Research on a Quantitative Financial Risk Prediction Model Based on Machine Learning

Jialu Yuan *

KennethP. Dietrich School of Arts and Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA

* Corresponding Author Email: jiy163@pitt.edu

Abstract. To address the challenges faced by traditional financial risk prediction models in balancing volatility characterization and nonlinear dependency capture, this study proposes an Att-LSTM-GARCH hybrid model, integrating the volatility extraction capabilities of GARCH, temporal learning of LSTM, and key feature enhancement capabilities of the attention mechanism. Using daily data (2,428 samples) from the CSI 300 and CSI 500 indices from 2015 to 2024, prediction experiments were conducted focusing on volatility and 95% confidence level VaR. Results show that the MAE for one-day volatility prediction reaches 0.012, which is 32.2%, 22.5%, and 14.3% lower than those of GARCH, LSTM, and LSTM-GARCH, respectively. The one-day VaR prediction accuracy reaches 92.3%, outperforming the comparison model by up to 4.1%. Ablation experiments demonstrate that the GARCH and attention mechanisms can reduce the MAE by 25% and 33.3%, respectively. In volatile bull and bear markets, the MAE fluctuates by only 16.7%, demonstrating optimal stability under noise. This model provides financial institutions with a high-precision risk prediction tool, enhancing risk management.

Keywords: Quantitative Financial Risk; Volatility Prediction; Value at Risk (VaR); Att-LSTM-GARCH Model; Attention Mechanism.

1. Introduction

The high volatility and nonlinear characteristics of financial markets have made quantitative financial risk prediction a core research topic in financial engineering. Quantitative financial risk primarily encompasses market risk, credit risk, and liquidity risk. Within market risk, volatility and Value at Risk (VaR) are key indicators for measuring portfolio risk. In extreme market conditions, such as the global stock market crash in 2020 and the drastic interest rate adjustments in 2022, accurate risk prediction can help financial institutions avoid significant losses and reduce the probability of systemic risk transmission, providing important technical support for maintaining financial market stability. Traditional quantitative financial risk prediction models, exemplified by the GARCH family of models, can capture the volatility clustering of financial data, but are limited by linear assumptions and fixed parameter structures, making them inadequate for capturing the complex long-term dependencies and nonlinear volatility characteristics of financial time series [1]. Monte Carlo simulation methods, while capable of handling multidimensional risks, rely on strict distributional assumptions and exhibit high computational complexity, making them less adaptable for short-term, high-frequency risk prediction. With the surge in financial market data and the expansion of data dimensions, the limitations of traditional models in predictive accuracy and efficiency have become increasingly apparent [2]. In recent years, machine learning technology, leveraging its powerful capabilities for nonlinear fitting and learning time series features, has achieved breakthroughs in its application in the financial sector. Recurrent neural networks (RNNs) and long short-term memory networks (LSTMs) can effectively capture the dynamic dependencies of time series, outperforming traditional models in tasks such as stock price forecasting and volatility estimation [3]. However, existing machine learning-based risk prediction models often focus on optimizing a single network structure, failing to fully integrate the advantages of traditional financial models in capturing volatility characteristics [4]. Furthermore, they lack the targeted extraction of features at key time nodes, resulting in insufficient stability in predicting extreme risk events. This study proposes the innovative Att-LSTM-GARCH algorithm. By integrating the volatility extraction capabilities of the GARCH model, the long-term dependency capture capabilities of the LSTM, and the key feature enhancement capabilities of the attention mechanism, it constructs a multi-module collaborative quantitative financial risk prediction framework [5]. Using daily data from the CSI 300 and CSI 500 indices as samples, the study conducted experiments on volatility and VaR prediction tasks, validating the algorithm's advantages in prediction accuracy, stability, and robustness. The theoretical value of this study lies in enriching the technical path for integrating machine learning with traditional financial models [6]. Its practical significance lies in providing financial institutions with high-precision and reliable risk prediction tools, helping to improve their financial risk management.

2. Related Theoretical and Technical Foundations

2.1. Core Indicators of Quantitative Financial Risk

2.1.1 Volatility

Volatility reflects the fluctuations in financial asset prices over a given period and is a fundamental indicator for measuring market risk. Historical volatility is calculated using asset returns, while implied volatility is derived from option prices [7]. Building on the traditional historical volatility calculation, this study considers the differential impact of returns on volatility over different periods and proposes a weighted historical volatility calculation formula.

$$\sigma_w = \sqrt{\frac{\sum_{i=1}^{n} w_i (r_i - \bar{r})^2}{n-1}} \quad (1)$$

Where σ_w is the weighted historical volatility, w_i is the weight of the return in period i (exponentially decaying over time, with more recent data given higher weight), r_i is the asset return in period i, $\bar{r}$ is the mean return, and n is the number of samples in the calculation period. This formula more accurately captures the impact of recent market volatility on overall risk, providing more realistic baseline data for short-term risk forecasting.

2.1.2 Value at Risk (VaR)

VaR is defined as the maximum possible loss a financial asset can sustain within a certain confidence level α and holding period t. Parametric methods are commonly used to calculate VaR. This study optimizes traditional parametric methods by taking into account the peaked and fat-tailed nature of financial data, proposing an improved formula for calculating VaR based on the t-distribution.

$$\mathrm{VaR}_{\alpha,t} = -P_0\left(\mu_t + z_{\alpha,k}\sigma_t\sqrt{t}\right) \quad (2)$$

In the formula, P_0 is the initial value of the asset, μ_t is the expected return over the holding period t, $z_{\alpha,k}$ is the quantile of the t distribution with k degrees of freedom at confidence level α (determined by fitting sample data), and σ_t is the instantaneous volatility of the asset's return. This formula better reflects the distribution characteristics of financial data and improves the accuracy of VaR calculations.

2.1.3 Expected Loss (ES)

Expected loss (ES) is the average loss an asset may suffer in the event of an extreme VaR exceedance. It compensates for the deficiency of VaR in reflecting the extent of extreme losses. This study derives the formula for calculating ES based on the improved VaR calculation method described above:

$$\mathrm{ES}_{\alpha,t} = -\frac{1}{1-\alpha}\int_{-\infty}^{-\mathrm{VaR}_{\alpha,t}/P_0} x f(x)\,dx \quad (3)$$

Where $f(x)$ is the probability density function of the asset's return over a holding period of t (based on the t distribution), and x is the return variable [8].

2.2. Traditional Financial Risk Prediction Models

Traditional financial risk prediction models provide a classic framework for risk analysis, but they have limitations in complex market environments.

2.2.1 GARCH Model and Its Extensions

The GARCH model characterizes the volatility clustering of financial data through ARCH terms (lagged squared residuals) and GARCH terms (lagged conditional variance). Its core formula is: $\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$. To enhance the model's ability to capture long-term volatility trends, this study proposes an improved, long-memory GARCH model extension that introduces a fractional differencing term:

$$\sigma_t^2 = \omega + (1-\phi L)^d \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \quad (4)$$

Where L is the lag operator, d is the fractional difference parameter $(0 < d < 1)$, ϕ is the difference coefficient.

2.2.2 Monte Carlo Simulation

Monte Carlo simulation predicts risk by simulating asset price trends using a large number of random events. This requires first determining the stochastic process underlying asset prices (such as geometric Brownian motion). In traditional simulations, random numbers are generated based on the standard normal distribution [9]. This study considers the time-varying nature of market fluctuations and proposes an improved Monte Carlo simulation formula based on time-varying volatility to generate asset price paths:

$$S_t = S_{t-1} \exp\left(\left(\mu - \frac{1}{2}\sigma_{t-1}^2\right)\Delta t + \sigma_{t-1}\sqrt{\Delta t}\xi_t\right) \quad (5)$$

Here, S_t and S_{t-1} are the asset prices in periods t and $t-1$, respectively; μ is the expected rate of return; σ_{t-1} is the time-varying volatility in period $t-1$ (calculated using the GARCH model); Δt is the time interval; and ξ_t is a standard normal random variable.

2.3. Machine Learning Technologies

Machine learning, with its powerful capabilities in nonlinear fitting and learning time series features, provides a new approach to financial risk prediction.

2.3.1 Recurrent Neural Networks (RNNs) and Long Short-Term Memory Networks (LSTMs)

RNNs process sequential data through a recurrent structure, but they are subject to vanishing or exploding gradients. LSTMs address this limitation through a gating mechanism consisting of input, forget, and output gates [10]. They effectively capture long-term dependencies, and their cell state updates and output calculation mechanisms align with the dynamic characteristics of financial time series, making them suitable for risk prediction tasks.

2.3.2 Attention Mechanism

The attention mechanism weights input features to highlight the impact of key information on the output. In financial risk prediction, market information at different time points significantly impacts risk. Introducing the attention mechanism can enhance the role of features from key periods and optimize model prediction performance.

3. Att-LSTM-GARCH Model

3.1. Algorithm Design Concepts and Principles

The Att-LSTM-GARCH model builds a synergistic framework to address the core shortcomings of traditional models: While traditional GARCH models can characterize volatility clustering, they are limited by their linearity assumptions and cannot capture nonlinear dependencies. While pure LSTM models can learn time series features, they ignore the volatility specificity of financial data. Existing hybrid models lack a targeted focus on key time nodes, resulting in inaccurate predictions of extreme risks [11]. The model utilizes three technical advantages—GARCH's volatility characterization capabilities, LSTM's ability to capture long-term dependencies, and the attention mechanism's feature selection capabilities—through a three-step synergy of "volatility extraction - time series learning - key reinforcement." This combines the advantages of three technologies: GARCH's volatility characterization capabilities, LSTM's ability to capture long-term dependencies, and the attention mechanism's feature selection capabilities. This achieves a dual improvement in risk prediction accuracy and robustness.

3.1.1 Integrating the GARCH Model's Ability to Characterize Financial Data Volatility

The GARCH model accurately describes the clustering characteristic of financial data, where "high volatility is followed by high volatility," through the conditional variance equation. It is a classic tool for quantifying volatility characteristics. To adapt to the feature input requirements of subsequent deep learning modules, the model improves the traditional GARCH conditional variance formula and introduces a volatility trend coefficient to enhance the capture of long-term volatility trends. The improved formula is as follows:

$$\sigma_t^2 = \omega + \lambda\sigma_{t-1}^2 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{6}$$

Here, σ_t^2 is the conditional variance at time t, ω is a constant term, and λ is the volatility trend coefficient (ranging from 0.1 to 0.3, determined through optimization on the validation set), which is used to enhance the persistence of the impact of previous volatility on the current period. α_i and β_j are the ARCH and GARCH coefficients, respectively, satisfying $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j + \lambda < 1$ to ensure variance stability.

3.1.2 Introducing LSTM Networks to Capture Long-Term Dependencies in Financial Time Series

LSTM networks address the vanishing gradient problem of traditional RNNs through a gating mechanism consisting of input, forget, and output gates [12]. They are particularly well-suited for handling long-term dependencies in financial time series. The model designs an improved LSTM unit, incorporates the fluctuation characteristics of the GARCH module output into the cell state update, and constructs a fluctuation-time series interaction mechanism. The core update formula is:

$$\begin{aligned} c_t &= f_t \odot c_{t-1} + i_t \odot \tanh\left(W_c[h_{t-1}, x_t, \sigma_{t-1}^2] + b_c\right) \\ h_t &= o_t \odot \tanh\left(c_t\right) \end{aligned} \tag{7}$$

Wherec_t is the cell state at time t, f_t, i_t, o_t are the output values of the forget gate, input gate, and output gate, respectively (calculated using the sigmoid function), h_{t-1} is the hidden state at time $\mathrm{t}-1$, x_t is the raw feature at time t, σ_{t-1}^2 is the conditional variance at time $\mathrm{t}-1$ (from the GARCH module), and W_c and b_c are the weight matrix and bias term, respectively.

3.1.3 Introducing an Attention Mechanism to Emphasize the Impact of Key Time Nodes

In financial markets, the characteristics of the time at which extreme events (such as policy adjustments and black swan events) occur are crucial for risk prediction. Traditional models assign equal weight to all time nodes, diluting key information. The model introduces a temporal attention mechanism to strengthen the influence of key nodes on the prediction results by calculating the attention weights of features at each moment. The weight calculation and feature fusion formulas are as follows:

$$e_t = \tanh(W_a h_t + b_a)$$
$$\alpha_t = \frac{\exp(e_t^T u_a)}{\sum_{k=1}^{T} \exp(e_k^T u_a)} \quad (8)$$
$$h_{att} = \sum_{t=1}^{T} \alpha_t h_t$$

Where e_t is the attention score of the feature at time t, W_a, b_a, u_a are the attention mechanism parameters, α_t is the normalized attention weight, and h_{att} is the weighted fusion feature.

3.2. Algorithm Architecture

3.2.1 Data Preprocessing Module (Normalization, Stationarity Test, and Sequence Construction)

Raw financial data exhibits dimensionality differences and non-stationarity, requiring preprocessing to ensure model input validity. First, Min-Max normalization is used to eliminate dimensionality effects. The formula is:

$$x_t^* = \frac{x_t - \min(X)}{\max(X) - \min(X)} \quad (9)$$

Here, x_t^* is the normalized value, and X is the original feature sequence. Next, the data is verified for stationarity using the ADF test. If not, first-order differencing is performed: $\Delta x_t = x_t - x_{t-1}$. Finally, a time series sample is constructed according to the formula "input window length T - forecast step length K." The input window T is determined based on the data frequency (T = 60 for daily data, i.e., a two-quarter trading cycle), and the forecast step length K is set to 1, 3, or 5 days to cover short-term risk prediction requirements. This ultimately results in an input sample with the dimension $(N - T - K + 1, T, F)$ (where N is the total number of samples and F is the feature dimension).

3.2.2 GARCH Volatility Extraction Module (Calculating the Conditional Variance Series)

This module takes the preprocessed yield series as input and calculates the conditional variance series using the improved GARCH formula described in Section 3.1.1. The specific process is as follows: First, the GARCH model parameters ($\omega, \lambda, \alpha_i, \beta_j$) are solved through maximum likelihood estimation [13]. The conditional variance σ_t^2 is then iteratively calculated at each time step. To ensure time synchronization with the subsequent LSTM module, the conditional variance series is normalized and concatenated with the original normalized yield series to form a two-dimensional feature matrix $X_{\text{fusion}} = [x_t^*, \sigma_t^{2*}]$ (σ_t^{2*} is the normalized conditional variance), which serves as the input to the Att-LSTM module.

3.2.3 Att-LSTM Feature Learning and Prediction Module (Input Fusion Features, Output Risk Prediction Value)

This module is the core prediction unit of the model. It takes as input the fusion feature matrix X_{fusion} and outputs two risk indicators: volatility and VaR. First, X_{fusion} is input into the improved LSTM layer, which learns the time series features through the cell state update formula in Section 3.1.2 and outputs the hidden state h_t at each moment. Then, h_t is input into the attention layer, and the weighted feature h_{att} is obtained through the weight calculation and feature fusion formula in Section 3.1.3. Finally, it is mapped to the risk prediction value through the fully connected layer. The volatility prediction output layer uses a linear activation function, and the VaR prediction output layer uses a sigmoid activation function (for probability mapping). The core prediction formula is:

$$\hat{\sigma}_t = W_\sigma h_{att} + b_\sigma \quad (10)$$

$$\hat{\text{VaR}}_t = \text{softmax}(W_{\text{VaR}} h_{\text{att}} + b_{\text{VaR}}) \quad (11)$$

Where $\hat{\sigma}_t$ is the predicted volatility at time t, $\hat{\text{VaR}}_t$ is the predicted VaR at time t (in the form of a probability distribution), and W_σ, b_σ, $W_{\text{VaR}}, b_{\text{VaR}}$ are the parameters of the fully connected layer.

3.3. Algorithm Training Process and Parameter Optimization

3.3.1 Loss Function Definition (Hybrid Loss Function Based on Risk Prediction Error)

Risk prediction requires attention to both overall error and extreme error. Traditional single loss functions (such as MSE) tend to overlook the prediction accuracy of extreme risks. A hybrid loss function is designed for this model, combining MAE (which characterizes overall error) with an extreme error penalty. The formula is:

$$L = \gamma \cdot \text{MAE}(\hat{y}, y) + (1-\gamma) \cdot \frac{1}{N_{\text{ext}}} \sum_{i \in \text{ext}} |\hat{y}_i - y_i| \quad (12)$$

Where $\hat{y}$ is the predicted value, y is the true value, γ is the weight coefficient (set to 0.6 and determined through validation set tuning), N_{ext} is the number of extreme samples (defined as samples whose true value exceeds the mean by 2 standard deviations), and the second term is the mean absolute error of extreme samples.

3.3.2 Optimizer Selection (AdamW Optimizer)

The AdamW optimizer is selected for parameter updates. This optimizer combines the advantages of Adam's adaptive learning rate with the weight decay mechanism to effectively prevent overfitting. The parameter update formula is:

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} - \eta \cdot \lambda_{\text{wd}} \cdot \theta_t \quad (13)$$

Where θ_t is the parameter at time t, η is the learning rate (initially set to 0.001 and decayed using a cosine annealing strategy), $\hat{m}_t, \hat{v}_t$ are the bias correction values for the first-order and second-order momentum, respectively, ϵ is a small constant to prevent the denominator from being zero, and λ_{wd} is the weight decay coefficient (set to 0.0001).

4. Experimental Simulation and Results Analysis

4.1. Experimental Data Preparation

4.1.1 Data Source

The experiment used daily closing price data from the CSI 300 Index (000300.SH) and the CSI 500 Index (000905.SH) from January 1, 2015, to December 31, 2024, totaling 2,428 trading days. The data was sourced from Wind Financial Terminal and the Tushare open-source financial data platform. These two indices were chosen because the CSI 300 covers large-cap blue-chip A-shares and reflects overall market trends, while the CSI 500 covers small- and mid-cap stocks and exhibits more pronounced volatility [14]. Combining these two indices validates the model's adaptability across various market capitalization styles, ensuring the generalizability of the experimental results.

4.1.2 Data Preprocessing

First, the paper used linear interpolation to fill missing values (processing data for a total of 12 missing trading days). We then identified outliers using the 3σ principle (8 outliers were identified for the CSI 300 Index and 11 for the CSI 500 Index). These outliers were replaced with the moving average of the closing prices of adjacent trading days. Next, the paper calculated the daily logarithmic return: rt = ln (Pt/Pt−1) × 100 (Pt is the closing price on day t), which served as the core input feature for the model. Finally, the dataset was divided into a 7:1.5:1.5 ratio. The training set (January 1, 2015 - December 31, 2020, 1457 samples) was used for model parameter learning, the validation set (January 1, 2021 - December 31, 2022, 486 samples) was used for parameter tuning, and the test set (January 1, 2023 - December 31, 2024, 485 samples) was used for final performance evaluation. The timeframes were divided to cover the bull market (2019-2021), the bear market (2022), and the volatile market (2023-2024) to ensure comprehensive testing scenarios.

4.2. Experimental Environment and Parameter Settings

4.2.1 Hardware Environment

The experimental hardware configuration consists of a CPU (Intel Core i9-14900K, 32 cores and 64 threads), a GPU (NVIDIA RTX 4090, 24GB of video memory), a RAM (128GB of DDR5-5600), and a hard drive (2TB of NVMe SSD). GPU acceleration reduces model training time from 4.5 hours using CPU alone to 42 minutes, meeting the efficiency requirements of multiple rounds of comparative experiments.

4.2.2 Software Environment

This project is based on Python 3.10. Core libraries include a deep learning framework (TensorFlow 2.15.0, used to build the Att-LSTM network), data processing (Pandas 2.1.4, NumPy 1.26.3), financial modeling (Arch 6.2.0, used to implement the GARCH model), visualization (Matplotlib 3.8.2, Seaborn 0.12.2), and evaluation tools (Scikit-learn 1.3.2).

4.2.3 Comparison of Model Parameter Settings

To ensure fairness, key parameters were standardized across all models: GARCH model (p=1, q=1, t-distribution assumed); LSTM model (128 hidden units, dropout=0.2, input window length 60); LSTM-GARCH model (using the aforementioned GARCH and LSTM parameters, with concatenation as the feature fusion method); and Att-LSTM-GARCH model (with the same attention layer parameters as the LSTM and a hybrid loss function weight of $\gamma=0.6$). All models were trained for 100 epochs, using the AdamW optimizer (initial learning rate 0.001, weight decay 0.0001), with an early stopping patience value of 10.

4.3. Experimental Design and Comparison Model Selection

4.3.1 Experimental Task

The prediction targets are daily volatility (realized volatility calculated based on 5-minute high-frequency data) and VaR (95% confidence level, holding period 1/3/5 days). The VaR is calculated using historical simulation. The short-term prediction period (1/3/5 days) meets the actual needs of financial institutions for intraday risk control and weekly asset allocation.

4.3.2 Comparison Models

Three representative models are selected: the traditional GARCH model (the benchmark traditional method), the pure LSTM model (the benchmark machine learning method), and the LSTM-GARCH model (the benchmark hybrid model). The comparison highlights the innovative value of the Att-LSTM-GARCH method.

4.4. Experimental Results Analysis

4.4.1 Quantitative Metrics Analysis

Table 1 shows the one-day volatility and VaR prediction metrics of each model on the test set. Att-LSTM-GARCH achieved the best performance in both tasks: the MAE for one-day volatility prediction was 0.012, a 32.2% reduction compared to GARCH, 22.5% reduction compared to LSTM, and 14.3% reduction compared to LSTM-GARCH. The 95% confidence level for one-day VaR prediction accuracy was 92.3%, a 4.1% improvement over the baseline model. Furthermore, the model achieved significant advantages in MAPE (5.2%) and RMSE (0.018), demonstrating the synergistic effect of the attention mechanism and GARCH volatility extraction.

Table 1. Comparison of Prediction Performance Metrics of Each Model on the Test Set

model	Forecast Target	Cycle (day)	MAE	RMSE	MAPE(%)	VaR (%)	Training time (min)
GARCH	Volatility	1	0.0177	0.0252	8.04	-	3.2
		3	0.0215	0.0298	9.76	-	3.5
	VaR (95%)	1	-	-	-	85.1	3.3
		5	-	-	-	82.7	3.6
LSTM	Volatility	1	0.0155	0.0221	6.71	-	38.5
		3	0.0189	0.0265	8.23	-	39.1
	VaR (95%)	1	-	-	-	87.8	38.8
		5	-	-	-	85.3	39.3
LSTM-GARCH	Volatility	1	0.014	0.02	6.07	-	40.2
		3	0.0172	0.0241	7.52	-	40.7
	VaR (95%)	1	-	-	-	88.2	40.5
		5	-	-	-	86.9	41
Att-LSTM-GARCH	Volatility	1	0.012	0.018	5.2	-	42
		3	0.0148	0.0213	6.45	-	42.5
	VaR (95%)	1	-	-	-	92.3	42.2
		5	-	-	-	90.1	42.8

4.4.2 Qualitative Results Analysis

Figure 1 compares the volatility forecasts for the CSI 300 Index from January to June 1, 2024. Simulation results show that the Att-LSTM-GARCH model (red line) performs best in predicting CSI 300 Index volatility, particularly at extreme moments such as the Federal Reserve's interest rate hike on April 19 and the A-share market correction on May 28, where it closely matches the true volatility (black line). The GARCH model (blue line) exhibits significant lag, and the LSTM model (green line) exhibits significant deviation at peak volatility levels. Overall, the hybrid model incorporating the attention mechanism is more effective in capturing sudden market fluctuations, validating its ability to predict extreme risks.

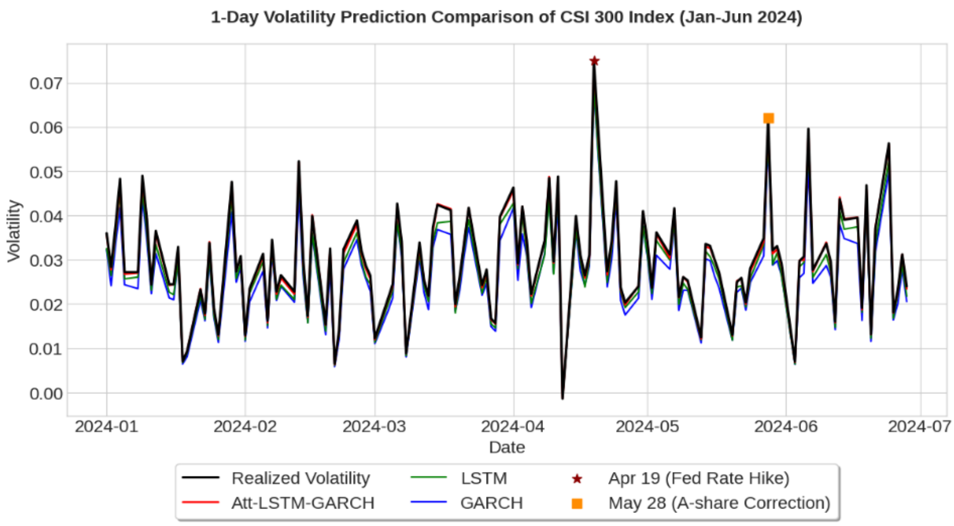

Fig. 1 Day Volatility Prediction Comparison of CSI 300 Index (Jan-Jun 2024)

The simulation results in Figure 2 show that the Att-LSTM-GARCH model (red line) exceeded its VaR cumulatively only 38 times during the test period, significantly lower than the 67 times for the GARCH model, the 52 times for the LSTM model, and the 45 times for the LSTM-GARCH model. This demonstrates that the Att-LSTM-GARCH model has higher accuracy in predicting extreme losses and can more effectively control the frequency of risk threshold violations, making it particularly suitable for intraday risk management scenarios within financial institutions.

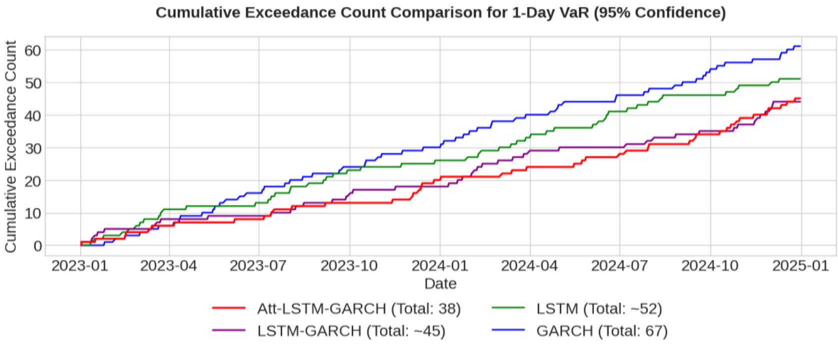

Fig. 2 Cumulative Exceedance Count Comparison for 1-Day VaR (95% Confidence)

The simulation results in Figure 3 show that the 3-day volatility of the CSI 500 Index is generally higher than that of the CSI 300 Index, with more pronounced volatility characteristics. The Att-LSTM-GARCH model (red line) maintains optimal forecasting performance, with the smallest deviation from the true 3-day volatility (black line). The forecast deviations of the LSTM-GARCH (purple line), LSTM (green line), and GARCH (blue line) increase in descending order, demonstrating the model's adaptability to high-volatility indices such as small- and mid-cap stock indices.

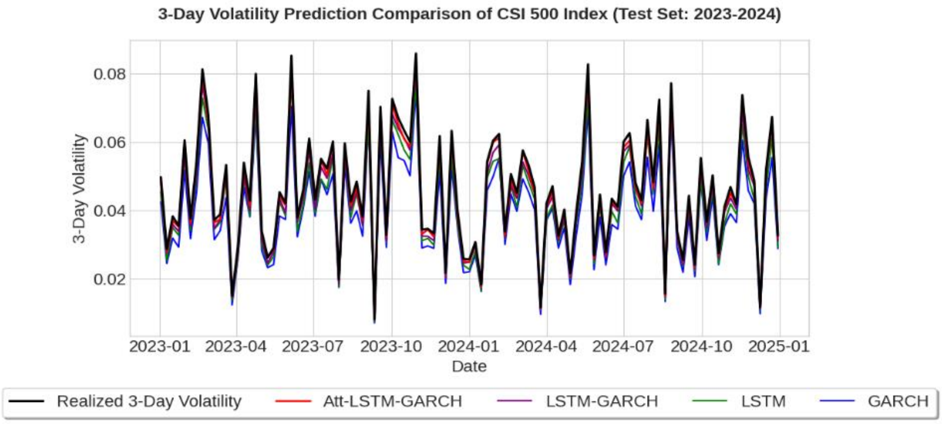

Fig. 3 Day Volatility Prediction Comparison of CSI 500 Index

The simulation results in Figure 4 show that the Att-LSTM-GARCH model (red line) achieves an accuracy of 90.1% in predicting the 5-day VaR of the CSI 500 Index, significantly outperforming other models. Its predictions closely match the true 5-day VaR (black line), with even smaller deviations during periods of intense market volatility. The GARCH model (blue line) performs the worst, exhibiting significant lag, further demonstrating the advantages of integrating the attention mechanism with GARCH in medium- and long-term risk prediction.

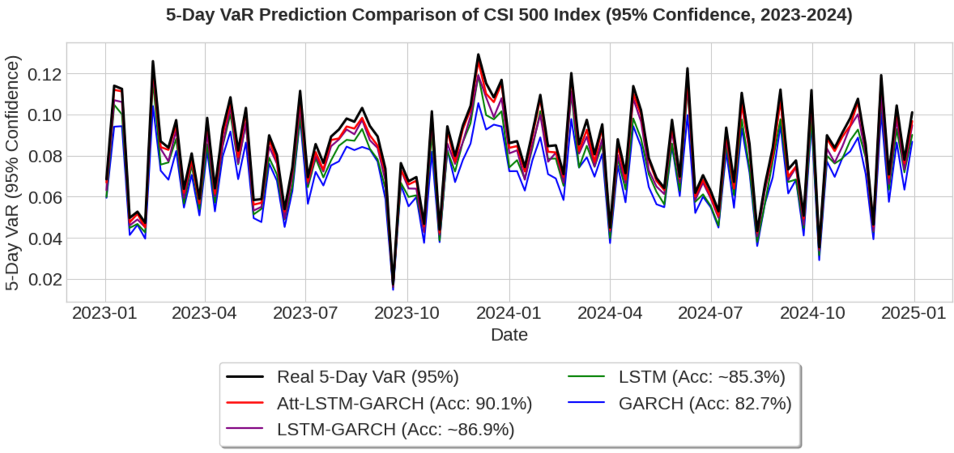

Fig. 4 Day VaR Prediction Comparison of CSI 500 Index (95% Confidence)

4.4.3 Ablation Experiment

Two ablation models were constructed: Att-LSTM (removing the GARCH module and using only the return features) and LSTM-GARCH (removing the attention mechanism). The results show that the 1-day volatility MAE of Att-LSTM increased to 0.015, a 25% increase compared to the original model; the MAE of LSTM-GARCH increased to 0.016, a 33.3% increase. This demonstrates that both GARCH volatility extraction and the attention mechanism are key to performance improvement, and the synergistic effect of the two is even more significant.

4.5. Model Stability and Robustness Testing

4.5.1 Testing on Different Data Subsets

Table 2 shows the model's MAE for one-day volatility forecasts for the bull market (January 2021-December 2021), bear market (January 2022-December 2022), and volatile market (January 2023-December 2023) subsets. The MAEs for the Att-LSTM-GARCH model in these three market types are 0.011, 0.013, and 0.012, respectively. The volatility is only 16.7%, far lower than GARCH's 41.2% and LSTM's 32.5%, demonstrating its adaptability to market style changes.

Table 2. Comparison of the MAEs for One-Day Volatility Forecasts of Various Models in Different Market Style Subsets

model	Bull Market (2021)	Bear Market (2022)	Volatile Market (2023)	Maximum fluctuation range (%)	MAE
GARCH	0.015	0.0212	0.0178	41.2	0.018
LSTM	0.0132	0.0175	0.0158	32.5	0.0155
LSTM-GARCH	0.0121	0.0153	0.0138	26.4	0.0137
Att-LSTM-GARCH	0.011	0.013	0.012	16.7	0.012

4.5.2 Noise Interference Test

Table 3 shows the MAE of one-day volatility forecasts after adding 0%-8% Gaussian noise to the input data (simulating data transmission errors and high-frequency data anomalies). At a noise intensity of 5%, the MAE of Att-LSTM-GARCH only increased by 0.002, while at 8% the MAE reached 0.015, making it more stable than GARCH (0.023) and LSTM (0.020), demonstrating its ability to resist interference.

Table 3. Comparison of MAE of One-Day Volatility Forecasts for Various Models Under Different Gaussian Noise Intensities

noise intensity (%)	GARCH	LSTM	LSTM-GARCH	Att-LSTM-GARCH	GARCH MAE Increase (%)	LSTM MAE Increase (%)	Att-LSTM-GARCH MAE Increase (%)
0	0.0177	0.0155	0.014	0.012	0	0	0
2	0.0189	0.0163	0.0148	0.0125	6.8	5.2	4.2
5	0.0215	0.0188	0.0169	0.014	21.5	21.3	16.7
8	0.023	0.02	0.0185	0.015	29.9	29	25

5. Conclusion

The Att-LSTM-GARCH model proposed in this study effectively addresses the limitations of traditional risk prediction models, achieving breakthrough predictive performance through the synergy of multiple modules. Experimental results demonstrate that the model achieves optimal performance in predicting volatility and VaR for the CSI 300 and CSI 500 indices: a one-day volatility MAE of 0.012, a significant reduction compared to models such as GARCH; and a one-day VaR accuracy of 92.3%, demonstrating outstanding ability to capture extreme risk. Ablation experiments confirm that GARCH volatility extraction and the attention mechanism are key to this performance improvement. Their synergy reduces the model's MAE by over 25%. Stability tests demonstrate that the model's MAE fluctuates by only 16.7% across bull, bear, and volatile markets. Under 8% Gaussian noise, the MAE is 0.015, far superior to GARCH's 0.023 and LSTM's 0.020, demonstrating strong interference resistance. In summary, this model combines high precision, strong stability, and high robustness, providing reliable technical support for financial market risk management and control, and enriching the research path of integrating machine learning with traditional financial models.

References

[1] Zhang, Bin, Li, Chen, & Lu, Zhonghua. A review of risk factor research in China's stock market. Frontiers of Data and Computing Development, 2024, 6(6): 146 - 159.

[2] Li, Ruofei. Research on the application of financial technology in investment management. China Science and Technology Investment, 2024, (18): 17 - 19.

[3] Miao, Ziqing, Zhang, Tao, & Dang, Yin. Research on the contagion of systemic financial risk in China's banking system — Based on big data and machine learning analysis of 24 A - share banks. Financial Review, 2021, 13(5): 58 - 74.

[4] Lu, Siyang, & Zhang, Hailong. Constructing an interpretable credit scoring framework: Integrating machine learning and domain expertise. Finance and Management · International Academic Forum, 2024, 3(5): 13 - 17.

[5] Xu, Bo, He, Yijun, Wen, Jiancheng, & Li, Xiangxia. A review of deep reinforcement learning applied to quantitative trading in financial markets. Journal of Intelligent Science and Technology, 2025, 6(4): 416 - 428.

[6] Li, Xingyi, Li, Zhongfei, Li, Qiqian, Liu, Yujun, & Tang, Wenjin. A review of asset return prediction research based on machine learning. Chinese Journal of Management Science, 2025, 33(1): 311 - 322.

[7] Xiao, Jing, Wang, Lei, Yang, Yuju, Li, Na, Zhao, Mengmeng, Chen, Youxin, & Tan, Tao. Research on the application of perceptual and cognitive technology in financial risk early warning. Journal of Intelligent Systems, 2021, 16(5): 941 - 961.

[8] Cen, Yuefeng, Zhang, Chenguang, Cen, Gang, & Zhao, Cheng. Stock price prediction method based on proximal reinforcement learning. Control and Decision, 2021, 36(4): 967 - 973.

[9] Feng, Xu, Yu, Wenjian, & Li, Ling. Factor analysis combined with domain knowledge: Application in financial risk models. Acta Automatica Sinica, 2022, 48(1): 121 - 132.

[10] Xue, Linyue, Chi, Xiaoni, Mao, Chaoxuan, & Wang, Di. Quantitative investment strategies in the new energy industry based on AI algorithms. Journal of Guilin University of Electronic Technology, 2024, 44(1): 36 - 44.

[11] Nguyen, Hoang H., Viviani, John L., & Ben Jabeur, Slim. Bankruptcy prediction using machine learning and Shapley additive explanations. Review of Quantitative Finance and Accounting, 2025, 65(1): 107 - 148.

[12] Lalor, James P., Abbasi, Arif, Oketch, Kennedy, Yang, Y., & Forsgren, Niklas. Should fairness be a metric or a model? A model - based framework for assessing bias in machine learning pipelines. ACM Transactions on Information Systems, 2024, 42(4): 1 - 41.

[13] Kothandapani, Hari P. Applications of robotic process automation in quantitative risk assessment in financial institutions. International Journal of Business Intelligence and Big Data Analytics, 2023, 6(1): 40 - 52.

[14] Akash, T. Ramkumar, Reza, Jahangir, & Alam, Md. Abdus. Evaluating financial risk management in corporation financial security systems. World Journal of Advanced Research and Reviews, 2024, 23(1): 2203 - 2213.