Prediction and Analysis of Financial Volatility Based on Implied Volatility and GARCH Model

Feng Lin
Wenzhou Business Department of CITIC Futures Co. LTD, Wenzhou, 325000, Zhejiang, China
DOI: 10.32629/memf.v3i1.650

Abstract: Based on the data information of Shanghai and Shenzhen CSI 300 stock index futures, the performance of the GARCH model and implied volatility prediction in different time periods were studied. This paper mainly discusses the volatility of index returns and uses Matlab and Minitab to measure the performance of the GARCH model and implied volatility model in volatility prediction, and then comments on the prediction results. The results show that the GARCH model has a good prediction effect in the short term, while the implied volatility has a good prediction power in the long term. Option prices can mirror market information in a more comprehensive way. As a result, implied volatility is more reasonable to predict future volatility.
***Keywords*:** Shanghai and Shenzhen options, GARCH model, implied volatility, conditional variance

1. Introduction

Volatility is a vital variable in financial research, which is indispensable for the investment portfolio, asset pricing, risk control and fiscal policy formulation. Volatility modelling is a primary daily task of the financial system and has attracted thousands of experts, academics and industry professionals over the last 20 years. They tried multiple methods to predict the fluctuations. However, with different models, the outcomes can be very diverse. These seemingly complex and changeable models can be divided into two categories. One is to predict future volatility by using historical information, commonly known as the "historical information method". For example, the ARCH family, the most common model family in the financial industry, and the stochastic volatility model (SV model) is gradually famous in recent years. Another type of model is the reverse parameter of option prices quoted in the market, also known as the implied volatility. In previous models, historical information attempts to predict future volatility by detecting trends in the past volatility samples.

This method has the following defects. Firstly, the result concluded from the sample is pseudo canonical, or there is the problem of over-fitting. Secondly, this method assumed that historical time must be repeated. That means the pattern found from historical samples must be applied in the future. This approach does not consider new information, such as changes in market conditions. In addition, as the information center of all participants, the prices in the financial system generated every day are the prediction made by the supply and demand parties. They trade with each other according to the historical information and the latest trends after obtaining a large amount of news, containing a large number of innovative predictions, which are constantly updated and adjusted dynamically. Therefore, the implied volatility method has its incomparable advantages. The premise of this method is that market participants are relatively objective, and prices can reflect investors' rational expectations for the future. Otherwise, there will be all kinds of noise in the stock index, and the implied volatility information cannot be obtained from the option price. Therefore, future volatility cannot be accurately predicted. Above all, the advantages and disadvantages of the two prediction methods become the research object of this paper.

Since stock index options in China are listed relatively late with small trading volume and short development period, it tends to lose the value of their fundamental theories and cannot be used to carry out scientific research on implied volatility. Therefore, the scientific research on volatility prediction and analysis is concentrated on the first category, namely time series analysis. In fact, the comparison of these two approaches has been a critical concern in overseas scientific research. With the gradual development of implied volatility research, the debate of whether its predictive analysis ability is better than the historical volatility has not stopped.

Since Engle's ARCH model (1982) and Bollerslev's GARCH model (1986) were proposed, the comparison between implied volatility and GARCH volatility has been in progress. Compared with the volatility of the time series analysis model, detecting the predictive analysis ability of implied volatility involves a complicated problem. Detection of the ability of implied volatility forecast analysis is collaborative since it tests whether an option market is reasonable and whether the option pricing model is correct. So when the implied volatility forecast analysis ability can not cover the historical volatility information(that is, the implied volatility does not contain all of the historical information), excluding sales market failure and traders' expectation, there is another possibility. The selling market is reasonable (the discrimination is accurate), but the option pricing formula is incorrect because many of the preconditions created by the index futures formula are assumed not to be established in reality. Therefore, the analysis of the information contained in implied volatility is very different in the long run. From multiple perspectives, academics choose various ways and tools to implement the gaps in the initial scientific research of the implied volatility to reduce the errors in the prediction and analysis, and improve the information quality. So far, the theoretical basis for volatility convergence is that the implied volatility contains effective information for the future and perhaps more information components of the total volatility. However, the opinions are still very different whether it includes historical volatility information and consists of the volatility information of time series analysis.

Early studies in this area focused on the stock market, such as Gemmill (1986), Lamoureux, and Lap-Strapes (1993). However, due to the low trading volume of individual stock options, its price contains a lot of noise, so the information contained in the implied volatility based on individual stock is not stable and reliable. Since the 1990s, the research on implied volatility began to focus on index options. All studies except Canina and Figlewski (1993) show that implied volatility contains valuable information for future volatility. However, there are different opinions on whether implied volatility includes all historical information and other time series volatility information. Ederington and Guan (2002, 2005) and Martens and Zein (2004) found that implied volatility did not fully include volatility of ARCH models and volatility prediction based on historical information. Blair et al. (2001) and Szakmary et al. (2003) found that the implied volatility of options performed better than the prediction of time series and completely contained all the information of time series volatility.

In addition, implied volatility behaves differently at historical stages. At the same time, Ederington and Guan (2002) also pointed out that the performance of implied volatility is susceptible to the length of the forecast period. And problems such as measurement errors do not significantly affect the performance of implied volatility. In these studies, the most common predicted interval length was one month. The studies in China only focus on historical information and the comparison of advantages and disadvantages of various models under this method, such as Wei Weixian and Zhou Xiaoming (1999), Zhou Jie and Liu Sanyang (2006), Xu Zhengguo and Zhang Shiying (2004), etc.

This paper compares the performance of the time series model represented by the GARCH family model and implied volatility model on different durations by using 300EF data from Shanghai and Shenzhen stock exchanges and discusses which model is more suitable for volatility prediction. The structure of the paper is as follows: The first part introduces the implied volatility model, the second part is the selection of data and calculation of different volatility, the third part presents empirical results, and the fourth part is the conclusion of this paper.

2. Implied volatility model and GARCH model

2.1 BS model of European option pricing and its implied volatility

In 1973, Black and Scholes published an option pricing formula (BS option pricing model). This formula is used to calculate the theoretical price of a European call option on an underlying asset. The derivation of the Black-Scholes formula requires many assumptions, which can be divided into two categories, namely, the assumption of stock price distribution and the assumption of trading background. The assumptions about stock price distribution include: (1) The logarithmic return rate (continuous compound return rate) of the underlying asset follows a normal distribution and is mutually independent in time. That is, the price of the underlying asset follows the geometric Brownian motion. (2) The volatility of the logarithmic return rate is known and is a constant. (3) The future stock dividend is known and is a fixed amount or a fixed rate of return.

Assumptions about the trading context include: (1) The risk-free rate is known, and it is a constant. (2) There is no transaction cost, and tax is not considered. (3) Investors can short futures or underlying assets locally and can borrow at a risk-free rate. (4) Under the above conditions, Black and Scholes derived the theoretical price of the European call option and put option (BS formula),

$$C\left(S,K,\sigma,r,T,q\right)=Se^{-qT}N\left(d_1\right)-Ke^{-rT}N\left(d_2\right) \tag{1}$$

$$d_1=\frac{\ln\left(\frac{S}{K}\right)+\left(r-q+\frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}} \tag{2}$$

$$d_2 = d_1 - \sigma\sqrt{T} \tag{3}$$

In which C represents the price of a call option at time t, S represents the stock price at time t, K represents the strike price, T represents the expiration date, r represents the risk-free interest rate from time t to time T, R represents the volatility of stock return rate, and N represents a normal distribution. Formula (1) is Black-Scholes European call option pricing formula, BS formula for short.

2.2 GARCH model

Bollerslev proposed the generalized autoregressive conditional heteroscedasticity model (GARCHModel) in 1986, in which one or more hysteresis values are used. The form is as follows.

For time series {a t}, if:

$$\begin{cases} a_t = \sigma_t u_t, u_t \sim i.\text{ i. d. N}(0,1) \\ \sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i a_{t-i}^2 \end{cases} \tag{4}$$

Then the sequence {a t} is an ARCH (q) sequence, and model (4) is an ARCH (q) model. Where σt2 is the variance, q is the order of the ARCH (q) model, and ut obeys the independent standard normal distribution, α0, αi ≥0i = 1, 2..., q. According to the above definition, it can be deduced:

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{5}$$

This is the commonly used GAREH (1, 1) model

3. Data selection and volatility calculation

Since the Huatai 300ETF option has the longest trading time and the largest trading volume, it is selected as the research object. From the history of Huatai 300ETF option trading, the early trading was relatively inactive, with the average daily trading volume not exceeding 5,000 contracts. At the end of 2012, the trading volume increased significantly, and after 2017, the average daily trading volume exceeded 10,000 contracts. Considering that the larger the trading volume, the more information contained. With the requirement of the number of samples for the empirical research, the sample period selected in this paper is from January 510300 to September 30, 2021.

3.1 Data selection and calculation of implied volatility

In order to study the difference between implied volatility and GARCH volatility when the duration of the forecast is different, the paper selects the daily closing price of Huatai options with the remaining maturity, respectively. Since Huatai options contracts always expire on the penultimate trading day of each contract month, there is no overlap in dates. In addition, there were two months of data unavailable, resulting in 455 daily volatilities.

Considering that a transaction with a larger trading volume represents more investors' recognition of the price and represents most traders' expectations of future volatility, this paper selects the option with the largest trading volume among the at-the-money options. The call option is used to calculate the implied volatility, and the at-the-money option is defined as the strike price K meets 01 97S0 < K < 11 03S0 (which means the strike price is larger than ¥1.97 and smaller than ¥11.03) option. The corresponding daily implied volatility is obtained by numerical method according to the BS formula.

3.2 Data selection and prediction of GARCH volatility

Considering that among numerous time series models, GARCH family models are still the most popular and commonly used models in finance. And complex models are not necessarily superior to simple models. This paper takes simple GARCH model volatility as the representative of time series models to compare with implied volatility. Ander Sen, Bollerslev and Lange (1999) believed that the higher the frequency of sample data relative to the prediction period, the higher the accuracy of volatility prediction. Daily data should therefore be better than monthly (weekly) data for predicting volatility over the next month (week). In fact, studies have found this to be true, with daily data significantly better than monthly (weekly) data.

In this paper, the in-sample data are used to build a model to predict the out-of-sample value. In this case, the out-of-sample interval 0 corresponds to the prediction interval of implied volatility. The rolling estimation method is adopted to the

GARCH model to obtain a better forecasting ability. Four hundred fifty-five daily return data before the week (month) are selected to build the GARCH model for prediction group by group, rather than building a model that predicts volatility over all weeks (or months). Since the parameters of the model estimated in different time periods will be different, if only one model is used to predict the weekly (monthly) volatility in the next nine years, the prediction accuracy of GARCH will be significantly reduced. In contrast, the rolling estimation method can improve the prediction ability of the model.

When using daily data to build the GARCH model, the corresponding heteroscedasticity is the variance of daily return. To obtain the variance of the following week (month), this model predicts the variance of each day in the next week (month) and scale them up. At this time, a dynamic prediction method is needed for the prediction.

3.3 Analysis of basic statistical characteristics of yield series

Minitab software was used to calculate the statistical characteristics of the return rate. As shown in Figure 2, it is the normality test through the quantile-quantile plot. The corresponding probability value of the probability statistic of the normal distribution of the return rate series is smaller than 0.05, indicating that the return series obeys the normal distribution. However, it is still necessary to test the stability of the yield series before establishing the model.

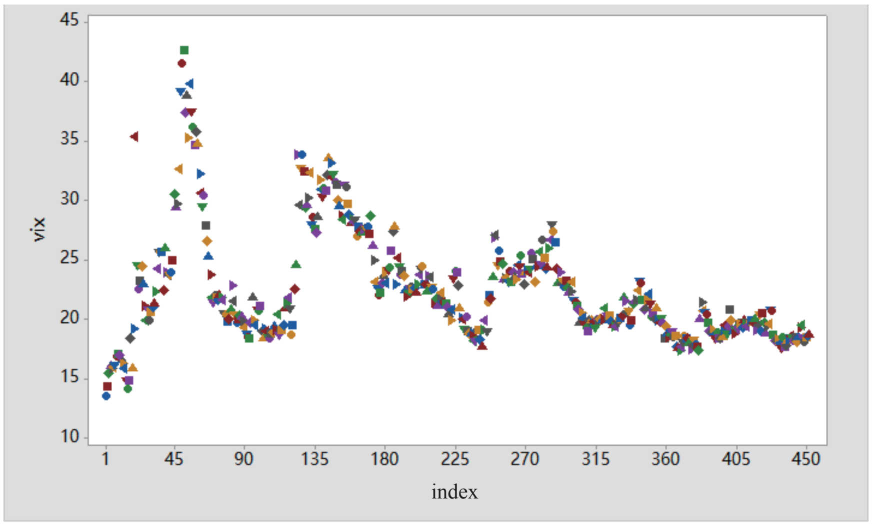

Figure 1. Yield sequence figure

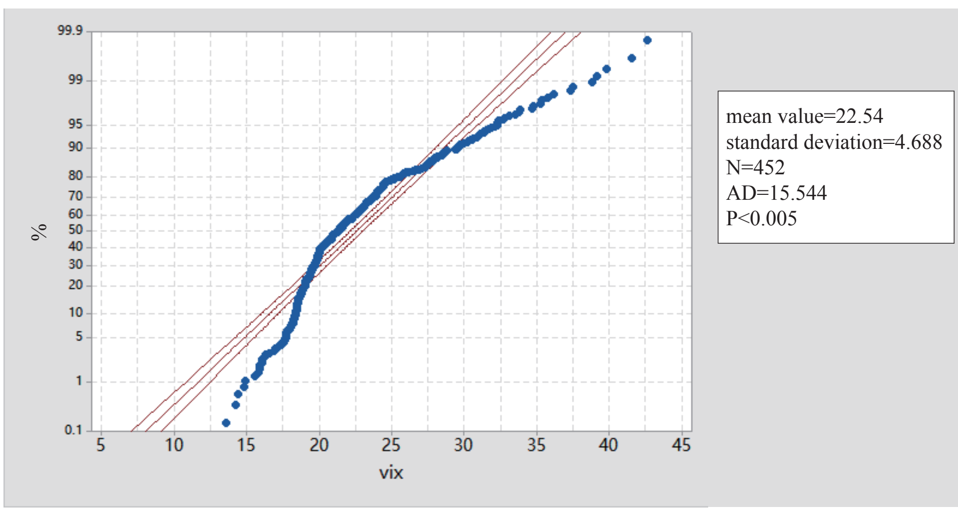

Figure 2. Normal distribution of yield

3.4 Stationarity test

Statistically, autocorrelation is defined as the Pearson correlation between values at different times in the same random process. The following results are obtained using the autocorrelation function and partial correlation function of the CSI

300ETF yield test. The autocorrelation function exceeds two standard deviations, and the partial autocorrelation tail exceeds the specified red line. If the autocorrelation function of the time series falls within the confidence interval when k>3, and gradually tends to zero, the time series is stationary. If the autocorrelation function of the time series falls more outside the confidence interval, the time series is not stationary. Therefore, when the rate of return is differentiated, it is stable after the first order differential.

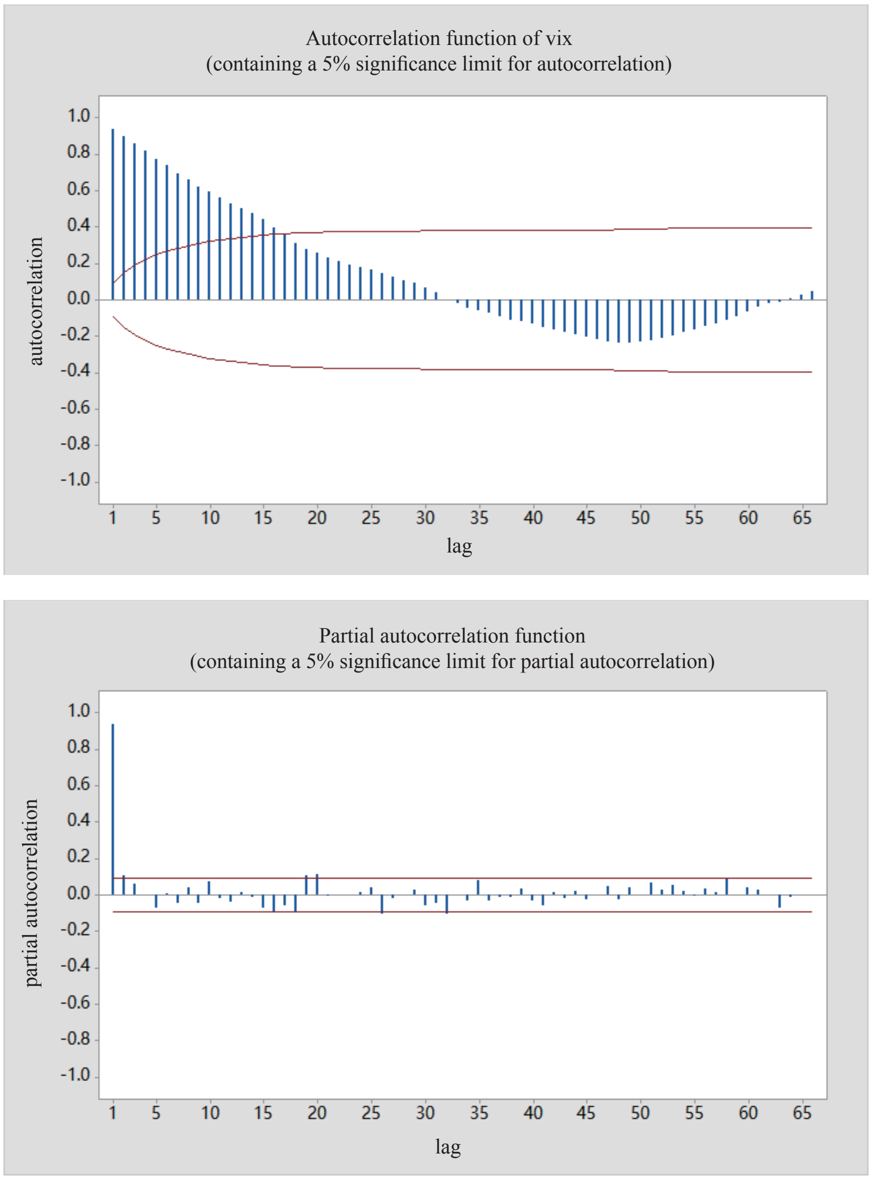

Figure 3. Autocorrelation partial correlation diagram of yield time series

Original hypothesis H0: ARCH effect does not exist in residual sequence, alternative hypothesis H1: ARCH effect exists in residual series. The ARCH LM test can be used to obtain the probability corresponding to the value of F-statistic and the value of T^*R^2, but the difference is not significant. Therefore, the null hypothesis is rejected, and the return series is considered to have an ARCH effect. The GARCH model can be established.

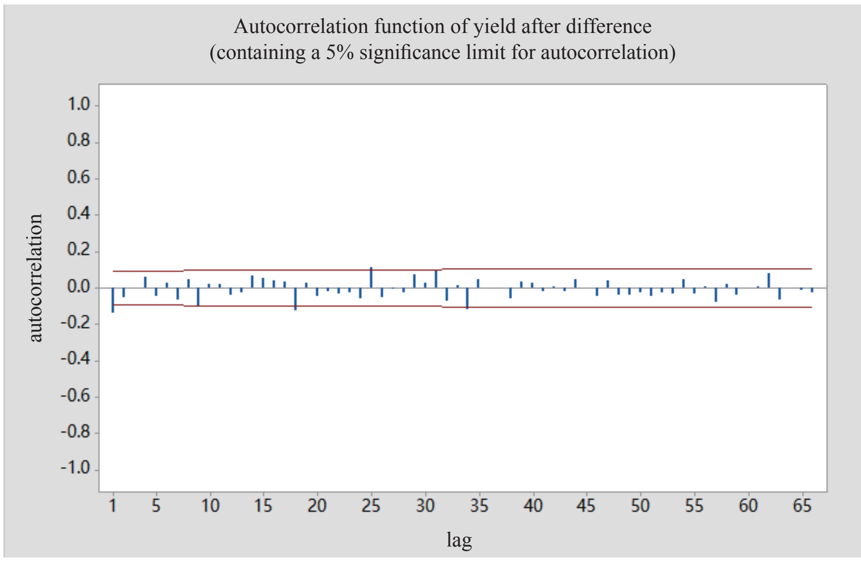

Figure 4. Differential primary diagram of yield

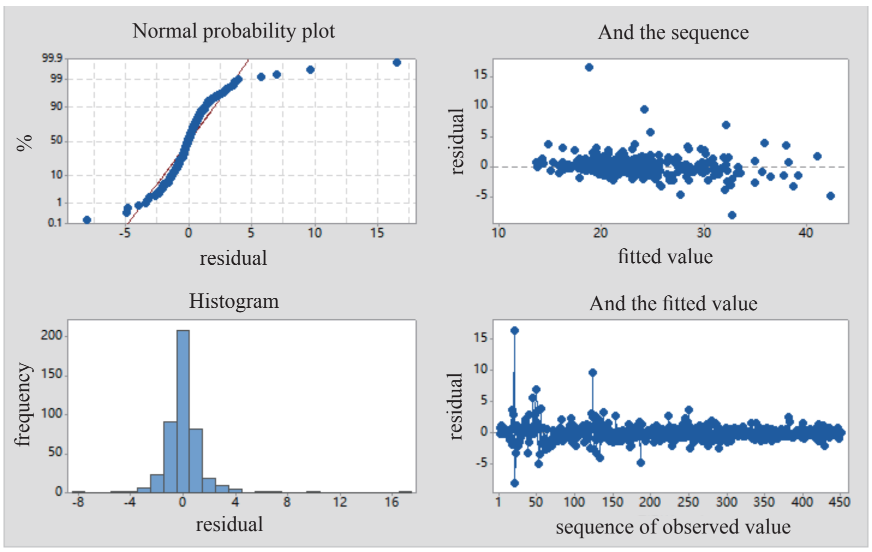

Figure 5. Residuals of yield series

3.5 Establishment of GARCH model

The minimum information criterion AIC and SC criterion are used, and the specific values are shown in Table 1. Then the GARCH (1-1)) model is established as the optimal model.

Table 1. GARCH model grading Table

Grading model		
GARCH	AIC	3.063091
(1-1)	Sc	3.108673
GARCH	Aic	3.065034
(2-1)	Sc	3.119732
GARCH	Aic	3.06492
(1-2)	Sc	3.119618
GARCH	Aic	3.069232
(2-2)	Sc	3.133046

Estimation model: GARCH = C(3) + C(4)*RESID(-1)^2 + C(5)*GARCH(-1).

Table 2. GARCH mean equation

Coefficient	Standard deviation	Error	P value
1.055735	0.189212	5.579629	0
0.94694	0.00883	107.2374	0

Through the coefficient of mean equation and P value in GARCH model, the significant results of this model are credible, and the parameters are significant at the significant level of 0.05.

Table 3. GARCH equation of variance

Conditional variance equation					
Constant c	C	0.206052	0.041261	4.993804	0
Residuals squared	RESID(-1)^2	0.667102	0.071643	9.311454	0
Variance	GARCH(-1)	0.40759	0.063698	6.398775	0

The GARCH model expression is shown below.

$$\begin{aligned} &M_t = 1.055 + 0.946 M_t(-4) + \varepsilon_t, u_t \sim N(0, \sigma_t^2) \\ &\sigma_t^2 = 0.0000 + 0.667 u_{t-1}^2 + 0.404 \sigma_{t-1}^2 \end{aligned} \tag{6}$$

Draw conditional variance volatility graph according to the model:

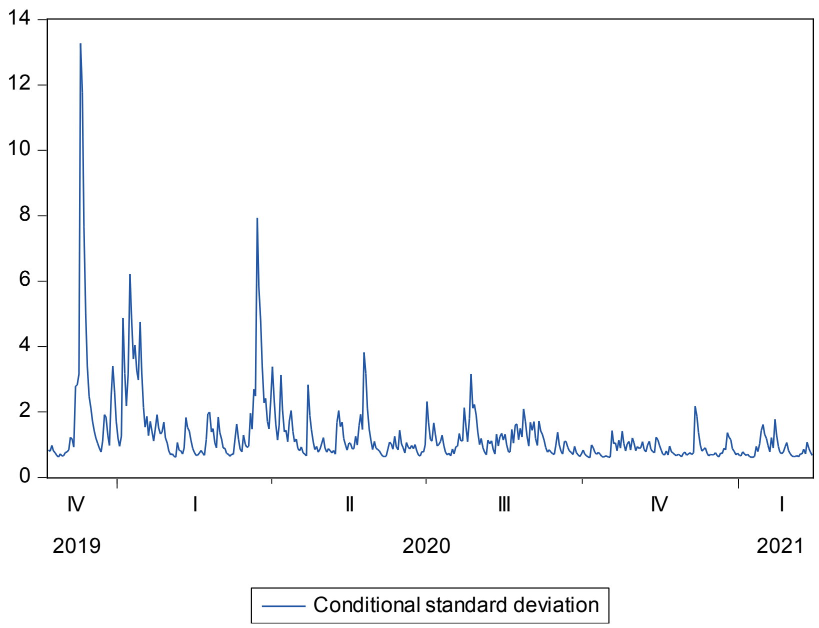

Figure 6. Yield volatility chart

3.6 Comparison between GARCH model and expected results of implied volatility

The GARCH model is compared with the implied volatility prediction, as shown in the following table. Table 4 shows the statistical characteristics of daily implied volatility and GARCH volatility. As can be seen from Table 4, the mean of both GARCH volatility and implied volatility in weekly volatility deviates far from the mean of realized volatility, which means that short-term prediction is more difficult than long-term prediction. Implied volatility and GARCH volatility are close to normal distribution.

Table 4. GARCH model and statistical characteristics of implied volatility

Variable	N	Mean	Standard error of mean	Standard deviation	Variance	Minimum value	Median	Maximum value
Implied volatility	452	0.013	0.0735	1.562	2.4398	-8.1373	-0.0658	16.5797
Conditional variance volatility	452	-0.0357	0.0788	1.6761	2.8093	-14.5045	-0.0675	15.0381

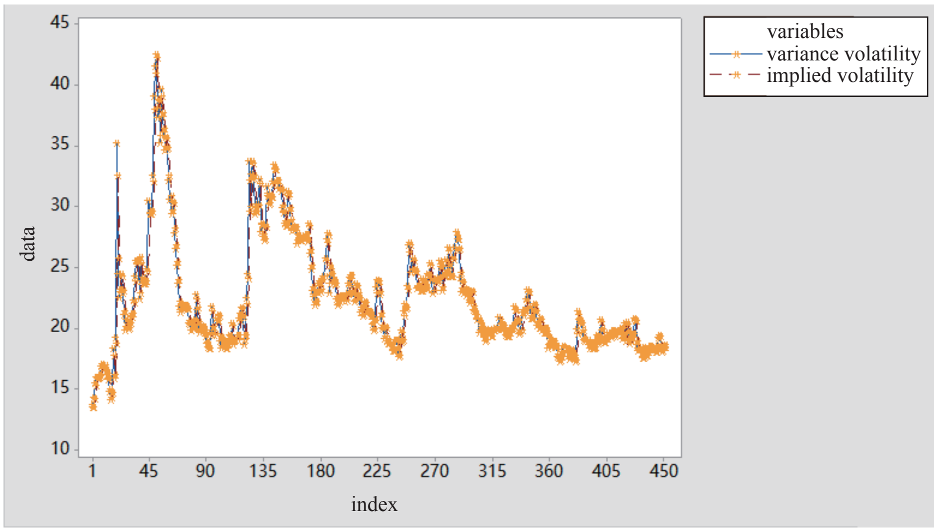

Figure 7. Comparison between GARCH model and implied volatility

The GARCH model performs better than implied volatility in predicting the volatility of the following weeks but worse than implied volatility in predicting the volatility of the following months.

GARCH model is made based on historical time information to distinguish the future, whether history repeats and when, which is itself an uncertain problem. When economic emergencies occur, prediction and analysis based on historical information become unreliable. Differing from GARCH volatility, implied volatility increases traders' judgment of the future based on the current economic and financial situation, instead of believing that the future will definitely repeat the historical time. Many investors have their own estimations for the future stock index futures. And from this price, the BS is used to calculate everyone's prediction for future volatility. Therefore, when the market participants are numerous, rational and mature, the option price generated by buying and selling will be more effective, and the implied volatility obtained from this option price is better than the time series analysis model in predicting future volatility. It can be seen that the more participants, the more transactions, the more accurate everyone would be in judging the future.

4. Conclusion

Implied volatility is better than the GARCH time series model when predicting one-month volatility. When predicting one-week volatility, it is the opposite. On the one hand, the option of index futures which has a one-month remaining period, is more active with larger trading volume but less trading volume for options with one-week life. Therefore, the previous option has more comprehensive volatility information, which is better than the estimation acquired from the GARCH model. On the other hand, it is more difficult for market participants to forecast the market condition for a short horizon. For example, predicting the index volatility for tomorrow is more challenging than predicting the index volatility for next month from option prices. In contrast, the time series model built on the historical volatility information assumes that the volatility would repeat the past pattern in the future. This assumption has a shortcoming that the model would fail when encountering unexpected situations. Thus, the longer horizon for the time series model, the larger probability it would collapse with poor estimation. Above all, the time series model and implied volatility have their own characteristics in predicting and analyzing volatility. And In this paper, the time series model is more suitable for short time horizon prediction, while implied volatility is more fitted for a longer forecasting window.

References

[1] Yang Xiaohui, Wang Yubin.Empirical analysis of volatility and implied volatility based on GARCH model: A case study of 50ETF options[J]. *Financial Theory & Practice*, 2019(05): 80-85.

[2] Qu Manxue, Wang Pengfei. Research on the forecasting ability of volatility index in China -- Information Comparison based on implied volatility[J]. *Economic Issues*, 2017(01): 60-66. DOI: 10.16011/j.cnki.jjwt.2017.01.011.

[3] Luo Hua, Wang Shuang. Forecasting financial volatility using GARCH model and implied volatility[J]. *Journal of Zhejiang Sci-Tech University (Natural Sciences Edition)*, 2016, 35(02): 322-326.

[4] Chen Yanhui. Prediction of implied volatility index based on ARMA-GARCH model and its application in option trading[J]. *Journal of Quantitative Economics*, 2014, 31(04): 27-35.

[5] Zheng Zhenlong, Huang Yizhou.Volatility prediction: GARCH model and the implied volatility [J]. *The Journal of Quantitative & Technical Economics*, 2010, 27(01): 140-150. DOI: 10.13653/j.cnki.jqte.2010.01.009.