

**ISM 6136 – Datamining/Predictive Analytics**

**Jacob Perrone**

**Class Assignment 10**

**Dr. Bharti Sharma**

**5 points**

**TASK: Time series forecasting – Data Mining Task using XLMiner**

**Perform time series forecasting on the Average Income of Tax payers in US dataset determine the best model for any ‘two’ of the states. Provide screen shots with your explanation below.**

1. Perform Data Partitioning on the dataset (select Training set at least 60 % or higher)

A 70/30 and 80/20 splits were used for each model.

Graphical user interface, application, Excel

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

2. Perform Lag analysis and explain the ACF and PACF plots. Take Training lag of 15 and Validation lag < 14 as the max lag (lag analysis window).

Chart

Description automatically generated

Chart, application

Description automatically generated

Chart

Description automatically generated

Graphical user interface, application, Word

Description automatically generated

Both chosen states ACF show a downward trend and exhibit a similar pattern to one another. Both chosen states PACF show a repeat in the first two lags which indicates seasonality. Both states show the same trend and seasonality so similar models can be used for both.

3. Develop 2 ARIMA models for each of the three states by changing the partitioning ‘or’ p,d,q parameters ‘or’ iterations. So total you should have four models (2 models for each State, choose the best model for each State). Create a table to show all the models.

Table

Description automatically generated

The best models per state would be model 3 for NE, and model 4 for OK. These models were chosen based on criteria that P value is lowest and the MSE values are low.

4. Compare the models and determine the best model for each of the two states - based on the following – provide screen shots and your comments for the following model selection criteria:

a) Forecast and Actual plot for Training

Model 1-4 respectively for each section

Chart

Description automatically generated

Chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

b) Forecast and Actual plot for Validation (select any year and compare the results with respect to that in all the models)

Chart, line chart

Description automatically generated Model 1: year 1990 Act:20674 Forecast:20964

Chart, line chart

Description automatically generated Model 2: year 1990 Act:16214 Forecast:16259

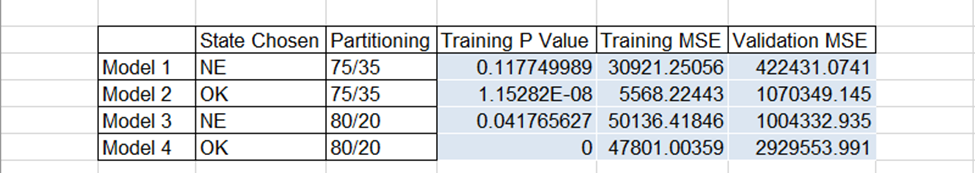
Chart, line chart

Description automatically generated Model 3: year 1990 Act:20674 Forecast:198387

Chart, line chart

Description automatically generated Model 4: year 1990 Act:16214 Forecast:17391

c) MSE training and validation error d) p-values of the coeff and AR1, AR2 etc



Table

Description automatically generated

Text, table

Description automatically generated

Text, table

Description automatically generated

Text, table

Description automatically generated