Predict_413_Discussion_Week_4

Singh, Gurjeet

February 2, 2018

Contents

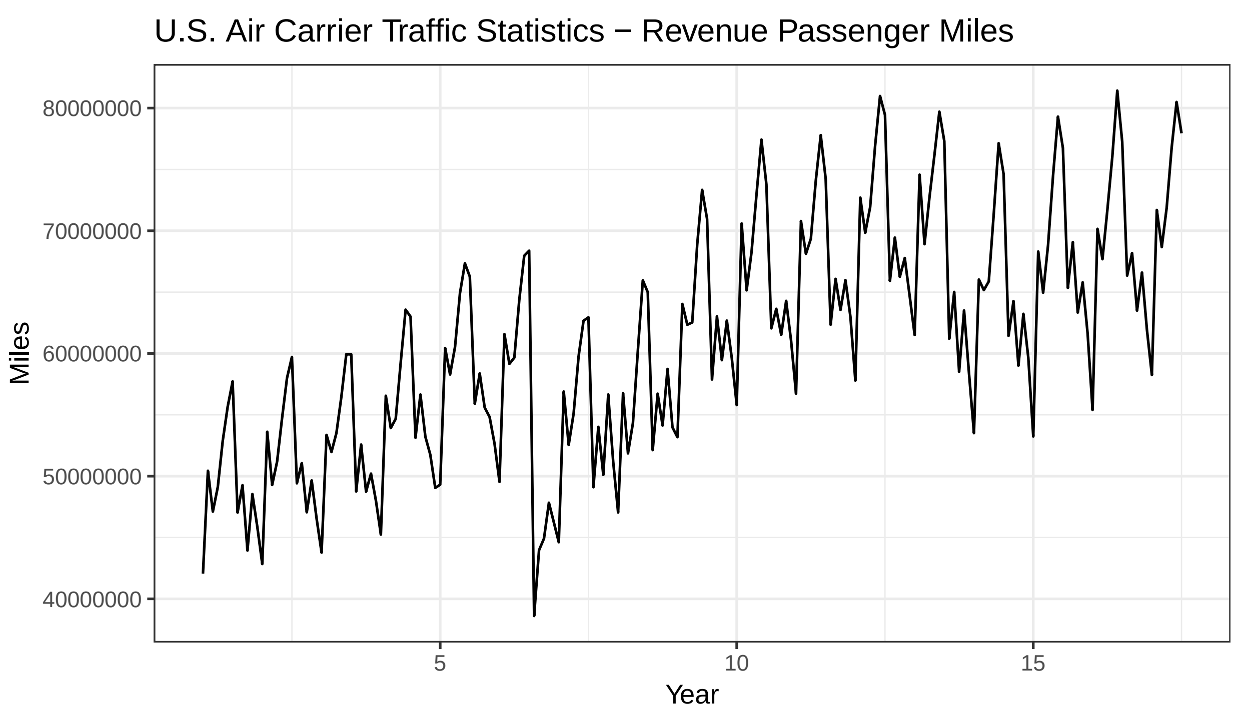
1. Introduction

For this week's discussion, we build an Auto Arima model as well as ETS model and test the performance of each model. We again run these two models on training and test dataset.

2. Data

For this week's discussion, I am using the same dataset (U.S Air Carrier Traffic Statistics – Revenue Passenger Miles) from the last week. The figure below shows that this data has strong seasonality within each year. Along with the seasonality, there is also an upward trend to it. We notice that every summer there's high traffic i.e. peak times. During the months of January and February, we see a dip in the traffic. This makes sense because a lot of people are back to their normal routine after the holidays. Therefore, not many people are travel in those months.

```
#Import the data
US.air.traffic  <- read.csv("us-air-carrier-traffic-statistic.csv")
colnames(US.air.traffic) <- c("YearMonth", "Revenue_Miles")
autoplot(ts(US.air.traffic[,"Revenue_Miles"],
              frequency = 12)) +
  ggtitle(paste0("U.S. Air ",
                   "Carrier Traffic Statistics ",
                   "- Revenue Passenger Miles")) +
  xlab("Year")+
  ylab("Miles")+
  theme_bw()
```



2.1 Split Data - Training/Test set

We split the data into the training set and test set. As per the requirement, we will hold out only 6 months of data for the test.

```
# #Split the data into training set and test set
data.ts<-ts(US.air.traffic[,"Revenue_Miles"], frequency=12)
train.ts<-ts(US.air.traffic[1:193,"Revenue_Miles"],end=c(2012,2),frequency=12)
test.ts<-ts(US.air.traffic[194:199,"Revenue_Miles"],start=c(2012,3),frequency=12)
```

3. Model

3.1 Auto Models - Arima & ETS Full Data

The table below shows the metrics of each model. We notice that Auto Arima model performed better and is the clear winner here. All the seven metrics are better for Auto Arima model.

```
#Auto Arima
traffic.arima <- auto.arima(data.ts)
ta.model.mets <- cbind("AIC" = traffic.arima$aic, "AICc"=traffic.arima$aicc,
                       "BIC" =traffic.arima$bic)
#ETS AUTO Model
traffic.ets <- ets(data.ts)
te.model.mets <- cbind("AIC" = traffic.ets$aic, "AICc"=traffic.ets$aicc,
                       "BIC" =traffic.ets$bic)
aa.model <- accuracy(traffic.arima)
est.model <- accuracy(traffic.ets)

rownames(ta.model.mets) <- "Auto Arima"
rownames(te.model.mets) <- "Auto ETS"
rownames(aa.model) <- "Auto Arima Model"
rownames(est.model) <- "Auto EST Model"

knitr::kable(rbind(ta.model.mets, te.model.mets),
             caption = "AIC, AICc, BIC for Arima and ETS Models")
```

Table 1: AIC, AICc, BIC for Arima and ETS Models

	AIC	AICc	BIC
Auto Arima	5968.417	5968.749	5984.573
Auto ETS	6833.503	6836.126	6882.902

```
knitr::kable(rbind(aa.model, est.model),
             caption = "Metrics for each model using full data")
```

Table 2: Metrics for each model using full data

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Auto Arima Model	17726.37	1895929	1139228	-0.0876403	1.997818	0.3670247	-0.0024269
Auto EST Model	123403.47	1881491	1256336	0.1346370	2.231370	0.4047536	0.0182516

3.2 Auto Model - Train & Test Data

We create the Auto Arima and ETS models using the training dataset. The first table shows the AIC, AICc, and BIC values of each model. Again, we notice that auto Arima performed better than auto ETS model.

The second table also shows that auto Arima outperformed the ETS model in the test set.

```r
#Auto Arima using training dataset
traffic.arima.train <- auto.arima(train.ts)
fcast.arima.train <- forecast(traffic.arima.train, h=6)
acc.arima <- accuracy(fcast.arima.train, test.ts)
rownames(acc.arima) <- c("Auto Arima - Train","Auto Arima - Test")
#Auto ETS using training dataset
traffic.ets.train <- ets(train.ts)
fcast.ets.train <- forecast(traffic.ets.train, h=6)
acc.ets <- accuracy(fcast.ets.train, test.ts)
rownames(acc.ets) <- c("Auto ETS - Train","Auto ETS - Test")
ta.train.mets <- cbind("AIC"=traffic.arima.train$aic, "AICc"=traffic.arima.train$aicc,
                       "BIC"=traffic.arima.train$bic)
te.train.mets <- cbind("AIC"=traffic.ets.train$aic, "AICc"=traffic.ets.train$aicc,
                       "BIC"=traffic.ets.train$bic)
rownames(ta.train.mets) <- "Auto Arima - Train"
rownames(te.train.mets) <- "Auto ETS - Train"
knitr::kable(rbind(ta.train.mets, te.train.mets),
             caption = "AIC, AICc, BIC for Arima and ETS Models")
```

Table 3: AIC, AICc, BIC for Arima and ETS Models

	AIC	AICc	BIC
Auto Arima - Train	5782.996	5783.339	5798.988
Auto ETS - Train	6627.111	6629.823	6676.052

```r
knitr::kable(rbind(acc.arima[,1:7], acc.ets[,1:7]),
             caption = "Metrics for each model")
```

Table 4: Metrics for each model

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Auto Arima - Train	16700.78	1923840	1161344.6	-0.0964374	2.043157	0.3653068	-0.0019718
Auto Arima - Test	-887716.19	1026463	887716.2	-1.1647129	1.164713	0.2792356	0.3591187
Auto ETS - Train	115236.10	1904285	1273352.1	0.1193085	2.269276	0.4005394	0.0192389
Auto ETS - Test	1912761.30	2031558	1912761.3	2.5449616	2.544962	0.6016688	0.3539079

4. Conclusion

In conclusion, we would select Auto Arima model because it performed better in full, training, and test dataset.

R Code available on my GitHub page (https://github.com/Singh-Gurjeet) in the Time Series and Forecasting repository.