**477/577 In-class Exercise 3 : Fitting ARMA(p,q)**

(due Wed Feb 28)

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.

D <- read.csv("http://gozips.uakron.edu/~nmimoto/pages/datasets/Steel.csv")

D1 <- ts(D[,2], start=c(1956,1), freq=12)

D2 <- window(D1, start=c(1969, 12))

Mav1 <- aggregate(c(D2), list(month=cycle(D2)), mean)$x

M.av <- ts(Mav1[cycle(D2)], start=start(D2), freq=frequency(D2))

D3 <- D2-M.av

Now your “D1” in R contains monthly production of raw steel in Australia (in 1000 tones) from January 1956 to Nov 1993. “D2” is the same data cut from Dec. 1969 to Nov 1993. “Mav1” is the monthly average calculated from “D2”. “M.av” is the same as Mav1, except that it is as long as D2. “D3” is the series when the monthly average is subtracted from D2.

1. Plot the series D1, D2, Mav1, and D3. Comment on the stationary of D1, D2, and D3. Briefly explain what you see in the plot of “D2” regarding the stationarity.

(Your code)

**(Your comment and plots)**

1. Note that D1 was cut after Dec 1969 to produce D2, which was used to produce D3. Does it make sense to do this? If this procedure is to be justified, what is the assumption made?

**(Your comment)**

1. Note that Monthly average was calculated from D2, in order to create D3. Does it make sense to do this? If this procedure is to be justified, what is the assumption made? Is there a way to verify that assumption?

**(Your comment)**

1. Model “D3” with ARMA(p,q) model using auto.arima() function in forecast package. Be sure to use d=0 option inside. What is the best fitting model selected by using minimum AICc? Are all parameter estimates significantly different from zero?

layout(matrix(1:4, 2,2, byrow=TRUE))

plot(D1)

plot(D2)

plot(Mav1, type="o")

plot(D3)

**(Your comment)**

1. Force to fit ARMA(p, q) model to D3, adding 1 to each p and q you found in #4. E.g if you found ARMA(1,1) to be the best model, for to fit ARMA(2,2) to D3. Use function Arima(D3, order=c(p, 0, q) ) . Comment on the parameter significance.

library(forecast)

Fit1 <- auto.arima(D3, d=0)

Fit1

**(Your comment)**

1. Check adequacy of the model you obtained in #4. (i.e. perform residual analysis)

source("http://gozips.uakron.edu/~nmimoto/477/TS\_R-90.txt")

Randomness.tests(Fit1$resid)

**(Your comment)**

1. Using model obtained in #4, produce 5-step ahead prediction for D3. Then use that to predict December 1993 steel production with 95% prediction interval. (c.f. p60 Chapter 3 slides)

Y <- D3

Fit1 <- Arima(Y, order=c(1,0,2) ) #- Force to fit ARMA(1,2)

Y.h <- predict(Fit1, n.ahead=10)

Yhat <- Y.h$pred

Yhat.CIu <- Yhat+1.96\*Y.h$se

Yhat.CIl <- Yhat-1.96\*Y.h$se

ts.plot(cbind(Y, Yhat, Yhat.CIu, Yhat.CIl ), type="o", col=c("black","red","blue","blue"))

abline(h=0)

abline(h=mean(Y), col="blue")

**(Your comment)**

1. Perform Rolling 1-step prediction of the last 48 month within the dataset, with window size of 240. Report root mean squared error of the prediction performance. (c.f. p61 and 62 Chapter 3 slides)

**(Your comment)**

#--- Rolling 1-step predicton of last 48 month

Rolling.len = 48

Window.size = 240

p = 1

d = 0

q = 2

Y <- D3

Yhat <- Yhat.CIu <- Yhat.CIl <- Y2<- 0 #- Initialize

for (i in 1:Rolling.len) {

window.bgn <- i

window.end <- i+Window.size-1

Fit1 <- Arima(Y[window.bgn:window.end], order=c(p,d,q) ) #- Force to fit AR(p)

Y.h <- predict(Fit1, n.ahead=1)

Yhat[i] <- Y.h$pred

Yhat.CIu[i] <- Yhat[i]+1.96\*Y.h$se

Yhat.CIl[i] <- Yhat[i]-1.96\*Y.h$se

}

X <- window(Y, start=time(Y)[1], end=time(Y)[Window.size])

Y2 <- window(Y, start=time(Y)[Window.size+1], end=time(Y)[Window.size+Rolling.len])

Yhat <- ts(Yhat, start=c(floor(time(Y)[Window.size+1]), cycle(Y)[Window.size+1]), freq=frequency(Y) )

Yhat.CIu <- ts(Yhat.CIu, start=c(floor(time(Y)[Window.size+1]), cycle(Y)[Window.size+1]), freq=frequency(Y) )

Yhat.CIl <- ts(Yhat.CIl, start=c(floor(time(Y)[Window.size+1]), cycle(Y)[Window.size+1]), freq=frequency(Y) )

Pred.error <- Y2-Yhat

Pred.rMSE = sqrt( mean( (Pred.error)^2 ) ) #- prediction root Mean Squared Error

Pred.rMSE

mean(Pred.error)

layout(matrix(c(1,1,1,2,2,3), 2, 3, byrow=TRUE))

plot(Y, type="o", col="blue", main="Rolling 1-step prediction with window=100" ) #- Entire dataset

lines(X, type="o")

lines(Yhat, type="o", col="red")

lines(Yhat.CIu, type="l", col="red", lty=2)

lines(Yhat.CIl, type="l", col="red", lty=2)

plot(Pred.error, type="o", main="Prediction Error (Blue-Red)")

abline(h=c(-1.96, 1.96), col="blue", lty=2)

acf(Pred.error)