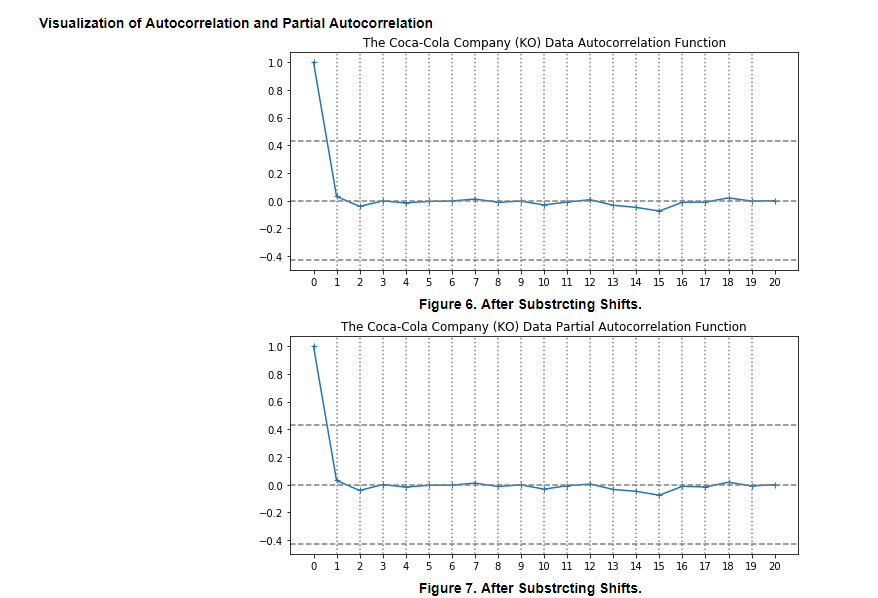
**Visualization of Autocorrelation and Partial Autocorrelation.**



Now that our time series has been stationarized by differencing, the next step to take in fitting an ARIMA model is to find out whether AR or MA terms are needed to correct any autocorrelation that remains in the differenced series.

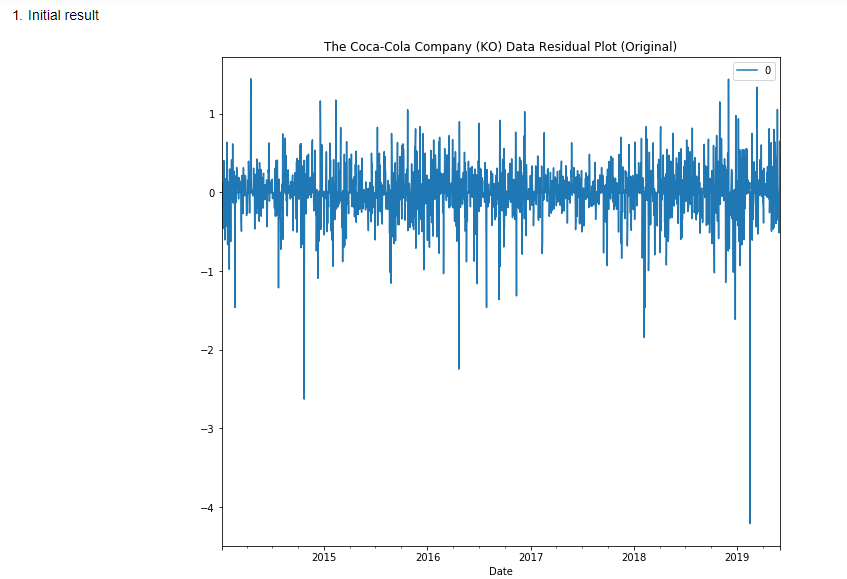
We look at the ACF & PACF graphs of the differenced series to see what number of AR (p)/MA(q) are needed.

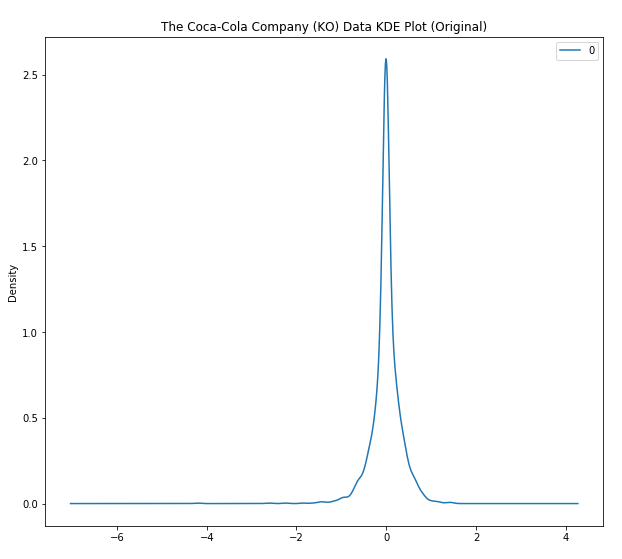
We know that ACF describes the autocorrelation between an observation and another observation at a prior time. [1] In the above plot for ACF function, the dotted lines represent the confidence band. The center line represents the mean and the lower and upper grey dotted lines represent the boundaries based off of 95% CI. The graph shows autocorrelation at the first lag is close to 0 or nearly 0. This obviously shows no integration of data, we have enough differencing and no further differencing is needed. We don’t need to apply a ‘d’ value to our model. It is a stationary model with no long term trend visible. From lag 1 onwards, there is hardly any linear decay visible. (how many differences did we take? Is there a way to know how much the std deviation changed?)

We will next look at the PACF plot to ascertain the most relevant lags.

The PACF describes the direct relationship between an observation and its lag. The first observation we make is the graph for ACF and PACF look nearly identical and that the PACF plot has a significant spike only at lag 1, this means that all the autocorrelations are effectively explained by the lag-1 autocorrelation.

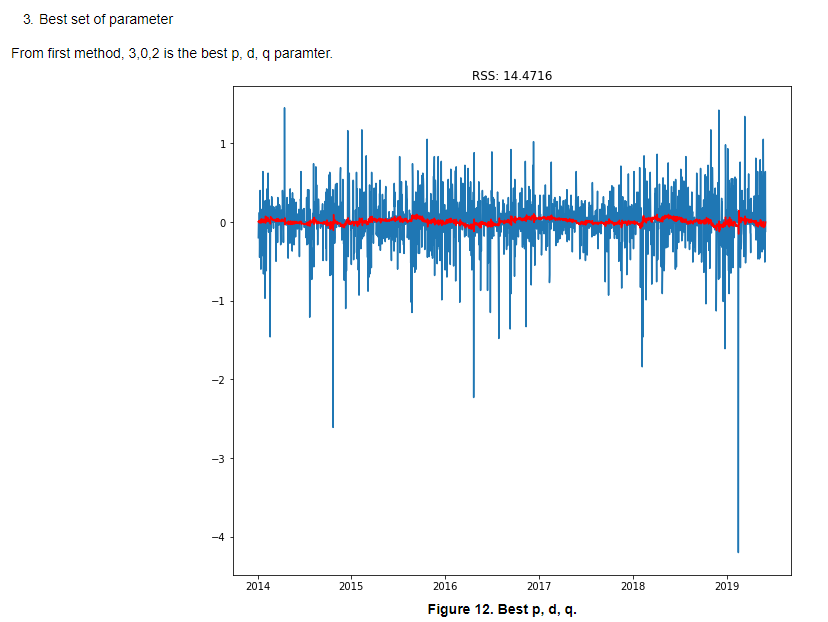
The PACF graph has a very large spike at lag 1 and no other significant spikes, indicating that in the absence of differencing an AR(1) model should be used. However, the AR(1) term in this model will turn out to be equivalent to a first difference, because the estimated AR(1) coefficient (which is the height of the PACF spike at lag 1) will be almost exactly equal to 1. [2] In other words, if we don't difference it, then we should fit an AR(1) model which will be equivalent to taking a first difference.





**Working on this**

The KDE curve echoes the same observation. It shows residuals of the model at 0 and it appears to be a bit skewed to the left



Using brute force, the best parameters that were uncovered were 3,0,2 for p,d,q with an RSS of 14.4716.

But this violates the 8th rule. Per the 8th rule, as stated previously “*Rule 8: It is possible for an AR term and an MA term to cancel each other's effects, so if a mixed AR-MA model seems to fit the data, also try a model with one fewer AR term and one fewer MA term--particularly if the parameter estimates in the original model require more than 10 iterations to converge. BEWARE OF USING MULTIPLE AR TERMS AND MULTIPLE MA TERMS IN THE SAME MODEL*.”

Keeping in mind the 8th rule violation, we select the next best parameters of 3,0,1 p,d,q with an RSS of 14.5279. This does not violate any of the rules.

ARIMA does not support seasonal data, and our data shows seasonality. Though we used ARIMA, and erased trending and seasonality from the series, we also employed SARIMA and used grid search to determine the best parameters that’ll produce the best fitting model for the series. From the second method used , the best parameters determined were: Best model ARIMA(0,1,1) x (1,0,0,12) and yelds the lowest AIC of 1194.5507507416933.

For evaluating the model, we shall use the AIC (Akaike Information Criterion) value, which is provided by ARIMA models fitted using statsmodels library. The Akaike information criterion (AIC) is an estimator of the relative quality of statistical models for a given set of data. Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models. Thus, AIC provides a means for model selection.[3] The AIC for the best parameters selected for the model using grid search was 1200.353. In contrast the AIC using SARIMA was 1194.55075. It achieves the same goodness of fit and is a smaller number than that for the ARIMA model

* **Error Metrics and Performance Evaluation**

AIC and BIC are both penalized-likelihood criteria. AIC helps quantify the goodness of fit and also the simplicity/parsimony of the model into a simple static. The generally acceptable principal is that when two models are being compared on the basis of the AIC, the model which minimizes AIC is considered better. AIC is an estimate of a constant plus the relative distance between the unknown true likelihood function of the data and the fitted likelihood function of the model, so a lower value for the AIC means a model is considered to be closer to the truth.[4]

The initial model’s AIC was 1200.353 and BIC was 1233.883

The final model selected had an AIC of 1194.5507

The final model, with a lower AIC than the initial model, may offer a better fit but that fit may not be worth the loss in parsimony imposed by the addition of two AR lags.

BIC is an estimate of a function of the posterior probability of a model being true, under a Bayesian setup, so that a lower BIC means that a model is considered to be more likely to be the true model.

The BIC for the initial model was 1233.883. Since we do not have BIC for the final model (is it possible to get it? ), we would be using AIC for determining the most likely model.

Conclusion: working on this

In conclusion, the final model was selected to predict the stock price of Coca Cola stock provided a useful model

**[1]** [**www.machinelearningmastery.com/gentle-introduction-autocorrelation-partial-autocorrelation**](http://www.machinelearningmastery.com/gentle-introduction-autocorrelation-partial-autocorrelation)

**[2]** <https://people.duke.edu/~rnau/411arim3.htm>

[3] <https://github.com/learn-co-students/dsc-3-26-08-sarima-models-lab-online-ds-ft-100118>

[4] [www.methodology.psu.edu/eresources/ask/sp07](http://www.methodology.psu.edu/eresources/ask/sp07) & [www.researchgate.net](http://www.researchgate.net)