ISSN 2959-6157

Apple Company Daily Closing Price Prediction Based on Time Series Analysis and Monte Carlo Method

Lixiang Chen[1*]

Kuang Yaming Honors School, Nanjing University, Nanjing, China
Corresponding author.
Email: 221840100@smail.nju.edu.cn

Abstract:

Accurate forecasting of stock prices is critical for quantitative investment and risk management. This paper investigates the methodology and practical application of time series analysis and the Monte Carlo method in predicting AAPL's daily closing prices. Time series are decomposed into several components so that models are built to capture the patterns of price change and generate forecasts. Empirical results show that the time series model can achieve acceptable prediction accuracy. But the overly smooth curve of price path indicate that the model fails to reflect fluctuations in real market environment. Meanwhile, the Monte Carlo method is used to extract statistical regularities and randomness from price changes. Specifically, it maps sample data to pseudo-random numbers, which are then employed for stochastic price simulation to generate future price paths. The model effectively simulates AAPL's general trend and realistic fluctuations over the coming trading days, yet its outputs rely on historical data and random sampling and thus should be used with caution. Despite their different mechanisms, both methods yield consistent overall trend results, offering valuable references for trading or investing AAPL stock.

Keywords: Time series analysis; Monte Carlo method; Daily closing price; Logarithmic return.

1. Introduction

The vitality of a financial market is reflected in the level of price fluctuations of its main asset prices. Whether in determining its short-term trading strategy, or measuring the value of its long-term investment, making an accurate prediction of asset price trend is of practical importance. At present, there are various methods for forecasting financial time series. However, the price of a financial asset is controlled by many factors, such as macroeconomic factors, market sentiment and policy factors [1,2]. Therefore, accurate and timely predictions are still challenging.

Among the various forecasting methods, time series analysis and Monte Carlo simulation are the two most typical prediction methods. Research on these two methods has been further developed in recent years. At present, in time series analysis, many new models have appeared. For example, MLP-based models such as TimeMixer use multiscale mixing or centralized channel mixing to improve efficiency and scalability [3]. Also, DecompSSM was proposed [4]. It combines state space models and decomposition. By modeling the time-series components separately, it can better model the variable dependency in the long sequence. Another learnable frequency-domain model called FreDN was presented [5]. The model used FFT to reduce spectral leakage and improve nonstationary time-series decomposition. As for Monte Carlo method, it is widely used in financial background. Kamila et al. used Monte Carlo simulation based on to the geometric Brownian motion model to predict the price of stocks and verify the efficiency of this method to capture the uncertainty of market [6]. Hainaut and Akbaraly propose a local least squares Monte Carlo approach which utilizes K-means clustering to group the simulated results and fit the separate regression in each cluster [7]. This proposed design overcomes the limitation of global regression in the fitting of nonlinear relationships.

There also exist attempts to combine time series analysis with Monte Carlo method. Singhal et al. proposed a hybrid forecasting model by combining the ARIMA model and Monte Carlo simulation [8]. They find that this combination effectively reduces the prediction risk of prediction by using single model. This study employs both time series analysis and Monte Carlo method to forecast the daily closing price of the same stock, Apple Inc. (AAPL), and discusses the performance of the two methods in the prediction process, providing reference for the selection and application of the forecasting methods in financial time series.

2. Time Series Analysis

2.1 Methodology

Time series analysis is a statistical method that can reveal the internal structure and dynamic patterns of a system through sequential data [9]. The core idea of time series analysis is to decompose a observed series into several components. An additive decomposition model is commonly adopted, with the basic form

$$H(t) = X(t) + P(t) + R(t) \tag{1}$$

Here $H(t)$ denotes the original series, i.e., the observed time series; $X(t)$ is the trend component; $P(t)$ represents the periodic component; and $R(t)$ is the random component.

The additive model applies when the components are independent and the amplitudes of fluctuations do not depend on the trend level. Among the three components, the trend component reflects the long-term evolution pattern of precipitation and is the most macro-guidance component in time series. Here a quadratic polynomial is used to fit the trend:

$$X(t) = b_0 + b_1 t + b_2 t^2 \tag{2}$$

where t is the time index (e.g., the index of the trading days), and b_0, b_1, b_2 are coefficients to be estimated.

To evaluate how well the trend component explains the original data, the correlation coefficient is calculated:

$$R = \sqrt{1 - \frac{\sum_{t=1}^{n}\left(x_t - \hat{x}_t\right)^2}{\sum_{t=1}^{n}\left(x_t - \bar{x}\right)^2}} \tag{3}$$

The closer the correlation coefficient R is to 1, the stronger the trend component's explanatory power and the better the fit. Generally, $|R| \geqslant 0.8$ is required to consider the trend component significant and worthy of inclusion in the model. If R is too small, it indicates no obvious trend in the series, and periodic component extraction can be considered directly.

After removing the trend component, the remaining series mainly contains periodic fluctuations and random noise. To identify periodic components, this study uses spectral analysis to get periodic terms by converting time series in the time domain into frequency domain and applying Fourier series expansion to capture hidden periodic patterns. The approximate periodic component is expressed as

$$P(t) = \frac{a_0}{2} + \sum_{k=1}^{L}\left(a_k cos\frac{2\pi kt}{n} + b_k sin\frac{2\pi kt}{n}\right) \tag{4}$$

The Fourier coefficients are calculated as:

$$\begin{cases} a_0 = \dfrac{1}{n}\sum_{t=1}^{n} P(t) \\ a_k = \dfrac{2}{n}\sum_{t=1}^{n} P(t) cos\dfrac{2\pi kt}{n} \\ b_k = \dfrac{2}{n}\sum_{t=1}^{n} P(t) sin\dfrac{2\pi kt}{n} \end{cases} \tag{5}$$

Not all harmonics deserve inclusion in the model. Only those with sufficiently large amplitudes are considered true periodic signals. This study uses the following statistic for significance testing:

ISSN 2959-6157

$$s_k^2 := a_k^2 + b_k^2 > \frac{4s^2 log\left(k/\alpha\right)}{n} \quad (6)$$

where s^2 is the series variance, and α is the significance level (usually 0.05).

After both the trend and significant periodic components are removed, the remaining series R(t) is considered a stationary stochastic process, meaning its statistical properties (e.g., mean, variance) do not change over time. Such series are commonly described using autoregressive models.

The autoregressive model assumes a linear relationship between the current value and values from several past time points, with the mathematical form:

$$R(t) = \epsilon_t + \Phi_1 R(t-1) + \Phi_2 R(t-2) + \ldots + \Phi_p R(t-p) \quad (7)$$

Determining the order p of the autoregressive model is crucial. An order too low results in underfitting, which does not capture the characteristics of series well; an order too high results in overfitting, decreasing generalization of the model. The Akaike Information Criterion (AIC) is employed in this study as an order selection criterion [10], i.e.

$$p_{best} = \underset{p}{argmin}\, AIC(p) = nlog\hat{\sigma}_p^2 + 2p \quad (8)$$

where $\hat{\sigma}_p^2$ is the variance of the residuals of the AR(p) model. By linearly superimposing the extracted trend, periodic, and random components, the author obtains the complete precipitation prediction model.

In order to evaluate the model's prediction performance, this study uses the posterior difference test with the following two indicators:

$$c = \frac{S_2}{S_1}, p = P\left(\left|q(k) - q\right| < 0.6745 S_1\right) \quad (9)$$

Based on the values of c and p, prediction performance is divided into four grades, see Table 1.

Table 1. Predicting effect of the posterior forecast method

Prediction Effect	Good	Qualified	Marginal	Unqualified
p	>0.95	>0.80	>0.70	≤0.70
c	<0.35	<0.50	<0.65	≥0.65

2.2 Numerical Results

In this part, empirical experiments are conducted. This research uses AAPL's daily closing stock price as the subject of empirical experiments. The data are from investing.com, from January 4, 2021, to March 16, 2026. There are 1,305 trading days' data in total. Although these trading days are not consecutive, for convenience, they can still be treated as consecutive indexes in practice because the trading behaviour itself is the driver of price change for an asset.

To construct the prediction formula and verify its predictive ability, the full sample data is divided into a modeling set (Estimation Period) and a testing set (Validation Period). Modeling set is used to estimate all the relevant parameters, which ends on December 31, 2025, consisting of 1,255 samples. The testing set is meant for evaluating the model's out-of-sample prediction accuracy. It contains 50 samples, covering the period from January 2, 2026, to March 16, 2026.

When using the model to predict future daily closing price data, the starting time of prediction is set to March 17, 2026, with a predictive horizon of 400 trading days. Based on the principle of least squares, the coefficients of the trend function are calculated. The values are shown in the following Table 2.

Table 2. The coefficients of the trend function

Coefficients	b_0	b_1	b_3
Values	136.4456	0.0233	0.0001

Hence, the trend function is:

$$X(t) = 136.4456 + 0.0233t + 0.0001t^2 \quad (10)$$

It can be observed that both the linear coefficient $b_1 = 0.0233$ and the quadratic coefficient $b_2 = 0.0001$ are relatively small. which means that the long-term variation trend of the stock price in the entire study period is not strong. Moreover, the quadratic coefficient is positive, which reflects a small upward acceleration. Other than that, the positive linear coefficient reflects an overall slow upward movement in the stock price.

The correlation coefficient of the trend component is R = 0.9040, and coefficient of determination is R^2 = 0.8172. This suggests that the trend component can explain about

81.72% of the variation of the original series, which shows a good fit.

Although the trend coefficients are small, this relatively high correlation coefficient proves that the quadratic polynomial form can effectively capture the long-term trend of the stock price series. This is the empirical basis for the following extraction of periodic and random components.

After removing the trend component, spectral analysis is used to obtain the energy s_k^2 of each harmonic term in the new series. The value is then compared with the significance threshold. Under the significance level $\alpha = 0.05$, the harmonic terms with an energy greater than the significance threshold are selected. There are a total of 9 harmonic components that pass the significance test. Their harmonic orders and corresponding harmonic coefficients are shown in Table 3.

Table 3. The coefficients of significant periodic terms

k	2	3	4	5	6
a_k	-7.2462	-4.661	13.3083	4.994	3.0764
b_k	5.7574	0.4834	0.4492	-6.7069	-3.8064
k	8	10	11	16	
a_k	-0.2119	3.1579	-1.1264	-0.8547	
b_k	-7.6434	1.9466	-3.2223	0.3433	

These harmonic components correspond to different cycle lengths. For the given series with sample size $n = 1255$, the cycle length corresponding to the k-th harmonic is n / k. Based on the selected harmonic orders, the periodic components are primarily concentrated in the low-frequency region, corresponding to longer fluctuation cycles. This indicates that the long-term, trend-like fluctuation characteristics of the stock price are more significant than its short-term random fluctuations. Specifically, there are modelable periodic patterns in the stock price series, rather than pure randomness.

An autoregressive model is constructed for the residual series $R(t)$ after removing the trend and cyclical components. The AIC is used to determine the model order. The order that minimizes the AIC values is selected as the optimal order. The results show that the optimal order for the AR model is $p = 3$. The coefficient estimates for the AR(3) model are as follows:

$$\Phi_1 = 0.9593, \Phi_2 = -0.0193, \Phi_3 = -0.0633 \tag{11}$$

Therefore, the specific form of the AR(3) model is:

$$R(t) = \Phi_1 R(t-1) + \Phi_2 R(t-2) + \Phi_3 R(t-3) + \epsilon_t \tag{12}$$

where $?_t$ is the noise sequence.

As shown in the estimation results, the first-order autoregressive coefficient $\Phi_1 = 0.9593$ is significantly positive and close to 1. This indicates that the residual series displays strong first-order autocorrelation, i.e., the current value of the series is highly correlated with its previous value.The second- and third-order autoregressive coefficients Φ_2 and Φ_3 are both negative but relatively small in magnitude. This implies that the second- and third-order lagged terms have a limited explanatory power for the current value. This finding aligns with common characteristics of financial time series, where the random fluctuation component of stock prices exhibits strong persistence, and past fluctuations have a significant impact on subsequent fluctuations.

Based on the model's performance during the Validation Period, the posterior error ratio is calculated as $c = 0.0799$, and the small error frequency is $p = 1.0000$. According to Table 1, it suggests that the model achieves extremely high fitting accuracy.

As Fig. 1 shows, during the Estimation period, the fitted value curve (red dashed line) closely aligns with the actual value curve (blue solid line), with the two curves almost completely overlapping, further confirming this point. The figure also shows that the prediction curve for the Validation Period (cyan solid line with plus markers) deviates only slightly from the actual curve (with errors within 10%), which suggests that the model has good out-of-sample predictive capability.

ISSN 2959-6157

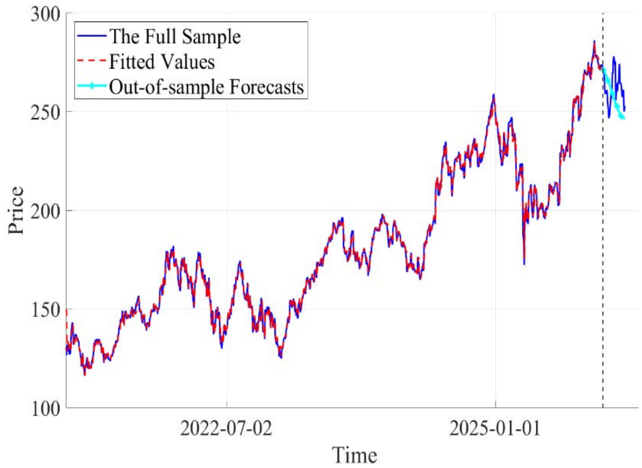

Fig. 1 Comparison of sample data and calculated values

The forecast results of the model are shown in Fig. 2. As can be observed, the plot illustrates the model's predictions for the general trend of AAPL's daily closing price over the next 400 trading days. According to its out-of-sample forecast performance, the model have good predictive effect in the short term. But in long-term forecasting, the model inadequately captures common price fluctuations. This is reflected by the excessively smooth predicted curve compared with historical price series. Hence, the model is suitable for estimating the general trend and may not be appropriate for precise price forecast.

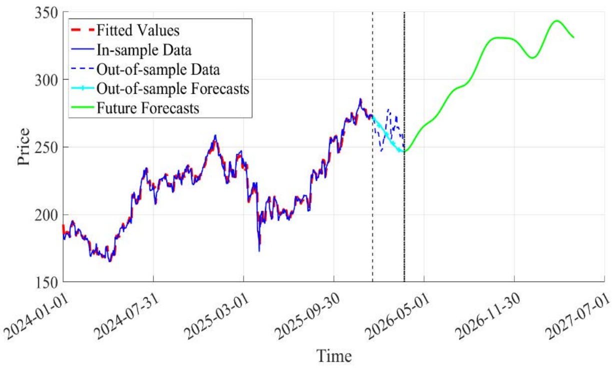

Fig. 2 The predicted curve of time series analysis

3. Monte Carlo Method

3.1 Data Preparation

The Monte Carlo method involves random sampling to obtain numerical results. Hence as a widely used random sampling technique, its main advantage is the high flexibility and general applicability of the method. This study employs a non-parametric Monte Carlo method as its main model. The key assumption of this approach is that the patterns of future price movements can be learned from the historical data. The model does not assume that returns follow any particular parametric distribution (for example, a normal distribution. Instead, it directly uses the empirical distribution of historical data for sampling.

In this section, the author used the same interval of AAPL daily closing price data as in the previous section, but with a different approach to data processing. However, the raw daily closing price data will not be used directly for modeling.

In financial time series analysis, raw price series $\{p_i\}_{i=1}^{n}$ are typically considered non-stationary, exhibiting trends and heteroskedasticity [11,12]. Price series of stocks often show a long-term trend, with the mean and variance changing over time. Therefore, a preprocessing step needs to be applied first, since it is difficult to capture stable statistical regularities by directly modeling raw prices.

For this reason, this study considered the logarithmic return series $\{r_i\}_{i=2}^{n}$ defined as:

$$r_i = log\left(p_i / p_{i-1}\right), i = 2,3,\cdots,n \tag{13}$$

Such series are generally weakly stationary. The stationarity of returns arises from differencing. To be more specific, the increment of the logarithmic price, i.e., the logarithmic return, removes the trend component from the price series. The newly formed time series of the logarithmic return typically fluctuates around zero, with roughly constant variance. Such properties meet the conditions of many statistical methods' underlying assumptions.

Augmented Dickey-Fuller (ADF) test results in Table 4 confirm that the logarithmic return of AAPL's daily closing stock price is a stationary stochastic process with constant mean and variance [13].

Table 4. ADF test results for logarithmic returns

Variable	Order of Differencing	t	P	AIC	Critical Values		
					1%	5%	10%
Logarithmic return	0	-35.33	0.000***	-6741.543	-3.435	-2.864	-2.568
	1	-13.612	0.000***	-6655.247	-3.435	-2.864	-2.568
	2	-16.567	0.000***	-6500.136	-3.435	-2.864	-2.568
Note: ***, **, and * denote significance levels of 1%, 5%, and 10%, respectively.							

3.2 Empirical Design

Similar to that of the time series analysis, the full dataset is split into a modeling set and a testing set. The modeling set, which ends on December 31, 2025, is used to construct the historical return distribution. Meanwhile the testing set consists of 50 observations and serves to evaluate the model's predictive performance.

Table 5. Sample values of logarithmic returns and corresponding probability intervals

Logarithmic Returns(%)	Probability Intervals	Logarithmic Returns(%)	Probability Intervals
-9.7013	[0.0000, 0.0004]	-0.6469	(0.2581, 0.3225]
-7.5680	(0.0004, 0.0009]	-0.4061	(0.3225, 0.3942]
-6.0472	(0.0009, 0.0013]	-0.1662	(0.3942, 0.4689]
-5.7701	(0.0013, 0.0021]	0.2990	(0.4689, 0.6254]
-5.0495	(0.0021, 0.0047]	0.7092	(0.6254, 0.7462]
-4.9325	(0.0047, 0.0068]	1.2521	(0.7462, 0.8272]
-4.2105	(0.0068, 0.0094]	1.7069	(0.8272, 0.8963]
-3.9037	(0.0094, 0.0162]	2.1166	(0.8963, 0.9411]
-3.7389	(0.0162, 0.0243]	2.6151	(0.9411, 0.9637]
-3.4469	(0.0243, 0.0320]	3.1679	(0.9637, 0.9787]
-3.1746	(0.0320, 0.0405]	3.6914	(0.9787, 0.9889]
-2.8805	(0.0405, 0.0503]	4.0963	(0.9889, 0.9940]
-2.6256	(0.0503, 0.0636]	4.6782	(0.9940, 0.9957]
-2.4176	(0.0636, 0.0768]	5.1052	(0.9957, 0.9966]
-2.0946	(0.0768, 0.0939]	5.9009	(0.9966, 0.9974]
-1.9025	(0.0939, 0.1152]	6.8791	(0.9974, 0.9983]
-1.6342	(0.1152, 0.1387]	7.1483	(0.9983, 0.9991]
-1.3795	(0.1387, 0.1681]	8.5236	(0.9991, 0.9996]
-1.1172	(0.1681, 0.2056]	14.2617	(0.9996, 1.0000]
-0.8740	(0.2056, 0.2581]		

After calculating the logarithmic return series from the modeling set, the empirical distribution function is derived. To make the simulation more effective, the logarithmic return values are grouped into bins. Values with similar magnitudes are aggregated and represented by their mean. The probability assigned to each bin is determined by the proportion of observations it contains relative to the total sample size. This procedure yields a discrete probability distribution model, in which each possible logarithmic return value r_i is associated with a sampling probability p_i, satisfying $\sum p_i = 1$. Based on that, corresponding cumulative probability intervals can be constructed for the random sampling in subsequent simulations, see Table 5. Note that maintaining consistency between the statistical properties of simulation series and historical data is essential for Monte Carlo method. Namely the estimators of simulation series should coincide with those of historical

ISSN 2959-6157

records. Hence when the model is applied for validation and forecast, if the mean logarithmic return of the predicted path deviates significantly from that of the historical data, the prediction is repeated until the two are approximately equal.

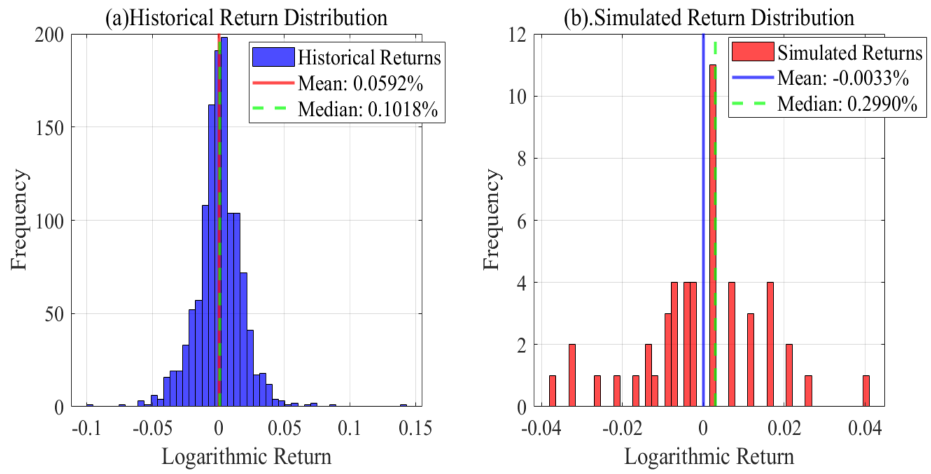

Fig. 3 Analysis of logarithmic return distribution

3.3 Results of Simulation

The model obtained through the above process performs well in the testing set. It demonstrates good predictive accuracy during the out-of-sample validation period, with an average relative error of 2.81%. Meanwhile, the following histogram (see Fig. 3) graphically presents the distribution of returns of the simulated price path in the validation period. By comparing sub figure (a) and (b) of the histogram, it can be seen that the distribution of returns in the whole validation period are similar to the distribution of historical returns. That is, the simulation process can achieve the statistical features of the historical returns, and the effectiveness of simulation has been verified.

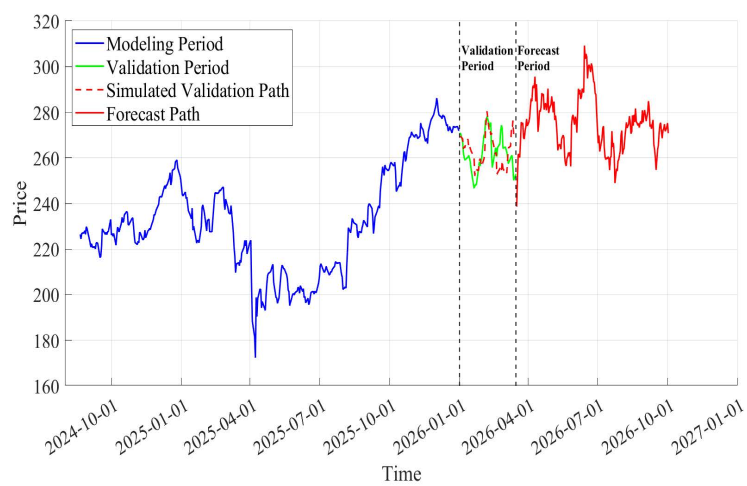

Fig. 4 Prediction results of Monte Carlo method

The Monte Carlo model's forecast of AAPL's daily closing price for the next 200 trading days (starting from March 17, 2026) is shown in the following Fig. 4. It can be observed that results obtained by Monte Carlo method based on logarithmic returns can also reproduce the overall trend of AAPL closing price. In addition, the generated price path exhibits short-term irregular fluctuations similar to those observed in real market conditions. Nevertheless, it should be noted that these results are derived from random sampling, and the distribution of logarithmic returns rely entirely on historical data. Therefore, the results should be interpreted with caution. The price path can only serve as a reference and should not be used as the solid basis for investment decisions.

1. Conclusion

Based on the above empirical experiments, following conclusions are drawn. For time series analysis, the model built in this paper is effective in capturing the trend and cyclical parts of time series. With relatively small short-term forecasting errors, this can generally predict the overall price trend accurately. However, this model cannot capture the irregular fluctuation of financial time series. The Monte Carlo method simulates the future price paths of stocks according to the probability distribution obtained from historical data. The method captures the overall trend and uncertainty of stock price for the forecast horizon. However, the simulation results strongly depend on historical data, which will greatly reduce prediction accuracy under extreme market circumstances. These two methods are different in their predictive goal and applicable scope. The time series analysis is suitable for precise forecasting and short-term decision-making support. The Monte Carlo simulation stresses the probabilistic characterization and is more applicable to medium and long term risk assessment. Over the same study period, the predicted trend from the two methods is fairly consistent. This means that despite their different underlying principles, the two methods could complement each other to describe the evolution trend of the electricity price.

Future research may adopt the two methods together in hybrid models. For instance, probability intervals from Monte Carlo simulation could function as a correction of error for time series models in order to strike a balance of prediction accuracy and risk control. Apart from that, future work may also look into incorporating more variables into the existing method, thus enriching the data. For the decomposition model of the time series, variables such as industry indexes can be added as the exogenous variables in the trend component, periodic component or random component. For Monte Carlo simulation, variable like VIX index or market attitude indicators can be used to calibrate the estimated probability distribution. The incorporated information may improve both forecast accuracy and robustness.

References

[1] Gui J, Naktnasukanjn N, Yu X, et al. Research on the impact of economic policy uncertainty and investor sentiment on the growth enterprise market return in China—An empirical study based on TVP-SV-VAR model. International Journal of Financial Studies, 2024, 12(4): 108.

[2] Darwish M, Hassanien E E, Eissa A H. Stock market forecasting: From traditional predictive models to large language models. Computational Economics, 2025, doi: 10.1007/s10614-025-11024-w.

[3] Wang S, Wu H, Shi X L, et al. TimeMixer: Decomposable multiscale mixing for time series forecasting. arXiv, 2024, arXiv:2405.14616.

[4] Nagashima S, Suzuki S, Koyama S, et al. A decomposition-based state space model for multivariate time-series forecasting. arXiv, 2026, arXiv:2602.05389.

[5] An Z, You J, Li J, et al. FreDN: Spectral disentanglement for time series forecasting via learnable frequency decomposition[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2026, 40(24): 19623-19631.

[6] Kamila I, Anggraeni F, Rumengan N R, et al. Application of Monte Carlo simulation in predicting stock prices at PT Bank Syariah Indonesia Tbk. J-KOMA: Jurnal Ilmu Komputer dan Aplikasi, 2025.

[7] Hainaut D, Akbaraly A. Risk management with local least squares Monte Carlo. ASTIN Bulletin, 2023, 53(3): 489-514.

[8] Singh S, Parmar K S, Kumar J. Development of multi-forecasting model using Monte Carlo simulation coupled with wavelet denoising-ARIMA model. Mathematics and Computers in Simulation, 2024.

[9] Polanco C. Mathematical foundations of time series analysis: Theory and cases. Bentham Books, 2025.

[10] Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control, 1974, 19(6): 716-723.

[11] Tsay R S. Analysis of financial time series. 3rd ed. Wiley, 2010.

[12] Dai M, Jin H, Kou S, et al. A dynamic mean-variance analysis for log returns. Management Science, 2020, 67: 1093-1108.

[13] Said S E, Dickey D A. Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika, 1984, 71(3): 599-607.