Individual Assignment 6

Ali Rafieepouralavialavijeh

4/23/2021

Problem statement and data sued

Dear manager, the Consumer Complaint Database is a collection of complaints about consumer financial products and services that are sent to companies for response. This database contains more than 2 million rows of data collected from 2011 to 2021. Each row corresponding to each single complaint, includes data about the request, type of the issue, date received and responded to, ID, etc. However, for this analysis the focus of the study would be on the complaints associated with the Bank of America which as a result consist 99,159 cases from 2011 to 2021. The research question is "**how many potential complaints would we get from consumers in 2022?**" In the figure below, it can be seen vertical lines showing the change of trend in the data. For the sake of this analysis and an appropriate forcast the highlighted data will be used in this analysis as it is near to contemporary time.

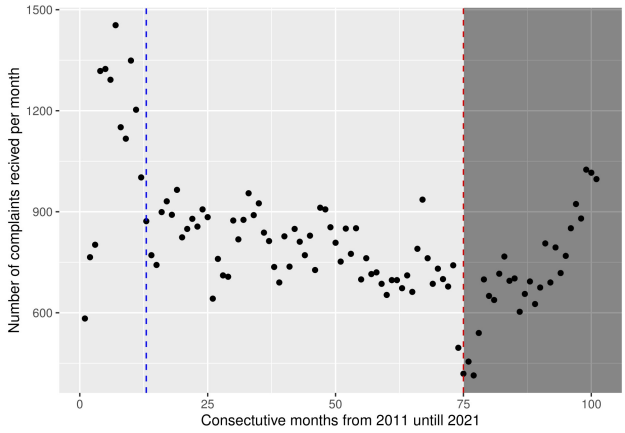

Figure 1: Distribution of the monthly complaints

Planning

Based on the specifics of this data set, which is a collected data through the time we should now investigate the data to see if we need to make some changes. Given the fact that the last data point is associated with March 2021 and that is collected in the middle of the month it seems to be too out of scale in comparison to other data points. As such, I would remove that to make a more consistent data set. Since visual inspection of the data does not show exaggerated effect of seasonality over time, I would use additive auto-correlation to study this time-series. The seasonal plot shows some variations in the fluctuations. In addition, the randomness seem to be constant over some time.

Analysis

Using ARIMA, I can now develop a forecasting model to predict the number of the complaints in 2022. ARIMA like any other forecasting model has a number of assumptions. 1) No unusual

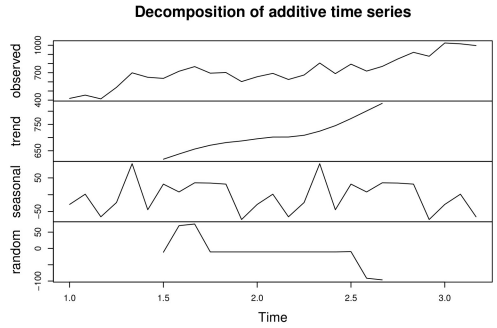

Figure 2: Decomposition plot of the time-series

observation is inspected in the range of data that is being studied 2) There is no evidence of changing variance, so we will not do a Box-Cox transformation. 3) The data are clearly non-stationary, as the series wanders up and down for long periods. Consequently, we will take a first difference of the data. 4) The PACF, and ACF shown in Figure shows an ARMA(1,1) 5) I will fit an ARIMA (1,1,1) 6) The ACF plot of the residuals from the ARIMA(3,1,1) model shows that all autocorrelations are within the threshold limits, indicating that the residuals are behaving like white noise.

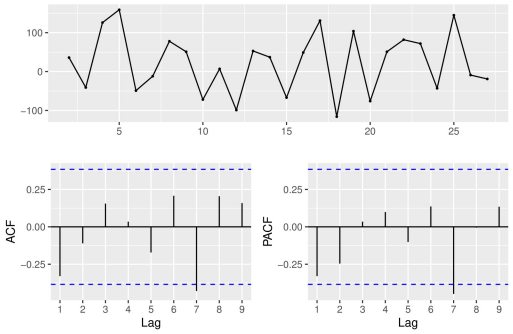

Figure 3: ACF and PACF plots after a 1 difference

Conclusion

Figure below shows an estimate of the predictions for the coming months to come. As can we noticed from the figure there are two confidence intervals for the estimates of the next points in the future, the highlighted one is an 80% confidence interval and the other one is a 95% confidence interval. Please see the results of the forecast for 2022 in the appendix. It is obvious that fractional numbers should be converted to integer numbers for the number of complaints.

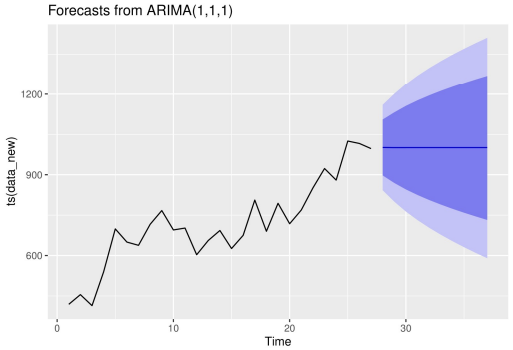

Figure 4: ARIMA(1,1,0) forecasting

Appendix

```
##    Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 28       1001.201 897.3396 1105.063 842.3586 1160.044
## 29       1000.673 869.4492 1131.898 799.9833 1201.363
## 30       1000.740 845.3758 1156.104 763.1311 1238.348
## 31       1000.731 824.6799 1176.783 731.4839 1269.979
## 32       1000.732 806.1609 1195.304 703.1610 1298.304
## 33       1000.732 789.2588 1212.206 677.3115 1324.153
## 34       1000.732 773.6109 1227.854 653.3801 1348.084
## 35       1000.732 758.9737 1242.491 630.9945 1370.470
## 36       1000.732 745.1735 1256.291 609.8889 1391.576
## 37       1000.732 732.0813 1269.383 589.8660 1411.598
## 38       1000.732 719.5981 1281.866 570.7746 1430.690
## 39       1000.732 707.6461 1293.818 552.4956 1448.969
## 40       1000.732 696.1628 1305.302 534.9334 1466.531
## 41       1000.732 685.0969 1316.368 518.0097 1483.455
## 42       1000.732 674.4061 1327.058 501.6595 1499.805
## 43       1000.732 664.0546 1337.410 485.8283 1515.636
## 44       1000.732 654.0121 1347.452 470.4695 1530.995
## 45       1000.732 644.2523 1357.212 455.5432 1545.921
## 46       1000.732 634.7527 1366.712 441.0148 1560.450
## 47       1000.732 625.4935 1375.971 426.8541 1574.610
## 48       1000.732 616.4573 1385.007 413.0345 1588.430
```