'C:/Users/u353822/Documents/Git/bin/GitHub/University of Chicago/Time Series and Forecasting Analysis/HW 4/

Adham Suliman

October 21, 2018

Load datasets

Part 1 - Use ARIMA(p,d,q) model to forecast. Find the model returned by R auto.arima(). Change the values of p and q and determine the best model using AICc and BIC. Do AICc and BIC select the same model as the best model?

Arima model on 180 E, 1EXIT auto.arima produces a pdq of (4,0,3) with an AICc of 4302.88 and BIC of 4337.34.

```
df <- rbind(jun_16[3:26,c(3,5,88)],
      jun_17[3:26,c(3,5,88)],
      jun_18[3:26,c(3,5,88)],
      jun_19[3:26,c(3,5,88)],
      jun_20[3:26,c(3,5,88)],
      jun_21[3:26,c(3,5,88)],
      jun_22[3:26,c(3,5,88)],
      jun_23[3:26,c(3,5,88)],
      jun_25[3:26,c(3,5,88)],
      jun_25[3:26,c(3,5,88)],
      jun_26[3:26,c(3,5,88)],
      jun_27[3:26,c(3,5,88)],
      jun_28[3:26,c(3,5,88)],
      jun_29[3:26,c(3,5,88)],
      jun_30[3:26,c(3,5,88)])
df1<- df
df1$Date <- format(as.Date(df1$Date,formate='%Y-%m-%d'), "%d")
jul_1$Date <- format(as.Date(jul_1$Date,formate='%Y-%m-%d'), "%d")
df1[,2] <- as.numeric(df1[,2])
jul_1[,2] <- as.numeric(jul_1$Date)
auto.arima(df1[,2])
```

```
## Series: df1[, 2]
## ARIMA(4,0,3) with non-zero mean
##
## Coefficients:
##          ar1      ar2     ar3      ar4      ma1     ma2      ma3      mean
##       2.2929  -2.1935  1.2401  -0.4151  -0.8957  0.3590  -0.4042  752.8263
## s.e.  0.1655   0.3735  0.3145   0.1073   0.1617  0.1638   0.1014    4.0907
##
## sigma^2 estimated as 8703:  log likelihood=-2142.18
## AIC=4302.36   AICc=4302.88   BIC=4337.34
```

Attempt to change parameters of auto.arima to find a lower AICc or BIC. A pdq of (8,0,2) produces a lower BIC of 4291.55 and a lower AICc of 4291.255.

```
#pdq of (1,0,3) produces AIC of 4308.36
arima(df1[,2],order=c(3,0,3))
```

```
##
## Call:
## arima(x = df1[, 2], order = c(3, 0, 3))
##
## Coefficients:
##          ar1      ar2     ar3      ma1      ma2      ma3  intercept
##       1.8237  -0.9091  0.0082  -0.4179  -0.2510  -0.2797   752.8191
## s.e.  0.1495   0.2784  0.1420   0.1403   0.0945   0.0747     3.6432
##
## sigma^2 estimated as 8704:  log likelihood = -2146.18,  aic = 4308.36
```

```
#pdq of (3,0,3) produces AIC of 4303
arima(df1[,2],order=c(4,1,5))
```

```
##
## Call:
## arima(x = df1[, 2], order = c(4, 1, 5))
##
## Coefficients:
##          ar1     ar2      ar3      ar4      ma1      ma2      ma3     ma4
##       1.1162  0.2214  -0.3280  -0.1523  -0.6853  -0.6970  -0.1362  0.2380
## s.e.  0.3546  0.6589   0.4084   0.1385   0.3463   0.5164   0.2054  0.0911
##          ma5
##       0.2911
## s.e.  0.0861
##
## sigma^2 estimated as 8728:  log likelihood = -2141.82,  aic = 4303.64
```

```
#pdq of (2,1,3) produces AIC of 5175.78
(better_model <- arima(df1[,2],order=c(8,0,2)))
```

```
##
## Call:
## arima(x = df1[, 2], order = c(8, 0, 2))
##
## Coefficients:
##          ar1     ar2     ar3      ar4      ar5      ar6     ar7      ar8
##       3.0366  -3.628  1.8603  -0.0568  -0.2714  -0.2354  0.4742  -0.2291
## s.e.  0.0565   0.175  0.2573   0.2715   0.2715   0.2565  0.1731   0.0555
##           ma1     ma2  intercept
##       -1.8005  0.8901   749.1443
## s.e.   0.0284  0.0279     8.0691
##
## sigma^2 estimated as 7070:  log likelihood = -2110.31,  aic = 4244.62
```

```
#BIC
AICc(better_model)
```

```
## [1] 4245.52
```

```
BIC(better_model)
```

```
## [1] 4291.255
```

Part 2 - Use day of the week seasonal ARIMA(p,d,q)(P,Q,D)s model to forecast for July 1 (which is a Monday) - note use the hourly data

```
x <- ts(df1[,2],frequency=24)
(arima <- auto.arima(x))
```

```
## Series: x
## ARIMA(2,0,1)(1,1,0)[24]
##
## Coefficients:
##          ar1      ar2      ma1     sar1
##       1.5001  -0.6863  -0.4740  -0.2874
## s.e.  0.0821   0.0653   0.1045   0.0580
##
## sigma^2 estimated as 7096:  log likelihood=-1966.22
## AIC=3942.44   AICc=3942.62   BIC=3961.52
```

```
(frcst_2 <- forecast(arima,24))
```

```
##          Point Forecast       Lo 80     Hi 80       Lo 95     Hi 95
## 16.00000       173.3505  65.397763  281.3032    8.251036  338.4500
## 16.04167       141.4607 -13.210497  296.1319  -95.088477  378.0099
## 16.08333       109.3426 -70.661932  289.3472 -165.950589  384.6358
## 16.12500       175.9076 -14.511917  366.3270 -115.313899  467.1290
## 16.16667       202.7298   9.963904  395.4957  -92.080187  497.5398
## 16.20833       340.4182 147.637958  533.1985   45.586247  635.2502
## 16.25000       488.8239 295.291672  682.3561  192.841904  784.8059
## 16.29167       598.1490 402.717018  793.5811  299.261554  897.0365
## 16.33333       727.5025 529.918449  925.0865  425.323793 1029.6811
## 16.37500       891.3455 692.189854 1090.5012  586.763205 1195.9279
## 16.41667      1015.4828 815.559676 1215.4060  709.726745 1321.2389
## 16.45833      1034.4256 834.294701 1234.5566  728.351784 1340.4995
## 16.50000      1036.8632 836.725340 1237.0011  730.778753 1342.9477
## 16.54167      1087.6504 887.477246 1287.8235  781.511990 1393.7888
## 16.58333      1096.5723 896.284395 1296.8602  790.258381 1402.8862
## 16.62500      1090.7718 890.340095 1291.2036  784.237945 1397.3057
## 16.66667      1114.4909 913.946045 1315.0358  807.783999 1421.1979
## 16.70833      1006.9292 806.324637 1207.5338  700.130979 1313.7275
## 16.75000      1027.0548 826.431755 1227.6778  720.228338 1333.8812
## 16.79167       893.1103 692.485964 1093.7346  586.281862 1199.9387
## 16.83333       770.6940 570.068160  971.3199  463.863246 1077.5248
## 16.87500       742.0631 541.430279  942.6959  435.221675 1048.9045
## 16.91667       651.2803 450.637779  851.9228  344.424057  958.1365
## 16.95833       501.5082 300.857517  702.1588  194.639486  808.3768
```

Part 3 - Use hour of the day seasonal ARIMA (p,d,q)(P,D,Q)s model to forecast for the hours 8:00, 9:00, 17:00 and 18:00 on July 1

```
df3 <- subset(df1, Time == '08:00' | Time == '09:00' | Time == '17:00' | Time == '18:00')
x <- ts(df3[,2],frequency=4)
(arima3 <- auto.arima(x))
```

```
## Series: x
## ARIMA(0,0,1)(0,0,1)[4] with non-zero mean
##
## Coefficients:
##          ma1    sma1       mean
##       0.5189  0.6414  1034.4394
## s.e.  0.1141  0.1525    51.2352
##
## sigma^2 estimated as 28309:  log likelihood=-392.3
## AIC=792.59   AICc=793.32   BIC=800.97
```

```
(frcst_3 <- forecast(arima3,4))
```

```
##       Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 16 Q1       818.6735 603.0489 1034.298 488.9042 1148.443
## 16 Q2       888.2686 645.3439 1131.193 516.7473 1259.790
## 16 Q3      1064.0886 821.1639 1307.013 692.5673 1435.610
## 16 Q4      1014.4445 771.5198 1257.369 642.9233 1385.966
```

```
frcst_3$mean
```

```
##         Qtr1      Qtr2      Qtr3      Qtr4
## 16  818.6735  888.2686 1064.0886 1014.4445
```

Part 4 - For the July 1 8:00, 9:00, 17:00 and 18:00 forecasts, which model is better (part 2 or part 3)

As can be seen below, the error used for part 3 equates to 239,997, which is much smaller than the error created utilizing the forecast for part 2 which is 933,765.

```
df4 <- jul_1[3:26,c(3,5,88)]
df4 <- subset(df4, Time == '08:00' | Time == '09:00' | Time == '17:00' | Time == '18:00')
df_final <- as.data.frame(cbind("Actual"=df4[,2],"Part_3"=as.numeric(frcst_3$mean),"Part_2"=as.numeric(frcst_2$fit
ted[c(8,9,17,18)])))
df_final1 <- as.data.frame(cbind(c(1233, 1110,1142,1129),c(818.7,888.3,1064.1,1014.4),c(442.6,585.4,1144.9,945.1))
)
colnames(df_final1) <- c('Actual',"Part_3", "Part_2")
df_final1$Part_3_diff <- (df_final1$Actual - df_final1$Part_3)^2
df_final1$Part_2_diff <- (df_final1$Actual - df_final1$Part_2)^2
df_final
```

```
##   Actual           Part_3           Part_2
## 1   1233 818.673507336535 442.557001831829
## 2   1110 888.268643233831 585.414002613855
## 3   1142 1064.08860245461 1144.85400591986
## 4   1129 1014.44454480644 945.054004999169
```

```
apply(df_final1,2,sum)
```

```
##      Actual      Part_3      Part_2 Part_3_diff Part_2_diff
##      4614.0      3785.5      3118.0    239997.0    933764.9
```