**Project Report: Time Series Forecasting Using ARIMA Model on Financial Portfolio Data**

**Moksha Gandhi**

**D019**

1. **Abstract**

This project applies the ARIMA (AutoRegressive Integrated Moving Average) model to forecast financial asset prices using time series data. The dataset was analyzed for stationarity using the Augmented Dickey-Fuller (ADF) test, and model parameters were selected based on ACF and PACF plots. The model's accuracy was evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The results indicate significant forecasting deviations, suggesting that financial data exhibits high volatility. The study highlights ARIMA’s limitations and suggests exploring advanced models like SARIMA or deep learning-based approaches for improved financial forecasting accuracy.

1. **Introduction**

Time series forecasting is an essential technique in financial analysis, helping investors and analysts predict future asset prices based on historical data. Accurate forecasting enables better decision-making in portfolio management, risk assessment, and investment strategies. This project applies the ARIMA (AutoRegressive Integrated Moving Average) model to a financial portfolio dataset to analyze asset price trends and predict future values. ARIMA is a widely used statistical model that effectively captures trends, seasonality, and residual errors in time series data.

The process begins with data preprocessing, where we check for missing values, convert date formats, and filter specific assets. We then assess the stationarity of the time series using the Augmented Dickey-Fuller (ADF) test, which is crucial since ARIMA assumes stationarity for optimal performance. If the data is non-stationary, differencing is applied to remove trends and stabilize variance.

After determining the appropriate parameters (p, d, q) using Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) plots, we train an ARIMA model and evaluate its performance using error metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The project aims to assess the accuracy of ARIMA in financial forecasting and explore its suitability for predicting future price trends in portfolio management.

1. **Literature Review**

Time series forecasting has been extensively studied in financial markets, given its critical role in stock price prediction, portfolio optimization, and economic forecasting. Box and Jenkins (1970) introduced the ARIMA model, which has since become a standard approach for analyzing and predicting time-dependent data. ARIMA models rely on three key components: autoregression (AR), differencing (I) to remove trends, and moving averages (MA) to smooth residual errors. This model has been widely used in financial applications due to its simplicity, interpretability, and effectiveness in short-term forecasting.

Several studies have demonstrated the effectiveness of ARIMA in financial forecasting. Chatfield (2000) discussed the importance of identifying stationarity in time series data before applying ARIMA models. Hyndman and Athanasopoulos (2018) highlighted ARIMA’s strength in capturing linear dependencies in time series data. However, financial markets are influenced by external factors such as economic policies, geopolitical events, and investor behaviour, which ARIMA alone may not fully capture.

Recent advancements integrate ARIMA with machine learning models to improve predictive accuracy. Hybrid models combining ARIMA with Long Short-Term Memory (LSTM) networks or GARCH models have shown promising results in capturing both linear and nonlinear relationships in financial time series. While ARIMA remains a fundamental technique for financial forecasting, its limitations suggest that integrating it with more complex models could enhance predictive performance. This study applies ARIMA to financial portfolio data to evaluate its accuracy in forecasting asset prices and to explore its real-world applicability in portfolio management.

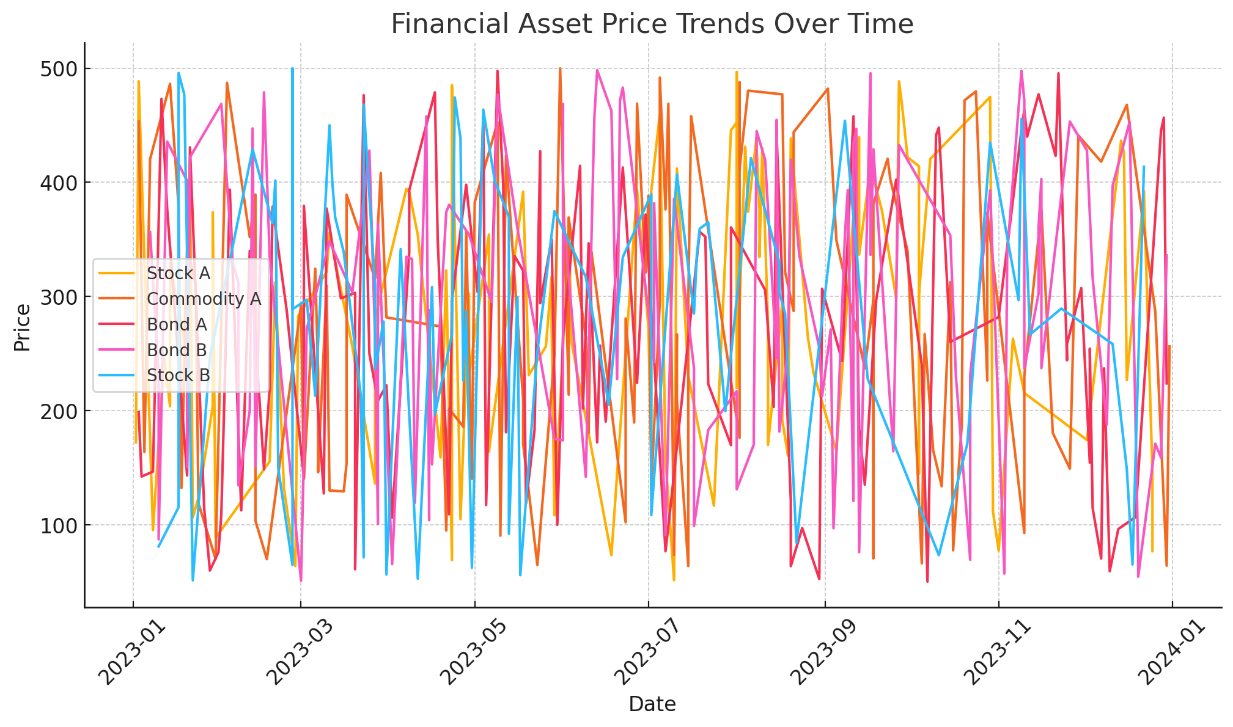
**Dataset Description**

The dataset used in this project contains historical financial portfolio data with the following attributes:

* Date: A timestamp for each recorded data point. This column was converted into a datetime format for accurate time series ordering.
* Asset: The dataset contains multiple financial assets. To apply ARIMA, we selected the asset with the most frequent records to ensure a continuous time series.
* Price: The closing price of the selected asset, which serves as the target variable for forecasting.

The dataset was first explored for missing values, inconsistencies, and outliers. An Augmented Dickey-Fuller (ADF) test was conducted to check for stationarity, which is essential for ARIMA modeling. The ADF test yielded a statistically significant result, indicating that the dataset was already stationary, and no differencing was required in ARIMA.

By understanding the dataset’s structure and statistical properties, we ensured the applied model was suitable for time series forecasting in financial analysis.



Time series line chart showing the price trends of different financial assets over time

1. **Methodology**

This project applies the ARIMA (AutoRegressive Integrated Moving Average) model to a financial portfolio time series dataset for forecasting future asset prices. The methodology follows a structured approach, including data preprocessing, stationarity testing, model selection, forecasting, and accuracy evaluation.

**4.1. Data Preprocessing**

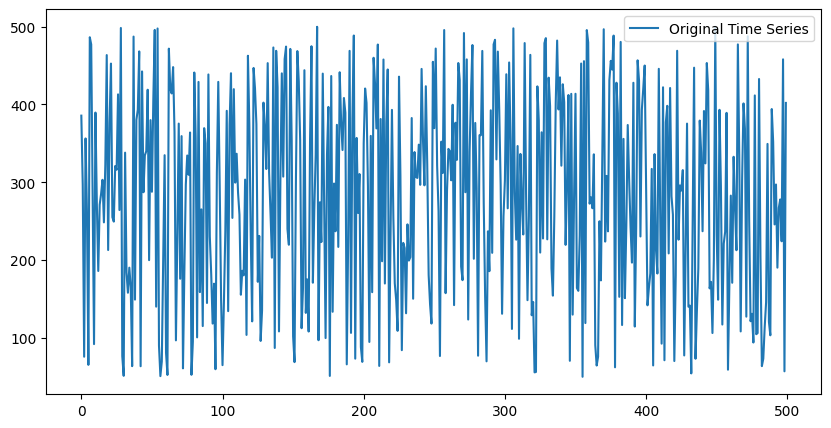
The dataset contains historical financial data, including asset prices and dates. The preprocessing steps involve:

* Loading the dataset: The dataset is read using Pandas, and the "Date" column is converted to a datetime format for proper time series handling.
* Filtering a specific asset: Since the dataset contains multiple assets, the most frequently occurring asset is selected for analysis.
* Sorting the data by date: Ensuring the time series is in chronological order is crucial for accurate forecasting.
* Setting the index: The "Date" column is set as the index to facilitate time series modelling.

**4.2. Stationarity Testing**

The ARIMA model assumes that the time series is stationary, meaning that its statistical properties (mean, variance, and autocorrelation) do not change over time. To check stationarity, we use:

* Augmented Dickey-Fuller (ADF) Test: If the p-value from the test is below 0.05, the data is stationary. Otherwise, differencing is applied to remove trends.
* Visual Inspection: A time series plot helps in visually identifying trends and seasonality.



Plot of original time series on Bond B

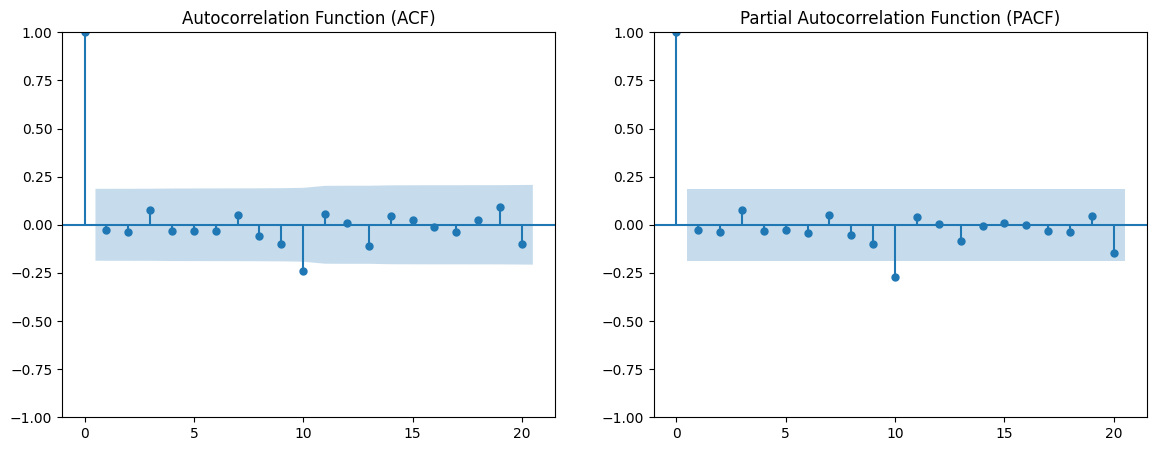
**4.3. Model Selection using ACF and PACF**

To determine the optimal parameters for ARIMA (p, d, q):

* Autocorrelation Function (ACF): Helps identify the moving average (q) component.
* Partial Autocorrelation Function (PACF): Helps determine the autoregressive (p) component.
* Differencing (d): If the data is non-stationary, differencing is applied (d=1).

Based on these plots, we choose an ARIMA(1,0,1) model, where:

* p = 1 (based on PACF)
* d = 0 (as the data is already stationary)
* q = 1 (based on ACF)



ACF and PACF plots

**4.4. Model Training and Forecasting**

The ARIMA(1,0,1) model is trained on historical data, and the fitted model is used to forecast the next 10 time steps. The forecasted values are compared with actual values for accuracy evaluation.

**4.5. Model Evaluation (Accuracy Metrics)**

The accuracy of the model is assessed using:

* Mean Absolute Error (MAE): Measures average absolute errors.

MAE=1n∑i=1n∣Actuali−Predictedi∣MAE = \frac{1}{n} \sum\_{i=1}^{n} |Actual\_i - Predicted\_i|

* Root Mean Squared Error (RMSE): Penalizes large errors.

RMSE=1n∑i=1n(Actuali−Predictedi)2RMSE = \sqrt{\frac{1}{n} \sum\_{i=1}^{n} (Actual\_i - Predicted\_i)^2}

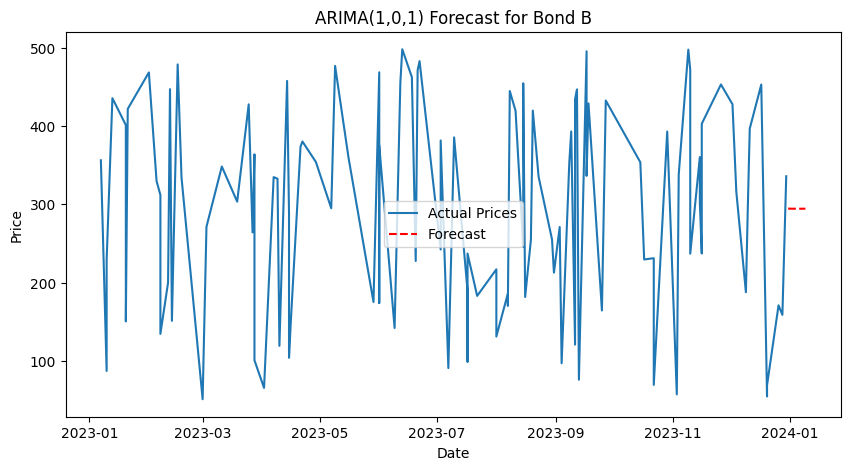
* Mean Absolute Percentage Error (MAPE): Measures the percentage error.

MAPE=100n∑i=1n∣Actuali−PredictediActuali∣MAPE = \frac{100}{n} \sum\_{i=1}^{n} \left| \frac{Actual\_i - Predicted\_i}{Actual\_i} \right|

**4.6. Time Series Plot of Actual vs. Forecasted Prices**

This plot compares the historical asset prices with the forecasted values from the ARIMA(1,0,1) model.

* The blue line represents actual prices.
* The red dashed line shows the predicted future values.
* The x-axis represents time (dates), while the y-axis represents asset prices.

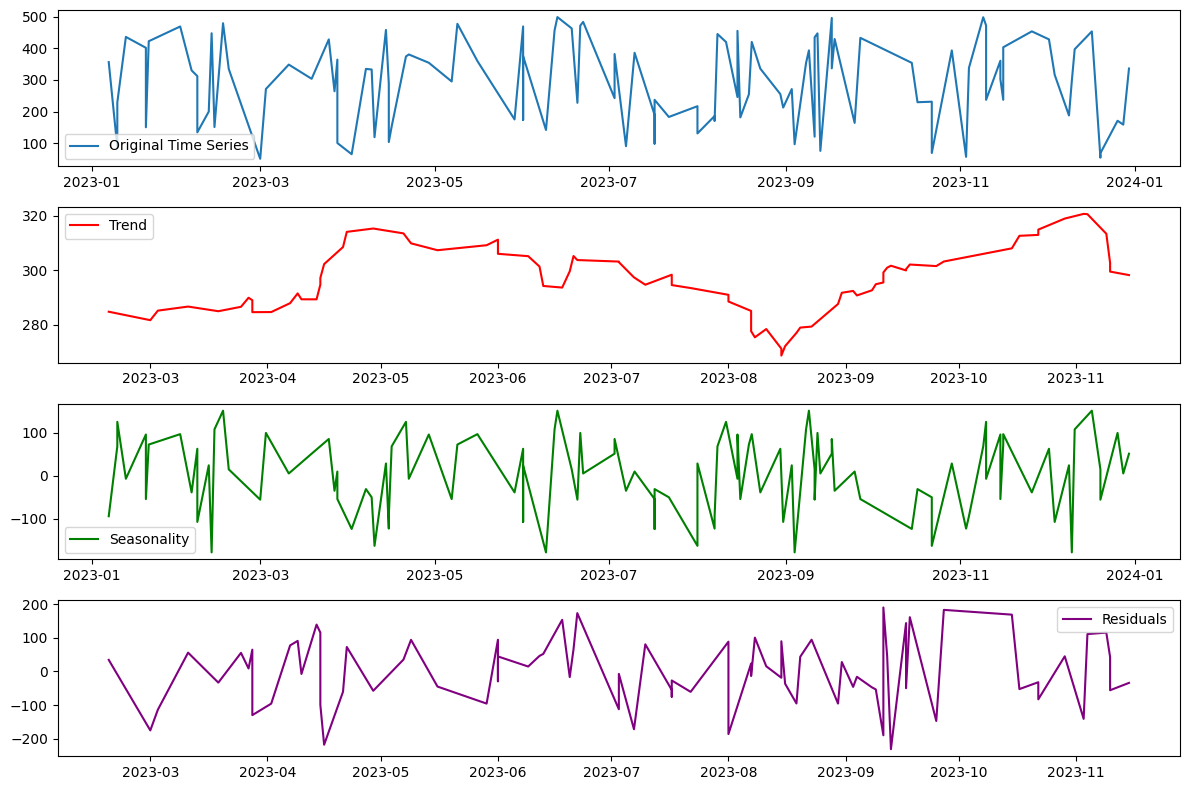


ARIMA(1,0,1) Model

**4.6. Decomposition**

To gain deeper insights into the structure of the financial time series data, seasonal decomposition was performed. This technique breaks down the time series into three fundamental components:

* Trend Component: Represents the long-term upward or downward movement in the price of the asset.
* Seasonal Component: Captures cyclic or recurring patterns observed at regular intervals.
* Residual Component: Accounts for the irregular fluctuations that cannot be explained by trend or seasonality.



Trend, Seasonality and Residual Component

* 1. **Results and Discussion**

The ARIMA(1,0,1) model was applied to forecast financial asset prices, and its performance was evaluated using **MAE, RMSE, and MAPE**. The obtained values were:

* **Mean Absolute Error (MAE):** 105.69
* **Root Mean Squared Error (RMSE):** 127.82
* **Mean Absolute Percentage Error (MAPE):** 62.72%

These results indicate that the model’s predictions have significant deviations from actual values, especially given the high **MAPE (62.72%)**, suggesting that the dataset exhibits high volatility, making it difficult for a linear model like ARIMA to capture complex patterns.

The stationarity test (ADF test) confirmed that the dataset was already stationary, meaning that **differencing was not required**. The ACF and PACF plots helped determine the AR and MA components. While ARIMA provided insights into time series forecasting, the results highlight the **limitations of traditional models** in predicting financial markets. Alternative approaches like **SARIMA or deep learning models (LSTMs, GRUs)** could improve prediction accuracy.

* 1. **Conclusion**

This project successfully applied the **ARIMA (1,0,1) model** to forecast financial asset prices using a time series dataset. The analysis began with **data preprocessing**, where missing values were handled, and the dataset was structured for time series modeling. **Stationarity testing** using the **Augmented Dickey-Fuller (ADF) test)** confirmed that the dataset was already stationary, eliminating the need for differencing. The **Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots** were used to determine the appropriate ARIMA parameters (**p=1, d=0, q=1**).

The trained ARIMA model was used to predict future asset prices for the next **10 time steps**, and its performance was evaluated using **Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE)**. The high MAPE value suggests that the model's predictions have significant deviations, likely due to high volatility in financial data.

While the ARIMA model provides a structured approach to forecasting, its assumptions about linearity and stationarity may limit its effectiveness in highly volatile financial markets. Future improvements could include **SARIMA (Seasonal ARIMA) or machine learning models** like LSTMs for better accuracy. Overall, this study demonstrates the power and limitations of ARIMA in financial forecasting.