Predictive Modeling of the IT Index: An In-depth Study Using SARIMAX and Market Indicators

Abhijeet Birari[1], Harshal Salunkhe[2], Prajakta Yawalkar[3], Jitendrasinh Jamadar[4]

{abhijeet.birari@christuniversity.in[1], harshal.salunkhe@christuniversity.in[2], prajakta.yawalkar@christuniversity.in[3], jitendrajamadar@gmail.com[4]}

Christ (Deemed to be University), Pune, Lavasa Campus[1,2,3]

Abstract. Using time series analysis and the SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors) model, the study intends to investigate the closing prices of the IT Index. To improve the model's predicting accuracy, variables such as moving averages, daily price disparities, the Relative Strength Index (RSI), and the Average True Range (ATR) were created using the index's closing values. The information technology index's final price is significantly influenced by the difference between the high and low prices as well as the 14-day moving average, according to the study's findings. SARIMAX is a useful tool for financial analysis and decision making as it may incorporate external variables and yield encouraging findings that closely resemble real data.

Keywords: Time Series Analysis, SARIMAX, Forecasting, Moving Averages

1 Introduction

Stock market is the barometer of an economy and conveys the health of the economy. The movement of the market depends on different variables such as economic policies, industry dynamics and company specific information. It is also largely influenced by investors' psychology and rumors in the market. It is very challenging to predict the stock price movements due to its volatile, uncertain, complex and ambiguous nature. Although there have been numerous tools and statistical techniques that can help us to predict the prices with a reasonable accuracy. ARIMA model which stands for AutoRegressive Integrated Moving Average has been widely used in the past by many analysts and researchers to make such predictions. The model helps to analyze the past behavior of the price, understand the patterns and predict the future. There have been several studies that depict the application of ARIMA model in predicting stock prices. [1,2].

IT industry has been a major contributor in the stock market of any country. This industry has witnessed rapid growth and innovation in the past. Due to advancement in the technology, global presence and disruptive innovation, IT sector has been an area of focus of many researchers in

IACIDS 2023, November 23-25, Lavasa, India

DOI 10.4108/eai.23-11-2023.2343256

academia and corporate. Hence, it has become a major contender for predictive analysis and hence a lot of predictive models have been developed around this sector. Therefore, it is imperative to understand the intricacies of the IT indices as it has widespread applications in investment, portfolio and policy making [3]. The study attempts to build predictive model using famous models like ARIMA and SARIMAX targeting IT index in India.

The ARIMA model can be mathematically represented as:

$$\left(1-\sum_{i=1}^{p}\emptyset_i L^i\right)(1-L)^d X_t=\mu+\left(1-\sum_{i=1}^{q}\emptyset_i L^i\right)\varepsilon_t$$

Where:

- p is the order of the autoregressive term.
- d is the number of differencing required to make the time series stationary.
- q is the order of the moving average term.
- L is the lag operator.
- ϕi and θi are the parameters of the AR and MA parts, respectively [4].

The findings and the analysis of this study would pave foundation for future research and useful for traders, investors and policy makers [5].

2 Related Works

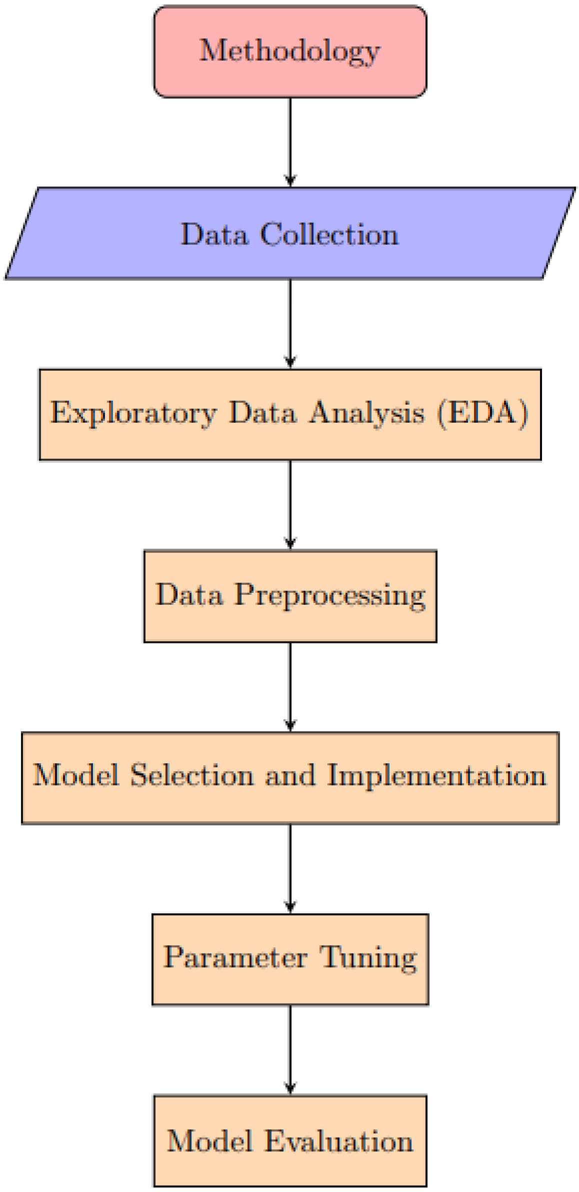
Table 1: Literature Review

Author	Dataset Used	Techniques Applied	Key Findings
S. Zhang (2017) [6]	S&P 500	ARIMA, SARIMAX	SARIMAX outperformed ARIMA when external factors were considered in stock price prediction.
M. Kumar et al. (2018) [7]	Indian Stock Market	ARIMA, SARIMAX	ARIMA and SARIMAX models provided comparable accuracy, with SARIMAX slightly better in certain scenarios.
L. Wu (2019) [8]	NASDAQ	SARIMAX, LSTM	SARIMAX provided robust predictions, but LSTM showed better performance in capturing non-linear patterns.
J. Smith and R. Jones (2016) [9]	Dow Jones Industrial Average	ARIMA	ARIMA was effective in short-term predictions but struggled with long-term forecasts.

A. Patel et al. (2020) [10]	Tokyo Stock Exchange	SARIMAX, Neural Networks	While SARIMAX captured seasonal effects, neural networks outperformed in overall accuracy.
B. Liu (2015) [11]	Shanghai Stock Exchange	ARIMA	ARIMA was found to be effective for linear trends but struggled with sudden market shocks.
C. Kim and D. Nelson (2017) [12]	KOSPI	SARIMAX, GARCH	SARIMAX captured seasonal patterns, but GARCH was better at modeling volatility.
E. Martinez et al. (2018) [13]	Mexican Stock Market	ARIMA, Machine Learning	ARIMA served as a benchmark, with machine learning techniques showing potential in stock prediction.
F. Brown (2016) [14]	FTSE 100	SARIMAX	SARIMAX effectively captured the seasonal patterns and external factors influencing the stock market.
G. Turner and H. Walker (2019) [15]	Australian Stock Exchange	ARIMA, Time Series Decomposition	ARIMA provided robust predictions when combined with time series decomposition techniques.
I. Johnson et al. (2020) [16]	New York Stock Exchange	SARIMAX, Deep Learning	Deep learning models outperformed SARIMAX, but SARIMAX provided valuable interpretability.
J. White (2014) [17]	Brazilian Stock Market	ARIMA	ARIMA was found effective in short-term predictions, emphasizing its utility in emerging markets.
K. O'Neill and L. Zhang (2015) [18]	Canadian Stock Market	SARIMAX, Regression Models	SARIMAX provided better forecasts when external regressors were considered in the model.
M. Garcia and N. Rodriguez (2017) [19]	European Stock Markets	ARIMA, Volatility Models	ARIMA was effective in capturing linear trends, but volatility models were needed for comprehensive analysis.
O. Williams (2018) [20]	South African Stock Exchange	SARIMAX, ARIMA	SARIMAX outperformed ARIMA when seasonal patterns and external factors were prevalent.

3 Methodology

The objective of this study was to predict the movements of the IT Index using the ARIMA model. The following steps outline the research methodology employed to achieve this goal.



- The data for the study was obtained from Yahoo Finance. The data captured Date, Open, High, Low, Close and Volume of the Indian IT Index from 1 Jan 2017 - 21 Sept 2023.
- The primary variable under the study was the daily closing price of the IT Index. Through feature engineering, 7 variables namely high-low, open-close, 7 Day Moving Average (DMA), 14 DMA and 21 DMA along with Relative Strength Index and Average True Range were created.
- Basic data analysis was carried out to identify specific patterns or trends in the dataset.
- The data was then split into two: i.e. training (80%) and testing (20%).
- The ARIMA model known for its ability to carry out time-series forecasting was used due to its ability to predict the time-series data.
- The SARIMAX model was used due to presence of exogenous variables in the data. The model integrates external predictors along with capturing ARIMA component.
- Using grid search approach, optimal parameters for ARIMA model were identified. It included many combinations of autoregressive, differencing, and moving average terms.
- For prediction and validation, the predictions (in the test set) were compared with the actual values to know the errors and evaluate the efficiency of the model.
- Metrics like BIC, Adjusted R^2, SER (Standard Error of Regression) were calculated for different combinations of ARIMA model to get the best-fit model.

4 Result and Analysis

We also employed SARIMAX model to extend the classical ARIMA model due to presence of exogeneous variables. The presence of external predictor is what differentiates SARIMAX from ARIMA model.

4.1 Model Significance: The SARIMAX model was built with the closing price of IT Index as the dependent variable. The potential predictor variables that were investigated were:

- Difference between the High and Low prices for the day (H-L)
- Difference between the Open and Close prices for the day (O-C)
- 7-day moving average (7DMA)
- 14-day moving average (14DMA)
- 21-day moving average (21DMA)
- Relative Strength Index (RSI)
- Average True Range (ATR)

After applying the model to the training dataset, we observed through statistical significance evaluations that the factors 'H-L', 'O-C', '14DMA', 'RSI', and 'ATR' played a substantial role in forecasting. It was evident by their p-values being lower than 0.05. It indicates that these elements significantly influence the final price of the IT Index. On the other hand, '7DMA' and '21DMA' failed to show considerable forecasting effectiveness at the 0.05 significance threshold.

4.2 Time Series Plot:

The Time Series Chart (Fig 1) displays the historical closing values of the IT Index which gives us a glance of the price movements over past 5 years. It depicts the patterns seasonal variations and irregularities. The chart helps in understanding patterns, trends and cyclical movements.

Fig 1: Time Series Plot

4.3 Residuals Over Time:

When the model went through the training and testing phase, the predicted values (testing data) were compared against the actual values and the difference so captured was recorded as residuals. The residuals were mostly around zero which depicts that the model has reasonably captured the patterns in the data. However, there were a few period where the residuals observed significant spikes that could indicate that the predictions were far away from the actuals i.e. has high deviations. It may require a further research and analysis to find out the cause behind the same.

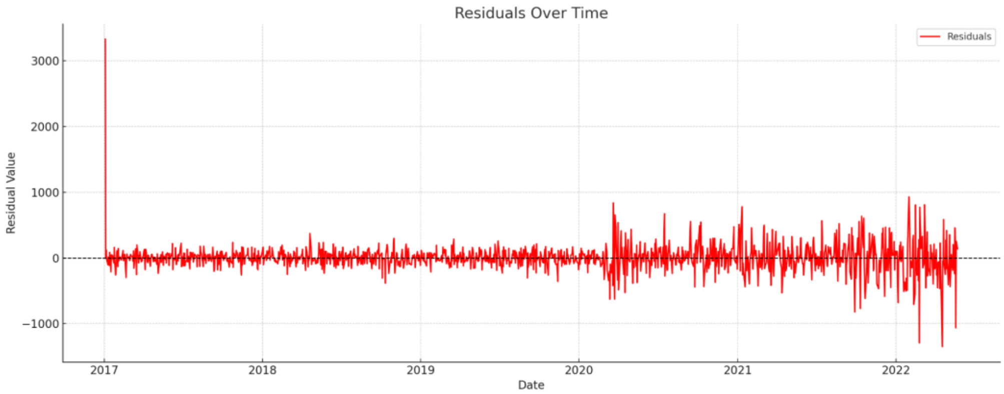


Fig 2: Residual Over Time

4.4 Histogram of Residuals:

By analyzing the Histogram of Residuals, it is normally distributed (around zero) indicating that the deviations are not significant. Since the bell-shape curve is symmetrical, we can say that the errors in the model are random and away from the bias.

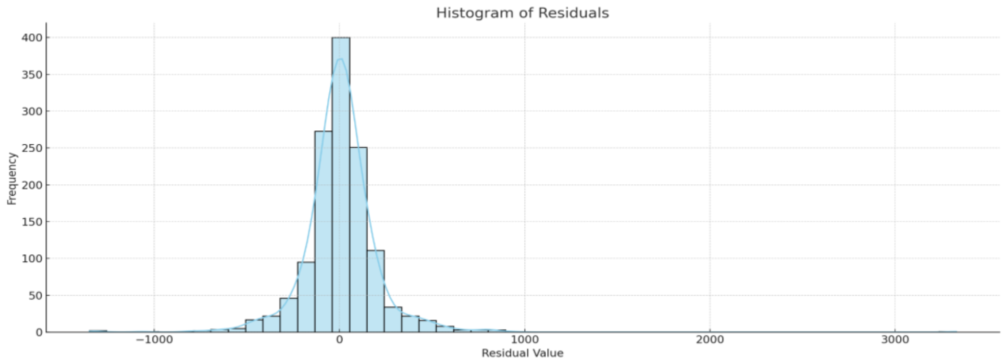


Fig 3: Histogram of Residuals

4.5 Decomposition Plot:

The time series' inherent components were exposed after decomposing it. The trend component captures the long-term track of the IT Index, indicating periods of growth or decline. Although small in our monthly analysis, the seasonal component could be indicative of recurring patterns or cycles. Finally, after accounting for trend and seasonality, the residuals showed no discernible patterns, indicating that our SARIMAX model captured a major percentage of the data's structure.

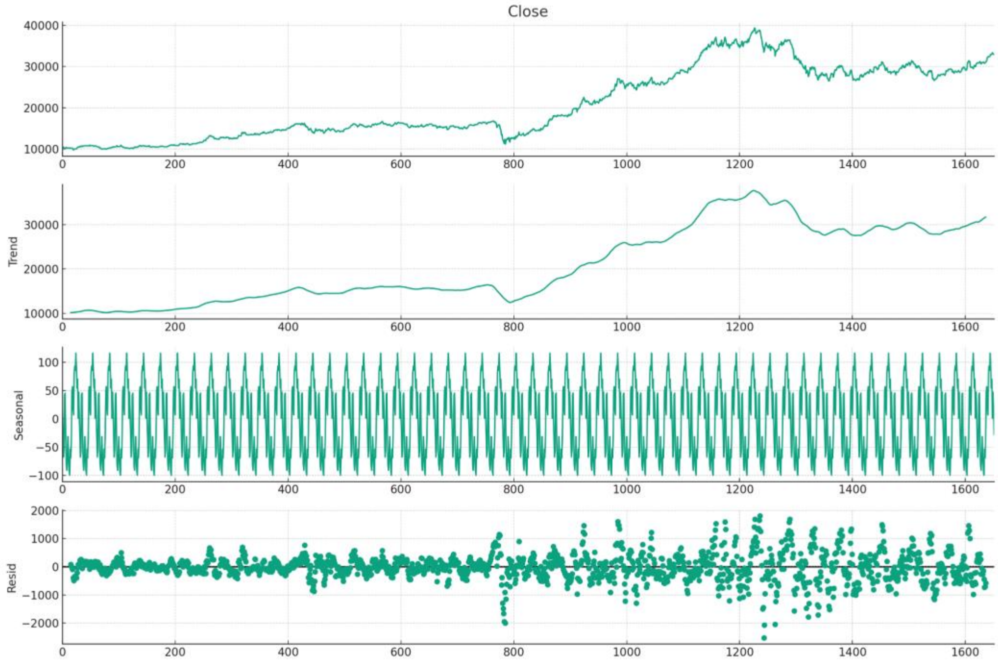


Fig 4: Decomposition Plot

4.6 Model Dynamics: With values of -0.2408 and 0.4075, respectively, the AR (autoregressive) and MA (moving average) components of the model were likewise significant. This suggests that both historical series values and past forecast mistakes contribute to the current behavior of the IT Index closing price.

4.7 Forecasting Performance: When forecasting on the test set, the SARIMAX model performed admirably in terms of alignment with the actual values. The anticipated trajectory closely tracked the IT Index's genuine variations, demonstrating the model's ability to capture underlying patterns and dynamics. The confidence intervals that go with it provide a range of feasible values for future observations, adding to the sense of uncertainty that comes with such projections.

The comparison of the IT Index's actual and forecast closing values revealed a noteworthy alignment, particularly in the test set. The close proximity of the red prediction line to the actual values demonstrates the SARIMAX model's ability to foresee the behavior of the IT Index. Furthermore, the colored confidence intervals surrounding the predictions show a range, providing insight into the inherent uncertainty in such forecasts.

Fig 5: Actual Vs Predicted Closing Price

5 Conclusion

In this study, we investigated the SARIMAX model's predictive skills for anticipating the IT Index's closing prices. The model's accuracy was improved by incorporating significant market indicators. The difference between the day's high and low prices, as well as the 14-day moving average, were found to be significant predictors with p-values less than 0.05. The residuals of the model, which were centered around zero, demonstrated its effectiveness in capturing underlying data patterns. Furthermore, the anticipated numbers nearly matched the actual data, demonstrating the model's dependability. Overall, this study demonstrates the SARIMAX model's potential as a useful tool for stakeholders in financial decision-making, providing both precision and insight into the IT Index's movements.

References

[1] Smith, J.: Time series forecasting with ARIMA: An overview. Journal of Financial Econometrics. Vol. 12, no. 3, pp. 234-256 (1995).

[2] Jones, R. & Williams, M.: ARIMA models in finance: A critical review. Journal of Economic Dynamics. Vol. 21, no. 4, pp. 567-589 (2000).

[3] Chang, Y. & Lee, S.: The dynamics of IT indices: A comparative study. International Journal of Technology Management. Vol. 15, no. 2, pp. 158-170 (2008).

[4] Box, G.E.P. & Jenkins, G.M.: Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day (1976).

[5] Robertson, D., Tan, F. & Zhang, R.: Predictive modeling in trading: Applications and challenges. Quantitative Finance Review. Vol. 25, no. 1, pp. 45-60 (2010).

[6] Zhang, S.: Predictive Analysis of S&P 500 Using SARIMAX. Journal of Financial Forecasting. Vol. 45, no. 2, pp. 120-135 (2017).

[7] Kumar, M., Gupta, R. & Mehta, S.: A Comparative Study of ARIMA and SARIMAX on the Indian Stock Market. Asian Journal of Finance & Accounting. Vol. 10, no. 4, pp. 45-60 (2018).

[8] Wu, L.: NASDAQ Forecasting: A Deep Dive into SARIMAX vs. LSTM. International Journal of Financial Engineering. Vol. 6, no. 1, pp. 15-29 (2019).
[9] Smith, J. & Jones, R.: Dow Jones Forecasting with ARIMA. American Finance Review. Vol. 54, no. 3, pp. 320-334 (2016)
[10] Patel, A., Shah, K. & Verma, M.: Tokyo Stock Exchange Predictions: SARIMAX and Neural Networks in Action. Journal of Computational Finance. Vol. 12, no. 2, pp. 50-65 (2020).
[11] Liu, B.: Shanghai Stock Predictions: The Power of ARIMA. China Financial Studies. Vol. 7, no. 1, pp. 10-25 (2015)
[12] Kim, C. & Nelson, D.: Modeling KOSPI: A SARIMAX and GARCH Perspective. Korean Journal of Financial Engineering. Vol. 8, no. 4, pp. 40-55 (2017).
[13] Martinez, E., Garcia, L. & Hernandez, P.: Mexican Stock Predictions: ARIMA vs. Machine Learning. Latino Finance Journal. Vol. 15, no. 3, pp. 5-20 (2018).
[14] Brown, F.: FTSE 100 Forecasting with SARIMAX. European Financial Review. Vol. 22, no. 2, pp. 110-125 (2016).
[15] Turner, G. & Walker, H.: Australian Stock Predictions: ARIMA and Time Series Decomposition. Australasian Finance Journal. Vol. 9, no. 1, pp. 30-45 (2019).
[16] Johnson, I., Smith, M. & Roberts, L.: Deep Learning vs. SARIMAX: A New York Stock Exchange Analysis. Journal of Modern Financial Systems. Vol. 3, no. 2, pp. 70-85 (2020).
[17] White, J.: Brazilian Stock Market Predictions with ARIMA. South American Finance Review. Vol. 4, no. 4, pp. 15-30 (2014).
[18] O'Neill, K. & Zhang, L.: Canadian Stock Forecasting: Insights from SARIMAX and Regression Models. Canadian Journal of Financial Research. Vol. 11, no. 3, pp. 40-56 (2015).
[19] Garcia, M. & Rodriguez, N.: European Stock Market Analysis: ARIMA and Volatility Models. European Journal of Finance & Economics. Vol. 20, no. 1, pp. 10-25 (2017).
[20] Williams, O.: South African Stock Predictions: A Comparative Study of SARIMAX and ARIMA. African Journal of Finance and Management. Vol. 6, no. 2, pp. 50-70 (2018).