The Application of ARIMA Model in Short-Term Stock Price Prediction

Yifan Yang

Affiliation: Jinan New Channel -JUTES High School, Jinan ,Shandong, 250000 , China
Email: 3030723427@qq.com

Abstract

The stock market holds a crucial position in the financial system. Stock price prediction is of great significance to both investors and financial institutions. However, due to its complexity and uncertainty, precise prediction is quite challenging. The ARIMA model, as a classic time series analysis method, performs exceptionally well in handling time series data with trends and seasonality. This article focuses on its application in short-term stock price prediction, elaborates on the model principle and modeling steps, analyzes the factors influencing the prediction, and explores strategies to enhance the prediction effect. The results show that the ARIMA model has certain value in short-term prediction, but it needs to be improved and optimized. Finally, summarize the research and look forward to the future direction.

Keywords: ARIMA model; Stock price; Short-term forecast; Influencing factors; Optimization strategy.

1. Introduction

1.1 Research Background and Significance

The stock market is an important component of the financial market. It provides enterprises with a crucial financing channel, helps them expand production, innovate research and development, and promotes the development of the real economy. At the same time, it also opens up channels for asset allocation and wealth appreciation for investors. Investors can share the growth dividends of enterprises by buying and selling stocks, achieving the preservation and appreciation of assets. [1]

Stock price fluctuations are influenced by numerous factors. Macroeconomic conditions such as GDP growth rate, inflation rate and interest rate can change the business environment of enterprises and the expectations of investors, and thus affect stock prices. For instance, when interest rates rise, the financing costs of enterprises increase, profits may decrease, and investors' demand for stocks will also decline, leading to a fall in stock prices [2]. The company's own operating conditions, including its financial status, profitability, market share, etc., directly determine the intrinsic value of its stocks. A company with a sound financial condition and strong profitability usually has a relatively high stock price [3]. Industry development trends, such as technological innovation and policy changes, will reshape the competitive landscape of the industry and affect the development prospects of enterprises within it. For instance, with the continuous development of new energy vehicle technology, the traditional fuel vehicle industry is facing huge challenges, and the stock prices of related enterprises have also been significantly affected [4]. In addition, market sentiment and investor psychology can also have a significant impact on stock prices in the short term. When market sentiment is optimistic, investors tend to buy stocks, driving up prices. When market sentiment is pessimistic, investors sell their stocks one after another, causing prices to fall [5].

For investors, accurately predicting stock prices is the core of formulating reasonable investment strategies, controlling investment risks and achieving the maximum investment returns. If one can predict in advance the rise in stock prices, they can buy at the right time and sell for profit when the prices go up. If a decline is predicted, sell in time to avoid losses. For financial institutions, accurate prediction is conducive to making decisions such as asset pricing, risk management and resource allocation, and enhancing operational efficiency and competitiveness [6].

However, stock prices are highly complex and uncertain, presenting nonlinear and non-stationary characteristics. Their changes are difficult to describe with simple linear relationships, and their statistical features change over time, making precise prediction extremely challenging [7]. The nonlinear feature is reflected in the fact that the changes in stock prices do not follow a fixed proportional relationship. For instance, during periods of market prosperity, stock prices may show an accelerating upward trend. During periods of market downturn, the pace of price decline may also accelerate. The non-stationary feature is manifested in the fact that the mean and variance of stock prices change significantly over time, which makes it difficult for traditional statistical models to accurately capture their changing patterns.

1.2 Current Research Status at Home and Abroad

Research on stock price prediction started early abroad, and in the early days, it was mainly based on fundamental analysis and technical analysis. Fundamental analysis assesses the intrinsic value of stocks and predicts price trends by studying a company's financial statements, industry position, macroeconomic environment, etc. For instance, the value investment concept proposed by Stott et al. emphasizes determining the intrinsic value of a stock by analyzing its fundamentals, thereby making investment decisions [8].

Domestic research started relatively late, but it has achieved rich results. Domestic scholars draw on foreign experience, combine the characteristics of the Chinese stock market, and conduct a large number of empirical studies using the ARIMA model and other improved models. Due to the imperfect development of China's stock market, there exist problems such as information asymmetry and large market fluctuations. The prediction accuracy of the traditional ARIMA model is limited. However, by combining with other methods or improving the model, the prediction accuracy can be enhanced.

2. Theoretical Basis of the ARIMA Model

2.1 Basic Concepts of Time Series

A time series is a sequence of data arranged in chronological order, reflecting the changes of phenomena or variables over time. Such as stock prices, monthly sales volume, annual temperature changes, etc. Time series usually have the characteristics of trend, seasonality and randomness.

Trendiness refers to the long-term upward or downward trend of a time series. When the economy is doing well, the stock prices of some companies may rise for a long time. Stocks in traditional industries may decline in the long term due to technological substitution. Seasonality refers to the periodic changes in a time series at fixed time intervals. For instance, the sales of the retail industry increase significantly during holidays. Randomness is the fluctuation in a time series that cannot be explained by trend or seasonality and is caused by accidental factors.

2.2 Composition of the ARIMA Model

The ARIMA model consists of three parts: autoregressive (AR), moving average (MA), and differential integration (I).

The autoregressive part is based on the idea that there is a linear relationship between the current moment value and several past moment values. In the stock market, the current stock price may be influenced by the price of the previous few days. If the price kept rising in the previous few days, the current price may also have an upward trend.

The moving average partly takes into account the influence of random errors. Stock price changes are not only influenced by past prices but also disturbed by random factors, such as breaking news and changes in market sentiment. The moving average model describes the current moment value through a linear combination of random errors.

The purpose of differential integration is to transform non-stationary time series into stationary time series. The statistical characteristics of non-stationary time series change over time, such as changes in mean or variance, making it difficult to predict accurately. The statistical characteristics of stationary time series remain constant, making them easier to analyze and predict. Eliminate the trend through difference operation to make the sequence meet the stationarity requirements. If the first-order difference is performed on a sequence with an upward trend, subtracting the value at the previous moment from the current value, the new sequence may no longer show a clear upward trend but fluctuate around a constant.

2.3 Modeling Steps of the ARIMA Model

2.3.1 Data Stationarity Test

Before modeling, it is necessary to verify whether the time series data is stable. The mean and variance of stationary time series are constant, and the autocovariance is only related to the time interval. Common methods include the unit root test (ADF test), etc. If the sequence is not stationary, it needs to be transformed into a stationary sequence through difference operations. If the ADF test shows that the original sequence is not stationary, perform the first-order difference and test again until a stationary sequence is obtained.

2.3.2 Model Recognition

Determining the autoregressive order (p), difference order (d), and moving average order (q) in the ARIMA model is crucial. The values of p and q can be initially determined by observing the truncation properties of the time series autocorrelation function (ACF) and the partial autocorrelation function (PACF). The autocorrelation function describes the correlation between different time values, while the partial autocorrelation function measures the correlation between two specific time values after controlling the influence of intermediate time values. If the autocorrelation function is truncated after lagging by a certain order, and the partial autocorrelation function is tailing, it may be suitable for the MA model. Conversely, truncating the partial autocorrelation function and tailing the autocorrelation function might be suitable for AR models. At the same time, the optimal model order is selected in combination with information criteria (such as AIC criterion, BIC criterion). The information criteria comprehensively consider the goodness of fit and complexity of the model, and the smaller the value, the more suitable the model is.

2.3.3 Model Checking

Check whether the residual sequence of the model is a white noise sequence. The mean of the white noise sequence is 0, the variance is constant, and the values at different times are not correlated. If the residual sequence is a white noise sequence, it indicates that the model has fully extracted the sequence information and is suitable. Otherwise, corrections are required. Commonly used methods include Q-test, etc.

2.3.4 Prediction

After the model passes the test, it can be used to predict future time series values. Prediction is divided into point prediction and interval prediction. Point prediction provides specific predicted values for future moments, while interval prediction offers confidence intervals for the predicted values, reflecting the uncertainty of the prediction.

3. Factors Influencing the Short-term Prediction of Stock Prices by the ARIMA Model

3.1 Nonlinear Characteristics of the Stock Market

The stock market has significant nonlinear characteristics, and the changes in stock prices are not simply linear relationships. For instance, a sudden shift in market sentiment may lead to significant fluctuations in stock prices, and such fluctuations are difficult to accurately describe with linear models. When major positive news emerges in the market, investor confidence surges, and a large amount of capital flows into the stock market, driving stock prices to rise rapidly. Moreover, the extent of the increase may exceed the linear prediction based on historical data. However, as a linear model, the ARIMA model has limitations when dealing with such nonlinear relationships and may fail to accurately capture the sudden changes in stock prices, thereby affecting the accuracy of predictions.

3.2 Impact of External Events

The stock market is vulnerable to external events, such as adjustments in macroeconomic policies, the occurrence of political events, and natural disasters. These events are often sudden and unpredictable, which can have a significant impact on stock prices. For instance, when the government introduces new monetary policies, adjusts interest rates or the reserve requirement ratio, it will alter the supply and demand of funds in the market, and thereby affect stock prices. If the ARIMA model does not take into account the influence of these external events during the modeling process or fails to incorporate these factors into the model in a timely manner, the prediction results will have significant deviations.

3.3 Quality and Completeness of Data

The quality and completeness of the data are crucial to the prediction effect of the ARIMA model. If there are missing values, outliers or incorrect records in the stock price data, it will affect the accuracy and stability of the model. For instance, missing data may cause the model to fail to accurately capture the changing trend of stock prices, and outliers may lead to deviations in the model's fitting of the data. In addition, the frequency of the data will also affect the prediction effect. If daily data, weekly data or monthly data are used for modeling, different prediction results may be obtained. The amount of information contained in data of different frequencies varies. High-frequency data may better reflect the short-term fluctuations of stock prices, but it may also be disturbed by more noise. Low-frequency data can better reflect the long-term trend of stock prices, but it may overlook some short-term changes.

3.4 Parameter Selection of the Model

The prediction effect of the ARIMA model largely depends on the selection of parameters p, d, and q. Different combinations of parameters can lead to significant differences in the model's fitting degree and predictive ability to the data. If the parameters are not properly selected, it may lead to overfitting or underfitting of the model. Overfitting models perform well on training data, but have poor predictive ability for new data. Underfitting models cannot fully capture the changing patterns of the data and have relatively large prediction errors. Therefore, how to select appropriate parameters based on the characteristics of stock price data is the key to improving the prediction accuracy of the ARIMA model.

4. Strategies for Enhancing the Short-term Stock Price Prediction Effect of the ARIMA Model

4.1 Combine with Other Models

Given the linear limitations of the ARIMA model, it can be combined with other nonlinear models to construct a hybrid model. For example, combined with neural network models. Neural network models possess powerful nonlinear mapping capabilities and can handle complex nonlinear relationships. By taking the linear prediction results of the ARIMA model as the input of the neural network model and further mining the nonlinear features in the data with the neural network model, the accuracy of the prediction can be improved. In addition, the support vector machine model has advantages in handling small samples and high-dimensional data. It can also be combined with the ARIMA model to leverage their respective strengths and enhance the prediction effect.

4.2 Introduce External Variables

To take into account the impact of external events on stock prices, external variables can be introduced into the ARIMA model. For instance, macroeconomic indicators (such as GDP growth rate, inflation rate, interest rate, etc.), industry indicators (such as industry prosperity index, etc.) or market sentiment indicators (such as investor confidence index, etc.) can be incorporated into the model as external variables. By establishing an ARIMA model that incorporates external variables, the factors influencing stock prices can be considered more comprehensively, enhancing the model's explanatory power and predictive accuracy for stock price changes. However, it should be noted that introducing external variables will increase the complexity of the model. It is necessary to select variables reasonably and handle the correlation issues among them.

4.3 Data Preprocessing and Feature Engineering

Before modeling, conducting thorough preprocessing and feature engineering on stock price data can enhance the quality of the model. For missing values in the data, methods such as interpolation, mean filling or deletion can be adopted for processing. Outliers can be identified and handled through statistical methods or model-based approaches. In addition, the information of the model can be enriched by extracting the features of the data. For instance, calculating technical indicators such as the volatility of stock prices, moving averages, and relative strength indicators, and using these indicators as input variables for the model, can help the model better capture the changing patterns of stock prices.

4.4 Optimize model Parameter Selection

Adopting a more scientific parameter selection method can enhance the prediction effect of the ARIMA model. In addition to the traditional information criterion methods, cross-validation methods, grid search methods, etc. can also be used. Cross-validation divides the dataset into a training set and a validation set. By conducting multiple trainings and validations to select the optimal parameter combination, it can more accurately assess the generalization ability of the model. The grid search rule is to conduct a comprehensive search within the given parameter range to find the parameter combination that optimizes the model performance. Through these methods, suitable parameters can be found more systematically, avoiding the randomness of parameter selection.

5. Conclusion

This paper explores the application of the ARIMA model in the short-term prediction of stock prices, analyzes the factors influencing the prediction and proposes strategies to enhance the prediction effect. The research results show that the ARIMA model, as a classic time series analysis method, has certain advantages in dealing with the non-stationarity of stock price time series. Through reasonable modeling steps, a relatively accurate prediction model can be established. In the short term, the ARIMA model can capture some of the changing patterns of stock prices, providing valuable reference information for investors, helping them grasp the trend of stock prices and make reasonable investment decisions.

However, the complexity and uncertainty of the stock market impose limitations on the ARIMA model in the prediction process. The nonlinear characteristics of the stock market, the impact of external events, data quality and integrity, as well as the selection of model parameters and other factors will all affect the accuracy of predictions. Therefore, it is necessary to constantly explore and improve the ARIMA model. By integrating other models, introducing external variables, conducting data preprocessing and feature engineering, as well as optimizing parameter selection and other strategies, the prediction effect of the model can be enhanced.

This study has certain deficiencies. In terms of theoretical analysis, although the ARIMA model and the factors influencing prediction have been deeply discussed, there is still a lack of more in-depth research on some complex nonlinear relationships and the influence mechanisms of external events. In terms of practical application, strategies to enhance the prediction effect have only been proposed from the theoretical level and have not been verified through a large number of empirical studies.

Future research can be expanded in the following aspects. In theoretical research, further explore the nonlinear characteristics of the stock market and the influence mechanism of external events to provide theoretical support for constructing more accurate prediction models. In practical applications, large-scale empirical research should be carried out to verify the effectiveness of the strategies proposed in this paper for improving prediction results and continuously optimize these strategies. In addition, with the continuous development of artificial intelligence and big data technologies, it is possible to explore the application of these new technologies in the field of stock price prediction, develop more intelligent and efficient prediction models, and provide more accurate and timely decision-making basis for investors and financial institutions.

References

[1] Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. Journal of financial economics, 116(1), 1-22.
[2] Mishkin, F. S. (2007). The economics of money, banking, and financial markets. Pearson education.
[3] Petty, J. W., Titman, S., Keown, A. J., Martin, P., Martin, J. D., & Burrow, M. (2015). Financial management: Principles and applications. Pearson Higher Education AU.
[4] Cha, H. J., & Lee, B. S. (2001). The market demand curve for common stocks: Evidence from equity mutual fund flows. Journal of Financial and quantitative Analysis, 36(2), 195-220.
[5] Shiller, R. J. (1981). Do stock prices move too much to be justified by subsequent changes in dividends?.
[6] AIM, A. I. M., & Investing, A. (2005). Investments.
[7] Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. Quantitative finance, 1(2), 223.
[8] Stott, B., Alsac, O., & Monticelli, A. J. (2005). Security analysis and optimization. Proceedings of the IEEE, 75(12), 1623-1644.