

Application of Time Series Models in Stock Price Prediction

-- Taking the CSI 300 Index as an Example

Runyi Liu

School of Mathematics and Computer Science, Shantou University, Shantou, China

Abstract: Stock price prediction has long been a challenge in financial research due to the high noise, non stationarity, and weak predictability of the data. This paper uses daily market data of the CSI 300 Index from January 2020 to December 2025 as a sample. Under a rolling expansion window framework that strictly eliminates forward looking bias, it systematically compares the predictive performance of five types of time series models: random walk, ARIMA, GARCH, XGBoost, and LSTM. The results show that: random walk constitutes a formidable baseline in short term prediction; ARIMA has notable cost effectiveness in stable phases but underperforms when facing sudden fluctuations; the GARCH family of models has irreplaceable value in risk management; XGBoost has the best robustness in medium term prediction; and LSTM leads in multi step prediction but has limited economic relevance. Further testing by market phase reveals that simple models have the highest cost effectiveness in low volatility periods, while the advantages of nonlinear models only become apparent in high volatility periods. Based on this, this paper proposes a hierarchical prediction framework of "baseline monitoring—volatility calibration—nonlinear correction," emphasizing that the true value of the model lies in probabilistic state reference rather than precise price prediction.

Keywords: Time series forecasting; Stock prices; Random walk; ARIMA model; GARCH model; Long short term memory network; Rolling validation.

1. Introduction

Predicting future stock price trends remains a core concern for investors. However, stock price sequences inherently possess characteristics such as high noise, non-stationarity, and time-varying structures, making it difficult for any single forecasting method to maintain long-term stability and effectiveness. Time series models, relying solely on historical price and volume sequences without requiring additional fundamental, fund flow, or sentiment data, have become the most fundamental and convenient tool in the field of quantitative forecasting. From the classic ARIMA model to the GARCH family of models specifically designed to characterize volatility clustering, and to the widely used machine learning and deep learning methods in recent years, various models continue to generate discussion in academic research and practical application.

However, in practical applications, some common design flaws often lead to unreliable research conclusions: using full sample information for standardization or feature extraction during data preprocessing can result in future information leakage; evaluating models solely based on single indicators such as root mean square error ignores the crucial significance of directional judgment and uncertainty quantification for trading decisions; and performing only a static sample partitioning instead of simulating daily rolling forecasts in real-world scenarios overestimates the model's generalization ability. These problems make it difficult to reproduce many seemingly excellent results in actual trading.

This paper attempts to circumvent these shortcomings by using the CSI 300 Index as the research object and conducting a fair comparison of five representative time series models under strictly unified rules. Instead of pursuing complex formula derivations, the paper uses accessible language to explain what difficulties each model addresses in stock price

data, how well it addresses them, and where its limitations lie. Discussions are further elaborated with specific market segments, aiming to provide quantitative strategy designers with a clear and reproducible reference.

2. The Core Challenges of Stock Price Prediction: Data Characteristics and Modeling Constraints

2.1. Distribution Characteristics of Returns

Logarithmic returns calculated using the daily closing prices of the CSI 300 Index from January 2, 2020 to December 31, 2025, exhibit typical characteristics of financial time series: a mean close to zero, a standard deviation of approximately 1.25%, negative skewness, and significantly higher kurtosis than a normal distribution. The Jarque-Bera test formally rejects the null hypothesis of a normal distribution. This heavy-tailed characteristic means that the probability of extreme negative volatility is far higher than expected under the normal assumption, fundamentally undermining the applicability of classical statistical inference methods relying on the normal distribution assumption in stock price prediction.

2.2. Non-stationarity And Stationarity Adjustment

The ADF unit root test shows that the price series cannot reject the null hypothesis of a unit root, meaning that prices themselves are not stationary; however, the logarithmic return series rejects the unit root hypothesis at a 1% significance level, indicating stationarity. This means that most time series models must apply differencing or take logarithmic returns of the price series before modeling. All models in this paper model the logarithmic return, and the predicted values are

then converted back to prices through inverse transformation.

2.3. Volatility Clustering and Failure of Constant Variance

Another significant characteristic of financial time series is the clustering of volatility—large fluctuations occur one after another, while small fluctuations are relatively stable [1]. This effect occurred multiple times during the sample period, most notably the explosive surge in late September 2024, in which the CSI 300 Index rose by more than 20% in five trading days. If a constant variance assumption is adopted, the prediction interval width remains unchanged, and it will be repeatedly breached when market sentiment changes drastically, thus losing its risk warning function. This indicates that dynamic modeling of volatility is a basic requirement at the risk management level.

2.4. Systematic Requirements of Data Characteristics for the Model

The above analysis sets forth four systematic requirements for the predictive model: first, it must be capable of handling non-stationary sequences; second, it should remain robust to extreme shocks under heavy-tailed distributions; third, volatility estimates should be dynamically updated, adapting to market conditions as they contract or expand; and fourth, it should possess mechanisms to capture nonlinear patterns to address the widespread state dependence and turning point effects in stock price sequences. These requirements collectively constitute the benchmark for subsequent model evaluation.

3. Model System and Methodological Logic

3.1. Baseline Model Layer: Random Walk and Linear Time Series

The random walk model assumes that the optimal prediction of future prices is the current price, corresponding to a weaker version of the efficient market hypothesis—that all historical price information is fully reflected in the current price. This model forms the baseline for financial forecasting; any complex model must first prove that it can surpass this naive strategy before it is worth discussing. The ARIMA model transforms non-stationary sequences into stationary ones through differencing, capturing short-term inertia with a linear combination of autoregressive and moving average terms. Parameters are automatically selected through information criteria, making it structurally transparent and computationally fast. However, this model assumes constant variance and a linear data generation process, resulting in a lag in response to sharp market fluctuations or trend reversals. The GARCH family of models does not directly improve point prediction accuracy but models the time-varying characteristics of volatility, allowing the prediction interval to contract or expand with market conditions. Its expanded form can also capture the leverage effect—a negative shock of equal magnitude triggers greater subsequent volatility than a positive shock.

3.2. Machine Learning Layer: Feature Engineering and Nonparametric Learning

Machine learning methods, exemplified by XGBoost, refactor prediction into supervised learning: lagged prices,

lagged volumes, and technical indicators such as MACD and RSI are constructed as features, and the algorithm automatically learns the nonlinear mapping between features and the target [2]. Its advantage lies in not pre-setting the function form and capturing the nonlinear interaction between features; its disadvantage is that the modeling of time-dependent structures is not natural enough, and the quality of feature construction determines the prediction effect. The ARIMAX model represents another approach—integrating external variables such as trading volume while maintaining the interpretability of the linear framework.

3.3. Deep Learning Layers: End-to-End Sequence Learning

LSTM networks alleviate the gradient vanishing problem through gating structures and can automatically extract long-range dependencies in sequences. The input is a multivariate sequence of multiple consecutive trading days, and the output is a prediction of the target day. No manual feature construction is required, which makes it possible to capture complex temporal patterns that are difficult for linear models to discover [3]. Temporal convolutional networks and Transformers represent the update direction of sequence modeling [4]. However, the cost of deep models is also obvious: training requires a lot of data and computing power; hyperparameters have a significant impact, the "black box" characteristic makes signal attribution difficult, and it is very easy to overfit on small samples.

4. Research Design and Fairness Assurance

4.1. Data Sources and Sample Description

This article selects daily trading data for the CSI 300 Index from January 2, 2020 to December 31, 2025, totaling approximately 1460 trading days. Fields include opening price, highest price, lowest price, closing price, and trading volume. This index was chosen because it covers a wide range of industries, has ample liquidity, and can comprehensively reflect the overall trend of the A-share market, while avoiding structural changes in individual stocks caused by events such as dividends, stock splits, mergers, and acquisitions. It should be noted that the periodic adjustments to the index constituents introduce survivorship bias, and caution should be exercised when generalizing the conclusions to individual stocks.

4.2. Rolling Expanded Window Validation

This paper adopts an extended window rolling prediction to simulate the information environment of real trading. Approximately 974 trading days from January 2020 to December 2023 are used as the initial training set, and predictions are made daily starting from January 2024. Each prediction uses only information up to the previous trading day; after completion, the data for that day is added to the training set, and the model is updated accordingly. ARIMA and GARCH are re-estimated, LSTM is incrementally fine-tuned, and XGBoost is retrained periodically. This design systematically eliminates forward-looking bias; all hyperparameters are determined through time-series cross-validation on the initial training set and are not adjusted during the testing period [5].

4.3. Three-Dimensional Design of Evaluation Dimensions

This paper evaluates the model from three dimensions: the point prediction dimension uses RMSE and MAPE to measure price error; the direction judgment dimension uses the accuracy of the upward or downward direction, which is directly related to the effectiveness of the trading signal; and the interval reliability dimension measures the actual coverage of 95% of the predicted interval, with an ideal value close to 95%. These three dimensions correspond to three decision-making scenarios: numerical reference, directional trading, and risk control. The multi-step prediction task is set up to predict the closing price for the next 1 day, 5 days, 10 days, and 20 days respectively.

5. Empirical Results and Analysis

5.1. Comparison of Point Prediction Accuracy

In one-step prediction, random walk and ARIMA perform similarly, while single-layer LSTM and XGBoost have a slight advantage, but the lead is limited. Differences emerge in multi-step prediction: as the step size increases, the error of linear models accumulates rapidly, while the error growth of LSTM and XGBoost is relatively gradual. XGBoost is most robust in five-day and ten-day predictions, while LSTM shows a more significant advantage in twenty-day predictions. The competitiveness of random walk in one-step prediction reminds us that the incremental information from historical prices in the very short term is extremely limited, and conclusions claiming significant superiority over random walk require careful examination to determine if there is a forward-looking bias.

5.2. Ability To Determine Direction

Overall, the directional accuracy is limited, with none of the models exceeding 60%. The advantage of LSTM over ARIMA is statistically identifiable, but its economic relevance is limited. When the step size is extended to ten days, the directional accuracy of all models decreases, with ARIMA showing a particularly pronounced decline. If transaction costs and slippage are factored in, even the best-performing model's advantage diminishes significantly after deducting costs. This indicates that relying solely on price series directional prediction signals is unlikely to generate robust excess returns in live trading.

5.3. Volatility Modeling and Interval Reliability

GARCH's unique contribution lies in its interval coverage. Under the constant variance assumption, ARIMA's actual coverage of 95% of the predicted intervals is significantly low, indicating a lack of adaptive interval capabilities. Introducing GARCH improves the coverage to near the nominal level. Taking a sudden surge in volatility during the testing period as an example: GARCH detected the increase in volatility and widened the interval before the volatility escalated, with subsequent sharp price fluctuations largely falling within the dynamic interval; while ARIMA's constant interval was breached in the early stages of the volatility escalation. This demonstrates that GARCH's dynamic quantification of risk levels is irreplaceable in practical decision-making.

5.4. Performance Differences by Market Stage

The testing period was divided into low-volatility and high-volatility periods for evaluation, revealing a clear pattern: In low-volatility environments, ARIMA's accuracy was not significantly different from that of complex models, while its computational cost was only a fraction of the latter, making it the most cost-effective; in high-volatility environments, ARIMA's directional accuracy dropped sharply, essentially losing its reference value, while LSTM and XGBoost, although also declining, remained at a certain level. XGBoost performed relatively evenly in both environments. This finding indicates that the judgment of model performance is highly dependent on market conditions, and a dynamic selection should be made rather than pursuing a single optimal solution [6].

6. Discussion: The Boundaries of Predictive Capability and Application Paths

6.1. The Ceiling of Pure Time Series Forecasting Capabilities

The above results reveal an unavoidable conclusion: on a daily timescale, relying solely on historical price and volume information makes it difficult to maintain a directional accuracy rate exceeding 60% in the long term, and the numerical error of multi-step predictions far exceeds the tolerance range of transaction costs. Understanding time series models as "tools that can provide precise buy and sell prices" is an unrealistic expectation. Their true value lies in providing probabilistic state references for investment decisions—whether the current market is more inclined to continue or reverse the trend, and whether volatility is escalating—this information can effectively assist in position management and risk budgeting.

6.2. Conception of a Hierarchical Prediction Framework

Based on the differentiated performance of various models, a pragmatic practical framework can be designed as a three-layer structure. The bottom layer, consisting of random walk and ARIMA running continuously with extremely low computational cost, provides a basic reference for trend direction. The middle layer uses GARCH to dynamically estimate risk levels, adjusting position size and stop-loss levels accordingly. The top layer invokes nonlinear correction signals from deep models such as LSTM when volatility significantly amplifies or the market enters atypical patterns. The three layers dynamically collaborate according to market conditions.

6.3. Methodological Limitations and Research Integrity

This study has several limitations that need to be openly acknowledged. First, it only used price and volume data, omitting fundamental, capital flow, and news sentiment information, thus lacking predictive power for long-term turning points driven by policy or fundamentals. Second, the periodic adjustments to the CSI 300 constituent stocks introduce survivorship bias. Third, the problem of error accumulation in multi-step predictions remains unresolved. Fourth, the conclusions of daily frequency analysis are not applicable to higher-frequency trading and stocks with poor

liquidity. These limitations are prerequisites that users must be aware of when referencing the relevant conclusions.

7. Conclusion

This paper systematically compares the performance of five types of time series models in predicting the CSI 300 index price within a rolling framework that strictly eliminates forward-looking bias. The results demonstrate the difficulty of surpassing the random walk as a short-term baseline, confirm the irreplaceable role of GARCH in volatility quantification, and verify the relative advantages of nonlinear models in multi-step prediction and complex market environments. The core finding is not "which model is best," but rather that model selection must match market conditions, and each model has a clear capability boundary. The difficulty in stock price prediction fundamentally stems from the extremely low signal-to-noise ratio of financial data; maintaining a clear understanding of this is more important than constantly piling on more complex models.

References

- [1] Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2), 223–236.
- [2] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *CoRR*, abs/1603.02754.
- [3] Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654–669.
- [4] Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied Soft Computing*, 90, 106181.
- [5] Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191, 192–213.
- [6] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PLOS ONE*, 13(3), e0194889, <https://doi.org/10.1371/journal.pone.0194889>.