ARIMA AND SARIMA TIME SERIES MODELS ON SEASONAL DATA

AND

VAR (Vector Auto regression) ON Multivariate Dataset

Appendix

Table of Contents:

Background Concepts behind PACF and ACF Plots:

PACF:

This plot helps in analysing the direct effects of the lagged values. Let us consider xt, xt-1, xt-2 time series. so while building an AR Model we need to know how previous lags are affecting the future values so that coefficients for the respective lags can be generated. we need to know how many previous lags to be considered for this case we need to find direct effect of lagged values on current value. PACF in above example can tell how xt-2 is affecting current time xt without the effect of xt-1. In this way By this direct effect values of previous lagged terms will be helpful in predicting p value for auto regression.

ACF:

This plot helps in analysing the indirect effect of previous lagged terms that means it gives us Pearson corelation of lagged values. This does not involve in removing the intermediate lags effect. So, this can be used for MA Model.

Various approaches in Handling and removing seasonality:

1, Differencing the time series values

2, Taking seasonal average if the seasons are very small.

3,Applying log on time series then taking seasonal average and subtracting it from the actual time series.

Detailed analysis on Augmented dickey fuller test:

This test is widely used for evaluating stationarity in time series. This test is based on unit root.

Unit Root:

$$Y_t = \alpha Y_{t-1} + \beta X_e + \epsilon$$

This means in above equation when alpha value is zero it indicated that it has unit root which means time series is stationary.

Dickey Fuller test:

This test depends on above Unit root this test evaluates whether there is presence of any unit root.

$$Y_t = \alpha Y_{t-1} + \beta X_e + \epsilon$$

In above equation details:

The null hypothesis is presence of root node thus the time series is stationary.

Augumented dickey fuller test:

$$y_t = c + \beta t + \alpha y_{t-1} + \phi_1 \Delta Y_{t-1} + \phi_2 \Delta Y_{t-2} .. + \phi_p \Delta Y_{t-p} + e_t$$

This test is similar to dickey fuller test but it has extra regressive terms in its equation it considers differenced terms coefficient in order to be more accurate.

Below is the equation for augmented dickey fuller test:

Null and alternate hypothesis are similar to Dickey fuller test.

Detailed Overview on granger causality tests:

In a multivariate time series this test helps to understand how other variables are affecting current time series variable. We cannot rely completely on this test for analysing cause and effect but these results are valid up to some extent.

Let us consider 2 time series equations below:

$$X_1(t) = \sum_{j=1}^{p} A_{11,j} X_1(t-j) + \sum_{j=1}^{p} A_{12,j} X_2(t-j) + E_1(t)$$

$$X_2(t) = \sum_{j=1}^{p} A_{21,j} X_1(t-j) + \sum_{j=1}^{p} A_{22,j} X_2(t-j) + E_2(t)$$

- Here p is number of lags, Matrix A contains Coefficients of the lagged values (like A1,2)
- E1,E2 are prediction errors
- If the variance of E1 and E2 is reduced by inclusion of the X2 or X1 in first or second equation then it is said x2 granger cause x1
- So in hypothesis for granger causality tests would be A12=0 (x2 does not granger cause x1)or A22=0(x1 does not granger cause x2)
- Above hypothesis can be evaluated by statistical tests

This test can be extended to any number of multivariate variables available in the dataset and also this can be considered for various lags by increasing number of lags and variables the equations become more complex. But extension is possible.

Choosing different p for VAR and VARMAX Concept;

In this project p value is chosen as 3 for VAR Based on BIC Values. P for Var model can be obtained using other values based other evaluation values other than BIC like AIC,FPE,HQIC.

```
sorted_order=model_1.select_order(maxlags=20)
print(sorted_order.summary())
```

```
 VAR Order Selection (* highlights the minimums)
==================================================
       AIC         BIC         FPE         HQIC
--------------------------------------------------
0        14.89       14.92   2.924e+06       14.90
1        14.31       14.41   1.644e+06       14.35
2        14.09       14.24   1.313e+06       14.15
3        13.98      14.19*   1.173e+06       14.06
4        13.95       14.23   1.147e+06       14.07
5        13.91       14.25   1.101e+06      14.05*
6        13.92       14.32   1.109e+06       14.08
7        13.91       14.38   1.101e+06       14.10
8       13.85*       14.38  1.039e+06*       14.07
9        13.87       14.46   1.054e+06       14.11
10       13.90       14.55   1.088e+06       14.16
11       13.94       14.65   1.129e+06       14.22
12       13.96       14.74   1.163e+06       14.28
13       13.96       14.80   1.157e+06       14.30
14       13.99       14.89   1.196e+06       14.36
15       13.98       14.94   1.178e+06       14.37
16       13.98       15.01   1.187e+06       14.40
17       14.00       15.09   1.212e+06       14.44
18       14.01       15.17   1.227e+06       14.48
19       14.01       15.22   1.222e+06       14.50
20       14.04       15.32   1.268e+06       14.56
--------------------------------------------------
```

If BIC is used p value will be 3,

If HQIC is used p value will be 5.

References:

https://www.machinelearningplus.com/time-series/augmented-dickey-fuller-test/

.