Group assignment 2, Group F

COCA-COLA TIME SERIES ANALYSIS

Our approach to find the best linear models was like the following:

- We looked at the plot of the time series and the ACF & PACF:
    - TS plot we could clearly see that its non-stationary in the mean as well as in the variance. For the variance we could see that we have a clear trend of increasing fluctuation and not actually just times of higher and lower volatility.
    - ACF: We saw that the decay was there but not as “fast” as it should be for a stationary data in the mean.
- To handle the non-stationary in the variance we took the log of the time series in order to achieve stationary in the variance.
- After getting a feeling of the non-stationarity of the data we ran the dicky fuller test and came to the result that the data is clearly not stationary in the mean and that we need to take one difference in order to achieve stationarity in the mean.
- Since we are having quarterly data, we set the seasonality coefficient to 4 (s=4). Following to that, we run the Osborn, Chui, Smith, and Birchenhall Test and found that we should take one seasonal difference.
- After that, we ran the SARIMA model, taking the time series transformed with the log. Our model has the following form:
  fit< -arima (z, order=c(0,1,0),seasonal=list(order=c(0,1,0),period=s))
- The methodology was to keep adding different lags to the model that we thought were going to be relevant for the predictions, looking at the PACF and ACF of the residuals from the model. Also, we have to test if these new introduced lags were significant, dividing the estimated value by their standard error, and keeping only those whose statistic was higher than 1.96.
- With this procedure, we were looking to obtain White Noise in the residuals, so to confirm it, we used the Box test.
- After trying out different combinations based on ACF and PACF, we found the following models to be appropriate:
    1. SARIMA (1,1,0)·(0,1,1)
    2. SARIMA (0,1,1)·(3,1,0)
    3. SARIMA (0,1,3)·(0,1,0)
    4. SARIMA (1,1,0)·(2,1,0)

What is the best model in terms of forecasting and why is this your choice?

- In order to test which model performs best, we left 24 observations for the prediction set and used the MAPE (Mean Absolute Percentage Error) in order to validate our predictions. Applying this method, we have come to the following conclusions:
    - All SARIMA models found have similar performance when it comes to predicting new data, with an error of around 50%. This is because our models have overestimated the predictions, with higher values than the real ones. Even so, we were able to capture the rising trend, so if our prediction as investors was that the earnings per share were going to increase, we would have made a profit, as our prediction was right. Also, all models have captured the cycles pretty well
    - Finally, the best model was based on the mean absolute percentage error is **the SARIMA (1,1,0) x (2,1, 0)**, with an error of 49,21%.
    - Errors of the rest of the models:

MODEL	MAPE
SARIMA (1,1,0) x (2,1, 0),	**49,21**
SARIMA (0,1,1) x (3,1, 0),	52,84
SARIMA (0,1,3) x (0,1, 0),	53,90
SARIMA (1,1,0) x (0,1, 1),	57,23

- Direct representation of our forecast vs real values: