ENV 797 - Time Series Analysis for Energy and Environment Applications | Spring 2024

Assignment 6 - Due date 02/28/24

Julia Kagiliery

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_Sp24.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.

```
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
```

```
library(cowplot)
library(fable) #chatgpt suggestion for using autoplot with series
```

```
## Loading required package: fabletools
```

This assignment has general questions about ARIMA Models.

Using ADF to determine if it is stochiastic or deterministic

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:AR models should have ACF that will decay exponentially with time and PACFs that identify the order of the AR model (this is an order two model so the PACF should be max 2)

- MA(1)

  Answer: MA models should have PACF that will decay exponentially with time and ACFs that identify the order of the AR model (this is a first order so ACF should max at 1)

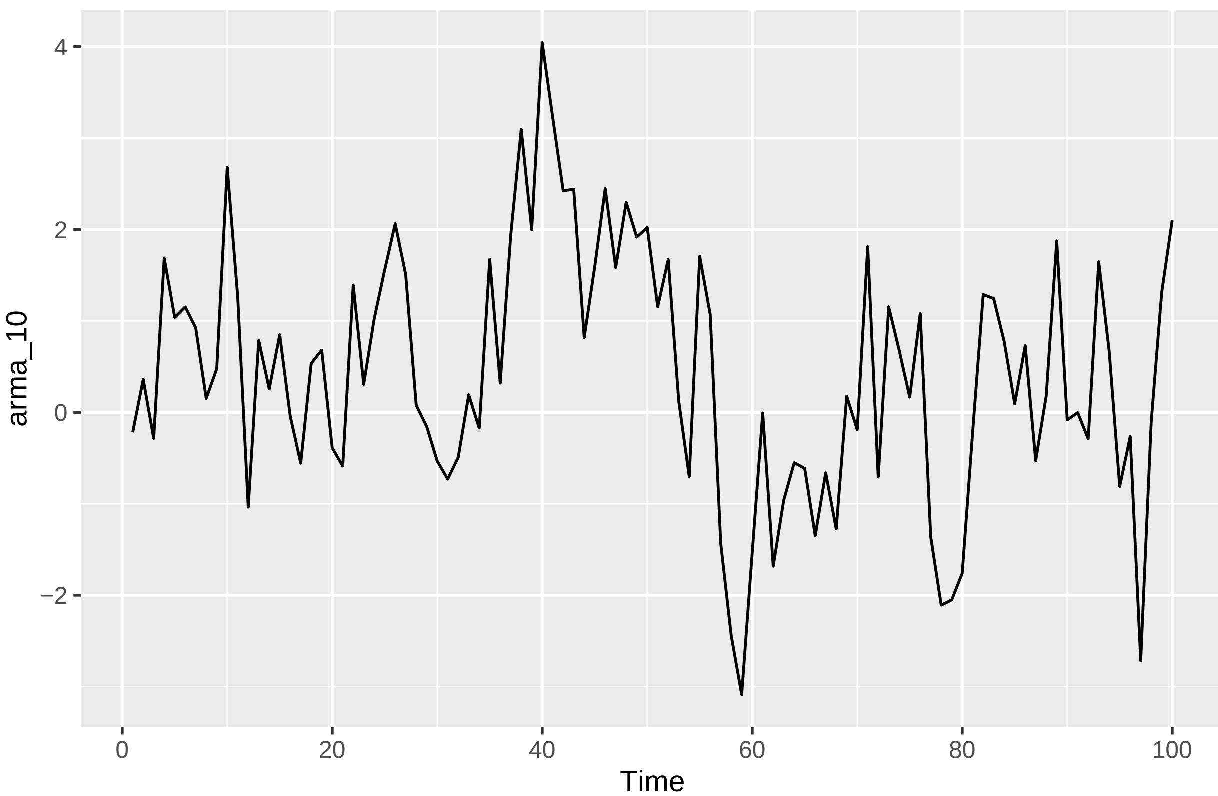
Q2

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).

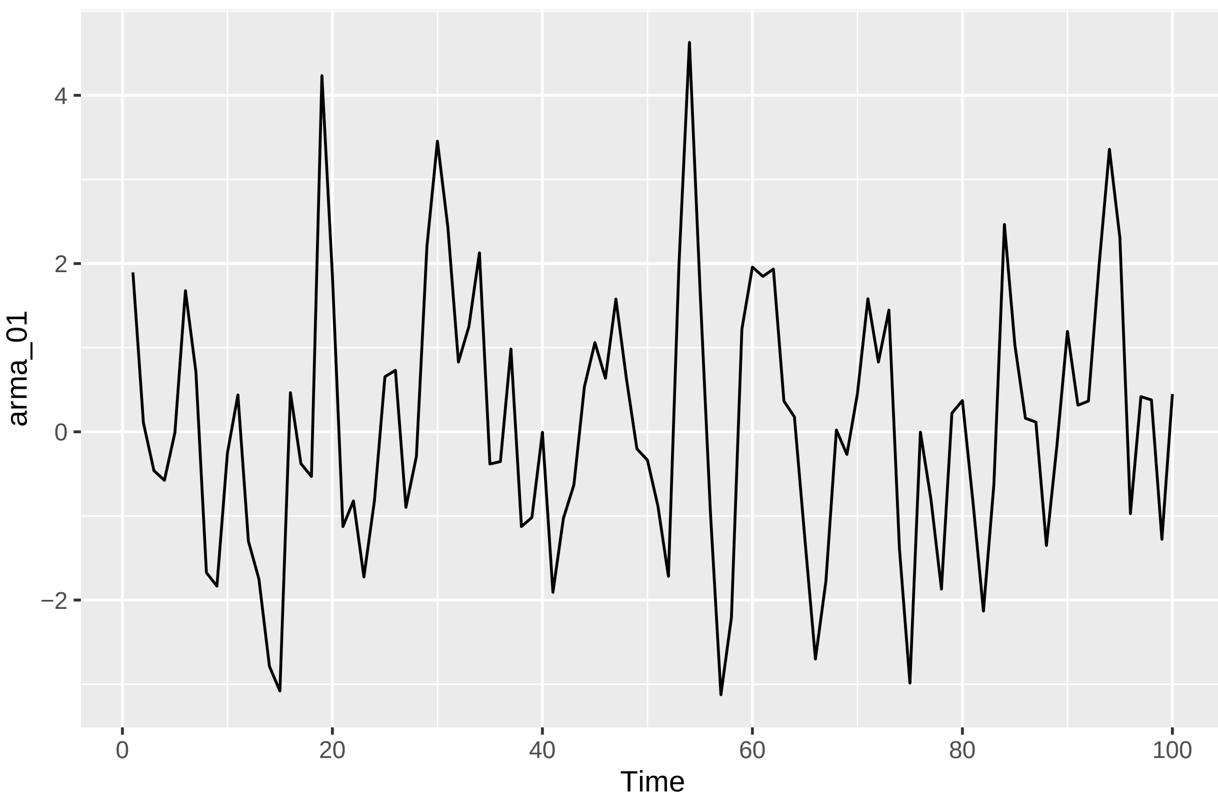
(a) Consider three models: ARMA(1,0), ARMA(0,1) and ARMA(1,1) with parameters $\phi = 0.6$ and $\theta = 0.9$. The ϕ refers to the AR coefficient and the θ refers to the MA coefficient. Use the `arima.sim()` function in R to generate $n = 100$ observations from each of these three models. Then, using `autoplot()` plot the generated series in three separate graphs.

```r
arma_10 <- arima.sim(list(ar = 0.6), 100)
arma_01 <- arima.sim(list(ma = 0.9), 100)
arma_11 <- arima.sim(list(ar = 0.6, ma = 0.9), 100)
```

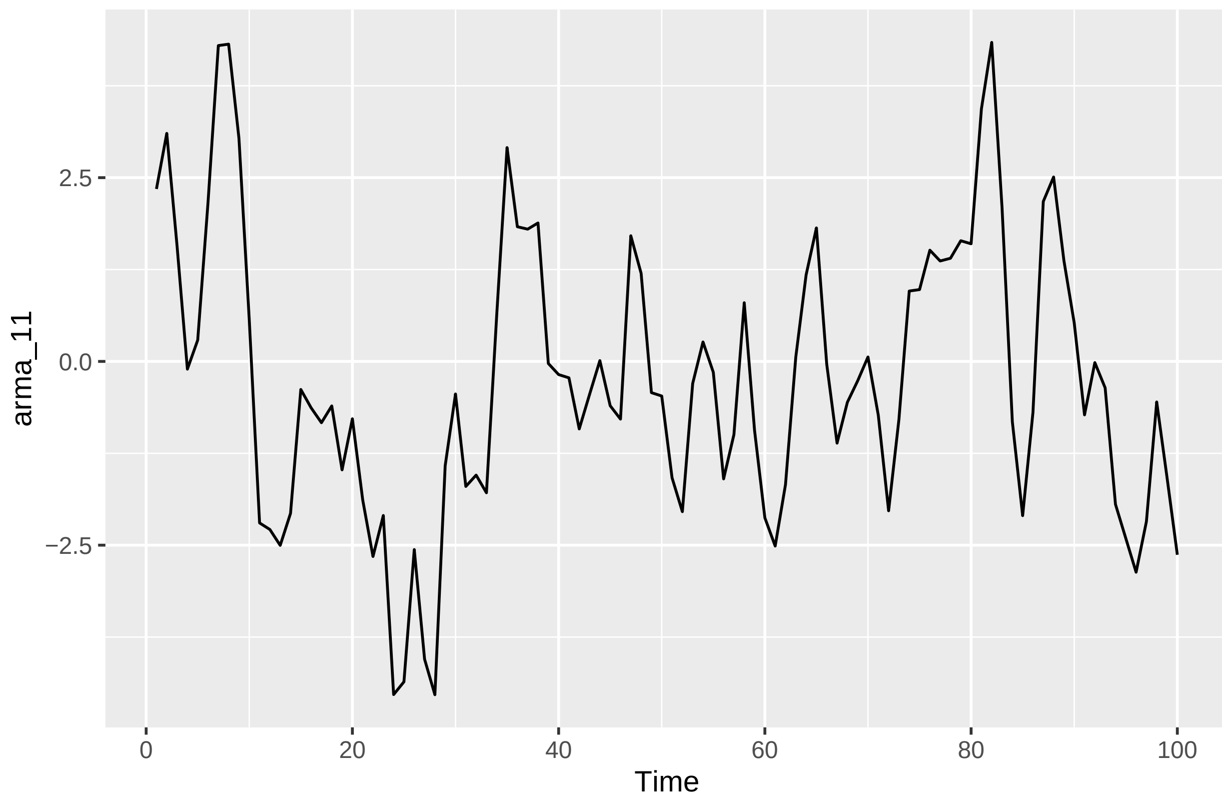
```r
autoplot(arma_10)
```




```r
autoplot(arma_01)
```

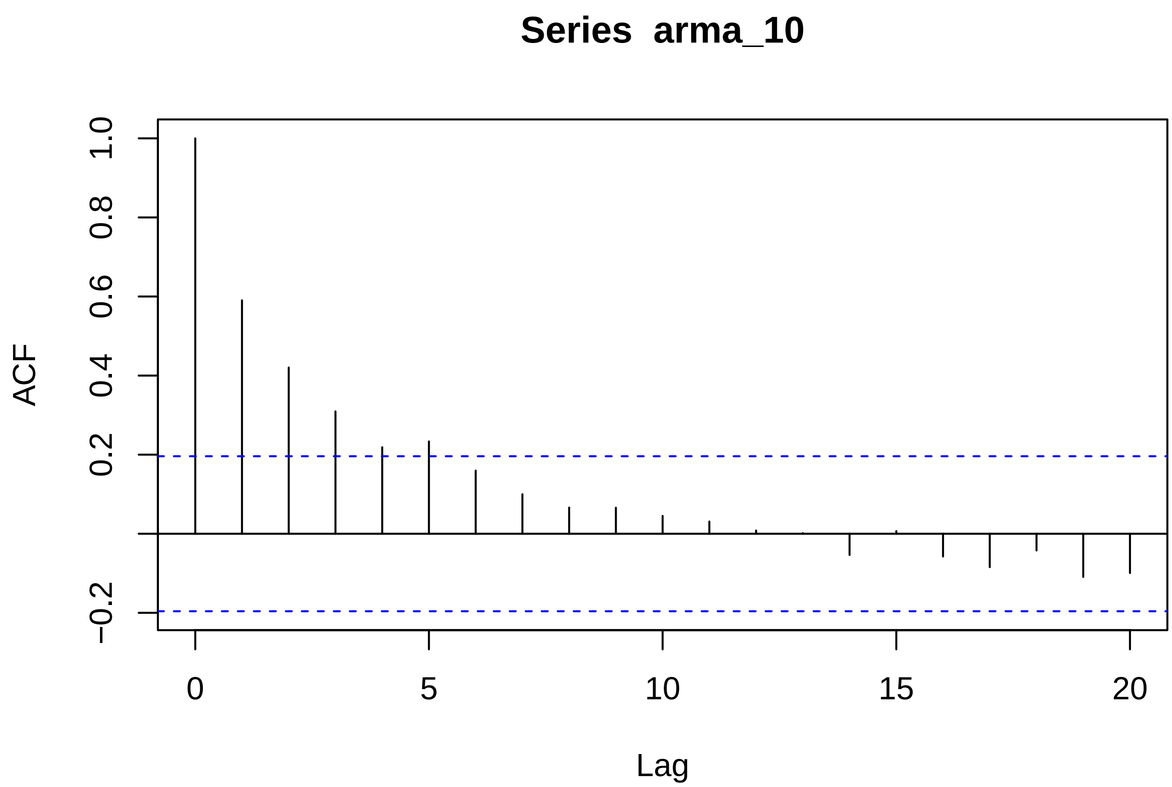



```r
autoplot(arma_11)
```

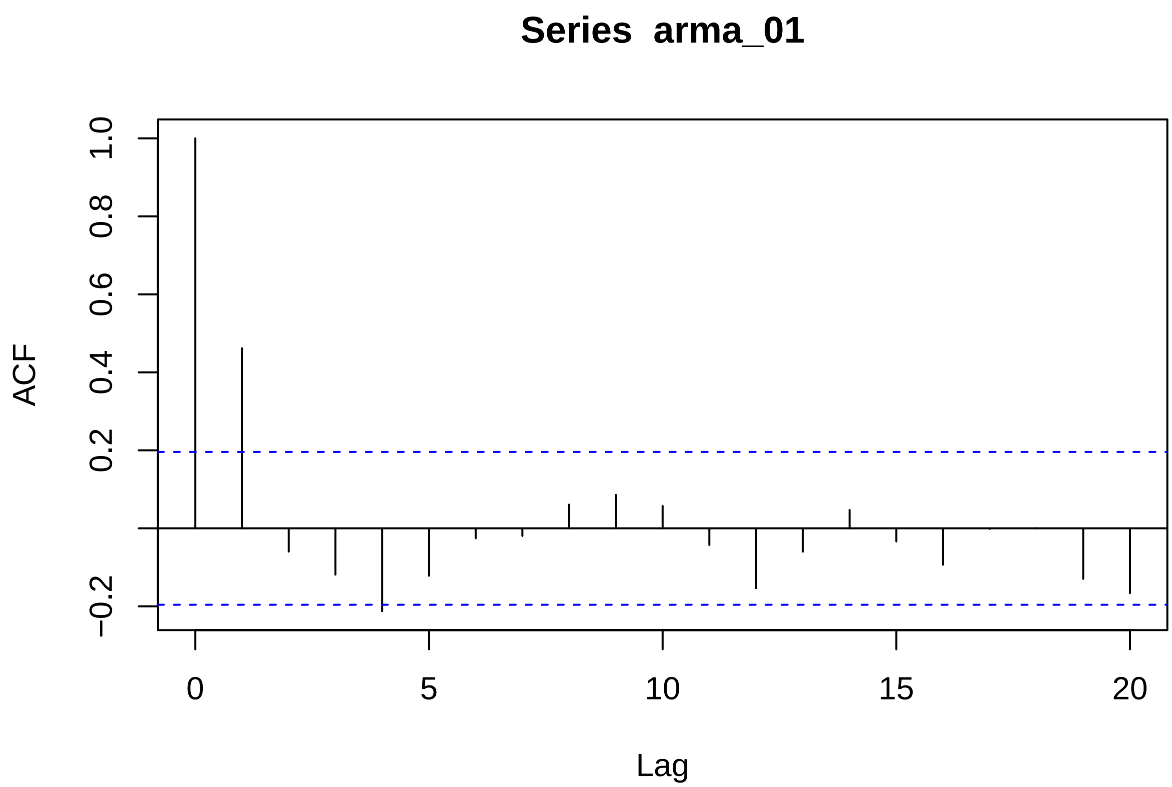


(b) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use `cowplot::plot_grid()`).

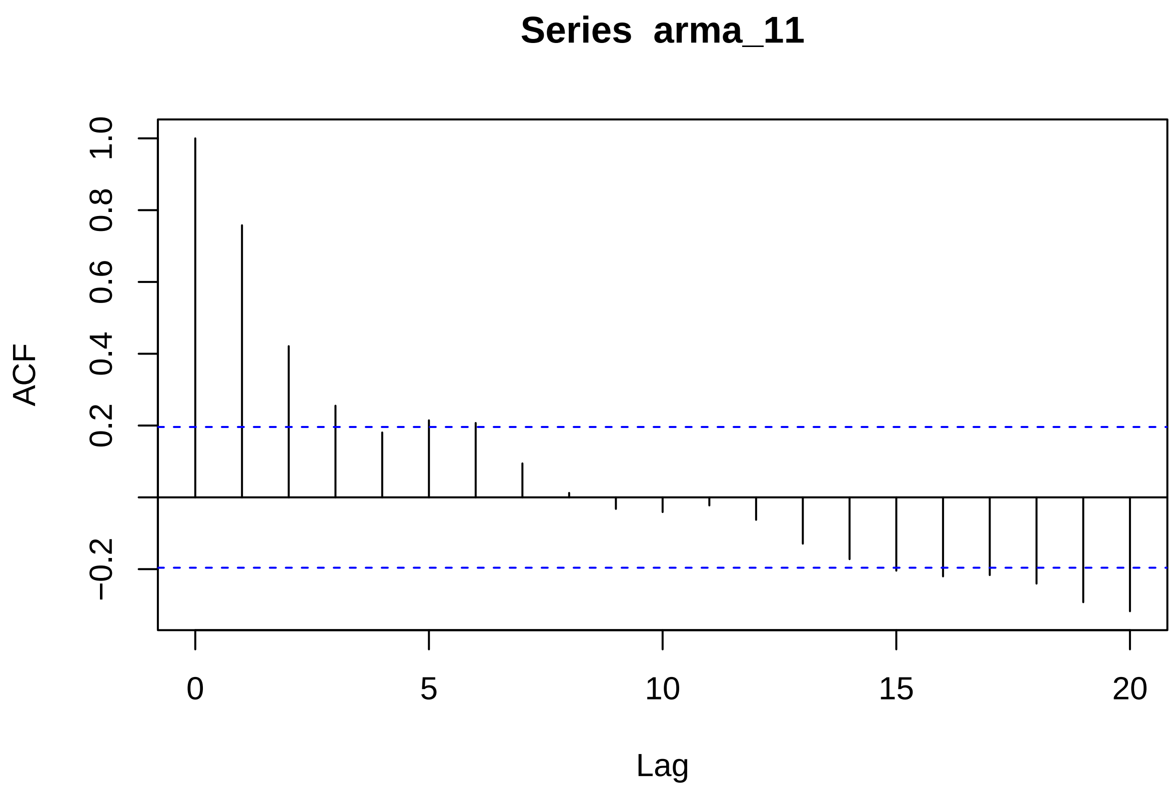
```
arma_10ACF <- autoplot(acf(arma_10))
```



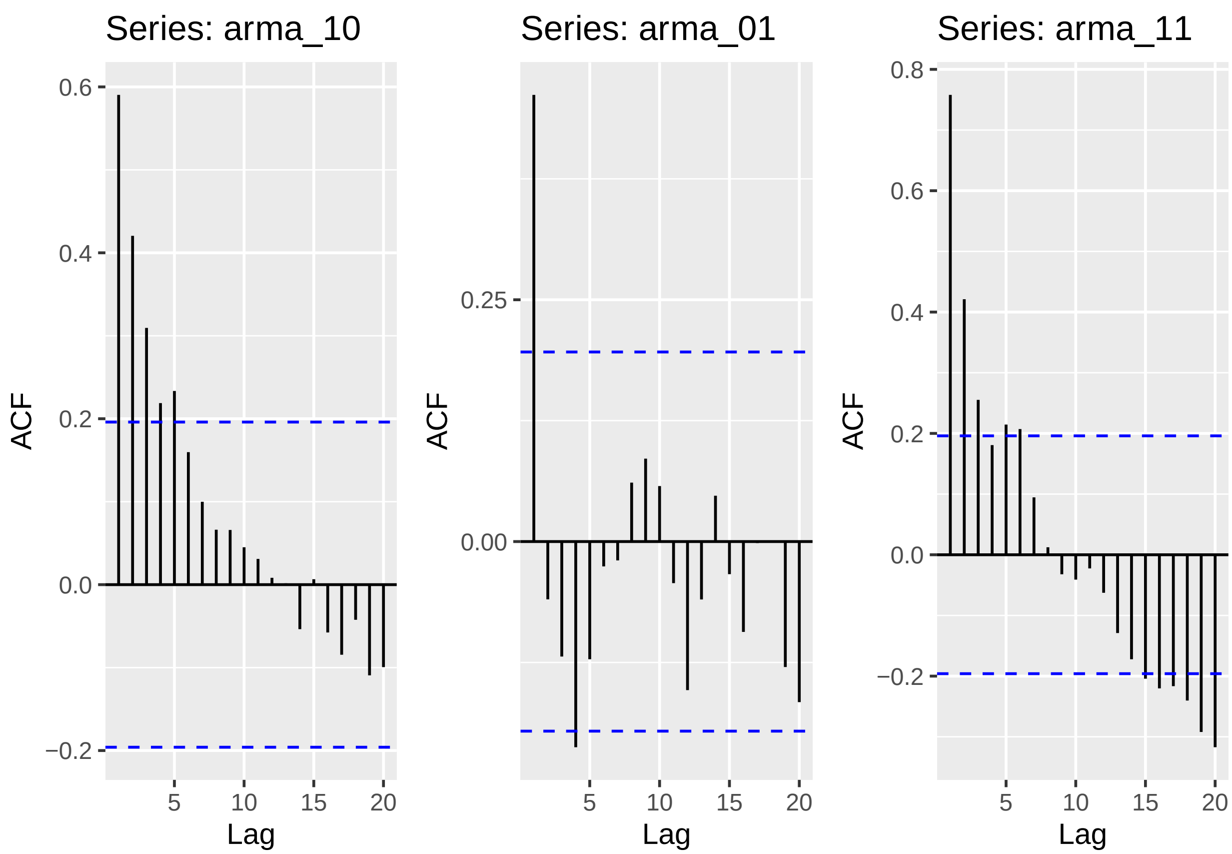

```
arma_01ACF <- autoplot(acf(arma_01))
```




```
arma_11ACF <- autoplot(acf(arma_11))
```

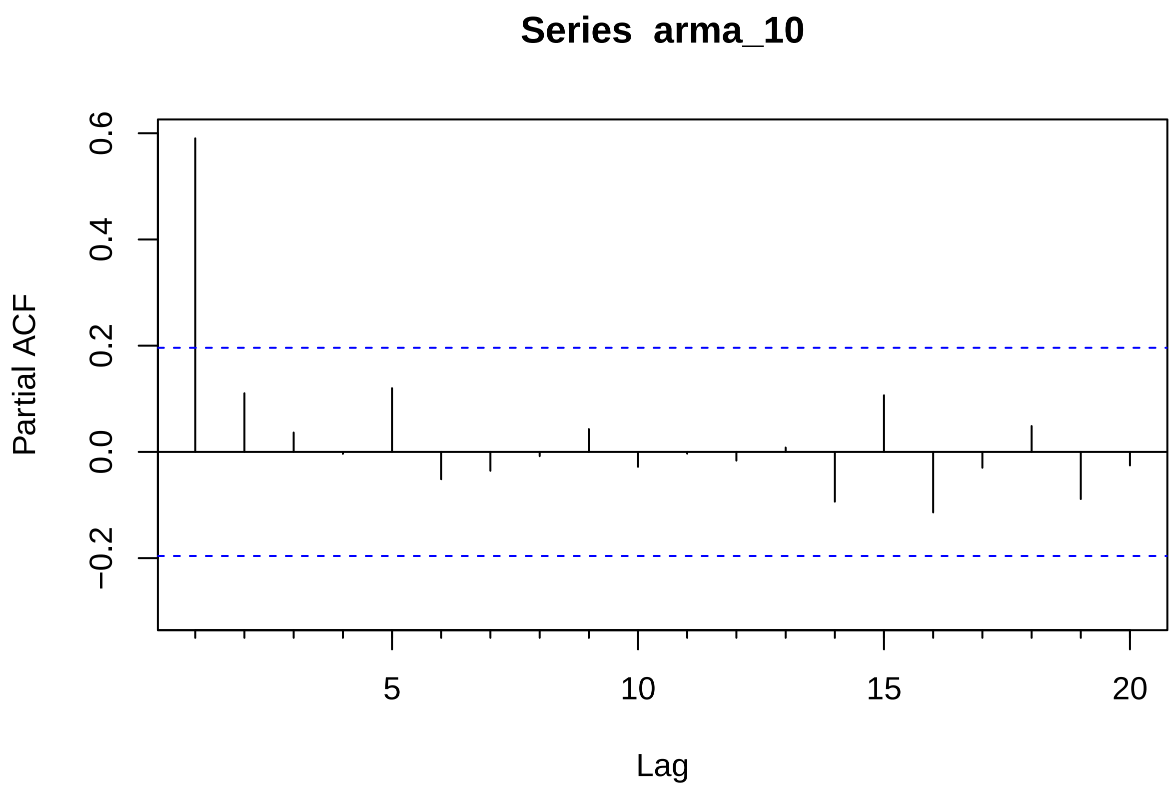



```
plot_grid(arma_10ACF, arma_01ACF, arma_11ACF, nrow = 1)
```

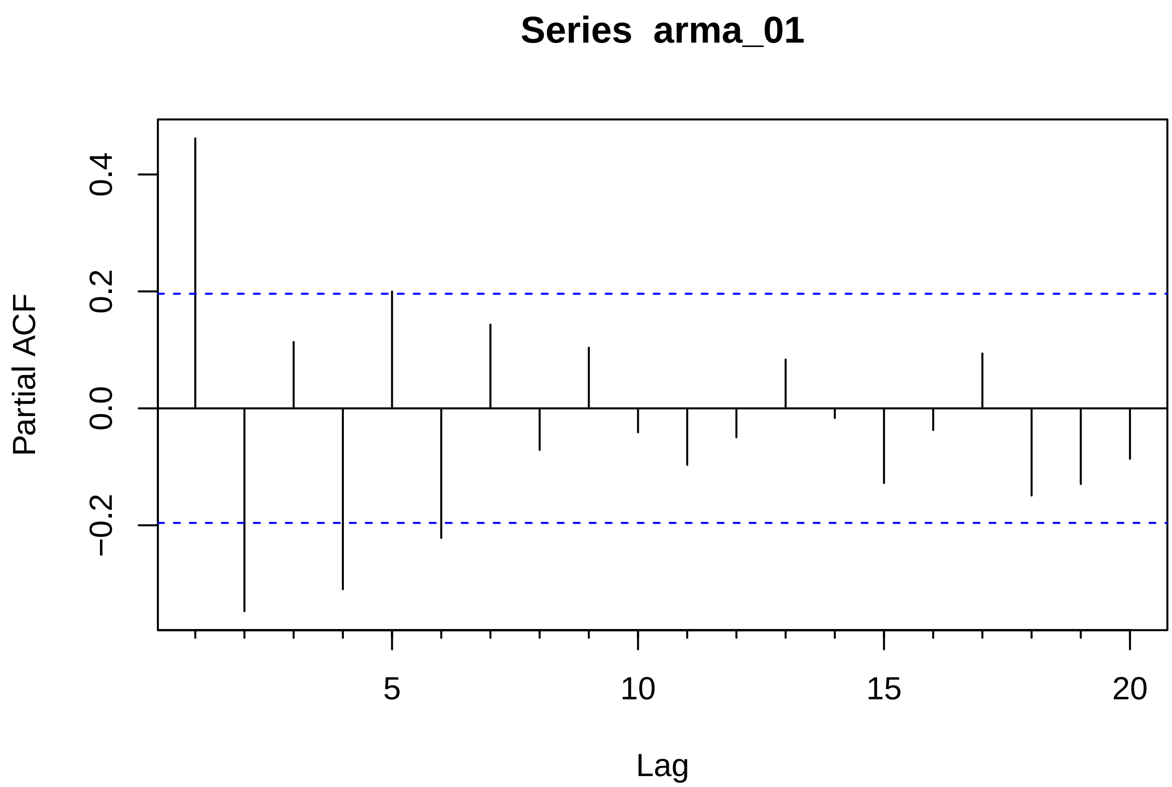


(c) Plot the sample PACF for each of these models in one window to facilitate comparison.

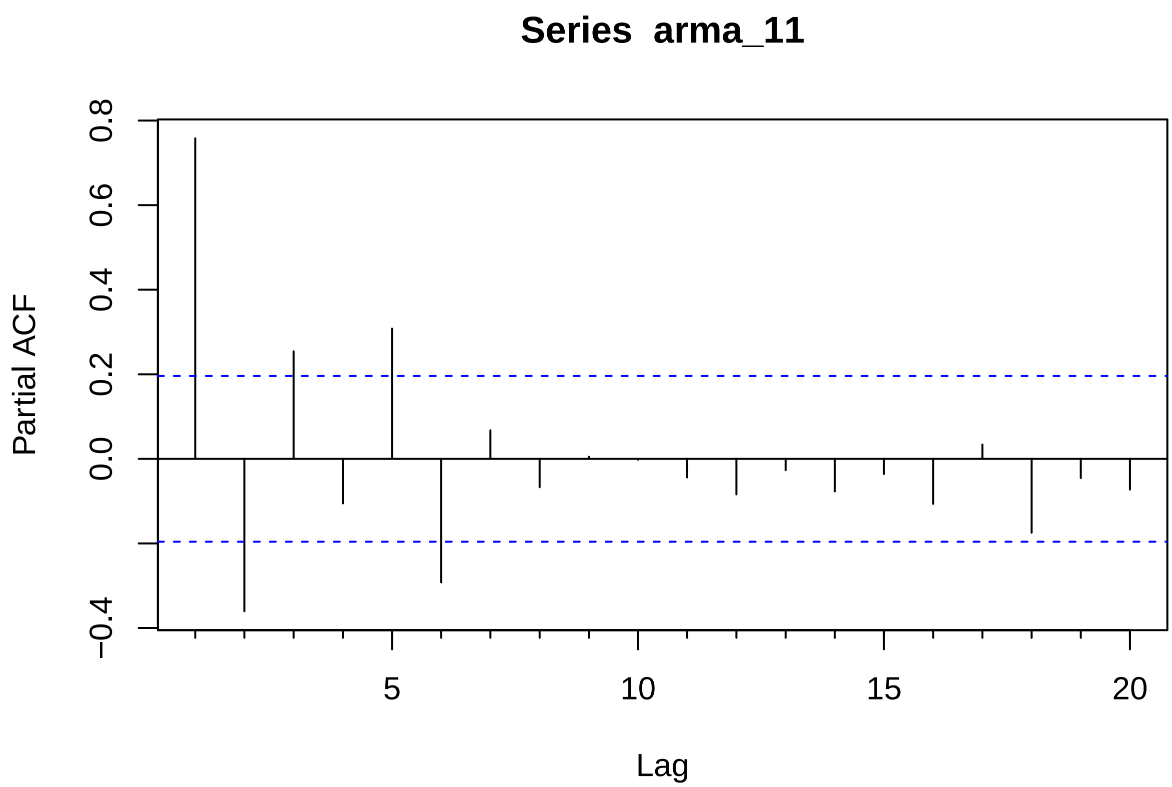
```
arma_10PACF <- autoplot(Pacf(arma_10))
```



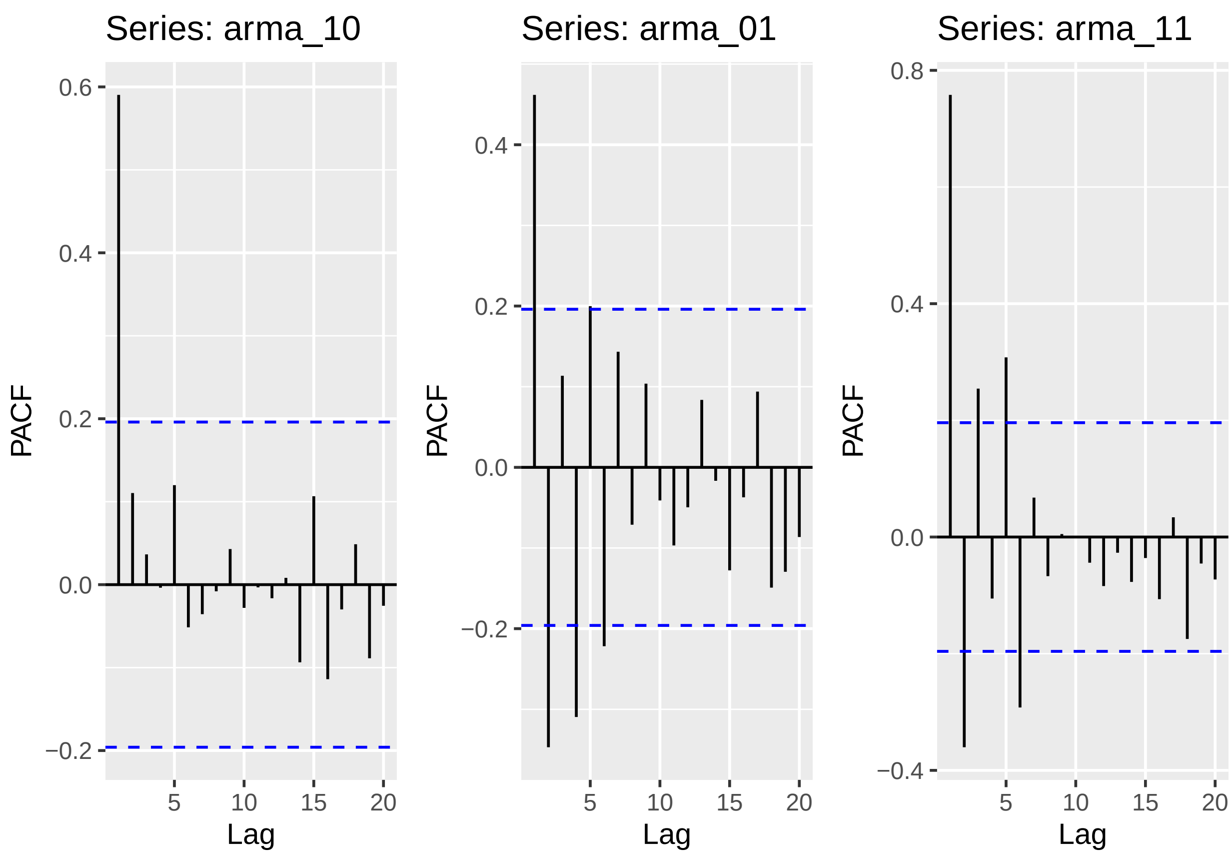

```
arma_01PACF <- autoplot(Pacf(arma_01))
```




```
arma_11PACF <- autoplot(Pacf(arma_11))
```




```
plot_grid(arma_10PACF, arma_01PACF, arma_11PACF, 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:

q = ACF for the MA component p = PACF for the AR component

I would be able to make an educated guess about the model for sure: ARMA(1,0): spike at lag 1 in the ACF plot, and the PACF should decay exponentially. ARMA(0,1): spike at lag 1 in the PACF plot, and the ACF should decay exponentially. ARMA(1,1): significant spikes at lag 1 in both ACF and PACF plots.

The magnitudes of these spikes should also be near the 0.8 and 0.5 values (which many of them are)

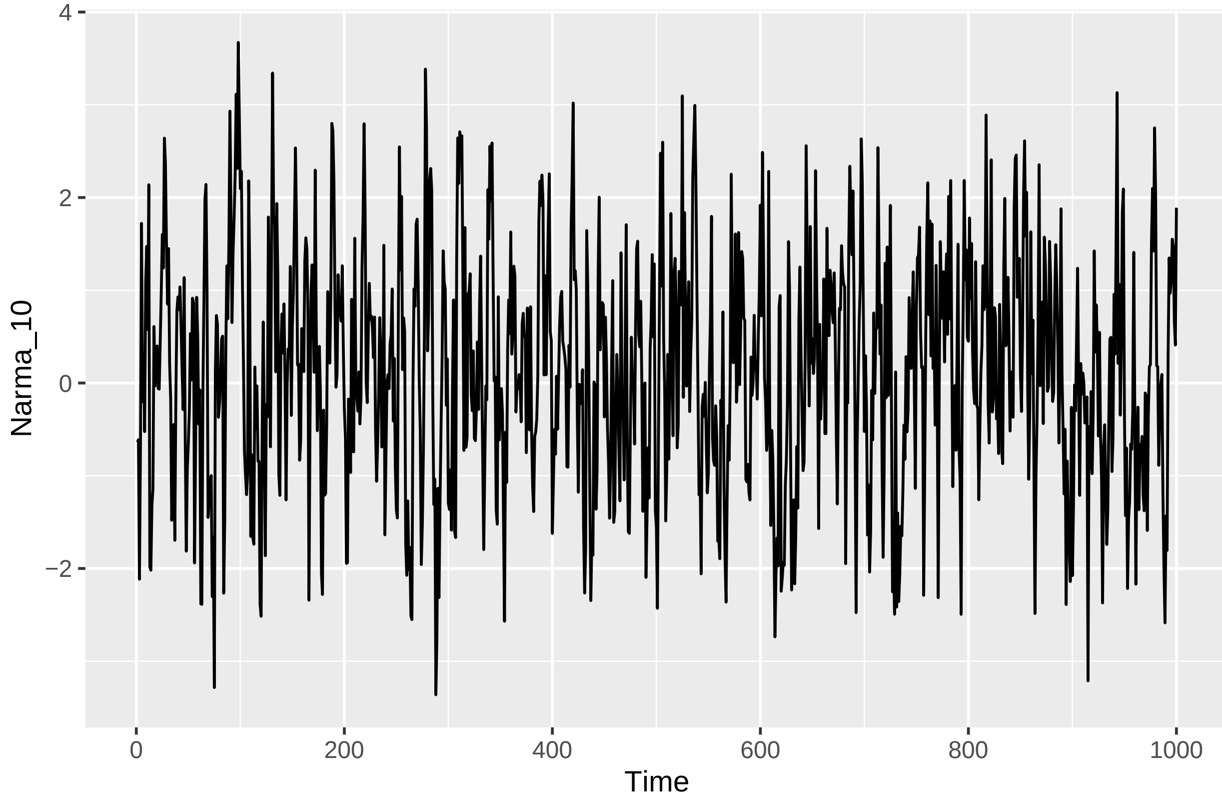
(e) Compare the PACF values R computed with the values you provided for the lag 1 correlation coefficient, i.e., does $\phi = 0.6$ match what you see on PACF for ARMA(1,0), and ARMA(1,1)? Should they match?

Answer:

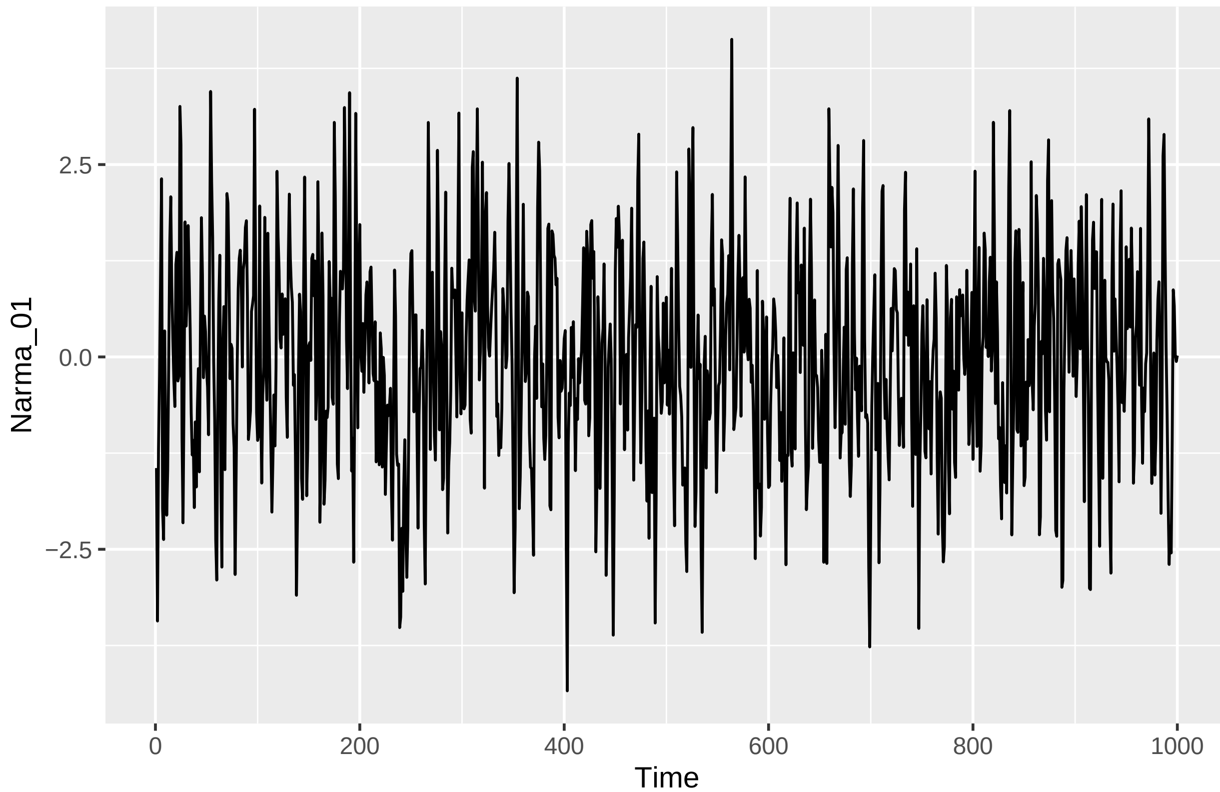
Yes they should match because p = 1 in so the spike should be in the first lag and equal in magnitude. I would have expected the magnitued to equal 0.6 but it is close. This could be due to maybe outliers or more difrencing needed. The results are not perfect but the do not imediately dismiss the model.

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

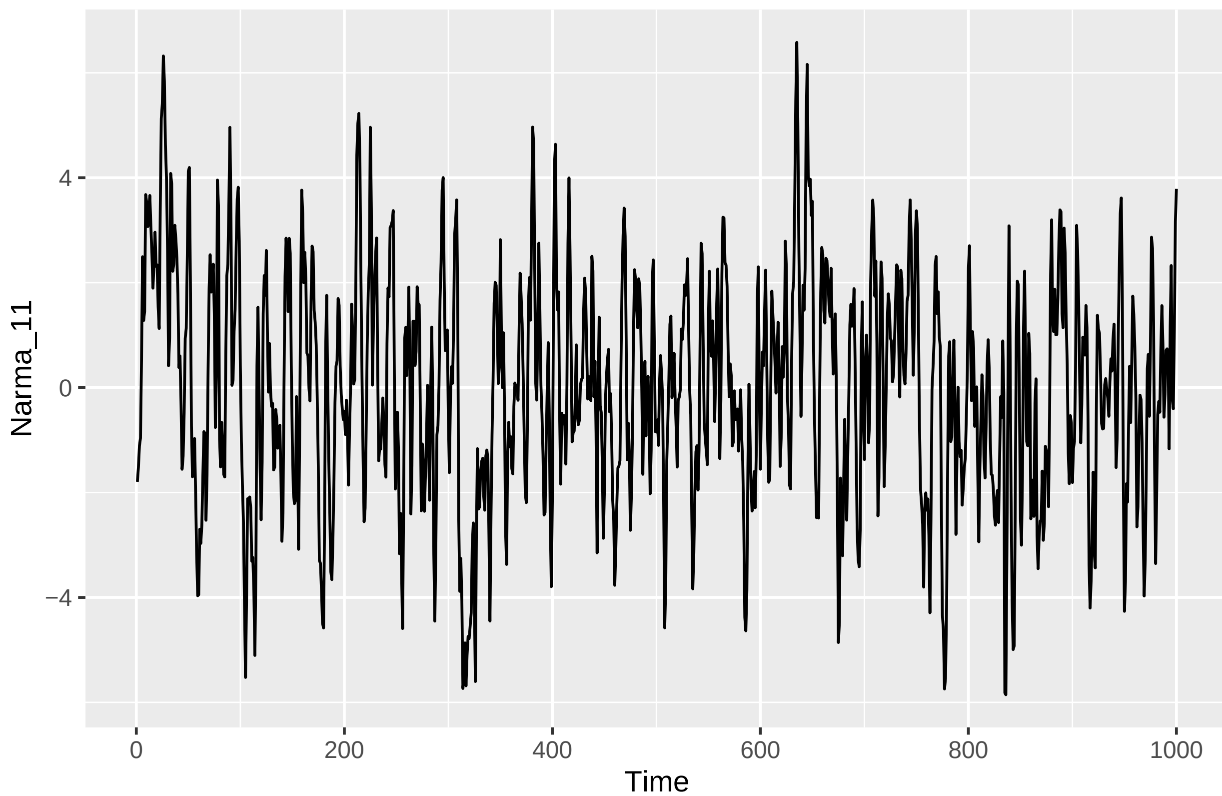
```
Narma_10 <- arima.sim(list(ar = 0.6), 1000)
Narma_01 <- arima.sim(list(ma = 0.9), 1000)
Narma_11 <- arima.sim(list(ar = 0.6, ma = 0.9), 1000)

autoplot(Narma_10)
```



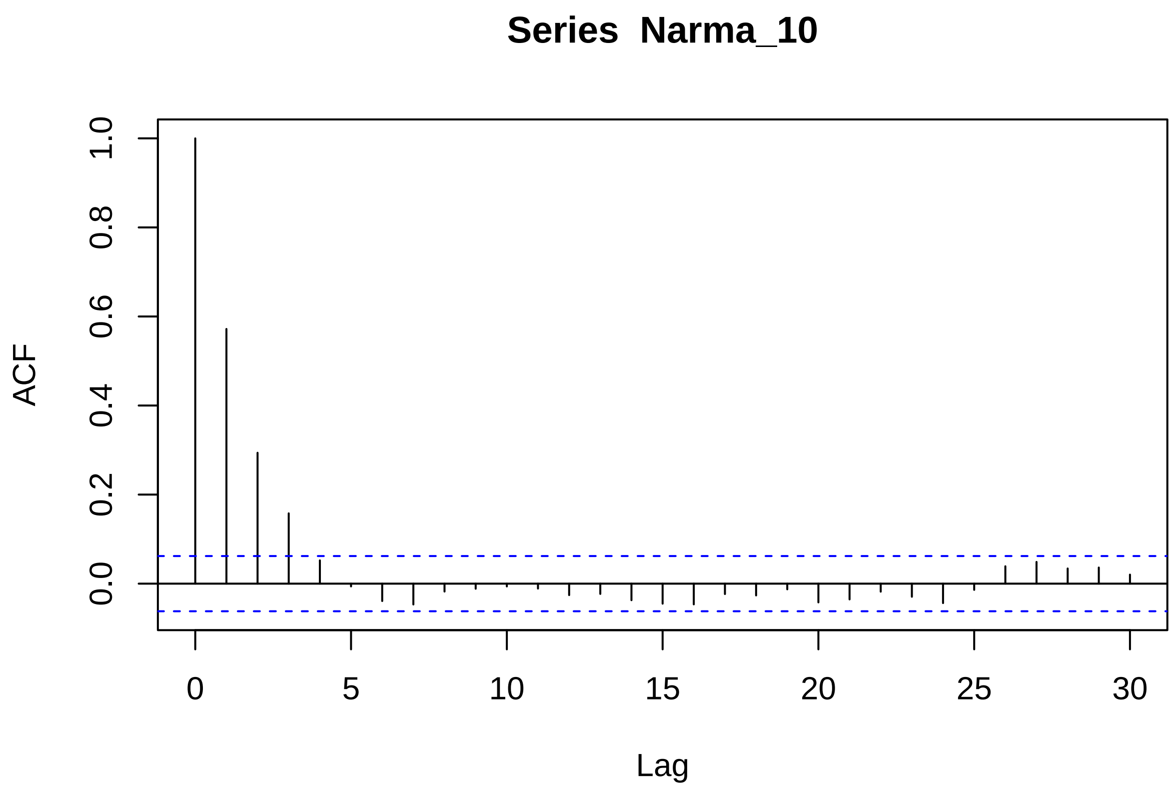

```
autoplot(Narma_01)
```



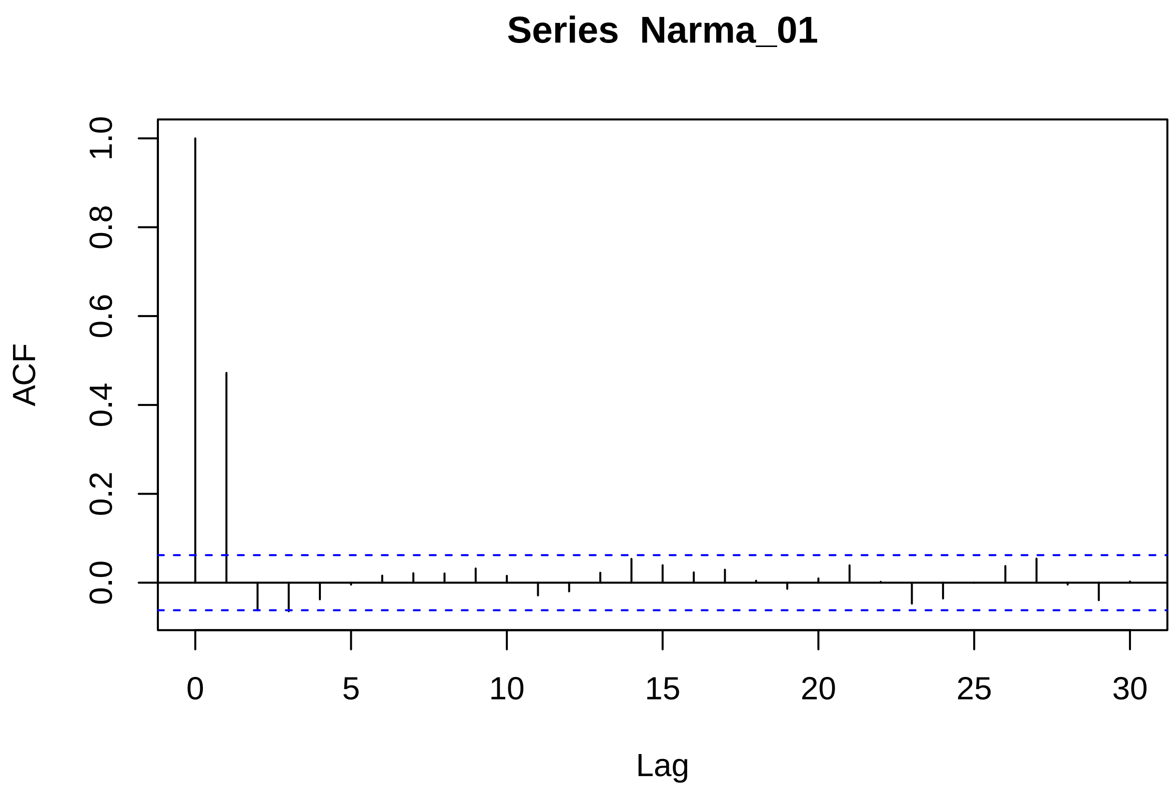

```
autoplot(Narma_11)
```



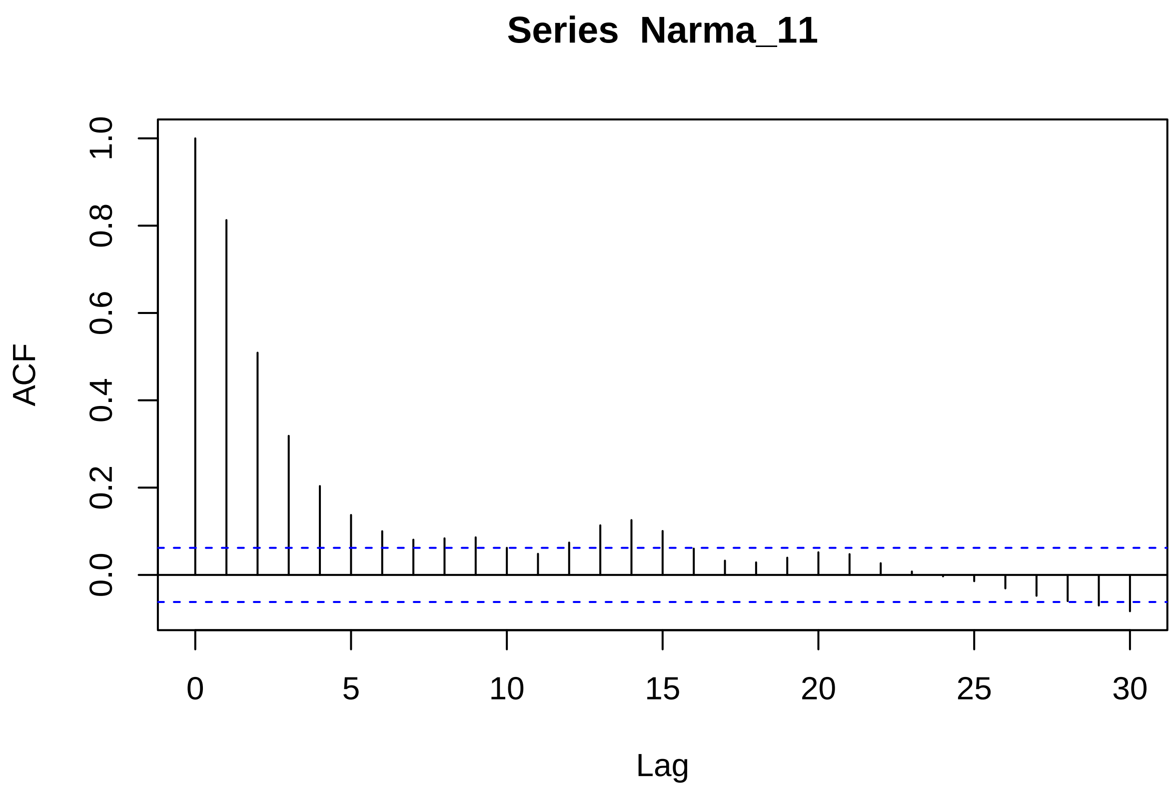

```
Narma_10ACF <- autoplot(acf(Narma_10))
```



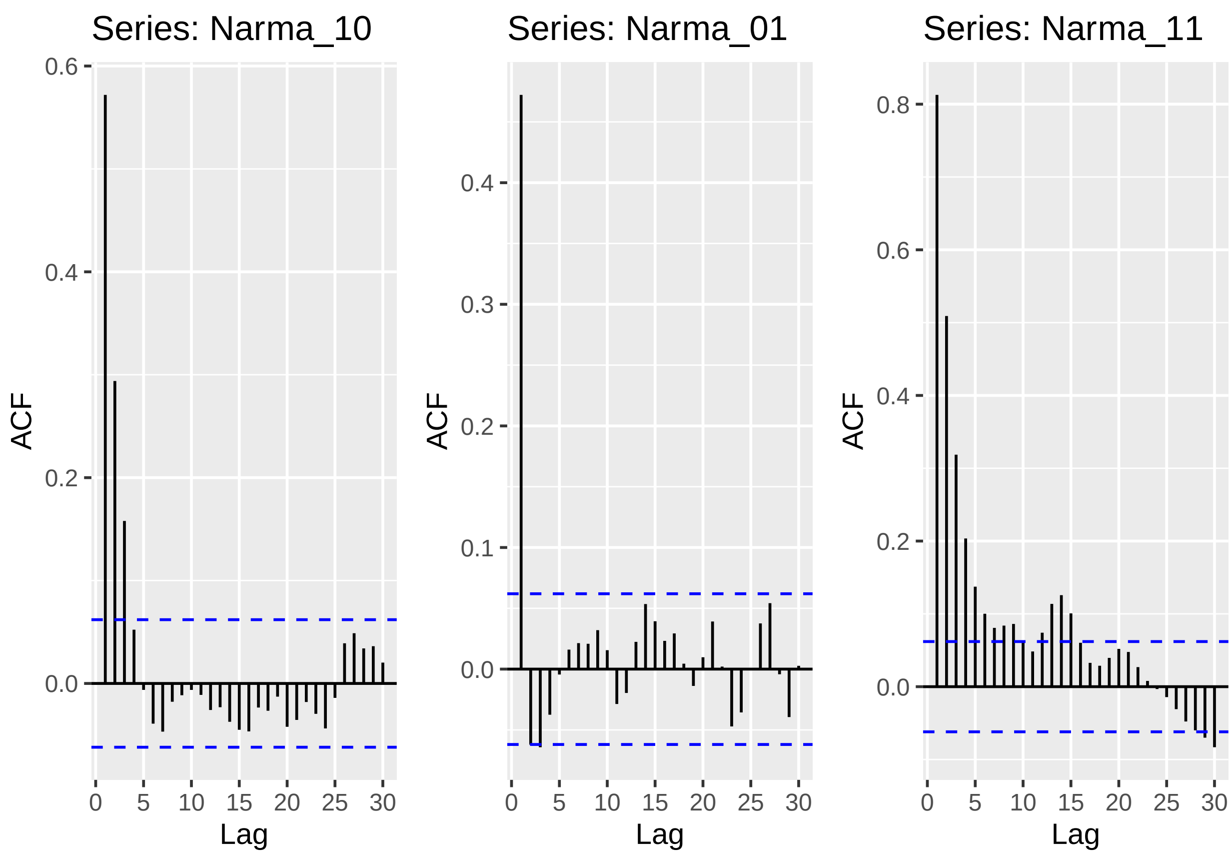

```r
Narma_01ACF <- autoplot(acf(Narma_01))
```



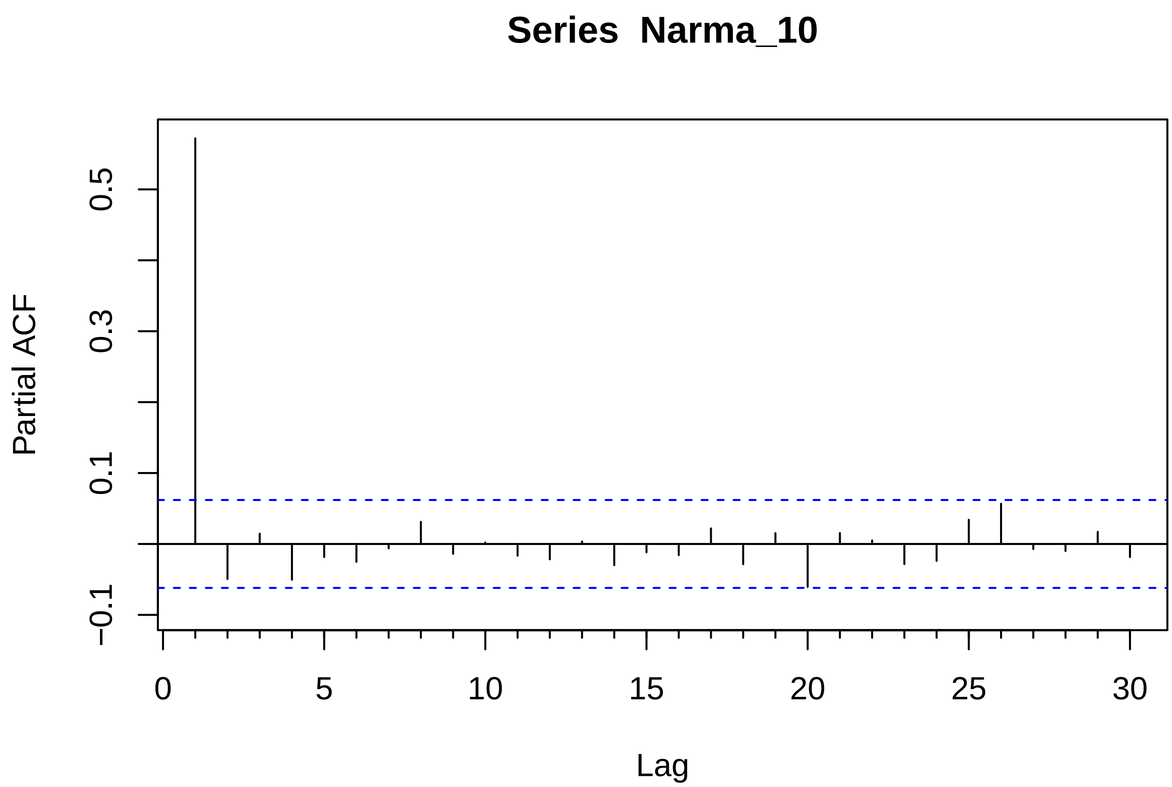

```
Narma_11ACF <- autoplot(acf(Narma_11))
```



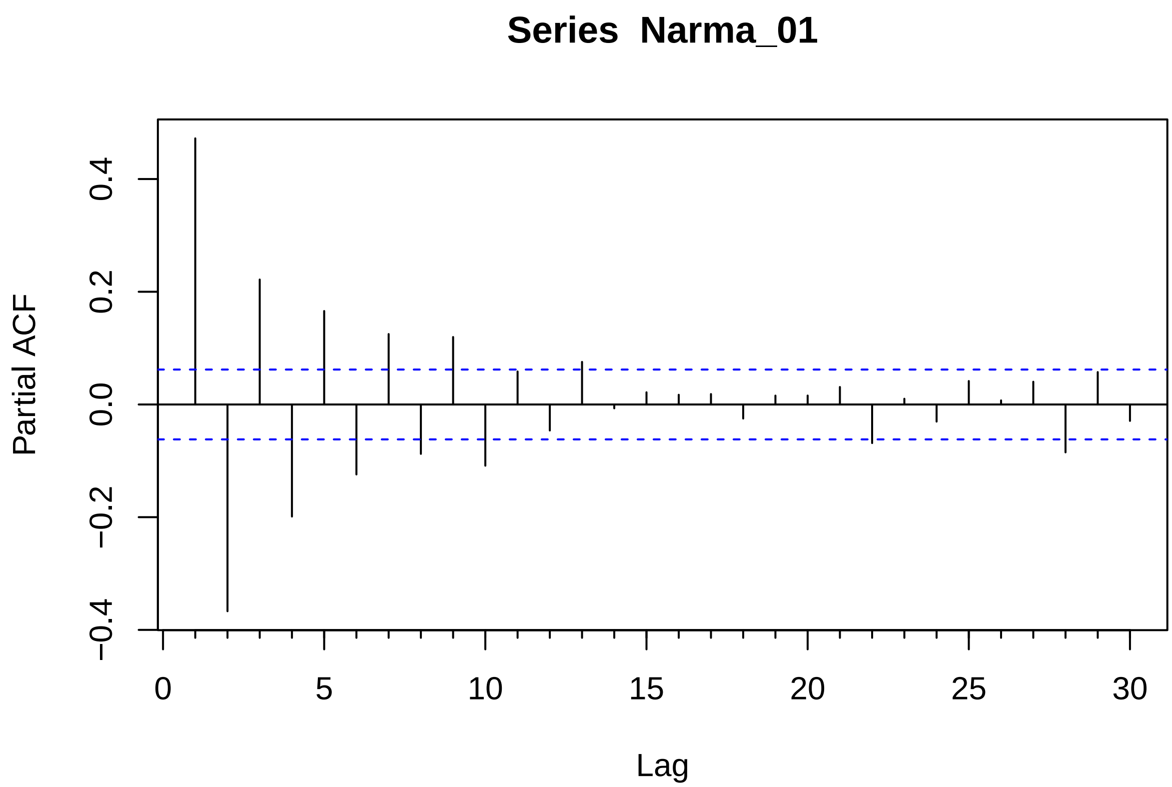

```
plot_grid(Narma_10ACF, Narma_01ACF, Narma_11ACF, nrow = 1)
```



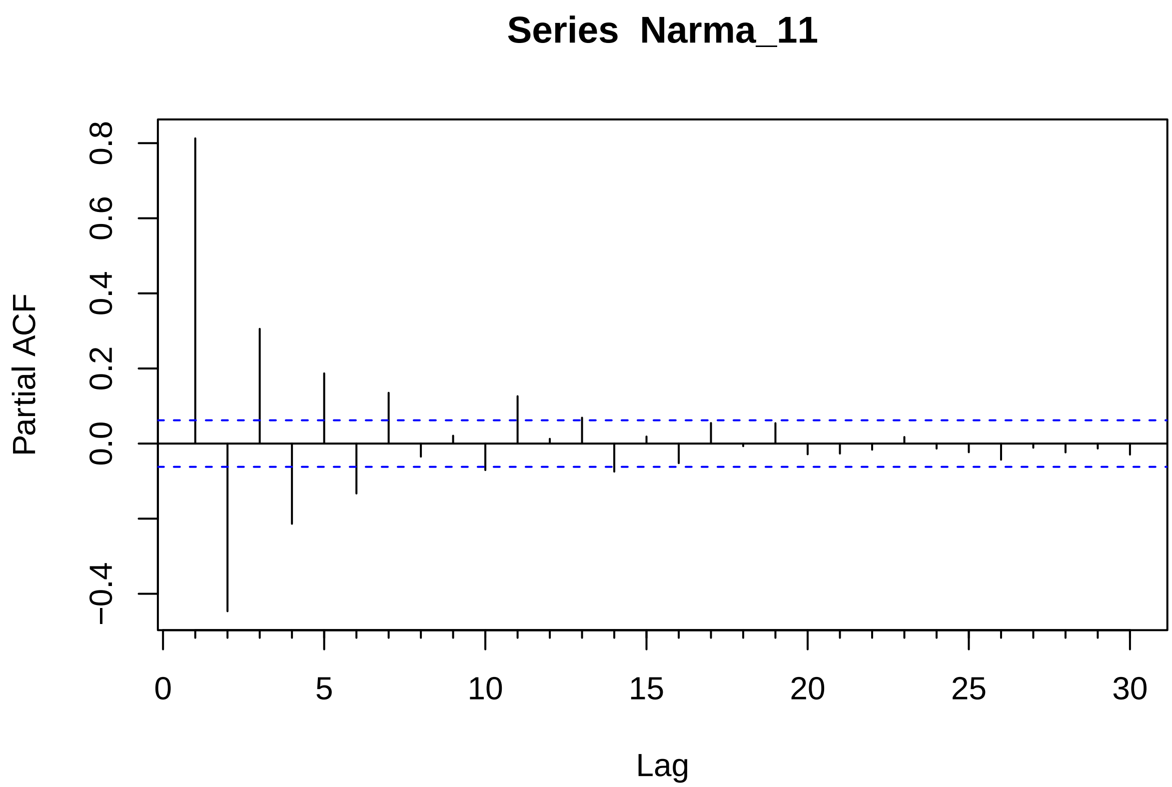

```
Narma_10pACF <- autoplot(Pacf(Narma_10))
```



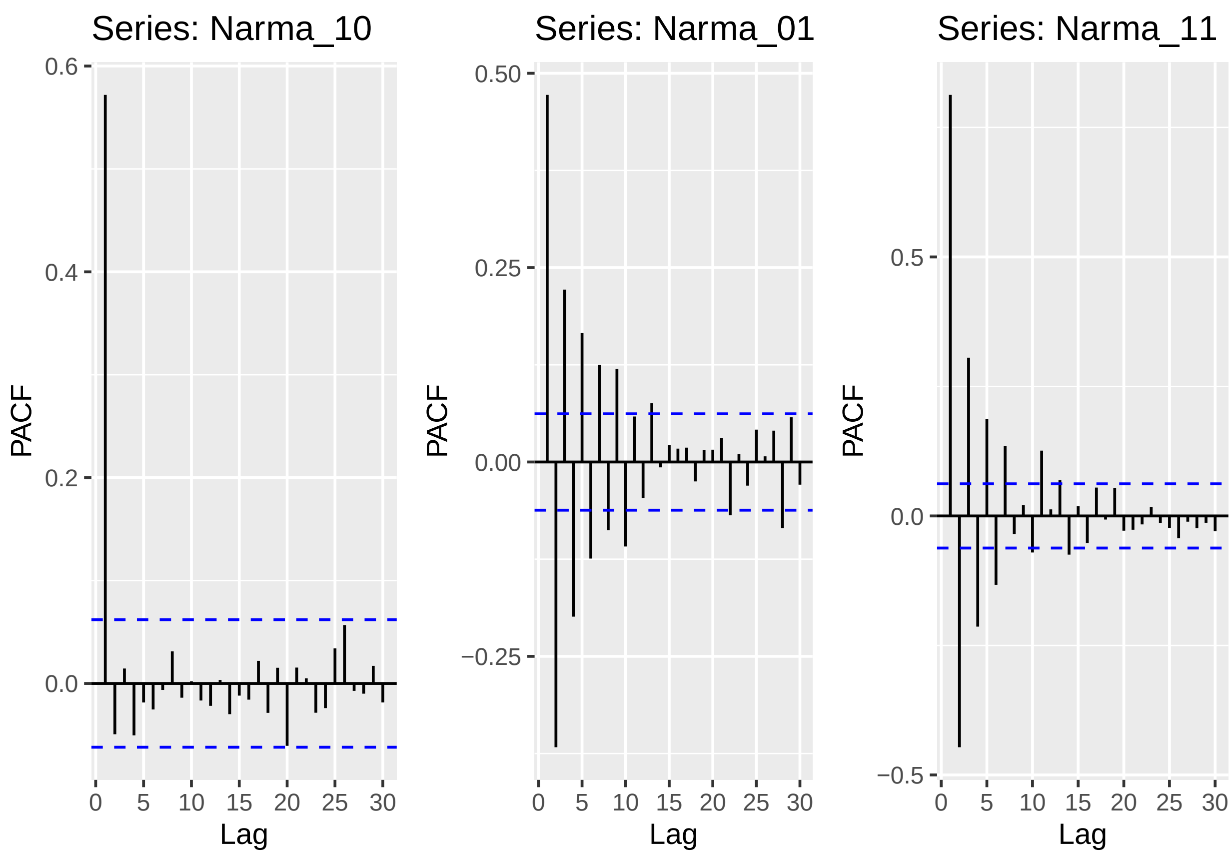

```
Narma_01pACF <- autoplot(Pacf(Narma_01))
```




```r
Narma_11pACF <- autoplot(Pacf(Narma_11))
```




```r
plot_grid(Narma_10pACF, Narma_01pACF, Narma_11pACF, nrow = 1)
```



Much of part d and e remains the same (copied below) but the increasing the n seems to have driven our PCAF and ACF to the given 0.6 and 0.8 values as is kind of expected (more repitition = better accuracy).

Answer:

q = ACF for the MA component p = PACF for the AR component

I would be able to make an educated guess about the model for sure: ARMA(1,0): spike at lag 1 in the ACF plot, and the PACF should decay exponentially. ARMA(0,1): spike at lag 1 in the PACF plot, and the ACF should decay exponentially. ARMA(1,1): significant spikes at lag 1 in both ACF and PACF plots.

The magnitudes of these spikes should also be near the 0.8 and 0.6 values (which many of them are),

(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:

Yes they should match because p = 1 in so the spike should be in the first lag and equal in magnitude. I would have expected the magnitued to equal 0.6 but it is close. This could be due to maybe outliers or more difrencing needed. The results are not perfect but the do not imediately dismiss the model.

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 because the 0.7 is applied to the past term so it is autoregressive aka the value of Yt is influenced by Yt-1.

d = 0 (looks like there is no seasonal difrencing, that is none within one year but we sort of reset something every year which plays into D)

q = 1 (represents a moving average which looks like it could be the at and at-1 components)

P = 1 because this requires me to look back at t-12

D = 0 (I don't really see evidence of diffrencing but what would it even look like?)

Q = 0 (there is no At-12 term, no component of A looks seasonal)

So: ARIMA$(1,0,1)(1,0,0)_1 2$

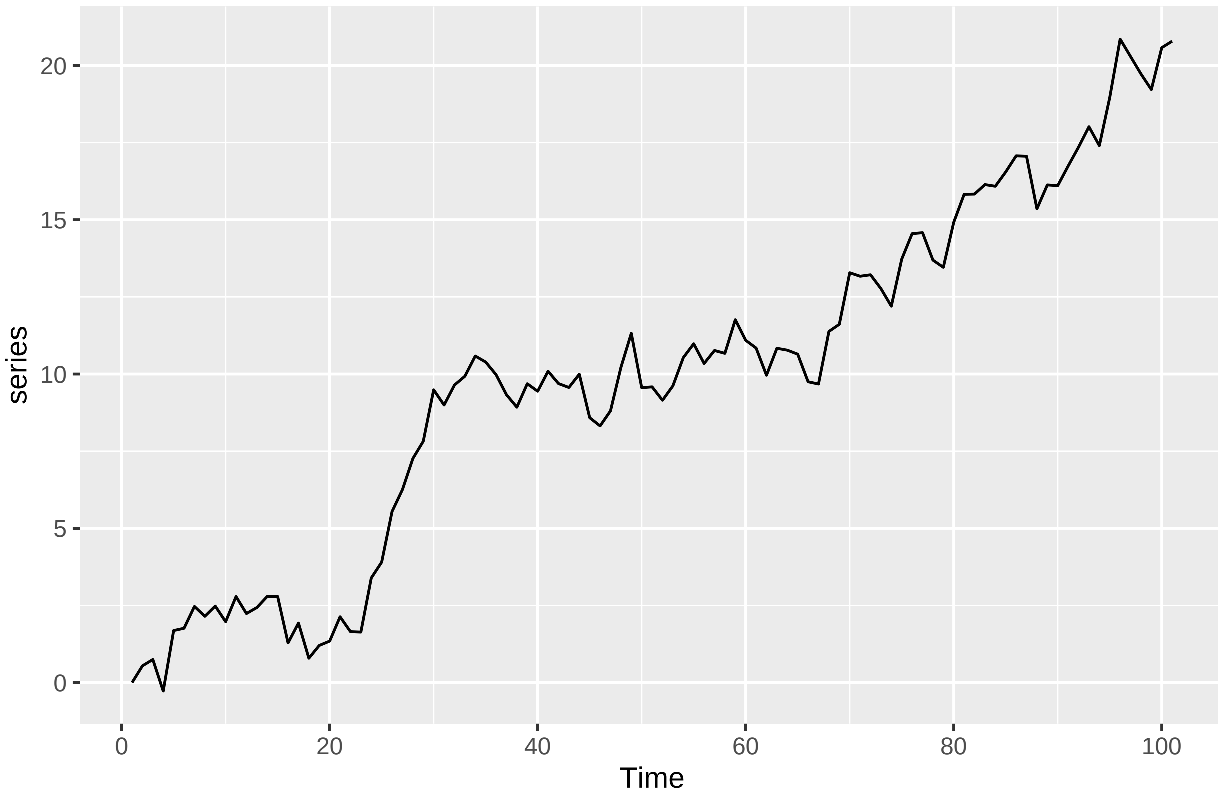
(b) Also from the equation what are the values of the parameters, i.e., model coefficients.

$\phi = 0.7$ $\Phi = -0.25$ $\theta = -0.1$

Q4

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
phi <- 0.8
theta <- 0.5
s <- 12
n <- 100

series <- ts(arima.sim(list(order = c(0, 1, 0), seasonal = list(order = c(1, 0, 0), period = s, phi = ph

autoplot(series)
```



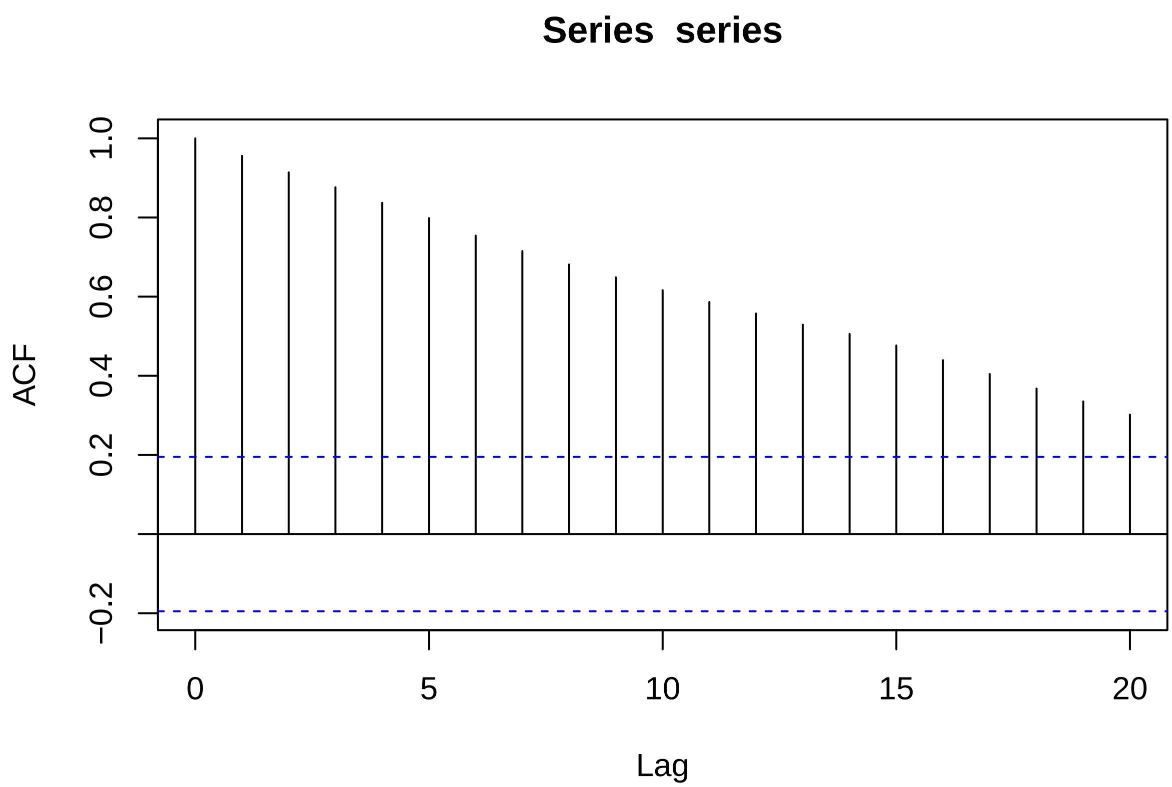
This does look seasonal to some degree; there is oscilation around 0 which makes me think there is some seasonality to it.

Q5

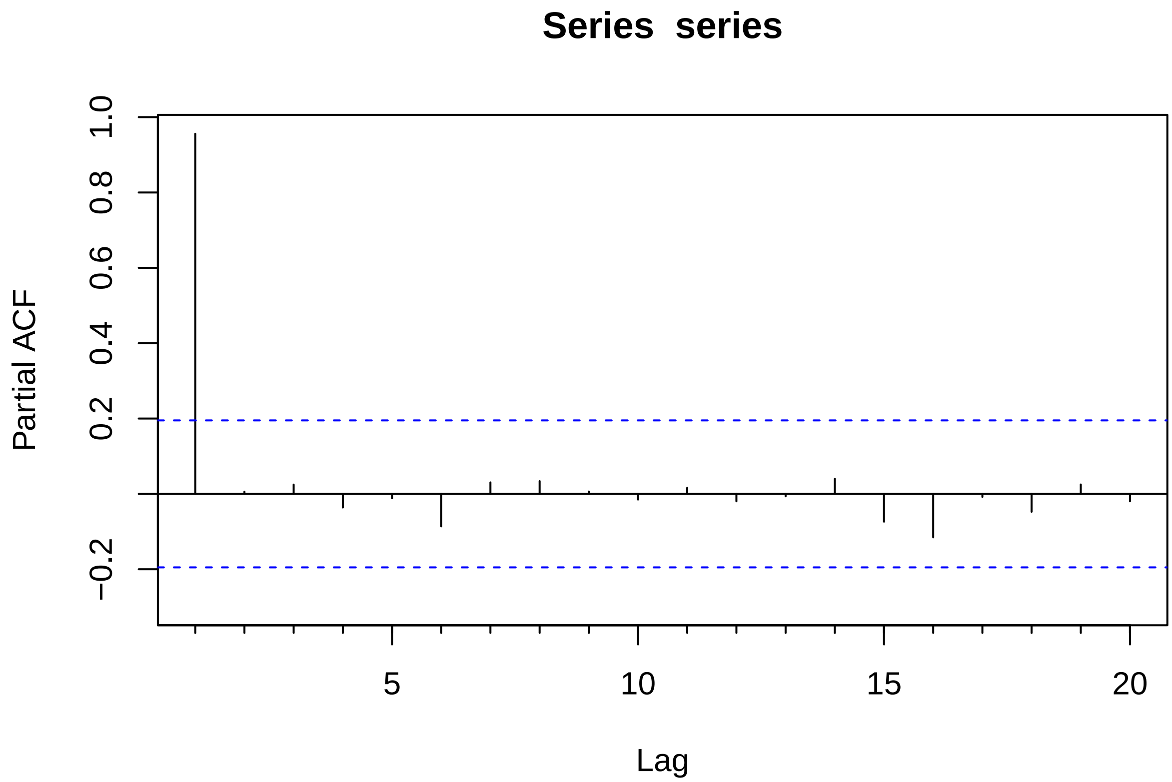
Plot ACF and PACF of the simulated series in Q4. Comment if the plots are well representing the model you simulated, i.e., would you be able to identify the order of both non-seasonal and seasonal components from the plots? Explain.

I think these ACF and PACF make sense. They are both showing first order (peaks around ~1) so it makes semse that our model has p = 1 and q =1. Per the last notes, The ACF and PACF from the seasonal adjusted series will help you specify components **p** and **q** of the ARIMA(p,d,q).

```r
acf(series)
```




```r
Pacf(series)
```


Series series
Partial ACF
1.0
0.8
0.6
0.4
0.2
−0.2
5
10
15
20
Lag