477/577 In-class Exercise 5 : Fitting Wine Sales

(due Fri 4/06/2017)

Name:

Use this file as a template for your report. Submit your code and comments together with (selected) output from R console.

- Your comments must be Arial font, and **BOLD FACED**.
- Your code must be Lucida Console font.

You must submit PRINTOUT of this file.

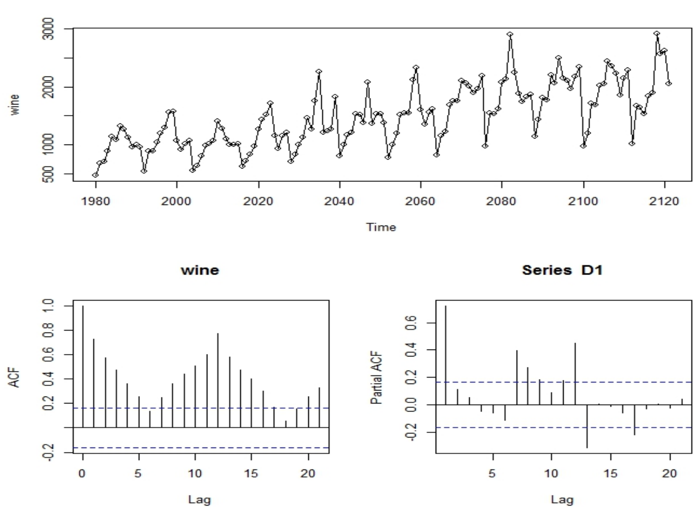
First, copy and paste below command in R console.

```
D0 <- read.csv("http://gozips.uakron.edu/~nmimoto/pages/datasets/wine.csv")
Wine <- ts(D0, start=1980)
```

Now your "D1" in R contains monthly Australian wine sales in 80's.

1. Plot D1, ACF and PACF of **Wine** data. Do you see seasonality? Is there an obvious trend to the data? What does ACF and PACF plots suggest about seasonality?

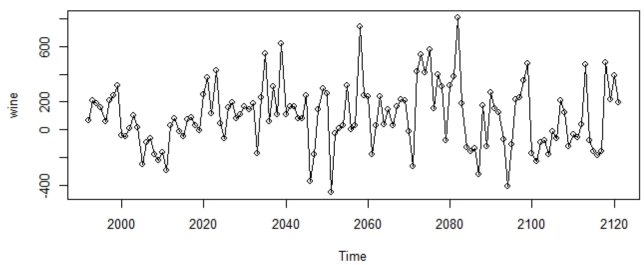
```
plot(D1, type="o")
acf(D1)
pacf(D1)
```

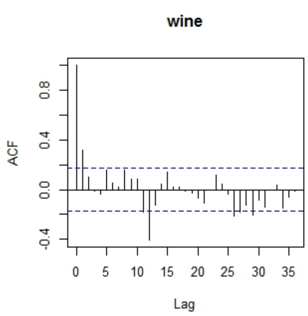


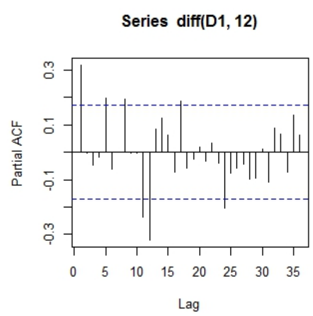
From the plot, there's obvious pattern of annual seasonality. ACF and PACF plots has large value at lag 12, also suggests seasonality. This suggests that seasonal differencing at lag 12 should be tried.

2. Take seasonal difference of **Wine** with lag 12, and plot the series, along with ACF and PACF of the series. Test for stationarity. State your conclusion about stationarity of seasonally differenced **Wine** data.

```
plot(diff(D1,12), type="o")
acf(diff(D1,12), lag.max=36)
pacf(diff(D1,12), lag.max=36)
source("http://gozips.uakron.edu/~nmimoto/477/TS_R-90.txt")
Stationarity.tests(diff(D1,12))
        KPSS    ADF    PP
p-val:  0.1 0.023 0.01
```







From the plot, linear trend and seasonality seems to be gone. All three stationarity tests indicate that diff(D1, 12) is stationary.

3. Given the ACF and PACF in (2), if ARMA model was fit to seasonally differenced **Wine** data, do you expect to see sAR term and/or sMA term?

   ACF still shows large correlation at lag 12, and PACF has large correlation at lag 12 and possibly at lag 24. This suggests sAR and/or sMA term can be present after the seasonal difference is taken.

4. Use **auto.arima(Wine, stepwise=FALSE)** to fit sARIMA model to the original (not differenced) Wine data. Did **auto.arima()** suggest seasonal model? Why or why not? Does it make sense to use this model?

```
  library(forecast)
source("http://gozips.uakron.edu/~nmimoto/477/TS_R-90.txt")
Fit1 <- auto.arima(D1)
Randomness.tests(Fit1$resid)

ARIMA(1,1,1)
         ar1      ma1
      0.4954  -0.9263
s.e.  0.0816   0.0263

BL15 BL20 BL25 ML15 ML20 JB       SD
[1,]    0    0    0    0    0   0 337.965
```

Auto.arima() is not picking up seasonal term, and Residual analysis shows model (1,1,1) is not fitting the data at all. This is because D1's frequency was set to 1.

5. Fix whatever necessary in the definition of **Wine**, so that **auto.arima()** considers seasonal model automatically. What is the suggested model now? How does the model look? Perform routine parameter significance check, and residual analysis. Make sure to use **stepwise=FALSE** option. Search for better choice of p,q,P,Q, around the values suggested by **auto.arima()**.

```
 D1 <- ts(D, start=c(1980,1), freq=12)
Fit5 <- auto.arima(D1, stepwise=FALSE)
Fit5
Randomness.tests(Fit5$resid)

ARIMA(1,0,1)(0,1,1)[12] with drift
         ar1      ma1     sma1   drift
      0.8146  -0.6230  -0.5854  8.6202
s.e.  0.2328   0.3294   0.0893  1.3150

sigma^2 estimated as 36413:  log likelihood=-867.66
AIC=1745.31   AICc=1745.79   BIC=1759.65

      BL15  BL20  BL25  ML15  ML20 JB       SD
[1,] 0.177 0.406 0.574 0.873 0.916  0 180.371
```

Now auto.arima() agrees to take seasonal difference, and suggests ARIMA(1,0,1)(0,1,1) with drift. MA1 parameter estimate is not significant and should be removed.

6. Are you happy with the value of d, and D, suggested in (5)? Why or why not? If unhappy, search for better value of d and/or D.

```
Fit6 <- Arima(D1, order=c(1,0,0), seasonal=c(0,1,1),
include.drift=TRUE)
Fit6
Randomness.tests(Fit6$resid)
```

```
Series: D1
ARIMA(1,0,0)(0,1,1)[12] with drift

Coefficients:
         ar1     sma1   drift
      0.2850  -0.5404  8.5504
s.e.  0.0844   0.0937  1.0175

sigma^2 estimated as 37075:  log likelihood=-868.9
AIC=1745.81   AICc=1746.13   BIC=1757.28

      BL15  BL20  BL25  ML15  ML20 JB      SD
[1,] 0.089 0.257 0.449 0.739 0.869  0 182.706
```

MA(1) term is removed from the previous model. This model seems to be fitting well. Including AR(2) term resulted in non-significant phi2 parameter, so (1,0,0)(0,1,1) seems to be the best model for (d=0, D=1) set. The fact #2 says that the series is stationary for (d=0,D=1) also makes the case strong for this model.

7. Now using the seasonally differenced **Wine** data with lag=12. Force fit MA(11), and sMA(1), and check for parameters being too close to unit root. What is the conclusion? What is the reason behind this check?

```
Fit7a <- Arima(D1, order=c(0,0,11), seasonal=c(0,1,0),
include.drift=FALSE)
Fit7a
Fit7b <- Arima(D1, order=c(0,0,0), seasonal=c(0,1,1),
include.drift=TRUE)
Fit7b
```

```
Series: D1
ARIMA(0,0,11)(0,1,0)[12]

Coefficients:
         ma1     ma2     ma3     ma4     ma5     ma6     ma7
ma8
      0.3968  0.4304  0.3041  0.2223  0.3078  0.1718  0.2102
0.5167
s.e.  0.1017  0.0920  0.1136  0.0939  0.0798  0.0975  0.0932
0.0862
         ma9    ma10    ma11
      0.4624  0.6602  0.4108
s.e.  0.1039  0.1023  0.1212
```

```
Series: D1
ARIMA(0,0,0)(0,1,1)[12] with drift

Coefficients:
         sma1   drift
      -0.5863  8.6404
s.e.   0.0963  0.7123
```

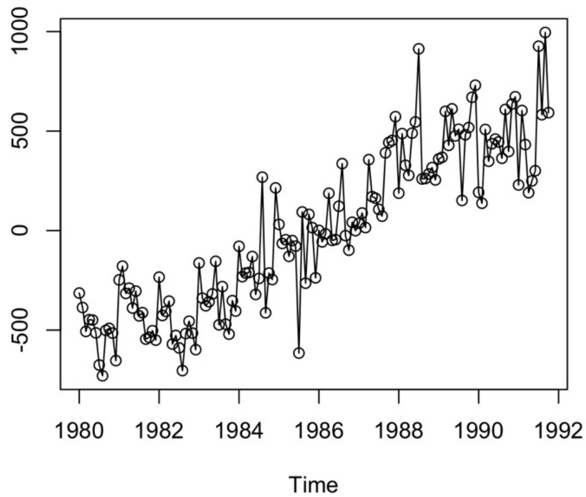
When MA(11) with drift was forced onto D=1 data, drift came up non-significant. Therefore, MA(11) is forced without drift above. No sign of all theta being 1.

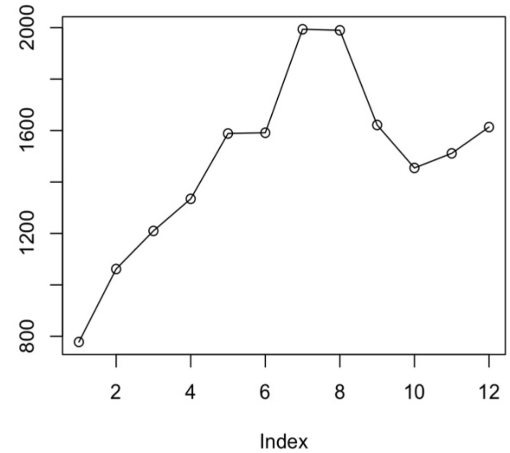
sMA(1) was forced to D=1 data with drift. Drift is significant, and should not be removed. sMA(1) parameter is not close to -1.

Neither test show no sign of problem with taking the seasonal difference (D=1). We have found no objecting evidence to the Model in #6.

8. Regardless of your conclusion in (7), subtract monthly average from the original **Wine** data. After the subtraction, fit linear trend with regression, then fit the residual with sARMA. Perform the routine parameter/residual analysis. Comment on the fit.

```
#--- Take Monthly Averages
Mav1  <- aggregate(c(D1), list(month=cycle(D1)), mean)$x
M.av1 <- ts(Mav1[cycle(D1)],
start=start(D1)freq=frequency(D1))
Ds.Wine <- D1-M.av1

plot(Mav1,   type="o")
plot(Ds.Wine,   type="o")
Fit8 <- auto.arima(Ds.Wine, xreg=time(D1), stepwise=FALSE)
Fit8
Randomness.tests(Fit8$residuals)


Series: Ds.Wine
Regression with ARIMA(0,0,2)(1,0,0)[12] errors

Coefficients:
         ma1     ma2    sar1  intercept      xreg
      0.2315  0.1975  0.2309  -209587.2  105.5398
s.e.  0.0833  0.0903  0.0898    14385.1    7.2437

sigma^2 estimated as 30533:  log likelihood=-932.52
AIC=1877.04   AICc=1877.66   BIC=1894.77

      BL15  BL20  BL25  ML15 ML20    JB      SD
[1,] 0.468 0.752 0.805 0.928 0.99 0.137 172.234
```





Plot on the right shows monthly average for 1 year. Plot on the left shows wine sales data after monthly average is subtracted.

When the series on the left is regressed with a line and fit with sARIMA, auto.arima shows (0,0,2)(1,0,0). Randomness.tests seems to indicate the fit is adequate.

9. Express your final model mathematically, using Y_t as your observations. Briefly, state the reason that you chose this model. Here's special character you may need (copy and paste to use) : $\bigtriangledown$, $\bigtriangledown_{12}$, ϕ_1, θ_1, Θ_1, Φ_1, X_t, e_t, σ^2. Match all the parameter estimates with their symbols.

The model in #6 (Seaonal difference) and model in #8 (Monthly average) are both good. We did not find any problem with either of the models. In this case, either model is a good choice. If you pick the final model to be #6: ARIMA(1,0,0) (0,1,1) [12] with drift,

Y_t = Obs.
$\bigtriangledown_{12}$ **$Y_t = c + X_t$**
$X_t = \phi X_{t-1} + e_t + \Theta e_{t-12}$

AR1 parameter $\phi = .2850$ (estimated)
drift c = 8.55 (estimated)
sMA1 parameter $\Theta = -.5404$ (estimated)
$E(e_t) = 0$, $Var(e_t) = 37075$ (estimated), e_t not normally distributed

```
ARIMA(1,0,0)(0,1,1)[12] with drift

         ar1     sma1   drift
      0.2850  -0.5404  8.5504
s.e.  0.0844   0.0937  1.0175

sigma^2 estimated as 37075
```