

Econometric Analysis using Time Series Modelling of Realty Stocks in India: A Data Science Perspective

Shweta P

Department of Mathematics(SAS)
Vellore Institute of Technology
Vellore, India
shweta.p2023@vitstudent.ac.in

Pradeep Kumar B

Department of Mathematics(SAS)
Vellore Institute of Technology
Vellore, India
pradeepkumar.b2023@vitstudent.ac.in

Dr. Jitendra Kumar

Department of Mathematics(SAS)
Vellore Institute of Technology
Vellore, India
jitendra.kumar@vit.ac.in

Abstract—This study aims to analyze the statistical patterns of Indian real estate stocks by employing econometric methods and time series modeling. Our objective is to reveal underlying trends and correlations within the Indian real estate stock market using historical data. Utilizing advanced statistical techniques like ARIMA modeling, our goal is to forecast stock prices accurately. The insights gained from this research will contribute to improving forecasting models and risk management strategies in the Indian financial markets. Our analysis includes data on open, high, low, and close stock prices.

Keywords—Time Series Analysis, Times Series Decomposition, ARIMA Modelling, Realty Stocks, trends, patterns, forecasting

I. INTRODUCTION

The real estate industry holds significant importance in driving the country's economic advancement and stands as a key contributor to employment, ranking second only to the agricultural sector. As per a report from the Confederation of Real Estate Developer's Association (CREDAI), the sector is anticipated to reach a market size of USD 1.3 trillion, constituting 13.8 percent of the projected GDP by FY2034[1]. Recent years have witnessed notable growth in the Indian real estate sector, driven by increased demand for both office and residential properties. Given the escalating investments in the real estate domain, it is important to analyze the sector's performance thoroughly and generate accurate forecasts to further stimulate investments.

This research analyses historical stock price data of 10 realty stocks in the Bombay Stock Exchange (BSE) and aims to generate accurate forecasts using ARIMA Modelling.

II. LITERATURE REVIEW

Using historical data from the National Stock Exchange, this research paper evaluates the accuracy of ARIMA modelling in predicting stock prices for 56 Indian equities in various industries. The methodology encompasses model fitting, selection, and forecasting, with particular emphasis on AIC (Akaike Information Criterion) for parameter estimation. The analysis using Mean Absolute Error (MAE) indicates promising results,

with the majority of industries achieving accuracy levels exceeding 85 percent [2].

This paper aims to utilize stochastic time series ARIMA modeling to predict changes in the stock market index direction and evaluate the stationarity of time series data, namely BSE close stock values and NSE close stock values. Through consideration of various evaluation metrics such as AIC, BIC (Bayesian Information Criterion), Root Mean Square Error(RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error(MAPE), Standard Error of Regression, and adjusted R2 values, the ARIMA (0,1,0) model was identified as the best fit for forecasting the values of time series[3].

This research focuses on modeling and forecasting stock returns in the Indian market using time series analysis techniques. The authors propose a hybrid model that combines autoregressive integrated moving average (ARIMA) and generalized autoregressive conditional heteroskedasticity (GARCH) models to capture the volatility and non-linearity of stock returns in India [4].

III. METHODOLOGY

A. Data Collection

The data for this study was obtained from BSE via Yahoo Finance and includes information on 10 companies within the realty sector. The dataset includes data points such as open, close, high, and low prices of realty stocks. The time frame covered in the dataset ranges 1st January 2019 to 18th March 2024.

B. Data Preprocessing

This involved identifying pertinent features, encoding categorical variables, addressing outliers, cleaning the data to remove errors and inconsistencies, scaling features for uniformity, and handling missing values.

C. Descriptive Statistics

We utilized Excel to calculate the coefficient of variation (CV) for the open, high, low, and close stock prices of all ten companies. Subsequently, we ranked them based on their

CV (referred to as consistency ranking) and computed the Spearman's correlation coefficient in order to determine the direction and strength of the correlations between the opening, closing, high, and low prices.

We also determined each of these price variables' mean and median values. These summary statistics shed light on the data distribution's central tendency.

D. Exploratory Data Analysis

We used Python to conduct various analytical tasks. We plotted the open, high, low, and close stock values against time for all companies included in the dataset. The graphs illustrated how these stock values fluctuate over time for each company individually. By visually examining these plots, we gained insights into the trends, patterns, and fluctuations in stock prices over the specified time period.

E. Time Series Analysis and ARIMA Modelling

Time Series Analysis and Autoregressive Integrated Moving Average (ARIMA) Modeling were conducted to comprehensively analyze the dataset. Time series decomposition was employed to identify trends and seasonality within the data, while also assessing the fluctuations in residuals to determine stationarity. Autocorrelation Function (ACF) plots were generated to visualize the autocorrelation structure of the data. Utilizing Python, an ARIMA (5,1,2) model was fitted to the time series data to facilitate forecasting. Additionally, Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) metrics were calculated to evaluate the accuracy of the forecasts.

The Python libraries StatsModels, scikit-learn (sklearn), NumPy, Matplotlib, and Pandas were used for analysis. These libraries provided the necessary functionality for time series analysis, ARIMA modeling, data manipulation, visualization, and evaluation of forecasting accuracy.

IV. RESULTS

(A) Descriptive Analysis: Table 1 shows the coefficient of variation of the realty stocks for opening, high, low and closing stock prices.

Table 1 - Coefficient of Variation

| COMPANY | OPEN | HIGH | LOW | CLOSE |
|---------------------|-------|-------|-------|-------|
| Phoenix Limited | 48.59 | 49.61 | 49.15 | 49.47 |
| Oberoi Realty | 38.96 | 37.5 | 39.6 | 39.38 |
| Prestige Group | 56.06 | 56.91 | 54.87 | 57.97 |
| Brigade Group | 58.07 | 57.28 | 58.49 | 58.49 |
| DLF | 50.71 | 49.77 | 52.89 | 52.62 |
| Godrej Properties | 33.66 | 33.61 | 33.93 | 33.66 |
| Sobha Limited | 45.42 | 46.88 | 43.58 | 47.95 |
| Mahindra Lifespaces | 61.51 | 60.01 | 62.15 | 60.92 |
| Sunteck Realty | 24.23 | 22.68 | 24.72 | 24.42 |
| Anant Raj Limited | 93.78 | 91.33 | 97.07 | 94.11 |

The consistency ranking (Table 2) was determined based on the coefficient of variation (CV) calculated for each company's stock prices. Companies with lower CV values were considered more consistent in their stock price movements, while those with higher CV values exhibited greater variability. This ranking approach allowed for the identification of companies with more stable stock price behaviors, providing valuable

insights into their performance within the market.

Table 2 - Consistency Ranking

| RANK | COMPANY |
|------|---------------------|
| 1 | Sunteck Realty |
| 2 | Godrej Properties |
| 3 | Oberoi Realty |
| 4 | Sobha Limited |
| 5 | Phoenix Limited |
| 6 | DLF |
| 7 | Prestige Group |
| 8 | Brigade Group |
| 9 | Mahindra Lifespaces |
| 10 | Anant Raj Limited |

(B) Inferential Analysis: The Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Squared Error (MSE) were computed for the open, high, low, and close stock prices of all companies. However, only the RMSE values are shown in Table 3, as RMSE is commonly considered a preferred metric for evaluating forecasting accuracy.

Table 3 - RMSE* Values for Open, High, Low and Close Stock prices

| COMPANY | OPEN | HIGH | LOW | CLOSE |
|---------------------|-------|-------|-------|-------|
| Phoenix Limited | 97.49 | 97.11 | 92.09 | 96.37 |
| Oberoi Realty | 24.19 | 21.82 | 21.3 | 21.98 |
| Prestige Group | 16.22 | 15.24 | 14.03 | 14.99 |
| Brigade Group | 12.83 | 12.31 | 10.51 | 10.68 |
| DLF | 9.72 | 8.77 | 8.98 | 9.44 |
| Godrej Properties | 42.5 | 38.46 | 36.39 | 39.78 |
| Sobha Limited | 24.75 | 23.22 | 22.94 | 23.65 |
| Mahindra Lifespaces | 8.78 | 8.39 | 7.12 | 7.74 |
| Sunteck Realty | 14.33 | 13.92 | 13.43 | 13.83 |
| Anant Raj Limited | 3.63 | 2.91 | 3.02 | 3.08 |

*Root Mean Square Error

The companies are ranked based on their RMSE values (Table 4), with lower RMSE indicating better forecast accuracy. This ranking allows for the identification of companies with more precise forecasting models, enabling stakeholders to assess the reliability of forecasts for each company's stock prices.

Table 4 – Ranking based on RMSE Values

| RANK | COMPANY |
|------|---------------------|
| 1 | Anant Raj Limited |
| 2 | Mahindra Lifespaces |
| 3 | DLF |
| 4 | Brigade Group |
| 5 | Sunteck Realty |
| 6 | Prestige Group |
| 7 | Oberoi Realty |
| 8 | Sobha Limited |
| 9 | Godrej Properties |
| 10 | Phoenix Limited |

The top three and bottom three companies were identified based on both consistency ranking and RMSE values (Table 5). For consistency ranking, the top three companies exhibit the least variability in their stock price movements, while the bottom three companies display higher levels of variability. For RMSE values,

the top three companies have the lowest RMSE, indicating more accurate forecasting models, whereas the bottom three companies have higher RMSE values, suggesting less accurate forecasts.

Table 5 - Top and Bottom 3 Companies based on Consistency Ranking and RMSE Values

| | Rank | Consistency Ranking | RMSE |
|---------------|------|---------------------|---------------------|
| Top | 1 | Sunteck Realty | Anant Raj Limited |
| | 2 | Godrej Properties | Mahindra Lifespaces |
| | 3 | Obero Realty | DLF |
| Bottom | 8 | Brigade Group | Sobha Limited |
| | 9 | Mahindra Lifespaces | Godrej Properties |
| | 10 | Anant Raj Limited | Phoenix Limited |

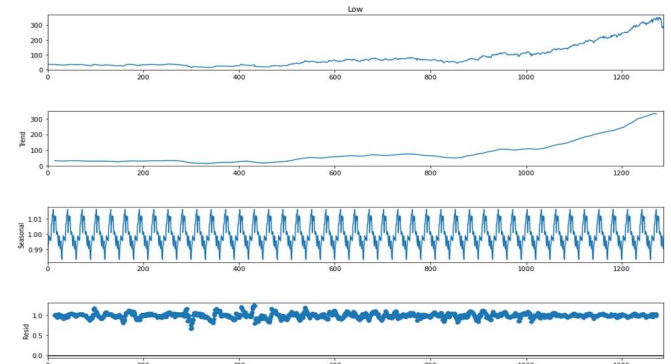
(C) Time Series Analysis and SARIMAX Results (ARIMA Modelling): The SARIMAX results for the top two companies from both consistency ranking and RMSE ranking include key metrics such as log likelihood, AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), and HQIC (Hannan-Quinn Information Criterion). Based on the values of these parameters, the attribute most suitable for stock price prediction can be recommended. This recommendation could be made by considering which model configuration yields the highest log likelihood, along with the lowest values of AIC, BIC, and HQIC. Lower AIC, BIC, and HQIC values indicate a better balance between model fit and complexity, while a higher log likelihood suggests a better fit of the model to the data.

Table 6: SARIMAX Results

| Sl. No | COMPANY | | LOG LIKELIHOOD | AIC | BIC | HQIC |
|--------|---------------------|-------|----------------|---------|---------|---------|
| 1 | Anant Raj Limited | Open | -3432.203 | 6880.40 | 6921.68 | 6895.90 |
| | | High | -3114.025 | 6244.05 | 6285.33 | 6259.54 |
| | | Low | -3171.367 | 6358.73 | 6400.01 | 6374.23 |
| | | Close | -3198.611 | 6413.22 | 6454.50 | 6428.71 |
| | | | HIGH | HIGH | HIGH | HIGH |
| 2 | Mahindra Lifespaces | Open | -4498.785 | 9013.57 | 9054.85 | 9029.06 |
| | | High | -4417.939 | 8851.87 | 8893.15 | 8867.37 |
| | | Low | -4154.335 | 8324.67 | 8365.95 | 8340.16 |
| | | Close | -4286.98 | 8589.96 | 8631.24 | 8605.45 |
| | | | LOW | LOW | LOW | LOW |
| 3 | Sunteck Realty | Open | -4865.873 | 9747.74 | 9789.02 | 9763.24 |
| | | High | -4776.969 | 9569.93 | 9611.21 | 9585.43 |
| | | Low | -4711.285 | 9438.57 | 9479.85 | 9454.06 |
| | | Close | -4777.489 | 9570.97 | 9612.25 | 9586.47 |
| | | | LOW | LOW | LOW | LOW |
| 4 | Godrej Properties | Open | -6525.766 | 13067.5 | 13108.8 | 13083.0 |
| | | High | -6347.85 | 12711.7 | 12752.9 | 12727.1 |
| | | Low | -6272.925 | 12561.8 | 12603.1 | 12577.3 |
| | | Close | -6411.418 | 12838.8 | 12880.1 | 12854.3 |
| | | | LOW | LOW | LOW | LOW |

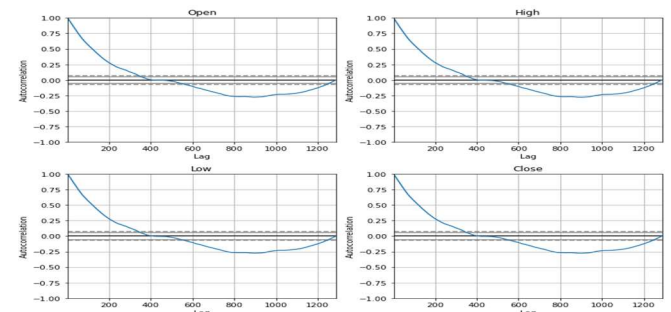
Based on the log likelihood, AIC, BIC, and HQIC values provided in Table 6, the SARIMAX model configuration for forecasting the high stock prices of Anant Raj Limited appears to be the most suitable, indicated by the highest log likelihood and the lowest values of AIC, BIC, and HQIC among the companies listed. Conversely, for Mahindra Lifespaces, Sunteck Realty, and Godrej Properties, the SARIMAX models for forecasting low stock prices seem to be more appropriate, as evidenced by comparatively higher log likelihood values and lower AIC, BIC, and HQIC values. These findings suggest that different companies may benefit from distinct SARIMAX model configurations tailored to their specific stock price behaviors.

Time Series Decomposition: Time series decomposition reveals that the data exhibits stationary characteristics, as indicated by minimal fluctuations in the residuals. This observation suggests lower associated risks with investing in the analyzed companies. Notably, the decomposition plots demonstrate similar patterns across all companies. For reference, the decomposition plot for Anant Raj Limited is provided below.



Graph 1 – Time Series Decomposition: Open Stock Prices (Anant Raj Limited)

Autocorrelation Plot (ACF): The Autocorrelation Function (ACF) plot depicts a slow decay, indicating a significant correlation between past and future values of the series. This suggests that the future movements of the series are heavily influenced by its past values. Notably, the ACF plot exhibits a similar pattern across all companies. For reference, the ACF plot for Anant Raj Limited is provided below.

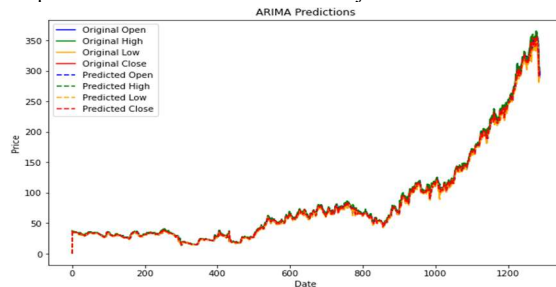


Graph 2 – Autocorrelation Plot(ACF)- Anant Raj Limited

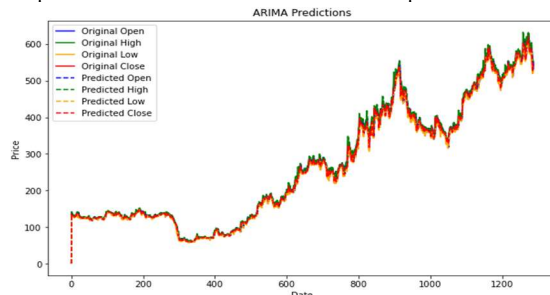
ARIMA Predictions: The ARIMA predictions for the open, high, low, and close stock prices of all 10 companies show a remarkable similarity between the predicted and original values, as evidenced by the almost overlapping graphs. This suggests that the ARIMA models effectively capture the underlying patterns and trends in the stock price data. For

reference, the graphs depicting the predicted and original values for Anant Raj Limited, Mahindra Lifespaces, Sunteck Realty, and Godrej Properties are provided below.

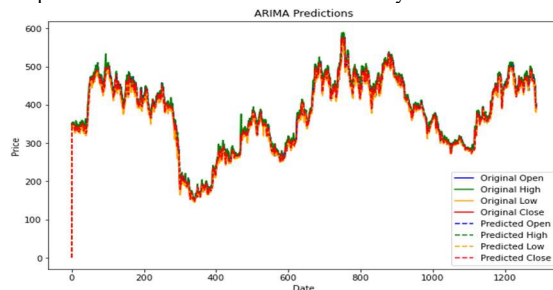
Graph 3 - ARIMA Predictions: Anant Raj Limited



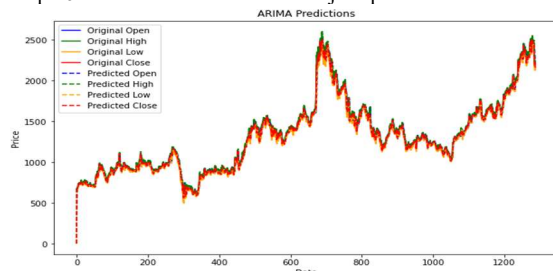
Graph 4 - ARIMA Predictions: Mahindra Lifespaces



Graph 5 - ARIMA Predictions: Sunteck Realty



Graph 6 - ARIMA Predictions: Godrej Properties



V. CONCLUSION

In conclusion, the study reveals several key insights. Firstly, time series decomposition and ACF plots indicate stationary data with significant autocorrelation, suggesting that past values heavily influence future movements, thereby providing valuable insights for forecasting. The consistency ranking based on the coefficient of variation (CV) along with ranking based on RMSE highlights companies with more stable stock price behaviors, aiding investors in assessing risk levels. SARIMAX modeling showcases the suitability of different model configurations for forecasting stock prices, with attributes like log likelihood, AIC, BIC, and HQIC guiding the selection process along with

demonstrating the effectiveness of the models in accurately capturing stock price trends and patterns, as evidenced by the close alignment between predicted and actual values. Overall, these findings provide valuable information for investors, enabling them to make informed decisions regarding investment strategies and risk management in the Indian realty sector.

VI. ACKNOWLEDGMENT

We would like to express our heartfelt gratitude to Prof. Dr. Jitendra Kumar, Department of Mathematics, for his invaluable guidance and support throughout our research, which has been instrumental in its successful completion.

VII. REFERENCES

- [1] Indian Real Estate Sector: Mapping India's Path to Economic Expansion. (2024, March 23). [www.business-standard.com. https://www.business-standard.com/content/press-releases-ani/indian-real-estate-sector-mapping-india-s-path-to-economic-expansion-124032300365_1.html](https://www.business-standard.com/content/press-releases-ani/indian-real-estate-sector-mapping-india-s-path-to-economic-expansion-124032300365_1.html)
- [2] C, V. R. (2019). Predicting the Stock Market Index Using Stochastic Time Series Arima Modelling: The Sample of BSE and NSE. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.>
- [3] Mondal, P., Shit, L., & Goswami, S. (2014, April 30). Study of Effectiveness of Time Series Modeling (Arima) in Forecasting Stock Prices. International Journal of Computer Science, Engineering and Applications, 4(2), 13–29. <https://doi.org/10.5121/ijcsea.2014.4202>
- [4] (n.d.). Time Series Analysis and Its Applications (3rd ed.). Springer. <https://mybiostats.files.wordpress.com/2015/03/time-series-analysis-and-its-applications-with-examples-in-r.pdf>. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.
- [5] Historical Stock Price Data. (n.d.). Yahoo Finance. Retrieved March 20, 2024, from <https://finance.yahoo.com/>