma_0)}{h_t^3(\theta_0)}\right\|\leqslant \mathrm{E}\left\|\frac{\partial h_t(\theta_0)}{\partial\gamma}\frac{1}{h_t(\theta_0)}\right\|^2<\infty.$$

Similarly, we can show that the other terms in $\mathrm{E}[\partial^2 l_t(\theta_0)/\partial\theta\,\partial\theta']$ are finite. Thus, the first part of (A.32) holds. The second part of (A.32) holds by the ergodic theorem.

By Taylor's expansion with each component, we have

$$\sum_{t=1}^{n}\frac{\partial\tilde{l}_t(\hat{\theta}_{sn})}{\partial\theta}=\sum_{t=1}^{n}\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta}+\sum_{t=1}^{n}\frac{\partial^2\tilde{l}_t(\theta_n^*)}{\partial\theta\,\partial\theta'}(\hat{\theta}_{sn}-\theta_0), \tag{A.33}$$

where θ_n^* lies between θ_0 and $\tilde{\theta}_{sn}$. By Lemma A.4(iii), Lemma A.7, and (A.32), we can show that

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\tilde{l}_t(\hat{\theta}_{sn})}{\partial\theta\,\partial\theta'}=-\frac{1}{2}\Sigma+\mathrm{o_p}(1)\quad\text{and}\quad \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\tilde{l}_t(\theta_n^*)}{\partial\theta\,\partial\theta'}=-\frac{1}{2}\Sigma+\mathrm{o_p}(1).$$

By Lemma A.4(ii), we can show that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial\tilde{l}_t(\hat{\theta}_0)}{\partial\theta}=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t(\theta_0)}{\partial\theta}+\mathrm{o_p}(1).$$

As in Ling and McAleer (2003a) and Francq and Zakoïan (2004), we can show that $\Sigma > 0$ and $\Omega > 0$. Furthermore, since $\sqrt{n}(\hat{\theta}_{sn} - \theta_0) = O_p(1)$, by (A.33), we have

$$\begin{aligned}\hat{\theta}_n &= \hat{\theta}_{sn} - \left[-\frac{1}{2}\Sigma + o_p(1)\right]^{-1}\left\{\frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t(\theta_0)}{\partial\theta} + \left[-\frac{1}{2}\Sigma + o_p(1)\right](\hat{\theta}_{sn} - \theta_0)\right\} \\ &\qquad + o_p\left(\frac{1}{\sqrt{n}}\right) \\ &= \theta_0 + \frac{2\Sigma^{-1}}{n}\sum_{t=1}^{n}\frac{\partial l_t(\theta_0)}{\partial\theta} + o_p\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

Finally, by the central limiting theorem, we can show that the conclusion holds. This completes the proof. □

References

Amemiya, T., 1985. Advanced Econometrics. Cambridge, Harvard University Press, Cambridge, MA.

Basrak, B., Davis, R.A., Mikosh, T., 2002. Regular variation of GARCH processes. Stochastic Processes and their Application 99, 95–115.

Berkes, I., Horvath, L., 2004. The efficiency of the estimates of the parameters in GARCH processes. Annals of Statistics 32, 633–655.

Berkes, I., Horvath, L., Kokoszka, P., 2003. GARCH processes: structure and estimation. Bernoulli 9, 201–207.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Bougerol, P., Picard, N.M., 1992. Stationarity of GARCH processes and of some nonnegative time series. Journal of Econometrics 52, 115–127.

Chan, N.H., Peng, L., 2005. Weighted least absolute deviations estimation for an AR(1) process with ARCH(1) errors. Biometrika 92, 477–484.

Davis, R.A., Knight, K., Liu, J., 1992. *M*-estimation for autoregressions with infinite variance. Stochastic Processes and their Application 40, 145–180.

Drost, F.C., Klaassen, C.A.J., Werker, B.J.M., 1997. Adaptive estimation in time series models. Annals of Statistics 25, 786–817.

Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of variance of U.K. inflation. Econometrica 50, 987–1008.

Francq, C., Zakoïan, J.M., 2004. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. Bernoulli 10, 605–637.

Hall, P., Yao, Q.W., 2003. Inference in ARCH and GARCH models. Eonometrica 71, 285–317.

Horvath, L., Liese, F., 2004. *Lp*-estimators in ARCH models. Journal of Statistics Plann. Inference 119, 277–309.

Jeantheau, T., 1998. Strong consistency of estimators for multivariate ARCH models. Econometric Theory 14, 70–86.

Koul, H.L. and Ling, S., 2006. Fitting an error distribution in some heteroscedastic time series models. Annals of Statistics 34, 994–1012.

Lee, S.-W., Hansen, R.E., 1994. Asymptotic theory for GARCH(1,1) quasi-maximum likelihood estimator. Econometric Theory 10, 29–52.

Li, W.K., Mak, T.K., 1994. On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. Journal of Time Series Analysis 15, 627–636.

Li, W.K., Ling, S., McAleer, M., 2002. A survey of recent theoretical results for time series models with GARCH errors. Journal of Economic Survey 16, 245–269.

Ling, S., 1999. On probability properties of a double threshold ARMA conditional heteroskedasticity model. Journal of Applied Probability 36, 688–705.

Ling, S., 2003. Self-weighted LSE and MLE for ARMA-GARCH Models. Unpublished working paper. HKUST.

Ling, S., 2004. Estimation and testing of stationarity for double autoregressive models. Journal of the Royal Statistical Society: Series B 66, 63–78.

Ling, S., 2005. Self-weighted LAD estimation for infinite variance autoregressive models. Journal of the Royal Statistical Society B 67, 381–393.

Ling, S., Li, W.K., 1997. Fractional autoregressive integrated moving-average time series conditional heteroskedasticity. Journal of American Statistical Association 92, 1184–1194.

Ling, S., McAleer, M., 2003a. Asymptotic theory for a new vector ARMA-GARCH model. Econometric Theory 19, 280–310.

Ling, S., McAleer, M., 2003b. On adaptive estimation in nonstationary ARMA models with GARCH errors. Annals of Statist. 31, 642–674.

Lumsdaine, R.L., 1996. Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. Econometrica 64, 575–596.

Mikosch, T., Gadrich, T., Klppelberg, C., Adler, R.J., 1995. Parameter estimation for ARMA models with infinite variance innovations. Annals of Statist. 23, 305–326.

Peng, L. and Yao, Q.W., 2003. Least absolute deviations estimation for ARCH and GARCH models. Biometrika 967–975.

Tsay, R.S., 1987. Conditional heteroscedastic time series models. Journal of American Statistical Association 7, 590–604.

Weiss, A.A., 1986. Asymptotic theory for ARCH models: estimation and testing. Econometrics Theory 2, 107–131.