The empirical research of ARMA-GARCH models based on high frequency data

Youning Guo¹, Tingting Li¹*

¹School of Sciences, Guilin University of Technology, Guilin Guangxi 541004, China
*Corresponding author’s e-mail: wsltt0621@163.com

Abstract. Based on high frequency data (5 minutes) and low frequency data (day), this paper models and analyses the Shanghai Stock Exchange Index (SH1A0001). Through the stationarity test and the ARCH effect test, the ARMA (1,1) – GARCH (1,1) model is established. Under the assumption of the error term obeys the Laplace distribution, the estimation is obtained. By the result of forward prediction, it is found that the result of model prediction in high frequency data is better than those in low frequency data.

1. Introduction

With the rapid development of electronic information technology, financial market transactions become more and more procedural, simplified and high-frequency. According to statistics, more than 65% of the current financial market is completed by high-frequency exchanges. Traditional financial data analysis is based on low frequency data, which are "day" as a unit, such as the opening price, closing price, the highest price, the lowest price and so on. The conclusion can only be used to predict the return of a day, but it is impossible to predict the trend of the stock market at a certain time. Since stocks have a price in every second, in order to facilitate the selection of data, this paper selects stock prices every five minutes as high-frequency data for analysis and modeling.

In 1970, Box and Jenkins [1] proposed an ARMA model in a book. And its modeling method can describe the dynamic structure of data fully and it can solve the problem of estimating many parameters in higher-order AR and MA models. It indicates that time series analysis has entered the era of parametric model. Engle (1982) [2] introduced the autoregressive conditional heteroscedasticity model, namely ARCH model, which is the first model to provide a systematic framework for volatility modeling. Bollerslev (1986) [3] extended the ARCH model to solve the multi-parameter problem of higher-order ARCH model, and established the generalized autoregressive conditional heteroscedasticity model, namely GARCH model.

Because ARMA-GARCH model can describe conditional mean and volatility clustering very well, more and more people study this model, making it a common model for analyzing financial data, see Nakat Suma (2000) [4], Verhoeven (2000) [5], Solbakke(2001) [6].

Ling and Li (1997) [7], Francq and Zakoian (2004) [8] established the asymptotic theory of quasi maximum likelihood estimation (QMLE) under some conditions for ARMA-GARCH model. Ling (2007) [9] proposed the progressive theory of global self-weighted QMLE and local QMLE for the ARMA-GARCH model. Zhu and Ling (2011) [10] discussed the asymptotic theory of global self-weighted quasi-maximum exponential likelihood estimation (QMELE) and local QMELE for ARMA-GARCH model. Visser (2011) [11] used high frequency data to estimate the parameters of GARCH
model under normal distribution. Huang and Chen (2014) [12] studied the quasi-maximum exponential likelihood estimation and its asymptotic properties of GARCH model with high frequency data.

ARMA-GARCH model has long been proved to have significant advantages in analyzing financial data. Therefore, the study of ARMA-GARCH model is of great significance. Nowadays, more and more investors prefer to make investment decisions by referring to daily high-frequency trading data. The analysis of ARMA-GARCH model based on high-frequency data can predict the fluctuation of financial data more accurately and flexibly, provide guidance for future financial transactions, and improve the management and control ability of future risks.

2. Model description

2.1. ARMA-GARCH model

ARMA model is composed of autoregressive (AR) model and moving average (MA) model. The form of ARMA (p, q) is as follows:

$$y_t = \mu + \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{j=1}^{q} \psi_j \varepsilon_{t-j} + \varepsilon_t$$  \hspace{1cm} (1)

The GARCH (r, s) model proposed by Bollerslev in 1986 is:

$$\varepsilon_t = \eta_t \sigma_t, \quad \sigma_t^2 = \omega_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2$$  \hspace{1cm} (2)

Where $y_t$ is the logarithmic return sequence, $\phi_{t-j}$ represents the information set at $t-1$ time, $\varepsilon_t$ is the innovation sequence with the mean value of 0, $\beta_j \geq 0 (j = 1, \cdots, p)$, $\alpha_i \geq 0 (i = 1, \cdots, q)$, $\omega_0 > 0$ are constant. $\eta_t$ is a sequence of i.i.d. random variables with $E\eta_t = 0$. As we all know, the necessary and sufficient conditions for the model to have a unique strictly stationary solution is as follows:

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

2.2. Parameter estimation

When using the method of maximum likelihood estimation to estimate the parameters of time series, one premise is that the distribution density function must be known, otherwise its likelihood function cannot be obtained. When calculating the likelihood function, sometimes it is not easy to get or the result is not very satisfactory. It can be replaced by other similar distributions, and then maximize the substituted likelihood function. To obtain the estimate, it is called quasi-maximum likelihood estimation.

For the ARMA-GARCH model, the error term is assumed to be the normal distribution, and the estimation is called the quasi-maximum likelihood estimation (QMLE-L). Both QMLE-N and QMLE-L have consistency and asymptotic normality, and QMLE-L has better fitting effect than other distributions (normal distribution, t distribution, GED distribution, etc.), see references [8-10].

If $\hat{\theta}$ is the QMLE-L, then $\hat{\theta} = \arg \min_{\theta \in \Theta} L_m(\theta)$, where

$$L_m(\theta) = \frac{1}{n} \sum_{t=1}^{n} l_i(\theta), \quad l_i(\theta) = \log \sqrt{h_i(\theta)} + \frac{|\hat{\gamma}(\theta)|}{\sqrt{h_i(\theta)}}$$
3. Empirical Analysis

3.1. Descriptive analysis of data
This paper chooses the closing price every five minutes of the Shanghai Stock Exchange Index (SH1A0001) from January 4, 2012 to December 31, 2012, with 11664 high-frequency data as the research sample. With the help of Eviews software, its properties are analyzed and the model is constructed.

Firstly, the figure 1 is a time chart of 11664 high frequency data. The logarithmic return rate is the first order difference of return, \( r_t = \ln P_t - \ln P_{t-1} \), where \( r_t \) is the logarithmic return rate, \( P_t \) is the closing price at \( t \) moment. Figure 2 shows the logarithmic return rate of the Shanghai Stock Exchange Index. From the graph, we can see that its fluctuation is characterized by aggregation.

As can be seen from figure 3, the mean value is close to 0; the skewness is slightly larger and 0, showing a right skewness; the kurtosis is much larger than 3, indicating that the sequence has a very obvious "peak and thick tail" feature; the J-Bera statistic is greater than the critical value 5.99 of significance level 0.05, so the logarithmic return series refuses to obey the assumption of normal distribution.

3.2. Stationarity test
We use the ADF unit root test to test the stability of the sequence. The hypothesis is as follows:

\[
\begin{align*}
H_0 &: \rho = 0 \\
H_1 &: \rho \neq 0
\end{align*}
\]
If $H_0$ is true, it is considered that the sequence has unit root, which is a non-stationary sequence. The $t$ test statistic is $t_{\rho} = \hat{\rho} / S(\hat{\rho})$, where $\hat{\rho}$ is the estimated value of $\rho$, $S(\hat{\rho})$ is the standard deviation of the estimated coefficient $\hat{\rho}$. The test results are shown in Figure 4.

| t-Statistic | Prob.* |
|-------------|--------|
| 16.23409    | 0.0000 |

Figure 4. ADF Unit Root Test.

As can be seen from figure 4, the absolute value of $t$ test statistic is greater than the absolute value of the critical value of the significance levels of 1%, 5% and 10%. Therefore, the original assumption that there is a unit root is rejected at the significance level of 1%. Therefore, the logarithmic return rate series is stationary.

### 3.3 Correlation test

As can be seen from figure 5, the values of AC function and PAC function.

From the results shown in figure 5, we can see that both autocorrelation function and partial correlation function of the return series have "tailing" phenomenon, so we consider the establishment of ARMA model. Try to build ARMA (2,2), ARMA (2,1), ARMA (1,2), ARMA (1,1), because the smaller the AIC and SC values, the better, we choose AMRA (1,1) finally.

### 3.4 ARCH effect test

As can be seen from figure 6, the probability of the statistic tends to zero, indicating that the original assumption that there is no heteroscedasticity is rejected. That is to say, the sequence has obvious ARCH effect, so conditional heteroscedasticity equation should be established. GARCH (1,1) is the most concise and widely used form, so we establish ARMA (1,1) - GARCH (1,1) to fit the logarithmic return rate series.
3.5. Estimation and comparison of models
The ARMA (1,1) - GARCH (1,1) model is obtained by solving with MATLAB as follows:
\[ y_t = 0.0000234 - 0.47 y_{t-1} + 0.52 \varepsilon_{t-1} + \varepsilon_t \]
\[ \sigma_t^2 = 0.0016 + 0.15 \varepsilon_{t-1}^2 + 0.6 \sigma_{t-1}^2 \]

From the estimated results \(0.15 + 0.6 < 1\), we can see that the model satisfies the constraints of stationarity, and shows that the volatility persistence of the Shanghai Stock Exchange Rate Index is not very long in this period, and there is a certain degree of "volatility clustering phenomenon".

Then select the daily closing price of the stock in 2012, a total of 243 data as low-frequency samples. ARMA (1,1) - GARCH (1,1) model is also applicable to low frequency data. Using the same method, the following models are obtained:
\[ y_t = 0.0000304 - 0.18 y_{t-1} + 0.16 \varepsilon_{t-1} + \varepsilon_t \]
\[ \sigma_t^2 = 0.001 + 0.2542 \varepsilon_{t-1}^2 + 0.7069 \sigma_{t-1}^2 \]

Using the results obtained, we predicted forward for 3 days, and compared with the real value, the average absolute error is as follows:

| Table 1. Comparative analysis of predicted result. |
|--------------------------------------------------|
| high frequency data                             |
| low frequency data                              |
| The average absolute error                      |
| 0.0017                                          |
| 0.0133                                          |

From table 1, we can get that the average absolute error of prediction based on high-frequency data is smaller than that based on low-frequency data, which shows that the estimation value is closer to the true value and the fitting effect is better when using high-frequency data to predict.

4. Conclusion
In this paper, we use Eviews to analyze the high frequency data (every 5 minutes) of Shanghai Stock Exchange Index. Through the stationarity test and the ARCH effect test, the ARMA (1,1) – GARCH (1,1) model is established. Under the assumption of the error term obeys the Laplace distribution, the estimation is obtained. At last, we use matlab to compare the fitting effect base on high-frequency data and low frequency data.

Therefore, we suggest that investors use as much high-frequency data as possible to predict when making investment decisions, and the results will also be relatively accurate and reliable.

Acknowledgments
This work was supported by Basic Competence Promotion Project for Young and Middle-aged Teachers in Guangxi, China (2018KY0262, 2019KY0269).

References
[1] Box G.E.P, Jenkins G.M. (2003) Time Series Analysis: Forecasting and Control. China Statistical Publishing House, Beijing.
[2] Engle R F. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. Econometrica, 50: 987-1008.
[3] Bollerslev T. (1987) A conditional heteroskedastic time series model for speculative prices and rates of return. The review of economics and statistics, 69: 542-547.
[4] Nakatsuma T. (2000) Bayesian analysis of ARMA-GARCH models: A Markov chain sampling approach. Journal of Econometrics, 95: 57-69.
[5] Verhoeven P. (2000) Modelling outliers and extreme observations for ARMA-GARCH processes. Econometric Society World Congress 2000 Contributed Papers, No:1922.
[6] Solibakke P.B. (2001) Efficiently ARMA-GARCH estimated trading volume characteristics in thinly traded markets. Financial Economics, 11: 539-556.

[7] Ling S, Li W.K. (1997) On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. Journal of American Statistical Association, 92: 1184-1194.

[8] Francq C, Zakoian J.M. (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. Bernoulli, 10: 605-637.

[9] Ling S. (2007) Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. Journal of Econometrics, 140: 849-873.

[10] Zhu K, Ling S. (2011) Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA-GARCH/IGARCH models. The Annals of Statistics, 39: 2131-2163.

[11] Visser M.P. (2011) GARCH Parameter Estimation Using High-Frequency Data[J]. Journal of Financial Econometrics, 9: 162-197.

[12] Huang G.J., Chen M. (2014) Quasi Maximum Exponential Likelihood Estimation of GARCH Model Based on High Frequency Data. Journal of Applied Mathematics, 37: 1005–1017.