ENV 797 - Time Series Analysis for Energy and Environment Applications | Spring 2025

Assignment 6 - Due date 02/27/25

Justin Maynard

Directions

You should open the .rmd file corresponding to this assignment on RStudio. The file is available on our class repository on Github.

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_Sp25.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: "ggplot2", "forecast", "tseries" and "sarima". 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.

```
#Load/install required package here
library(ggplot2)
library(forecast)
```

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

```
library(tseries)
library(sarima)
```

```
## Loading required package: stats4
```

```
##
## Attaching package: 'sarima'
```

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

```r
library(cowplot)
```

This assignment has general questions about ARIMA Models.

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: The ACF in an AR model will decay exponentially with time, and the PACF will spike and identify the order of the model.

- MA(1)

  Answer: The opposite of the above is true. The PACF will decay exponentially with time, and the ACF will spike and identify the order of the model.

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).

(a) 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 the `arima.sim()` function in R to generate $n = 100$ observations from each of these three models. Then, using `autoplot()` plot the generated series in three separate graphs.

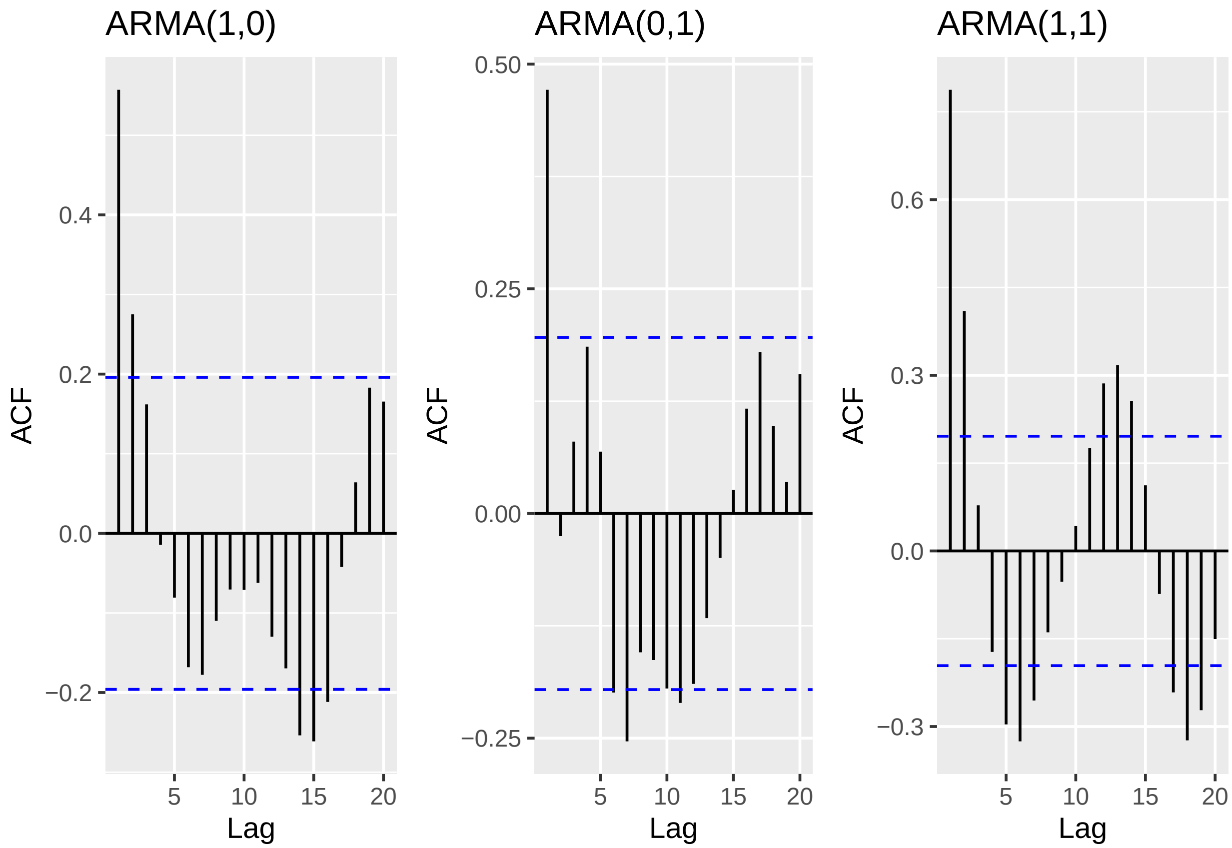
```r
arma1_0 = autoplot(arima.sim(list(order = c(1,0,0), ar = .6), n = 100)) +
  ggtitle("ARMA(1,0)") + ylab("")

arma0_1 = autoplot(arima.sim(list(order = c(0,0,1), ma = .9), n = 100)) +
  ggtitle("ARMA(0,1)") + ylab("")

arma1_1 = autoplot(arima.sim(list(order = c(1,0,1), ar = .6, ma = .9), n = 100)) +
  ggtitle("ARMA(1,1)") + ylab("")
```

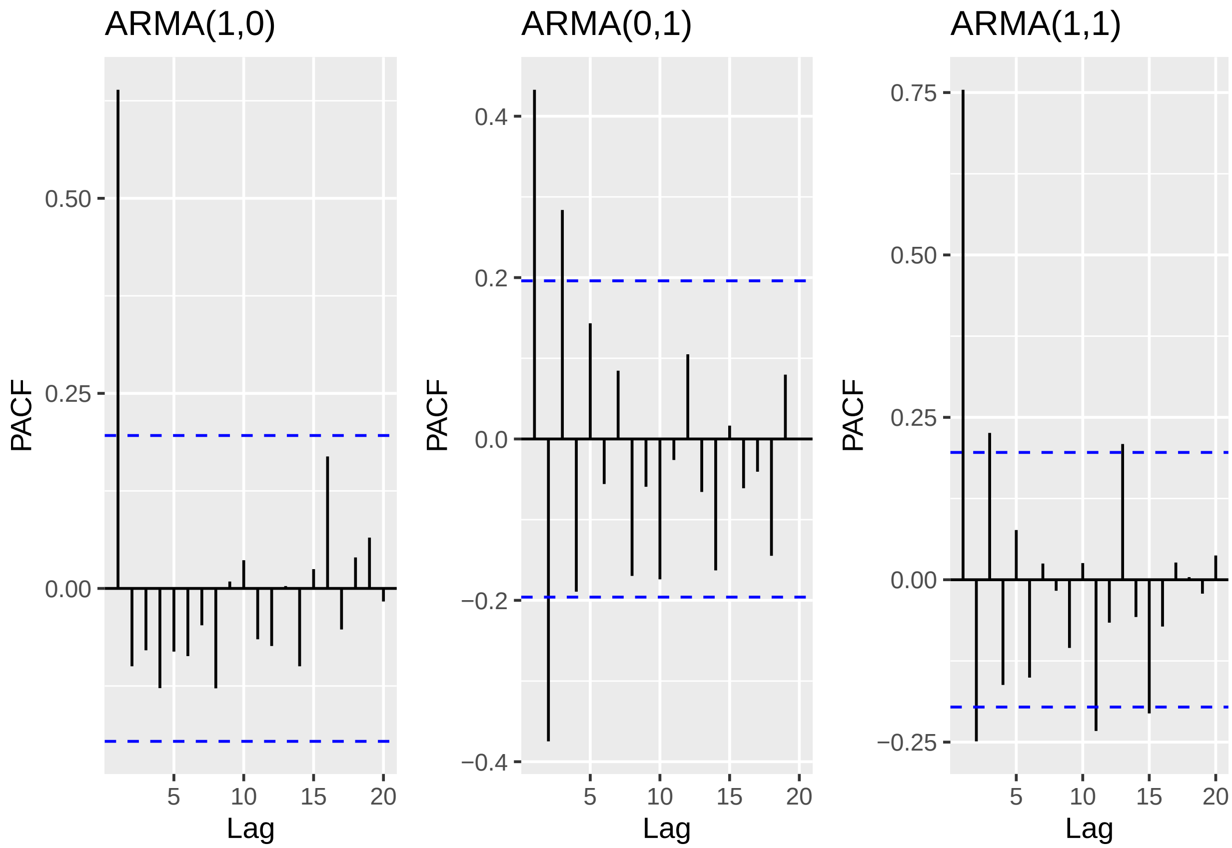
(b) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use `cowplot::plot_grid()`).

```r
plot_grid(qb_acf_1, qb_acf_2, qb_acf_3,
          nrow = 1 )
```



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

```
plot_grid(q3_1, q3_2, q3_3,
          nrow = 1 )
```



(d) 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 able identify them correctly? Explain your answer.

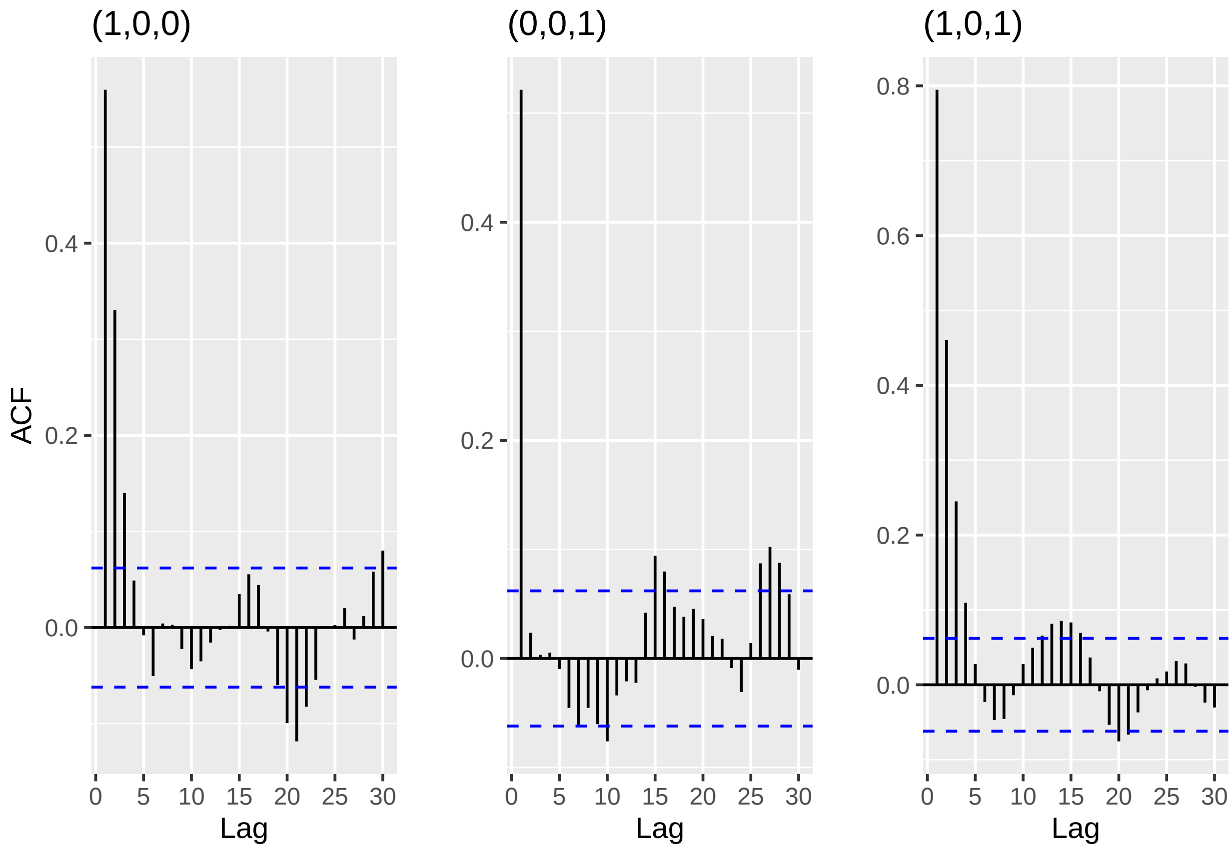
Answer: I would be able to identify the AR process as you can see a clear decay over time in the ACF and a cutoff at lag 1 in the PACF. In the MA process the ACF does identify a clear cutoff and there is delay over time in the PACF. The ARMA process can be identified by a MA order in the ACF (cutoff) with a decay from the AR, with the opposite for PACF.

(e) Compare the PACF values R computed with the values you provided for the lag 1 correlation coefficient, i.e., does $\phi = 0.6$ match what you see on PACF for ARMA(1,0), and ARMA(1,1)? Should they match?

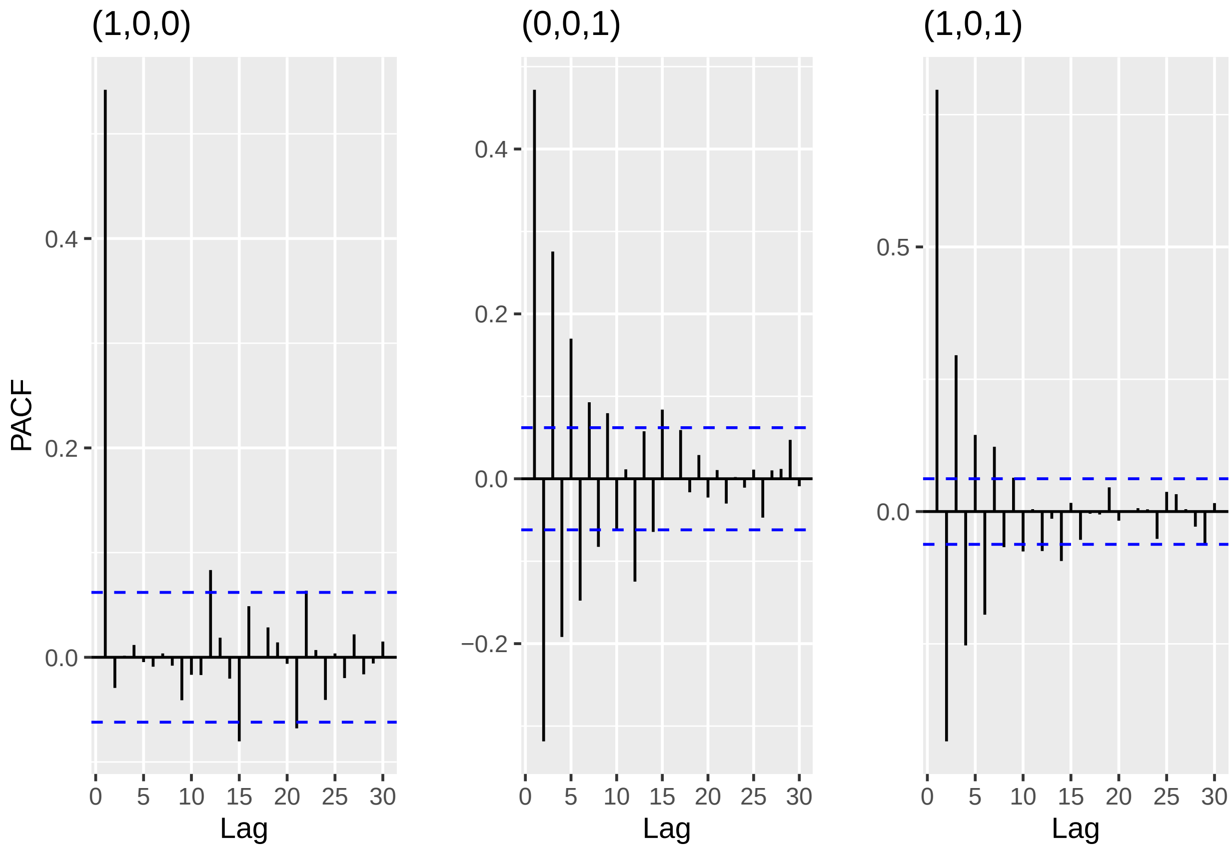
Answer: Phi does not appear to match for the ARMA(1,0) modela st he PACF appears to be lower than .6. The PACF does appear to match for the ARMA(1,1) model. These vales may differe due to the moving average component.

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

```
plot_grid(model_1_acf + ggtitle("(1,0,0)"), model_2_acf + ggtitle("(0,0,1)") + ylab(""), model_3_acf + g
          nrow = 1 )
```




```
plot_grid(model_1_pacf+ ggtitle("(1,0,0)"),model_2_pacf + ggtitle("(0,0,1)")+ ylab(""),model_3_pacf  + g
          nrow = 1 )
```



After increasing the number of observations the differences between the models become more pronounced and easily identifiable. The AR decay and MA cutoff in the ACF is more apparent, as is the AR cutoff and MA decay in the PACF. The value of Phi appears to be much closer to .6 in the AR ACF, and is closer to .8 in the ARMA ACF.

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: q:1 P:1 D: Q:0 s:12

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

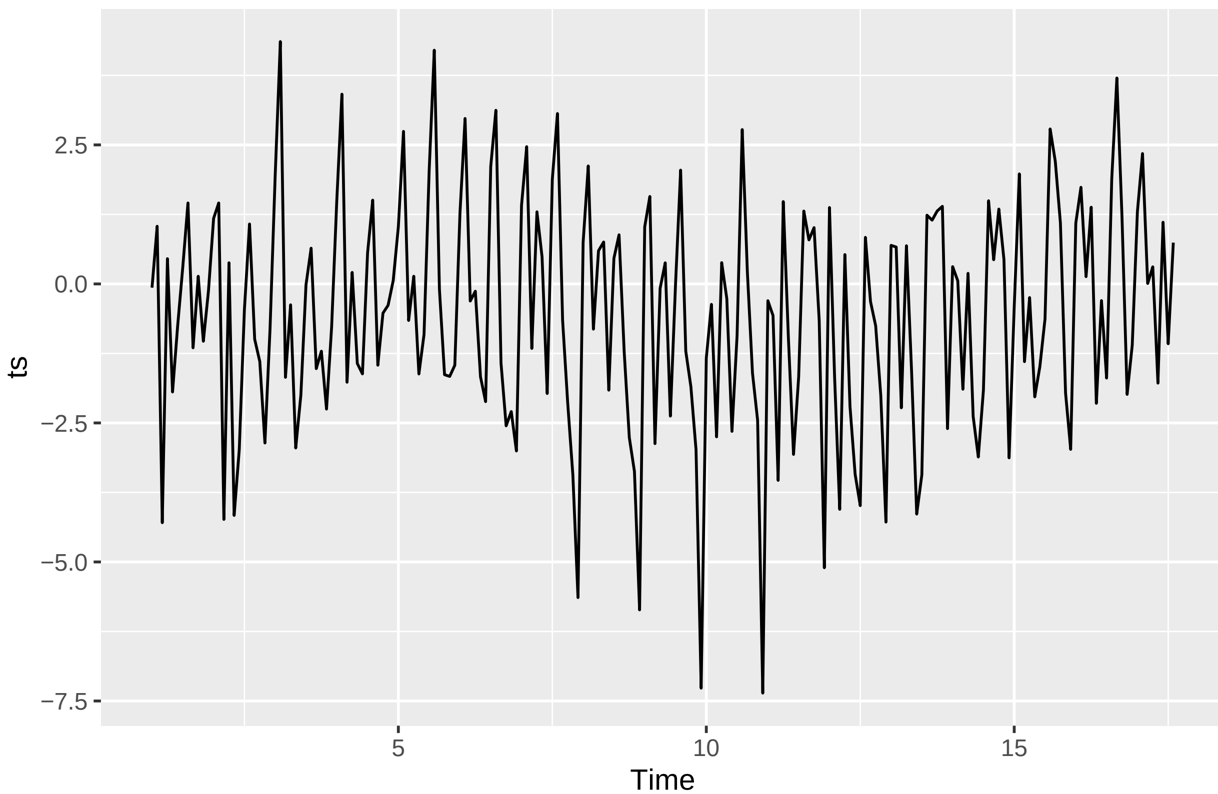
Q4

phi 1 = .7 phi 12 = .25 theta 1 = .1

Simulate a seasonal ARIMA$(0, 1) \times (1, 0)_{12}$ model with $\phi = 0.8$ and $\theta = 0.5$ using the `sim_sarima()` function from package `sarima`. 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 the generated series using `autoplot()`. Does it look seasonal?

```r
data <- sim_sarima(n = 200, model = list(ma = .5, sar = .8, nseasons = 12))
ts <- ts(data, frequency = 12)

autoplot(ts)
```

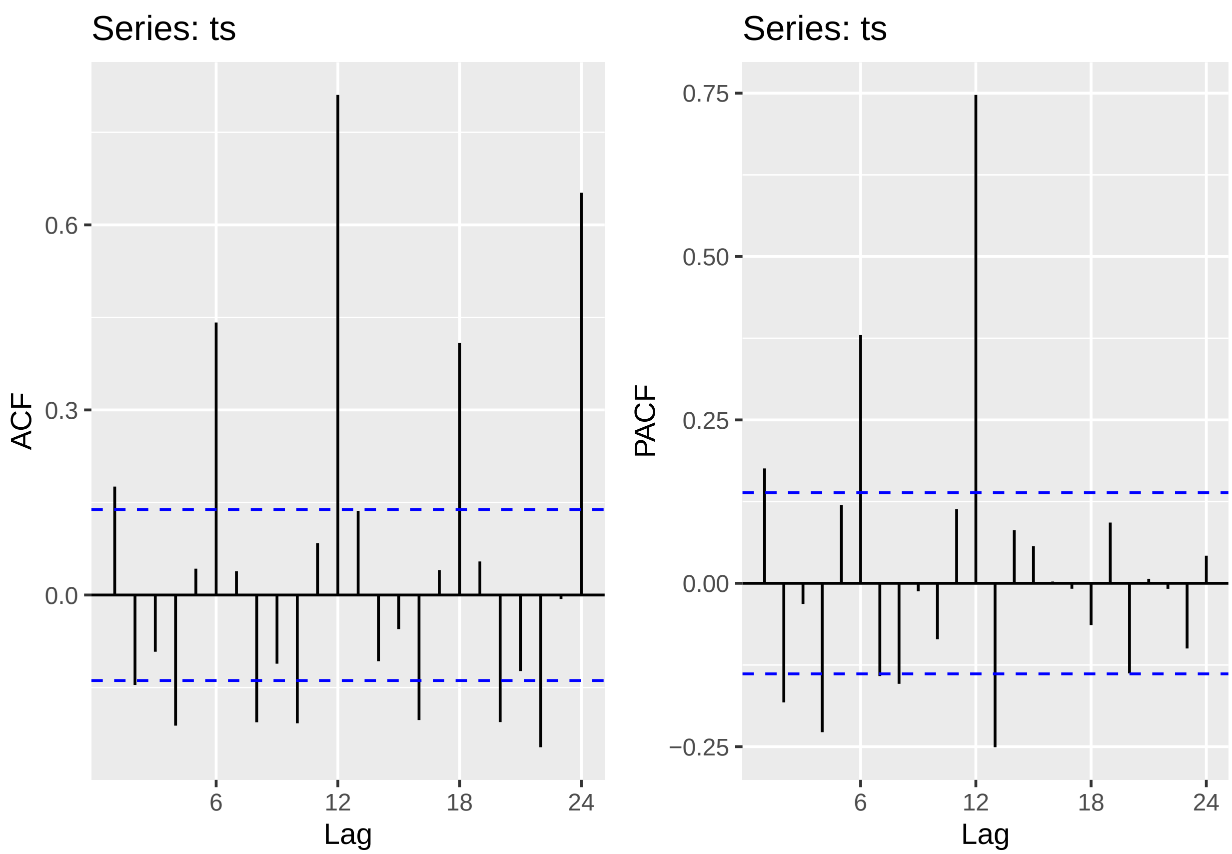


It does appear to be seasonal, with monthly swings in the data.

Q5

Plot ACF and PACF of the simulated series in Q4. 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
plot_grid(q5acf,q5pacf)
```



Seasonal AR can be identified by multiple spikes at the 12 lag lntervals in the ACF, but the MA process can not be identified.