IE School of Human Sciences & Technology

Master's in Big Data & Business Analytics

Course Name:

BIG DATA & ARTIFICIAL INTELLIGENCE IN OPERATIONS MANAGEMENT

Supervised by:

Prof. LORENZO PASCUAL CANEIRO

Submitted by:

Group H

Member Names:

JUAN LUIS BETANCUR BOTERO

MANUEL MARINA HERRERA

ALICE SEYNAEVE

KRISHNA AGRAWAL

ALI ELMONOFY

THOMAS AUGUST MARIA STENGER

ANTOINE DE MAINTENANT

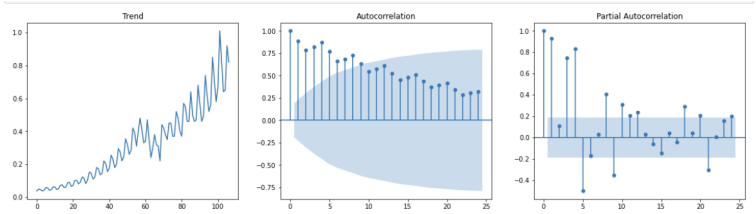
Table Content

1. Model Selection

1.1 Original Time Series

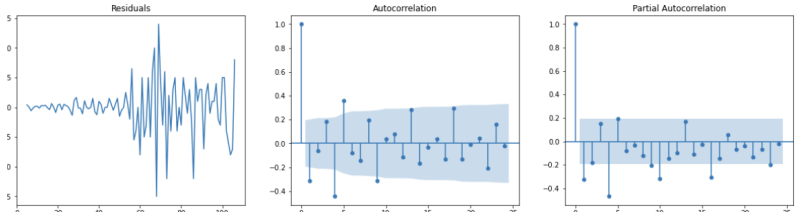
The quarterly earnings (Seasonality = 4) for the Coca Cola company are not ready to be entered into a Sarima model, because it's not stationary in the mean (slow decay in the ACF graph).

1. Coca Cola earnings without transformation



Using the formulas ndiffs and nsdiff the time series need a 1 difference operator in both seasonal and regular trend. The Sarima model with those differences ((0,1,0)(0,1,0)) has out of bound points in the acf and pacf plots, meaning that the model can be improved.

2. Coca Cola earnings with 1 seasonal and 1 regular difference



1.2 Proposed Model

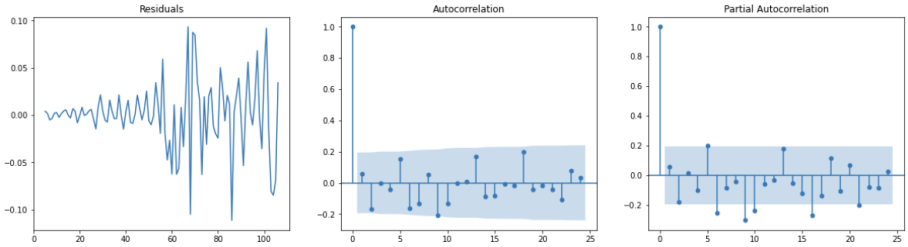
1.2.1 Sarima (0,1,1) (0,1,1)

The first model implemented was the Sarima (0,1,1)(0,1,1). As shown in the ACF chart above there are 3 clear points out of bounds 1,4 and 5. Therefore, the Sarima model would be with (0,1,1) in the regular part accounting for points 1 and 5, and in the seasonal part (0,1,1) accounting for point number 4 that is the first seasonal element.

As shown in the chart number 3, the proposed model residuals have a mean of 0 and are not autocorrelated (ADF, PACF and box text p_val = 23%) so they are white noise (not gaussian because residuals are not normal), meaning that this model can predict the future and cannot be improved in a linear way. The only caveat is that the residuals for this model looks with unequal variance. The first part

of the series is small, and in the second part the variance increases. For that reason, the next model would be implemented with a Log transformation.

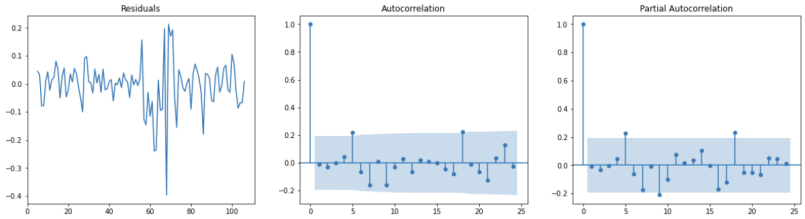
3. Residuals, ADF and PACF (0,1,1)(0,1,1)



1.2.2 Sarima (0,1,1) (0,1,1) with log transformation

By implementing a log transformation on the previous sarima we are now capable to state that we have a stable variance even though there are some points that are showing some peaks, as shown in the chart number 4. The proposed model residuals have a systematic mean of 0 and there isn't any point which is clear out of bounds (there are only two points 5 and 18 which are borderline). So we could say that it is not autocorrelated (ADF, PACF and box text p_val = 60%) and therefore, they are white noise. This model can predict the future and cannot be improved further in a linear way.

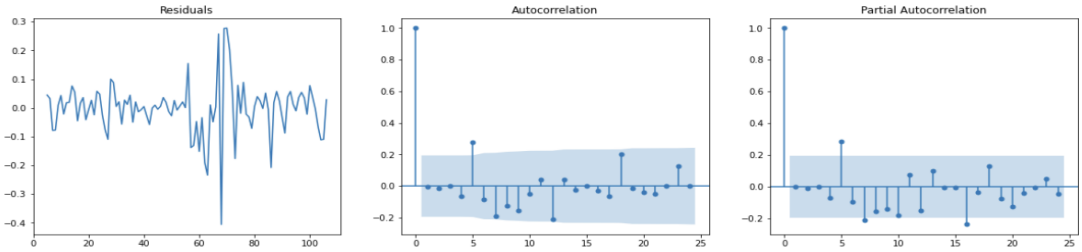
4. Residuals, ADF and PACF (0,1,1)(0,1,1) log



1.2.3 Sarima (1,1,1) (2,1,0) with log transformation

After implementing the log transformation on the original sarima, we have achieved the stationarity in the mean and the variance (with some outlier values with higher peaks). As shown in the chart number 5, the proposed model residuals have a mean of 0 and are not autocorrelated (ADF, PACF and box text p_val = 26%). Hence they are white noise (not gaussian because residuals are not normal), meaning that this model can predict the future and cannot be improved in a linear way.

5. Residuals, ADF and PACF (1,1,1)(2,1,0) log



2. Model Prediction

3 models were found that can predict the future and which residuals are white noise. For calculating the best performing model, the last 24 quarters were excluded for the training test and used as a test set. The chosen metric to validate the best model was mean average percentage error (MAPE), because it gives an easy understanding of how good the model is at predicting.

The performance of each model was measured for 1, 4 and 24 steps. For the first two options the MAPE was calculated using an expanding method.

MAPE for predictions using expanding method			
Sarima	24Q	4Q (Year)	1Q
log:no, Regular: (0, 1, 1), Seasonal: (0, 1, 1, 4)	8.26%	7.47%	6.05%
log:yes, Regular: (0, 1, 1), Seasonal: (0, 1, 1, 4)	6.50%	6.49%	5.22%
log:yes, Regular: (1, 1, 1), Seasonal: (2, 1, 0, 4)	9.92%	6.76%	5.12%

Note: The MAPE for the 4Q and 1Q is the average error of the predictions vs test

The model with the **best** MAPE is the sarima **(1, 1, 1) (2, 1, 0, 4)** with logarithm in a 1 step forecasting approach with a MAPE of 5.12%. However, if it is necessary to predict for larger steps (1 or more years) the best model is the sarima **(0, 1, 1) (0, 1, 1, 4)** with a log.

6. Prediction for one step ahead for Sarima (1,1,1)(2,1,0)