Data Science in Finance and Economics

http://www.aimspress.com/journal/dsfe

DSFE, 1(2): 165–179.
DOI: 10.3934/DSFE.2021009
Received: 29 July 2021
Accepted: 27 September 2021
Published: 08 October 2021

Research article

Daily LGARCH model estimation using high frequency data

Xiaoling Chen , Xingfa Zhang*, Yuan Li and Qiang Xiong

School of Economics and Statistics, Guangzhou University, Guangzhou, 510006, China

* **Correspondence:** Email: xingfazhang@gzhu.edu.cn.

Abstract: In this paper, we introduce the intraday high frequency data to estimate the daily linear generalized autoregressive conditional heteroscedasticity (LGARCH) model. Based on the volatility proxies constructed from the intraday high frequency data, the quasi maximum likelihood estimation (QMLE) of the daily LGARCH model and its asymptotic distribution are studied under some regular assumptions. One criterion is also given to choose the optimal volatility proxy according to the asymptotic results. Simulation studies show that the QMLE of the parameters performs well. It is also found that introducing the intraday high frequency data can significantly improve the estimation precision. The proposed method is applied to analyze the SSE 50 Index, which consists of the 50 largest and most liquid A-share stocks listed on Shanghai Stock Exchange. Empirical results show the method is of potential application value.

Keywords: LGARCH model; intraday high frequency data; quasi maximum likelihood estimation

JEL Codes: C13, C22

1. Introduction

Volatility clustering is a well-known characteristic of financial time series. Accurately describing volatility is helpful for pricing and risk management of financial assets. Many conditional heteroscedasticity models have been proposed to describe the time varying volatility. Among them, the autoregressive conditional heteroscedasticity (ARCH) model proposed by Engle (1982) and the generalized autoregressive conditional heteroscedasticity (GARCH) model proposed by Bollerslev (1986) have been widely studied, especially in the financial industry. For example, Nelson (1991) applied GARCH model to asset pricing, Zou et al. (2015) used GARCH model to estimate the combination of market investment risk value, and De Davide (2019) used GARCH model to analyze and predict S & P 500 index. In many cases, the impact of assets on the market is asymmetric. That is to say, investors react differently to the same amount of good news and bad news. Therefore, in order to reflect this phenomenon, more and more scholars are involved in the study of asymmetric GARCH model, see, e.g., Hentschel (1995), Pan et al. (2008), Gyamerah (2019) and Linton et al. (2020). In addition, considering the cyclical factors of market fluctuations, many scholars began to study periodic GARCH model, such as Zhao et al. (2016).

As noted by Duffie and Pan (1997), maximum likelihood estimation of the GARCH type model has the potential disadvantage of being overly sensitive to extreme returns. For example, if we consider a market crash, then extreme daily absolute returns may be 10–20 times the normal daily fluctuation, so the quadratic form of GARCH model yields a return effect that is 100–400 times the normal variance, resulting in excessive fluctuation prediction. Therefore, in order to avoid the impact of extreme returns, Xiao and Koenker (2009) proposed a LGARCH model. It has been shown that LGARCH model can produce more robust inferences, compared to the previous GARCH type models. The specific form of LGARCH (1, 1) model is as follows:

$$y_t = h_t \varepsilon_t, \tag{1}$$

$$h_t = \omega + \alpha |y_{t-1}| + \beta h_{t-1}, \tag{2}$$

where, $\omega > 0, \alpha, \beta \geqslant 0$, $\{\varepsilon_t\}_{t=1}^{\infty}$ is an independent identically distributed sequence with mean 0 and variance 1, namely, $\{\varepsilon_t\} \sim i.i.d(0,1)$, and y_s is independent of $\{\varepsilon_t : t \geq 1\}$ for $t > s$. Let $\mathcal{F}_t$ be the $\sigma-$field generated by $\{\varepsilon_t, \ldots, \varepsilon_1, y_0, y_{-1} \ldots\}$. Given $\mathcal{F}_{t-1}$, the conditional mean of y_t is $E(y_t|\mathcal{F}_{t-1}) = 0$, and the conditional variance of y_t is $Var(y_t|\mathcal{F}_{t-1}) = E(h_t^2 \varepsilon_t^2|\mathcal{F}_{t-1}) = h_t^2 E(\varepsilon_t^2|\mathcal{F}_{t-1}) = h_t^2$.

With the development of electronic information technology, it is easier to obtain intraday high frequency data in the financial market, and such data usually contain lots of useful information and are valuable in improving model estimation. To achieve a more precise parameter estimator of common GARCH model, Visser (2011) proposed a volatility proxy model, embedding intraday high frequency data into the framework of daily GARCH model. The volatility proxy model not only maintains the parameter structure of daily GARCH model, but also introduces the intraday high frequency data. Available results show that the variance of parameter estimator in GARCH (1,1) model can be reduced 20 times by selecting an appropriate volatility proxy, which greatly improves the estimation accuracy of model parameters. Many scholars have further extended the results of Visser (2011) to other cases. For example, Wang et al. (2018) proposed a compound quantile regression (CQR) method to estimate the GARCH model based on high frequency data, and proved the asymptotic normality of the estimators without strong moment conditions; Wu et al. (2018) studied the quasi maximum exponential likelihood estimation (QMELE) of non-stationary GARCH (1,1) model under high frequency data, and obtained the limiting properties under weak moment conditions; Fan et al. (2017) studied the VaR estimation based on periodic GARCH model with high frequency data; Deng et al. (2020) studied the parameter test of GARCH model, where the parameter estimators were obtained from intraday high frequency data, and the corrected likelihood ratio test and Wald test statistics were further investigated.

In the literature, few studies have been done about introducing the intraday high frequency data to the estimation of daily LGARCH model. However, as mentioned above that LGARCH model is more robust than other GARCH type models. Therefore, it makes sense to introduce the intraday high frequency data to estimate the daily LGARCH model, which is a main contribution of this paper. Another contribution is that the proposed estimation method is adopted to all the parameters of the model and it is not necessary to set ω in (2) to be 1 as before, see, for example, Visser (2011) and Wang et al. (2018). The rest of this paper is organized as follows. In Section 2, we introduce the volatility proxy model and estimators. In Section 3, we derive the asymptotic results of the model estimator.

Simulations and empirical studies are respectively shown in Section 4 and Section 5. We conclude the article in Section 6.

2. Model and estimation

2.1. Volatility proxy model

Let $\theta = (\omega, \alpha, \beta)'$ be the parameter vector for model (1)–(2). In order to introduce intraday high-frequency data, it is necessary to extend the LGARCH model to the volatility proxy model. Denote $Y_t(u)$ to be the logarithmic return of an asset at time u on day t, where the time of each trading day is standardized to the interval of [0, 1]. From Visser (2011), we firstly consider the following scale model:

$$Y_t(u) = h_t Z_t(u), \tag{3}$$

$$h_t = \omega + \alpha|y_{t-1}| + \beta h_{t-1}. \tag{4}$$

here, $0 \leq u \leq 1$, h_t is the volatility of day t. Standard process $Z_t(u)$ satisfies: when $t \neq s$, $Z_t(u)$ is independent of $Z_s(u)$ and has the same distribution as $Z_s(u)$. When $u = 1$, ($y_t \equiv Y_t(1)$, $\varepsilon_t \equiv Z_t(1)$, $EZ_t^2(1) = 1$) , the model (3)–(4) degenerates into model (1)–(2). It is easy to see that the scale model introduces intraday data information $Y_t(u)$, and it retains the parameter structure of the daily LGARCH model. Unfortunately, model (3)–(4) can not be directly estimated due to the inconsistent frequency between h_t and $Y_t(u)$. In order to estimate θ, we further need to construct the intraday high-frequency data into a daily volatility proxy.

The volatility proxy is a daily sequence based on the intraday data. That is, for intraday yield process $Y_t(u)$, let $H_t \equiv H(Y_t(u))$ be a volatility proxy for $Y_t(u)$, where $H(\cdot)$ is a given function. Common volatility proxies include realized volatility and intraday price range. Positive homogeneity is an important property of volatility proxy. Namely, for a non-zero constant ρ ($\rho > 0$), the following equality holds:

$$H(\rho Y_t(u)) = \rho H(Y_t(u)) > 0. \tag{5}$$

Consider a given function $H(\cdot)$ satisfying the positive homogeneity and apply $H(\cdot)$ to equation 3. Then it is obtained that

$$H_t \equiv H(Y_t(u)) = H(h_t Z_t(u)) = h_t H(Z_t(u)) > 0. \tag{6}$$

Define

$$z_{H,t} = H(Z_t(u)), \mu = E z_{H,t}^2, \varepsilon_t^* = \frac{z_{H,t}}{\sqrt{\mu}}. \tag{7}$$

Because the standard process $Z_t(\cdot)$ is independently and identically distributed, hence ε_t^* is an i.i.d. sequence with $E\varepsilon_t^{*2} = 1$. Then, combined with equations (3)–(7), we have the following volatility proxy model,

$$H_t = h_t z_{H,t} = h_t \sqrt{\mu} \cdot \varepsilon_t^*, \tag{8}$$

$$h_t = \omega + \alpha|y_{t-1}| + \beta h_{t-1}. \tag{9}$$

In the above model, a redundant parameter μ appears due to the setting of ε_t^*, which makes it impossible to directly apply the QMLE method to estimate θ and μ simultaneously. Next, we give an indirect approach to estimate the parameters. Let $h_t^* = h_t \sqrt{\mu}$, then (8)–(9) can be rewritten as follows:

$$H_t = h_t^* \varepsilon_t^*, \tag{10}$$

$$h_t^* = \omega^* + \alpha^*|y_{t-1}| + \beta^* h_{t-1}^*, \tag{11}$$

where

$$\omega^* = \omega\sqrt{\mu}, \alpha^* = \alpha\sqrt{\mu}, \beta^* = \beta. \tag{12}$$

When volatility proxy $H_t = |y_t|$, it is easy to have $\mu = 1, h_t^* = h_t$. Therefore, model (1)–(2) is a special case of model (10)–(11). According to (10)–(11), we can use the QMLE method to estimate the $\theta^* = (\omega^*, \alpha^*, \beta^*)'$. Once an estimator for μ is given, then we can get the estimation of the parameter $\theta = (\omega, \alpha, \beta)'$ by using (12). The detailed estimation procedures are given in the next section.

2.2. *Parameter estimation*

Define $\theta^* = (\omega^*, \alpha^*, \beta^*)' \in \Theta$, where $\Theta \subseteq \mathcal{R}^3$ is a parameter space for model (10)–(11). In addition, suppose that $\theta_0^* = (\omega_0^*, \alpha_0^*, \beta_0^*)'$ is the true value of the parameter θ^*, which is an interior point of the parameter space Θ.

Following the convention in the literature, see Visser (2011), we consider the quasi conditional log-likelihood function (apart from a constant term).

$$L_T(\theta^*) = \frac{1}{n}\sum_{t=1}^{n} l_t(\theta^*), \quad l_t(\theta^*) = \log h_t^{*2}(\theta^*) + \frac{H_t^2}{h_t^{*2}(\theta^*)}. \tag{13}$$

Then, the QMLE of parameter θ^* can be defined as follows:

$$\hat{\theta}^* = \underset{\theta^*\in\Theta}{argmin}\, L_T(\theta^*). \tag{14}$$

According to (12), to further estimate θ, we need to know μ. After $\hat{\theta}^*$ is estimated, the fitting sequence $\{\hat{h}_t^*\}$ is obtained from (11). It is already known that the absolute value of return $|y_t|$ can also be regarded as a special volatility proxy ($H_t = |y_t|$, $\varepsilon_t^* = |\varepsilon_t|$). When $H_t = |y_t|$, the estimated parameter obtained by the likelihood function (13) is actually the estimator of θ in the LGARCH (1,1) model (1)–(2), which only uses the information of daily data $\{y_t\}$ and no intraday high frequency data is introduced. When $H_t = |y_t|$, the estimated value of θ is denoted as $\tilde{\theta} = (\tilde{\omega}, \tilde{\alpha}, \tilde{\beta})'$ and the corresponding fitting series for h_t is denoted as $\{\tilde{h}_t\}$. According to $h_t^* = h_t\sqrt{\mu}$ or $\mu = h_t^{*2}/h_t^2$, we can get an estimator of μ as followed:

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T}\frac{\hat{h}_t^{*2}}{\tilde{h}_t^2}. \tag{15}$$

Finally, the parameter estimators of LGARCH model with high-frequency information, denoted by $\hat{\theta} = (\hat{\omega}, \hat{\alpha}, \hat{\beta})'$, are given by

$$\hat{\omega} = \frac{\hat{\omega}^*}{\sqrt{\hat{\mu}}},\ \hat{\alpha} = \frac{\hat{\alpha}^*}{\sqrt{\hat{\mu}}},\ \hat{\beta} = \hat{\beta}^*. \tag{16}$$

Different from the usual estimator $\tilde{\theta} = (\tilde{\omega}, \tilde{\alpha}, \tilde{\beta})'$ corresponding to $H_t = |y_t|$, the estimator $\hat{\theta} = (\hat{\omega}, \hat{\alpha}, \hat{\beta})'$ generally contains intraday high frequency data information and hence is expected to have a better performance.

3. Asymptotic theory

Before stating the asymptotic results for $\hat{\theta}^*$, we firstly make the following assumptions.
A1 The parameter space Θ is compact, and θ_0^* is an interior point of Θ.
A2 The i.i.d. random sequence $\{\varepsilon_t^*\}$ satisfies $E(\varepsilon_t^{*4}) < \infty$, and there exists a positive and continuous probability density function almost everywhere.
A3 The series $\{H_t, h_t^*\}$ generated from model (10)–(11) are strictly stationary and geometrically ergodic for the considered parameter space Θ.

Based on assumptions (A1)–(A3), adopting similar arguments to Visser (2011), it is not difficult to prove the asymptotic distribution of $\hat{\theta}^*$,

$$\sqrt{T}(\hat{\theta}^* - \theta_0^*) \xrightarrow{L} N(0, \Sigma^*), \quad T \longrightarrow \infty, \tag{17}$$

where $\Sigma^* = \Omega_I^{-1}\Omega_S\Omega_I^{-1}$,

$$\Omega_I = E\left(\frac{\partial^2 l_t(\theta_0^*)}{\partial\theta_i^*\partial\theta_j^*}\right), \tag{18}$$

$$\Omega_S = E\left(\frac{\partial l_t(\theta_0^*)}{\partial\theta_i^*}\frac{\partial l_t(\theta_0^*)}{\partial\theta_j^*}\right). \tag{19}$$

To obtain the asymptotic variance of $\hat{\theta}^*$, we give the partial derivatives of the likelihood function $l_t(\theta^*)$ with respect to θ^* as follows:

$$\frac{\partial l_t(\theta^*)}{\partial\theta^*} = 2(1 - \frac{H_t^2}{h_t^{*2}(\theta^*)})\frac{1}{h_t^*(\theta^*)}\frac{\partial h_t^*(\theta^*)}{\partial\theta^*} = 2(1 - \varepsilon_t^{*2})\frac{1}{h_t^*(\theta^*)}\frac{\partial h_t^*(\theta^*)}{\partial\theta^*}. \tag{20}$$

$$\begin{aligned}\frac{\partial^2 l_t(\theta^*)}{\partial\theta_i^*\partial\theta_j^*} &= -\frac{2}{h_t^{*2}(\theta^*)}\left(1 - \frac{3H_t^2}{h_t^{*2}(\theta^*)}\right)\frac{\partial h_t^*(\theta^*)}{\partial\theta_i^*}\frac{\partial h_t^*(\theta^*)}{\partial\theta_j} + \frac{2}{h_t^*(\theta^*)}\left(1 - \frac{H_t^2}{h_t^{*2}(\theta^*)}\right)\frac{\partial^2 h_t^*(\theta^*)}{\partial\theta_i^*\partial\theta_j^*}\\ &= -\frac{2}{h_t^{*2}(\theta^*)}\left(1 - 3\varepsilon_t^{*2}\right)\frac{\partial h_t^*(\theta^*)}{\partial\theta_i^*}\frac{\partial h_t^*(\theta^*)}{\partial\theta_j} + \frac{2}{h_t^*(\theta^*)}\left(1 - \varepsilon_t^{*2}\right)\frac{\partial^2 h_t^*(\theta^*)}{\partial\theta_i^*\partial\theta_j^*}.\end{aligned} \tag{21}$$

Further,

$$\Omega_I = E\left(\frac{\partial^2 l_t(\theta_0^*)}{\partial\theta_i^*\partial\theta_j^*}\right) = 4E\left(\frac{1}{h_t^{*2}(\theta_0^*)}\frac{\partial h_t^*(\theta_0^*)}{\partial\theta_i^*}\frac{\partial h_t^*(\theta^*)}{\partial\theta_j^*}\right). \tag{22}$$

$$\Omega_S = E\left(\frac{\partial l_t(\theta_0^*)}{\partial\theta_i^*}\frac{\partial l_t(\theta_0^*)}{\partial\theta_j^*}\right) = 4E\left(\varepsilon_t^{*2} - 1\right)^2 E\left(\frac{1}{h_t^{*2}(\theta_0^*)}\frac{\partial h_t^*(\theta_0^*)}{\partial\theta_i^*}\frac{\partial h_t^*(\theta^*)}{\partial\theta_j^*}\right). \tag{23}$$

In terms of (7), $E\varepsilon_t^{*2} = 1$, $E\left(\varepsilon_t^{*2} - 1\right)^2 = Var(\varepsilon_t^{*2})$. Therefore,

$$\Sigma^* = \frac{1}{4}Var(\varepsilon_t^{*2})G(\theta_0^*)^{-1}, \tag{24}$$

where,

$$G(\theta_0^*)_{i,j} = E\left(\frac{1}{h_t^{*2}(\theta_0^*)}\frac{\partial h_t^*(\theta_0^*)}{\partial\theta_i^*}\frac{\partial h_t^*(\theta^*)}{\partial\theta_j^*}\right). \tag{25}$$

According to (11), it can be obtained that:

$$h_t^*(\theta^*) = \frac{\omega^*}{1-\beta^*} + \alpha^* \sum_{j=0}^{\infty} (\beta^*)^j |y_{t-j-1}|. \tag{26}$$

Further,

$$\frac{\partial h_t^*(\theta^*)}{\partial \omega^*} = \frac{1}{1-\beta^*}, \tag{27}$$

$$\frac{\partial h_t^*(\theta^*)}{\partial \alpha^*} = \sum_{j=0}^{\infty} (\beta^*)^j |y_{t-j-1}| = \sum_{j=1}^{\infty} (\beta^*)^j |y_{t-j}|, \tag{28}$$

$$\frac{\partial h_t^*(\theta^*)}{\partial \beta^*} = \frac{\omega^*}{(1-\beta^*)^2} + \alpha^* \sum_{j=1}^{\infty} j(\beta^*)^{j-1} |y_{t-j-1}| = \sum_{j=1}^{\infty} (\beta^*)^{j-1} h_{t-j}^*(\theta^*). \tag{29}$$

Let $s_{\omega^*}^2$, $s_{\alpha^*}^2$, $s_{\beta^*}^2$ be the asymptotic variances of ω^*, α^*, β^* respectively. Based on the asymptotic property of $\hat{\theta}^*$, we can get the following conclusion provided the parameter μ is given:

$$\begin{aligned} \sqrt{T}(\hat{\omega} - \omega_0) &\xrightarrow{L} N\left(0, \frac{s_{\omega^*}^2}{\mu}\right), \\ \sqrt{T}(\hat{\alpha} - \alpha_0) &\xrightarrow{L} N\left(0, \frac{s_{\alpha^*}^2}{\mu}\right), \\ \sqrt{T}(\hat{\beta} - \beta_0) &\xrightarrow{L} N\left(0, s_{\beta^*}^2\right). \end{aligned} \tag{30}$$

In practice, it is important to choose a proper volatility proxy H_t for parameter estimation. According to (24), it can be seen that the impact of H_t to the asymptotic variance is based on $E\,(\varepsilon_t^{*4})$ which equals $E(z_{H,t}^4)/[E(z_{H,t}^2)]^2$ from (7). Consequently, it is expected to choose a volatility proxy H_t with small value for $E(\varepsilon_t^{*4})$. From (6)–(7), similar to Liang et al. (2021), we can obtain

$$\frac{EH_t^4}{(EH_t^2)^2} = \frac{E(h_t^4)E(z_{H,t}^4)}{[E(h_t^2)]^2[E(z_{H,t}^2)]^2} = \frac{E(h_t^4)}{[E(h_t^2)]^2}\frac{E(z_{H,t}^4)}{[E(z_{H,t}^2)]^2} = c \cdot \frac{E(z_{H,t}^4)}{[E(z_{H,t}^2)]^2}. \tag{31}$$

Here, $c = E(h_t^4)/[E(h_t^2)]^2$ is a positive constant and hence $E(z_{H,t}^4)/[E(z_{H,t}^2)]^2$ is proportional to $EH_t^4/(EH_t^2)^2$. Let

$$MH = (EH_t^4)/(EH_t^2)^2. \tag{32}$$

Then we have smaller $MH \longleftrightarrow$ smaller $E(z_{H,t}^4)/[E(z_{H,t}^2)]^2 \longleftrightarrow$ smaller $Var(\varepsilon_t^{*2})$.

From the above, among several candidates, one can choose the volatility proxy H_t according to its MH value. The optimal H_t should have the smallest MH value.

4. Simulations

In this section, we carry out Monte Carlo experiments to assess the finite-sample performance of the proposed estimators. In order to simulate the process of (3)–(4), we refer to Visser's (2011) example to simulate the intraday standard stochastic process $Z_t(u)$, which is produced by the following stochastic difference equations:

$$d\Gamma_t(u) = -\delta(\Gamma_t(u) - u_\Gamma)du + \sigma_\Gamma dB_t^{(2)}(u), \tag{33}$$

$$dZ_t(u) = \exp(\Gamma_t(u))dB_t^{(1)}(u), \quad u \in [0, 1]. \tag{34}$$

here, $B_t^{(1)}$ and $B_t^{(2)}$ are two uncorrelated Brownian motions, $Z_t(0) = 0$, and $\Gamma_t(0)$ is randomly generated from the stationary distribution $N(\mu_\Gamma, \sigma_\Gamma^2)$. The time interval [0,1] within the day is equally divided into 240 cells to correspond to the frequency of 1 min in a real trading day. The settings for parameters in (33)–(34) are $\delta = 1/2$, $\sigma_\Gamma = 1/4$ and $\mu_\Gamma = -1/16$. To further generate the stochastic process $Y_t(u)$, the parameters in equation (4) are set in the following two cases: $\theta_0 = (0.1, 0.4, 0.2)'$ and $\theta_0 = (0.1, 0.3, 0.5)'$.

We consider realized volatility (RV) as the volatility proxies H_t in (8) under different frequencies, namely, 1-minute (RV1), 5-minute (RV5), 10-minute (RV10), 15-minute (RV15) and 30-minute (RV30). For RV1, the formula is given by:

$$H_t = RV1_t = \left(\sum_{i=1}^{240} [Y_t(u_i) - Y_t(u_{i-1})]^2 \right)^{1/2}, \tag{35}$$

where, the value of $Y_t(u_0)$ is replaced by $Y_t(0) = 0$. Other volatility proxies can be computed similarly. For comparison, we also consider the case $H_t = |y_t|$ where the estimator is reduced to the usual estimator which only uses the daily data. The sample sizes are $T = 500, 1000$ and 1500 and the replication time is set to be 1000. For each H_t, its MH value in (32) is estimated as:

$$\widehat{\mathrm{MH}} = \frac{T^{-1} \sum_{t=1}^{T} H_t^4}{\left(T^{-1} \sum_{t=1}^{T} H_t^2\right)^2}. \tag{36}$$

The mean of 1000 estimated MH values is taken as the final estimated value for MH and is used to judge whether the volatility proxy is optimal.

Tables 1 and 2 summarize the empirical biases (Bias), empirical standard deviations (SD), asymptotic standard deviations (AD) and the value of MH ($\widehat{\mathrm{MH}}$) for $\hat{\theta} = (\hat{\omega}, \hat{\alpha}, \hat{\beta})'$. It can be seen that the Bias for each case is generally small. SD and AD of all parameters are very close and decrease when the sample size becomes large, which is consistent with the asymptotic results. Compared to the case $H_t = |y_t|$, the SD and AD of the parameters estimated by RV are significantly smaller, which implies that introducing the intraday high frequency data can effectively improve the precision of the estimator.

By comparing the MH values, it can be found that in the simulation examples considered, the optimal order of different volatility proxies is: $RV1_t > RV5_t > RV10_t > RV15_t > RV30_t > |y_t|$. Namely the estimator under $RV1_t$ shows the best performance. The simulation results show that the volatility proxies model with intraday high frequency data have a good effect on parameter estimation, which is helpful to improve the estimation accuracy of LGARCH model.

Table 1. Bias, SD and AD of QMLEs and $\widehat{\mathrm{MH}}$, $\theta_0 = (0.1, 0.4, 0.2)'$.

$\theta_0 = (0.1, 0.4, 0.2)'$			$\lvert y_t \rvert$	$RV1_t$	$RV5_t$	$RV10_t$	$RV15_t$	$RV30_t$
$T = 500$	$\hat{\omega}$	Bias	0.0030	−0.0002	−0.0000	0.0000	0.0002	0.0006
		SD	0.0227	0.0083	0.0088	0.0092	0.0099	0.0117
		AD	0.0210	0.0080	0.0085	0.0099	0.0116	0.0131
	$\hat{\alpha}$	Bias	−0.0043	−0.0007	−0.0010	−0.0012	−0.0024	−0.0012
		SD	0.0685	0.0255	0.0268	0.0285	0.0300	0.0352
		AD	0.0851	0.0249	0.0264	0.0308	0.0347	0.0378
	$\hat{\beta}$	Bias	−0.0136	−0.0024	−0.0032	−0.0031	−0.0033	−0.0061
		SD	0.1228	0.0407	0.0440	0.0466	0.0514	0.0621
		AD	0.1082	0.0480	0.0506	0.0590	0.0668	0.0739
	$\widehat{\mathrm{MH}}$		5.9783	2.2481	2.3255	2.4233	2.5264	2.8541
$T = 1000$	$\hat{\omega}$	Bias	0.0022	0.0002	0.0001	0.0002	0.0001	0.0006
		SD	0.0167	0.0057	0.0061	0.0064	0.0069	0.0081
		AD	0.0200	0.0053	0.0058	0.0062	0.0071	0.0082
	$\hat{\alpha}$	Bias	0.0012	0.0009	0.0007	0.0005	0.0004	0.0011
		SD	0.0473	0.0179	0.0191	0.0202	0.0214	0.0251
		AD	0.0490	0.0155	0.0168	0.0177	0.0199	0.0236
	$\hat{\beta}$	Bias	−0.0105	−0.0013	−0.0011	−0.0010	−0.0009	−0.0035
		SD	0.0925	0.0274	0.0301	0.0327	0.0357	0.0431
		AD	0.1023	0.0295	0.0324	0.0345	0.0396	0.0450
	$\widehat{\mathrm{MH}}$		5.9892	2.1409	2.2125	2.3067	2.4062	2.7358
$T = 1500$	$\hat{\omega}$	Bias	0.0011	0.0000	-0.0000	-0.0000	-0.0000	0.0002
		SD	0.0140	0.0046	0.0048	0.0051	0.0054	0.0066
		AD	0.0133	0.0038	0.0041	0.0045	0.0048	0.0056
	$\hat{\alpha}$	Bias	0.0005	−0.0005	−0.0004	−0.0002	−0.0005	0.0000
		SD	0.0423	0.0151	0.0163	0.0172	0.0185	0.0217
		AD	0.0463	0.0125	0.0135	0.0148	0.0154	0.0186
	$\hat{\beta}$	Bias	−0.0060	−0.0011	−0.0008	−0.0008	−0.0005	−0.0021
		SD	0.0766	0.0231	0.0247	0.0269	0.0292	0.0357
		AD	0.0719	0.0213	0.0231	0.0253	0.0268	0.0318
	$\widehat{\mathrm{MH}}$		5.9796	2.0980	2.1750	2.2722	2.3712	2.7094

Table 2. Bias, SD and AD of QMLEs and $\widehat{\text{MH}}$ value, $\theta_0 = (0.1, 0.3, 0.5)'$.

$\theta_0 = (0.1, 0.3, 0.5)'$			$\lvert y_t\rvert$	$RV1_t$	$RV5_t$	$RV10_t$	$RV15_t$	$RV30_t$
$T = 500$	$\hat{\omega}$	Bias	0.0072	−0.0011	−0.0008	−0.0007	−0.0006	0.0001
		SD	0.0375	0.0103	0.0110	0.0119	0.0128	0.0156
		AD	0.0436	0.0097	0.0112	0.0112	0.0135	0.0144
	$\hat{\alpha}$	Bias	0.0008	0.0005	0.0009	0.0010	0.0007	−0.0001
		SD	0.0647	0.0217	0.0227	0.0246	0.0259	0.0305
		AD	0.0681	0.0198	0.0220	0.0230	0.0263	0.0292
	$\hat{\beta}$	Bias	−0.0202	0.0008	−0.0001	−0.0006	−0.0004	−0.0019
		SD	0.1237	0.0334	0.0357	0.0390	0.0424	0.0529
		AD	0.1394	0.0324	0.0372	0.0378	0.0448	0.0487
	$\widehat{\text{MH}}$		5.2476	1.9006	1.9645	2.0490	2.1307	2.4028
$T = 1000$	$\hat{\omega}$	Bias	0.0037	0.0000	0.0001	0.0003	0.0005	0.0007
		SD	0.0243	0.0074	0.0080	0.0086	0.0094	0.0115
		AD	0.0260	0.0064	0.0082	0.0091	0.0096	0.0131
	$\hat{\alpha}$	Bias	0.0013	0.0004	0.0006	0.0006	0.0010	0.0009
		SD	0.0459	0.0157	0.0167	0.0181	0.0190	0.0223
		AD	0.0431	0.0152	0.0169	0.0188	0.0201	0.0275
	$\hat{\beta}$	Bias	−0.0114	−0.0014	−0.0017	−0.0023	−0.0030	−0.0034
		SD	0.0816	0.0239	0.0263	0.0287	0.0312	0.0388
		AD	0.0830	0.0251	0.0278	0.0310	0.0326	0.0441
	$\widehat{\text{MH}}$		5.4240	1.8962	1.9615	2.0488	2.1387	2.4244
$T = 1500$	$\hat{\omega}$	Bias	0.0027	−0.0000	0.0000	0.0001	0.0001	0.0003
		SD	0.0199	0.0061	0.0065	0.0071	0.0078	0.0095
		AD	0.0146	0.0055	0.0057	0.0061	0.0072	0.0082
	$\hat{\alpha}$	Bias	0.0014	0.0002	0.0005	0.0004	0.0007	0.0001
		SD	0.0377	0.0126	0.0135	0.0143	0.0155	0.0186
		AD	0.0353	0.0113	0.0122	0.0132	0.0149	0.0174
	$\hat{\beta}$	Bias	−0.0077	−0.0002	−0.0005	−0.0007	−0.0008	−0.0009
		SD	0.0675	0.0195	0.0211	0.0232	0.0256	0.0316
		AD	0.0557	0.0201	0.0210	0.0223	0.0261	0.0302
	$\widehat{\text{MH}}$		5.4975	1.9082	1.9754	2.0684	2.1681	2.4631

5. Empirical studies

In this section, model (3)–(4) is applied to study a real data set. We analyze SSE 50 Index (Shanghai Stock Exchange 50 Index) return series over the period April 28, 2008–July 24, 2013. The data set consists of the closing prices of the SSE 50 Index at one minute intervals, with a total of 1273 trading days and 240 observations per day. Denote the price sequence as $\{P_t(u), t \in [0, 1273], u \in [0, 1]\}$. Then the intraday return at u time on day t can be calculated as

$$Y_t(u) = [\log P_t(u) - \log P_{t-1}(1)] \times 100. \tag{37}$$

For demonstration, we plot $Y_t(1)$ in Figure 1.

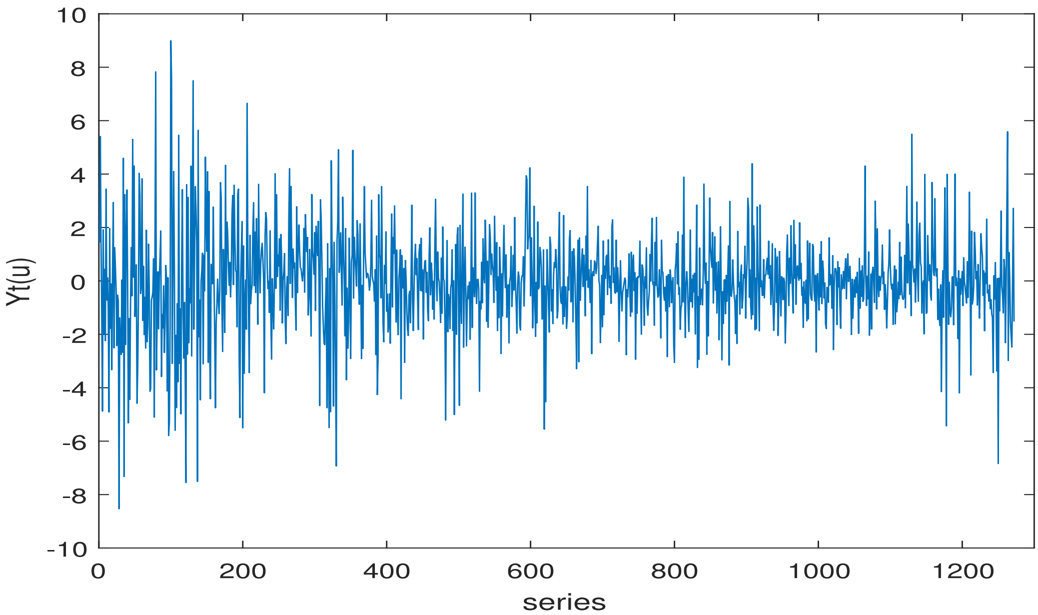

Figure 1. Time series plot of $\{Y_t(1)\}_{t=1}^{1273}$.

To estimate the model, we choose realized volatility as the volatility proxy based on different intraday sampling frequency: 1-minute, 5-minute, 10-minute, 15-minute and 30-minute. Accordingly, the realized volatility of different frequencies are recorded as $RV1$, $RV5$, $RV10$, $RV15$ and $RV30$. Similar to the simulation part, for comparison, we also consider the volatility proxy $H_t = |y_t|$ which corresponds to the estimator without using high frequency data. Figure 2 shows the time series plots of different volatility proxies.

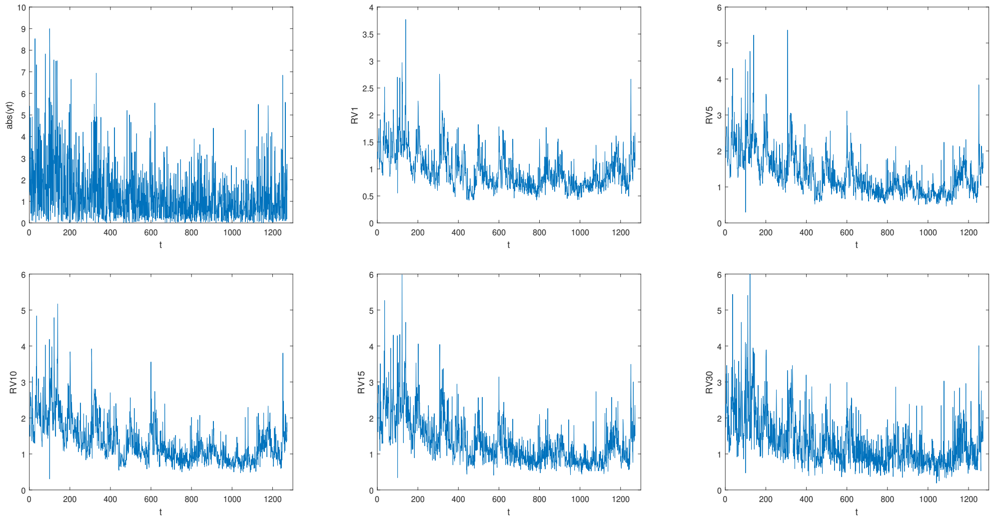

Figure 2. The time series of different volatility proxies

The LGARCH (1,1) model estimation results among different H_t are given in Table 3. The AD in the table means the asymptotic variance of the estimated parameter which is calculated according to the formula (17) and (30).

Table 3. Parameter estimation of LGARCH (1,1) model based on different volatility proxies.

H_t	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	AD($\hat{\omega}$)	AD($\hat{\alpha}$)	AD($\hat{\beta}$)	$\widehat{\mathrm{MH}}$	$\widehat{Var}(\varepsilon_t^{*2})$
$\lvert y_t \rvert$	0.0220	0.0501	0.9499	0.0104	0.0144	0.0144	5.5530	3.5947
$RV1_t$	0.0836	0.1102	0.8685	0.0094	0.0083	0.0104	1.8320	0.3154
$RV5_t$	0.0239	0.1175	0.8996	0.0061	0.0099	0.0088	2.2553	0.5578
$RV10_t$	0.0311	0.1137	0.8980	0.0061	0.0091	0.0085	2.2245	0.4724
$RV15_t$	0.0217	0.1154	0.9024	0.0058	0.0097	0.0085	2.4282	0.5581
$RV30_t$	0.0326	0.1116	0.8986	0.0082	0.0121	0.0113	2.6533	0.8288

It can be observed from Table 3 that the $\widehat{\mathrm{MH}}$ values of realized volatility are generally much smaller than that of $|y_t|$, where $RV1_t$ is the smallest and the corresponding $\widehat{Var}(\varepsilon_t^{*2})$ is also the smallest, which is consistent with the previous theoretical results. Compared with the estimation results of $RV1_t$ and $|y_t|$, the asymptotic variance of $\hat{\omega}$, $\hat{\alpha}$ and $\hat{\beta}$ corresponding to $RV1_t$ are obviously smaller.

From Table 3, when $H_t = |y_t|$, the estimated LGARCH (1,1) model is

$$\begin{aligned} y_t &= h_t \varepsilon_t, \\ h_t &= 0.0220 + 0.0501|y_{t-1}| + 0.9499 h_{t-1}. \end{aligned} \tag{38}$$

When $H_t = RV1_t$, the fitting model is

$$\begin{aligned} y_t &= h_t \varepsilon_t, \\ h_t &= 0.0836 + 0.1102|y_{t-1}| + 0.8685 h_{t-1}. \end{aligned} \tag{39}$$

To further compare the estimation effect between $|y_t|$ and $RV1_t$, the 95% confidence intervals of parameter estimators are calculated based on the AD in Table 3. Let $\hat{\theta}_L = (\hat{\omega}_L, \hat{\alpha}_L, \hat{\beta}_L)'$ and $\hat{\theta}_U = (\hat{\omega}_U, \hat{\alpha}_U, \hat{\beta}_U)'$ be the lower bound and the upper bound respectively. Then we can calculate the upper and lower bounds of h_t as follows:

$$h_{Lt} = \hat{\omega}_L + \hat{\alpha}_L|y_{t-1}| + \hat{\beta}_L h_{L,t-1}, \quad h_{Ut} = \hat{\omega}_U + \hat{\alpha}_U|y_{t-1}| + \hat{\beta}_U h_{U,t-1}. \tag{40}$$

For comparison, we plot the computed h_t, h_{Lt} and h_{Ut} based on equations (38) and (39) in Figure 3. It can be seen from Figure 3 that the estimated h_t from two models are basically close. However, the interval $[h_{Lt}, h_{Ut}]$ (circle) of model (39) is significantly narrower than that of model (38) (triangle).

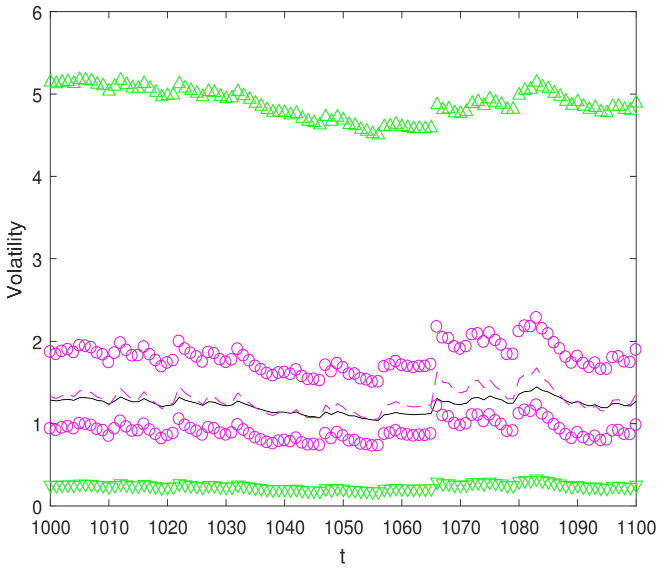

Figure 3. Time series plots of h_t, h_{Lt} and h_{Ut}: for model (38), h_t (real line), h_{Lt} and h_{Ut} (triangle) ; for model (39), h_t (dashed line), h_{Lt} and h_{Ut} (circle).

It is of some interest to compare the performance between $RV1_t$, $RV5_t$, $RV10_t$, $RV15_t$ and $RV30_t$. Due to limitations of space, we only show the comparisons between $RV1_t$ and $RV30_t$. We draw the 95% confidence intervals of h_t calculated by the $RV1_t$ and $RV30_t$ in Figure 4. We can find that the confidence interval under $RV1_t$ is narrower than that under $RV30_t$, which means that the model estimation effect is better with $RV1_t$.

To sum, from the estimation results of Table 3 and the plots in Figure 3, it is shown that introducing the intraday high frequency data can help to improve the LGARCH model estimation for the considered data. Hence the proposed approach is of certain practical value.

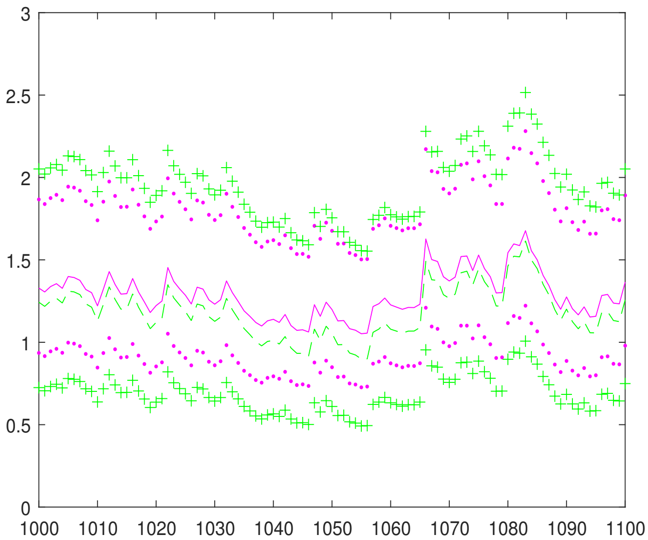

Figure 4. Time series plots of h_t, h_{Lt} and h_{Ut}: for model under $RV1_t$(real line, purple,·) and $RV30_t$(dashed line, green,+).

6. Conclusions

With the motivation that to obtain more precise estimators for the well-known daily LGARCH model, this paper studies how to introduce intraday high frequency data into the model estimation. Based on the existing volatility proxy idea, the estimation methods of all parameters of the model and the corresponding asymptotic properties are given. The selection criteria of the optimal volatility proxy is also proposed. The simulation results show that the parameter estimators perform well under finite samples. The empirical study based on SSE 50 index shows that using the high frequency data can significantly improve the estimation accuracy of daily frequency LGARCH model, compared to the usual method which only uses the daily data.

Acknowledgments

The work is partially supported by National Natural Science Foundation of China 11731015,11701116, Natural Science Foundation of Guangdong Province 192 2018A030310068, Innovative Team Project of Ordinary Universities in Guangdong Province 2020WCXTD018, Science and Technology Program of GuangZhou 202102020368 and Guangzhou University Research Funding YG2020029,YH202108.

Conflict of interest

The authors declare no conflict of interest.

References

Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. *J Econ* 31: 307–328.

De Davide D (2019) Forecasting volatility using combination across estimation windows: An application to S&P500 stock market index. *Math Biosci Eng* 16: 7195–7216.

Deng CL, Zhang XF, Li Y, et al. (2020) Garch Model Test Using High-Frequency Data. *Mathematics* 8: 1922–1939.

Engle RF (1982) Autoregressive conditional heteroscdeasticity with estimates of the variance of united kingdom inflation. *Econometrica* 50: 987–1007.

Fan P, Lan Y, Chen M, et al. (2017) The estimating method of VaR based on PGARCH model with high-frequency data. *Syst Eng Theory Pract* 37: 2052–2059.

Hentschel L (1995) All in the family Nesting symmetric and asymmetric GARCH models. *J Financ Econ* 39: 71–104.

Linton O, Wu J (2020) A coupled component DCS-EGARCH model for intraday and overnight volatility. *J Econ* 217: 176–201.

Liang X, Zhang XF, Li Y, et al. (2021) Daily nonparametric ARCH(1) model estimation using intraday high frequency data. *AIMS Math* 6: 3455–3464.

Nelson DB (1991) Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica* 59: 347–370.

Duffie D, Pan J (1997) An Overview of Value at Risk. *J Deriv* 4: 7–49.

Pan J, Wang H, Tong H (2008) Estimation and tests for power-transformed and threshold GARCH models. *J Econ* 142: 352–378.

Gyamerah SA (2019) Modelling the volatility of Bitcoin returns using GARCH models. *Quant Financ Econ* 3: 739–753.

Visser MP (2011) GARCH parameter estimation using high-frequency data. *J Financ Econom* 9: 162–197.

Wang M, Chen Z, Wang CD (2018) Composite quantile regression for GARCH models using high-frequency data. *Econ Stat* 7: 115–133.

Wu S, Feng M, Zhang H, et al. (2018) Quasi-maximum exponential likelihood estimation for non-stationary GARCH (1,1) models with high-frequency data. *Science in China* 48: 443–456.

Xiao Z, Koenker R (2009) Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models. *J Am Stat Assoc* 104: 1696–1712.

Zou Y, Yu L, He K (2015). Estimating Portfolio Value at Risk in the Electricity Markets Using an Entropy Optimized BEMD Approach. *Entropy* 17: 4519–4532.

Zhao B, Chen Z, Tao GP, et al. (2016) Composite quantile regression estimation for P-GARCH processes. *Sci China Math* 59: 977–998.