PSTAT 174/274 Spring 2021: Homework 5

1. Create a glossary of R-commands for time series. It should contain all commands that you learned so far in the labs, doing homework, and reviewing posted lecture slides. At the minimum, the glossary should include commands that allow

define working directory:
setwd(dir=)
or go to session, set working directory
read and plot data
plot()
simulate and plot ARMA models
ex. For ARMA(p,q): Xt- phi_1X(t-1) + phi_2X(t-2) = Zt + Theta_1Z(t-1)
x <- arima.sim(model=list(ar=c(phi_1, phi_2), ma=c(Theta_1)), n=n, sd = 1)
ts.plot(x, main="simulated data from ARMA(p,q), phi=(phi_1, phi_2), theta=Theta_1")
add trend and mean line to the original time series plot
trend:
abline(lm(x ~ as.numeric(1:length(x))))
mean:
abline(h=mean(x), col="red")
calculate and plot theoretical acf/pacf for ARMA models
ACF:
plot(0:100,ARMAacf(ar=c(phi_1, -0.7), ma=c(Theta_1), lag.max=100), xlim=c(1,40),ylab="r",type="h", main="ACF for ARMA(2,1) ar c=(phi_1, phi_2), ma c= Theta_1") ; abline(h=0)
PACF:
plot(ARMAacf(ar=c(phi_1, phi_2), ma=c(Theta_1)), lag.max=40, pacf=TRUE),type="h", xlab="lag", ylim=c(-.8,1));
abline(h=0)
calculate and plot sample acf/pacf
ACF:
plot(acf(x, lag.max=40), main="acf for ARMA(p,q)")
PACF:
plot(pacf(x, lag.max=40), main="pacf for ARMA(p,q)")
check whether a particular model is causal/invertible (R commands to find and plot roots of polynomials)
# code to check AR part of both models for invertibility and causality: source("plot.roots.R")
plot.roots(NULL,polyroot(c(phi_0, phi_1, phi_2)), main="roots of ar part")
plot.roots(NULL,polyroot(c(phi_0, phi_1, phi_2)), main="roots of ma part")
plot.roots(NULL,polyroot(c(phi_0, phi_1, phi_2, Theta_1, Theta_2)), main="roots of ar part")
OR
install.packages("UnitCircle")
library(UnitCircle)
uc.check(pol_ = c(phi_0, phi_1, phi_2), plot_output = TRUE)

perform Box-Cox transforms
ex. For using iowa data from tsdl

```
library(tsdl)
library(forecast)
iowa_ts <- tsdl[[1]]
#Box-Cox
library(MASS)
t <- 1:length(iowa_ts)
fit <- lm(iowa_ts ~ t)
bcTransform <- boxcox(iowa_ts ~ t,plotit = TRUE)
lambda <- bcTransform$x[which(bcTransform$y == max(bcTransform$y))]
lambda

iowa_bc <- (1/lambda)*(iowa_ts^lambda - 1)
ts.plot(iowa_bc,main = "Box-Cox tranformed data", ylab = expression(Y[t]))

#log transform
iowa_log <- log(iowa_ts)
# square root transform
iowa_sqrt <- sqrt(iowa_ts)
#Plot Box-Cox, log and square root transformed data
par(mfrow=c(2,2))
ts.plot(iowa_bc, main = "Box-Cox Transform")
ts.plot(iowa_log, main = "Log Transform")
ts.plot(iowa_sqrt, main = "Square Root Transform")
```

perform differencing data at lags 1 and 12
differencing data at lag 1

```
datadiff1 <- diff(data)
plot.ts(datadiff1)
abline(lm(datadiff1 ~ as.numeric(1:length(datadiff1))) )
```

differencing data at lag 12

```
AP=read.table(".txt")
plot.ts(AP)
APdiff12 <- diff(AP, lag=12, differences = 1)
plot.ts(APdiff12)
abline(lm(APdiff12 ~ as.numeric(1:length(APdiff12))) ) #regression line
```

perform Yule-Walker estimation and find standard deviations of the estimates.

```
x = ar(x, aic = TRUE, order.max = NULL, method = c("yule-walker"))
x$x.mean #mean estimate
sqrt(diag(x$asy.var.coef))  #st. errors
```

perform MLE and check AICC associated with the model

```
x_fit = arima(x, order = c(,,), method = "ML")
library(qpcR)
AICc(x_fit)
```

2. Choose a dataset that you will be interested to analyze for your class final project. URLs of time series libraries are posted on Gaucho Space. Provide the following information about the project:

(a) Data set description: briefly describe the data set you plan to use in your project.

The data I chose gives "Monthly car sales in Quebec 1960-1968"

(b) Motivation and objectives: briefly explain why this data set is interesting or important. Provide a clear description of the problem you plan to address using this dataset (for example to forecast).

I chose this data set to observe if there are any trends or seasonality within these years of the car sales and to predict future sales.

(c) Plot and examine the main features of the graph, checking in particular whether there is (i) a trend; (ii) a seasonal component, (iii) any apparent sharp changes in behavior. Explain in detail.

There seems to be a linear trend and a seasonal component, the acf also shows seasonality, while the pacf shows sharp changes at lags .3, .7, .8 and .9

(d) Use any necessary transformations to get stationary series. Give a detailed explanation to justify your choice of a particular procedure. If you have used transformation, justify why. If you have used differencing, what lag did you use? Why? Is your series stationary now?

I have to difference at lag(1) to remove the linear trend and difference at lag(12) to remove seasonality.

(e) Plot and analyze the ACF and PACF to preliminary identify your model(s): Plot ACF/PACF. What model(s) do they suggest? Explain your choice of p and q here.

This is most likely a SARIMA(p, d, q) × (P, D, Q)s model. I applied one seasonal differencing so D = 1 at lag s = 12. The ACF shows a strong peak at h = .9s so the MA part could be Q = 1. The PACF shows a strong peak at h = .9s so the AR part could be P = 1. The ACF and PACF graphs, suggest p = 0 and q = 0.

A possible model could be:

(1 − 0.140B^12)Yt = (1 − 0.766B^12)Zt

Or

(1 − 0.140B^12)(1 – B^12)(1 − B)Xt = (1 − 0.766B^12)Zt

3. An ARMA(3, 0) model is fit to the following quarterly time series:

Year	Quarter 1	Quarter 2	Quarter 3	Quarter 4
2018	3.53	1.33	1.85	0.61
2019	0.98	3.61	3.44	3.38
2020	2.91	2.12	4.62	2.93

The estimated coefficients are:

ar1	ar2	ar3	intercept
0.252	0.061	-0.202	2.637

Forecast the value for Quarter 1 of 2021. Give full explanation on how you arrived to your answer. Show calculations.

A. Less that 3.00 B. At least 3.00, but less than 3.25 C. At least 3.25, but less than 3.50
D. Atleast 3.50, but less than 3.75 E. At least 3.75.

Important: For models with AR part, the "intercept" reported in standard output of R is a misnomer. It is actually the mean of the process, so that the fitted model is

$X_t - 2.637 = 0.252(X_{t-1} - 2.637) + 0.061(X_{t-2} - 2.637) - 0.202(X_{t-3} - 2.637) + Z_t.$

$X_t - 2.637 = 0.252(X_{t-1} - 2.637) + 0.061(X_{t-2} - 2.637) - 0.202(X_{t-3} - 2.637) + Z_t$

$y_t = 2.637 + 0.252x_1 + 0.061x_2 - 0.202x_3$

Q1: 2018, 2019, 2020

$= 2.637 + 0.252(3.53) + 0.061(0.98) - 0.202(2.91)$

$= 2.637 + 0.88956 + 0.05978 - 0.58782$

$= 2.99852 > 3.00$

A

4. You are given the following information about an AR(1) model with mean 0: $\rho(2) = 0.215$, $\rho(3) = -0.100$, $X_T = -0.431$. Calculate the forecasted value of X_{T+1}.

AR(1) $\mu = 0$ $\rho(2) = 0.215$ $X_T = -0.431$

$\rho(3) = -0.100$ $X_{T+1} =$

$\hat{X}_{T+1} = \hat{\phi} X_T$

$\hat{X}_{T+1} = -0.4637(X_T)$

$= -0.4637(-0.431)$

$= .1998547$

$\hat{\phi}_2 = \rho(2) = 0.215$

$\hat{\phi} = -\sqrt{0.215}$

$= -0.4637$

$\hat{X}_{T+1} \approx 0.1999$

5. The five models, AR(1), ARMA(1, 1), ARMA(1, 2), ARMA(2, 3), and ARMA(4, 3) are fitted to the same time series. The models are ranked using Akaike Information Criterion (AIC): AIC $= -2\times$ log-likelihood$+2\ \times(p+q+2)$.

You are given the following information:

Model	Loglikelihood
AR(1)	-650
ARMA(1, 1)	-641
ARMA(1, 2)	-636
ARMA(2, 3)	-630
ARMA(4, 3)	-629

Determine the best model.

Model	Loglikelihood	AIC
AR(1)	-650	1306
ARMA(1, 1)	-641	1290
ARMA(1, 2)	-636	1282
ARMA(2, 3)	-630	1274 ← lowest
ARMA(4, 3)	-629	1276

AR(1) $AIC = -2(-650)+2(1+0+2)$
$p=1, q=0$
$= 1300+6$
$= 1306$

ARMA(1,1) $AIC = -2(-641)+2(1+1+2)$
$p=1, q=1$
$= 1282+8$
$= 1290$

ARMA(1,2) $AIC = -2(-636)+2(1+2+2)$
$p=1, q=2$
$= 1272+10$
$= 1282$

ARMA(2,3) $AIC = -2(-630)+2(2+3+2)$
$p=2, q=3$
$= 1260+14$
$= 1274$

ARMA(4,3) $AIC = -2(-629)+2(4+3+2)$
$p=4, q=3$
$= 1258+18$
$= 1276$

ARMA(2,3) is the best model