HW 5

Adham Suliman

October 25, 2018

Import data

```
data(beersales)
#df <-
df <- melt(beersales, value.name = 'Months')
```

Part 1 - use ARIMA(p,d,q) model to forecast beer sales for all months of 1990.

1A - Use the h-period in forecast() to forecast each month of 1990.

```
train_df <- ts(df[1:180,1], frequency=12)
(arima_1990 <- auto.arima(train_df,seasonal=FALSE))
```

```
## Series: train_df
## ARIMA(1,1,3)
##
## Coefficients:
##          ar1     ma1     ma2     ma3
##      -0.3636  0.3530  0.3702  0.6659
## s.e.  0.1142  0.0856  0.0563  0.0626
##
## sigma^2 estimated as 1.111:  log likelihood=-262.29
## AIC=534.58   AICc=534.93   BIC=550.52
```

```
Period12 <- as.data.frame(forecast(arima_1990,h=12))
```

1B - Use the monthly data as a continuous time series. Forecast for 1990 Jan, Plug forecast into the time series to forecast for 1990 Feb. And so on and so forth. In other words, h=1 in all the forecasts.

```
Jan <- forecast(auto.arima(ts(df[1:180,1],frequency=1)),h=1)$mean
Feb <- forecast(auto.arima(ts(df[1:181,1],frequency=1)),h=1)$mean
Mar <- forecast(auto.arima(ts(df[1:182,1],frequency=1)),h=1)$mean
Apr <- forecast(auto.arima(ts(df[1:183,1],frequency=1)),h=1)$mean
May <- forecast(auto.arima(ts(df[1:184,1],frequency=1)),h=1)$mean
Jun <- forecast(auto.arima(ts(df[1:185,1],frequency=1)),h=1)$mean
Jul <- forecast(auto.arima(ts(df[1:186,1],frequency=1)),h=1)$mean
Aug <- forecast(auto.arima(ts(df[1:187,1],frequency=1)),h=1)$mean
Sep <- forecast(auto.arima(ts(df[1:188,1],frequency=1)),h=1)$mean
Oct <- forecast(auto.arima(ts(df[1:189,1],frequency=1)),h=1)$mean
Nov <- forecast(auto.arima(ts(df[1:190,1],frequency=1)),h=1)$mean
Dec <- forecast(auto.arima(ts(df[1:191,1],frequency=1)),h=1)$mean
(No_Period <- rbind(Jan,Feb,Mar,Apr,May,Jun,Jul,Aug,Sep,Oct,Nov,Dec))
```

```
##         [,1]
## Jan 11.97945
## Feb 14.01984
## Mar 14.19137
## Apr 16.48635
## May 15.10367
## Jun 16.72150
## Jul 16.44643
## Aug 17.49438
## Sep 16.71857
## Oct 15.21228
## Nov 14.55158
## Dec 14.05728
```

1C - which of the two above approaches yield the better results in terms of Mean Squared Error 1990? The method utilized in part B yielded a lowere SSE than the one utilzied in A.

```
final_df <- as.data.frame(cbind(df[181:192,1],Period12[,1], No_Period))
colnames(final_df) <- c('actual','Period12','No_Period')
final_df$season_diff <- (final_df$actual-final_df$Period12)^2
final_df$No_Period_diff <- (final_df$actual-final_df$No_Period)^2
apply(final_df,2,sum)
```

```
##        actual      Period12      No_Period    season_diff No_Period_diff
##      185.2354      140.3623      182.9827       191.5640        18.5666
```

Part 2 - use month of the year seasonal ARIMA(p,d,q)(P,Q,D)s model to forecast beer sales for all the months of 1990.

```
x <- ts(df[1:180,1],frequency = 12)
(Seasonal_arima <- auto.arima(x,D=1))
```

```
## Series: x
## ARIMA(4,1,2)(2,1,2)[12]
##
## Coefficients:
##          ar1      ar2     ar3      ar4      ma1     ma2     sar1     sar2
##       0.5103  -0.1662  0.1032  -0.3966  -1.1757  0.3125  0.6838  -0.592
## s.e.  0.1453   0.0986  0.0863   0.0789   0.1492  0.1421  0.1451   0.165
##          sma1    sma2
##       -1.1967  0.5849
## s.e.   0.1394  0.2087
##
## sigma^2 estimated as 0.2837:  log likelihood=-134.55
## AIC=291.1   AICc=292.81   BIC=325.4
```

```
(Seasonal_arima_forecast <- as.data.frame(forecast(Seasonal_arima,12)))
```

```
##        Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Jan 16       13.81601 13.13331 14.49871 12.77191 14.86011
## Feb 16       13.07707 12.35715 13.79698 11.97605 14.17808
## Mar 16       14.96181 14.23546 15.68817 13.85095 16.07268
## Apr 16       15.58503 14.83785 16.33220 14.44232 16.72774
## May 16       17.24847 16.49698 17.99996 16.09917 18.39777
## Jun 16       16.86360 16.10993 17.61727 15.71096 18.01624
## Jul 16       16.95571 16.19987 17.71156 15.79974 18.11168
## Aug 16       17.02231 16.26451 17.78012 15.86336 18.18127
## Sep 16       14.28619 13.51600 15.05638 13.10828 15.46410
## Oct 16       14.55136 13.75967 15.34304 13.34057 15.76214
## Nov 16       12.89695 12.09174 13.70216 11.66548 14.12841
## Dec 16       12.30127 11.48554 13.11699 11.05372 13.54881
```

Part 3 - Which model (Part 1 or Part 2) is better to forecast beer sales for each month of 1990 (Jan, Feb, …, Dec) ? As can be seen below, the method used in part 2 allows for an SSE of 6.78 while the method used in 1C allows for a SSE of 19.861969.

```
final_df1 <- final_df[,c(1,3,5)]
final_df1$Seasonal_arima_forecast <- Seasonal_arima_forecast[,1]
final_df1$seas_frcst_diff <- (final_df1$actual-final_df1$Seasonal_arima_forecast)^2
final_df1$No_Period_diff <- (final_df1$actual-final_df1$No_Period)^2
apply(final_df1,2,sum)
```

```
##                  actual               No_Period          No_Period_diff
##              185.235400              182.982697               18.566603
## Seasonal_arima_forecast         seas_frcst_diff
##              179.565771                6.780025
```