Empirical Analysis of Constructing GARCH Model to Predict Stock Prices with Trading Volume

Yifan Li(✉)

Beijing Normal University-Hong Kong Baptist University United International College (UIC), Zhuhai, China
evelinee6611@outlook.com

Abstract. The immature stock capital market exhibits a kind of instability and immaturity, which can cause strong volatility in the Chinese stock market. Under such background conditions, how to describe as well as predict the price of the Chinese stock market has become a popular topic of concern for scholars in the financial sector. GARCH family model is a more popular model for studying financial time series in recent years, and with the development of academic research, more scholars have tried to incorporate external influence factors into the model to form an improved GARCH model to improve the fitting and forecasting ability of the GARCH model. Inspired by the above research, this paper will analyse the factors affecting stock price volatility and conduct an empirical study on stock prices and trading volumes in the Chinese stock market. Using daily data on the Shanghai Composite index and trading volumes in the Chinese stock market from 2 January 2020 to 1 December 2022, this paper selects elements that fit the daily data on stock prices to construct a GARCH family model. The GARCH (1,1), TGARCH, and EGARCH models with volume factors are used to estimate, analyse and forecast each time series. The final results show that the best fit and forecast results are obtained for the SSE index return series based on the EGARCH model with the introduction of volume factors.

Keywords: empirical analysis · GARCH models · Chinese stock index · prediction · time series · trading volume · stock price · leverage effect · data model diagnostic analysis · hypothesis testing

1 Introduction

According to Clark's Mixture distribution hypothesis (MDH) [1], information is the driving force behind price changes. In the financial markets, new information is used to understand how the stock market is affected by the movement of stock prices. Trading volume is seen as a reflection of the degree of variation in an investor's information agreement, and as an important factor in investor behaviors, it can largely influence the movement and volatility of stock prices.

The relationship between stock trading volume and price is not only an important means of studying arbitrage opportunities, but also a means of studying the effectiveness

Y. Jiang et al. (Eds.): ICFIED 2023, AEBMR 237, pp. 589–602, 2023.
https://doi.org/10.2991/978-94-6463-142-5_65

of stock markets, and an effective way of understanding the laws and structure of financial markets. In general, the intrinsic driving force of the stock market is the volume of stock transactions, which can visually reflect the supply and demand situation of the stock market and, in certain models, can determine the trend of stock price movements. The relationship between volume and price has long been a popular area of research in finance, with price and volume as two important pieces of information which, to a certain extent, reflect the flow of information on stock fluctuations.

This paper builds on the quantitative strategy methods used, such as ARCH and GARCH models, to combine forecasting stock prices with stock trading volume factors and to explore the internal relationship between the two. The addition of a trading volume factor to the base model allows for a new integrated model to be obtained which can be further extended and new and improved ideas can then be fitted to a more accurate model. The GARCH family of models is based on historical data and different models are used to fit different market stages in the Chinese stock market, and it is the final objective of this paper to select the model that produces the most desirable forecasting results.

Therefore, this paper selects the price, trading volume, and other data of the Shanghai Composite index from the year 2020 to 2022 to construct a GARCH family model, filter the best-fit and prediction model, introduce an equal trading volume process based on equal time process, and form a stock price series analysis method based on the trading volume process. The results of this study contribute to a better understanding of the relationship between trading volume and price volatility in the financial system, and the empirical study for the Chinese market has important reference value for practical investment and risk management.

The paper is structured as follows: Sect. 2 introduces the relevant models; Sect. 3 presents the empirical analysis of selected stocks in the Chinese stock market, etc. Section 4 presents the selection of the best forecast model and its corresponding results; Sect. 5 presents the conclusions of the paper.

2 Methodology

2.1 GARCH Family Models

2.1.1 GARCH Model

Bollerslev (1986) [2] proposed a useful extension, the GARCH model. The mean-corrected log returns are set and the specific model form is as follows:

$$Y_t = \mu_t + \varepsilon_t \tag{1}$$

$$\varepsilon_{\mathrm{t}} | I_{t-1} \sim \mathrm{N}\left(0, \sigma_t^2\right) \tag{2}$$

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{3}$$

where Y_t denotes the return at day t, μ_t denotes the conditional mean of the return given under the information set I_{t-1} for the past t-1 periods, σ_t^2 is the variance of the distribution of the random error ε_t given the conditions, and in order to ensure that the conditional variance σ_t^2 is always positive, it is usually assumed that the condition

$$\omega > 0;\ \alpha, \beta \geq 0 \tag{4}$$

With only three unknown factors, this model definition often performs quite well and is simple to estimate: ω, α, β. One of the most popular models for financial time series is GARCH (1,1) since it can manage highly broad volatility structures with just three parameters (Cipra, 2008) [3]. The GARCH model expresses long-memory properties with a smaller lag order, which can improve the ARCH model to some extent.

2.1.2 Threshold GARCH (TGARCH) Model

Since ARCH and GARCH models produce the same volatility for the same level of favorable and harmful shocks, however, in real stock markets, observers find that harmful shocks have a greater impact than favorable shocks of the same magnitude. To address the 'leverage effect' that GARCH fail to describe, Zakoian (1994) and Glosten et al. (1993) [4] added a multiplicative dummy variable to the model equation to represent the difference between a negative and a positive shock.

They proposed a TGARCH (1,1) model with the following equation:

$$\sigma_t^2 = \omega + \alpha\mu_{t-1}^2 + \gamma\mu_{t-1}^2 d_{t-1} + \beta\sigma_{t-1}^2 \tag{5}$$

Dummy variable = $\{1, 0\}$, which takes the value 1 if $\mu_{t-1} < 0$, and 0 otherwise.

When $\mu_{t-1} < 0$, $d_{t-1} = 1$, then the equation can be changed to $\sigma_t^2 = \omega + (\alpha+\gamma)\mu_{t-1}^2 + \beta\sigma_{t-1}^2$, according to the equation it is known that unfavorable information can cause multiplier $(\alpha + \gamma)$ shocks;

When $\mu_{t-1} > 0, d_{t-1} = 0$, then the equation may become $\sigma_t^2 = \omega + \alpha\mu_{t-1}^2 + \beta\sigma_{t-1}^2$, and according to the equation it is known that favorable information causes a multiple α of the shock;

If $\gamma \neq 0$, reflects the difference in the impact of positive and negative shocks on volatility and their magnitude;
If $\gamma = 0$, there is no leverage effect: adverse news and favourable news shocks have the same impact;
If $\gamma > 0$, the volatility is leveraged: the impact of the shock to asset returns from unfavorable news is greater than the impact of the shock from favorable news.

2.1.3 Exponential GARCH (EGARCH) Model

To deal well with the biased nature of the return distribution and the fact that the model is too constrained by the non-negativity of the coefficients, the exponential GARCH (EGARCH) model was first proposed by Nelson (1991) [5], which has a variance equation of:

$$ln\sigma_t^2 = \omega + a\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| + \gamma\frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \beta ln\sigma_{t-1}^2 \tag{6}$$

If $\gamma \neq 0$, then the fluctuations are asymmetric;
If $\gamma < 0$, then there is a leverage effect of fluctuations;
If $\varepsilon_{t-1} < 0$, representing the effect of adverse news, the conditional variance tends to increase, and the effect of adverse information on log volatility is $\alpha(1-\gamma)|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}|$;
If $\varepsilon_{t-1} > 0$, representing the effect of favorable news, the conditional variance tends to decrease and the effect of favorable information on log volatility is $\alpha(1+\gamma)|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}|$.

2.2 GARCH Family Models with Trading Volumes

Clark (1973) first proposed the Mixture Distribution Hypothesis (MDH), based on information flow theory, which states that trading volume is a positive function of the logarithm of the number of traders and a positive function of the logarithm of information intensity. If all traders receive information at the same time, the absolute value of volume and price change are negatively correlated; if information arrives continuously (different traders receive information at different times), the absolute value of volume and price change are positively correlated [6]. The MDH model has the longest history of any model that examines the volatility of stock prices and trading volumes, and has been the theoretical basis for much of the relevant empirical research.

According to Clark's (1973) mixture of distribution hypothesis (MDH), a key factor in determining how returns on financial assets are affected is the variance of daily price movements, which is controlled by the daily price-relevant information's random number as the mixing variable. The core tenet of MDH is that the random number of daily price-relevant bits of information acting as the mixing variable determines the variance of daily price fluctuations.

Thus, volume volatility may have the ability to predict current volatility and replace the importance of GARCH in affecting daily returns (Lamoureux and Lastrapes, 1990a) [7]. Based on the mixed distribution model, this part uses the volume variable as a proxy for the information arrival process. This paper introduces the volume factor into the GARCH family of models. Each model introduces the volume factor into the GARCH family model.

The equations for the variance equation changes for each model are as follows:

- GARCH (1,1)-V model

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1} + \theta V_t \tag{7}$$

- TGARCH (1,1)-V model

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \gamma\varepsilon_{t-1}^2 d_{t-1} + \beta\sigma_{t-1}^2 + \theta V_t \tag{8}$$

- EGARCH (1,1)-V model

$$ln\sigma_t^2 = \omega + a\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| + \gamma\frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \beta ln\sigma_{t-1}^2 + \theta V_t \tag{9}$$

In the above model, $\theta \geq 0$, this parameter represents the degree of influence of the volume factor on volatility and the other parameters are consistent with the original model.

3 Empirical Analysis

3.1 Data Selection

This paper uses the closing price of the Shanghai Composite Index and the daily data of the volume as the price data and trading volume data of the Chinese stock market respectively, and stores these data in different EXCEL workbooks to complete the collection of data, while completing the data standardization of these data. The data used for the empirical study were the daily log return series and the corresponding daily volume change series, which spanned from 2 January 2020 to 1 December 2022, with a total of 1414 transactions. The data was sourced from Choice Financial Terminal and analyzed using the software Eviews 10.0.

3.2 Calculation of Yield and Change Rates of Volume

To maintain the continuity and comparability of stock prices, we have compounded the closing prices. The daily yield of a stock is defined as the first-order difference between the logarithms of the closing prices of the two adjacent trading days [8] (current price $\mathrm{P_t}$ and previous day's price $\mathrm{P_{t-1}}$):

$$R_t = (log(P_t) - log(P_{t-1})) \tag{10}$$

Define the rate of change in the daily trading volume of an individual stock as the first order difference between the logarithmic values of two adjacent trading days' volumes [8] (the current volume $\mathrm{Vol_t}$ and the previous day's volume $\mathrm{Vol_{t-1}}$):

$$V_t = (log(Vol_t) - log(Vol_{t-1})) \tag{11}$$

3.3 Statistical Characteristic

The historical data obtained for the Shanghai Composite Index was data visualized to obtain a statistical chart as well as the average daily trading volume for the period 2020–2022. As this time series of closing prices and trading volumes are volatile, the daily closing prices and daily trading volumes are logarithmized and then first order differentially differentiated to obtain the corresponding time series of returns and volatility of trading volumes, the Figs. 1 and 2 shows the visualisation results.

According to Figs. 1 and 2, the time series of returns and volume volatility become smoother than before the log-differentiation process, but the volatility is still high.

According to Table 1, the skewness of the yield series is -0.8174 $<$ 0, which indicates that it is left-skewed; while the skewness of the volume volatility series is 0.298474 $>$ 0, which indicates that it is right-skewed. For both, their p-values of Jarque-Bera statistic are less than 0.05, which means that they should reject the null hypothesis and the yield and volume volatility series obeys normal distribution at 5% significant level.

Therefore, the statistical eigenvalues also indicate that the daily return series of the Shanghai Composite Index and the volume volatility series do not follow the normal distribution of the efficient market hypothesis.

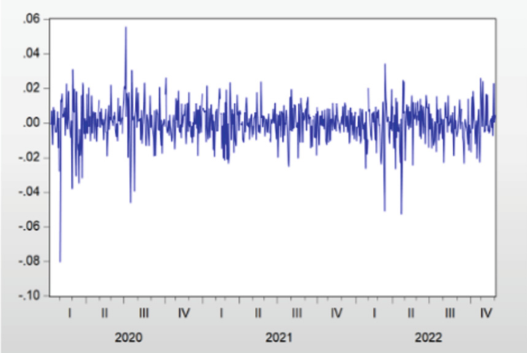

Fig. 1. Yield time series

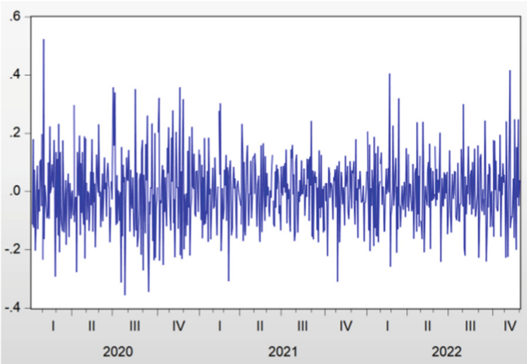

Fig. 2. Volume volatility time series

3.4 Stationarity Test

The Granger causality test and the GARCH model both require the time series to be stationary, so this paper tests the stationarity of the stock return and volume volatility series. The most used test for smoothness is the unit root test, where the test for smoothness is the Augmented Dickey-Fuller (ADF) (1979) and Phillips-Perron (1988) tests [9] of the unit root verification method. The results of the unit root test for both methods are shown in the following graph:

In the unit root test, its null hypothesis is that the series has a unit root. The results of the two methods were summarized and compared and the results are shown in Table 2.

As can be seen from Table 2, the statistics under both tests are much smaller than the critical values at the 1% test level. Therefore, we can reject the null hypothesis: (H0: there is no unit root in the return series as well as the volume volatility series). And we can conclude that the test results show that both the daily return series of the Shanghai Composite Index and the volume volatility series are stationary, and thus further empirical analysis can be conducted.

Table 1. Statistical properties of the yield series and the volatility of trading volume

	YIELD	VOLUME VOLATILITY
MEAN	3.64e-05	0.000397
MEDIAN	0.000413	-0.000589
MAXIMUM	0.055542	0.552409
MINIMUM	-0.080392	-0.355401
STD.DEV.	0.011331	0.120427
SKEWNESS	-0.817420	0.298474
KURTOSIS	8.666691	3.783670
JARQUE-BARA	1023.232	28.54848
PROBABLILTY	0.000000	0.000001

Table 2. Unit Root Test Results

VARIABLE	ADF TEST		PHILLIPS- PERRON TEST	
	t-Statistic	Prob	Adj.t-Stat	Prob
Rt	-26.08668	0.0000	-26.17589	0.0000
Vt	-19.41133	0.0000	-43.0975	0.0001

Table 3. Ljung- Box test result

Auto/PARTIAL Correlation	LJUNG- BOX test	
	Q-stat	Prob
1	0.1847	0.667
2	0.4360	0.804
3	0.8020	0.849
4	3.5889	0.464
5	4.2132	0.519

3.5 Autocorrelation Test

The volatility clustering characteristics of financial time series data suggest the existence of autocorrelation and heteroskedasticity, which are tested in this paper. Two methods are used to test the residuals for autocorrelation: the Ljung- Box test and the LM test [10]. The results of the Ljung- Box test is shown in Table 3.

Table 4. LM test result

LM TEST	F-STATISTIC	PROB.F	OBS*R-SQUARED	PROB.CHI-SQUARED
VALUE	0.269348	0.8475	0.811715	0.8467

The p-values of the autocorrelation coefficient and the partial autocorrelation coefficient of Ljung-Box Q of the yield series were both within the 5% significance level, so we could not reject the null hypothesis that the residuals were not autocorrelated.

The LM test, the Lagrange multiplier test, is used to test whether there is serial correlation in the series of model residuals. The null hypothesis is that there is no serial correlation; the alternative hypothesis is that there is p-order autocorrelation. Unlike the D-W test, which can only test for first-order autocorrelation, the LM statistic can test for both first-order and higher-order autocorrelation, so we prefer to use the LM test for residual autocorrelation. The test statistic asymptotically follows a chi-square distribution and if the calculated p-value is too large then the null hypothesis is rejected and serial correlation is considered to exist.

The LM test was then applied to test the autocorrelation of the residuals, and the results of the LM test are shown in Table 4.

The results of the LM test show that the p-value of Prob. Chi-Square is greater than 0.05, so we cannot reject the null hypothesis that there is no autocorrelation. Therefore, for the mean equation, we do not need to make a correction for serial correlation and do not need to build an ARMA model for a correction first on this basis, and can continue to build a GARCH family model on this basis.

3.6 Testing for ARCH Effects

ARCH stands for Autoregressive Conditional Heteroskedasticity and the ARCH effect test is a test for the presence of autocorrelation of multiple orders in the squared residual disturbance terms [11]. The presence of the ARCH effect is the basis for modeling the GARCH family of models. The mean equation for this series was established to test for the ARCH effect and the results are shown in Fig. 3.

As shown in Fig. 3, we can summarize the result as shown in Table 5.

From the heteroskedasticity test, the ARCH test results show that the Prob.F and Prob.Chi-Square values are both less than 0.05, so we can reject the null hypothesis and conclude that there is an ARCH effect in the SSE return time series. (H0: there is no ARCH effect for this series) Since there is an ARCH effect for this series, we can continue to build the ARCH model as well as the GARCH model on this basis.

3.7 Model Fitting

In conjunction with the previous tests described above, this section will apply a GARCH family model to the time series of the Shanghai Composite Index returns for empirical analysis. On this basis, the volume factor will be introduced into the GARCH family model for fitting and analysis, comparing the advantages and disadvantages of fitting

Heteroskedasticity Test: ARCH

F-statistic	3.043489	Prob. F(5,694)	0.0100
Obs*R-squared	15.01967	Prob. Chi-Square(5)	0.0103

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 12/14/22 Time: 15:38
Sample (adjusted): 1/13/2020 12/01/2022
Included observations: 700 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.99E-05	1.62E-05	6.155226	0.0000
RESID^2(-1)	0.112813	0.037958	2.972032	0.0031
RESID^2(-2)	0.070695	0.038196	1.850832	0.0646
RESID^2(-3)	0.013014	0.038330	0.339530	0.7343
RESID^2(-4)	0.019963	0.038237	0.522077	0.6018
RESID^2(-5)	0.008309	0.038001	0.218651	0.8270

R-squared	0.021457	Mean dependent var	0.000129
Adjusted R-squared	0.014407	S.D. dependent var	0.000354
S.E. of regression	0.000352	Akaike info criterion	-13.05987
Sum squared resid	8.58E-05	Schwarz criterion	-13.02086
Log likelihood	4576.955	Hannan-Quinn criter.	-13.04479
F-statistic	3.043489	Durbin-Watson stat	2.001070
Prob(F-statistic)	0.010003		

Fig. 3. ARCH effect test results

Table 5. ARCH effect test results

ARCH TEST	F-STATISTIC	PROB.F (5694)	OBS*R-SQUARED	PROB.CHI-SQUARED (3)
VALUE	3.043489	0.0100	15.01967	0.0103

and forecasting effects of the three GARCH family models under the condition that the standardized residuals of the return series follow a normal distribution, where the three GARCH family models are GARCH, TGARCH, and EGARCH [12].

The fitted results of the GARCH family variance equation model for the Shanghai Composite Index under normal distribution conditions and the model with the addition of the volume factor are shown in Table 6.

3.8 Optimal Model Analysis and Selection

By comparing the information criteria results of each model, AIC, SC, and HQ, it can be determined that EGARCH-V has the smallest AIC, SC, and HQ values among all the models studied under normal distribution, i.e. AIC = -6.412769, SC = -6.374019, and HQ = -6.397796.

Therefore, based on these three information criteria for the best model selection, we finally choose the EGAECH model and add the volume factor to form an improved model: EGARCH-V model for the fitting of the Shanghai Composite index return series. The introduction of the volume factor to the original model results in an improved

Table 6. Fitting results of the GARCH family model under the normal distribution

Model	ω	α	β	γ	θ	AIC	SC	HQ
GARCH	1.76E-05*	0.189413*	0.684070*			-6.228918	-6.203085	-6.204120
GARCH-V	1.91E-05*	0.186626*	0.661882*		0.000355*	-6.357816	-6.325524	-6.345338
TGARCH	2.41E-05*	0.083447*	0.599986*	0.280174*		-6.250388	-6.218096	-6.237910
TGARCH-V	2.33E-05*	0.013693	0.609398*	0.358187*	0.000337*	-6.399979	-6.361229	-6.385005
EGARCH	-2.054028*	0.355075*	0.802988*	-0.167806*		-6.248437	-6.216145	-6.235959
EGARCH-V	-1.458340*	0.214442*	0.859421*	-0.330379*	3.991169*	-6.412769	-6.374019	-6.397796

Remarks: *, **, *** denote significance at the 1%, 5%, and 10% levels of significance, respectively, and unlabeled as insignificant. where G/T/EGARCH-V is the improved model of the original model with the addition of the trading volume factor, respectively.

GARCH family of models, most of the coefficients of which are significant at the 1% level, and can therefore show that the introduction of the volume factor has good explanatory power for volatility. The asymmetry coefficients are all significant at the 1% level of significance after the introduction of the volume factor, and can therefore indicate that the introduction of the volume factor improves the portrayal of volatility asymmetry in the GARCH family of models.

The sum of the residual disturbance term **α** and the conditional variance lagged first-order term **β** is significantly less than 1, thus it can indicate that the introduction of the volume factor reduces the persistence of volatility [13]. The conditional variance lagged first-order term represents that the stock market has not fully absorbed the shock of the previous period's information, if this coefficient becomes smaller, it means that the volume factor can absorb the impact of the previous part of the shock on the current period.

In summary, this yield volatility has a long memory, aggregation as well as asymmetry. The introduction of the volume factor improves the fit of the GARCH family of models. The overall empirical results show that EGARCH-V has the smallest AIC, SC, and HQ values under a normal distribution, and therefore the EGARCH-V model has a better fit than other models. In addition, the EGARCH-V model eliminates the ARCH effect from the residuals of the equation and verifies that the Shanghai Composite Index volatility is leveraged and that the model is reasonable.

4 Model Prediction

4.1 Classification of the Training and Test Sets

The analysis of the empirical results in the previous section under the assumption of normal distribution shows that the introduction of the volume factor in the EGARCH model can better fit the Shanghai index return volatility. In this section, the modeling training set will be used as the basis for the model based on the above optimal model EGARCH-V under different distribution assumptions, such as normal distribution, t-distribution, and GED-distribution, and the modeling training interval selected for this empirical analysis is 2020.1.2–2022.1.2, with a total of two years of observations. For

Table 7. Comparison of EGARCH-V forecasting models

DISTRIBUTION	MODEL FORM	RMSE	MAE	MAPE	TIC
NORMAL	EGARCH-V	36.85229	27.04853	0.844141	0.005689
T	EGARCH-V	36.95453	27.07181	0.845009	0.005703
GED	EGARCH-V	36.91964	27.06034	0.844620	0.005698

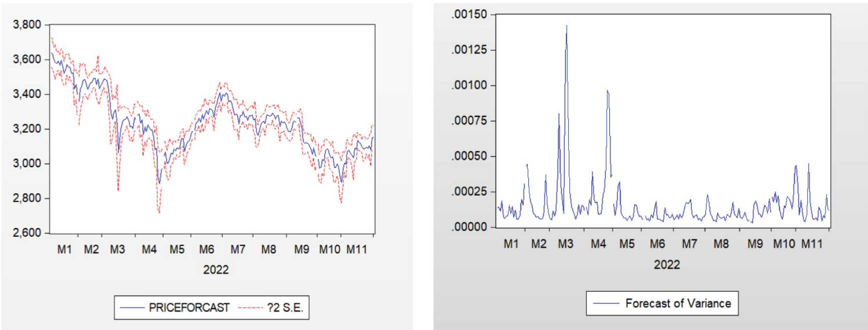

Fig. 4. EGARCH-V forecasting model results

the testing interval, the selected forecast interval is 2022.1.3–2022.12.1, for a total of approximately one year of observation, where the predicted values are compared with the true values to determine the predictive effectiveness of the model. Under different distribution assumptions, the one with a high level of parity in the forecast evaluation indicators was selected as the best forecast model distribution.

4.2 Selection and Comparison of Predictive Model Evaluation Indicators

In this paper, four prediction evaluation indicators, namely Root Mean Squared Error, Mean Absolute Error, Mean Absolute Percent Error, and Theil Inequality Coefficient will be used to compare EGARCH-V under different hypothetical distributions to select the optimal prediction model and distribution.

According to Table 7, we can find that based on the EGARCH model introducing the trading volume factor, under the normal hypothesis distribution, its forecast evaluation indicators are the smallest forecast evaluation indicators of the three distribution methods.

4.3 Presentation of the Results of the Best Forecasting Model

Based on the best forecasting model selected in the previous section, the EGARCH-V model with normal distribution is chosen for forecasting, and the modeling interval chosen is 2020.1.2–2022.1.2. The forecasting results obtained by the model are shown in Figs. 4 and 5.

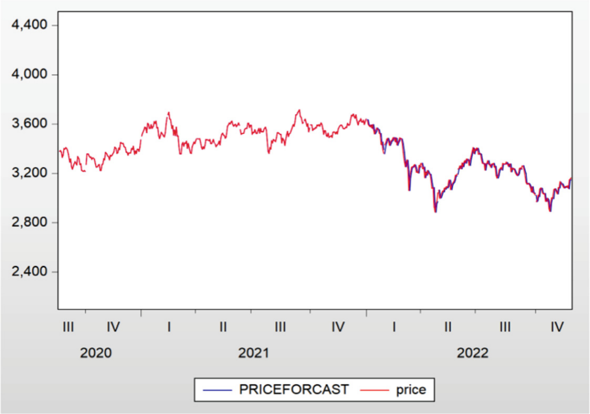

Fig. 5. Real SSE Index Price vs Forecast

As shown Figs. 4 and 5, the red dash is the real stock price from 2020.1.2 to 2022.12.1, and the blue dash is the forecast stock price from 2022.1.3–2022.12.1. This comparison chart shows that the EGARCH-V model with normal distribution can accurately predict the stock prices and trends of the SSE from 2022.1.3 to 2022.12.1.

5 Conclusion

This paper constructs a GARCH family model based on historical data of the Shanghai Composite index return series for empirical analysis. In this paper, the returns are fitted in GARCH (1,1), TGARCH (1,1), EGARCH (1,1), and the improved models GARCH-V, TGARCH-V, and EGARCH-V formed by adding the trading volume factor respectively, and EGARCH-V is judged to be the best-fitted model according to the AIC, SC, and HQ model evaluation indexes. In the asymmetric study of TGARCH and TGARCH-V, the estimated value of the coefficient of the asymmetry term is $\gamma > 0$, indicating that there is a significant "asymmetry effect" in the series, which is expressed as a "leverage effect" [14]. It can therefore be shown that there is a significant presence in the volatility of our Shanghai Composite Index, i.e. good news and bad news have different degrees of impact on the conditional variance. To the same extent, "bad news" has a greater impact on stock market volatility than "good news" [15]. The sum of the residual disturbance term $\boldsymbol{\alpha}$ and the lagged first-order term $\boldsymbol{\beta}$ of the conditional variance is significantly less than one, thus suggesting that the introduction of the volume factor reduces the persistence of volatility. The volume factor is an important proxy for market information flow and its introduction can significantly reduce the GARCH effect in the volatility model.

The EGARCH-V model is fitted to the training set to predict the test set. The best EGARCH-V distribution was selected for the Shanghai Composite index returns based on four forecast evaluation indicators: RMSE, MAE, MAPE, and TIC, among the normal, t-distribution, and GED distribution [16]. The results of the ARCH-LM test on the best model and distribution show that the model residuals do not have the AECH effect, i.e. the model forecasts and fitting are reasonable. Therefore, the results show that the best forecasts can be obtained for the Shanghai Composite index return series under a normal distribution with EGARCH model with the introduction of volume factors.

References

1. Clark P.K.A (1973), Subordinated stochastic process model with finite variance for speculative prices. Econometrica, 41: 135-155.
2. Bollerslev T. (1986), Generalized Auto regressive Conditional Heterosc edasticity. Journal of Econometrics, 31: 307-327.
3. Cipra, T. (2008). Finanční ekonometrie. The dynamic relationship among return, volatility and trading volume in China stock market – an empirical study based on quantile regression. In: Proceedings of the Sixth International Symposium on Corporate Governance, 75–85.
4. Xiaoting G. (2006) Volatility characteristics of the Chinese securities investment fund market, Chinese Journal of Management Science pp.,1003-207
5. Nelson D. (1990),Conditional heteroskedasticity in asset returns: a new approach, Econometrica, 59: 347-370.
6. Shengrong C., Hongzhen G. (2004) Analysis of GARCH model of trading volume and price volatility, New Inquiry into Theory, No. 7, 2004
7. Christopher G. Lamoureux and William D. Lastrapes (1990), Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects, The Journal of Finance, 45: 221-229.
8. Feng F., Hongliang W. (2005) Application of GARCH model in studying the relationship between stock market trading volume and stock price, Finance and Economics, 12:0033-35
9. Plachý, R. (2014). Impact of Trading Volume on Prediction of Stock Market Development. Acta Univ. Agric. Silvic. Mendel. Brun., *62*(6), 1373-1380. DOI: https://doi.org/10.11118/actaun201462061373
10. Huijian G. (2008), Application and empirical analysis of the volume price relationship in GARCH-like models, Qingdao: Shandong University of Science and Technology
11. Ling L. (2016) Research on the predictive ability of stock market volatility fitting based on improved GARCH model, Chengdu: Southwest University of Finance and Economics
12. Ximei L., Xinyu L., Yuhe C., Tianli Z. (2021), A study on DOI the volatility of SZSE index based on GARCH family model, Financial Sight,28:94-97. DOI: https://doi.org/10.19699/j.cnki.issn2096-0298.2021.08.094.
13. Gao J. (2021)Research on Stock Price Forecast Based on ARIMA-GARCH Model, E3S Web of Conferences 292, 02030 (2021)
14. Haiwei P., Zudi L., (2009), Nonlinear analysis of financial systems: nonlinear effects of trading volume on stock volatility, Systems Science and Mathematics, 11: 1527-1541
15. Nguyen T. and Bui T.T.M. (2018), Modeling the Volatility and Forecasting the Stock Price of the German Stock Index (DAX30), International Journal of Economics and Financial Research, pp. 72-92
16. Yilin Z., Yuan Z. (2012) A study on the dynamic relationship between trading volume and stock price volatility, Fudan Journal (Natural Sciences), Vol. 51 No. 4