Time Series Modeling of the S&P 500 And Quantitative Trading Strategy Optimization

Cirui Wang *

University of California, Berkeley, College of Computing, Data Science, and Society, Berkeley, CA, 94704, United States

* Corresponding Author Email: rjzjbd@berkeley.edu

Abstract. The stock market trend is an important indicator for evaluating the overall economy and the state of the financial market. The S&P 500, as one of the main stock indexes, is widely used by investors, fund managers, and policymakers. Using daily data of the S&P 500, this paper builds an ARIMA model and compares it with the simple Buy & Hold strategy to examine how useful it is for trading. Although the ARIMA model has a high mean absolute percentage error (MAPE), it performs better than Buy & Hold in total return and Sharpe ratio, suggesting higher risk-adjusted returns. However, it shows little progress in controlling volatility and maximum drawdown, which means the model has limits when the market is unstable. This study shows both the value and the limits of using a classic time series model like ARIMA for financial prediction. The results also provide a basis for later work to build stronger hybrid models.

Keywords: S&P 500; ARIMA Model; Time Series Forecasting; Trading Strategy; Financial Modeling.

1. Introduction

1.1. Research Background

The Standard & Poor's 500 (S&P 500) is one of the key indexes of the U.S. stock market. It covers 500 major companies from many industries, so it reflects both market trends and the wider economy. Because of its scope and impact, it is closely observed by investors, policymakers, and firms. Forecasting the S&P 500 has high value: it helps investors design trading strategies, guides asset managers in portfolio allocation, and signals possible economic downturns. Even small gains in forecast accuracy can bring large financial benefits for big portfolios and support better decisions during unstable periods. Movements of the index also influence retirement funds, consumer confidence, and economic planning. Therefore, reliable prediction models can reduce uncertainty, improve financial outcomes, and contribute to overall economic stability.

1.2. Literature Review

The ARIMA model is often used in financial forecasting. Ariyo, Adewumi, and Ayo studied its use in stock price prediction and pointed out the need for a clear process in choosing the right model. They used methods like the Bayesian Information Criterion (BIC) to compare models, residual checks to test fit, and differencing to make the data stable. This gave a simple and structured way to set parameters for ARIMA in financial work [1]. Benvenuto et al. showed with COVID-19 case data that ARIMA can track short-term changes in complex time series. Their results show that ARIMA works well with unstable data and is still useful for short-term forecasting, such as predicting S&P 500 returns [2]. This shows that ARIMA is a strong option for modeling the uncertain moves of the S&P 500. Millantika used ARIMA in portfolio planning by making return forecasts for Jakarta Islamic Index stocks and adding them to a mean-variance model. The study found that the ARIMA-based portfolio had slightly better weekly returns, lower risk, and a higher Sharpe ratio. These results show that using forecasting models in portfolio building can improve risk-adjusted returns [3].

While the existing literature, as discussed, demonstrates the efficacy of ARIMA models in generating accurate point forecasts and even hints at their utility in portfolio construction, a significant gap remains. Most research focuses on generating single numeric forecasts with ARIMA or other time series models in financial markets. However, relatively few works extend these predictions into optimized quantitative strategies that are further integrated into portfolio optimization.

1.3. Research Framework

This research follows a simple and straightforward research framework. First, the daily closing prices of the S&P 500 index are preprocessed and modeled using the ARIMA approach, with the data divided into training and testing sets. The fitted ARIMA model is then used to generate short-term forecasts of the index price. Second, these forecasts are used to set a quantitative trading signal: These forecasts are converted into a trading signal: a predicted positive return generates a long position, while a predicted non-positive return results in closing the position and holding cash. The strategy is then tested on historical data to see how it performs, mainly by looking at overall returns and the Sharpe ratio. Finally, the performance of this ARIMA-based trading strategy is rigorously benchmarked against a passive buy-and-hold strategy on the S&P 500 index.

2. Method

In this research, all analyses are implemented in R. The data used in this study were obtained from the Federal Reserve Bank of St. Louis (FRED) database, specifically the S&P 500 Index (code: SP500). It is measured in index points, with a daily frequency based on closing prices, up to August 13, 2025 [4]. The raw index price series was converted into a log-return series for analysis and modeling, as financial returns are typically stationary and more amenable to time series modeling. The research adopts the time series approach, with a focus on the ARIMA model as the main forecasting tool. The analysis begins with visual inspection and stationarity tests of the raw data series to decide whether differencing is required. Next, the autocorrelation and partial autocorrelation functions are used to provide the choices for model orders. Information criteria, including AIC and BIC, are verified to select the optimal parameter combination. The residual diagnostics check the fitted model to ensure that the residuals behave like white noise.

At the forecasting stage, the model produces the short-term predicted values of the index price, and thus the predicted returns by computing log differences of the forecasted index levels. These forecasts are directly translated into trading signals, with a positive predicted return indicating a long position and a predicted non-positive return resulting in an exit from the market position (holding cash). The Sharpe ratio is a measure of the risk-adjusted performance, which reflects the excess return earned per unit of risk [5]. The Sharpe ratio is selected as the main evaluation metric in evaluating the performance of this ARIMA-based strategy.

In this research, the period from August 14 to August 21, 2025, served as the test set for the ARIMA model. Based on these forecasts, the trading strategy was constructed, and its cumulative return was compared with a simple buy-and-hold measure. The outcomes assess whether the forecasting model offers improvements in investment performance.

3. Empirical Results and Analysis

3.1. Arima Model Result Analysis

3.1.1 Time series characteristics

The initial step in ARIMA modeling involves diagnosing the stationarity of the time series. Figure 1 illustrates that the S&P 500 closing price shows a long-term upward trend, with a certain level of volatility [6]. The autocorrelation and partial autocorrelation functions for the original price series display strong, gradually diminishing correlations, confirming its non-stationary nature. This observation aligns with the typical behavior of stock indices.

First-order differencing was applied to address the non-stationarity. This transformation successfully removes the underlying stochastic trend. The autocorrelation and partial autocorrelation functions for the differenced series are significantly weaker, with most lag terms falling within the significance bounds. These results indicate that the differenced S&P 500 series can be considered stationary, a prerequisite for applying an autoregressive integrated moving average (ARIMA) model. Figure 1a and Figure 1b demonstrate that this transformation effectively removes the trend, thereby improving the suitability of the data for modeling.

Fig. 1a S&P 500 Stock Price (Photo Credit: Origin)

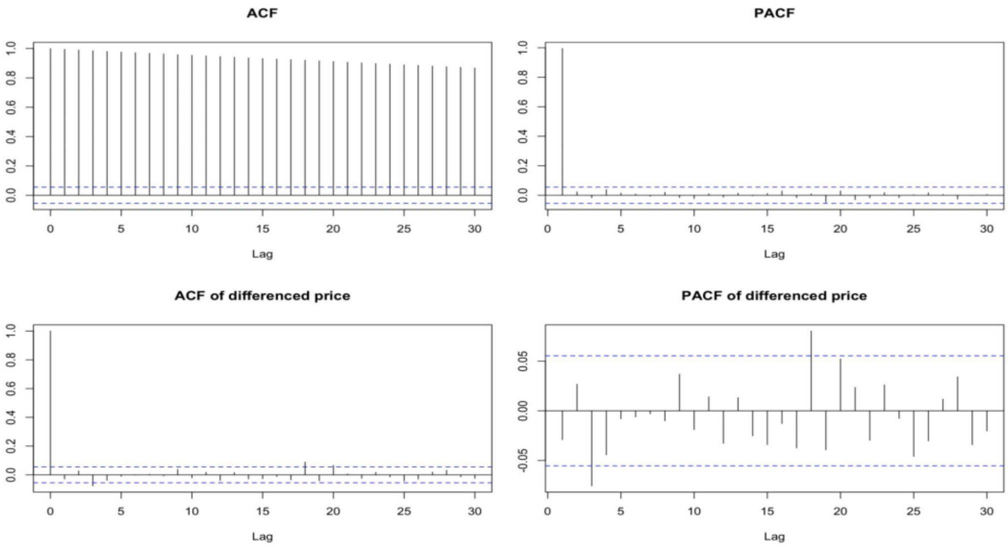

Fig. 1b ACF and PACF Plots of Original and Differenced Series (Photo Credit: Origin)

3.1.2 Model selection

Based on the differenced data, the ARIMA model automatically selected the ARIMA (0,1,4) form with a drift term as its optimal one. Table 1 shows the error metrics and residual diagnostics for this model.

First, from the perspective of residual bias, the mean error is only 0.0056, and the mean percentage error is also close to zero, indicating that the model does not systematically overestimate or underestimate prices. Second, the p-value of the Ljung–Box test is 0.9116, which is far above the commonly used significance level, indicating that the residual sequence can be considered white noise. Meanwhile, the first-order autocorrelation is nearly zero, further confirming that there is no significant correlation between the residuals. In terms of forecast accuracy, based on the root mean square error, which is approximately 49, the average absolute error, which is 35, and the average absolute percentage error, which is as high as 78.6%, this model is statistically valid but still limited in terms of predictive accuracy. An Absolute Percentage Error of this magnitude is common when forecasting volatile financial returns, but indicates that the forecasts should be interpreted with caution and are likely more useful for discerning directionality (i.e., the sign of the return) than precise price levels.

The Ljung–Box test (p = 0.9116) fails to reject the null hypothesis of white noise, and the first-lag autocorrelation (ACF1 = 0.0003) is nearly zero, confirming that the residuals are uncorrelated.

Table 1. Forecast Accuracy and Diagnostic Statistics

ME	-0.005583955
RMSE	49.27182
MAE	35.33415
MPE	-0.01016453
MAPE	0.7858633
MASE	1.001203
Ljung-Box test (p-value)	0.9116
ACF1	0.0002845363

3.1.3 Forecasting result analysis

Fig. 2 ARIMA Forecasting Result (Photo Credit: Origin)

Figure 2 presents the ARIMA-based forecast of the S&P 500 closing price for the next 14 trading days. The forecast extrapolates the historical upward trend, with the predicted mean following a persistent upward trajectory. The blue forecast line follows an upward path, displaying smooth and continuous growth. In the short term, the predicted values closely align with recent trends, indicating reasonable model performance. As this model predicts further into the future, the prediction intervals (PIs) diverge, reflecting the increasing uncertainty inherent in the forecasts as the forecast horizon extends. This shows that uncertainty increases over longer periods, which therefore reduces the reliability of long-term forecasts. Consequently, while the model is capable of generating short-term trend forecasts, its utility for long-term prediction is limited, as it cannot precisely capture future market volatility and abrupt fluctuations.

3.2. Performance Evaluation: ARIMA Trading Strategy vs. Buy-and-Hold

To assess the significance of the forecast results in actual investments, this research adopts a simple trading strategy, which was compared with the buy-and-hold measure. Figure 3 shows the cumulative return performance of the two strategies during the test period. Both curves generally trend upward, reflecting the overall upward trend of the S&P 500. However, it is clear that the trading strategy based on ARIMA signals outperforms the buy-and-hold strategy. This suggests that despite its limitations in point forecasting accuracy (as indicated by the high MAPE), the model captures enough directional information to generate economically valuable trading signals.

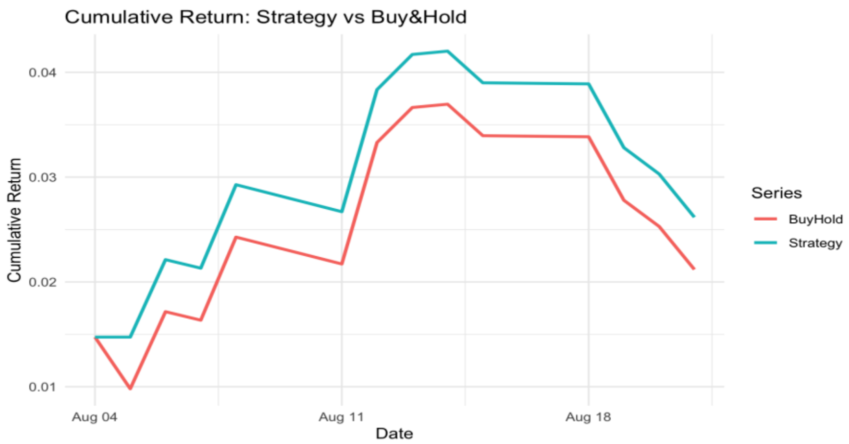

Fig. 3 Cumulative return comparison between the proposed strategy and the Buy-and-Hold (Photo credit: Origin)

Table 2 reports the detailed performance measures. The ARIMA strategy achieved a cumulative return of 0.0262, higher than the buy-and-hold return of 0.0212. At the same time, its volatility was 0.00616, slightly lower than the 0.00640 of the buy-and-hold strategy, suggesting that the ARIMA strategy provided higher returns without increasing risk. The maximum drawdown was the same for both strategies, meaning that the worst losses during the sample were similar. The Sharpe ratio shows the most obvious difference. The ARIMA strategy reached 4.81, while the buy-and-hold strategy reached 3.76. This indicates that the ARIMA approach generated higher risk-adjusted returns, yielding a greater return per unit of risk.

Table 2. Performance Metrics of ARIMA Strategy vs. Buy & Hold

Model	Cumulative Return	Volatility	Maximum Drawdown	Sharpe Ratio
ARIMA Strategy	0.02617125	0.006159356	-0.01520745	4.805531
Buy and Hold	0.02118624	0.006404436	-0.01520745	3.761561

The model's utility can explain this divergence between high forecast error and strong strategy performance in predicting return direction rather than magnitude. For the trading rule employed, which depends solely on the sign of the forecasted return, accurately predicting the sequence of positive and negative return days is more critical than precisely forecasting the price level. The results imply that the ARIMA model, while flawed for point forecasts, retains significant value for binary classification (up/down) in this context.

Overall, the results suggest that even though the ARIMA model is limited in making precise forecasts, it can still generate useful trading signals and improve returns compared to a simple passive strategy over a short time horizon.

4. Discussion and Implications

4.1. ARIMA Model Predictive Power and Limitations

The empirical analysis reveals a nuanced picture of the ARIMA model's utility: while it generates statistically significant trading signals that improve risk-adjusted returns (Section 3.2), its practical predictive power is constrained by several fundamental limitations. The results in section 3.2 indicate that the ARIMA model can capture short-term patterns in financial time series, as reflected in the higher Sharpe Ratio of the ARIMA-based strategy compared with the Buy-and-Hold strategy. This suggests that ARIMA is able to improve the returns relative to risk in the short run. Nevertheless, the inherent characteristics of equity markets, which are full of noise, volatility, and semi-efficiency, limit their predictability. In particular, while ARIMA captures short-term patterns, its long-term advantage is not evident. Based on Table 2, the model does not substantially outperform in terms of volatility reduction and drawdown control. This observation aligns with Fama's efficient market hypothesis, which argues that prices fully reflect available information, leaving little room for consistent predictability, making it difficult for statistical models to generate continued excess returns [7].

4.2. Relating ARIMA to Established Trading Approaches

Based on Table 2, when focusing solely on volatility, the difference between the two strategies is relatively minor, and the maximum drawdown experienced under the ARIMA model is only slightly reduced. These results imply that the ARIMA strategy enhances performance primarily through minor adjustments rather than major risk control. In other words, while the strategy improves the return-to-risk tradeoff, it still leaves investors exposed to high risks when the market falls.

This outcome—improved risk-adjusted returns without a radical reduction in absolute risk exposure—is not unique to ARIMA-based strategies. This research's findings are consistent with evidence from Gurrib, who shows that an optimized moving average crossover strategy yields lower absolute returns than Buy-and-Hold but achieves a higher Sharpe ratio by reducing drawdowns [8]. The parallel suggests that improvements in risk-adjusted performance need not rely solely on forecasting models such as ARIMA. Instead, even simple timing rules can deliver similar benefits by limiting exposure to large losses. For our study, this comparison highlights that ARIMA contributes primarily through predictive power, whereas moving average crossovers achieve gains through timing discipline, pointing to complementary pathways for enhancing portfolio outcomes.

4.3. Optimization Pathways and Future Improvements

Notwithstanding its utility, the ARIMA model's limitations in capturing market nonlinearities and structural breaks, as noted in Section 4.1, present clear opportunities for improvement. The ARIMA model assumes linear relationships and relies on fixed parameters, thus restricting its ability to capture sudden market structural changes, nonlinear patterns, or extreme market events that often symbolize turbulent markets. As a result, forecasts may systematically differ from actual prices, leading to poor risk control and higher forecast errors during periods of instability. This drawback suggests that the pricing strategy should move beyond fixed historical parameters based solely on the ARIMA model. Instead, it needs to adopt flexible methods that can adapt adeptly to market shifts.

While the ARIMA framework demonstrates predictive value, other researchers have suggested that improvements in the prediction accuracy often come from combining forecasting with risk management. Moreira and Muir show that simple volatility-managed portfolios, where trade amounts are reduced when volatility is high and increased when volatility is low, can largely improve Sharpe Ratios and reduce drawdowns [9]. This suggests that combining ARIMA forecasts with a volatility-based risk control allows a trading strategy ensuring profits during stable periods while reducing exposure in unstable periods. Another promising aspect is the development of hybrid models that integrate ARIMA with machine learning components. Recent research suggests that hybrid ML-TS models that use ARIMA residuals as features in the LSTM model can outperform either model alone, particularly when market conditions vary randomly [10]. Although such an extension is beyond the scope of the current study, it represents a logical step toward addressing the limitations observed in the ARIMA method under volatile market conditions.

5. Conclusion

This study demonstrates that while autoregressive integrated moving average (ARIMA) models are statistically valid and can generate economically valuable trading signals for the S&P 500, their practical application is circumscribed by inherent limitations in forecasting precision and risk management. This research analyzes the performance of an ARIMA-based trading strategy against a simple Buy & Hold strategy for forecasting the S&P 500 returns. The results show that the ARIMA strategy generates marginally higher cumulative returns and a notably improved Sharpe ratio, displaying better performance compared to the Buy & Hold strategy. However, volatility reduction is modest, and maximum drawdown remains unchanged, suggesting that the ARIMA model does not provide enhanced downside protection during periods of market stress. As discussed, these findings highlight both the strengths and the limitations of the ARIMA-based strategy: It captures short-term trends effectively in stable regimes but is susceptible to significant forecasting errors and offers limited risk control during periods of high volatility and structural market breaks. Future improvements may involve adding adaptive parameters or machine learning techniques that can dynamically respond to changes in market conditions and help enhance models' robustness and ability of risk controls.

This study contributes to the field of financial investments by showing the effectiveness of an ARIMA-based forecasting strategy in predicting S&P 500 returns. Accurate forecasting of stock market movements has both business and social value. For investors and financial institutions, ARIMA-based forecasts can inform portfolio allocation decisions and improve profits. From a broader social perspective, enhancing forecasting accuracy also contributes to the overall market efficiency by reducing information asymmetry.

By identifying the limitations of the ARIMA model under volatile market conditions, this research discovers opportunities for combined models that integrate the more nuanced statistical and machine learning approaches.

References

[1] Adebiyi Ayodele Ariyo, Adewumi Aderemi Oluyinka, Ayo Charles Korede. Stock Price Prediction Using the ARIMA Model/ 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation. Cambridge, UK: IEEE, 2014: 106-112. DOI:10.1109/UKSim.2014.67.

[2] Benvenuto Domenico, Giovanetti Marta, Vassallo Lazzaro, Angeletti Silvia, Ciccozzi Massimo. Application of the ARIMA model on the COVID-2019 epidemic dataset. Data in Brief, 2020, 29: 105340. DOI: 10.1016/j.dib.2020.105340

[3] Salwa Cendikia Millantika. Portfolio Optimization by Considering Return Predictions Using the ARIMA Method on Jakarta Islamic Index Sharia Stocks. International Journal of Quantitative Research and Modeling, 2025, 6(2): 248-254.

[4] Ta, Van-Dai, Chuan-Ming Liu, and Direselign Addis Tadesse. Portfolio optimization-based stock prediction using long-short term memory network in quantitative trading. Applied Sciences, 2020, 10(2): 437.

[5] Wang, Jujie, Zhenzhen Zhuang, and Liu Feng. Intelligent optimization based multi-factor deep learning stock selection model and quantitative trading strategy. Mathematics, 2022, 10 (4): 566.

[6] Michańków, Jakub, Paweł Sakowski, and Robert Ślepaczuk. Hedging Properties of Algorithmic Investment Strategies using Long Short-Term Memory and Time Series models for Equity Indices. arXiv preprint arXiv:2309.15640, 2023.

[7] Fama, Eugene Francis. Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance, 1970, 25(2): 383-417. DOI:10.2307/2325486.

[8] Gurrib Ikhlaas. The Moving Average Crossover strategy: Does it work for the S&P500 Market Index. SSRN Electronic Journal, 2015. DOI:10.2139/ssrn.2578302.

[9] Moreira A, Muir T. Volatility-Managed Portfolios. Journal of Finance, 2017, 72(4): 1611-1644. doi:10.1111/jofi.12511.

[10] Kamil Kashif, Robert Ślepaczuk. 2024. LSTM-ARIMA as a Hybrid Approach in Algorithmic Investment Strategies [EB/OL]. (2024-06-26) [2025-09-12]. arXiv: 2406. 18206. DOI:10.48550/ arXiv.2406.18206. https://arxiv.org/abs/2406.18206