

Implied Volatility Prediction Based on Different Term Structures: An Empirical Study of the SSE 50 ETF Options Market from High-Frequency Data.

Wenqi Yang^{1,a}, Jingkun Ma^{2,b}

¹Shandong University, Center for Economic Research, Jinan, China

²Hong Kong University of Science and Technology, School of Science, Hong Kong, China

Abstract. This article focuses on the implied volatility forecast of the SSE 50 ETF options market from June 1, 2017, to August 30, 2019, and constructs AR (1) model and ARMA-GARCH model based on liquidity characteristics to compare and analyze the prediction effect of implied volatility on different option types and term structures. The results show that, during the sample period of the SSE 50 ETF options market, the effect of model fitting of the ARMA-GARCH model is significantly better than the AR (1) model; the fitting sequences predicted by the two models have typical time-varying and synchronization characteristics, and the prediction effect of the ARMA-GARCH model in the whole period is significantly better than the AR (1) model.

1 Introduction

The volatility of financial asset prices reflects the uncertainty of their price changes and plays a very important role in risk supervision, derivative product pricing and investors' portfolio selection. Therefore, predicting and analyzing the volatility of financial asset prices for both financial market participants and managers are of great significance.

Earlier studies on the volatility of financial asset prices were mainly about historical volatility models based on low-frequency data. For example, French et al. ^[1] used the ARIMA (0,1,3) model to study the monthly return rate volatility of the S&P Composite Index, Hui Xiaofeng et al. ^[2] used the GARCH family model to effectively predict the volatility of the RMB exchange rate after the reform of RMB exchange rate. Today, the GARCH models are widely used in various research fields including the volatility of financial asset price.

However, Efimova and Serletis^[3] conducted volatility prediction based on the GARCH family model and found that the volatility of financial asset prices has a periodic time-varying law, and even there is a clustering of volatility, so the use of low-frequency data to research the market volatility may ignore a large amount of in-market information about intraday trading. On this occasion, Andersen and Bollerslev ^[4] first proposed the use of high-frequency daily data to predict market volatility, becoming the first to use high-frequency data to study the volatility of financial asset prices. They provide empirical findings that the realized volatility is a consistent unbiased estimator of integral volatility, which can greatly reduce the impact

of noise in model regression.

In terms of predicting the volatility of asset prices, Black and Scholes^[5] used derivative prices of financial assets as observation variables, and proposed the Black-Scholes option pricing theory including stock dividends, using asset prices to reversely predict volatility. They believe that using the implied volatility model can better reflect the liquidity characteristics of the market, and can keenly describe the impact of sudden major events on the price of financial assets. Since then, Christofersen and Jacobs ^[6] found through empirical research that the predicting effect of the implied volatility model is significantly better than that of the stochastic volatility model.

Some scholars have found that the prediction effect of implied volatility is generally significantly better than historical volatility models based on low-frequency data. For example, Godbey and Mahar ^[7] took the call and put options of 460 stocks in the S&P 500 Index as the research object and found that the implied volatility contains a large amount of information of market volatility and the prediction effect of GARCH model based on implied volatility is significantly better than the single-formed GARCH model.

From the empirical results, it is obvious that there are many differences in the market internal information contained in the implied volatility of different types of options. Moriggia et al. ^[8] found that the implied volatility models based on call and put options can predict significantly different volatility sequences, and they have typical periodic characteristics.

There are more than one hundred option product varieties based on the SSE 50 Index as the subject matter

^abert1204@126.com, ^bmjk550349403@126.com

in China. This provides a wealth of sample data for the study of volatility issues. Therefore, based on the above empirical research, we select the implied volatility of the SSE 50 ETF option yield as the research objects, use the AR (1) model and the ARMA-GARCH model based on liquidity characteristics to predict the volatility of the SSE 50 ETF options, and analyze the differences in the prediction effects of the above models.

Compared with the existing research, the possible marginal contribution of this paper mainly includes the following three points: first, the previous literature rarely certifies the SSE 50 ETF options market as the research object, and we fill the research gap in related academic fields; second, considering the continuous price adjustments and confidence updates in the financial market may cause market volatility to exhibit volatility and short-term characteristics, in this paper, the daily frequency trading data is used to construct relevant indicators in the empirical process, which overcomes the possibility of low-frequency data to measure implied volatility causing the problem of "summary fallacy"; third, in terms of predicting the market's implied volatility indicators, previous studies have mainly focused on a single GARCH model, but the GARCH model is not sensitive to the recognition of changes in market liquidity, to this end, we improve the original GARCH model, using the liquidity-based GARCH model to predict the implied volatility, and solve the problem of insensitivity in characterizing the implied volatility based on changes in market liquidity.

The rest of the paper is organized as follows. In section 2, the indicator design and data description are introduced. In section 3, the models are presented. In section 4, the result of the empirical study is reported. Finally, section 5 concludes the paper.

2 Indicator Design and Data Description

2.1 Implied Volatility Indicator

Black and Scholes^[6] used the Black-Scholes formula for the first time to calculate the implied volatility:

$$C_t = S_t N(d_1) - X_t e^{-r_t T_t} N(d_2) \quad (1)$$

$$P_t = X_t e^{-r_t T_t} N(-d_2) - S_t N(-d_1) \quad (2)$$

$$d_1 = \frac{\ln(S_t / X_t) + \left(r_t - q_t + \frac{\sigma_t^2}{2} \right) T_t}{\sigma_t \sqrt{T_t}} \quad (3)$$

$$d_2 = d_1 - \sigma_t \sqrt{T_t} \quad (4)$$

where X_t and S_t are respectively the strike price and current price at time t , r_t and q_t are respectively the risk-free interest rates and dividends at time t , σ_t and T_t are respectively the implied volatility and time to the maturity date at time t , C_t and P_t are respectively the call option and put option price at time t .

2.2 Description of Data

We select the daily frequency time-series data of SSE 50

ETF options market from June 1, 2017 to August 30, 2019 for the construction of related indicators and the estimation of the model. The original data are all from the Wind database. As of August 30, 2019, after removing the lack of trading data due to stock suspension and the rolling time window required for model calculation, the total sample has 6636 daily frequency data, including the strike price, current price, risk-free interest rate, dividends, time to the maturity date, call option price, put option price and trading volume of the SSE 50 ETF options market. What's more, we directly delete trading days without trading data and corresponding trading day market data.

Table1. test results of ADF unit root

Implied volatility sequence	t-value	5% critical value	conclusion
call options based on the long-term	-3.18	-3.42	instability
lag of 1 period	-17.72*	-3.42	stability
call options based on the short-term	-2.92	-3.42	instability
lag of 1 period	-25.61*	-3.42	stability
put options based on the long-term	-3.67*	-3.42	stability
put options based on the short-term	-3.65*	-3.42	stability

a. The symbols * indicates that the regression coefficients are significant at 5% levels.

Besides, to calculate the success rate of the model prediction, we divide the sample data into two parts: in-sample data and out-of-sample data. The selection range of the in-sample data is 433 daily data from June 1, 2017, to April 9, 2019, and the selection range of out-of-sample data is 100 daily data from April 10, 2019, to August 30, 2019. At the same time, we use the in-sample data to predict the future implied volatility series and use the out-of-sample data to evaluate the prediction results. If on the t day in the rolling regression, the future implied volatility series predicted based on the in-sample data falls in the 95% confidence interval of implied volatility calculated based on the out-of-sample data, indicating the forecast is successful at that day. Based on this method, the success rate of the model in predicting the implied volatility of the entire out-of-sample data interval can be calculated.

To remove the influence of unit roots and prevent the pseudo-regression problem of the model, we firstly use the ADF unit root method to perform unit root tests on the four groups of implied volatility indicators. According to Table 1, the first-order difference of call options based on the long-term, the first-order difference of call options based on the short-term, put options based on the long-term and put options based on the short-term are under 5% significance level which means they all reject the original hypothesis that unit-roots exist, so before the empirical research, we do the first-order lag of call options based on the long-term and call options based on the short-term.

3 Model

The implied volatility of different term structures may have different prediction effects on the future implied volatility series. Godbey and Mahar^[7] believe that the prediction effect based on the ARMA-GARCH model is significantly better than the AR model. In this part, we propose an AR model and an ARMA-GARCH model based on the term structure and liquidity characteristics of different implied volatility. Finally, complete content and organizational editing before formatting.

3.1 AR (1) model

The general form of the AR model based on implied volatility is as follows:

Table2. test results of AIC criterion

call options based on the long-term					
lag order	1	2	3	4	5
average AIC value	-264.0	-263.2	-262.1	-261.2	-259.8
call options based on the short-term					
lag order	1	2	3	4	5
average AIC value	-269.1	-268.2	-267.5	-266.3	-265.1
put options based on the long-term					
lag order	1	2	3	4	5
average AIC value	-267.2	-266.4	-265.4	-264.5	-263.2
put options based on the short-term					
lag order	1	2	3	4	5
average AIC value	-289.5	-289.0	-288.3	-287.5	-287.0

$$IV_t = \alpha_0 + \sum_{i=1}^p \alpha_i IV_{t-i} + \varepsilon_t \quad (5)$$

where IV_t is the implied volatility at time t and ε_t follows the independent and identical distribution, and $E(\varepsilon_t) = 0$. After that, we select the autoregressive lag order of the model through the AIC criterion. The specific formula of the AIC criterion is as follows:

$$AIC = 2k + n \ln\left(\frac{SSR}{n}\right) \quad (6)$$

where k is the lag order, n is the number of observations and SSR is the sum of squared residuals.

Table 2 gives the results of the autoregressive lag order. It is found that when 1 is the number of lag items, the average AIC value can reach the minimum value. Therefore, we select 1 as the optimal number of lag items, that is, selects the AR (1) model to predict the implied volatility series. The specific form is as follows:

$$IV_{i,t} = \alpha_0 + \alpha_1 * IV_{i,t-1} + \varepsilon_t \quad i = 1, 2, 3, 4 \quad (7)$$

where $IV_{1,t}$ is the implied volatility of call options based on the long-term; $IV_{2,t}$ is the implied volatility of call options based on the short-term; $IV_{3,t}$ is the implied volatility of put options based on long-term; $IV_{4,t}$ is the implied volatility of put options based on the short-term.

3.2 ARMA-GARCH model

According to the opinion of Godbey and Mahar^[7], to

improve the prediction effect of implied volatility, we introduce the short-term implied volatility and the long-term implied volatility in the AR (1) model. Besides, Black and Scholes^[8] proposed that the implied volatility sequence of the market can well explain the liquidity characteristics of the market, and can keenly describe the impact of sudden major events on the market. Therefore, we add dummy variables and liquidity variables to the AR (1) model to build an ARMA model:

$$IV_{i,t} = \alpha_0 + \alpha_1 * IV_{i,t-1} + \beta_1 IV_{i,t-1}^s + \theta_1 IV_{i,t-1}^l + \omega_1 D_{i,t-1} Vol_{i,t-1} + \varepsilon_i \quad (8)$$

where IV^s is the daily frequency short-term implied volatility of options within one month from the expiration date, IV^l is the daily frequency long-term implied volatility of options that are more than three months from the expiration date. Vol_t is the trading volume at time t. D_t is the dummy variable to explain the dummy variable for the interference of major emergencies, when Vol_t is in the quantile interval of 20% to 80%, $D_t = 0$; When Vol_t is outside the quantile interval of 20% to 80%, $D_t = 1$.

However, in the regression process of ARCH models, the impact of regression errors is usually neglected, and the GARCH models will further model the variance of the error terms in the ARCH model. We focused on the moving average component of model heteroscedasticity for the impact of the prediction effect, thereby further improving the prediction effect of the implied volatility. Therefore, we perform a rolling regression of the GARCH family model based on 100 window periods for the implied volatility series. Before rolling regression, we use the AIC criterion to determine the lag order of the regression parameters. It is found that when most of the regression parameters select the number of lag items when $p=1$ and $q=0$, the average AIC value can reach the minimum value, so the specific form of the GARCH model constructed in this article is as follows:

$$r_t = \mu_t + h_t \varepsilon_t \quad (9)$$

$$h_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \quad (10)$$

where h_t^2 is the conditional variance. Formulas (8)-(10) constitute the ARMA-GARCH model based on the term structure and liquidity characteristics of different implied volatility built in this paper.

4 Empirical Analysis

Considering that the time series data in the sample have heteroscedasticity and autocorrelation, the characteristics will reduce the effectiveness of the estimator in the model. Based on the autocorrelation heteroscedasticity consistency test method proposed by Newey and West^[9], we adopt Newey-West's model regressions and further adjusts the t-statistic value to further Reduce the autocorrelation and heteroscedasticity of time series data, thereby improving the effectiveness of parameter estimates.

Figures 1 to 4 show the time-varying fitting trend graphs of the AR (1) model and the ARMA-GARCH model based on rolling regression over the entire sample

period. It can be seen from the four figures that during most of the entire sample period from June 1, 2017, to June 30, 2019, the results of R-square of the four types of implied volatility series predicted by the ARMA-GARCH model are significantly better than the AR (1) model, especially the implied volatility prediction sequence based on the long-term call options. The results of R-square predicted by the ARMA-GARCH model are average significantly superior to AR (1) model during all trading days in the entire sample period. And the R-square sequences predicted by the two models have typical time-varying characteristics and synchronization characteristics.

Table 3 gives the statistical results of each prediction model under the entire sample period. In Table 3, from the statistical results of the implied volatility prediction of call options based on the long-term, the average R-square value predicted by the ARMA-GARCH model is 12.9% higher than that of the AR (1) model, and the predicted success rate is increased from 2% to 37%. From the statistical results of the implied volatility prediction of call options based on the short-term, the average R-square predicted by the ARMA-GARCH model is 12.5% higher than that of the AR (1) model, and the success rate of prediction is increased from 0% to 87%. From the statistical results of the implied volatility prediction of put options based on the long-term, the average R-square predicted by the ARMA-GARCH model is 19% higher than that of the AR (1) model, and the success rate of prediction is increased from 0% To 49%. From the statistical results of the implied volatility prediction of put options based on the short-term, the average R-square value predicted by the ARMA-GARCH model is 17.3% higher than that of the AR (1) model, and the predicted success rate is increased from 0% to 84%. In particular, the success rate of the ARMA-GARCH model in predicting the implied volatility sequence of call options based on the short-term and put options based on the short-term is more than 80%, indicating that the implied volatility sequence based on two option types and term structures can be predicted by the ARMA-GARCH model very efficiently.

Table3. statistical results of prediction models during the whole period

call options based on the long-term		
parameter	average R-square	predicted success rate
values: AR (1) model	0.74	2%
values: GARCH model	0.85	37%
call options based on the short-term		
parameter	average R-square	predicted success rate
values: AR (1) model	0.70	0%
values: GARCH model	0.80	87%
put options based on the long-term		
parameter	average R-square	predicted success rate
values: AR (1) model	0.68	0%

call options based on the long-term		
parameter	average R-square	predicted success rate
values: GARCH model	0.84	49%
put options based on the short-term		
parameter	average R-square	predicted success rate
values: AR (1) model	0.62	0%
values: GARCH model	0.75	84%

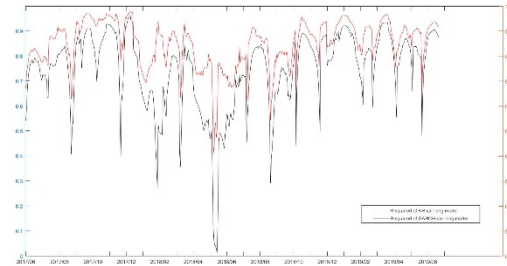


Fig1. R-square of two models of call options based on the long-term.

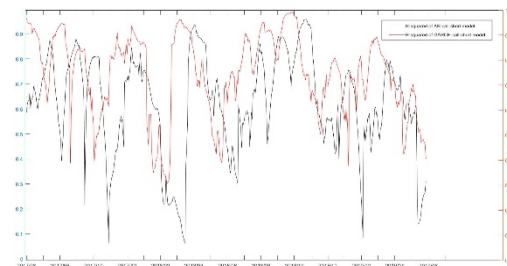


Fig2. R-square of two models of call options based on the short-term.

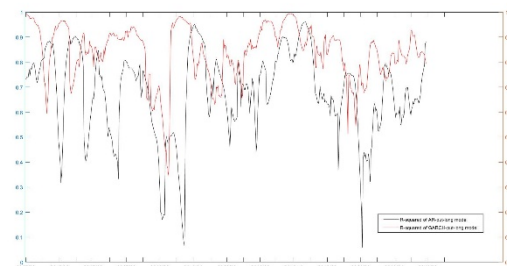


Fig3. R-square of two models of put options based on the long-term.

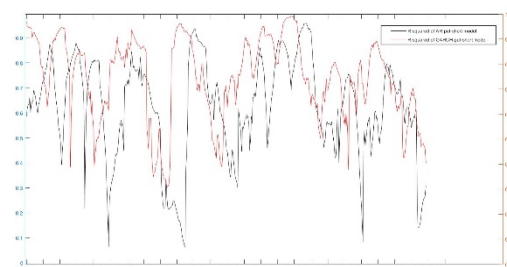


Fig4. R-square of two models of put options based on the short-term.

5 Conclusion

We take the SSE 50 ETF options market as the research object. Based on the sample of all high-frequency daily trading data from June 1, 2017, to August 30, 2019, AR (1) models and ARMA-GARCH models based on liquidity characteristics are built and compared the prediction effects of the two models on implied volatility under different option types and different term structures. The conclusions are as follows: For most of the entire sample period, the effect of model fitting results of the four types of implicit volatility series predicted by the ARMA-GARCH model are significantly better than the AR (1) model; the fitting sequence predicted by two models has typical time-varying and synchronic characteristics; the prediction effect of the ARMA-GARCH model based on the four implied volatility sequences is significantly better than the AR (1) model, verifying Godbey and Mahar^[7] research opinions on the prediction effect of the ARMA-GARCH model.

References

1. French, K R, Schwert G W, Stambaugh R F. Expected Stock Returns and Volatility[J]. *Journal of Financial Economics*,1987, Vol.19, No.1, pp. 3–29.
2. Hui Xiaofeng, Liu Hongsheng, Hu Wei, He Danqing. Prediction of RMB exchange rate based on time series GARCH model [J]. *Finance Research*, 2003(05): 99-105.
3. Efimova O, Serletis A. Energy Markets Volatility Modelling using GARCH[J]. *Energy Economics*,2014, 264-273.
4. Andersen T G, Bollerslev T. Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts[J]. *International Economic Review*, 2001, 39(4):885–905.
5. Black, F, Scholes M S. The Pricing of Options and Corporate Liabilities[J]. *Journal of Political Economy*,1973, 81(3), 637-654.
6. Christoffersen P F, Jacobs K. The Importance of the Loss Function in Option Pricing[J]. *Journal of Financial Economics*,2004, 72, 291–318.
7. Godbey J M, Mahar J W. Forecasting Power of Implied Volatility: Evidence from Individual Equities[J]. *B>Quest*,2007, University of West Georgia.
8. Moriggia V, Muzzioli S, Torricelli C. Call and Put Implied Volatilities and the Derivation of Option Implied Trees[J]. *Front. Finan. Econ*,2007, 4, 35–64.
9. Newey, West. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix [J]. *Econometrica*, 1987, 55(3):703-708.