Week 1 Assignment Report
Time Series Forecasting Using ARIMA

Mahant

1 Introduction

Time series forecasting plays a crucial role in financial analytics by enabling the modeling and prediction of asset prices over time. One of the most widely used classical approaches for univariate time series forecasting is the Autoregressive Integrated Moving Average (ARIMA) model. ARIMA is particularly suitable for data exhibiting non-stationarity, which is a common characteristic of financial time series.

In this assignment, an ARIMA model is applied to forecast the stock prices of Tesla Inc. (TSLA). The primary objectives are to collect real-world financial data, analyze stationarity, perform model selection using grid search, generate forecasts, and evaluate model performance using standard error metrics.

2 Data Collection

Historical stock price data for Tesla Inc. (TSLA) was obtained using the Yahoo Finance API through the `yfinance` Python library. The dataset contains daily stock prices over the past two years, providing sufficient observations for time series modeling.

Among the available attributes (Open, High, Low, Close, Volume), the closing price was selected for analysis, as it reflects the final agreed-upon market value for each trading day.

3 Data Preprocessing

The time series was extracted from the closing price column and indexed by date. A check for missing values was performed, and any missing observations were handled using forward filling. This method is appropriate for financial data, where prices generally persist across non-trading days.

The dataset was divided into a training set (80%) and a testing set (20%) to ensure that the forecasting model was evaluated on unseen data and to prevent information leakage.

4 Stationarity Analysis

Stationarity is a key requirement for ARIMA modeling. To assess stationarity, the Augmented Dickey-Fuller (ADF) test was applied to the training dataset.

The null hypothesis of the ADF test states that the series is non-stationary. The resulting p-value exceeded the 0.05 significance level, indicating that the null hypothesis could not be rejected. Therefore, the series was considered non-stationary.

To address this, first-order differencing was applied, and the differencing parameter was set to:

$$d = 1$$

5 Model Development and Selection

A grid search approach was employed to identify the optimal ARIMA parameters. The following parameter ranges were explored:

- $p \in \{0, 1, 2, 3\}$
- $d = 1$
- $q \in \{0, 1, 2, 3\}$

Each candidate model was fitted on the training data, and the Akaike Information Criterion (AIC) was used to evaluate model quality. AIC penalizes overly complex models and balances goodness of fit with model simplicity. The ARIMA model with the lowest AIC value was selected as the final model.

6 Model Training and Forecasting

The selected ARIMA model was trained exclusively on the training dataset. Forecasts were then generated for the entire testing period. The predicted values were compared against the actual stock prices to assess the model's forecasting behavior.

The forecast produced by the model appeared flat across the prediction horizon. This outcome is expected in financial time series modeling and occurs when the selected ARIMA model behaves as a random walk without drift. In such cases, the expected future change is zero, and the optimal forecast corresponds to the most recent observed value. This reflects the inherent uncertainty and weak predictability of stock price movements rather than a modeling error.

7 Model Evaluation

The forecasting performance was evaluated using standard error metrics:

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)

These metrics quantify the deviation between the predicted and actual values. RMSE, in particular, penalizes larger errors more heavily and provides insight into the model's overall predictive accuracy. The evaluation results indicate that while ARIMA serves as a reasonable baseline model, its forecasting capability for stock prices is inherently limited.

8 Residual Analysis

Residual analysis was performed to validate the adequacy of the fitted ARIMA model. The residuals were examined for randomness and absence of systematic patterns.

The residuals closely resembled white noise, suggesting that the model successfully captured the underlying structure of the time series and that no significant information remained unexplained.

9 Conclusion

This assignment demonstrated the application of the ARIMA model for forecasting Tesla stock prices using real-world financial data. The process included data preprocessing, stationarity testing, model selection via grid search, forecasting, and evaluation.

The flat forecast observed in the results highlights a key property of financial time series: stock prices often behave like random walks, making precise future predictions challenging. Consequently, ARIMA models provide a strong baseline but may be complemented by more advanced techniques such as SARIMA or LSTM models for improved performance.