2)

Summary of the Training Dataset

The data was from January 5, 2009, to September 14, 2010. Istanbul Stock Exchange (ISE) index had a minimum value observed is -0.062208, the maximum value is 0.068952, and the mean value is 0.001964. Similarly, the Standard & Poor's 500 (S&P 500) index had a minimum of -0.0542620, a maximum value of 0.0683664, and a mean value of 0.0004331. The Financial Times Stock Exchange (FTSE) index, reflecting the performance of the UK stock had a minimum of -0.0548160, a maximum value of 0.0503227, and a mean value of 0.0004665.

The Nikkei index, signifying the performance of the Japanese stock market had a minimum value of -0.05045, a maximum value of 0.06123, and a mean value of 6.542e-05. The Bovespa index, which indicates the performance of the Brazilian stock market. It exhibits a minimum value of -0.0538495, a maximum value of 0.0637915, and a mean value of 0.0012178. The European Union (EU) stock market index had a minimum value of -0.0488168, a maximum value of 0.0670425, and a mean value of 0.0004315. The Emerging Markets index had a minimum value of -0.038564, a maximum value of 0.047805, and a mean value of 0.001087.

Outliers

The results showed that the outliers ranged between 9 and 3. With 9, Dax had the highest outliers followed with FTSE, MSCI Emerging Markets, S&P 500 and Euro Stoxx 50 which had, 7, 7, 6 and 6 respectively. Nikkei 225 and Bovespa each had 5 outliers while Istanbul Stock Exchange had 3.

Missing data

The results showed that training dataset does not have any missing data.

Distribution of Data

Histograms show data is approximately normally distributed.
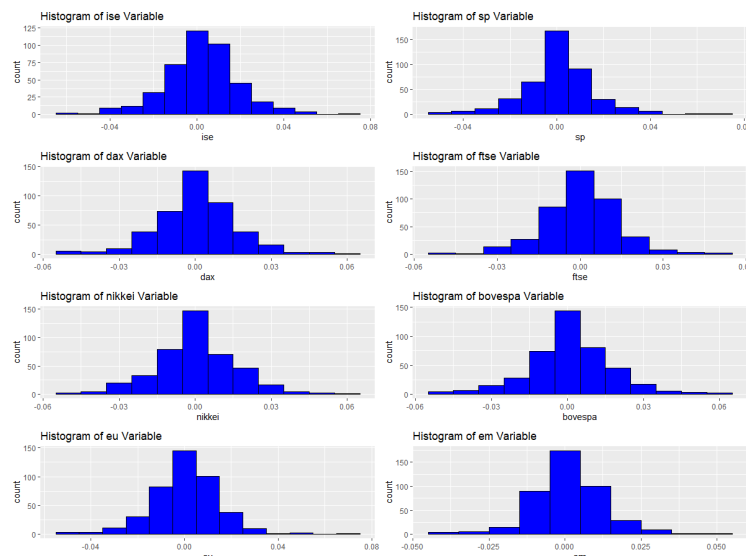
Correlation

Highly correlated variables include FTSE and EU (0.952), DAX and EU (0.940), and DAX and FTSE (0.877), showing strong positive associations. Conversely, weaker correlations are seen between EU and SP (0.132), EU and Bovespa (0.192), and Nikkei and EU (0.305), suggesting less pronounced positive links.

3)

a) Modelling

The data was checked for stationarity using augmented dicky fuller test. With a highly negative Dickey-Fuller test statistic of -12.567. and a low p-value of 0.01 the null hypothesis of non-stationarity. Is rejected. The data can be used for forecasting.

For the ARIMA (0,0,0) model, the AIC value is -2270.644. This value quantifies the trade-off between model complexity and goodness-of-fit. The negative value indicates that this model's fit is better relative to a baseline model.

For the Auto ARIMA model, the AIC value is -2253.60. Auto ARIMA automatically searches for the best combination of ARIMA parameters based on the data. In this case, it has found a combination that results in a slightly better AIC value compared to the ARIMA (0,0,0) model. The Auto ARIMA model has a slightly better AIC, suggesting that it might be preferred due to its automatic parameter selection process.

ACF of Differenced Time Series **PACF of Differenced Time Series**
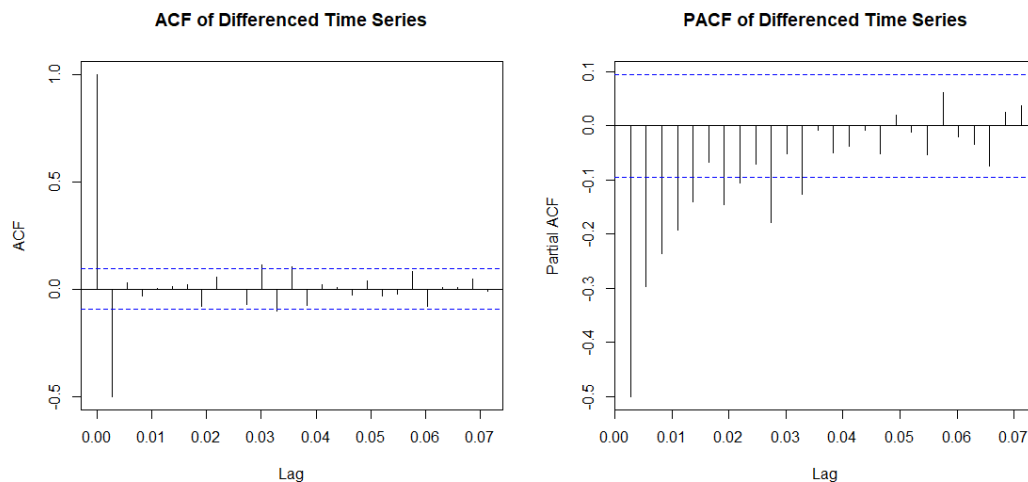
b) Assumptions

- Stationarity: The time series should be approximately stationary,
- Autoregressive Terms: The relationship between the current value of the time series and its past values is linear.
- Moving Average (MA) Terms: The relationship between the current value of the time series and its past residuals is linear.
- Seasonality: The model should account for seasonality

- Weak Dependency: The residuals should be uncorrelated and exhibit no discernible pattern.

PACF of Residuals
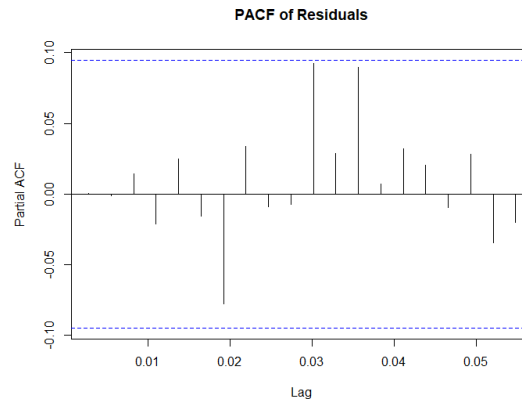
c) Forecast the Istanbul Stock Exchange 100 Index for the test dataset

The ARIMA (0,0,0) had forecasted mean value of 0.001993884. The auto ARIMA had a forecasted value of 0.001948731. The RMSE of 0.0130 and MAE of 0.0099 suggests that, on average, the forecasted values are accurate within this range. The MPE of 100% indicates that, on average, the forecasted values are 100% higher than the actual values. A value of 100% signifies perfect match. Additionally, the analysis confirms that the ISE data shows *white noise* characteristics, supporting the effectiveness of the selected forecasting approach.

Forecast using ARIMA (0,0,0)
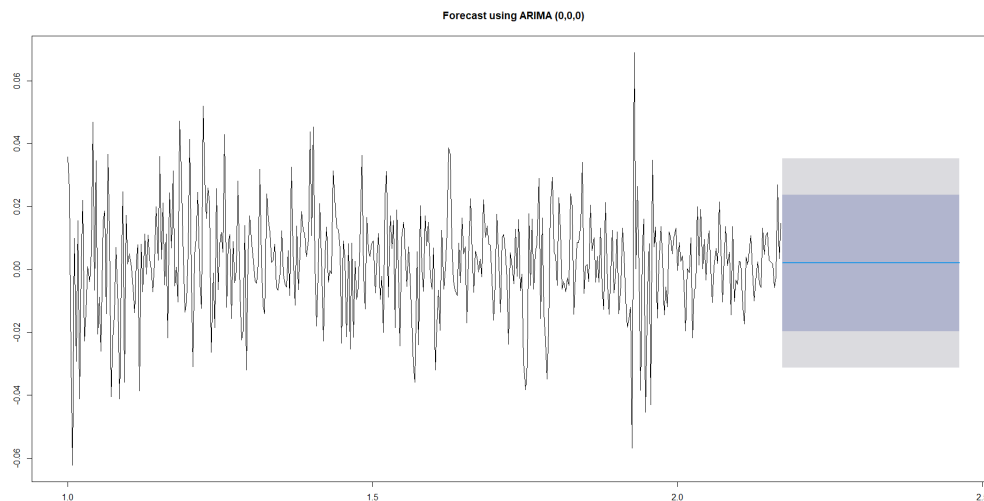
d) Weakness of the Analysis

- The data is not from a wide timeframe hence does not capture other cyclic realities
- The model is sensitive to outliers
- Requires selecting p, d, q parameters, which can be challenging and impact accuracy.
- Assumes linear relationships, not suitable for all data patterns.
- Does not effectively capture complex long-term trends.
- Struggles with data with sudden structural changes.
- Quality and quantity of historical data affect performance.