

Comparative Analysis of Linear Regression, Polynomial Regression, and ARIMA Model for Short-term Stock Price Forecasting

Hanlin Mo (Ankaa)^{1,a,*}

¹*College of Art and Science, Syracuse University, Syracuse, the U.S.*

a. hmo@syr.edu

**corresponding author*

Abstract: This research investigates the effectiveness of three prominent stock price prediction methodologies: Linear Regression, Polynomial Regression, and AutoRegressive Integrated Moving Average (ARIMA) model. The study leverages one and a half years of historical data from Apple, Tesla, Amazon, and Nike stocks to predict average prices over the ensuing 14 days. The predictive efficacy of each model is tested against actual data, revealing their respective strengths and limitations. Linear Regression offers an overview of stock trends with limited intricacy, while Polynomial Regression delivers a more comprehensive understanding of price variations and cyclical trends. However, Polynomial Regression's reliability for predictions remains uncertain. In contrast, the ARIMA model, designed explicitly for short-term forecasting, demonstrates superior accuracy, correctly predicting seven out of eight scenarios. It should be noted that this is despite its assumptions of linearity and stationarity. The findings underscore the complexity of accurate stock market prediction and highlight the ARIMA model's reliability for short-term forecasts. Therefore, understanding the strengths and weaknesses of each model is crucial for improving stock price prediction techniques and for better grasping the complex dynamics of volatile stock markets.

Keywords: stock price forecasting, linear regression, polynomial regression, ARIMA model

1. Introduction

Market analysts, investors, and academics worldwide share a collective objective, which is the ability to effectively predict future stock prices with precision. The significance of dependable forecasting cannot be overemphasized, considering its extensive implications for various stakeholders such as public and private investors, governments, and corporations. Accurate projections are essential for economic planners, risk managers, and individuals involved in investment decision-making processes. Accurate stock price forecasting, however, remains tricky, necessitating sophisticated techniques and procedures due to the complexity and volatility of stock markets.

Numerous prediction approaches and algorithms have been developed within this context, each possessing distinct advantages and disadvantages. The present study aims to investigate three widely used methodologies for forecasting stock prices, namely linear regression, polynomial

regression, and the Autoregressive Integrated Moving Average (ARIMA) model. The primary objective is to enhance understanding of these techniques by conducting a comprehensive analysis. The selection of these models was based on their proven track records in financial forecasting, their user-friendly nature, and their ability to provide detailed insights into stock price dynamics.

Stocks from four major companies will be the subject of this study: Apple (AAPL), Tesla (TSLA), Amazon (AMZN), and Nike (NKE). The selection criteria include market capitalization, trading volume, and representation across industries. This ensures high-quality data collection from which to draw conclusions.

To start, the linear regression analysis, a simple yet effective method for identifying patterns in stock price data, is to be conducted. However, linear models may fall short in capturing the highly nonlinear structure of financial markets. Therefore, the use of polynomial regression becomes a more suitable alternative. Polynomial regression has more predictive power because of its ability to account for complicated stock price changes and cyclical patterns. Finally, the ARIMA model is employed for conducting time series analysis. The ARIMA model is a suitable approach for short-term stock price prediction due to its ability to incorporate temporal price dynamics and the inherent volatility of financial markets. While these models may offer insights into a given issue, it is crucial to acknowledge their inherent imperfections. Acknowledging these constraints not only demonstrates our dedication to scholarly criteria, but also motivates us to persist in the pursuit of enhanced prediction models.

The research process encompasses a series of sequential steps, which include Subsequently, a comprehensive analysis will be conducted to compare linear regression, polynomial regression, and ARIMA models, along with their respective applications. Emphasis will be placed on evaluating the advantages and limitations associated with each approach. This review provides crucial contextual information and rationale for our study, situating our work within the broader academic discourse. Following a comprehensive examination of the existing body of literature, we proceed to elucidate our experimental methodology, encompassing the techniques and instruments employed in our investigation. This section provides a comprehensive overview of the research methodology employed, thereby establishing a foundation for the subsequent presentation of findings and conclusions.

2. Systematic Literature Review

2.1. Application of Linear Regression

The utilization of linear regression analysis in the finance industry for the purpose of forecasting future stock prices has been extensively practiced, as it serves as a fundamental component in various financial models.

The application of linear regression has been extensively employed in a multitude of research investigations within the field of finance. In their seminal work, Bhuriya et al. conducted a noteworthy empirical investigation in the field of asset pricing. The researchers employed the method of linear regression to establish a comprehensive three-factor model [1]. The researchers utilized market equity, book-to-market ratio, and company size as independent variables to predict forthcoming stock returns. This analysis demonstrates the efficacy of linear regression in capturing the relationship between multiple predictors and stock prices. Gururaj et al. conducted supplementary research wherein they utilized linear regression analysis with support vector machines to investigate the relationship between market-wide liquidity and stock prices [2]. The researchers made a significant finding that highlights the importance of liquidity in predicting variations in stock prices. This underscores the efficacy of linear regression in understanding the influence of market-wide factors on the prices of individual stocks.

In recent times, a substantial amount of scholarly investigation has been devoted to examining the incorporation of linear regression with other methodologies with the aim of augmenting its predictive capacities. Seethalakshmi conducted a study wherein a hybrid model integrating linear regression and artificial neural networks was utilized for the purpose of predicting stock market prices [3]. The study's results indicated that the hybrid model displayed superior performance in comparison to the conventional linear regression approach. The proposed study suggests a promising avenue for future research by seeking to capitalize on the strengths of linear regression while addressing its limitations through the integration of alternative methodologies. The process plays a crucial role as an initial phase in exploratory analysis and provides significant insights into the interconnections among variables.

The inherent simplicity of linear regression confers a notable advantage; however, it also engenders specific limitations. The limitation mentioned above is highlighted in the study conducted by Gururaj et al., wherein it is contended that stock markets often exhibit complex and nonlinear dynamics that cannot be effectively captured by linear models [4]. Furthermore, it is frequently observed that there are variations in the volatility of stock prices over a period, which is referred to as heteroscedasticity [5]. Linear regression models may face challenges in effectively addressing this issue.

In conclusion, linear regression has been instrumental in the field of financial modeling and has proven to be a valuable tool for predicting stock prices. Although it has certain limitations, the capacity to integrate with other methodologies implies its enduring significance in forthcoming financial market research. As we delve further into the realm of advanced models, it is crucial to acknowledge the fundamental significance of linear regression and the valuable insights it has provided us regarding the dynamics of financial markets.

2.2. Application of Polynomial Regression

Polynomial regression expands upon the functionalities of linear regression, enabling a more versatile approach to modeling the relationships between variables. This is achieved by incorporating polynomial terms into the regression equation, thereby capturing intricate and non-linear relationships that frequently manifest in real-world scenarios. This section provides a comprehensive literature review on the utilization of polynomial regression in the domain of stock price prediction. It examines the application, advantages, and constraints associated with this approach.

Within the domain of financial market analysis, polynomial regression has emerged as a pivotal tool for capturing the intricate fluctuations in stock prices. One of the advantages of this method over linear regression is its capability to incorporate more intricate, curved relationships that closely resemble real-world phenomena. Durairaj and BH have employed polynomial regression models for the purpose of examining and predicting various financial data, including stock prices [6]. These models have demonstrated significant predictive abilities, surpassing linear models in terms of accuracy. Lahmiri conducted a study on the utilization of polynomial regression in the domain of finance [7]. The researcher employed quadratic regressions to analyze projected stock returns in a cross-sectional study. The application of non-linear regression enabled the researchers to effectively capture complex associations between the predictors and stock returns that might have been disregarded by a linear model.

The versatility of polynomial regression is evident in its utilization for the detection of trends and patterns in stock prices. Misra et al. conducted a notable study wherein they utilized polynomial models integrated with machine learning techniques to examine the volatility of the stock market [8]. The incorporation of polynomial terms in the model improved its capacity to capture the phenomenon of volatility clustering that is frequently observed in stock markets. This phenomenon

is characterized by the propensity for periods of elevated volatility to be followed by subsequent periods of elevated volatility, and conversely. However, it is important to acknowledge that polynomial regression does possess certain limitations. One of the primary factors to be taken into account when employing polynomial regression is the inherent possibility of overfitting. Koehrsen argues that as a model becomes overly complex, it tends to overfit the noise inherent in the data, thereby failing to accurately capture the underlying trend. This phenomenon frequently produces noteworthy results when implemented on training data, yet it tends to demonstrate restricted efficacy when applied to unfamiliar, unobserved data [9]. The subject matter discussed holds considerable significance within the field of stock price prediction, where the primary aim often involves the forecasting of future prices through the analysis of past data [10].

Furthermore, the interpretation of the relationship between predictors and the outcome in polynomial regression can become increasingly difficult, especially as the degree of the polynomial increases. According to Sakhare and Imambi, the coefficients in linear regression can be interpreted in a direct manner [11]. Specifically, a unit change in the predictor variable leads to a corresponding change in the outcome variable equal to the value of the coefficient. However, the interpretation in polynomial regression becomes more complex and less straightforward. In spite of the aforementioned challenges, polynomial regression has exhibited its effectiveness in capturing complex relationships within financial data. The efficacy of polynomial regression in capturing the non-linear relationship between stock prices and factors like implied volatility has been demonstrated in prior research conducted by Shakhla et al. [12]. This modeling methodology provides significant insights that cannot be obtained through linear models.

In summary, polynomial regression has made noteworthy contributions to the field of financial analysis and the prediction of stock prices. The capacity to model intricate relationships provides potential benefits compared to linear regression, especially in volatile markets characterized by non-linear dynamics frequently observed in price movements. Polynomial regression, when applied with caution and careful consideration of its limitations, presents promising opportunities for future research in the analysis of financial markets.

2.3. Application of AutoRegressive Integrated Moving Average (ARIMA)

The AutoRegressive Integrated Moving Average (ARIMA) model has gained significant popularity as a statistical technique for forecasting time-series data, especially in the financial industry. The purpose of this section is to present a comprehensive literature review on the utilization of ARIMA models in the prediction of stock prices. It will explore the application, benefits, and constraints associated with this approach. ARIMA, short for Autoregressive Integrated Moving Average, is a time series model that integrates the principles of autoregression, differencing, and moving average to make predictions about future data points within the series. One notable benefit of this tool lies in its ability to proficiently depict and analyze a wide range of time series structures [13]. The characteristics enhance the tool's appeal in the realm of stock price prediction, as they frequently exhibit temporal patterns that can be accurately captured by a properly calibrated ARIMA model.

Numerous studies have employed ARIMA models in the context of stock price forecasting, resulting in promising results. Wang and Guo conducted a study on the application of ARIMA in the finance domain. They advocated for the adoption of ARIMA models to forecast economic and financial time series [14]. The research conducted by the authors established the fundamental basis for the extensive utilization of ARIMA in financial forecasting and yielded a multitude of applications within the domain of stock price prediction. In their study, Meher et al. employed ARIMA modeling techniques to make predictions regarding the closing stock prices of the Nigerian stock market [15]. The study's results indicated that the ARIMA model effectively captured the temporal patterns observed in the stock prices and produced reliable short-term forecasts.

According to Pillay, ARIMA models have demonstrated significant efficacy in capturing diverse time series phenomena, including trends and seasonal patterns, commonly observed in stock price data [16]. Furthermore, the ARIMA models possess the ability to incorporate historical values and errors, thereby effectively capturing the temporal dependencies frequently observed in financial time series data [17]. Due to its adaptability and robustness, the ARIMA model remains extensively employed for time series forecasting. The present body of research has been focused on improving the ARIMA modeling methodology by incorporating additional techniques that complement its capabilities. Pandey and Bajpai proposed a hybrid model that integrates the ARIMA method with artificial neural networks in order to enhance the accuracy of stock price prediction [18]. The study exhibited improved performance when compared to the sole utilization of the ARIMA model.

Nevertheless, it is important to acknowledge that ARIMA models do possess certain limitations. According to Mashadihasanli, one of the main limitations pertains to the assumption of linearity, which is akin to that of linear regression models [19]. The validity of the assumption may not always be upheld in financial markets, as they are frequently impacted by a multitude of nonlinear factors. These factors can include abrupt market fluctuations resulting from geopolitical events or financial crises. In addition, it should be noted that ARIMA models rely on the assumption of stationarity in the underlying time series. However, this assumption may not hold true for stock prices, as they frequently demonstrate trends and fluctuating volatilities throughout time [20]. Another challenge that arises is the selection of optimal parameters for the ARIMA model, a task that demands expert knowledge and can be time-consuming. Incorrectly specifying the model parameters can result in suboptimal forecast performance.

In summary, ARIMA models have made substantial contributions to the field of stock price forecasting. The ability to capture temporal dependencies in time series data renders them a highly effective tool for predicting stock prices. Although they have certain limitations, when the appropriate model selection and parameter tuning are applied, they can offer valuable insights for making informed investment decisions.

3. Experiment and Results Discussion

This research first pulls out the data of stock prices from 2021-07-01 to 2022-07-01 (Full-Year), and from 2021-07-01 to 2022-01-01 (Half-Year). Then, using linear regression and polynomial regression to analyze the stock price trend and figure out the best degree for polynomial regression to suit this size of data (see Figure 1). After that, it makes prediction on the average stock price of the next 14 days (which is 2022-01-01 to 2022-01-15, and 2022-07-01 to 2022-07-15) using linear, polynomial regression and the ARIMA model.

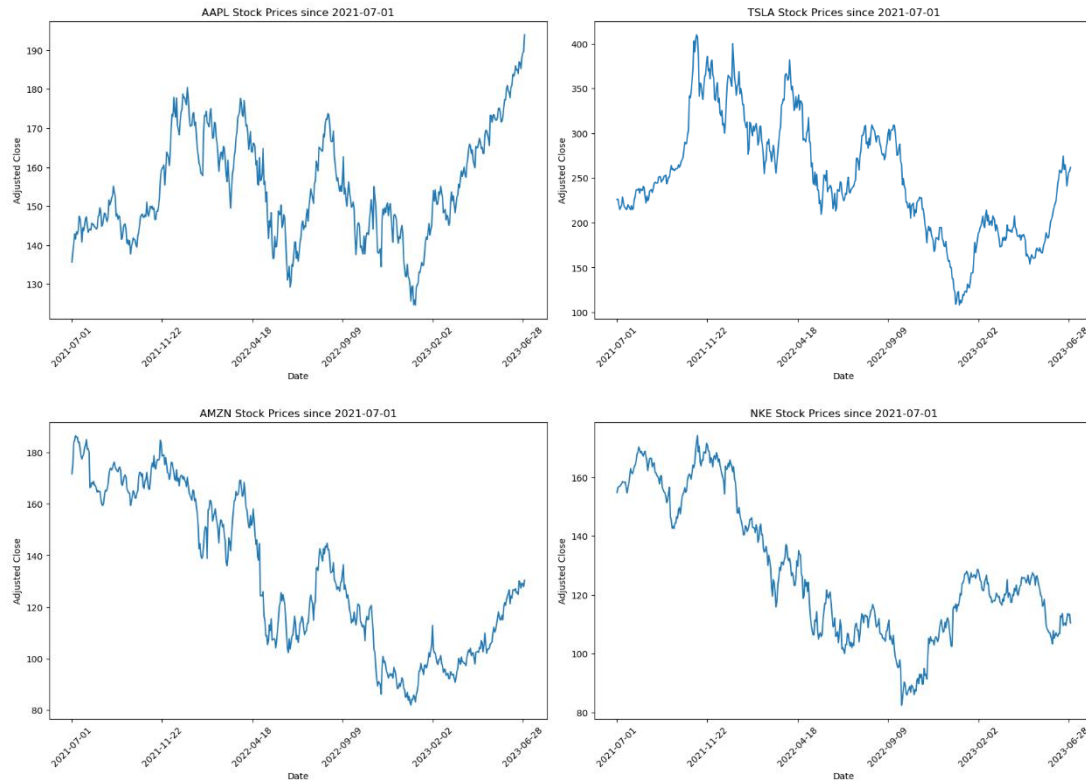


Figure 1: AAPL, TSLA, AMZN, and NKE stock prices since 2021-07-01.

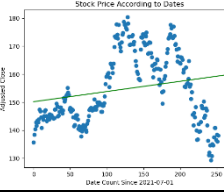
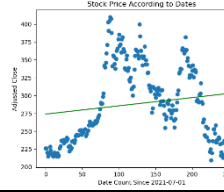
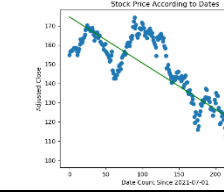
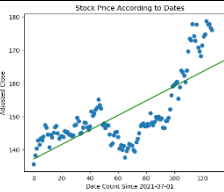
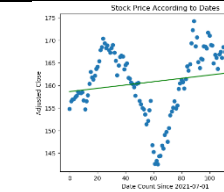
Firstly, the historical data of the next 14 days is pulled out. By computing the average of the 14 days and comparing with the starting date, I define that the average price higher than the starting price indicates “UP” and the average price lower than the starting price indicates “DOWN” (see Table 1).

Table 1: Actual stock prices of 14 days after.

	AAPL	TSLA	AMZN	NKE
2022-07-01				
Start	138.10	227.26	109.56	99.99
Mean	144.30	237.02	112.48	103.53
Trend	UP	UP	UP	UP
2022-01-01				
Start	180.43	399.93	170.40	161.94
Mean	173.37	361.302	164.35	153.62
Trend	DOWN	DOWN	DOWN	DOWN

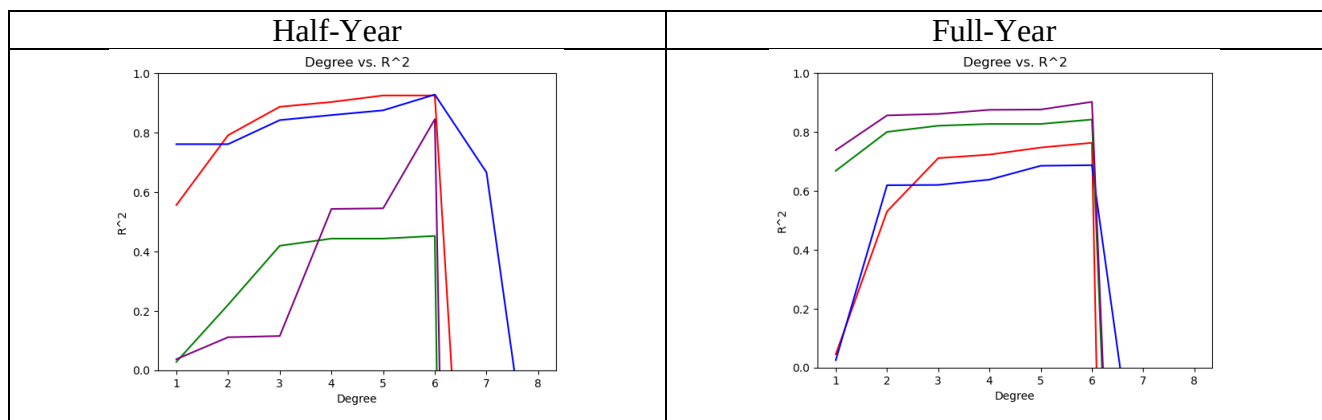
At the next step, using OLS method, the linear regression coefficients, and R^2 are being computed. Then, I graphed the function according to 8 stock price data. Extending the line by 14 days on x axis, I make the prediction and compare it with the actual data. It's obvious that the prediction is not quite accurate with 3 correct predictions out of 8. However, it's easy to see the general overall trend of the stock is going up or down by just looking at the slope of the linear regression line. A positive slope indicates that the stock price is generally increasing during this period selected (see Table 2).

Table 2: Linear regression in stock prices.

LIN	AAPL	TSLA	AMZN	NKE
Full-Year				
R^2	0.045	0.025	0.668	0.738
Prediction	UP	UP	DOWN	DOWN
Result	TRUE	TRUE	FALSE	FALSE
Half-Year				
R^2	0.556	0.761	0.028	0.037
Prediction	UP	UP	DOWN	UP
Result	FALSE	FALSE	TRUE	FALSE

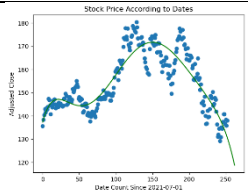
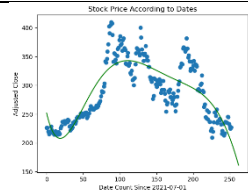
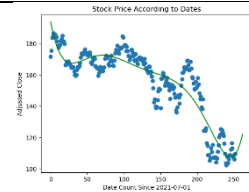
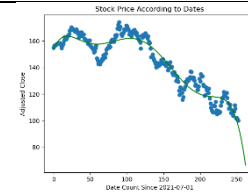
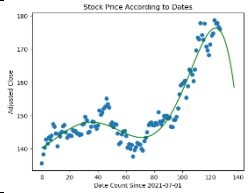
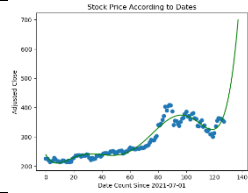
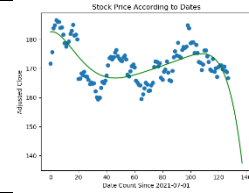
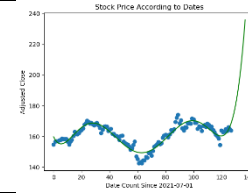
In the following, the research first did experiments on 1 to 8 degrees on polynomial regression according to the Stock Prices data. And compared the R-squared (An indicator that tells us how fit a function is according to the data provided). In the table below, it shows that all 4 stocks we initially selected (AAPL, TSLA, AMZN, NKE) reach the highest and most reasonable R^2 at degree 6. For this reason, the research continues to use degree 6 as the primary regression function for polynomial regression (see Table 3).

Table 3: Performance of polynomial degrees.



In the below polynomial regression experiments data table 4, it's observable that polynomial regression does a better job at describing the highs and lows of the stock price data than linear regression. For example, a vertex of the parabola is shown on the period around 150 days on the graph of AAPL's Full Year, it indicates the stock was experiencing a high tide and it's dropping down after. However, the linear regression data of AAPL's Full Year only shows the graph is generally raising during the period of a full year. In comparison, the linear regression graph neglects the highs and low tides but directly reflects the stock pricing is rising or decreasing for a year. Speaking of prediction results, polynomial regression is also not a stable and convincing method for stock price prediction, while 4 of the predictions are correct out of 8.

Table 4: Polynomial regression in stock prices.

	AAPL	TSLA	AMZN	NKE
Full-Year				
R ²	0.763	0.687	0.842	0.902
Degree	6	6	6	6
Prediction	DOWN	DOWN	UP	DOWN
Result	FALSE	FALSE	TRUE	FALSE
Half-Year				
R ²	0.925	0.928	0.452	0.845
Degree	6	6	6	6
Prediction	DOWN	DOWN	DOWN	UP
Result	TRUE	TRUE	TRUE	FALSE

Finally, the research selects one of the best time series regression methods currently – ARIMA and tests its performance on the data. In this test, the research is using (1,0,1) as the p, d, and q for the ARIMA order. According to the Prediction Result, ARIMA scored 7 out of 8 tests. This indicates that ARIMA is quite accurate on just predicting “UP” or “DOWN” for 14 days after the periods selected. The ARIMA model is obviously a better model on future price prediction compared to linear and polynomial regression models (see Table 5).

Table 5: ARIMA prediction on stock prices.

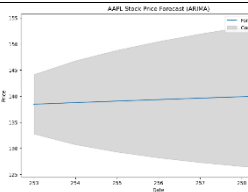
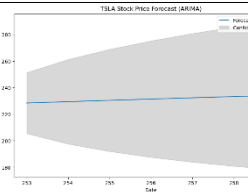
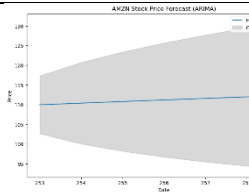
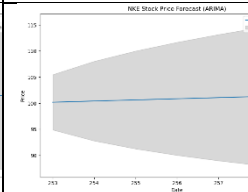
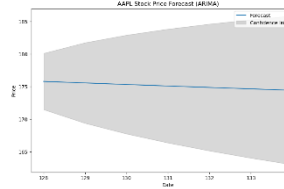
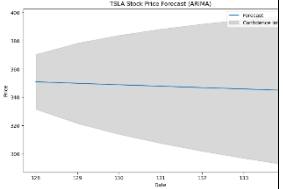
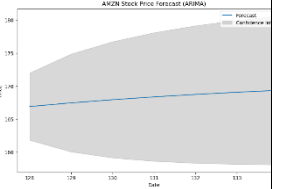
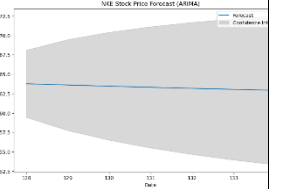
	AAPL	TSLA	AMZN	NKE
P-Graph (Year)				
Prediction	UP	UP	UP	UP

Table 5: (continued).

Result	True	True	True	True
P-Graph (Half)				
Predicti on	DOWN	DOWN	UP	DOWN
Result	TRUE	TRUE	TRUE	TRUE

4. Conclusion

In conclusion, Linear regression is underfitted, Polynomial regression gives a good/robust fit. Linear regression provides a direct “UP” or “Down” for the stock price in general, Polynomial aids to observe highs and low tides on the market, and ARIMA is the best at making predictions on future stock prices.

The future goal of this research is look more into the ARIMA ordering models and try using different types of regressions and data (Volume, etc.) to form a more formulated system of functions that aims to both having high accuracy on future stock price prediction and appropriate fit on historic data of different kinds of periods.

References

- [1] Riyadh, D., Kaushal, G., Sharma, A., & Singh, U. (2017) Stock market predication using a linear regression. In 2017 international conference of electronics, communication and aerospace technology (ICECA), 2, 510-513. IEEE.
- [2] Seethalakshmi, R. (2018) Analysis of stock market predictor variables using Linear Regression. International Journal of Pure and Applied Mathematics, 119(15), 369-378.
- [3] Gururaj, V., Shriya, V. R., and Ashwini, K. (2019) Stock market prediction using linear regression and support vector machines. Int J Appl Eng Res, 14(8), 1931-1934. http://ripublication.com/ijaerv14n8_24.pdf
- [4] Gupta, B. K., Mallick, M. K., and Hota, S. (2021) Survey on stock price forecasting using regression analysis. In Intelligent and Cloud Computing: Proceedings of ICICC 2019, 2, 147-156.
- [5] Cattaneo, M. D., Jansson, M., and Newey, W. K. (2018) Inference in linear regression models with many covariates and heteroscedasticity. Journal of the American Statistical Association, 113(523), 1350-136.
- [6] Durairaj, M., and BH, K. M. (2021) Statistical evaluation and prediction of financial time series using hybrid regression prediction models. International Journal of Intelligent Systems and Applications in Engineering, 9(4), 245-255.
- [7] Lahmiri, S. (2018) Minute-ahead stock price forecasting based on singular spectrum analysis and support vector regression. Applied Mathematics and Computation, 320, 444-451.
- [8] Misra, M., Yadav, A. P., and Kaur, H. (2018) Stock market prediction using machine learning algorithms: a classification study. In 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE), 2475-2478). IEEE.
- [9] Koehrsen, W. (2018) Overfitting vs. underfitting: A complete example. Towards Data Science, 405.
- [10] Amrutphale, J., Rathore, P., and Malviya, V. (2020) A novel approach for stock market price prediction based on polynomial linear regression. In Social Networking and Computational Intelligence: Proceedings of SCI-2018, 161-171.
- [11] Sakhare, N. N., and Imambi, S. S. (2019) Performance analysis of regression based machine learning techniques for prediction of stock market movement. International Journal of Recent Technology and Engineering, 7(6), 655-662.
- [12] Shakhla, S., Shah, B., Shah, N., Unadkat, V., and Kanani, P. (2018) Stock price trend prediction using multiple linear regression. International Journal of Engineering Science Invention (IJESI), 7(10), 29-33.

- [13] Almasarweh, M., and Alwadi, S. (2018) *ARIMA model in predicting banking stock market data*. *Modern Applied Science*, 12(11), 309.
- [14] Wang, Y., and Guo, Y. (2020) *Forecasting method of stock market volatility in time series data based on mixed model of ARIMA and XGBoost*. *China Communications*, 17(3), 205-221.
- [15] Meher, B. K., Hawaldar, I. T., Spulbar, C. M., and Birau, F. R. (2021) *Forecasting stock market prices using mixed ARIMA model: A case study of Indian pharmaceutical companies*. *Investment Management and Financial Innovations*, 18(1), 42-54.
- [16] Pillay, S. (2020) *Determining the optimal arima model for forecasting the share price index of the johannesburg stock exchange*. *Journal of Management Information and Decision Sciences*, 23(5), 527-538.
- [17] Dhyani, B., Kumar, M., Verma, P., and Jain, A. (2020) *Stock market forecasting technique using arima model*. *International Journal of Recent Technology and Engineering*, 8(6), 2694-2697.
- [18] Pandey, V. S., and Bajpai, A. (2019) *Predictive efficiency of ARIMA and ANN models: a case analysis of nifty fifty in Indian stock market*. *International Journal of Applied Engineering Research*, 14(2), 232-244.
- [19] Mashadihasanli, T. (2022) *Stock Market Price Forecasting Using the Arima Model: an Application to Istanbul, Turkiye*. *Journal of Economic Policy Researches*, 9(2), 439-454.
- [20] Uma Maheswari, B., Sujatha, R., Fantina, S., and Mansurali, A. (2021) *ARIMA versus ANN—A comparative study of predictive modelling techniques to determine stock price*. In *Proceedings of the Second International Conference on Information Management and Machine Intelligence: ICIMMI 2020*, 315-323.