ARIMA Model for Time Series Forecasting

Nikhila Pothamsetty

December 21, 2025

1 Introduction

The ARIMA model is used for analyzing and forecasting time series data.It is composed of three main components: Autoregression (AR), Integrated (I), and Moving Average (MA).

2 Theoretical Concepts of ARIMA

2.1 Components of ARIMA

- **Autoregressive (AR) -** p: This component indicates that current value of the time series is linearly dependent on its own previous values. parameter p is the number of lag observations included in the model
- **Integrated (I) -** d: This component refers to the use of **differencing** to make the time series stationary. The parameter d is the number of non-seasonal differences needed for stationarity.
- **Moving Average (MA) -** q: This component indicates that the current value is linearly dependent on the residual errors from a moving average of past forecast errors.parameter q is the size of the moving average window.

2.2 Key Concepts

- **Stationarity**: A time series is stationary if its statistical properties, such as mean, variance, and autocorrelation, are constant over time. ARIMA models assume the underlying series is stationary.
- **Differencing**: This transformation calculates the difference between consecutive observations to remove trends.
- **ACF and PACF Plots**: The Autocorrelation Function (ACF) helps determine the q parameter, while the Partial Autocorrelation Function (PACF) helps determine the p parameter.

3 Implementation and Results

3.1 Data and Preprocessing

The dataset used was the historical closing price of **Microsoft Corporation (MSFT)** stock from January 2023 to January 2025.

1. **Stationarity Test**: The **Augmented Dickey-Fuller (ADF) test** yielded a p-value of 0.189 (non-stationary).

2. **Differencing**: First-order differencing ($d = 1$) was applied, resulting in a stationary p-value of 0.000.

3. **Data Split**: The series was split into training (80%) and testing (20%) sets.

3.2 Model Selection and Fitting

A grid search was performed to find the optimal model order based on the lowest **Akaike Information Criterion (AIC)**.

Parameter	**Value**
Ticker	MSFT
Differencing Order (d)	1
Best ARIMA Order (p, d, q)	**(2, 1, 1)**
Best AIC	2422.08

Table 1: ARIMA Model Selection Results

3.3 Model Evaluation

Metric	**Value**
Mean Absolute Error (MAE)	419.94
Mean Squared Error (MSE)	176489.53
Root Mean Squared Error (RMSE)	420.11

Table 2: Forecast Error Metrics

4 Visualization and Analysis

4.1 Time Series and Forecast Plot

The forecast generated a near-straight line projection. This highlights the limitation of ARIMA in capturing the high volatility and non-linear trends inherent in MSFT stock data over long horizons.

4.2 Residual Analysis

- **Trend in Residuals**: Residuals exhibited a strong trend, indicating model misspecification.

- **Autocorrelation**: ACF/PACF plots of residuals showed spikes outside confidence intervals, meaning the model failed to capture all temporal information.