BOVESPA Index Forecasting

Khanh Tran, Roberto Primo
Department of Mathematics and Statistics and Department of Computing Science
Thompson Rivers University, Kamloops, Canada
trank22@mytru.ca, sanchesr22@mytru.ca

Abstract

This report examines forecasting the IBOVESPA index using two approaches: mean forecasting and volatility forecasting. Mean forecasting includes ARIMA, Dynamic Regression, NNAR, and STL models, while volatility forecasting involves Naïve, Mean, Drift, NNAR, Prophet, and GARCH models. The study analyzes the IBOVESPA index's behavior from 2017 to 2022, incorporating economic indicators to enhance forecasting accuracy. NNAR modelling proved to be the most consistent for mean forecasting, with the lowest Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), although Dynamic Regression also provided robust outputs for balancing complexity and fit. The Prophet model has the highest accuracy for volatility forecasting, with an RMSE of 0.0152 and a MAE of 0.01207. The findings aim to provide insights into the Brazilian stock market's trends and volatility, offering a resource for financial analysis and decision-making.

Index Terms

Index Forecasting, Volatility Modeling, Times Series Analysis, Machine Learning

I. Introduction

This report centers on the critical task of forecasting the IBOVESPA index, the most prominent stock market index in Brazil, which represents a broad spectrum of companies and industries. Forecasting stock indices, especially in volatile and diverse markets like Brazil, presents a complex challenge with significant implications for investors, policymakers, and the broader economy.

The Brazilian stock market has significant growth potential but is also renowned for its volatility due to a variety of factors, such as changes in economic policy, political unrest, and dynamics in the global market. Accurate forecasting of the IBOVESPA index is crucial for various stakeholders. Investors depend on these forecasts for informed decisions about asset allocation, risk management, and investment strategy.

This report aims to compare and evaluate mean forecasting and volatility forecasting – two prominent methods in financial market prediction – in the context of the IBOVESPA index. Our objective is to determine the efficacy of each method in capturing the index's behavior. Through the analysis of historical data, economic indicators, and contemporary forecasting techniques, this study strives to provide a valuable tool for predicting market trends in Brazil.

II. Background

The IBOVESPA index, is the primary benchmark index of the São Paulo Stock Exchange in Brazil, established in 1968. It is a total return index weighted by traded volume and is comprised of the most liquid stocks on the BM&F Bovespa. Reflecting not only stock price variations but also the distribution of dividends and benefits, the IBOVESPA is updated every four months, ensuring it accurately mirrors current market dynamics. This composition has remained consistent since its inception, making the IBOVESPA a reliable indicator of market trends and economic health. Its sector representation offers a comprehensive overview of Brazil's economic activities, proving invaluable for investors and analysts. The index serves as more than just a market indicator; it is a reflection of Brazil's economic stability and growth potential, influenced by factors including domestic policies, global market trends, political stability, and international conditions.

Forecasting the IBOVESPA's movements is a complex endeavour due to its inherent volatility and the myriad of factors influencing it. These include both internal economic conditions and external global influences, requiring sophisticated modeling and a deep understanding of market dynamics. The index's consistent methodology since 1968 further underscores the challenge of predicting its trajectory amidst evolving economic landscapes.

III. Data

The primary dataset for the IBOVESPA index was obtained from Yahoo Finance, accessed through the 'quantmod' package in R, a tool for quantitative financial modeling and trading strategy development. The period considered spans from January 1, 2017, to December 31, 2022. This five-year timeframe was chosen to provide a substantial temporal scope, allowing for the analysis of both short-term fluctuations and longer-term trends in the index. This period also captures a range of economic conditions, including two presidential elections, offering a robust dataset for modeling.

In addition to the index data, covariate data was included in our analysis, providing insights into various economic factors that might influence the IBOVESPA index. This data was sourced from the 'rbcb' package in R, designed specifically for interacting with the Brazilian Central Bank's API, ensuring direct access to official economic indicators. The covariates selected were IPCA, Brazil's primary measure of inflation; SELIC, the Brazilian Central Bank's primary interest rate, instrumental in monetary policy decisions; SPREAD, the difference between the borrowing and lending rates, indicating the cost of credit in the economy; and CDI, a benchmark interest rate for interbank loans in Brazil, reflecting the cost of borrowing money in the financial markets.

Given the varying periods of availability for these covariates, data transformation was necessary to align them with the daily frequency of the IBOVESPA index data. This alignment was critical to ensuring consistency in our analysis, allowing for accurate correlation and causation assessments between the IBOVESPA index and the selected economic indicators.

$$r_n = (1 + r_m)^{\frac{pm}{pn}} - 1 \tag{1}$$

Where:

- r_m is the original rate for frequency m (e.g., annual rate).
- r_n is the desired new rate for frequency n (e.g., monthly or daily rate).
- p_m is the number of periods in the original frequency m (e.g., 1 for an annual rate).
- p_n is the number of periods in the new frequency n (e.g., 12 for monthly, 252 for daily if considering 252 trading days in a year).

In addition, when using a volatility approach, we considered the log returns as the volatility measure.

$$R_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{2}$$

- R_t is the log return at time t.
- P_t is the price of the asset (e.g., IBOVESPA index value) at time t.
- P_{t-1} is the price of the asset at time $t-1$ (the previous period).
- ln denotes the natural logarithm.

IV. Method

This section outlines the methodologies employed in forecasting the IBOVESPA index, divided into two primary approaches: mean forecasting and volatility forecasting. Each subsection details the specific models and techniques used, along with key parameters and algorithm choices.

A. Mean Forecasting

For the mean forecasting of the IBOVESPA index, a combination of statistical and machine learning models was utilized to capture the underlying trends and patterns.

1) **ARIMA:** This model was employed for its effectiveness in modeling time series data with trends and seasonality. Key parameters such as the order of autoregression (p), the degree of differencing (d), and the order of moving average (q) were optimized based on the AICc (Akaike Information Criterion Correction).
2) **Dynamic Regression:** This approach incorporated the selected economic covariates (IPCA, SELIC, Spread, CDI) into a regression framework, allowing for the analysis of their impact on the IBOVESPA index. The model's coefficients were estimated using maximum likelihood, and the significance of each covariate was assessed.
3) **NNAR (Neural Network Autoregression):** The NNAR model, a form of nonlinear autoregressive model using neural networks, was applied to capture complex patterns in the data. The network architecture, including a hidden layer that includes several neurons, was determined through cross-validation.

B. Volatility Forecasting

Volatility forecasting involved a series of models, in addition to the ARIMA and NNAR methods that were also applied to the mean forecasting, we included some additional models that were implemented to establish baseline forecasts. These models provided a preliminary understanding of the index's volatility patterns. Followed by a focus on GARCH (Generalized Autoregressive Conditional Heteroskedasticity) models.

1) **Naïve, Mean and Drift:** The naïve model utilizes today's volatility as the predictor for tomorrow's, offering a direct approach, while the mean model, on the other hand, forecasts future volatility as the historical average, assuming a reversion to the long-term mean. Lastly, the drift model incorporates the average historical change in volatility, adding a trend perspective. These models collectively provide foundational insights into the IBOVESPA index's volatility patterns, forming the basis for comparison with more advanced models.
2) **Prophet:** Designed for forecasting time series data with strong seasonal patterns and several seasons of historical data. It is robust to missing data and shifts in the trend, making it suitable for daily stock index data like the IBOVESPA. The model includes components for trend, seasonality, and holidays, allowing for flexible yet interpretable modeling of time series volatility.
3) **GARCH:** Selected for its proficiency in modeling time-varying volatility, the GARCH model is particularly adept at handling the volatility clustering commonly observed in financial time series. Parameters, including the order of GARCH and ARCH terms, were optimized for the best performance. The model's effectiveness was then compared against the simpler baseline models to assess its capability to accurately capture the volatility of the IBOVESPA index.

All methods and models were implemented using the R programming language. The complete code, along with associated files and detailed documentation, is available in the project repository on DASC6510-01-Project-GroupG GitHub Repository.

V. Results

In analyzing the BOVESPA index's absolute value through various forecasting models, the NNAR(8,4) model emerges as the most accurate, as evidenced by its lowest RMSE and MAE values. This performance indicates that for the dataset and period in question, the NNAR model adeptly captures the BOVESPA index's underlying patterns and trends. Meanwhile, the ARIMA(3,1,3) and ARIMA(3,2,3) models show comparable accuracy, though the former slightly edges out the latter in terms of RMSE. The Linear Model

(LM) with ARIMA(2,1,0) errors, despite its low AICc suggesting an optimal balance between complexity and fit, lags in predictive accuracy, as reflected in its RMSE and MAE scores.

Selecting the most suitable model involves more than just evaluating accuracy metrics. The lower AICc value of the LM w/ ARIMA(2,1,0) model highlights its efficiency in balancing complexity and fit, crucial for practical, long-term forecasting. On the other hand, the higher AICc values for the ARIMA models, especially ARIMA(3,1,3), might hint at overfitting, potentially limiting their applicability to future data. The NNAR(8,4) model, notwithstanding its lack of AICc for complexity assessment, stands as a viable option for scenarios where forecasting precision is critical.

TABLE I
ARIMA MODEL COMPARISON FOR ABSOLUTE VALUE

Model	RMSE	MAE	AICc
ARIMA(3,2,3)	1477.360	1036.091	25163.25
ARIMA(3,1,3)	1476.021	1039.545	25169.68
LM w/ ARIMA(2,1,0)	1484.28	1040.855	25117.94
NNAR(8,4)	1356.331	1002.539	Not Applicable

TABLE II
MODEL COMPARISON FOR VOLATILITY

Model	RMSE	MAE
ARIMA	0.01537	0.01227
Drift	0.02065	0.01702
Naive	0.02087	0.0172
Mean	0.01534	0.01221
NNAR	0.01534	0.01220
Prophet	0.01524	0.01207
GARCH	0.01849	0.01499

In a similar vein, assessing the volatility forecast accuracy of models like traditional time series and GARCH on the Brazilian stock index requires a focus on key metrics like RMSE and MAE. Table II summarizes the performance of these models in terms of volatility forecasting. The ARIMA model demonstrates a commendable fit, with relatively low RMSE and MAE, efficiently capturing the index's variability. Conversely, the Drift and Naive model, with higher RMSE and MAE, shows limitations in accurately depicting the index's dynamics. Baseline models like Mean, closely rivaling ARIMA, and the NNAR model also perform well, highlighting their ability to capture non-linear patterns. The Prophet model offers competitive accuracy with an RMSE of 0.01524 and an MAE of 0.01207. GARCH, primarily designed for volatility dynamics rather than point estimates, shows moderately accurate RMSE and MAE compared to some traditional models, yet offers unique insights into volatility forecasting.

VI. DISCUSSION

In this report's comparative analysis, a deeper understanding of various models' forecast accuracy for the IBOVESPA index is unveiled, focusing on both mean and volatility forecasting. The analysis reveals that while models like ARIMA, NNAR, and Prophet show commendable performance in predicting the index's value, the GARCH model distinctly stands out in its ability to capture the nuanced dynamics of market volatility. This distinction is crucial, as the choice of the optimal forecasting model hinges on specific needs. Whether the focus is on pinpointing the index's value or understanding its volatility patterns, each model offers unique strengths.

The results of this study are not just academic but carry substantial practical implications. They equip market analysts and participants with an enhanced toolkit of robust methodologies for forecasting, crucial in a market characterized by its complexity and volatility. In particular, the insights gleaned from the GARCH model's performance emphasize the value of specialized models, especially in scenarios where understanding volatility is key to making informed decisions.

Looking ahead, there is fertile ground for further research. Opportunities lie in fine-tuning the parameters of the existing models, exploring the potential of ensemble methods that combine the strengths of various models, or integrating additional macroeconomic factors to enrich the analysis. Such advancements could lead to even more precise and reliable forecasts, ultimately benefiting a broad range of stakeholders in the financial market. This continuous evolution of forecasting methodologies underlines the dynamic nature of financial markets and the ongoing need for adaptive and sophisticated analytical tools.