Time_Series_HW_5

Patrick Kelly

Thursday, May 21, 2015

Assignment #5 - ARIMA Beer Sales

Load data from TSA package (the package is written by textbook authors Jonathan Cryer and Kung-Sik Chan).

```
library("TSA")
library("forecast")
library("zoo")
library("hydroGOF")

data(beersales)
```

The data is the monthly beer sales in millions of barrels, 01/1975 - 12/1990.

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

```
#limit data to 1990
(beer.2 <- window(beersales, 1975, c(1989,12)))
```

```
##          Jan     Feb     Mar     Apr     May     Jun     Jul     Aug
## 1975 11.1179  9.8413 11.5732 13.0097 13.4182 14.4418 14.7534 13.8816
## 1976 10.8633 11.0000 10.9934 12.9140 13.5853 14.1553 15.0056 14.8590
## 1977 10.0067 10.4321 14.5477 14.2748 14.9986 15.7100 14.7980 14.6431
## 1978 10.6897 11.0093 14.7983 13.5984 14.9606 15.8187 15.2871 16.2773
## 1979 12.3244 12.0133 15.0094 14.9562 15.9268 15.5702 15.1282 15.5625
## 1980 12.5357 12.6446 14.0848 14.3271 16.1862 16.6604 17.0810 16.2811
## 1981 12.0798 12.4126 15.0092 15.4733 16.9966 17.2933 17.3701 16.2422
## 1982 11.9036 12.9126 15.6815 15.8119 16.5611 17.2255 16.1033 16.2590
## 1983 12.5696 12.6644 15.0723 15.5742 16.8397 17.0121 16.8476 17.3471
## 1984 12.4214 12.5443 15.3242 15.0629 16.8656 17.2300 17.3288 16.9654
## 1985 13.5114 12.7501 14.4642 15.8558 17.6043 16.1731 16.6319 16.0352
## 1986 13.9861 13.0120 14.6625 16.0165 17.1046 16.5952 17.0626 16.3092
## 1987 13.6094 13.7362 15.3119 15.9071 16.1350 16.6147 17.0362 15.8162
## 1988 13.8006 13.9416 15.2575 15.2452 16.4849 17.0435 16.4097 16.2246
## 1989 14.0913 13.1950 15.4059 14.8754 16.7768 16.9378 16.2259 17.4078
##          Sep     Oct     Nov     Dec
## 1975 12.5123 11.8983 10.6088 11.5874
## 1976 13.4387 12.2184 10.5208 10.8335
## 1977 12.8878 11.6235 11.4853 11.5065
## 1978 13.9370 13.3270 12.0353 11.5670
## 1979 13.7112 13.6425 12.5158 11.7629
## 1980 14.5118 14.1594 12.5120 12.3830
## 1981 14.6808 13.8444 12.3871 12.9072
## 1982 14.8834 13.8291 13.1376 12.2662
## 1983 14.8442 13.8536 12.7904 11.9797
## 1984 13.6582 14.2932 12.4037 11.3818
## 1985 13.5914 14.0102 12.3939 12.1101
```

```
## 1986 14.0156 14.6417 12.4761 12.8391
## 1987 14.3066 14.4671 12.5856 12.3225
## 1988 14.4386 13.9469 13.2062 12.2347
## 1989 14.7684 14.3167 13.4048 12.0999
```

```
#fit the ARIMA(p,d,q)
(fit.beer <- auto.arima(beer.2, seasonal = FALSE))
```

```
## Series: beer.2
## 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 1.086:  log likelihood=-262.29
## AIC=534.58   AICc=534.93   BIC=550.52
```

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

```
#forecast all months of 1990
(pred.1990 <- forecast(fit.beer,h=12))
```

```
##          Point Forecast     Lo 80    Hi 80     Lo 95    Hi 95
## Jan 1990       11.97945 10.643826 13.31507 9.9367909 14.02210
## Feb 1990       11.71412  9.835280 13.59296 8.8406832 14.58756
## Mar 1990       11.65137  9.034864 14.26788 7.6497691 15.65297
## Apr 1990       11.67419  8.035420 15.31295 6.1091739 17.23920
## May 1990       11.66589  7.376389 15.95539 5.1056639 18.22612
## Jun 1990       11.66891  6.770653 16.56716 4.1776739 19.16014
## Jul 1990       11.66781  6.243221 17.09240 3.3716165 19.96400
## Aug 1990       11.66821  5.759124 17.57729 2.6310435 20.70537
## Sep 1990       11.66806  5.312961 18.02317 1.9487729 21.38735
## Oct 1990       11.66812  4.895744 18.44049 1.3106660 22.02557
## Nov 1990       11.66810  4.502907 18.83329 0.7098843 22.62631
## Dec 1990       11.66810  4.130473 19.20574 0.1402922 23.19592
```

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.

```
#initialize vectors to loop through each month
pred.loop.1990  <- vector(mode="list", length=13)
beer.loop <- vector(mode="list", length=13)
fit.beer.loop <- vector(mode="list", length=13)

beer.loop[[1]] <- beer.2
fit.beer.loop[[1]] <- fit.beer

#loop through each month, creating the forecast then merging the value onto the ts
for (i in 2:13 ) {
```

```
  #forecast on previous beer table
  (pred.loop.1990[[i]] <- forecast(fit.beer.loop[[i-1]],h=1)$mean)
  #merge on new forecasted value
  beer.loop[[i]] <- as.ts(c(as.zoo(beer.loop[[i-1]]),as.zoo(pred.loop.1990[[i]])))
  #fit a new model with the new value included
  (fit.beer.loop[[i]] <- auto.arima(beer.loop[[i]], seasonal = FALSE))

}

#check the final product
beer.loop[[13]]
```

```
##           Jan      Feb      Mar      Apr      May      Jun      Jul
## 1975 11.11790  9.84130 11.57320 13.00970 13.41820 14.44180 14.75340
## 1976 10.86330 11.00000 10.99340 12.91400 13.58530 14.15530 15.00560
## 1977 10.00670 10.43210 14.54770 14.27480 14.99860 15.71000 14.79800
## 1978 10.68970 11.00930 14.79830 13.59840 14.96060 15.81870 15.28710
## 1979 12.32440 12.01330 15.00940 14.95620 15.92680 15.57020 15.12820
## 1980 12.53570 12.64460 14.08480 14.32710 16.18620 16.66040 17.08100
## 1981 12.07980 12.41260 15.00920 15.47330 16.99660 17.29330 17.37010
## 1982 11.90360 12.91260 15.68150 15.81190 16.56110 17.22550 16.10330
## 1983 12.56960 12.66440 15.07230 15.57420 16.83970 17.01210 16.84760
## 1984 12.42140 12.54430 15.32420 15.06290 16.86560 17.23000 17.32880
## 1985 13.51140 12.75010 14.46420 15.85580 17.60430 16.17310 16.63190
## 1986 13.98610 13.01200 14.66250 16.01650 17.10460 16.59520 17.06260
## 1987 13.60940 13.73620 15.31190 15.90710 16.13500 16.61470 17.03620
## 1988 13.80060 13.94160 15.25750 15.24520 16.48490 17.04350 16.40970
## 1989 14.09130 13.19500 15.40590 14.87540 16.77680 16.93780 16.22590
## 1990 11.97945 11.71413 11.65140 11.67432 11.66595 11.66896 11.66782
##           Aug      Sep      Oct      Nov      Dec
## 1975 13.88160 12.51230 11.89830 10.60880 11.58740
## 1976 14.85900 13.43870 12.21840 10.52080 10.83350
## 1977 14.64310 12.88780 11.62350 11.48530 11.50650
## 1978 16.27730 13.93700 13.32700 12.03530 11.56700
## 1979 15.56250 13.71120 13.64250 12.51580 11.76290
## 1980 16.28110 14.51180 14.15940 12.51200 12.38300
## 1981 16.24220 14.68080 13.84440 12.38710 12.90720
## 1982 16.25900 14.88340 13.82910 13.13760 12.26620
## 1983 17.34710 14.84420 13.85360 12.79040 11.97970
## 1984 16.96540 13.65820 14.29320 12.40370 11.38180
## 1985 16.03520 13.59140 14.01020 12.39390 12.11010
## 1986 16.30920 14.01560 14.64170 12.47610 12.83910
## 1987 15.81620 14.30660 14.46710 12.58560 12.32250
## 1988 16.22460 14.43860 13.94690 13.20620 12.23470
## 1989 17.40780 14.76840 14.31670 13.40480 12.09990
## 1990 11.66829 11.66814 11.66819 11.66813 11.66816
```

1C - which of the two above approaches yield the better results in terms of Mean Squared Error 1990?

```
#create ts of 1990 actuals
beer1990 <- window(beersales,1990)
```

```r
#calc MSE for first forecast
mse(pred.1990$mean,beer1990)
```

```
## [1] 15.96367
```

```r
#calc MSE for second forecast
mse(window(beer.loop[[13]],1990),beer1990)
```

```
## [1] 15.96326
```

Based on the MSE calculations above, the 2 forecasts product very similiar results but the first forecase (forecasting all 12 months at once) is marginally better.

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.

```r
#fit the ARIMA(p,d,q)(P,Q,D)s
(fit.s.beer <- auto.arima(beer.2, seasonal = TRUE))
```

```
## Series: beer.2
## 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 0.2667:  log likelihood=-134.55
## AIC=291.1   AICc=292.81   BIC=325.4
```

```r
#forecast all months of 1990 with seasonality
(pred.s.1990 <- forecast(fit.s.beer,h=12))
```

```
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Jan 1990       13.81601 13.15408 14.47794 12.80367 14.82835
## Feb 1990       13.07707 12.37905 13.77508 12.00955 14.14458
## Mar 1990       14.96181 14.25756 15.66607 13.88474 16.03888
## Apr 1990       15.58503 14.86058 16.30947 14.47709 16.69297
## May 1990       17.24847 16.51985 17.97709 16.13414 18.36280
## Jun 1990       16.86360 16.13286 17.59434 15.74602 17.98117
## Jul 1990       16.95571 16.22286 17.68856 15.83491 18.07651
## Aug 1990       17.02231 16.28757 17.75706 15.89862 18.14601
## Sep 1990       14.28619 13.53943 15.03295 13.14412 15.42826
## Oct 1990       14.55136 13.78375 15.31896 13.37741 15.72530
## Nov 1990       12.89695 12.11623 13.67766 11.70295 14.09094
## Dec 1990       12.30127 11.51036 13.09218 11.09168 13.51086
```

Part 3 Which model (Part 1 or Part 2) is better to forecast beer sales for each month of 1990 (Jan, Feb, . . . , Dec) ?

```r
#calc MSE for part 1 forecast
mse(pred.1990$mean,beer1990)
```

```
## [1] 15.96367
```

```r
#calc MSE for part 2 forecast
mse(pred.s.1990$mean,beer1990)
```

```
## [1] 0.5650021
```

Based on the MSE calculations above, the part 2 forecast (the seasonal ARIMA) is considerably better. Using the seasonal ARIMA drastically reduces the MSE to ~0.57 (compared to the ~15.96 MSE from part 1)