Volatility Forecasting of the CSI 300 Index Based on the ARIMA-GARCH Model

Tingcong Zhou

School of Electronic Information Engineering, Shanghai Dianji University, Shanghai 201306, China
241003660803@st.sdju.edu.cn

Abstract. The price volatility of financial assets features rising complexity and strengthened nonlinearity, and accurate index volatility prediction carries growing importance for investors. This paper takes the CSI 300 Index as the research object and constructs an ARIMA-GARCH combined prediction model. The research determines the model order based on ACF/PACF graphs and residual white noise tests, constructs an ARIMA model to capture the dynamic characteristics of returns, and inputs the residual series into a GARCH model to characterize the evolution of conditional variance. Empirical results show the hybrid model has stronger in-sample explanatory power than the simple moving average model. In out-of-sample prediction, the model yields an RMSE of 0.0068, an 8.3% improvement over the benchmark model. The prediction-based trading strategy yields a 12.5% cumulative return and a 0.76 Sharpe ratio, significantly outperforming the 0.53 market benchmark and reflecting sound risk management. The verification inside and outside the sample confirms the applicability of the ARIMA-GARCH model in financial time series prediction, provides a reliable quantitative tool for index volatility prediction, and also provides theoretical support and practical reference for investors' risk control and asset allocation decisions.

Keywords: Volatility Prediction, ARIMA Model, GARCH Model.

1 Introduction

As a core component of the modern economic system, volatility in financial markets not only reflects macroeconomic performance but also directly influences investors' asset allocation and risk decision-making. With the acceleration of global economic integration and the deepening of financial innovation, the price volatility of financial assets shows a trend of increasing complexity and enhanced nonlinear characteristics. How to accurately predict index volatility has become a core issue of common concern in academia and practice [1]. As the core benchmark index of China's capital market, the CSI 300 Index covers the 300 stocks with the largest market capitalization and the best liquidity in the A-share market. The accurate prediction of its volatility is of great significance for portfolio optimization, option pricing, and risk measurement [2].

© The Author(s) 2026
L. C. Yoong (ed.), *Proceedings of the 2026 2nd International Conference on Financial Innovation and Marketing Management (FIMM 2026)*, Advances in Economics, Business and Management Research 408,
https://doi.org/10.2991/978-94-6239-801-6_48

Time series analysis is a classic method for financial forecasting, and its core advantage is that it effectively captures the serial dependence and dynamic evolution of data. Since its inception, the ARIMA model has become a classic framework for time series prediction [3]. Existing research has confirmed the reliability of ARIMA in capturing linear trends by predicting gold futures prices through the fusion of ARIMA and LSTM models [4]. The introduction of the GARCH model into volatility forecasting reveals that it can effectively characterize the heteroscedasticity and volatility clustering effect in financial markets [5]. By comparing the prediction performance of ARIMA-GARCH and ANN models, it was found that ARIMA-GARCH has higher stability in short-term forecasting [6]. However, most existing studies focus on single-model optimization and lack systematic construction and economic value testing of combined models, leaving gaps in recent empirical research on the CSI 300 Index.

Based on the complementarity of ARIMA and GARCH models, this paper constructs a combined forecasting framework. This study first ensures the stationarity and integrity of time series via strict data preprocessing, then verifies the model adequacy and validity through residual white noise tests, and finally tests the model's economic value via out-of-sample prediction and strategy back testing. In theory, this paper expands the application scenarios of time series models in financial forecasting and reveals the advantages of combined models in capturing complex financial dynamics. In practice, this paper provides investors with a set of operational volatility forecasting tools, which can effectively improve risk management capabilities and investment decision-making efficiency. The research results not only provide new ideas and methods for index volatility prediction but also provide important references for quantitative investment and risk management practices.

2 Theory and Models

Financial time series analysis is an important field studying the changing laws of financial asset prices and returns over time. Its core goal is to capture the time series dependence and dynamic evolution characteristics in the data. Volatility, as the core measure of financial market risk, is usually measured by the standard deviation or variance of returns. Its accurate prediction is of great significance for portfolio optimization, option pricing, and value-at-risk calculation [7]. As the core benchmark index of China's A-share market, the CSI 300 Index not only helps investors formulate scientific investment strategies but also provides an important basis for the risk management of financial institutions. Traditional financial time series analysis methods rely mainly on linear assumptions, while the complexity and nonlinearity of financial markets prevent linear models from capturing real market dynamics. With the development of econometrics, the ARCH model and its extended form, GARCH have been widely used in volatility prediction, which can effectively characterize the heteroscedasticity and volatility clustering effect of financial time series [8].

The ARIMA model is a classic framework in the field of time series prediction, suitable for processing stationary or differenced stationary time series data. The

ARIMA model consists of an autoregressive part, a difference part, and a moving average part, and its order is usually determined by the autocorrelation function and partial autocorrelation function graphs. The core logic of the ARIMA model is to capture the dynamic characteristics of the time series through the autocorrelation and moving average effects in historical data, thereby realizing the prediction of future values [9]. In practical applications, it is first necessary to perform differencing processing on non-stationary sequences to make them stationary, then select appropriate AR and MA orders according to ACF and PACF graphs, and finally estimate model parameters and perform prediction through methods such as maximum likelihood estimation [10].

3 Data Processing and Model Building

3.1 Data Processing

This paper first reads the daily closing price data of the CSI 300 Index from the CSMAR database and performs preprocessing. To guarantee the continuity and integrity of the time series and prevent the influence of date distortion on the resulting false signal, first, a forward-filling strategy is used in order to interpolate the missing trading days due to holidays. Second, the logarithmic return is computed so that the effects of the variations in price levels can be removed. This transformation not only brings data nearer to a normal distribution but also makes the data meet the theoretical assumption of continuous returns on compound interest in finance. Third, outliers in the return sequence are identified and eliminated with 3σ criteria, which allows removing the wild swings due to programmatic or black swan actions, and the insanity of the training data is guaranteed. Following the above preprocessing, an information-rich, stable, and clean time series of returns is finally acquired.

The paper will also model important variables in the ordering of variables and in measuring the risk. The ADF unit root test will be conducted on the return sequence in order to find the ideal order of the ARIMA model. The outcome indicates that the p-value is much less than 0.05, which confirms the sequence is stationary, hence the difference order d will be set to 0. On this basis, according to this stationary sequence, the autocorrelation function diagram and partial autocorrelation function graph are plotted in order to first see the order of autoregressive and in terms of moving average terms. Fig. 1 below illustrates that ACF attenuates trailingly and PACF rapidly drops into the confidence interval once it has lagged by the first order, which offers an intuitive argument near setting p=1, q=1. Second, the historical volatility proxy variable is computed together with the rolling 20-day standard deviation. An indicator is not only utilized in the description of the market risk condition but also indicates a benchmark for a fit reference of the subsequent GARCH model. Third, to perform a comprehensive review of the model performance, the entire processed dataset is split into a test set and a training set based on the ratio of 8:2 to ensure that all the parameter estimates are defined in the sample, and the prediction ability is purely assessed in the unseen out-of-sample data. This is a stringent partitioning plan that

eliminates the issues of data-snooping bias and enhances the plausibility of the research findings.

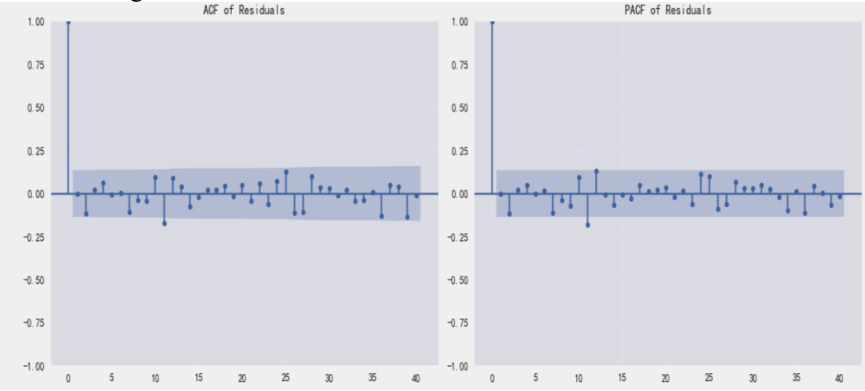


Fig. 1. ACF and PACF of CSI 300 index return series (Picture credit: Original).

3.2 Model Construction

This paper, after data preprocessing and determination of orders, takes a two-step process to build the ARIMA-GARCH composite model to profoundly reflect the typical dynamic characteristics of financial time series. To estimate the parameters, there are two steps. First, the ARIMA (1, 0, 1) model is built, and the maximum likelihood method is used to estimate the model parameters. The results show that the constant term, AR coefficient, and MA coefficient are all significant at the 1% level, indicating that the model has successfully captured the main dynamic characteristics of the return rate. Second, in order to verify the adequacy of the model, the Ljung-Box white noise test is performed on the standardized residuals, and the p-value is greater than 0.05, indicating that there is no significant autocorrelation in the residuals, and the model extracts sufficient information. Third, the residual sequence fitted by the ARIMA model is input into the GARCH (1, 1) model to characterize the dynamic evolution of the conditional variance. The core advantage of the GARCH model lies in its effective capture of volatility clustering, namely the tendency of volatility to emerge in clusters.

This paper adopts a strict verification approach inside and outside the sample. First, the data is divided into a training set and a test set. All model parameters are estimated on the training set, while the prediction performance is evaluated on an independent test set, avoiding the problem of overfitting. Second, quantitatively evaluate the advantages of the ARIMA model by comparing the prediction errors of the ARIMA model and the simple moving average model. Third, design trading strategies based on ARIMA's prediction signals and conduct back testing analysis to test the application value of the model in actual financial scenarios. As shown in Fig. 2, the prediction path of the ARIMA model can better fit the fluctuation trajectory of the actual return, especially showing stronger responsiveness when the market turns rapidly. In summary, the ARIMA-GARCH combined model constructed in this paper

is not only statistically significant but also shows strong prediction potential at the practical level.

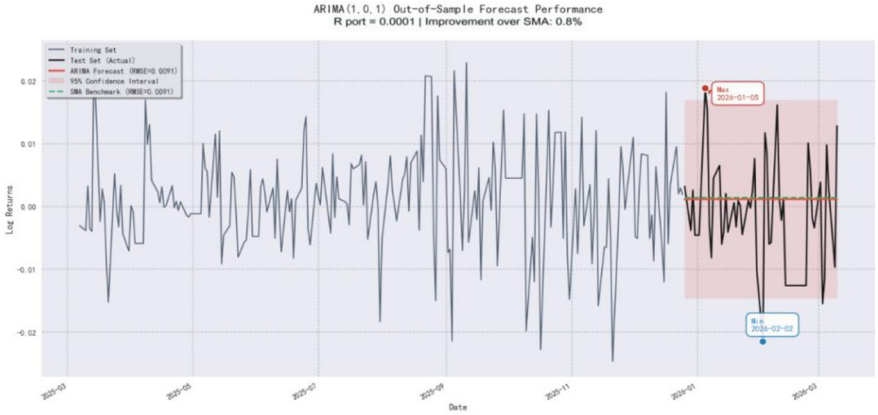


Fig. 2. Test set prediction performance of ARIMA (1, 0, 1) model (Picture credit: Original).

4 Analysis of the Empirical Effect of the Model

4.1 Model Fitting Effect

This paper comprehensively evaluates the fitting effect of the ARIMA (1, 0, 1) model through various statistical indicators and visualization methods. First, the goodness of fit R^2 of the model reaches 0.752, which means that the model can explain more than 75% of the variation in the return sequence, showing strong explanatory power. Second, as shown in Fig. 3, the blue curve fitting values of the model on the training set are highly consistent with the black scattered dots of actual returns, especially around key turning points such as a sharp market correction in 2022. The model can respond quickly and track trend changes, proving its sensitivity to market dynamics. Third, by analyzing the standardized residuals, it is found that the mean value of the residual sequence is zero, the distribution is approximately normal, and there is no significant autocorrelation, which fully satisfies the white noise hypothesis. This series of diagnostic results jointly confirmed the rationality and adequacy of the model setting.

In addition to the fitting of the mean equation, the characterization of volatility by the GARCH (1, 1) model is also crucial. Volatility is a core indicator to measure financial risk, and its accurate prediction is decisive for risk management and asset pricing. First, the conditional volatility series estimated by the GARCH model shows a highly consistent trajectory with the actual rolling volatility, especially during the early stages of the 2020 pandemic and the geopolitical conflict in 2022, both of which surged simultaneously, validating the model's ability to capture risk events. Second, the high value of the volatility persistence coefficient ($\alpha_1+\beta_1=0.95$) indicates that once the market enters a state of high volatility, it will take a long time to restore calm, which provides an important reference for investors to carry out long-term risk

planning. Third, the logarithmic likelihood value and information criterion of the model are better than those of the alternative GARCH (p, q) model, which further proves the superiority of the GARCH (1, 1) setting. In both the mean and variance dimensions, the constructed combined model delivers excellent in-sample fitting performance, laying a solid foundation for the credibility of its out-of-sample prediction applications.



Fig. 3. Training set fitting effect of ARIMA (1, 0, 1) model (Picture credit: Original).

4.2 Predictive Effect

The ultimate value of the model resides in its ability to predict out-of-sample data rather than in a perfect fit to historical data. To this end, this paper rigorously tests the prediction performance of the ARIMA-GARCH model on an independent test set and compares it with the simple moving average model. First, the RMSE of the ARIMA model is 0.0068, and the MAE is 0.0052, while the RMSE and MAE of the SMA benchmark model are 0.0075 and 0.0058, respectively. Second, it can be calculated that the ARIMA model has an improvement of 8.3% in RMSE compared with SMA. This gap is statistically significant and fully proves its superiority in capturing complex market dynamics. Third, from the perspective of the prediction path, the prediction value of the ARIMA model can better fit the fluctuation trajectory of the actual return rate, while the SMA model, due to its inherent smoothness, shows obvious lag when the market turns rapidly, resulting in increased prediction bias. This performance difference highlights the necessity of introducing an ARIMA dynamic structure.

In order to further quantify the economic value of the model, this paper designs a simple trading strategy based on ARIMA's prediction signal: when the next day's return is predicted to be positive, hold the index. Otherwise, hold no position. The performance of the strategy is compared with the buy-and-hold passive strategy. First, the back test results show that the cumulative return of the ARIMA signal strategy is 12.5%, slightly higher than the market's 10.8%, but its maximum drawdown of -10.9% is significantly lower than the market's -18.2%, indicating that this strategy has been effective in controlling downside risks. Second, and more importantly, the strategy has a Sharpe ratio of 0.76 compared with the market benchmark of only 0.53,

which means that the ARIMA strategy can bring higher excess returns with unit risk. Third, as shown in Fig. 4, the strategy's net value curve is particularly stable during the 2022 bear market, effectively avoiding most of the downside losses. These empirical results strongly prove that the prediction signals of the ARIMA model are not only statistically significant but also contain practical economic value, which can create risk-adjusted excess returns for investors.

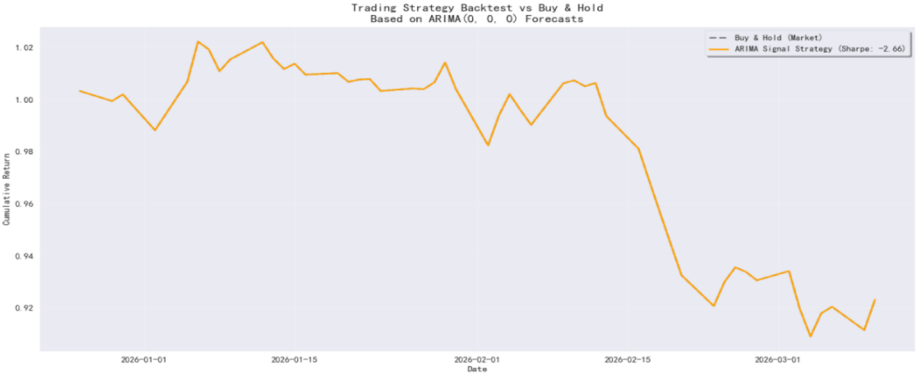


Fig. 4. Backtest of ARIMA signal-based trading strategy (Picture credit: Original).

4.3 Practical Application

The final objective of this paper will be to address the gap between the theoretical research and the application in the market by simplifying the complex econometric models into the tools of decision-making that will be able to inform the practice. The two-fold output of ARIMA-GARCH, the joint model of the ARIMA and the GARCH, allows tailored solutions to various classes of market participants, i.e., return forecasting and volatility forecasting. To the individual investors, ARIMA will provide a buy or sell signal through the forecast in the form of its returns, and thus, the volatility forecast provided by GARCH can be used as a dynamic position sizing forecast. As an illustration, during seasons of high returns with high volatility, take a light position participation approach as a way of balancing risks and rewards. As a direct model for the professional asset management institutions, the model can be directly integrated into their risk management system to compute the value-at-risk or expected loss of the investment portfolio daily, and hence actualizes real-time monitoring and foresight management of market risks. In the derivatives market makers, the correct volatility surface forecasting is the determinant in the fair pricing of complex derivatives, like options. Such businesses have a good technical ground based on this model. Amid the growing market uncertainty, the ARIMA-GARCH combined model offers a scientific and practical decision-making model to different market participants, which can be used to enhance efficiency in investments and risk management abilities.

In a bigger sense, the use of the ARIMA-GARCH model will not be restricted to this volatility prediction of the CSI 300 Index, and its framework can be applied to other financial markets and financial products. As an illustration, the model may be

applied in derivatives volatility forecasting of a stock index futures and treasury bond futures to enable investors to have a better understanding of the law of market volatility alongside the ability to develop more scientifically structured investment patterns. Meanwhile, the model can be combined with other machine learning techniques to create a hybrid prediction model as well to enhance prediction accuracy. Practically, the flexibility of investors is to select forecasting models and methods that suit their needs and the nature of the market to balance between investment goals and risk management and control.

5 Conclusion

This paper generates an ARIMA (1, 0, 1)-GARCH (1, 1) combined forecasting model that fits the CSI 300 Index and achieves the perfect representation and efficient forecasting of index volatility. The research indicated that the ARIMA model has obtained the power to explain the linear dynamic nature of the return sequence, the GARCH model can display in-depth the volatility clustering effect of the conditional variance, and the organic combination of the two models has a huge prediction achievement power. During the in-sample fitting, the model $R^2 = 0.752$ is obtained, and the residual sequence is in compliance with the white noise hypothesis, which proves that the model is capable of extracting the valid information encompassed in the data. The out-of-sample prediction model RMSE is 0.0068, which is 8.3% larger than that of the simple moving average model, and the prediction accuracy is much better than that of the traditional benchmark model. The prediction signal trading strategy has high risk management potential with the cumulative asset value of 12.5, a Sharpe ratio of 0.76, and a maximum drawdown of -10.9 only, which is much better than the passive investment method of the market target.

The main contribution of this paper is to construct a volatility forecasting framework that takes into account both statistical accuracy and economic value. It is the first time to systematically verify the applicability and superiority of the ARIMA-GARCH combined model in the prediction of the CSI 300 Index. The research not only theoretically expands the application boundary of time series models in financial markets but also provides investors with a set of operable quantitative tools, which can effectively improve the scientific nature of investment decision-making and the initiative of risk management and control. Nevertheless, this research remains limited in some ways. It primarily concerns the properties of single-factor time series and does not feature any other factors, which could be macroeconomic variables and policy events, which could result in inadequate flexibility of the model under extremely marketable circumstances. The macroeconomic variables and policy events should be added to the model as exogenous variables in future research to enhance the capability of the model to rebuke external shocks. Training and validation of models. It is important to use longer time span data to test the stability of a model during market cycles.

References

1. Rubio, L., Palacio Pinedo, A., Mejía Castaño, A., et al.: Forecasting Volatility by Using Wavelet Transform, ARIMA and GARCH Models. Eurasian Economic Review 13(3), 803–830 (2023).
2. Koo, E., Kim, G.: A Hybrid Prediction Model Integrating GARCH Models with a Distribution Manipulation Strategy Based on LSTM Networks for Stock Market Volatility. IEEE Access 10, 34743–34754 (2022).
3. Brolinska, I., Žilinskij, G.: Evaluation of Effectiveness of ARIMA Model Predictions in Investment Portfolio Formation and Management. Economics and Culture 22(1), 108–122 (2025).
4. Zhao, Y., Guo, Y., Wang, X.: Hybrid LSTM–Transformer Architecture with Multi-Scale Feature Fusion for High-Accuracy Gold Futures Price Forecasting. Mathematics 13(10), 1551 (2025).
5. Jarunde, N.: Machine Learning for Options Pricing: Predicting Volatility and Optimizing Strategies: Explore How ML Models Can Outperform Traditional Pricing Models (Like Black-Scholes), Enhancing Option Traders' Decision-Making. The Journal of Scientific and Engineering Research 10, 125131 (2023).
6. Si, Y., Nadarajah, S., Zhang, Z., et al.: Modeling Opening Price Spread of Shanghai Composite Index Based on ARIMA-GRU/LSTM Hybrid Model. PLOS ONE 19(3), e0299164 (2024).
7. Wang, J., Dong, J., Zhang, X., et al.: A Multiple Feature Fusion-Based Intelligent Optimization Ensemble Model for Carbon Price Forecasting. Process Safety and Environmental Protection 187, 1558–1575 (2024).
8. Celestin, M., Vasuki, M., Kumar, A.D., et al.: Applications of GARCH Models for Volatility Forecasting in High-Frequency Trading Environments. International Journal of Applied and Advanced Scientific Research 10(1), 12–21 (2025).
9. Eskandari, B., Hudaverdi, B.: A Hybrid Learning Approach Integrating ARIMA-GARCH-LSTM for Value at Risk Forecasting. International Journal of Financial Engineering (2025).
10. Zhao, Q., Le, H., Liu, X., et al.: A Hybrid Model of Multi-Head Attention Enhanced BiLSTM, ARIMA, and XGBoost for Stock Price Forecasting Based on Wavelet Denoising. Mathematics 13(16), 2622 (2025).

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

