DE GRUYTER OPEN Open Math. 2017; 15: 1539–1548

Open Mathematics

Research Article

Haiyan Xuan*, Lixin Song, Muhammad Amin, and Yongxia Shi

Quasi-maximum likelihood estimator of Laplace (1, 1) for GARCH models

https://doi.org/10.1515/math-2017-0131
Received April 20, 2017; accepted November 9, 2017.

Abstract: This paper studies the quasi-maximum likelihood estimator (QMLE) for the generalized autoregressive conditional heteroscedastic (GARCH) model based on the Laplace (1,1) residuals. The QMLE is proposed to the parameter vector of the GARCH model with the Laplace (1,1) firstly. Under some certain conditions, the strong consistency and asymptotic normality of QMLE are then established. In what follows, a real example with Laplace and normal distribution is analyzed to evaluate the performance of the QMLE and some comparison results on the performance are given. In the end the proofs of some theorem are presented.

Keywords: Asymptotic normality, GARCH model, Laplace (1,1), Quasi-maximum likelihood estimator, Strong consistency

MSC: 62M10, 91G70

1 Introduction

The ARCH model has been widely used ever since it was first proposed by Engle (1982)[1] because this model was able to address the volatility in the forecasting of Britain's inflation rate. In many statistical applications, particularly finance, the ARCH model is the leading way to explain changes in the conditional variance of the error term over time. In recent years, the ARCH model has been extended to the generalized-ARCH (GARCH) model (see Bollerslev (1986)[2]). Since the GARCH model can explain the phenomena of volatility convergence and the thick tail of the rate of return (see David (2014)[3] and Yang (2008)[4]) it has drawn widespread concern from many scholars and has many applications.

Recently, some advances have been made for the structure and parameter estimation of the GARCH model. Weiss (1986)[5] established some results on the asymptotic properties of the QMLE depending on assumptions of moment conditions. Lee (1994)[6] and Lumsdaine (1996)[7] studied the asymptotic properties of the QMLE for the GARCH model. Berkes et al. (2003)[8] studied the structure and estimator of GARCH. Berkes and Horvath (2003, 2004)[9,10] provided consistency convergence rate that is QMLE and validity of parameter estimation for general GARCH(r,s). Francq and Zakoian (2004)[11] studied the QMLE for GARCH(r, s). Straumann (2006)[12] presented the QMLE by a stochastic recurrence method, which includes GARCH(r, s). Ling (2007)[13] proposed a self-weighted QMLE, and Zhu (2011)[14] investigated the local QMLE for IGARCH models under a fractional moment condition only. The theoretical properties of the QMLE in the GARCH model need to be developed further, especially in statistical applications, to include situations where these sorts of moment conditions are not satisfied. Han and Kristensen (2014)[15] applied the asymptotic properties of

***Corresponding Author: Haiyan Xuan:** School of Mathematical Sciences, Dalian University of Technology, Dalian, China and School of Economics and Management, Lanzhou University of Technology, Lanzhou, China, E-mail: haiyanxuan@gmail.com
Lixin Song: School of Mathematical Sciences, Dalian University of Technology, Dalian, China, E-mail: lxsong@dlut.edu.cn
Muhammad Amin: Nuclear Institute for Food and Agriculture (NIFA) Peshawar, Pakistan, E-mail: aminkanju@gmail.com
Yongxia Shi: School of Sciences, Lanzhou University of Technology, Lanzhou, China, E-mail: yongxiashi@126.com

Gaussian QMLE to the GARCH model with an additional explanatory variable, and showed that the QMLE of the parameters for the volatility equation is consistent and mixed-normally distributed in large samples.

Although the literature on classical GARCH models is quite rich, most of it is based on residuals of GARCH model, which follow a normal distribution, as noted by Francq and Zakoian (2004)[11]. Nelson (1991)[16] used other distributions to investigate the GARCH model. In this paper, we consider the Laplace distribution since this distribution is worthy of being studied because it describes the fat-tail feature of financial market data. This paper mainly investigates the QMLE for the GARCH model based on Laplace distribution. The theoretical results on strong consistency and asymptotic normality of the QMLE are established. A performance comparison between Laplace distribution and normal distribution is made to show that the former is superior to the later.

The article is organized as follows. The main results for the QMLE of GARCH(r, s) are given based on Laplace distribution in the second section. In the third section a practical instance is described. The proofs of two theorems are in the end.

2 Main results

In this section we investigate the quasi-maximum likelihood estimator (QMLE) for the generalized autoregressive conditional heteroscedastic (GARCH) model based on the Laplace (1,1) residuals to propose some theoretical results.

The GARCH (r,s) model has the following form:

$$\varepsilon_t = \eta_t \sigma_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2, \tag{1}$$

Here, η_t is a sequence of independent and identical distributed (i.i.d.) random variables, $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, 2, \cdots, r$; $\beta_j \geq 0$, $j = 1, 2, \cdots, s$. According to Bougerol and Picard(1992)[17], the sufficient and necessary condition for the strictly stationary solution of GARCH (r, s) model is

$$\sum_{i=1}^{r} \alpha_i + \sum_{j=1}^{s} \beta_j < 1. \tag{2}$$

Assume that $\theta = (\alpha_0, \alpha_1, \cdots, \alpha_r, \beta_1, \beta_2, \cdots, \beta_s)'$ is the parameter vector of formula (1) and its true parameter vector is θ_0. Let $l = r + s + 1$, then θ is l dimension vector. The parameter vector space is Θ, $\Theta \subset R_0^{r+s+1}$, $R_0 = [0, \infty)$. By assumption that a Laplace distribution has the density $f(x) = 0.5e^{-|x-1|}$ for η_t and conditionally on initial values $\varepsilon_0, \cdots, \varepsilon_{1-r}, \sigma_0^2, \cdots, \sigma_{1-s}^2$, then the Laplace quasi-likelihood is

$$L_n(\theta) = L_n(\theta; \varepsilon_0, \cdots, \varepsilon_n) = \sum_{t=1}^{n} \frac{1}{2\sqrt{\sigma_t^2}} exp(-\frac{|\varepsilon_t - 1|}{\sqrt{\sigma_t^2}})$$

with regard to $t \geq 1$, where

$$\sigma_t^2 = \sigma_t^2(\theta) = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2.$$

We select the initial values

$$\varepsilon_0 = \cdots = \varepsilon_{1-r} = \sqrt{\omega}, \ \sigma_0^2 = \cdots = \sigma_{1-s}^2 = \omega \tag{3}$$

or

$$\varepsilon_0 = \cdots = \varepsilon_{1-r} = \varepsilon_1, \ \sigma_0^2 = \cdots = \sigma_{1-q}^2 = \varepsilon_1^2. \tag{4}$$

As a result, $\hat{\theta}_n$ is named the QMLE for θ and has the following form

$$\hat{\theta}_n = \arg\max_{\theta \in \Theta} L_n(\theta) = \arg\min_{\theta \in \Theta} I_n(\theta), \tag{5}$$

where

$$I_n(\theta) = n^{-1}\sum_{t=1}^{n} l_t, \quad l_t = l_t(\theta) = \log\sigma_t(\theta) + \frac{|\varepsilon_t - 1|}{\sigma_t(\theta)}. \tag{6}$$

Denote $\alpha(z) = \sum_{i=1}^{r}\alpha_i z^i$, $\beta(z) = 1 - \sum_{j=1}^{s}\beta_j z^j$. If $r = 0, \alpha(z) = 0$; if $s = 0, \beta(z) = 1$. Before providing main results, we introduce firstly the following assumptions.

Assumption 1. $\theta_0 \in \Theta$, Θ is compact and θ_0 is an inner dot.

Assumption 2. $\sum_{j=1}^{r}\beta_j < 1$.

Assumption 3. $E[\frac{|\varepsilon_t-1|}{\sigma_t}] = 1$.

Assumption 4. If $s > 0$,there are no common roots for $\alpha(z)$ and $\beta(z)$, $\alpha(1) \neq 0$, $\alpha_r + \beta_s \neq 0$.

Assumption 5. $\tau = E[\frac{|\varepsilon_t-1|}{\sigma_t}]^2 < \infty$.

Assumption 1 ensures the parameter vector space is compact and is required to prove asymptotic normality. Assumption 2 and 4 are the identifiability conditions for model (1). Assumption 3 is a necessary condition to prove the strong consistency and Assumption 5 is asymptotic normality.

Actually, the initial values of ε_t and σ_t^2 are unknown when $t \leq 0$. Let $\tilde{\varepsilon}_t(\theta)$ and $\tilde{\sigma}_t^2(\theta)$ be $\varepsilon_t(\theta)$ and $\sigma_t^2(\theta)$, respectively, when ε_t and $\sigma_t^2(\theta)$ are constants when $t \leq 0$. The formula (6) can be modified as

$$\tilde{I}_n(\theta) = \tilde{I}_n(\theta; \varepsilon_n, \varepsilon_{n-1}, \cdots) = n^{-1}\sum_{t=1}^{n}\tilde{l}_t, \tag{7}$$

$$\tilde{l}_t = \tilde{l}_t(\theta) = \log\sqrt{\tilde{\sigma}_t^2(\theta)} + \frac{|\varepsilon_t - 1|}{\sqrt{\tilde{\sigma}_t^2(\theta)}}. \tag{8}$$

In what follows we establish the main results of this paper.

Theorem 2.1. *Under the initial values (3) or (4), if the Assumptions 1-5 hold, then there exists a sequence of minimizers $\hat{\theta}_n$ of $I_n(\theta)$ such that*

$$\hat{\theta}_n \to \theta_0 \quad a.s., \quad as\ n \to \infty.$$

Theorem 2.2. *If the Assumptions 1-5 hold, then*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to N(0, MJ^{-1}) \quad as\ n \to \infty,$$

where

$$M = \frac{\tau - 1}{4}, \quad J = E_{\theta_0}\left(\frac{\partial^2 l_t(\theta_0)}{\partial\theta\partial\theta^T}\right) = E_{\theta_0}\left(\frac{1}{\sigma_t^4(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta^T}\right). \tag{9}$$

3 Applications and comparison

In this section, the China Securities Index 800 (CSI 800) from January 12, 2007 to December 31, 2008 is studied. There are 482 data points. The descriptive statistics of data subjected to differential, denoted by $\{y_t\}_{t=1}^{481}$ are shown in Figure 1.

As shown in Figure 1, the mean was near 0. At the 0.05 significance level, the value of J-Bera statistic is greater than the critical value. This indicates that the regression may not follow the normal distribution. It can be initially determined that the distribution of the return presents "fat tail" feature.

When the significance level is 1%, 5% and 10%, the value of the test statistic t is smaller than the critical value in Table 1. In this way, the sequence rejects the null hypothesis, which the unit root exists. It is also

Fig. 1. Log-return descriptive statistics of CSI 800

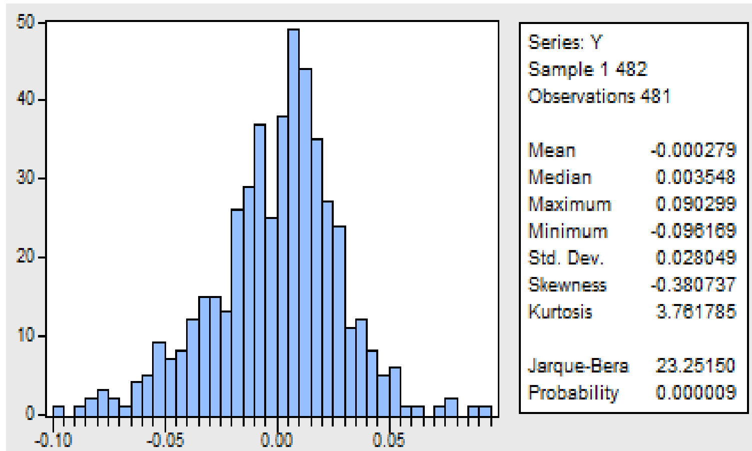

Table 1. ADF unit root test

		t-Statistic	Prob.*
Augmented Dickey-Fuller	test statistic	-21.50265	0.0000
Test critical values:	1% level	-3.443748	
	5% level	-2.867342	
	10% level	-2.569922	

Table 2. ARCH test

F-statistic	3.282313	Prob. F(1,478)	0.0707
Obs*R-squared	3.273568	Prob. Chi-square(1)	0.0704

a stationary series. As shown in Table 2, $n * R^2 = 3.273568 > \chi^2_{0.1}(1)$, which indicates the sequence has heteroskedasticity.

In this way, the GARCH(1,1) model was established according to $\{y_t\}$. When η_t obeyed the Laplace (1, 1), the estimation of the parameters vector was performed by using the QMLE at MATLAB.

The results were $\alpha_0 = 0.0701, \alpha_1 = 0.1381, \beta_1 = 0.4605$. As $\alpha_1 + \beta_1 = 0.5986 < 1$, which satisfies the strictly stationary condition. Next, the sample biases, the sample standard deviations (SD) and the asymptotic standard deviations (AD) of the estimation on the parameter vector were given when η_t obeyed the N(0, 1) and η_t obeyed the Laplace(1, 1).

The bias is a technical index that reflects the degree of violation of the stock price and the moving average in the process of fluctuation. The computational formula is:

$$bias = \frac{Closingprice - MovingaveragepriceinNdays}{MovingaveragepriceinNdays} \times 100\%$$

SD is the square root of the arithmetic mean of deviation from the mean square. AD is the standard deviation of the asymptotic distribution of deviation. They all reflect the degree of dispersion among individuals in the group. Therefore, these indexes can be used to analyze the accuracy of the estimation on the parameter vector. In general, the estimation is much more accurate only if the values of bias, AD and SD are much smaller.

As shown in Table 3, when $\eta_t \sim$ Laplace(1, 1), the values of the bias, the SD, and the AD were all smaller than the ones when $\eta_t \sim$ N(0,1). This indicated that the fitting effect of Laplace distribution is better than that of normal distribution. It is hereby suggested that instead of the Normal distribution the Laplace distribution is much more effective for the data from financial markets.

Table 3. Estimators for GARCH(1, 1)

	QMLE-N			QMLE-L		
	$\bar{\alpha}_0$	$\bar{\alpha}_1$	$\bar{\beta}_1$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$
Bias	**-0.0005**	**1.4103**	**-0.6701**	**0.0001**	**-0.3619**	**-0.0698**
SD	**0.0035**	**0.1912**	**0.0256**	**0.0025**	**0.0042**	**0.0175**
AD	**0.0000**	**0.0695**	**0.0451**	**0.0000**	**0.0491**	**0.0159**

4 Proofs

In this section we will present the proof of Theorem 2.1 and Theorem 2.2.

Proof of Theorem 2.1. The formula $\sigma_t^2 = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2$ in model (1) can be rewritten in vector form as

$$\underline{\tilde{\sigma}}_t^2 = \underline{\tilde{c}}_t + B\underline{\tilde{\sigma}}_{t-1}^2, \tag{10}$$

where

$$\underline{\tilde{\sigma}}_t^2 = \begin{pmatrix} \sigma_t^2 \\ \sigma_{t-1}^2 \\ \vdots \\ \sigma_{t-s+1}^2 \end{pmatrix}, \underline{\tilde{c}}_t = \begin{pmatrix} \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, B = \begin{pmatrix} \beta_1 & \beta_2 & \cdots & \beta_r \\ 1 & 0 & \cdots & 0 \\ \vdots & & & \\ 0 & \cdots & 1 & 0 \end{pmatrix}. \tag{11}$$

Then following intermediate results can be used to prove Theorem 2.1.

(i) $\lim_{n\to\infty} \sup_{\theta\in\Theta} |I_n(\theta) - \tilde{I}_n(\theta)| = 0 \;\; a.s.$
(ii) $\exists t \in Z$ makes $\sigma_t^2(\theta) = \sigma_t^2(\theta_0) \;\; P_{\theta_0} \;\; a.s. \Rightarrow \theta = \theta_0$.
(iii) $E_{\theta_0}|l_t(\theta_0)| < \infty$, and if $\theta \neq \theta_0$, $E_{\theta_0} l_t(\theta) > E_{\theta_0} l_t(\theta_0)$.
(iv) For $\theta \neq \theta_0$, there is a neighbourhood $V(\theta)$ making $\liminf_{n\to\infty} \inf_{\theta^*\in V(\theta)} \tilde{I}_n(\theta^*) > E_{\theta_0} l_1(\theta_0) \;\; a.s.$

(i) As Assumption 1,

$$\sup_{\theta\in\Theta} \rho(B) < 1. \tag{12}$$

Iterating (10) produces the following:

$$\underline{\sigma}_t^2 = \underline{c}_t + B\underline{c}_{t-1} + B^2\underline{c}_{t-1} + \cdots + B^{t-1}\underline{c}_1 + B^t\underline{\sigma}_0^2 = \sum_{k=0}^{\infty} B^k \underline{c}_{t-k}. \tag{13}$$

It is supposed here that $\underline{\tilde{\sigma}}_t^2$ may be the vector obtained by replacing σ_{t-i}^2 by $\tilde{\sigma}_{t-i}^2$. Let $\underline{\tilde{c}}$ be the vector obtained by replacing $\varepsilon_0^2, \cdots, \varepsilon_{1-p}^2$ with the initial values (3) or (4). We have

$$\underline{\tilde{\sigma}}_t^2 = \underline{c}_t + B\underline{\tilde{c}}_{t-1} + \cdots + B^{t-r-1}\underline{\tilde{c}}_r + 1 + B^{t-r}\underline{\tilde{c}}_r + \cdots + B^{t-1}\underline{\tilde{c}}_1 + B^t\underline{\tilde{\sigma}}_0^2. \tag{14}$$

Through (12)-(14), it is almost certain the following is true:

$$\begin{aligned} \sup_{\theta\in\Theta} \|\underline{\sigma}_t^2 - \underline{\tilde{\sigma}}_t^2\| &= \sup_{\theta\in\Theta} \left\| \left\{ \sum_{k=1}^{r} B^{t-k}(\underline{c}_k - \underline{\tilde{c}}_k) + B^t(\underline{\sigma}_0^2 - \underline{\tilde{\sigma}}_0^2) \right\} \right\| \\ &\leq K\rho^t, \quad \forall t. \end{aligned} \tag{15}$$

Hence,

$$\begin{aligned} \sup_{\theta\in\Theta} |I_n(\theta) - \tilde{I}_n(\theta)| \leq\; & n^{-1} \sum_{t=1}^{n} \sup_{\theta\in\Theta} \left\{ \left| \frac{\tilde{\sigma}_t - \sigma_t}{\sigma_t \tilde{\sigma}_t} (\varepsilon_t - 1) \right| + \frac{1}{2} \left| \log\left(1 + \frac{\sigma_t^2 - \tilde{\sigma}_t^2}{\tilde{\sigma}_t^2}\right) \right| \right\} \\ \leq\; & n^{-1} \sum_{t=1}^{n} \sup_{\theta\in\Theta} \left\{ \sqrt{\left| \frac{\tilde{\sigma}_t^2 - \sigma_t^2}{\tilde{\sigma}_t^2 \sigma_t^2} (\varepsilon_t - 1)^2 \right|} + \frac{1}{2} \left| \log\left(1 + \frac{\sigma_t^2 - \tilde{\sigma}_t^2}{\tilde{\sigma}_t^2}\right) \right| \right\} \end{aligned}$$

$$\leq \frac{\sqrt{K}}{n}\left\{\sup_{\theta\in\Theta}\frac{1}{\alpha_0}\right\}\sum_{t=1}^{n}\sqrt{\rho^t(\varepsilon_t-1)^2}+\frac{K}{2n}\left\{\sup_{\theta\in\Theta}\frac{1}{\alpha_0}\right\}\sum_{t=1}^{n}\rho^t.$$

By the Markov inequality, the following equation can be determined:

$$\sum_{t=1}^{\infty}P(\rho^t(\varepsilon_t-1)^2>\epsilon)\leq\sum_{t=1}^{\infty}\frac{E(\rho^t(\varepsilon_t-1)^2)^m}{\epsilon^m}<\infty.$$

From the Borel-Cantelli lemma, (i) is obtained.

(ii) It's obvious that the result (ii) can be easily proved with Assumption 2 and Assumption 4.

(iii) Because of $E_{\theta_0}l_t^-(\theta)\leq E_{\theta_0}\ln^-\sigma_t^2\leq\max\{0,-\ln\omega\}<\infty$. It remains to be shown that $E_{\theta_0}l_t^+(\theta)<\infty$. By Jenson inequality,

$$E_{\theta_0}\log\sigma_t(\theta_0)=\frac{1}{2}E_{\theta_0}\log\sigma_t^2(\theta_0)=\frac{1}{2}E_{\theta_0}\frac{1}{m}\log\{\sigma_t^2(\theta_0)\}^m\leq\frac{1}{2m}\log E_{\theta_0}\{\sigma_t^2(\theta_0)\}^m<\infty,$$

Thus,

$$E_{\theta_0}l_t(\theta_0)=E_{\theta_0}\left\{\frac{1}{2}\log\sigma_t^2(\theta_0)+\frac{|\varepsilon_t-1|}{\sigma_t(\theta_0)}\right\}=1+\frac{1}{2}E_{\theta_0}\log\sigma_t^2(\theta_0)<\infty.$$

Therefore,

$$E_{\theta_0}l_t(\theta)-E_{\theta_0}l_t(\theta_0)=E_{\theta_0}\ln\frac{\sigma_t(\theta)}{\sigma_t(\theta_0)}+E_{\theta_0}\frac{\sigma_t(\theta_0)}{\sigma_t(\theta)}-1.$$

For all $x>0$, $\log x\leq x-1$, where the equality is true if and only if $x=1$. Thus, it is true that

$$E_{\theta_0}l_t(\theta)-E_{\theta_0}l_t(\theta_0)\geq E_{\theta_0}\left\{\log\frac{\sigma_t(\theta)}{\sigma_t(\theta_0)}+\log\frac{\sigma_t(\theta_0)}{\sigma_t(\theta)}\right\}=0 \tag{16}$$

It is noted that the equality in (16) holds if $\sigma_t(\theta_0)=\sigma_t(\theta)$.

(iv) From result (i),

$$\liminf_{n\to\infty}\inf_{\theta^*\in V_k(\theta)\cap\Theta}\tilde{I}_n(\theta^*)\geq\liminf_{n\to\infty}\inf_{\theta^*\in V_k(\theta)\cap\Theta}I_n(\theta^*)-\limsup_{n\to\infty}\sup_{\theta\in\Theta}|\tilde{I}_n(\theta)-I_n(\theta^*)|$$

$$\geq\liminf_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\inf_{\theta^*\in V_k(\theta)\cap\Theta}l_t(\theta^*)$$

Based on the ergodic theorem, Beppo-Levi theorem and the formula (16), the result (iv) can be proved.

By compactness theory, the proof of Theorem 2.1 is finished. □

Proof of Theorem 2.2. Through a Taylor expansion at θ_0,

$$\begin{aligned}0&=n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}l_t(\hat{\theta}_n)\\&=n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}l_t(\theta_0)+\left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta_i\theta_j}l_t(\theta_{ij}^*)\right)\sqrt{n}(\hat{\theta}_n-\theta_0),\end{aligned}$$

which indicates that both

$$n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}l_t(\theta_0)\Rightarrow N(0,(\frac{\tau-1}{4})J) \tag{17}$$

and

$$n^{-1}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta_i\theta_j}l_t(\theta_{ij}^*)\to J(i,j)\quad in\ probability. \tag{18}$$

hold. The proof of Theorem 2.2 is divided into the following six conclusions.

(i) $E_{\theta_0}\|(\partial l_t(\theta_0)/\partial\theta)(\partial l_t(\theta_0)/\partial\theta^T)\| < \infty$, $E_{\theta_0}\|\partial^2 l_t(\theta_0)/\partial\theta\partial\theta^T\| < \infty$.
(ii) J is nonsingular, $var_{\theta_0}\{\partial l_t(\theta_0)/\partial\theta\} = MJ$.
(iii) There exists a neighborhood $V(\theta_0)$ of θ, with regard to $i, j, k \in \{1, \cdots, r+s+1\}$, such that

$$E_{\theta_0} \sup_{\theta \in V(\theta_0)} \left| \frac{\partial^3 l_t(\theta)}{\partial\theta_i \partial\theta_j \partial\theta_k} \right| < \infty.$$

(iv) $\|n^{-1/2}\sum_{t=1}^n\{\partial l_t(\theta_0)/\partial\theta - \partial\tilde{l}_t(\theta_0)/\partial\theta\}\| \to 0$ when $n \to \infty$, $\sup_{\theta\in V(\theta_0)}\|n^{-1}\sum_{t=1}^n\{\partial^2 l_t(\theta)/\partial\theta\partial\theta^T - \partial^2\tilde{l}_t(\theta)/\partial\theta\partial\theta^T\}\| \to 0$ in probability.
(v) $n^{-1/2}\sum_{t=1}^n \partial l_t(\theta_0)/\partial\theta \Rightarrow N(0, MJ)$.
(vi) $n^{-1}\sum_{t=1}^n \partial^2 l_t(\theta_{ij}^*)/\partial\theta_i\partial\theta_j \to J(i,j)\ \ a.s.$
(i) Because of $l_t(\theta) = \ln\sqrt{\sigma_t^2} + |\varepsilon_t - 1|/\sqrt{\sigma_t^2}$, it is true that

$$\begin{aligned}\frac{\partial l_t}{\partial\theta} &= \frac{1}{2\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta} + \left(-\frac{|\varepsilon_t - 1|}{2}\frac{1}{(\sigma_t^2)^{3/2}}\right)\frac{\partial\sigma_t^2}{\partial\theta} \\ &= \frac{1}{2}\left(1 - \frac{|\varepsilon_t - 1|}{\sqrt{\sigma_t^2}}\right)\left(\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}\right),\end{aligned} \tag{19}$$

$$\begin{aligned}\frac{\partial^2 l_t}{\partial\theta\partial\theta^T} &= \frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta^T}\left(\frac{1}{2\sigma_t^2} - \frac{|\varepsilon_t - 1|}{2}\frac{1}{\sqrt{\sigma_t^6}}\right) + \frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta^T}\left(\frac{3|\varepsilon_t - 1|}{4}\frac{1}{\sqrt{\sigma_t^{10}}} - \frac{1}{2}\frac{1}{\sigma_t^4}\right) \\ &= \frac{1}{2}\left(\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta^T}\right)\left(1 - \frac{|\varepsilon_t - 1|}{\sqrt{\sigma_t^2}}\right) + \frac{1}{2}\left(\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta^T}\right)\left(\frac{3}{2}\frac{|\varepsilon_t - 1|}{\sqrt{\sigma_t^2}} - 1\right).\end{aligned} \tag{20}$$

For $\theta = \theta_0$, we have

$$E_{\theta_0}\left\|\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}(\theta_0)\right\| < \infty,\ \ E_{\theta_0}\left\|\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta^T}(\theta_0)\right\| < \infty,$$
$$E_{\theta_0}\left\|\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta^T}(\theta_0)\right\| < \infty.$$

The proof of (i) is finished.

(ii) From (i), it have

$$E_{\theta_0}\left\{\frac{\partial l_t(\theta_0)}{\partial\theta}\right\} = E_{\theta_0}\left(\frac{1}{2} - \frac{|\varepsilon_t - 1|}{2\sqrt{\sigma_t^2}}\right)E_{\theta_0}\left\{\frac{1}{\sigma_t^2(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\right\} = 0.$$

Since (20), (i) and (9), the following must also be determined.

$$\begin{aligned}var_{\theta_0}\left\{\frac{\partial l_t(\theta_0)}{\partial\theta}\right\} &= E_{\theta_0}\left\{\frac{\partial l_t(\theta_0)}{\partial\theta}\frac{\partial l_t(\theta_0)}{\partial\theta'}\right\} \\ &= E\left\{\frac{(1 - \frac{|\varepsilon_t - 1|}{\sigma_t})^2}{4}\right\}E_{\theta_0}\left\{\frac{1}{\sigma_t^4(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}\right\} \\ &= \left(\frac{\tau - 1}{4}\right)J.\end{aligned} \tag{21}$$

This shows that J is non-singular. So we establish the conclusion of (ii).

(iii)It is shown in (20) that $l_t(\theta)$ is differentiated. Then

$$\begin{aligned}\frac{\partial^3 l_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k} &= \frac{1}{2}\left(1 - \frac{|\varepsilon_t - 1|}{\sqrt{\sigma_t^2}}\right)\left(\frac{1}{\sigma_t^2}\frac{\partial^3\sigma_t^2}{\partial\theta_i\partial\theta_j\partial\theta_k}\right) \\ &+ \left(1 - \frac{15}{8}\frac{|\varepsilon_t - 1|}{\sqrt{\sigma_t^2}}\right)\left(\frac{1}{\sigma_t^6}\frac{\partial\sigma_t^2}{\partial\theta_i}\frac{\partial\sigma_t^2}{\partial\theta_j}\frac{\partial\sigma_t^2}{\partial\theta_k}\right)\end{aligned}$$

$$+\frac{1}{2}\frac{1}{\sigma_t^4}\frac{\partial^2\sigma_t^2}{\partial\theta_i\partial\theta_j}\frac{\partial\sigma_t^2}{\partial\theta_k}\left(\frac{3}{2}\frac{|\varepsilon_t-1|}{\sqrt{\sigma_t^2}}-1\right)$$
$$+\frac{1}{2}\frac{1}{\sigma_t^4}\frac{\partial^2\sigma_t^2}{\partial\theta_i\partial\theta_k}\frac{\partial\sigma_t^2}{\partial\theta_j}\left(\frac{3}{2}\frac{|\varepsilon_t-1|}{\sqrt{\sigma_t^2}}-1\right)$$
$$+\frac{1}{2}\frac{1}{\sigma_t^4}\frac{\partial^2\sigma_t^2}{\partial\theta_j\partial\theta_k}\frac{\partial\sigma_t^2}{\partial\theta_i}\left(\frac{3}{2}\frac{|\varepsilon_t-1|}{\sqrt{\sigma_t^2}}-1\right).$$

Since

$$E_{\theta_0}\sup_{\theta\in V(\theta_0)}\frac{|\varepsilon_t-1|}{\sigma_t}<\infty\ ,\quad E_{\theta_0}\sup_{\theta\in V(\theta_0)}\left|\frac{1}{\sigma_t^2}\frac{\partial^3\sigma_t^2}{\partial\theta_i\partial\theta_j\partial\theta_k}\right|<\infty,$$
$$E_{\theta_0}\sup_{\theta\in V(\theta_0)}|\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_i}|<\infty\ ,\quad E_{\theta_0}\sup_{\theta\in V(\theta_0)}\left|\frac{1}{\sigma_t^4}\frac{\partial^2\sigma_t^2}{\partial\theta_i\partial\theta_j}\right|<\infty,$$

we have

$$E_{\theta_0}\sup_{\theta\in V(\theta_0)}\left|\frac{\partial^3 l_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right|<\infty,$$

which shows that the conclusion of (iii) is true.

(iv)It follows from (3), (4), (13) and (14) that

$$\sup_{\theta\in\Theta}\left\|\frac{\partial\sigma_t^2}{\partial\theta}-\frac{\partial\tilde{\sigma}_t^2}{\partial\theta}\right\|<K\rho^t,\ \sup_{\theta\in\Theta}\left\|\frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta^T}-\frac{\partial^2\tilde{\sigma}_t^2}{\partial\theta\partial\theta^T}\right\|<K\rho^t,\quad\forall t,\tag{22}$$

which yields

$$\left|\frac{1}{\sigma_t^2}-\frac{1}{\tilde{\sigma}_t^2}\right|=\left|\frac{\tilde{\sigma}_t^2-\sigma_t^2}{\sigma_t^2\tilde{\sigma}_t^2}\right|\le\frac{K\rho^t}{\sigma_t^2},\quad\frac{\sigma_t^2}{\tilde{\sigma}_t^2}\le 1+K\rho^t.\tag{23}$$

Because of

$$\frac{\partial l_t(\theta)}{\partial\theta}=\frac{1}{2}\left(1-\frac{|\varepsilon_t-1|}{\sqrt{\sigma_t^2}}\right)\left(\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta}\right),\quad\frac{\partial\tilde{l}_t(\theta)}{\partial\theta}=\frac{1}{2}(1-\frac{|\varepsilon_t-1|}{\sqrt{\tilde{\sigma}_t^2}})(\frac{1}{\tilde{\sigma}_t^2}\frac{\partial\tilde{\sigma}_t^2}{\partial\theta}),$$

it is true that

$$\left|\frac{\partial l_t(\theta_0)}{\partial\theta}-\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta}\right|=\frac{1}{2}|\left\{\frac{|\varepsilon_t-1|}{\sqrt{\tilde{\sigma}_t^2}}-\frac{|\varepsilon_t-1|}{\sqrt{\sigma_t^2}}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_i}\right\}$$
$$+\left\{1-\frac{|\varepsilon_t-1|}{\sqrt{\tilde{\sigma}_t^2}}\right\}\left\{\frac{1}{\sigma_t^2}-\frac{1}{\tilde{\sigma}_t^2}\right\}\left\{\frac{\partial\sigma_t^2}{\partial\theta_i}\right\}$$
$$+\left\{1-\frac{|\varepsilon_t-1|}{\sqrt{\tilde{\sigma}_t^2}}\right\}\left\{\frac{1}{\tilde{\sigma}_t^2}\right\}\left\{\frac{\partial\sigma_t^2}{\partial\theta_i}-\frac{\partial\tilde{\sigma}_t^2}{\partial\theta_i}\right\}|(\theta_0)$$
$$\le\frac{1}{2}K\rho^t(1+\frac{|\varepsilon_t-1|}{\sigma_t})\left|1+\left\{\frac{1}{\sigma_t^2(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta_i}\right\}\right|.$$

Thus,

$$\left|n^{-1/2}\sum_{t=1}^n\left\{\frac{\partial l_t(\theta_0)}{\partial\theta}-\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta}\right\}\right|$$
$$\le\frac{K}{2}n^{-1/2}\sum_{t=1}^n\rho^t(1+\frac{|\varepsilon_t-1|}{\sigma_t})\left\{1+\frac{1}{\sigma_t^2(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta_i}\right\}.$$

Similarly to the proof (i), according to the Markov inequality and the independent relationship between η_t and $\sigma_t^2(\theta_0)$, for all $\varepsilon>0$ we have

$$P\left(n^{-1/2}\sum_{t=1}^n\rho^t(1+\frac{|\varepsilon_t-1|}{\sigma_t})\left\{1+\frac{1}{\sigma_t^2(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\right\}>\varepsilon\right)$$

$$\leq \frac{2}{\varepsilon}\left(1 + E_{\theta_0}\left|\frac{1}{\sigma_t^2(\theta_0)}\frac{\partial \sigma_t^2(\theta_0)}{\partial \theta}\right|\right) n^{-1/2}\sum_{t=1}^{n}\rho^t \to 0. \tag{24}$$

Thus, the first part of (iv) was obtained. Due to (20), (22), and (23), we have

$$\begin{aligned}
&\sup_{\theta\in V(\theta_0)}\left|n^{-1}\sum_{t=1}^{n}\left\{\frac{\partial^2 l_t(\theta)}{\partial\theta_i\partial\theta_j} - \frac{\partial^2 \tilde{l}_t(\theta)}{\partial\theta_i\partial\theta_j}\right\}\right| \\
&\leq \frac{n^{-1}}{2}\sum_{t=1}^{n}\sup_{\theta\in V(\theta_0)} |\left\{\frac{|\varepsilon_t - 1|}{\sqrt{\tilde{\sigma}_t^2}} - \frac{|\varepsilon_t - 1|}{\sqrt{\sigma_t^2}}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta_i\partial\theta_j}\right\} \\
&+ \left\{1 - \frac{|\varepsilon_t - 1|}{\sqrt{\tilde{\sigma}_t^2}}\right\}\left\{\left(\frac{1}{\sigma_t^2} - \frac{1}{\tilde{\sigma}_t^2}\right)\frac{\partial^2\sigma_t^2}{\partial\theta_i\partial\theta_j} + \frac{1}{\tilde{\sigma}_t^2}\left(\frac{\partial^2\sigma_t^2}{\partial\theta_i\partial\theta_j} - \frac{\partial^2\tilde{\sigma}_t^2}{\partial\theta_i\partial\theta_j}\right)\right\} \\
&+ \left\{\frac{3}{2}\frac{|\varepsilon_t - 1|}{\sqrt{\sigma_t^2}} - \frac{3}{2}\frac{|\varepsilon_t - 1|}{\sqrt{\tilde{\sigma}_t^2}}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_i}\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_j}\right\} \\
&+ \left\{\frac{3}{2}\frac{|\varepsilon_t - 1|}{\sqrt{\tilde{\sigma}_t^2}} - 1\right\}\left\{\left(\frac{1}{\sigma_t^2} - \frac{1}{\tilde{\sigma}_t^2}\right)\frac{\partial\sigma_t^2}{\partial\theta_i} + \frac{1}{\tilde{\sigma}_t^2}\left(\frac{\partial\sigma_t^2}{\partial\theta_i} - \frac{\partial\tilde{\sigma}_t^2}{\partial\theta_i}\right)\right\}\left\{\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_j}\right\} \\
&+ \left\{\frac{3}{2}\frac{|\varepsilon_t - 1|}{\sqrt{\tilde{\sigma}_t^2}} - 1\right\}\left\{\frac{1}{\tilde{\sigma}_t^2}\frac{\partial\tilde{\sigma}_t^2}{\partial\theta_i}\right\}\left\{\left(\frac{1}{\sigma_t^2} - \frac{1}{\tilde{\sigma}_t^2}\right)\frac{\partial\sigma_t^2}{\partial\theta_j} + \frac{1}{\tilde{\sigma}_t^2}\left(\frac{\partial\sigma_t^2}{\partial\theta_j} - \frac{\partial\tilde{\sigma}_t^2}{\partial\theta_j}\right)\right\}| \\
&\leq \frac{1}{2}Kn^{-1}\sum_{t=1}^{n}\rho^t N_t.
\end{aligned}$$

As a consequence,

$$N_t = \sup_{\theta\in V(\theta_0)}\left\{1 + \frac{|\varepsilon_t - 1|}{\sqrt{\sigma_t^2}}\right\}\left\{1 + \frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta_i\partial\theta_j} + \frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_i}\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_j}\right\}.$$

By (iii) and Holder inequality, N_t was integrable for some neighbourhood $V(\theta_0)$. So, it follows from the Markov inequality that the second part of (iv) is true.

(v) The proof of (v) is easily obtained from the central limit theorem for martingale difference. Suppose that $\mathcal{F}_t$ is σ-domain generated from varibles $\{\varepsilon_{t-i}, i \geq 0\}$. As $E_{\theta_0}(\partial l_t(\theta_0)/\partial\theta|\mathcal{F}_t) = 0$, $var_{\theta_0}(\partial l_t(\theta_0)/\partial\theta)$ exists. From Assumption 3 and (ii), $0 < \tau - 1 < \infty$ and J is non-singular. Hence, the matrix $var_{\theta_0}(\partial l_t(\theta_0)/\partial\theta)$ is non-degenerate. Thus, $\lambda \in R^{p+q+1}$, $\{\lambda^T(\partial/\partial\theta)l_t(\theta_0), \mathcal{F}_t$ is a martingale difference sequence. From the central limit theorem and the Wold-Cramer device, the asymptotic normality result (v) is established.

(vi) To prove (vi), we firstly prove that the second-order derivatives of $l_t(\theta)$ exists. For all i, j,

$$\begin{aligned}
n^{-1}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta_i\partial\theta_j}l_t(\theta_{ij}^*) = n^{-1}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta_i\partial\theta_j}l_t(\theta_0) + n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial\theta^T}\{\frac{\partial^2}{\partial\theta_i\partial\theta_j} \\
l_t(\tilde{\theta}_{ij})\}(\theta_{ij}^* - \theta_0),
\end{aligned} \tag{25}$$

Here, $\tilde{\theta}_{ij}$ locates between θ_{ij}^* and θ_0. As $\tilde{\theta}_{ij}$ almost certainly converges to θ_0, it follows from the ergodic theorem and (iii) that

$$\begin{aligned}
&\lim_{n\to\infty}\sup\left\|n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial\theta^T}\left\{\frac{\partial^2}{\partial\theta_i\partial\theta_j}l_t(\tilde{\theta}_{ij})\right\}\right\| \\
&\leq \lim_{n\to\infty}\sup n^{-1}\sum_{t=1}^{n}\sup_{\theta\in V(\theta_0)}\left\|\frac{\partial}{\partial\theta^T}\left\{\frac{\partial^2}{\partial\theta_i\partial\theta_j}l_t(\theta)\right\}\right\| \\
&= E_{\theta_0}\sup_{\theta\in V(\theta_0)}\left\|\frac{\partial}{\partial\theta^T}\left\{\frac{\partial^2}{\partial\theta_i\partial\theta_j}l_t(\theta)\right\}\right\| < \infty.
\end{aligned}$$

Since $\|\theta_{ij}^* - \theta_0\| \to 0 \ a.s.$, the second term on the right-hand side of (25) converges to 0 with probability 1. The first term on the right-hand side of (25) is also proved by the ergodic theorem. As a result, the conclusion of (vi) is obtained immediately.

Finally, the Slutsky lemma, (iv), (v), and (vi) are used to produce (17) and (18), i.e. the conclusion of Theorem 2.2 is established. Here, we complete the proof. □

Conflict of interest statement

We declare that we have no commercial or associative interest conflicts of interest in this work, and we have no financial and personal relationships with other people or organizations which can inappropriately influence our work, the manuscript which have no conflict of interest, entitled "Quasi-maximum Likelihood Estimator of Laplace (1, 1) for GARCH Models".

Acknowledgement: The authors would like to thank the reviewers for their valuable comments and suggestions. This work was supported by National Natural Science Foundation of China (No.11371077).

References

[1] Engle R. F., Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, Econometrica, 1982, 50, 987-1007.
[2] Bollerslev T., Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics, 1986, 31, 307-327.
[3] David A., Lennart F. H., GARCH models for daily stock returns impact of estimation frequency on Value-at-Risk and expected shortfall forecasts, Economics Letters, 2014, 123, 187-190.
[4] Yang Y. L., Chang C. L., A double-threshold GARCH model of stock market and currency shocks on stock returns, Mathematics and Computers in Simulation, 2008, 79, 458-474.
[5] Weiss A. A., Asymptotic theory for ARCH models: estimation and testing, Econometric Theory, 1986, 2, 107-131.
[6] Lee S. W., Hansen B. E., Asymptotic theory for the GARCH(1, 1) quasi-maximum likelihood estimator, Econometric Theory, 1994, 10, 29-52.
[7] Lumsdaine B. L., Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1, 1) and covariance stationary GARCH(1, 1) models, Econometrica, 1996, 64, 575-596.
[8] Berkes I., Horvath L., Kokoszka P., GARCH processes: structure and estimation, Bernoulli, 2003, 9, 201-227.
[9] Berkes I., Horvath L., The rate of consistency of the quasi-maximum likelihood estimator, Statistics and Probability Letters, 2003, 61, 133-143.
[10] Berkes I., Horvath L., The efficiency of the estimators of the parameters in GARCH processes, The Annals of Statistics, 2004, 32, 633-655.
[11] Francq C., Zakoian J. M., Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes, Bernoulli, 2004, 10, 605-637.
[12] Straumann D., Mikosch T., Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equation approach, The Annals of statistics, 2006, 34, 2449-2495.
[13] Ling S. Q., Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models, Journal of Econometrics, 2007, 140, 849-873.
[14] Zhu K., Ling S. Q., Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA-GARCH/IGARCH models, The Annals of Statistics, 2011, 39, 2131-2163.
[15] Han H., Kristensen D., Asymptotic Theory for the QMLE in GARCH-X Models With Stationary and Nonstationary Covariates, Journal of business & economic statistics, 2014, 32, 416-429.
[16] Nelson D. B., Conditional heteroskedasticity in asset returns: a new approach, Econometrica, 1991, 59, 347-370.
[17] Bougerol P., Picard N., Stationarity of GARCH processes and of some nonnegative time series, Econometrics, 1992, 52, 115-127.