Exploring the Efficacy of State Space Models in Forecasting: A Comparative Study with ARIMA

Tuhin Patra
patra.tuhin2020@gmail.com

July 30,2024

Abstract

Our project explores the effectiveness of State Space Models (SSMs) for forecasting and compares their results with those of the traditional ARIMA model. The primary objective is to assess the forecasting accuracy of SSMs, focusing on their ability to capture complex patterns and seasonal variations in time series data. We apply SSMs both with and without exogenous variables and juxtapose their performance against an ARIMA model. The evaluation is conducted using several error metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). Through this comparative analysis, we aim to emphasize the strengths of State Space Models in enhancing forecasting accuracy and to provide insights into their advantages over ARIMA in various forecasting scenarios.

1 Introduction

Accurate forecasting of financial time series data, such as stock market indices, plays a crucial role in investment decisions and economic planning. The Nifty 50, one of the major stock market indices in India, serves as a significant benchmark for the Indian stock market's performance. Given the dynamic and volatile nature of financial markets, reliable forecasting models are essential for predicting future trends and making informed decisions.

In this project, we focus on enhancing forecasting accuracy using State Space Models (SSMs), which are known for their flexibility in capturing complex patterns and seasonal variations in time series data. We aim to improve the SSM by incorporating various state components and examining their impact on forecasting performance. Specifically, we explore the effectiveness of SSMs with different state configurations and compare their results with those of the traditional ARIMA model, which has been widely used for time series forecasting.

The ARIMA model, though established and reliable, may not always capture the intricate patterns present in financial time series data as effectively as SSMs. By comparing the performance of SSMs with ARIMA, we seek to evaluate the advantages and limitations of each approach. This comparison will be based on several error metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), providing a comprehensive analysis of forecasting accuracy.

Through this project, we aim to highlight the potential of State Space Models in enhancing forecasting accuracy and to offer insights into their relative effectiveness compared to ARIMA. This analysis will contribute to a deeper understanding of forecasting methodologies and their applicability in financial contexts.

2 Literature Review

Forecasting financial time series data is a well-established area of research, given its critical importance in investment and economic planning. Various models have been proposed and extensively studied to predict stock prices and market indices, with State Space Models (SSMs) and ARIMA being among the most prominent. In this section, we have examined prior studies conducted by various researchers, critically evaluating their methodologies and findings to identify potential areas for future research. Through this analysis, we seek to identify areas where improvements can be made to enhance the results obtained by previous studies.

2.1 Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter.*

Harvey's seminal work[4] provides a comprehensive introduction to State Space Models (SSMs) and their application in time series forecasting. The book presents the theoretical foundation of SSMs, including local level models, local trend models, and seasonal models. Harvey's discussion on the Kalman filter is particularly relevant, as it offers a practical method for estimating the parameters of SSMs. This work establishes the versatility of SSMs in handling complex time series data, highlighting their ability to model various components such as trend and seasonality. Harvey's contribution is foundational for understanding how SSMs can be applied to forecasting tasks, making it a key reference for any comparative study involving SSMs.

2.2 Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control.*

Box and Jenkins'[1] classic text introduces the ARIMA (Autoregressive Integrated Moving Average) model, a cornerstone of time series analysis. The book details the methodology for identifying, estimating, and validating ARIMA models. It provides a systematic approach to modeling time series data by combining autoregressive and moving average components with differencing to achieve stationarity. This work has been pivotal in establishing ARIMA models as a standard tool for forecasting, particularly in economic and financial contexts. Box and Jenkins' comprehensive approach to ARIMA modeling makes it an essential reference for understanding its application and limitations in comparison to more flexible models like SSMs.

2.3 Zivot, E., & Wang, J. (2006). *Modeling Financial Time Series with S-Plus.*

Zivot and Wang's book[5] offers an in-depth exploration of financial time series modeling using S-Plus, with a particular focus on State Space Models. The authors discuss various extensions of SSMs, including those that incorporate exogenous variables and deal with non-stationary data. Their work demonstrates the effectiveness of SSMs in capturing structural changes and complex patterns in financial data. The comparative analysis provided in this book highlights the advantages of SSMs over traditional models like ARIMA, especially in scenarios involving structural breaks and irregular patterns. This reference is valuable for understanding the practical application of SSMs in financial forecasting and provides a basis for evaluating their performance against ARIMA models.

3 Methodology Adapted

In this section we illustrate some methodology which is used in this project.

3.1 State Space Model

State Space Models (SSMs) decompose time series data into various components using two primary equations:

- **Observation Equation:** Links observed data to latent state variables.
- **State Equation:** Describes the evolution of state variables over time.

Observation Equation:

$$y_t = Z_t \alpha_t + \epsilon_t$$

where:

- y_t is the observed data at time t.
- Z_t is the observation matrix.
- α_t is the state vector at time t.
- $\epsilon_t \sim N(0, H_t)$ is the observation error.

State Equation:

$$\alpha_t = T_t \alpha_{t-1} + R_t \eta_t$$

where:

- T_t is the transition matrix.
- R_t is the control matrix.
- $\eta_t \sim N(0, Q_t)$ is the state error.

Local Level Model with Seasonal Components
A local level model with seasonal components is employed:
Local Level Model:

$$y_t = \mu_t + \epsilon_t$$

$$\mu_t = \mu_{t-1} + \xi_t$$

where:

- μ_t is the local level at time t.
- $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ is the observation error.
- $\xi_t \sim N(0, \sigma_\xi^2)$ is the level error.

Seasonal Component:

$$y_t = \mu_t + \gamma_t + \epsilon_t$$

$$\gamma_t = \gamma_{t-s} + \omega_t$$

where:

- γ_t is the seasonal component at time t.
- $\omega_t \sim N(0, \sigma_\omega^2)$ is the seasonal error.

Incorporating Exogenous Variables
To enhance the model, exogenous variables (e.g., the Bank Nifty index) are included:

$$y_t = \mu_t + \gamma_t + X_t\beta + \epsilon_t$$

where:

- X_t is the matrix of exogenous variables.
- β is the vector of coefficients for the exogenous variables.

3.2 ARIMA Model

The Autoregressive Integrated Moving Average (ARIMA) model is a widely-used method for time series forecasting. The ARIMA model is characterized by three main components:

- **Autoregressive (AR) part:** This component models the relationship between an observation and a number of lagged observations (previous time steps).
- **Integrated (I) part:** This component involves differencing the raw observations to make the time series stationary.
- **Moving Average (MA) part:** This component models the relationship between an observation and a number of lagged forecast errors.

The ARIMA model combines these three components to capture different aspects of the time series data, making it a powerful tool for forecasting future values based on past data.

3.3 Kalman Filtering

Kalman filtering is a recursive algorithm used for estimating the state of a linear dynamic system from a series of incomplete and noisy measurements. In the context of State Space Models, Kalman filtering is essential for:

- **Parameter Estimation:** Updating the estimates of the state variables and model parameters.
- **Forecasting:** Generating future predictions based on the current state estimates.

The Kalman filter was used to estimate the parameters of the State Space Models and to refine forecasts. It allows for efficient handling of time-varying data and provides updated predictions as new data becomes available.

3.4 Evaluation Metrics

To compare the performance of the SSM and ARIMA models, several error metrics were calculated:

- **Mean Absolute Error (MAE):** Measures the average magnitude of the errors in a set of forecasts, without considering their direction.
- **Mean Squared Error (MSE):** Calculates the average of the squared differences between forecasted and actual values, giving more weight to larger errors.
- **Root Mean Squared Error (RMSE):** Provides the square root of the MSE, offering an interpretable measure of the average error magnitude.

- **Mean Absolute Percentage Error (MAPE):** Calculates the average absolute percentage error between forecasted and actual values, expressing accuracy as a percentage.

These metrics were used to assess and compare the accuracy of the forecasts generated by both the SSM and ARIMA models.

3.5 Implementation and Tools

The models were implemented using Python, with the 'statsmodels' library utilized for both the SSM and ARIMA models. Data handling and manipulation were performed with 'pandas', and visualizations were created using 'matplotlib'. The 'sklearn' library was used to compute evaluation metrics.

4 Experimental Evaluation

4.1 Data Collection

For this project, financial time series data was collected from Yahoo Finance[3] to evaluate and compare forecasting models. The data used includes adjusted close prices for several major financial indices.

Primary Dataset: Initially, data was collected for the Nifty 50 index, which serves as a benchmark for the Indian equity market. This dataset spans from January 1, 2010, to January 1, 2024, and includes monthly adjusted close prices.

Additional Datasets: To enhance the analysis, additional datasets were incorporated:

- **S&P 500 Index:** This dataset includes monthly adjusted close prices for the S&P 500 index, covering the same period as the Nifty 50 dataset. The S&P 500 serves as an exogenous variable to improve the forecasting models.
- **Bank Nifty Index:** Data for the Bank Nifty index was also collected, spanning from January 1, 2010, to January 1, 2024. This index is a subset of the Nifty 50 and focuses on the banking sector.
- **Sensex Index:** Monthly adjusted close prices for the Sensex index were obtained for the same period. The Sensex is another major Indian stock market index and provides additional context for the analysis.

4.2 Data Processing

- **Handling Missing Values:** All datasets were processed to handle missing values by forward filling. This technique ensures that any missing values are replaced with the most recent available data.
- **Resampling:** The data was resampled to a monthly frequency to standardize the time series. Monthly adjusted close prices were calculated by averaging daily prices within each month.

- **Training Set:** Includes data from January 1, 2010, to December 31, 2022. This subset was used to train the forecasting models, including both the state space model and ARIMA model.
- **Test Set:** Covers the period from January 1, 2023, to December 31, 2023. This subset was used to evaluate the performance of the models and to compare their forecasts with actual observed data.

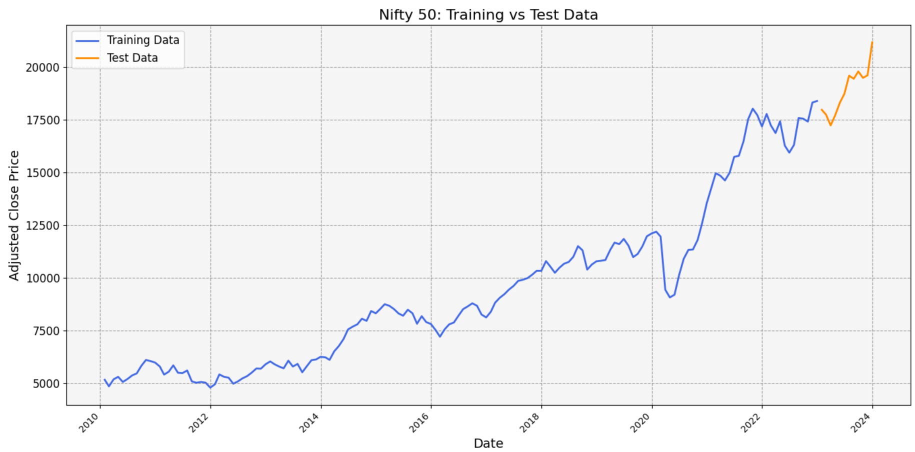

Figure 1: Nifty Data

The datasets were carefully aligned to ensure that the training and test sets are contiguous and to facilitate accurate forecasting and evaluation.

4.3 Result

In this section we will describe result obtained by state space models in various settings. We will also implement the ARIMA model on the nifty data to compare the results.

4.3.1 State Space Model

We have train the state model on training data and forecast the result on the test data. To evaluate we mainly consider Mean Absolute Percentage Error(MAPE). We have forest for one year that is 2023. We got lowest MAPE of 5.44%. To understand better we have plot the acutal and forcasted price.

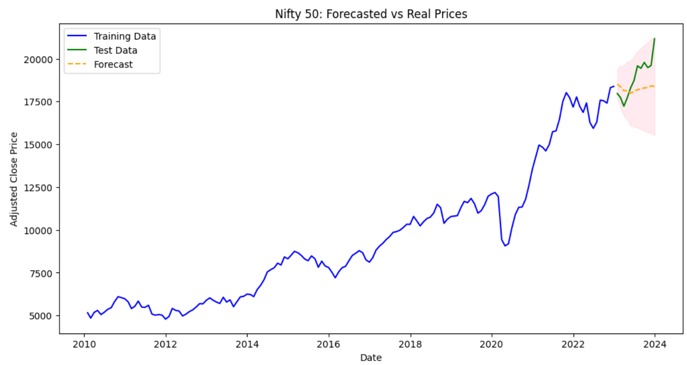

Figure 2: Forecasted and Acutal Price in State Space Model

4.3.2 State Space Model With S&P500

We have train the state model on combining S&P500 data with Nifty50 as a state and forecast the result on the test data. To evaluate we mainly consider Mean Absolute Percentage Error(MAPE). We have forest for one year that is 2023. We got lowest MAPE of 3.44%. To understand better we have plot the acutal and forcasted price.

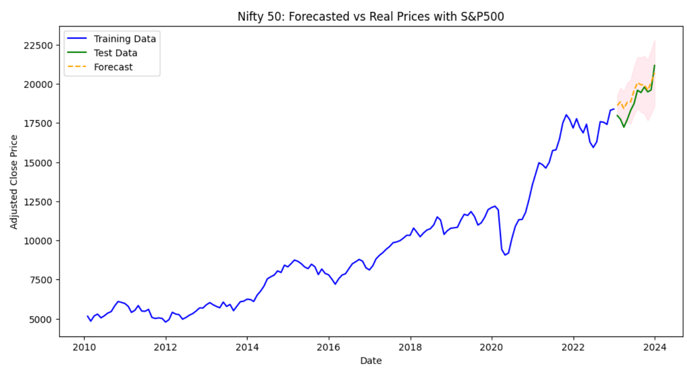

Figure 3: Forecasted and Acutal Price in State Space Model With S&P500

4.3.3 State Space Model With Bank Nifty and Sensex

We have train the state model on combining Bank Nifty and Sensex data with Nifty50 as a state and forecast the result on the test data. To evaluate we mainly consider Mean Absolute Percentage Error(MAPE). We have forest for one year that is 2023. We got lowest MAPE of 0.78%. To understand better we have plot the acutal and forcasted price.

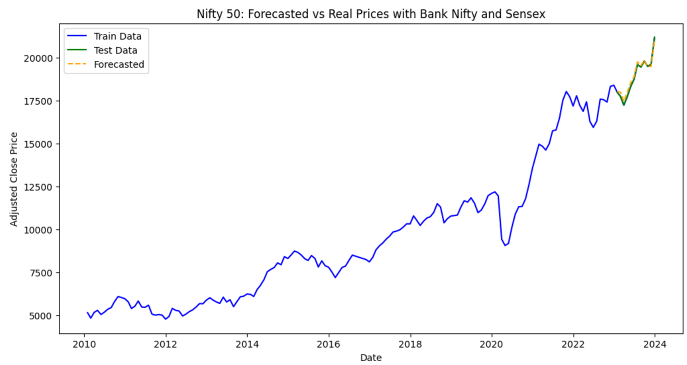

Figure 4: Forecasted and Acutal Price in State Space Model With Bank Nifty and Sensex

4.3.4 ARIMA Model

We have train the ARIMA model to compare with state model and forecast the result on the test data. To evaluate we mainly consider Mean Absolute Percentage Error(MAPE). We have forest for one year that is 2023. We got lowest MAPE of 5.28%. To understand better we have plot the acutal and forcasted price.

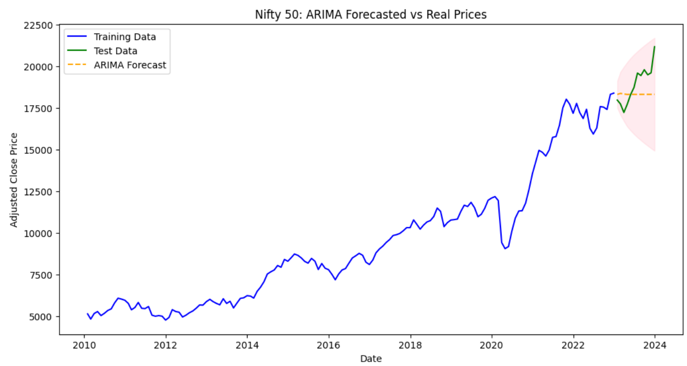

Figure 5: Forecasted and Acutal Price in ARIMA Model

4.3.5 Comparison Between Models

Here is the summary of results of various model.

Model	MAPE (%)
SSM	5.44
SSM With S&P	3.44
SSM With Bank Nifty And Sensex	0.78
ARIMA	5.28

Table 1: Summary of Models and their MAPE Values

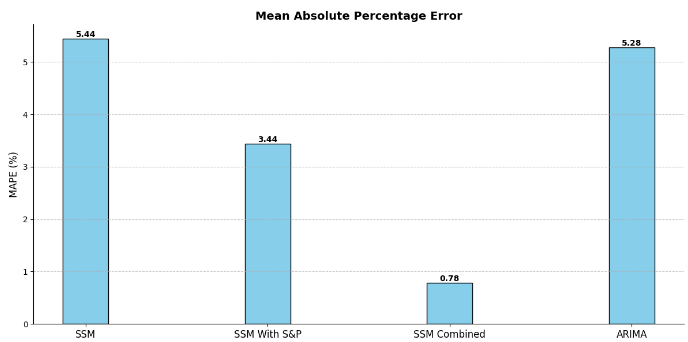

Figure 6: Comparison Between Models

So we can conclude that combined state space models with multiple states giving better result that normal ARIMA model.

5 Conclusion

This project evaluated the effectiveness of State Space Models (SSMs) for forecasting the Nifty 50 index and compared their performance with the ARIMA model. Our results showed that SSMs, especially when combined with exogenous variables like the S&P 500, Bank Nifty, and Sensex, significantly outperformed ARIMA. The best-performing SSM, incorporating Bank Nifty and Sensex data, achieved a MAPE of 0.78%, compared to 5.28% for ARIMA.

These findings highlight the superior ability of SSMs to capture complex patterns and external influences in financial time series data. The study underscores the potential of SSMs in enhancing forecasting accuracy, making them valuable for investment decisions and economic planning. Future research could further explore SSMs with additional exogenous variables and more sophisticated configurations to continue improving forecast performance.

Overall, this project demonstrates the advantages of State Space Models over traditional ARIMA models in financial forecasting, providing valuable insights for their application in various financial contexts.

6 References

[1] G. E. P. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, 1976.

[2] Peter J Brockwell and Richard A Davis. *Introduction to time series and forecasting*. Springer, 2002.

[3] Yahoo Finance. Yahoo finance. `https://finance.yahoo.com/`, 2024.

[4] A. C. Harvey. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, 1989.

[5] E. Zivot and J. Wang. *Modeling Financial Time Series with S-Plus*. Springer, 2006.