ISSN: 2692-7608 DOI: 10.6981/FEM.202406_5(6).0016

Forecasting Stock Market Volatility based on GARCH-like Models -- An Example of CSI 300 Index

Xinyue Wang*

School of Mathematics and Statistics, Beijing Technology and Business University, Beijing, China

*Corresponding author: xinyuewangg@outlook.com

Abstract

Stock market volatility prediction has been a topic of great interest in the financial field. Stock market volatility prediction methods based on GARCH-like models have been widely used in academia and practice. Taking CSI 300 index for the research target, this paper explores the application and effect of TGARCH model, EGARCH model and GARCH-M model in stock market volatility forecasting. Firstly, this paper introduces the basic principles of GARCH-type models and the application background in the financial field. Subsequently, by analyzing the historical data of CSI 300 index, TGARCH model, EGARCH model and GARCH-M model are constructed, and the fitting effects of different models are compared and evaluated, and some conclusions and insights are drawn. The research of this paper can provide investors with a more scientific method for predicting the volatility of the stock market and provide reference for the stable development of the financial market.

Keywords

GARCH Family Model; Stock Market; CSI 300; Volatility.

1. Introduction

Stock market volatility is a very important indicator in the financial market, which is significant for investors, policy makers, and academics. Forecasting volatility in the stock market is a challenging problem because changes in volatility can directly affect the risk and return of investors. Therefore, forecasting the volatility of the stock market by building effective models is important for investors' decision making.

GARCH-type models, as a commonly used volatility prediction model, have been widely used in the financial field. Among them, TGARCH model, EGARCH model and GARCH-M model are three common GARCH-like models. While these models share common features in capturing the dynamic characteristics of volatility, they each differ in their treatment of asymmetric effects as well as their estimation methods. Among them, the TGARCH and EGARCH models are specifically designed to capture the asymmetric effects of the impact of short versus good news on volatility. Although the GARCH-M model additionally considers how volatility directly affects returns, providing investors with a basis for measuring risk premiums. Although they have some differences in model structure and estimation methods, they can all better capture the characteristics of stock market volatility, and their forecasting effectiveness can be assessed by comparing and analyzing them.

In this paper, we take CSI 300 index as an example and use TGARCH, EGARCH and GARCH-M models. Through the comparative analysis of these three models, we can get a more profound comprehension of the forecasting effect of different models on market volatility as well as their respective advantages and shortcomings. Based on academic research and theory, this paper

ISSN: 2692-7608 DOI: 10.6981/FEM.202406_5(6).0016

will explore the applicability and effectiveness of these models to provide investors with more useful market forecasting information, hoping that the research results can provide new perspectives and inspirations for volatility forecasting research in the stock market.

2. Introduction to the Model

2.1. ARCH Model

The following is the basic idea behind the 1982 proposal of the ARCH model:

The mean value equation:

$$x_t + \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \ldots + \beta_p x_{t-p} + u_t \tag{1}$$

variance equation:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i u_{t-i}^2 + \sum_{j=1}^{q} \beta_j \delta_{t-j}^2 \tag{2}$$

The (p+1) parameters in the variance equation: α_0 , α_1 , α_2 ,..., α_q , and in the mean equation:β_0, β_1, β_2,...,β_q, need to be estimated using the great likelihood estimation.

This is because the parameters in the ARCH model control the trend and persistence of volatility. When the parameter value is large, the change of volatility will be more dramatic; when the parameter value is small, the change of volatility will be more gentle. Thus, the most important aspect of the ARCH model is that it can capture the aggregation of volatility, which means that periods of high volatility generally last for a period of time. Additionally, times of low volatility last for a while. [1].

Due to the large number of parameter estimates required in the ARCH model, make ensure that every parameter satisfies the criteria is really challenging, and the ARCH model implies that both positive and negative shocks have a comparable effect on the model during the building phase, so there is no leverage effect in the study of volatility.[2] Therefore, Engle's students, Bollerslev and Taylor, in order to solve the deficiencies in the model created by their teacher, they proposed the GARCH model in this regard.

2.2. GARCH Model.

The underlying principle of the GARCH model is outlined as follows:

The mean value equation:

$$x_t + \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_p x_{t-p} + u_t \tag{3}$$

variance equation:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i u_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{4}$$

The generalized autoregressive conditional heteroskedasticity model, or GARCH model, adds an exponentially weighted average term of past time fluctuations to the basis of the ARCH model, enabling the long-term volatility memory properties to be more accurately captured by the GARCH model.

ISSN: 2692-7608 DOI: 10.6981/FEM.202406_5(6).0016

The variance equation of the GARCH model shows that just three parameters are needed to express the GARCH model, while the ARCH model exists an infinite number of parameter equations, which undoubtedly provides some convenience in terms of parameter estimation of the model. And the use of GARCH model can play the role of order reduction and improve the accuracy of the whole model. Later, Zakoian and Olosten found that there is asymmetry in the New York stock market volatility study. Since the current GARCH model finds it difficult to explain the existence of asymmetry, these two scientists developed the TGARCH model based on the GARCH model in order to address the issue.

2.3. TGARCH Model

Threshold ARCH model, also referred to as TGARCH model. It was independently put up by Glasten, Jaganathan, and Runkle (1994) and Zakaran (1990).The model is as follows:

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

Where d_{t-1} is a dummy variable and $d_{t-1} = 1$ when $u_{t-1} < 0$, otherwise $d_{t-1} = 0$. Asymmetry exists as long as $\gamma \neq 0$. The asymmetric term in this equation is $\gamma \times u_{t-1}^2 d_{t-1}$. Upon scrutiny of the model, it becomes apparent that positive news ($u_{t-1} > 0$) and negative news ($u_{t-1} < 0$) exerts distinct effects on the model. Since $d_{t-1} = 1$ when $u_{t-1} < 0$, otherwise $d_{t-1} = 0$. For good news, $u_{t-1} > 0, d_{t-1} = 0$, which means that $\gamma \times u_{t-1}^2 d_{t-1}$ is 0. Therefore, the good news brings an α-fold shock to the model, while the bad news brings an (α+γ)-fold shock to the model. Therefore, in the event that γ is greater than zero, the asymmetric effect leads to a market that progressively exhibits heightened volatility; conversely, in the event that γ is less than zero, the asymmetric effect results in a market that progressively experiences decreased volatility.[2].

2.4. EGARCH Model

Given the limitations of the aforementioned three models in adequately explaining the presence of the leverage effect, Nelson introduced the EGARCH model in 1991, which is expressed as follows:

$$\ln(\sigma_t^2) = \omega + \sum_{j=1}^{q} \beta_j \ln(\sigma_{t-j}^2) + \sum_{i=1}^{p} \alpha_i \left| \frac{u_{t-i}}{\sigma_{t-i}} - E\left(\frac{u_{t-i}}{\sigma_{t-i}}\right) \right| + \sum_{k=1}^{r} \gamma_k \frac{u_{t-k}}{\sigma_{t-k}} \tag{6}$$

From the formula of this model, it can be seen that the leverage effect is exponential, and when γ<0, the leverage effect is more obvious, for the stock market, it is subject to negative shocks resulting in its volatility is greater than the volatility caused by the positive shocks it is subject to, with leverage effect. The advantage of this model over GARCH and ABSGARCH is that it can differentiate between the effects of positive and negative news rates. Positive new interest rates are "good" and negative new interest rates are "bad". Despite the fact that the magnitudes of positive and negative news interest rates are identical, the EGARCH model is capable of discerning the disparate impacts of positive and negative news interest rates on volatility. [3].

2.5. GARCH-M Model

The GARCH-M model, denoting the conditional mean of returns with the designation "M," is referred to as GARCH (GARCH in the mean). A basic GARCH(1,1)-M model can be formulated as:

$$\gamma_t = \mu + c\sigma_t^2 + \alpha_t \tag{7}$$

$$\alpha_t = \sigma_t \varepsilon_t \tag{8}$$

ISSN: 2692-7608 DOI: 10.6981/FEM.202406_5(6).0016

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (9)$$

where μ and c are constants, the parameter c is known as the risk premium parameter. A positive value of parameter c indicates a positive correlation between returns and volatility, indicating a higher likelihood of market volatility and increased investment risk.

3. Data Sources and Variable Design

3.1. Data Source

The sample interval is from January 2, 2018 to November 3, 2023 and the data source is Choice.

3.2. Variable Design

In this study, the hs300 variable is employed to proxy for the closing price of the CSI 300 index during the data analysis phase. The CSI 300 index is a joint initiative of the Shanghai Stock Exchange and the Shenz Stock Exchange, comprising 300 equities from China's A-share market that are characterized by their size and trading liquidity. The index is designed to mirror the collective performance of large-italization stocks listed on the Shanghai and Shenzhen stock exchanges.

4. Empirical Analysis

4.1. Stability Test

Firstly, the closing price series is tested for smoothness and the results are as follows:

Table 1. Closing price series for smoothness test Table

Exogenous: Constant			
Lag Length: 0 (Automatic - based on SIC, maxlag=23)			
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.517427	0.5247
Test critical values:	1% level	-3.434759	
	5% level	-2.863375	
	10% level	-2.567795	
*MacKinnon (1996) one-sided p-values.			

Based on the results from the Augmented Dickey-Fuller (ADF) test, it is observed that the P-values of the hs300 series exceed the 0.05 threshold. However, the corresponding T-statistic value surpasses the critical T value at the 5% significance level, suggesting a lack of adherence to the assumption of smoothness within the series. To address this issue and ensure the smoothness of the financial time series, it is imperative to transform the original series by applying a logarithmic transformation and conducting a first-order differencing process to generate the dlnhs300 series. Subsequently, a follow-up ADF test is necessary to verify the attainment of a smooth series.

Upon examination of the results from the Augmented Dickey-Fuller (ADF) test, it is apparent that the p-value associated with the dlnhs300 sequence falls significantly below 0.05. Furthermore, the corresponding T-statistic value is below the critical T value at the 5% significance level. Consequently, the null hypothesis is rejected, affirming that the sequence complies with the smoothness assumption and mitigates the likelihood of pseudo-regression. This finding will be further scrutinized in subsequent analysis.

Frontiers in Economics and Management Volume 5 Issue 6, 2024
ISSN: 2692-7608 DOI: 10.6981/FEM.202406_5(6).0016

Table 2. ADF test table

Null Hypothesis: DLNHS300 has a unit root			
Exogenous: Constant			
Lag Length: 0 (Automatic - based on SIC, maxlag=23)			
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-37.62458	0.0000
Test critical values:	1% level	-3.434762	
	5% level	-2.863376	
	10% level	-2.567796	
*MacKinnon (1996) one-sided p-values.			

4.2. Time Series Plots

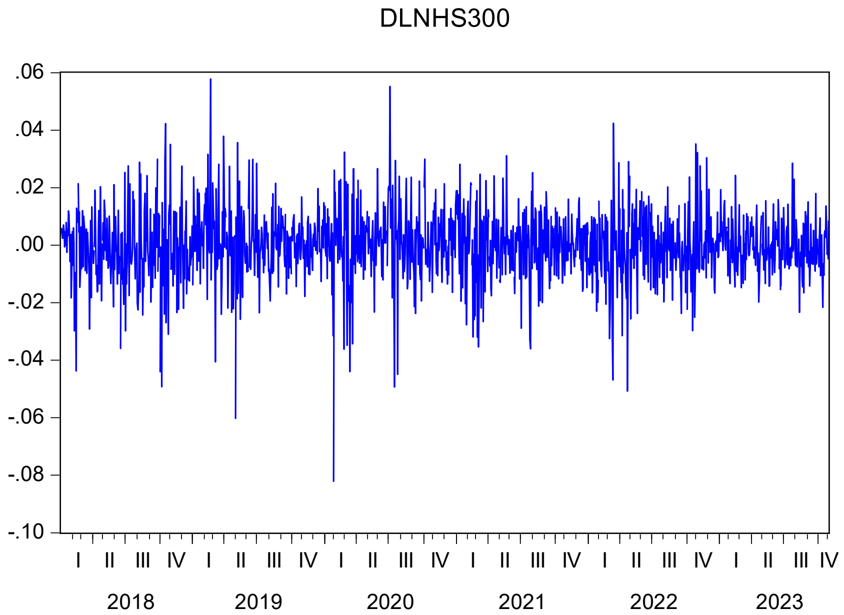

Figure 1. Time series plot

Upon visual inspection of the time series plot depicting the variable returns, an initial assessment suggests that the series exhibits a stable, non-directional pattern, devoid of a discernible upward or downward trend.

4.3. Descriptive Statistical Analysis

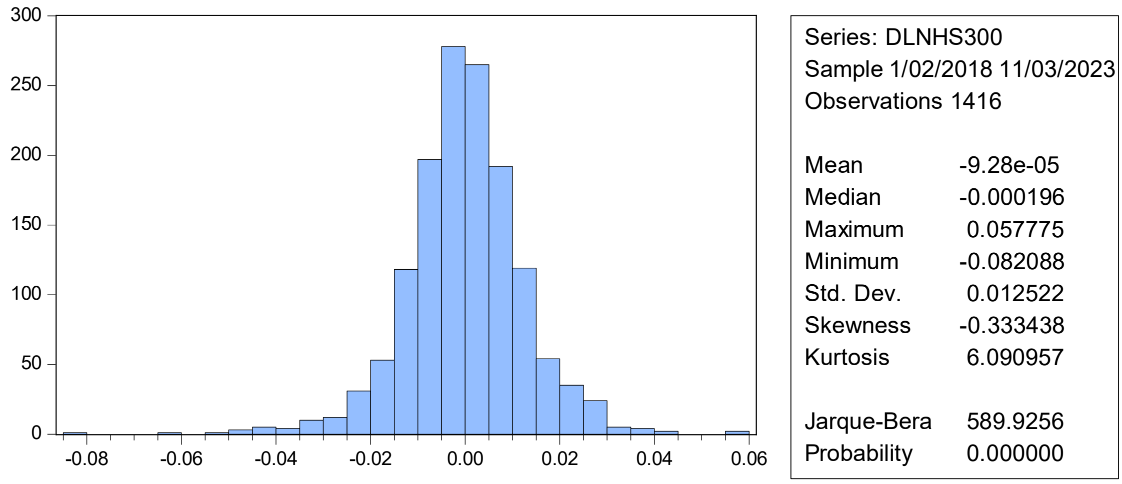

Figure 2. Descriptive statistical analysis chart

ISSN: 2692-7608 DOI: 10.6981/FEM.202406_5(6).0016

The descriptive statistics analysis reveals that the skewness of the dlnhs300 yield series is -0.333438, with a median of -0.000196, and a mean of -9.28e-05. The sample has a left-skewed distribution with a spiky thick-tailed state; the skewness is -0.333438, which is less than 0, and the kurtosis is 6.090957, which is larger than 3. Simultaneously, the sequence exhibits non-conformance to a normal distribution, while the Jarque-Bera (J-B) statistic and the associated probability approach zero. This trend is consistent with the prevalent characteristics found in financial time return series data.

4.4. ARCH Effect Test

Table 3. ARCH effect test table

Heteroskedasticity Test: ARCH			
F-statistic	15.73315	Prob. F(1,1413)	0.0001
Obs*R-squared	15.58192	Prob. Chi-Square(1)	0.0001

The ARCH effect refers to the heteroskedasticity of stock or financial asset price fluctuations.ARCH stands for Autoregressive Conditional Heteroskedasticity. In finance, the ARCH effect suggests that the volatility of asset prices may be unstable over time, i.e., the variance of the volatility is not constant, but is correlated with past volatility. The model posits that the current volatility is reliant on the historical volatility, thereby enabling a more precise estimation of the prospective volatility levels.

The model exhibits a pronounced ARCH effect, highlighting the volatile nature of the variable's return fluctuations. The outcomes of the sequential ARCH-LM test indicate that the significance of the lagged squared residual term is statistically significant, with the likelihood of Obs*R-squared converging towards zero. Additionally, the P-value associated with the joint significance of the model's F-statistics is substantially below 0.05. Therefore, the rationale for constructing a GARCH model to characterize the yield changes is duly justified.

4.5. GARCH Model and Standardized Residual Analysis

Using the aforementioned formula, this work has built a GARCH(1,1) model to better characterize the time-varying nature of serial returns and the accumulation of volatility. The results of the estimation are shown below. It is clear from the study of the descriptive statistics that the series does not follow a normal distribution. As a result, we have selected the T distribution/GED distribution as the error distribution. These are the outcomes that were obtained.

Table 4. Table of GARCH test results

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000127	0.000284	0.444901	0.6564
	Variance Equation			
C	5.37E-06	1.94E-06	2.769618	0.0056
RESID(-1)^2	0.059801	0.014677	4.074353	0.0000
GARCH(-1)	0.905988	0.022877	39.60305	0.0000
T-DIST. DOF	6.393770	1.057013	6.048903	0.0000
R-squared	-0.000307	Mean dependent var		-9.28E-05
Adjusted R-squared	-0.000307	S.D. dependent var		0.012522
S.E. of regression	0.012524	Akaike info criterion		-6.056887
Sum squared resid	0.221937	Schwarz criterion		-6.038329
Log likelihood	4293.276	Hannan-Quinn criter.		-6.049954
Durbin-Watson stat	2.000911			

ISSN: 2692-7608 DOI: 10.6981/FEM.202406_5(6).0016

The GARCH model, an acronym for Generalized Autoregressive Conditional Heteroskedasticity, is an advanced extension of the ARCH model utilized to characterize the presence of heteroscedasticity, or volatility instability, observed within financial time series data. Introduced by Tim Bollerslev in 1986, the GARCH model, akin to its predecessor, the ARCH model, incorporates the influence of past volatility levels on the current volatility dynamics.The difference is that the GARCH model introduces the squared volatility of the past moment as an influencing factor, making the model more flexible in capturing the dynamic nature of volatility.

The analysis from the table above reveals that the series successfully passes the test at the 5% confidence level, thus justifying the overall joint significance of the model's lagged terms. Moreover, the coefficients associated with both the ARCH and GARCH terms within the model exhibit statistical significance at the 1% significance level and exhibit positive values, consistent with the prescribed criteria for GARCH model parameters. Moreover, the combined sum of these coefficients falls below and closely approximates 1, adhering to parameter constraints and indicating persistent price volatility. This result underscores the crucial role of currently available information in predicting future risk levels.

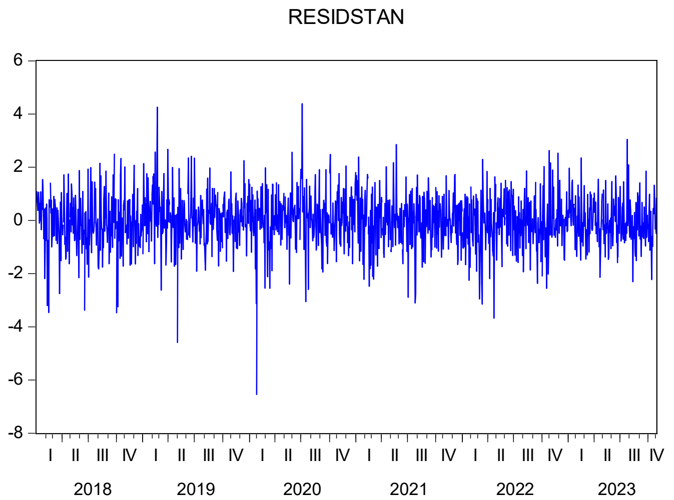

Figure 3. Plot of standardized residuals of the GARCH model

The plot displaying the standardized residuals derived from the GARCH model is presented in Figure. In the standardized residuals, the mean value is 0, and the residuals show a white noise state, indicating that the GARCH model fits better.

4.6. TGARCH Model

The TGARCH model, as an extension of the GARCH model, incorporates the concept of Leverage Effect into its framework. Leverage effect refers to the phenomenon of increased volatility in financial markets when asset prices fall.The TGARCH model takes this leverage effect into account and considers it in its modeling.The full name of TGARCH is Threshold GARCH, which is a threshold GARCH model. In the TGARCH model, a threshold variable is introduced so that when the asset price falls below a certain threshold, the change in volatility is affected by different parameters to better capture the leverage effect.

The table provided above illustrates the TGARCH model for the hs300 index, with the term RESID(-1)^2*(RESID(-1)<0) serving as the leverage component. This term quantifies the asymmetric impact within the data, and its statistical significance is indicated as 0.0031, which falls below the threshold of 0.01. Consequently, it satisfies the criteria for significance at the 1% level, suggesting that the hs300 index exhibits a leverage. Specifically, this implies that the hs300 index reacts differently to positive and negative news, with the former having a more favorable impact than the latter.

ISSN: 2692-7608 DOI: 10.6981/FEM.202406_5(6).0016

Table 5. Table of TARCH test results

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-8.39E-05	0.000280	-0.300039	0.7641
	Variance Equation			
C	7.48E-06	2.35E-06	3.188610	0.0014
RESID(-1)^2	0.043282	0.016114	2.685942	0.0072
RESID(-1)^2*(RESID(-1)<0)	0.064353	0.021734	2.960904	0.0031
GARCH(-1)	0.876060	0.025310	34.61273	0.0000
GED PARAMETER	1.368564	0.066763	20.49892	0.0000
R-squared	-0.000001	Mean dependent var		-9.28E-05
Adjusted R-squared	-0.000001	S.D. dependent var		0.012522
S.E. of regression	0.012522	Akaike info criterion		-6.054448
Sum squared resid	0.221869	Schwarz criterion		-6.032178
Log likelihood	4292.549	Hannan-Quinn criter.		-6.046128
Durbin-Watson stat	2.001524			

4.7. EGARCH Model

Table 6. Table of EGARCH test results

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.000107	0.000277	-0.385889	0.6996
	Variance Equation			
C(2)	-0.703731	0.180086	-3.907751	0.0001
C(3)	0.183826	0.034884	5.269660	0.0000
C(4)	-0.057740	0.017855	-3.233787	0.0012
C(5)	0.936158	0.019290	48.53149	0.0000
GED PARAMETER	1.362926	0.066434	20.51538	0.0000
R-squared	-0.000001	Mean dependent var		-9.28E-05
Adjusted R-squared	-0.000001	S.D. dependent var		0.012522
S.E. of regression	0.012522	Akaike info criterion		-6.051250
Sum squared resid	0.221870	Schwarz criterion		-6.028981
Log likelihood	4290.285	Hannan-Quinn criter.		-6.042930
Durbin-Watson stat	2.001523			

The EGARCH model is similar to the TGARCH model in that both measure the leverage effect of the variables.The full name of EGARCH is Exponential GARCH, or Exponential GARCH model. Unlike traditional GARCH models, the conditional expectation of the innovation variance of the EGARCH model takes the form of an exponential function, which allows for more flexibility in dealing with asymmetries in volatility.

The EGARCH model, short for Exponential GARCH, and the TGARCH model, also known as Threshold GARCH, are distinct classes of GARCH models employed for modeling the volatility of financial time series. The primary differentiating factor between these models lies in the treatment of volatility asymmetry. Overall, EGARCH focuses on capturing leverage effects and asymmetries in volatility, while TGARCH focuses on significant changes in volatility under specific conditions. The choice of which model to use usually depends on the understanding of the actual situation and the effectiveness of the model fit. [4].

The EGARCH model demonstrates that the asymmetric term (leverage term) deviates significantly from the original hypothesis at a 5% significance level, presenting its noteworthy

ISSN: 2692-7608 DOI: 10.6981/FEM.202406_5(6).0016

statistical relevance. This outcome suggests the presence of a leverage effect within the model, pointing to distinct impacts of positive and negative news on the CSI 300 index.

Table 7. EGARCH model residual ARCH effect test

Heteroskedasticity Test: ARCH			
F-statistic	0.133377	Prob. F(1,1296)	0.7150
Obs*R-squared	0.133569	Prob. Chi-Square(1)	0.7148

The fitted EGARCH model is tested for the residual ARCH effect, at which time the P-value accepts the original hypothesis at a significance level of 5%, indicating that there is no ARCH effect on the residuals, and useful information has been extracted by the established EGARCH model, and the model is established appropriately for subsequent prediction analysis.

4.8. Volatility Analysis

The volatility images fitted against the EGARCH model are presented below:

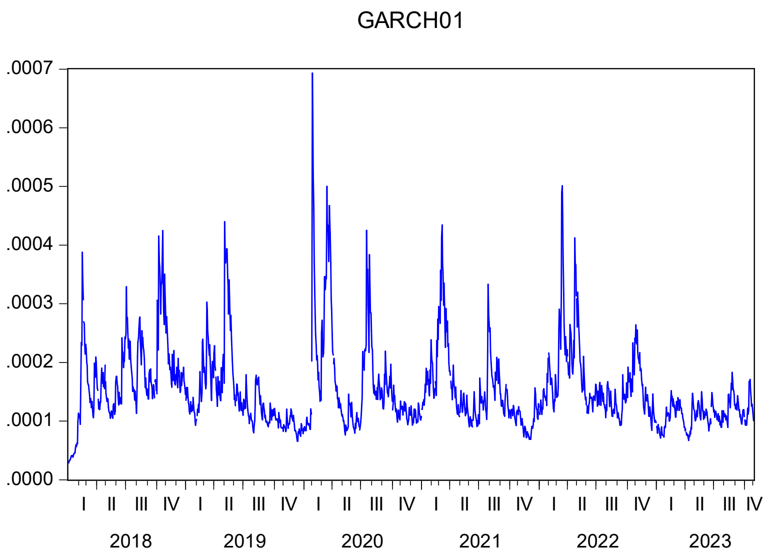

Figure 4. EGARCH model volatility image

4.9. GARCH-M Model

The GARCH-M model is a hybrid volatility model that integrates the conventional GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model with diverse other volatility models to effectively capture and model the volatility dynamics inherent in financial time series data. In a GARCH-M model, "M" stands for "mixed", implying that the model combines different forms of conditional heteroskedasticity models. Typically, such mixed models combine volatility models at different stages or under different conditions to better capture the complex volatility characteristics in financial markets. [5].

By combining these components, the GARCH-M model is able to adapt more flexibly to changes in volatility under different market conditions. Such models typically require more parameter estimation, but in some cases can provide more accurate volatility forecasts. The choice to use a GARCH-M model usually depends on the understanding of market characteristics and the need for model fitting.

The GARCH-M model integrates the M term within the mean equation. Examination of the results presented in the aforementioned table reveals that the M term's significance value is 0.8036, surpassing the threshold of 0.1. This implies acceptance of the null hypothesis at the 10% significance level, thereby indicating that the term lacks statistical significance.

ISSN: 2692-7608 DOI: 10.6981/FEM.202406_5(6).0016

Table 8. Table of results of GARCH-M test

Variable	Coefficient	Std. Error	z-Statistic	Prob.
@SQRT(GARCH)	-0.030949	0.124425	-0.248738	0.8036
C	0.000388	0.001420	0.273559	0.7844
	Variance Equation			
C	5.78E-06	1.96E-06	2.953844	0.0031
RESID(-1)^2	0.070171	0.015015	4.673341	0.0000
GARCH(-1)	0.893045	0.022004	40.58466	0.0000
GED PARAMETER	1.366659	0.063706	21.45253	0.0000
R-squared	-0.000550	Mean dependent var		-9.28E-05
Adjusted R-squared	-0.001258	S.D. dependent var		0.012522
S.E. of regression	0.012530	Akaike info criterion		-6.050048
Sum squared resid	0.221991	Schwarz criterion		-6.027779
Log likelihood	4289.434	Hannan-Quinn criter.		-6.041728
Durbin-Watson stat	2.001190			

5. Conclusion

Overall, stock market volatility forecasting based on GARCH-like models can provide investors and managers with more accurate market trends and risk assessments, leading to more informed investment decisions. By using different GARCH-like models, we can better understand the dynamics of the stock market and prepare for future market changes. In this study, we use TGARCH, EGARCH and GARCH-M models to forecast the volatility of CSI 300 index and conduct a comparative analysis of these three models. In the comparative analysis of the models, we found that each model has its unique advantages and shortcomings, but in general, the GARCH-M model has a greater advantage in predicting market volatility, and can capture the volatility characteristics of the market more accurately. Upon examination of the outcomes, it is evident that positive news exerts a greater influence on the stock market compared to negative news. Furthermore, based on the residual series plot of the model, it is apparent that the volatility of CSI 300 stock prices exhibits the presence of volatility clustering. This study provides investors with more accurate and effective market forecasting information, and provides a new theoretical and empirical basis for the volatility forecasting research of the stock market. It is hoped that the results of this research can provide reference for investors' decision-making and provide new insights for research in related fields.

References

[1] Engle, Ng, Rothschild. (1990) Asset pricing with a factor ARCH covariance structure: empirical estimates for treasury bills. Journal of Econometrics., (45): 277-299.

[2] Yu, K.X., Li, X.Y. (2023) Analysis of the volatility of China's stock market. Cooperative Economy and Technology., (23): 49-51.

[3] He, X.Q., Sun, Q.Y. (2003) Leverage and risk-return trade-offs in the Chinese stock market. Southern Economy., (09): 62-65.

[4] Zhang, P. (2020) A comparative study on the volatility of CSI 300 index and Hang Seng index based on GARCH family model. Thesis of Jiangxi University of Finance and Economics.

[5] Chen, X. (2023) Mean-variance effect analysis based on investor sentiment. Productivity Research., (11): 137-141.