ENV 790.30 - Time Series Analysis for Energy Data | Spring 2021

Assignment 6 - Due date 03/16/22

Emre Yurtbay

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 project open the first thing you will do is change "Student Name" on line 3 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. Rename the pdf file such that it includes your first and last name (e.g., "LuanaLima_TSA_A06_Sp22.Rmd"). Submit this pdf using Sakai.

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.\

```r
#Load/install required package here
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(tseries)
```

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

```r
library(forecast)
```

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: In the AR(2) process, we should expect to see a exponentially decaying ACF plot and no significant spikes after 2 lags in the PACF plot

MA(1)

> Answer: In the MA(1) process, we should expect to see a exponentially decaying PACF plot and no significant spikes after 1 lag in the ACF plot

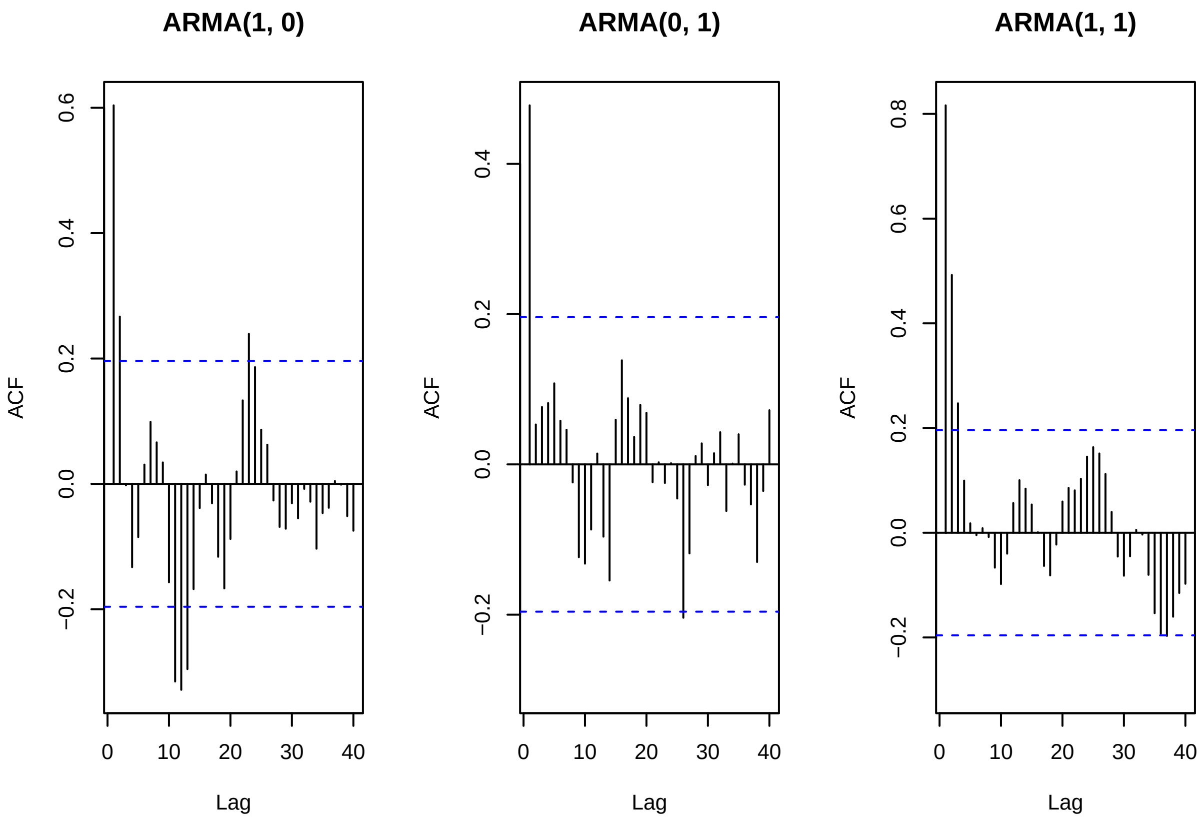
Q2

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

```r
set.seed(69)
arma10 = arima.sim(n = 100, list(ar = c(0.6)))
arma01 = arima.sim(n = 100, list(ma = c(0.9)))
arma11 = arima.sim(n = 100, list(ar = c(0.6), ma = c(0.9)))
```

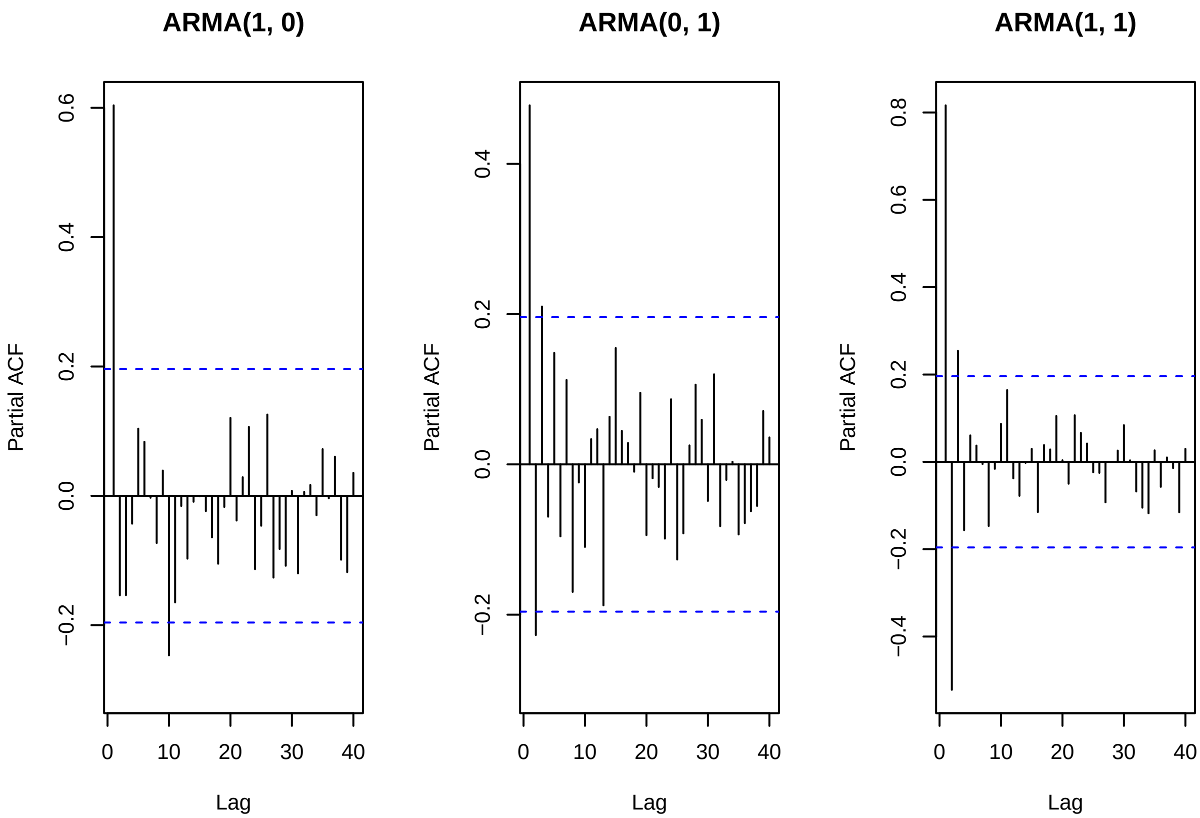
Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use command $par(mfrow = c(1, 3))$ **that divides the plotting window in three columns).**

```r
par(mfrow = c(1, 3))
arma10 %>%
  Acf(lag.max = 40,
      main = "ARMA(1, 0)")
arma01 %>%
  Acf(lag.max = 40,
      main = "ARMA(0, 1)")
arma11 %>%
  Acf(lag.max = 40,
      main = "ARMA(1, 1)")
```



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

```
par(mfrow = c(1, 3))
arma10 %>%
  Pacf(lag.max = 40,
      main = "ARMA(1, 0)")
arma01 %>%
  Pacf(lag.max = 40,
      main = "ARMA(0, 1)")
arma11 %>%
  Pacf(lag.max = 40,
      main = "ARMA(1, 1)")
```



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: For the ARMA(1, 0), we should expect to see an exponetially tapering ACF plot and no significant spikes after 1 lag in the PACF plot. We seem to see this behavior, but its not entirely clear. For the ARMA(0, 1), we should expect to see an exponetially tapering
> PACF plot and no significant spikes after 1 lag in the ACF plots. Again, this looks about like what is happening but I think we need more data to tell. Its a bit harder to tell with the PACF, but the ACF behavior looks correct. The ARMA(1,1) ACF and PACF seem to just tail off, but they also seem to resemble those of the ARMA(0, 1), so it would be hard to tell that the data generating process is an ARMA(1, 1).

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

```
arma10 %>%
  Pacf(lag.max = 5,plot = F)
```

```
##
## Partial autocorrelations of series '.', by lag
##
##      1      2      3      4      5
##  0.604 -0.154 -0.154 -0.043  0.104
```

```r
arma01 %>%
  Pacf(lag.max = 5,plot = F)
```

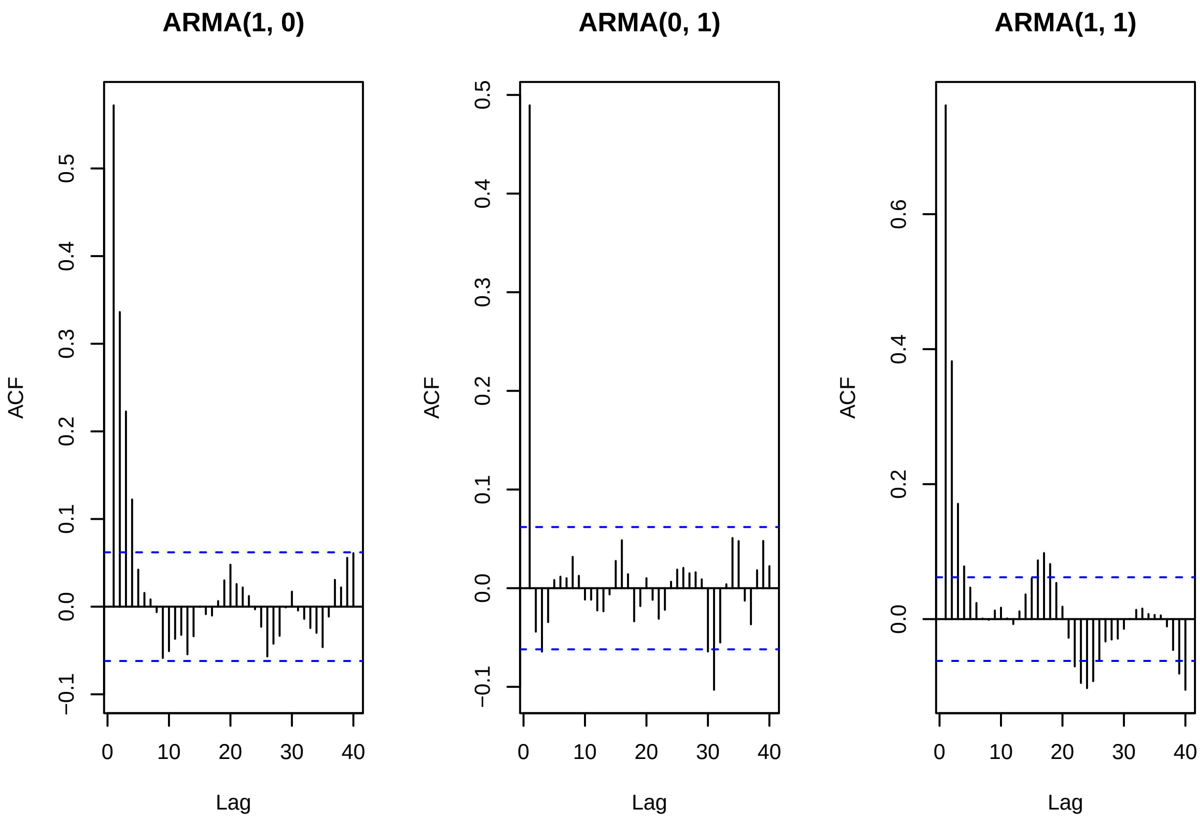
```
##
## Partial autocorrelations of series '.', by lag
##
##      1      2      3      4      5
##  0.478 -0.227  0.210 -0.070  0.148
```

Answer: For the arma(1, 0), ϕ should be the lag 1 partial autocorrelation . The theoretical value is 0.6, and the computed value is 0.604, so they are actually pretty close. For the ARMA(0, 1) Using the formula $\phi = \theta/(1 + \theta^2)$, we get a value for $\theta = 0.74$, which is a lower than the truth.

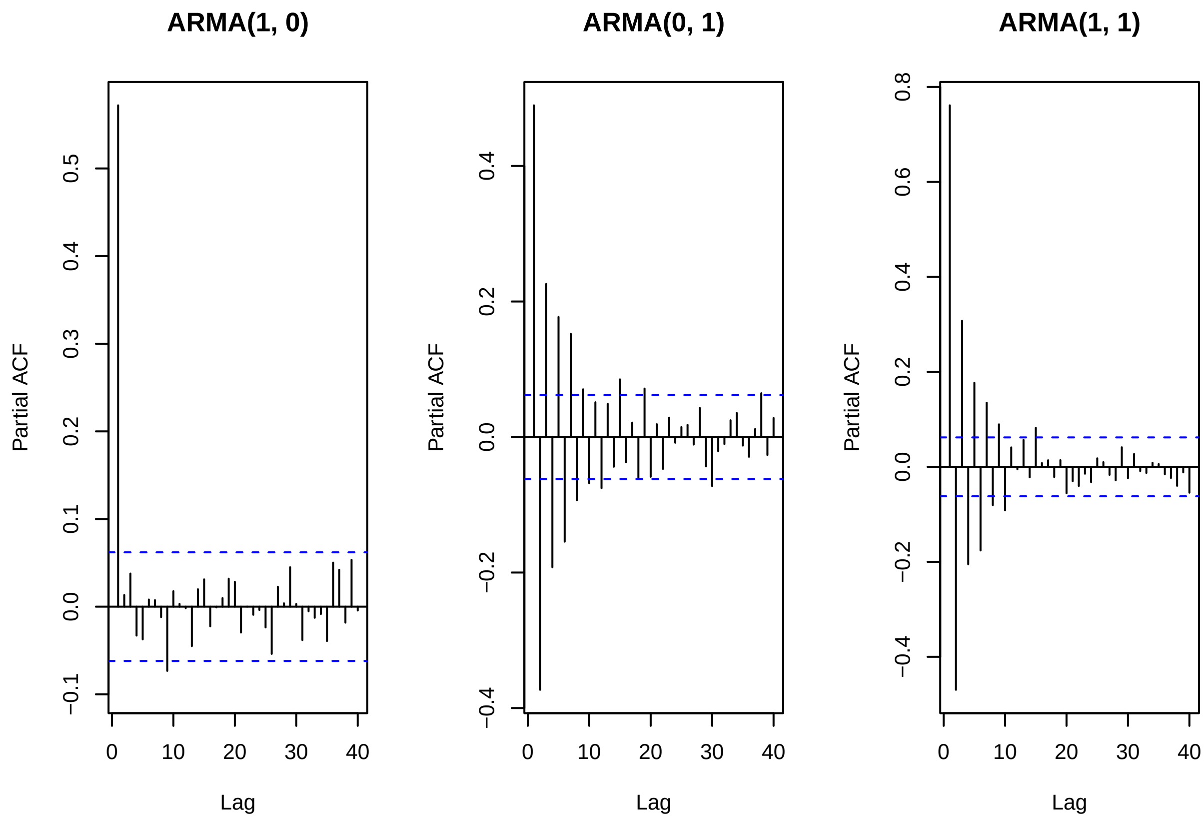
Increase number of observations to $n = 1000$ and repeat parts (a)-(d).

```r
set.seed(100)
arma10_2 = arima.sim(n = 1000, list(ar = c(0.6)))
arma01_2 = arima.sim(n = 1000, list(ma = c(0.9)))
arma11_2 = arima.sim(n = 1000, list(ar = c(0.6), ma = c(0.9)))
```

```r
par(mfrow = c(1, 3))
arma10_2 %>%
  Acf(lag.max = 40,
      main = "ARMA(1, 0)")
arma01_2 %>%
  Acf(lag.max = 40,
      main = "ARMA(0, 1)")
arma11_2 %>%
  Acf(lag.max = 40,
      main = "ARMA(1, 1)")
```




```
par(mfrow = c(1, 3))
arma10_2 %>%
  Pacf(lag.max = 40,
      main = "ARMA(1, 0)")
arma01_2 %>%
  Pacf(lag.max = 40,
      main = "ARMA(0, 1)")
arma11_2 %>%
  Pacf(lag.max = 40,
      main = "ARMA(1, 1)")
```




```r
arma10_2 %>%
  Pacf(lag.max = 5,plot = F)
```

```
##
## Partial autocorrelations of series '.', by lag
##
##      1      2      3      4      5
##  0.572  0.013  0.038 -0.033 -0.037
```

```r
arma01_2 %>%
  Pacf(lag.max = 5,plot = F)
```

```
##
## Partial autocorrelations of series '.', by lag
##
##      1      2      3      4      5
##  0.489 -0.373  0.226 -0.192  0.177
```

Answer: For the ARMA(1, 0), we should expect to see an exponetially tapering ACF plot and no significant spikes after 1 lag in the PACF plot. We seem to see this behavior better here. For the ARMA(0, 1), we should expect to see an exponetially tapering
PACF plot and no significant spikes after 1 lag in the ACF plots. This looks good as well. The ARMA(1,1) ACF and PACF seem to just tail off, governed by the AR and MA orders and coefficients, but its always difficult to tell an ARMA(1, 1) based off the plots. ϕ should be the lag 1 partial autocorrelation . The theoretical value is 0.6, and the computed value is 0.57, which again is pretty close. Using the formula $\phi = \theta/(1 + \theta^2)$, we get a value for $\theta = 0.8$, which is a closer to the truth than before.

Q3

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

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.

$$ARIMA(1, 0, 1)(1, 0, 0)_{12}$$

Also from the equation what are the values of the parameters, i.e., model coefficients.

$$\phi_1 = 0.7, \theta_1 = 0.1, \phi_{12} = -0.25$$

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.

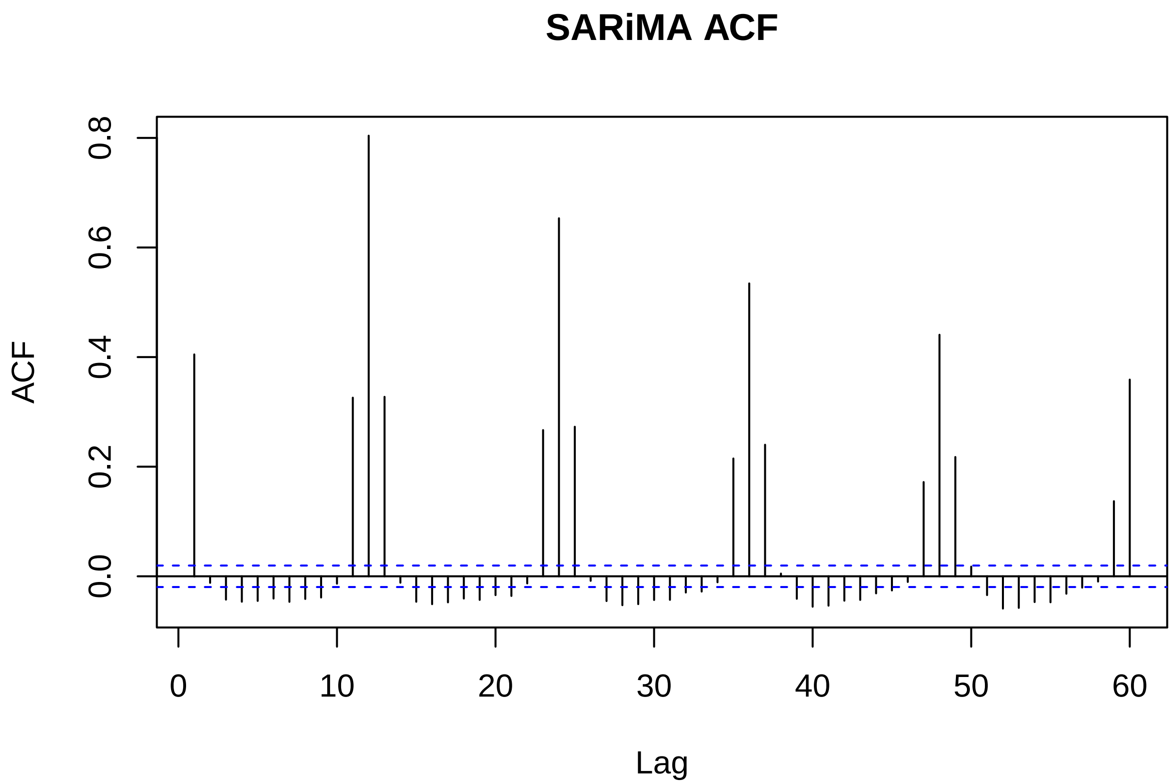
$$y_t = a_t - 0.5a_{t-1} + 0.8y_{t-12}$$

```
set.seed(999)
library(sarima)
```

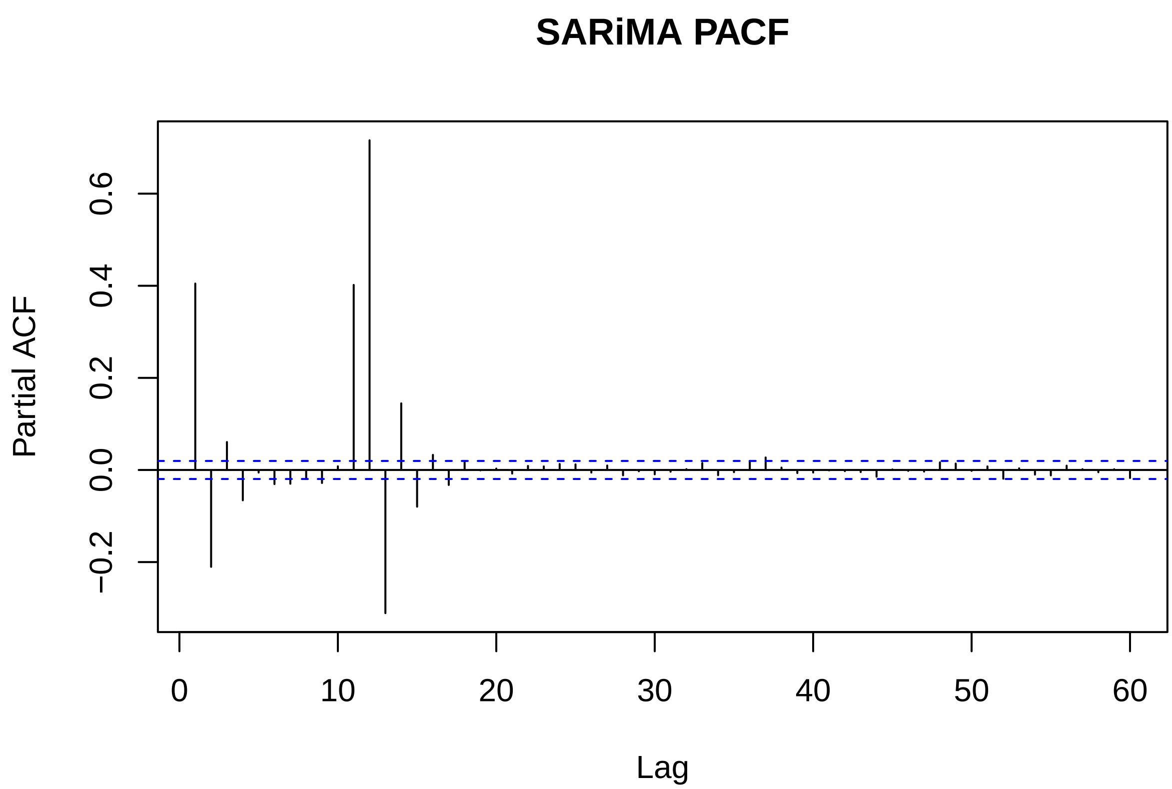
```
## Loading required package: stats4

##
## Attaching package: 'sarima'

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

```
x = sim_sarima(n = 10000, model= list(ma = 0.5, sar = 0.8, nseasons=12))
```

```
x %>%
  Acf(lag.max = 60,
      main = "SARiMA ACF")
```




```r
x %>%
  Pacf(lag.max = 60,
      main = "SARiMA PACF")
```



We see the peaks at seasonal spots in both graphs- around multiples of 12, so this is promising. Based on the

ACF, I would be able to tell the non seasonal AR order is 0, but the cutoff behavior at the beginning of the ACF would point the the MA order maybe being 1. The cutoff behavior after seasonal spikes on the PACF makes it harder to tell the SAR order. Using PACF and ACF plots for SARIMA data is pretty difficult.