Scope
- Train set period : '2020-10-01 ' - '2023-09-30'
- Test set = '2023-10-01' - '2023-11-29'
- Close price

Introduction
Problem
- Our main research question is which forecasting algorithm is the best for this type of data set/series. What we want to achieve is to have a good comparison of different forecasting algorithms to find out which of the algorithms has the best prediction accuracy for the investigated data set.

Who cares
- DJI influences daily life. Nowadays, whether you eat well, land a job, or have good mental health depends on which country you live. Thus, DJI is a good indicator to decide where to live to have a better life/cost of living.

Data set analysis:
- Where it obtained, what is it
- mean, median, standard deviation, and skewness to describe central tendency, variability, and distribution shape
- Outliers
- Decomposition
  - Has a trend and remainder, thus non stationary.

Key process of experiment

Model	Predictors	Train RMSE	Test RMSE
TSLM	Visa Close	1493.8102	1493.8102
DR - ARIMA(0,1,0)	Visa Close	227.1161	467.969
DR - ARIMA(3,1,3)	FTSE Close	272.4554	1099.418
DR - ARIMA(0,1,1)	N225 Close	303.3204	731.225
DR - ARIMA(0,1,3)	FTSE + N225 Close	271.5634	1006.004

- Used linear and dynamic regression
1. Model & predictors selection reason :
    a. Predictors
        i. Visa company belongs to 4% of DJI. Thus Visa has correlation with DJI.
        ii. An effect of big markets (FTSE, N225, HSI) can affect other big markets.
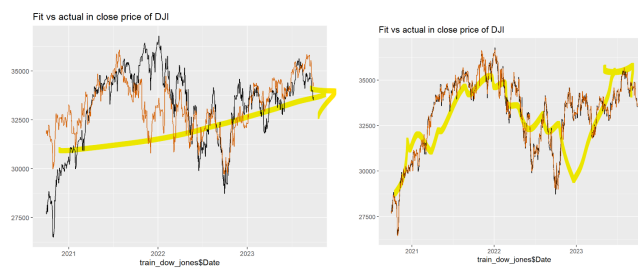   b. Model
        i. Dynamic is a backup for regression to capture TSLM leftover residuals
            E.g. DJI plot shows a upward trend, but with some random walk pattern in
   ¾ period.
1. Key results (why one was suitable/not suitable)
    a. Group Green (a company that belongs to DJI is used as predictors)
        i. TSLM only captures an upward trend.
        ii. [best] ARIMA(0,1,0) [Random walk, can't predict based on past] captures
            not only trends, but autocorrelations.

b. Group Yellow (other international markets (Japanese, HongKong, British) are used as predictors)

2023-08-01 to 2023-11-29 Dow Jones (forecast)
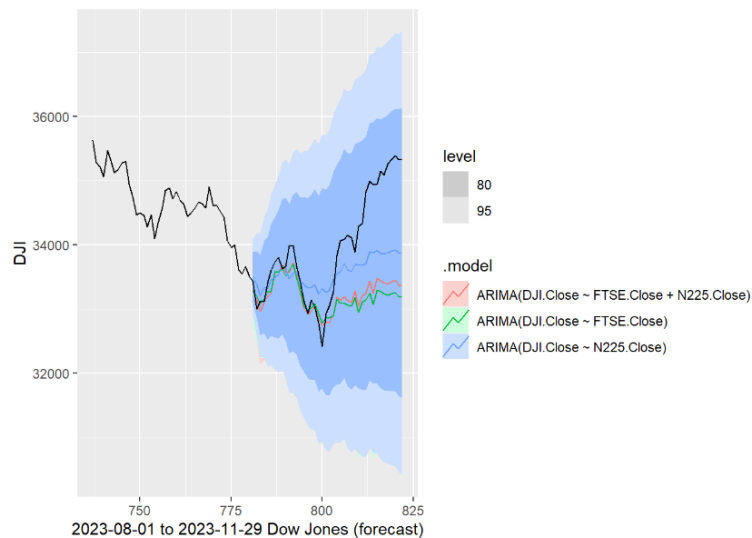
Key highlights: some models were able/not able to capture upward trend from day800th

Skip analysis. Read the red one below

Analysis

we know differencing was to capture the dynamics more.

ARIMA(3,1,3) errors with FTSE :

1. FTSE was not a good predictor, but it didn't have enough correlation with DJI.

2. AR of lag 3 was not effective, thus DJI can be random walk.

ARIMA(0,1,1) errors with N225 :

1. N225 is a good predictor for DJI.

2. fixing the error for the model is best done by only considering lag 1. a large MA lag value doesn't work.

ARIMA(0,1,3) errors with FTSE & N225

1. FTSE was not a good predictor, N225 solely itself as predictor was effective.

2. MA of lag 3 was not effective

Conclusion

1. Effective predictor :DJI seems to be influenced by N225 more than FTSE.

2. Short lags are effective : Short term movements in DJI series are best captured by immediate past, high order lag don't significantly improve the model.