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

Abhijeet Birari¹, Harshal Salunkhe², Prajakta Yawalkar³, Jitendrasinh Jamadar⁴

{abhijeet.birari@christuniversity.in¹, harshal.salunkhe@christuniversity.in²,
prajakta.yawalkar@christuniversity.in³, jitendrajamadar@gmail.com⁴}

Christ (Deemed to be University), Pune, Lavasa Campus¹,²,³

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
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:

\[
\left(1 - \sum_{i=1}^{p} \phi_i L^i\right) (1 - L)^d X_t = \mu + \left(1 - \sum_{i=1}^{q} \theta_i L^i\right) \epsilon_t
\]

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.
- \(\phi_i\) and \(\theta_i\) 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>
<thead>
<tr>
<th>Author</th>
<th>Dataset Used</th>
<th>Techniques Applied</th>
<th>Key Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>M. Kumar et al. (2018) [7]</td>
<td>Indian Stock Market</td>
<td>ARIMA, SARIMAX</td>
<td>ARIMA and SARIMAX models provided comparable accuracy, with SARIMAX slightly better in certain scenarios.</td>
</tr>
<tr>
<td>L. Wu (2019) [8]</td>
<td>NASDAQ</td>
<td>SARIMAX, LSTM</td>
<td>SARIMAX provided robust predictions, but LSTM showed better performance in capturing non-linear patterns.</td>
</tr>
<tr>
<td>J. Smith and R. Jones (2016) [9]</td>
<td>Dow Jones Industrial Average</td>
<td>ARIMA</td>
<td>ARIMA was effective in short-term predictions but struggled with long-term forecasts.</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Stock Market</td>
<td>Model(s)</td>
<td>Notes</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------------------</td>
<td>------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Networks</td>
<td>ARIMA was found to be effective for linear trends but struggled with sudden market shocks.</td>
</tr>
<tr>
<td>B. Liu (2015)</td>
<td>Shanghai Stock Exchange</td>
<td>SARIMAX,</td>
<td>SARIMAX captured seasonal patterns, but GARCH was better at modeling volatility.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GARCH</td>
<td>ARIMA served as a benchmark, with machine learning techniques showing potential in stock prediction.</td>
</tr>
<tr>
<td>C. Kim and D. Nelson</td>
<td>KOSPI</td>
<td>SARIMAX,</td>
<td>SARIMAX effectively captured the seasonal patterns and external factors influencing the stock market.</td>
</tr>
<tr>
<td>(2017)</td>
<td></td>
<td>GARCH</td>
<td>ARIMA provided robust predictions when combined with time series decomposition techniques.</td>
</tr>
<tr>
<td>E. Martinez et al.</td>
<td>Mexican Stock Market</td>
<td>ARIMA, Machine</td>
<td>Deep learning models outperformed SARIMAX, but SARIMAX provided valuable interpretability.</td>
</tr>
<tr>
<td>(2018)</td>
<td></td>
<td>Learning</td>
<td>ARIMA was found effective in short-term predictions, emphasizing its utility in emerging markets.</td>
</tr>
<tr>
<td>F. Brown (2016)</td>
<td>FTSE 100</td>
<td>SARIMAX</td>
<td>SARIMAX provided better forecasts when external regressors were considered in the model.</td>
</tr>
<tr>
<td>G. Turner and H.</td>
<td>Australian Stock</td>
<td>ARIMA, Time</td>
<td>ARIMA was effective in capturing linear trends, but volatility models were needed for comprehensive analysis.</td>
</tr>
<tr>
<td>Walker (2019)</td>
<td>Exchange</td>
<td>Series</td>
<td>SARIMAX captured seasonal patterns and external factors were prevalent.</td>
</tr>
<tr>
<td>I. Johnson et al.</td>
<td>New York Stock</td>
<td>SARIMAX, Deep</td>
<td>SARIMAX outperformed ARIMA when seasonal patterns and external factors were prevalent.</td>
</tr>
<tr>
<td>(2020)</td>
<td>Exchange</td>
<td>Learning</td>
<td>ARIMA provided robust predictions when combined with time series decomposition techniques.</td>
</tr>
<tr>
<td>J. White (2014)</td>
<td>Brazilian Stock</td>
<td>ARIMA</td>
<td>ARIMA was found effective in short-term predictions, emphasizing its utility in emerging markets.</td>
</tr>
<tr>
<td>K. O’Neill and L.</td>
<td>Canadian Stock</td>
<td>SARIMAX,</td>
<td>SARIMAX provided better forecasts when external regressors were considered in the model.</td>
</tr>
<tr>
<td>Zhang (2015)</td>
<td>Market</td>
<td>Regression</td>
<td>ARIMA was effective in capturing linear trends, but volatility models were needed for comprehensive analysis.</td>
</tr>
<tr>
<td>M. Garcia and N.</td>
<td>European Stock</td>
<td>SARIMAX,</td>
<td>SARIMAX outperformed ARIMA when seasonal patterns and external factors were prevalent.</td>
</tr>
<tr>
<td>Rodriguez (2017)</td>
<td>Markets</td>
<td>ARIMA, Volatility</td>
<td>ARIMA was found effective in short-term predictions, emphasizing its utility in emerging markets.</td>
</tr>
<tr>
<td>O. Williams (2018)</td>
<td>South African Stock</td>
<td>SARIMAX, ARIMA</td>
<td>SARIMAX provided better forecasts when external regressors were considered in the model.</td>
</tr>
</tbody>
</table>
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.

![Time Series Plot](image)

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.
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.

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.
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.


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