ENV 790.30 - Time Series Analysis for Energy Data | Spring 2023

Assignment 6 - Due date 03/06/23

Kexin Huang

Directions

You should open the .rmd file corresponding to this assignment on RStudio. The file is available on our class repository on Github. And to do so you will need to fork our repository and link it to your RStudio.

Once you have the file open on your local machine the first thing you will do is rename the file such that it includes your first and last name (e.g., "LuanaLima_TSA_A06_Sp23.Rmd"). Then change "Student Name" on line 4 with your name.

Then you will start working through the assignment by **creating code and output** that answer each question. Be sure to use this assignment document. Your report should contain the answer to each question and any plots/tables you obtained (when applicable).

When you have completed the assignment, **Knit** the text and code into a single PDF file. Submit this pdf using Sakai.

R packages needed for this assignment: "xlsx" or "readxl", "ggplot2", "forecast","tseries", and "Kendall". Install these packages, if you haven't done yet. Do not forget to load them before running your script, since they are NOT default packages.

Questions

This assignment has general questions about ARIMA Models.

Packages needed for this assignment: "forecast","tseries". Do not forget to load them before running your script, since they are NOT default packages.\

```
#Load/install required package here
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
library(tseries)
```

Q1

Describe the important characteristics of the sample autocorrelation function (ACF) plot and the partial sample autocorrelation function (PACF) plot for the following models:

- AR(2)

  Answer:

For the AR(2) model, the ACF will decay exponentially with time, and there will be a cutoff at lag=2 (after which the values will be within the significant range) in the PACF plot indicating the model order.

- MA(1)

  Answer:

For the MA(1) model, the PACF will decay exponentially with time, and there will be a cutoff at lag=1 in the ACF plot indicating the model order.

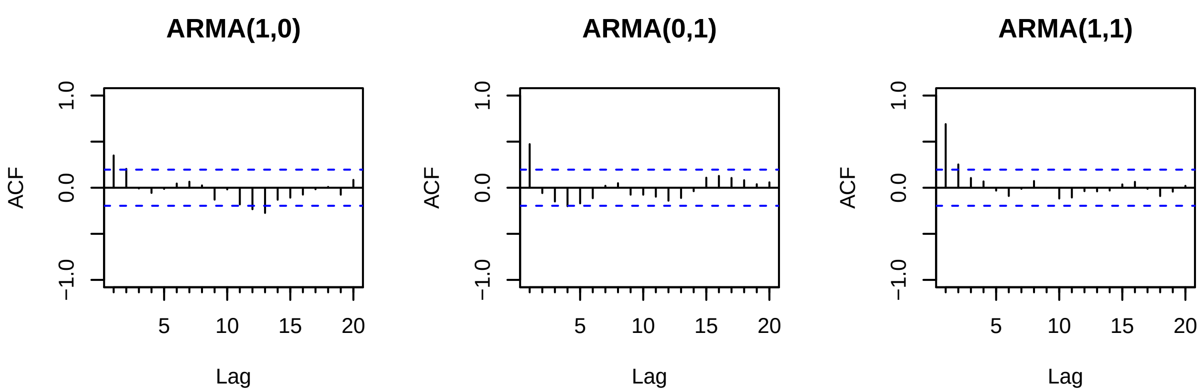
Q2

Recall that the non-seasonal ARIMA is described by three parameters ARIMA(p, d, q) where p is the order of the autoregressive component, d is the number of times the series need to be differenced to obtain stationarity and q is the order of the moving average component. If we don't need to difference the series, we don't need to specify the "I" part and we can use the short version, i.e., the ARMA(p, q). Consider three models: ARMA(1,0), ARMA(0,1) and ARMA(1,1) with parameters $\phi = 0.6$ and $\theta = 0.9$. The ϕ refers to the AR coefficient and the θ refers to the MA coefficient. Use R to generate $n = 100$ observations from each of these three models

```
set.seed(100)

#ARMA(1,0)
ARMAmodel_10<- arima.sim(model=list(ar=0.6), n=100)

#ARMA(0,1)
ARMAmodel_01<- arima.sim(model=list(ma=0.9), n=100)

#ARMA(0,1)
ARMAmodel_11<- arima.sim(model=list(ar=0.6, ma=0.9), n=100)
```

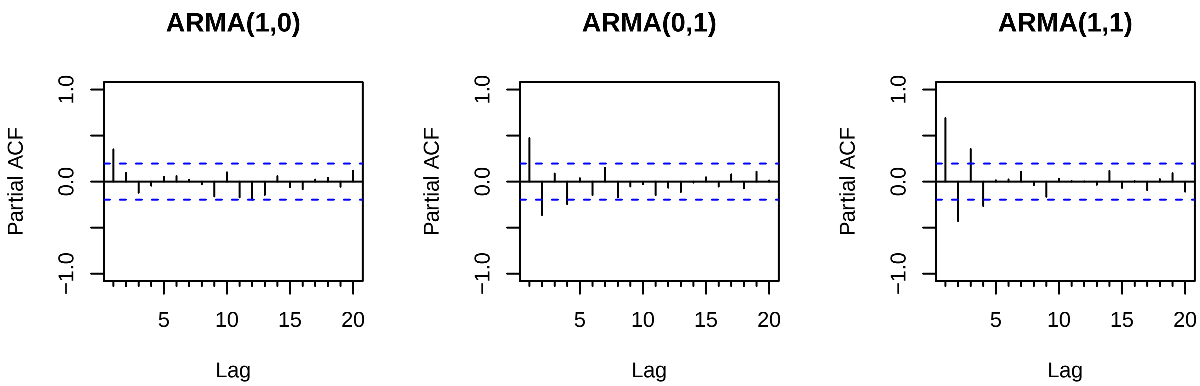
(a) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use command $par(mfrow = c(1, 3))$ that divides the plotting window in three columns).

```
par(mfrow=c(2,3))
Acf(ARMAmodel_10, main = "ARMA(1,0)", ylim = c(-1, 1))
Acf(ARMAmodel_01, main = "ARMA(0,1)", ylim = c(-1, 1))
Acf(ARMAmodel_11, main = "ARMA(1,1)", ylim = c(-1, 1))
```



(b) Plot the sample PACF for each of these models in one window to facilitate comparison.

```
par(mfrow=c(2,3))
Pacf(ARMAmodel_10, main = "ARMA(1,0)", ylim = c(-1, 1))
Pacf(ARMAmodel_01, main = "ARMA(0,1)", ylim = c(-1, 1))
Pacf(ARMAmodel_11, main = "ARMA(1,1)", ylim = c(-1, 1))
```



(c) Look at the ACFs and PACFs. Imagine you had these plots for a data set and you were asked to identify the model, i.e., is it AR, MA or ARMA and the order of each component. Would you be identify them correctly? Explain your answer.

Answer:

To identify the first model, I would first look at the ACF plot. The ACF plot decays exponentially, which is a strong indication of an AR process, so q should be 0. Then, looking at the PACF plot, the cut-off point occurs at lag=1, so p=1. So the first model should be ARMA(1,0).

As for the second model, in the ACF plot, the negative value at lag=1 indicates that it could be an MA process. However, from my perspective, the "exponential decay" pattern is not clear in the PACF plot, which makes it difficult for me to identify the model if I were given only this PACF plot. To identify the order, I'll look at the ACF plot and find the cutoff point, which occurs at 1, so q=1.

For the third model, I don't think I could identify it correctly because the cutoff in the PACF plot doesn't match the fact that p=1.

(d) Compare the ACF and PACF values R computed with the theoretical values you provided for the coefficients. Do they match? Explain your answer.

Answer:

```
# Fit ARMA(1,0) model to simulated data
fit_10 <- arima(ARMAmodel_10, order = c(1, 0, 0))
cat("phi_10 =", fit_10$coef["ar1"] )
```

```
## phi_10 = 0.3479548
```

```
# Fit ARMA(0,1) model to simulated data
fit_01 <- arima(ARMAmodel_01, order = c(0, 0, 1))
cat("theta_01 =", fit_01$coef["ma1"] )
```

```
## theta_01 = 0.9610674
```

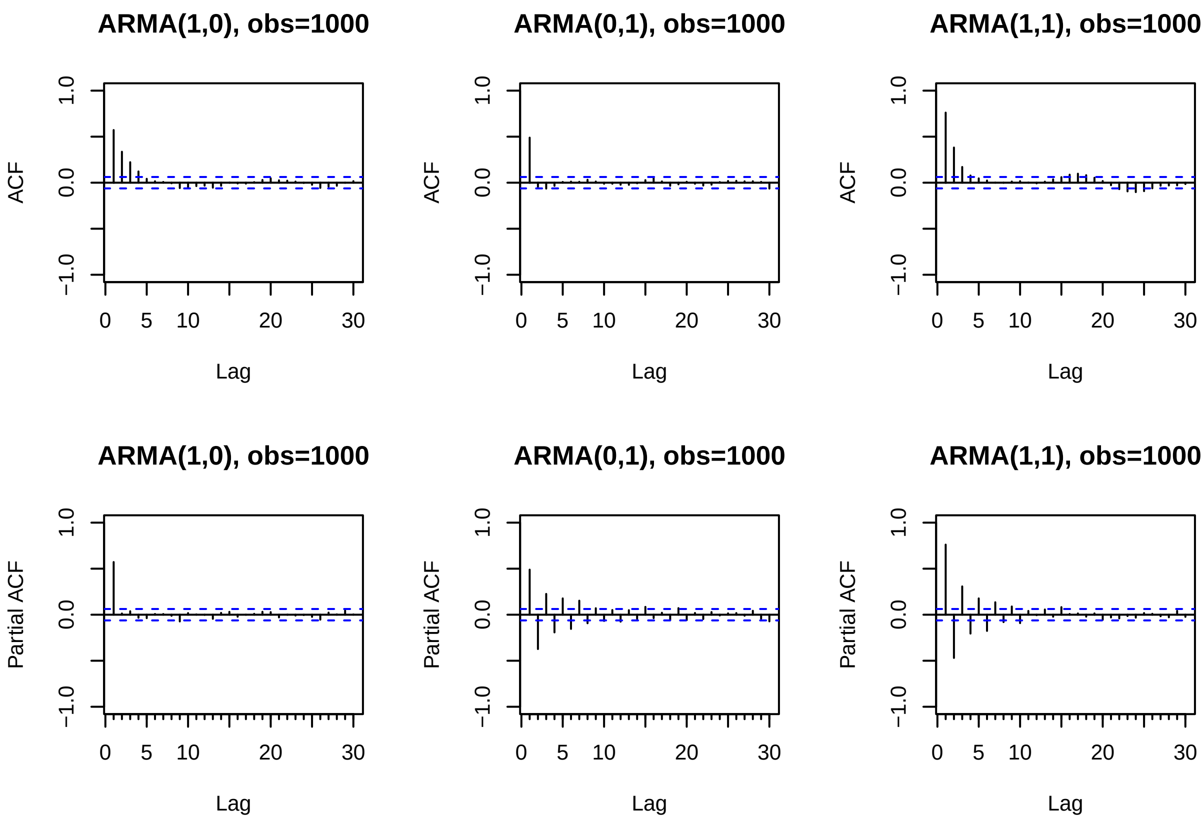
```
# Fit ARMA(1,1) model to simulated data
fit_11 <- arima(ARMAmodel_11, order = c(1, 0, 1))
cat("phi_11 =", fit_11$coef["ar1"],"\n","theta_11 =", fit_11$coef["ma1"])
```

```
## phi_11 = 0.5127021
##  theta_11 = 0.8392864
```

As the result shows, the calculated values are close to the theoretical ones, but not exactly the same (phi=0.6, theta=0.9).

(e) Increase number of observations to $n = 1000$ and repeat parts (a)-(d).

```r
set.seed(100)

#ARMA(1,0)
ARMAmodel_10_1000<- arima.sim(model=list(ar=0.6), n=1000) #the AR coefficient is 0.6
#ARMA(0,1)
ARMAmodel_01_1000<- arima.sim(model=list(ma=0.9), n=1000) #the MA coefficient is 0.9
#ARMA(0,1)
ARMAmodel_11_1000<- arima.sim(model=list(ar=0.6, ma=0.9), n=1000)

#ACF
par(mfrow=c(2,3))
Acf(ARMAmodel_10_1000, main = "ARMA(1,0), obs=1000", ylim = c(-1, 1))
Acf(ARMAmodel_01_1000, main = "ARMA(0,1), obs=1000", ylim = c(-1, 1))
Acf(ARMAmodel_11_1000, main = "ARMA(1,1), obs=1000", ylim = c(-1, 1))

#PACF
Pacf(ARMAmodel_10_1000, main = "ARMA(1,0), obs=1000", ylim = c(-1, 1))
Pacf(ARMAmodel_01_1000, main = "ARMA(0,1), obs=1000", ylim = c(-1, 1))
Pacf(ARMAmodel_11_1000, main = "ARMA(1,1), obs=1000", ylim = c(-1, 1))
```




```r
#check the values
fit_10_1000 <- arima(ARMAmodel_10_1000, order = c(1, 0, 0))
cat("phi_10 =", fit_10_1000$coef["ar1"] )
```

```
## phi_10 = 0.5717121
```

```r
fit_01_1000 <- arima(ARMAmodel_01_1000, order = c(0, 0, 1))
cat("theta_01 =", fit_01_1000$coef["ma1"] )
```

```
## theta_01 = 0.9200618
```

```r
# Fit ARMA(1,1) model to simulated data
fit_11_1000 <- arima(ARMAmodel_11_1000, order = c(1, 0, 1))
cat("phi_11 =", fit_11_1000$coef["ar1"],"\n","theta_11 =", fit_11$coef["ma1"])
```

```
## phi_11 = 0.5479517
##  theta_11 = 0.8392864
```

The parameters in all three models are closer to the "true value" I specified in the simulation (phi=0.6, theta=0.9) compared to the previous attempt. I think this is because increasing the sample size generally leads to a more accurate estimation of the parameters.

Q3

Consider the ARIMA model $y_t = 0.7 * y_{t-1} - 0.25 * y_{t-12} + a_t - 0.1 * a_{t-1}$

(a) Identify the model using the notation ARIMA$(p, d, q)(P, D, Q)_s$, i.e., identify the integers p, d, q, P, D, Q, s (if possible) from the equation.

p=1, d=1, q=1 P=1, D=0, q=0 (because there's no constant, so I assume the d= 1.) s=12

(b) Also from the equation what are the values of the parameters, i.e., model coefficients. AR term: phi_1=0.7 SAR term: phi_12=-0.25 MA term: theta_1=0.1

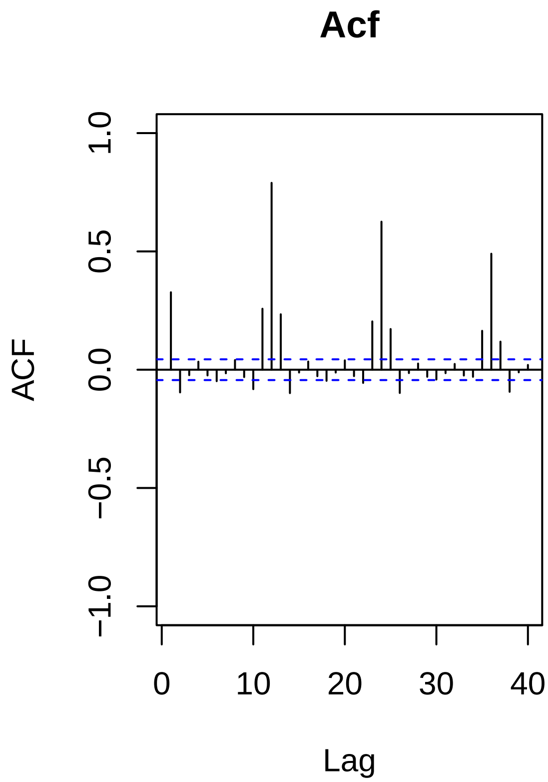
Q4

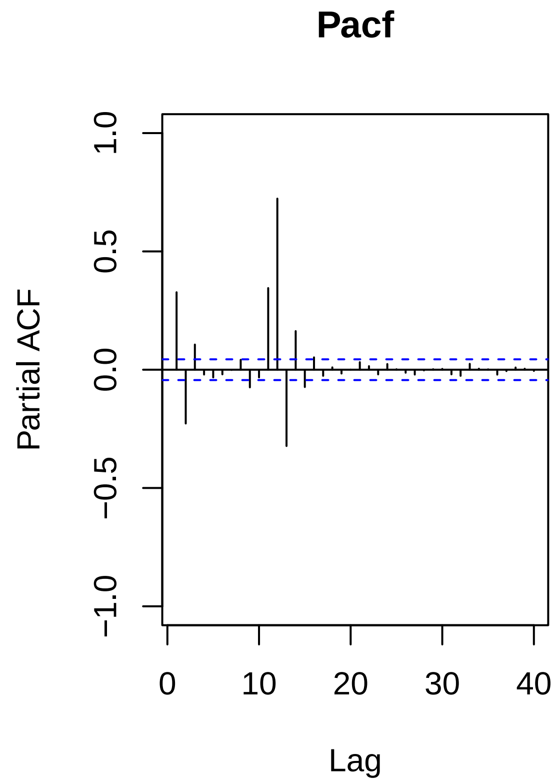
Plot the ACF and PACF of a seasonal ARIMA$(0, 1) \times (1, 0)_{12}$ model with $\phi = 0.8$ and $\theta = 0.5$ using R. The 12 after the bracket tells you that $s = 12$, i.e., the seasonal lag is 12, suggesting monthly data whose behavior is repeated every 12 months. You can generate as many observations as you like. Note the Integrated part was omitted. It means the series do not need differencing, therefore $d = D = 0$. Plot ACF and PACF for the simulated data. Comment if the plots are well representing the model you simulated, i.e., would you be able to identify the order of both non-seasonal and seasonal components from the plots? Explain.

```r
set.seed(100)
library(sarima)
```

```
## Loading required package: stats4

##
## Attaching package: 'sarima'

## The following object is masked from 'package:stats':
##
##     spectrum
```

```r
SARIMAmodel_10_01<- sim_sarima(model=list(ma=0.5,sar=0.8, nseasons=12), n=2000)

par(mfrow=c(1,2))
Acf(SARIMAmodel_10_01,main="Acf",ylim=c(-1,1),lag.max=40)
Pacf(SARIMAmodel_10_01,main="Pacf",ylim=c(-1,1),lag.max=40)
```





I'll first look at lag 1 through 11 to identify the p,q terms: since the value at lag=1 is negative in the ACF plot and there is no exponential decay, I think this series should contain an MA process. Then I'll look for the MA order (q) in the ACF plot. It cuts off at lag=1, then q=1.

For the seasonal component, there are several positive peaks in the ACF plot at lag=12, 24, 36 and only one peak at PACF. Therefore, it should contain a SAR process and P=1.