Financial Engineering and Risk Management (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/ferm.2025.080124
ISSN 2523-2576 Vol. 8 Num. 1

Stock Market Price Prediction Based on GARCH-BO-LSTM

Chen Ting*

School of Economics, Shanghai University, Shanghai, China
**Corresponding author: 1738876167@qq.com*

Keywords: Stock Prediction, Long Short-Term Memory Network (LSTM), Bayesian Optimization, Volatility Modeling

Abstract: Current stock market price prediction research mainly uses machine and deep learning for historical data, sentiment, and macro - indicators to boost accuracy. Prediction is crucial for investors, risk management, and market stability. LSTM has strengths like handling long - term dependencies in price sequences but has long training times and high resource use. This paper gets index volatility and return data via GARCH, then uses Bayesian optimization on LSTM to enhance prediction. It validates the model by comparing four metrics with others. Using the CSI 1000 Index, the Bayesian - optimized LSTM reduces RMSE by 0.169% compared to the basic LSTM.

1. Introduction

Stock prediction means forecasting future stock price trends by analyzing historical stock market data. It has long been a key research area in financial markets. There are mainly two ways for stock price prediction: linear and nonlinear prediction [1]. Common linear prediction models like the ARIMA and GARCH models have been vital in financial stock prediction development. But because of the high noise and nonlinear nature of financial time series data, relying only on linear models has reduced stock price forecast accuracy [2]. Nonlinear models have come about to overcome these drawbacks, enhancing prediction accuracy based on linear models. Among them, machine learning neural networks are being more and more used in stock prediction.

Long Short-Term Memory (LSTM), a classic time-series prediction network, shows great performance in stock price forecasting [3]. Sun et al. (2022) used the LSTM network to extract data's temporal features and predicted the Shanghai Stock Exchange Index 000001, getting better results than traditional statistical methods. In recent years, scholars have been improving LSTM. Bao Zhenshan et al. (2020) upgraded LSTM's output module and verified the model with CSI 500 Index's intraday minute-level data, finding the improved model superior to the basic LSTM [4]. Hu Yuwen (2021) proposed a stock price prediction model combining PCA, LASSO and LSTM neural networks, enhancing prediction accuracy. Han Ying et al. (2023) integrated broad learning and deep learning based on LSTM, introduced Complementary Ensemble Empirical Mode Decomposition (CEEMD) for noise reduction to deal with stock sequences' non-stationary features, and proposed the CEEMD-LSTM-BLS (C-L-B) stock prediction model [5].

In recent years, LSTM has many diverse extended forms for complex financial time - series data.

But in practice, LSTM has issues like numerous parameters (manual tuning reduces efficiency) and difficulty in capturing nonlinear data relationships, limiting its financial forecasting use. To address these, this paper studies the GARCH - BO - LSTM model. The GARCH model captures financial time - series volatility well, providing useful features for LSTM. Bayesian Optimization (BO) automatically tunes LSTM's hyperparameters, enhancing training efficiency and prediction accuracy [6]. By combining these, the paper aims to overcome traditional LSTM's limitations in financial forecasting, offering a more accurate, efficient, and practical method for stock price prediction, with the hope of achieving significant results in financial research and practice to drive financial forecasting technology development [7].

2. Model Construction and Trading Strategy

2.1 Principles of the GARCH Model

The GARCH model is a statistical model used to estimate and predict the volatility of time series data [8-10]. The GARCH model, an extension of the ARCH model, effectively addresses the heteroskedasticity issue in financial time series residuals, which the ARCH model cannot resolve. Its mathematical expression is as follows:： In the equation, σ_t represents the conditional variance, and u_t~N(0,1) is a standard normal random variable, where σ_t and u_t are independent of each other. Additionally, the coefficients A_i and G_i must satisfy the following constraints:

$$y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \alpha y_{t-1} + \varepsilon_t + \sum_{i=1}^{q} \beta_i \varepsilon_{t-1} \tag{1}$$

$$\varepsilon_t = u_t \sigma_t \tag{2}$$

$$\sigma^2 = k + \sum_{i=1}^{p} A_i \varepsilon_{t-1}^2 + \sum_{i=1}^{q} G_i \sigma_{t-i}^2 \quad k > 0, A_i \geq 0, G_i \geq 0 \tag{3}$$

$$\sum_{i=1}^{p} A_i + \sum_{i=1}^{q} G_i < 1 \tag{4}$$

2.2 Bayesian Optimization

The core steps of Bayesian optimization are building a Gaussian surrogate model and optimizing the acquisition function.

2.2.1 Constructing Gaussian Surrogate Model

The Gaussian process, defined by a mean and covariance function, constructs sample set D = {(X, λ)}, where X is the training set and λ is a mapping function. Closely located input points x have similar mapping function values, enhancing prediction accuracy when training and test samples are near. The covariance function, typically the squared exponential one, gauges point similarity. Its formula contains l, the characteristic length scale, which dictates function smoothness.

$$k(x, x') = exp(-\frac{(x-x')^2}{2l^2}) \tag{5}$$

$$\lambda \sim N(0, k(x, x')) \tag{6}$$

2.2.2 Optimizing the Acquisition Function

The Gaussian surrogate model links hyperparameters to model performance.The acquisition function identifies the next optimal sampling point. Often, the Expected Improvement function E(x) is used, calculating the expected improvement over the current best to efficiently locate the optimum while balancing search cost and performance.

$$E(x)=\begin{cases}(\alpha(x)-y(x^{+})-\xi)*\varphi\left(\frac{\alpha(x)-y(x^{+})-\xi}{\tau(x)}\right)+\sigma(x)*\phi\left(\frac{\alpha(x)-y(x^{+})-\xi}{\tau(x)}\right), \; if\ \tau(x)>0 \\ 0, if\ \tau(x)=0\end{cases} \quad (7)$$

Where, y(x) represents the expected value at the optimal training sample x, $\alpha(x)$ is the mean of the current training sample x, $\tau(x)$ is the variance of x,and ξ is a balancing parameter. φ and ϕ are the probability density function (PDF) and cumulative distribution function (CDF) of the standard normal distribution, respectively.

2.3 LSTM

The LSTM network is an improved RNN version that addresses the vanishing gradient problem via a self-looping mechanism. The internal cell unit mainly consists of three key gate units: the forget gate, the input gate, and the output gate.Its structure is illustrated in Fig. 1.

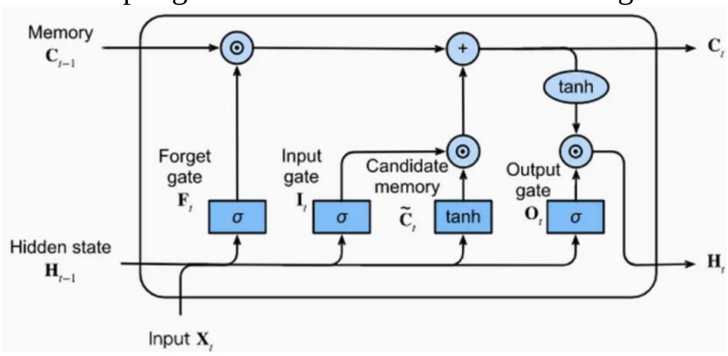

Fig. 1. LSTM Structure Diagram.

2.3.1 Forget Gate

The specific calculation formula is as follows, f_t is the output state of the forget gate, σ is the Sigmoid activation function, W_f is the weight matrix , b_f is the bias term, W_f and b_f determine the extent to which information is retained or discarded.

$$f_t=\sigma*(W_f*[h_{t-1},X_t]+b_f) \quad (8)$$

2.3.2 Input Gate

The input gate initially updates the information via the Sigmoid function, as depicted in Equation (9). Subsequently, it quantifies the extent of new information to be integrated, as illustrated in Equation (10). Ultimately, by combining the two components, the input gate discards the filtered - out info from the forget gate and integrates the processed info into the long - term state, as shown in Equation (11).

$$i_t=\sigma*(W_i*[h_{t-1},X_t]+b_i) \quad (9)$$

$$c_t=\sigma*(W_c*[h_{t-1},X_t]+b_c) \quad (10)$$

$$C_t=i_t*c_t+f_t*C_{t-1} \quad (11)$$

2.3.3 Output Gate

The output ate specific calculation formulas are as follows,W_o and $b_o$represent the weight matrix and bias term of the output gate, respectively, and h_t is the output value of the current cell.

$$o_t = \sigma * (W_o * [h_{t-1}, X_t] + b_o) \tag{12}$$

$$h_t = o_t * tanh^{-1}(c_t) \tag{13}$$

3. Empirical Analysis

3.1 Data Source

This paper conducts empirical analysis on the CSI 1000, SZCI, and SHCI. The sample spans from January 2016 to December 2024, with data from https://cn.investing.com/. The CSI 1000 has 1,000 stocks outside the CSI 800, representing small-cap and liquid stocks. The SZCI is a Shenzhen Stock Exchange index with 500 representative listed firms, calculated by weighted average of circulating shares. The SHCI is an early and authoritative Chinese stock index reflecting all listed stocks on the Shanghai Stock Exchange. Testing multiple indices ensures the model's predictive effectiveness.

Price trend graphs in Figs. 2-4 indicate significant price fluctuations for each index, with the SZCI having the largest variation amplitude.

Fig. 2. Price Trend Chart of the CSI 1000.

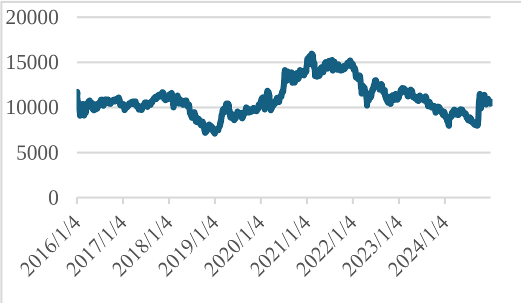

Fig. 3. Price Trend Chart of the SZCI.

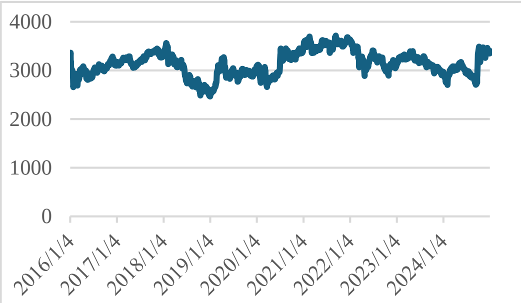

Fig. 2. Price Trend Chart of the SHCI.

3.2 Data Processing

In this paper, to smooth the optimization process, the data is normalized using a transformation function. Data normalization refers to scaling the data proportionally so that it falls within a specific small interval, typically between 0 and 1 or between -1 and 1. The purpose is to eliminate the influence of data dimensions, making data with different features comparable. The transformation formula is as follows, where x_{max} and x_{min} represent the maximum and minimum values within the component, respectively.

$$x = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{14}$$

3.3 Model Evaluation Metrics

This paper uses RMSE, MSE, MAE, and MAPE to evaluate model prediction results. MSE is the average of squared differences between predicted and actual values. RMSE, the square root of MSE, represents the standard deviation of these differences. MAE is the average of absolute differences, and MAPE is the average of absolute percentage differences. Generally, lower values of these four metrics indicate better model prediction performance. The calculation formulas for these metrics are shown in Table 1.

Table 1 Calculation Formulas for Metrics.

RMSE	MSE	MAE	MAPE
$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}$	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}$	$\frac{1}{n}\sum_{i=1}^{n}\lvert y_i-\hat{y}_i\rvert$	$\frac{100\%}{n}\sum_{i=1}^{n}\lvert y_i-\hat{y}_i\rvert$

3.4 Model Construction and Prediction

3.4.1 GARCH Modeling

First, the GARCH model is used to model the return rates of the stock closing prices, resulting in the return rate curves. Figs. 5-7 show the price volatility of the three indices

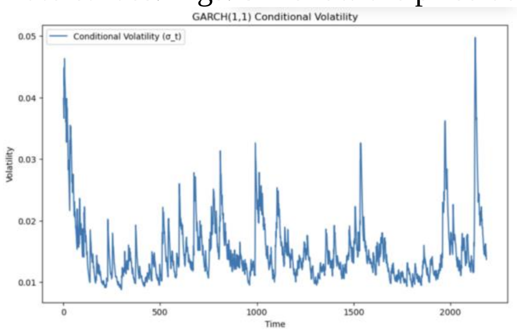

Fig. 5. Volatility of the CSI 1000.

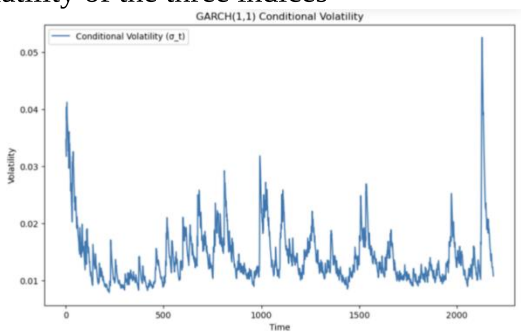

Fig. 3. Volatility of the SZCI.

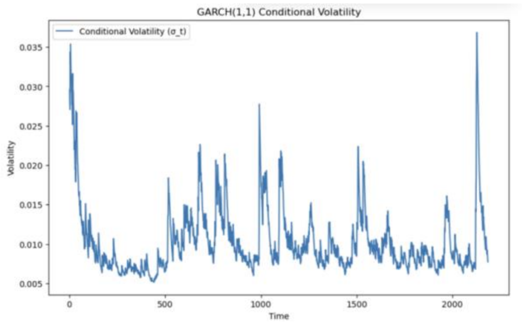

Fig. 4. Volatility of the SHCI.

Based on the descriptive statistics and GARCH-modeled volatility graphs, the volatility of the three indices stays relatively high, with the CSI 1000 Index being the most volatile. In a regression model, "Coef" represents the coefficients (weights or slopes), showing how much the independent variables affect the dependent variable. "Stderr" is the standard error of these coefficients, which gauges the precision of coefficient estimates; a lower Stderr means more accurate estimates. "T" is the ratio of the coefficient to its standard error, used to check if the coefficient differs significantly from zero. A larger absolute t-value implies a more substantial influence of the independent variable on the dependent one. "P" is for hypothesis testing, usually to see if the coefficient is significantly different from zero; a smaller p-value gives stronger grounds to reject the null hypothesis that the coefficient equals zero.

Taking the CSI 1000 Index as an illustration: The coefficient of omega is 8.026e-06, accompanied by a standard error of 1.280e-09, a t-value of 6270.295, and a p-value of 0.000. Among the three indices, this shows the CSI 1000 Index has the largest coefficient, the smallest standard error, and the highest t-value. The alpha coefficient is 0.0992, with a standard error of 3.632e-02, a t-value of 2.731, and a p-value of 6.323e-03. Since the alpha coefficient is positive, it means past errors positively affect current volatility. The beta coefficient is 0.8687, having a standard error of 2.867e-02, a t-value of 30.298, and a p-value of 1.202e-201. The positive beta coefficient implies that past volatility impacts current volatility. In general, the p-values of omega, alpha, and beta for the CSI 1000 Index are all low, signifying that the results are statistically significant.

3.4.2 Building the LSTM Model and Performing Bayesian Optimization

After conducting GARCH modeling to get the conditional volatility and logarithmic returns of the three indices, an LSTM model is constructed based on that, and then Bayesian Optimization is carried out. In this paper, 75% of the data is chosen as the training set, while 25% serves as the test set. Once the LSTM modeling results are obtained, Bayesian Optimization is applied to the model. This process uses a Gaussian process to figure out the optimal learning rate, dropout rate, number of LSTM neurons, and validation loss for each index.

After performing Bayesian Optimization, the model training begins, with RMSE, MSE, MAE, and MAPE selected as the evaluation metrics for the training set. To more clearly demonstrate the model's predictive performance, Figs. 8-10 show the predicted values versus the true values for the three indices.

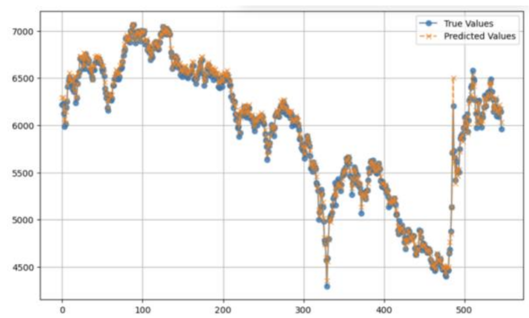

Fig. 5. Prediction Results for the CSI 1000.

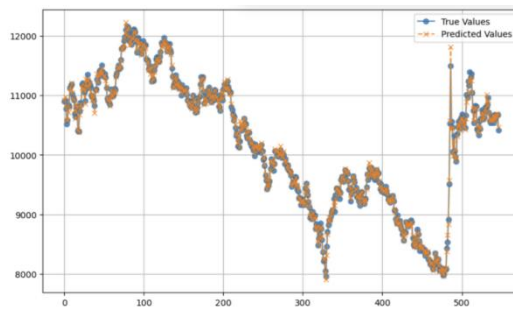

Fig. 6. Prediction Results for the SZCI.

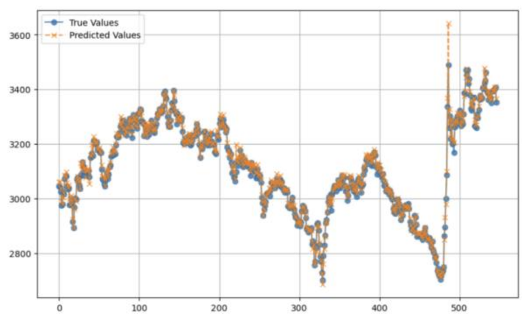

Fig. 7. Prediction Results for the SHCI.

4. Result Comparison and Analysis

To check if the GARCH-BO-LSTM model in this paper has better predictive ability, this study compares its results with those of the BPNN, XGBOOST, RNN, SVR models and the LSTM model without Bayesian Optimization. Findings reveal that the Gaussian-process-optimized LSTM model's predictions for the three chosen indices outdo other models in all four metrics (RMSE, MSE, MAE, MAPE), proving the GARCH-BO-LSTM model's effectiveness.

From the GARCH-BO-LSTM model's predictions for different indices, the CSI 1000 has the best performance, with the lowest RMSE, MSE, MAE and MAPE values. When comparing different models for the three indices, the CSI 1000 and SZCI show big performance differences among models. For the SHCI, except for the SVR model, the performance differences among models are smaller. Compared with other indices, the SHCI has the smallest mean, standard deviation and range, which is due to the authority and comprehensiveness of its data.

5. Conclusion

In financial investment, investors highly care about accurately predicting stock price trends. Existing single deep learning prediction models can forecast stock prices to a degree, but have limitations like non-trivial prediction errors and overfitting risks. This paper first builds a GARCH model to catch data's nonlinear relationships, boosting the LSTM model's predictive power. Then, it applies the Bayesian Optimization algorithm to the LSTM model. Comparisons with other models show that after Gaussian process optimization, the LSTM model effectively solves problems like lag and poor fitting in the basic LSTM, thus improving prediction accuracy. With technological progress, future stock market predictions require more advanced models to enhance precision and efficiency, offering stronger support for market analysis and investment decisions.

References

[1] Huang Houju, Li Bo. Stock Price Prediction Based on VMD-CSSA-LSTM Combined Model [J]. Journal of Nanjing University of Information Science & Technology, 2024, 16(03): 332-340.

[2] Hu Yuwen. Stock Prediction Based on Optimized LSTM Model [J]. Computer Science, 2021, 48(S1): 151-157.

[3] Sun Y, Sun Q, Zhu S. Prediction of Shanghai Stock Index Based on Investor Sentiment and CNN-LSTM Model [J]. Journal of Systems Science and Information, 2022, 10(06): 620-632.

[4] Bao Zhenshan, Guo Junnan, Xie Yuan, et al. Stock Price Rise and Fall Prediction Model Based on LSTM-GA [J]. Computer Science, 2020, 47(S1): 467-473.

[5] Han Ying, Zhang Dong, Sun Kaiqiang, et al. Research on a New Stock Prediction Model Combining Long Short-Term Memory Network and Broad Learning [J]. Operations Research and Management Science, 2023, 32(08): 187-192.

[6] Zhao Chengyang. Parallelized Generalized EGO Algorithm and Its Applications [D]. Central China Normal University, 2022.

[7] Zhang Xiuli, Liu Yingfen. Research on the Dynamic Correlation and Spillover Effects Between Shanghai Crude Oil Futures and Domestic Agricultural Futures [J]. China Securities and Futures, 2024, (05): 23-35.

[8] Zhang Chen, Liu Yulin, Li Hua, et al. A Parallel Traffic Real-Time Prediction Model for Cloud-Network Integration [J]. Computer and Digital Engineering, 2024, 52(11): 3350-3355.

[9] Xu Wei, Luo Jianping, Li Xia, et al. Multi-Task Model Based on Correlation Learning and Its Applications [J]. Journal of Shenzhen University (Science and Engineering Edition), 2013, 40(04): 494-503.

[10] Qin Kun, Liu Liqun, Wu Qingfeng, et al. Short-Term Electricity Price Prediction Based on Improved CEEMDAN-BO-LSTM [J]. Journal of Shaanxi University of Science & Technology, 2025, 43(01): 169-176.