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

Assignment 6 - Due date 02/28/24

Yilun Zhu

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.

```
#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: For AR(2), the ACF plot will show decaying exponentially with time. The Pacf plot will show it's order and since p=2, in pacf, its significant covariants will cut off at lag 2.

- MA(1)

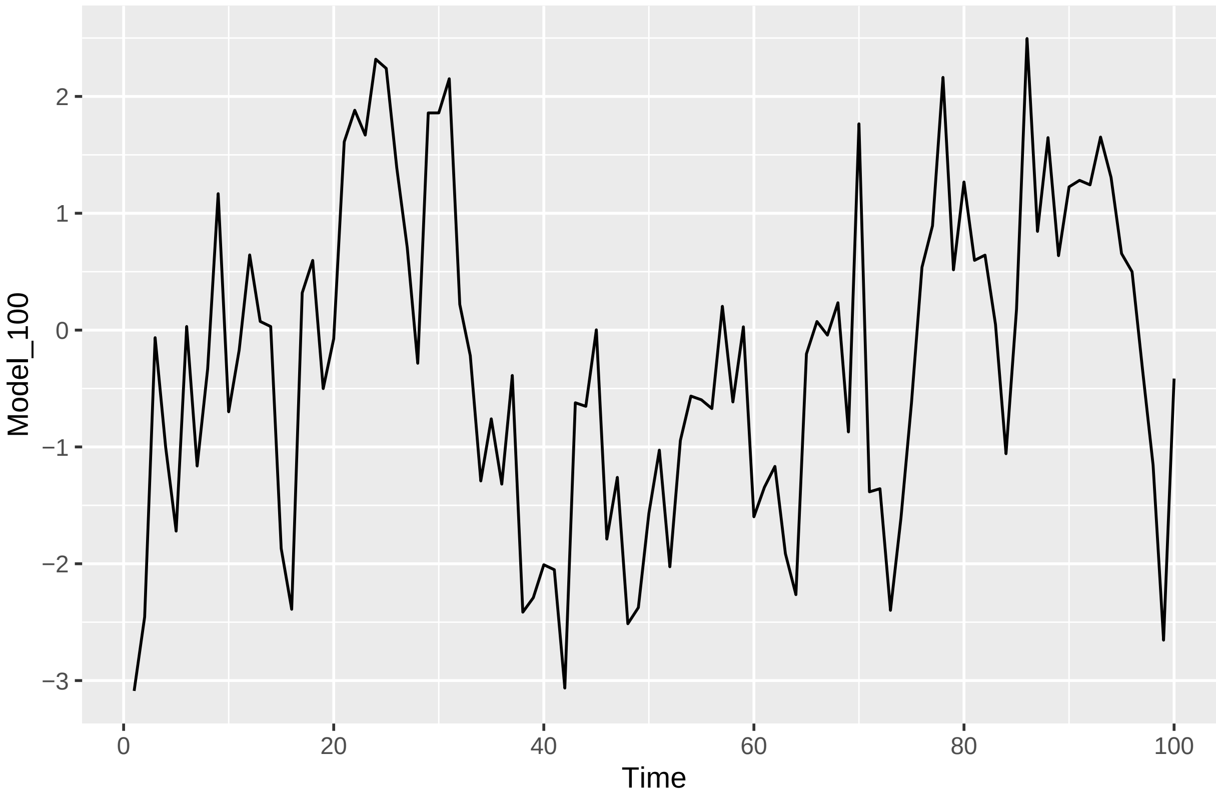
  Answer: For MA(1), the ACF plot will show it's order and since q=1, in acf, its significant autocovariances will cut off at lag 1. In Pacf, MA plot will decay exponentially with time.

Q2

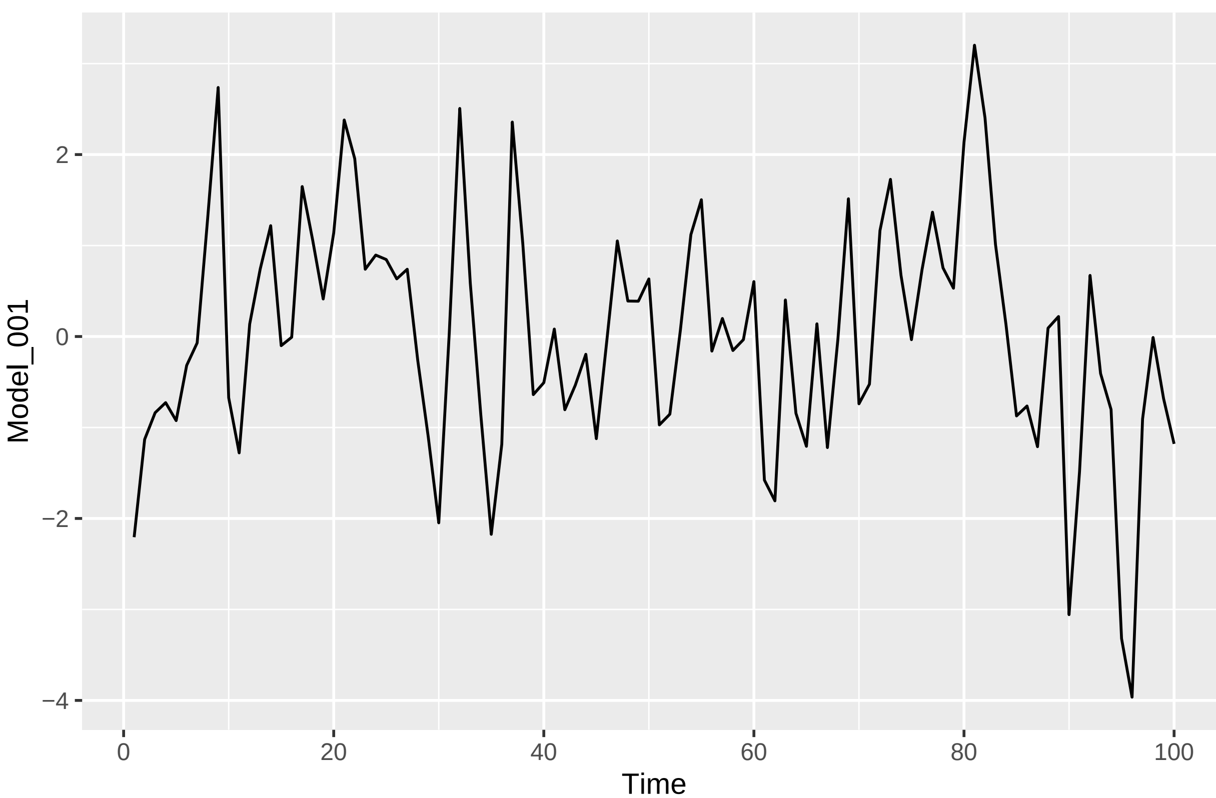
Recall that the non-seasonal ARIMA is described by three parameters ARIMA(p, d, q) where p is the order of the autoregressive component, d is the number of times the series need to be differenced to obtain stationarity and q is the order of the moving average component. If we don't need to difference 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.

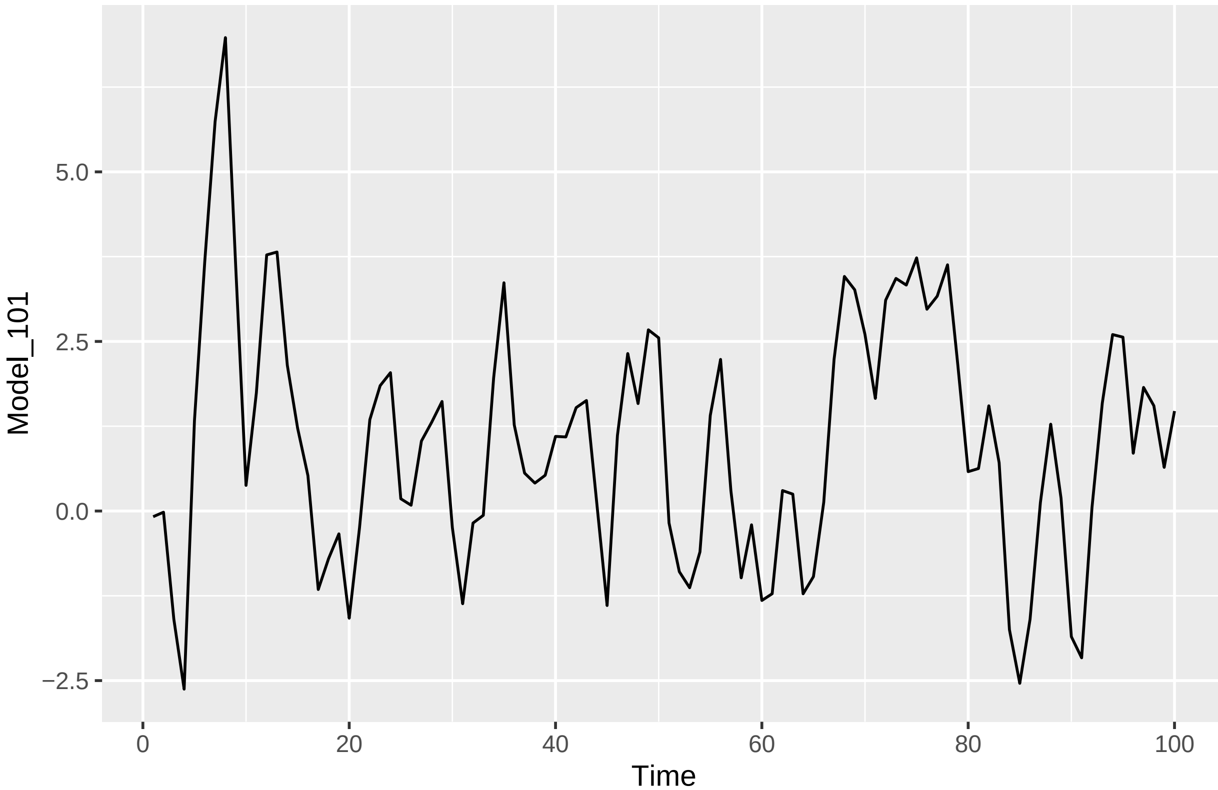
```r
Model_100 <- arima.sim(n = 100, list(order=c(1,0,0),ar = 0.6))
Model_001 <- arima.sim(n = 100, list(order=c(0,0,1),ma = 0.9))
Model_101 <- arima.sim(n = 100, list(order=c(1,0,1),ar = 0.6, ma = 0.9))

autoplot(Model_100)
```




```
autoplot(Model_001)
```



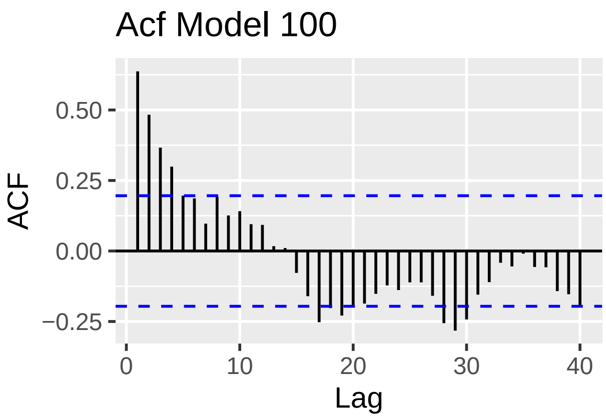
```r
autoplot(Model_101)
```

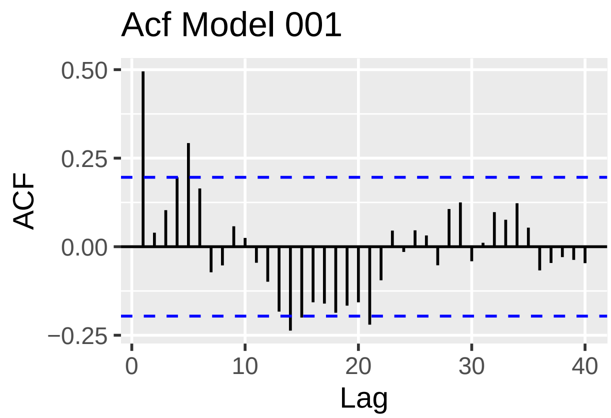


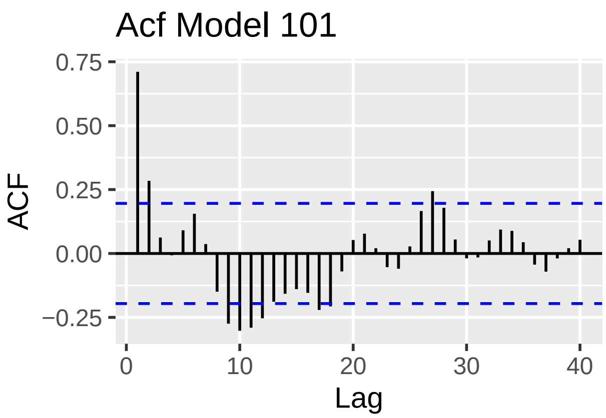
(b) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use `cowplot::plot_grid()`).

```r
plot_grid(
  autoplot(Acf(Model_100, lag = 40, plot=FALSE),
                 main = "Acf Model 100"),
  autoplot(Acf(Model_001, lag = 40, plot=FALSE),
                 main = "Acf Model 001"),
  autoplot(Acf(Model_101, lag = 40, plot=FALSE),
                 main = "Acf Model 101")
)
```

```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown parameters: ‘main‘
## Ignoring unknown parameters: ‘main‘
## Ignoring unknown parameters: ‘main‘
```



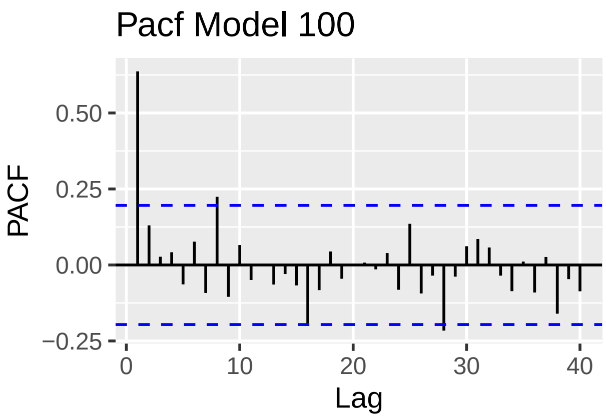


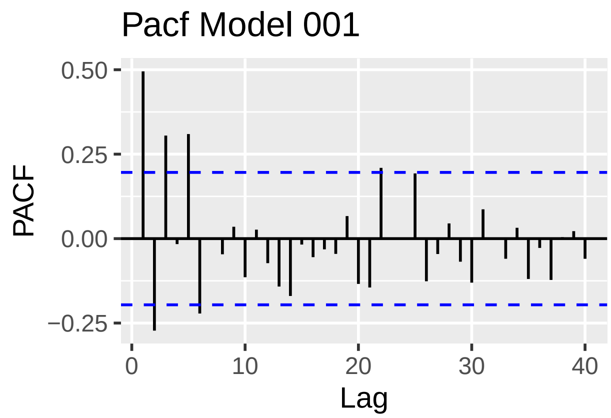


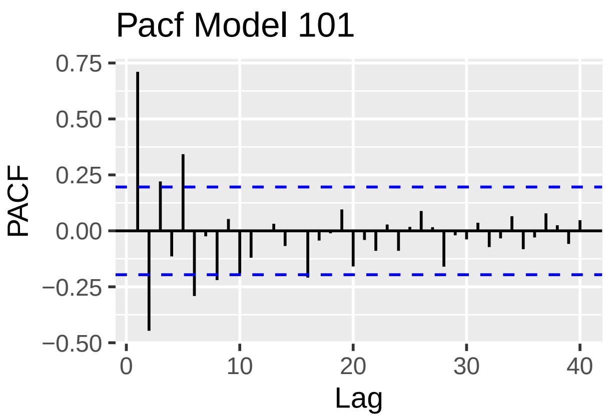
(c) Plot the sample PACF for each of these models in one window to facilitate comparison.

```
plot_grid(
  autoplot(Pacf(Model_100, lag = 40, plot=FALSE),
                 main = "Pacf Model 100"),
  autoplot(Pacf(Model_001, lag = 40, plot=FALSE),
                 main = "Pacf Model 001"),
  autoplot(Pacf(Model_101, lag = 40, plot=FALSE),
                 main = "Pacf Model 101")
)
```

```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown parameters: ‘main‘
## Ignoring unknown parameters: ‘main‘
## Ignoring unknown parameters: ‘main‘
```







(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.

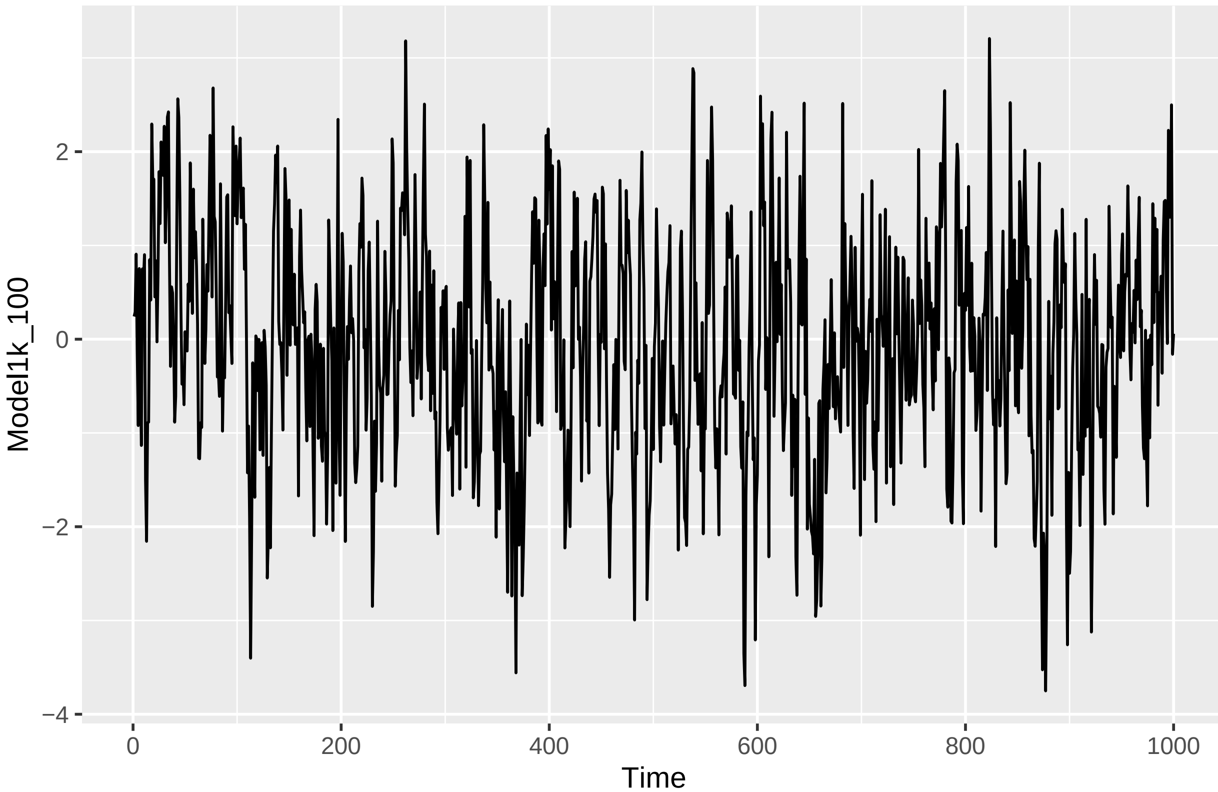
Answer: For AR, in acf, it should be decaying slowly and floating around 0 in large lag, in pacf, it should cut off at lag p. For MA, in acf, it cuts off at lag q and show decaying in pacf. For ARMA, it shows tailing off in both acf and pacf. (But I found that sometimes these random generated Model didn't cuts off at 1 as it was set AR1/MA1.)

(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?

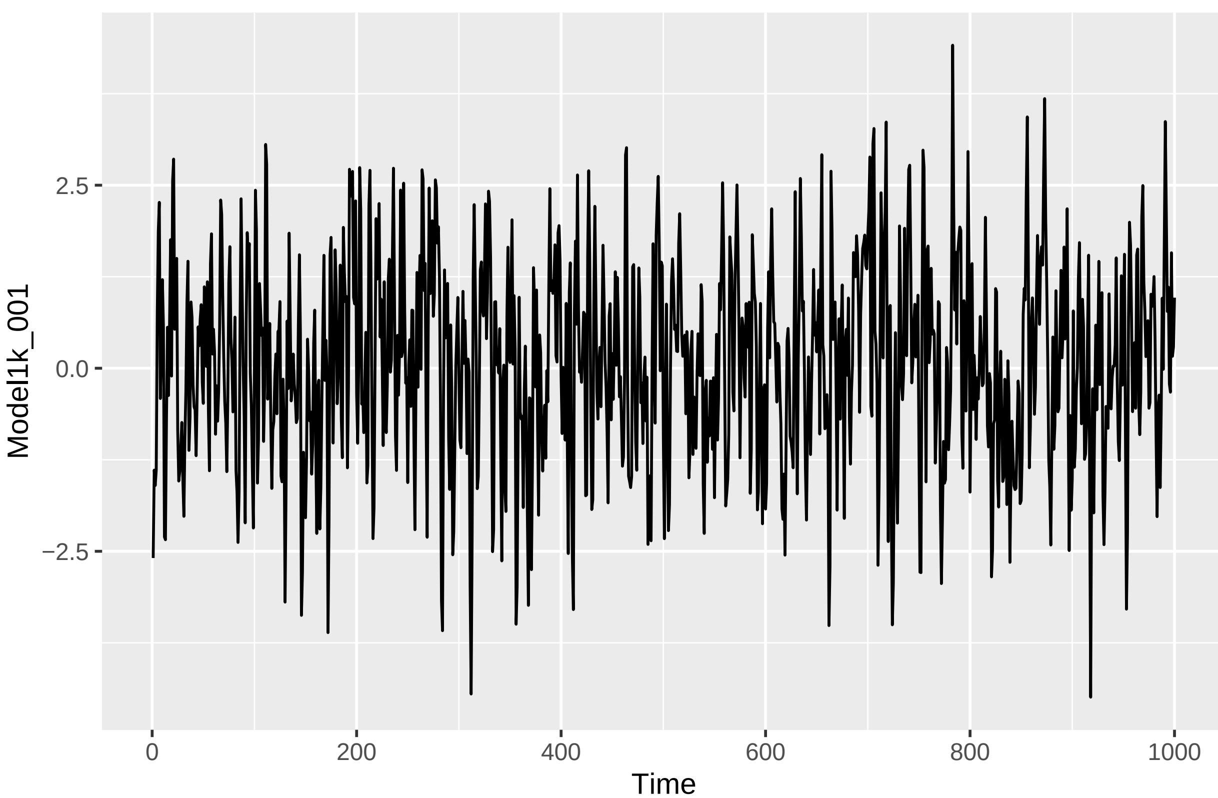
Answer: ϕ didn't match with AR(1) and ARMA(1,1) but is close in pacf. Because the regression model has the constant C and error term a_t, the outcome must be a value $= \phi +$ (C+a_t)/y_0 $> \phi$. So AR(1) should match with 0.6 but ARMA should have some difference since ARMA is combination of AR and MA.

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

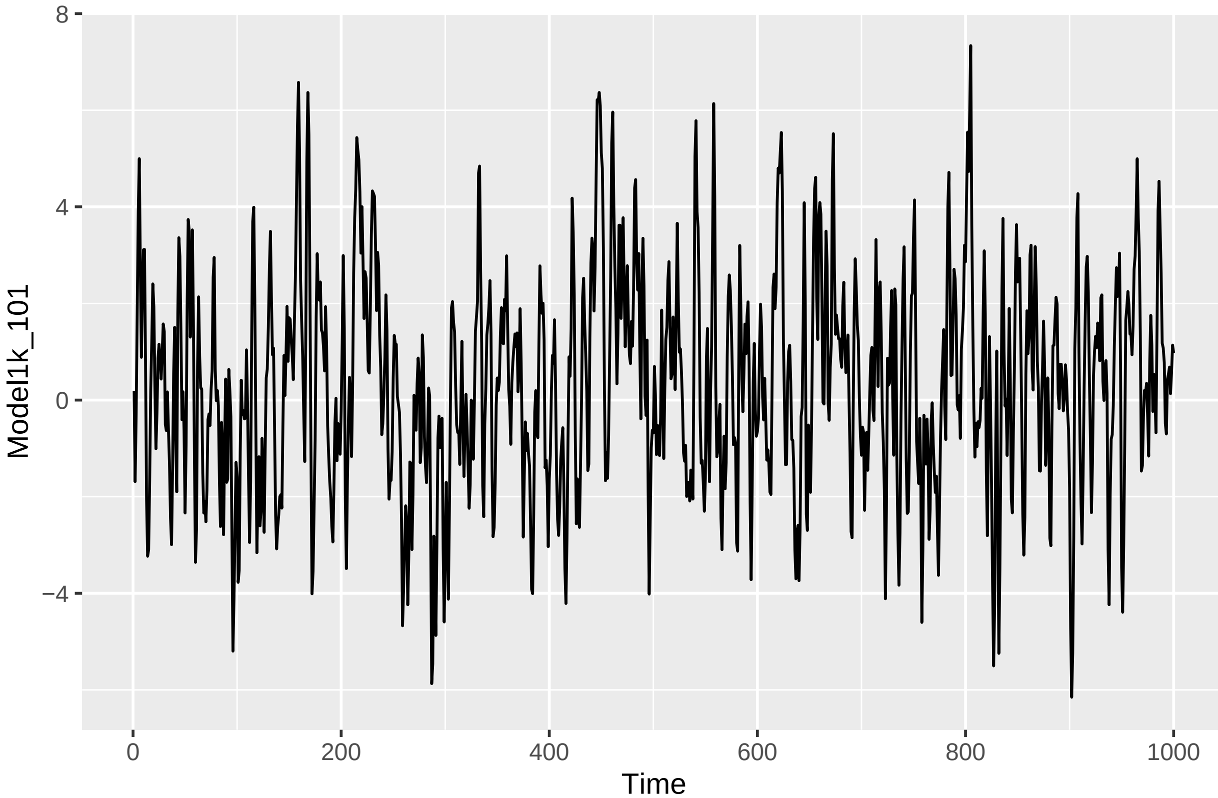
```
Model1k_100 <- arima.sim(n = 1000, list(order=c(1,0,0),ar = 0.6))
Model1k_001 <- arima.sim(n = 1000, list(order=c(0,0,1),ma = 0.9))
Model1k_101 <- arima.sim(n = 1000, list(order=c(1,0,1),ar = 0.6, ma = 0.9))

autoplot(Model1k_100)
```




```r
autoplot(Model1k_001)
```

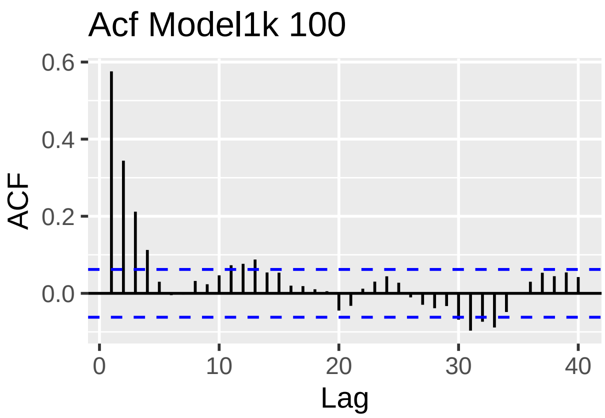


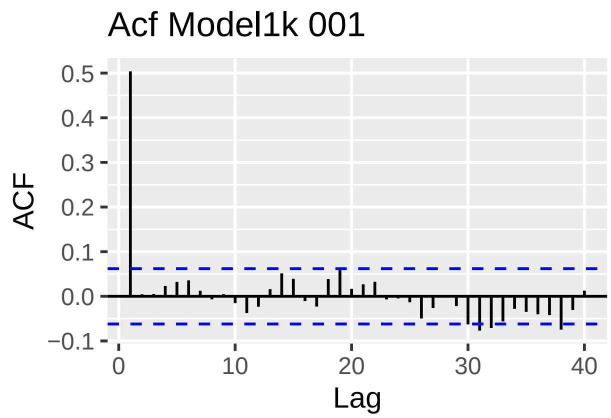
```r
autoplot(Model1k_101)
```

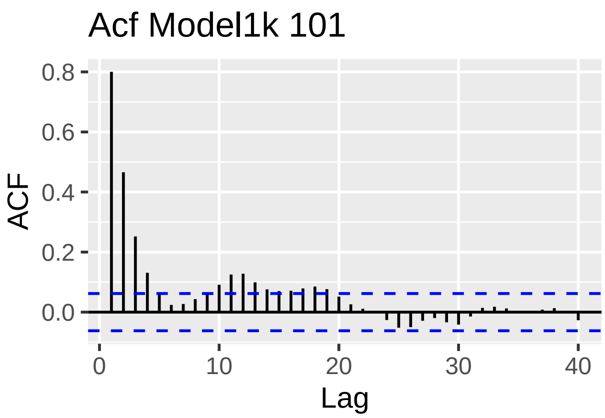



```r
plot_grid(
  autoplot(Acf(Model1k_100, lag = 40, plot=FALSE),
                main = "Acf Model1k 100"),
  autoplot(Acf(Model1k_001, lag = 40, plot=FALSE),
                main = "Acf Model1k 001"),
  autoplot(Acf(Model1k_101, lag = 40, plot=FALSE),
                main = "Acf Model1k 101")
)
```

```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown parameters: ‘main‘
## Ignoring unknown parameters: ‘main‘
## Ignoring unknown parameters: ‘main‘
```

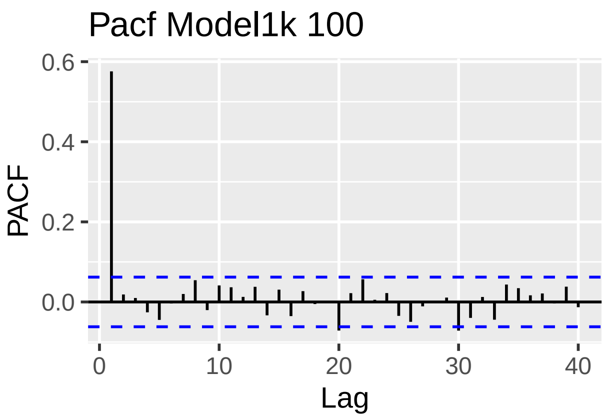


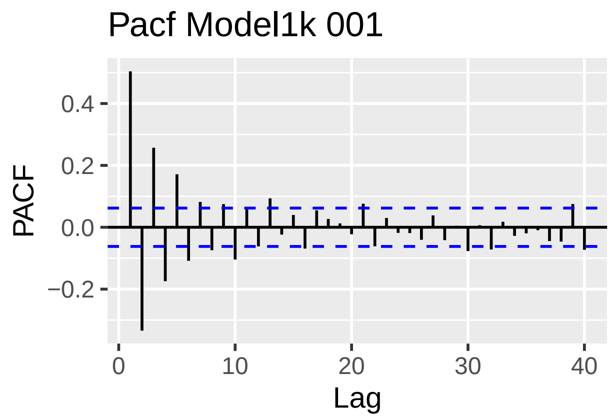


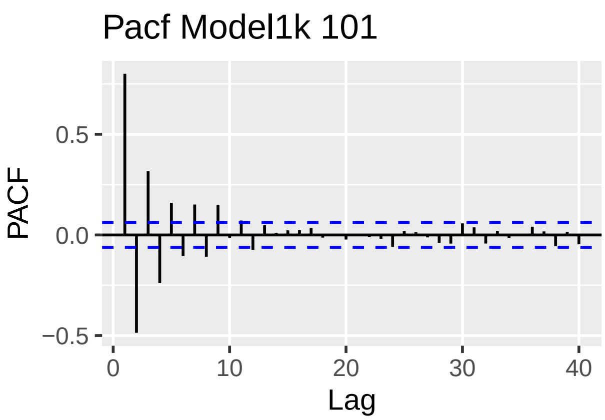



```
plot_grid(
  autoplot(Pacf(Model1k_100, lag = 40, plot=FALSE),
                   main = "Pacf Model1k 100"),
  autoplot(Pacf(Model1k_001, lag = 40, plot=FALSE),
                   main = "Pacf Model1k 001"),
  autoplot(Pacf(Model1k_101, lag = 40, plot=FALSE),
                   main = "Pacf Model1k 101")
)
```

```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown parameters: ‘main‘
## Ignoring unknown parameters: ‘main‘
## Ignoring unknown parameters: ‘main‘
```







>

Answer: When observation increases to 1000, the outcome becomes much more clear. The cuts off and tails off of AR and MA is easy to identify. Acf of ARMA is similar to Acf of AR and Pacf of ARMA is similar to Pacf of MA. The value of first lag of AR(1,0) in Pacf is almost equal to ϕ=0.6

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.

Answer: This is the ARIMA$(1, 0, 1)(1, 0, 0)_1$2.

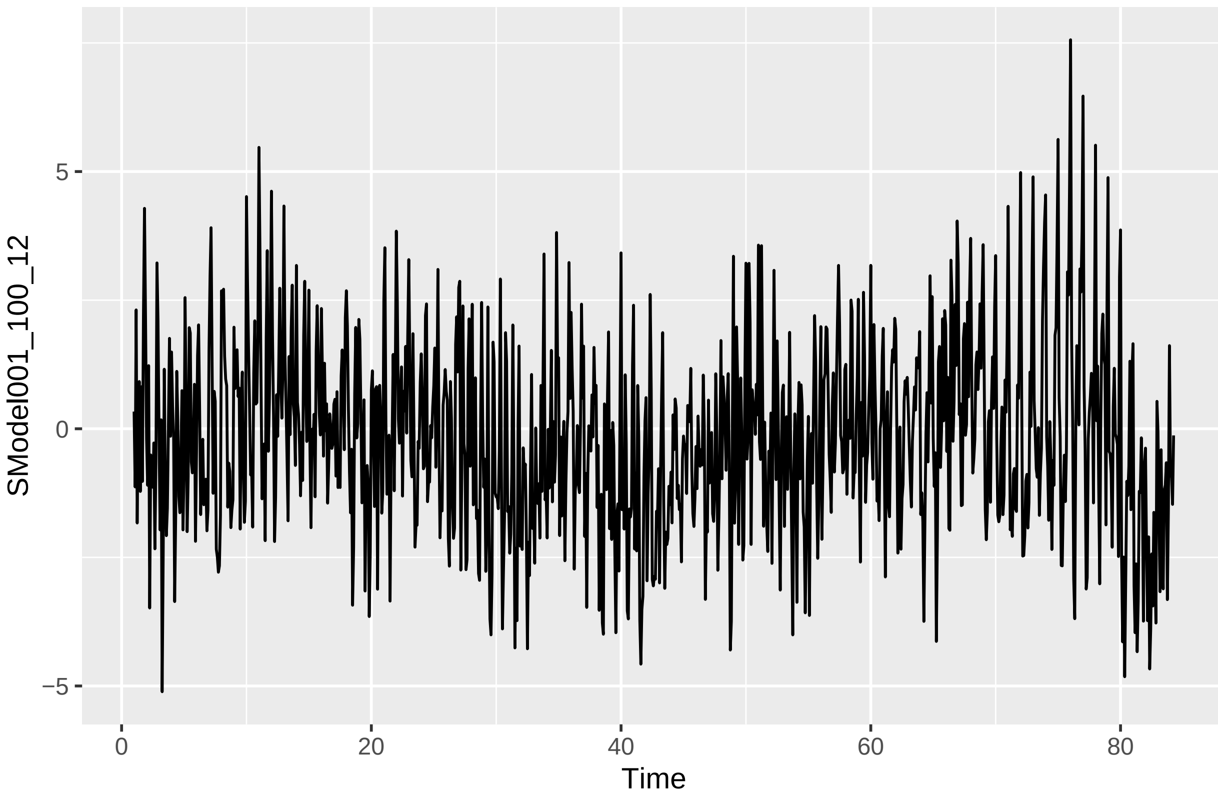
(b) Also from the equation what are the values of the parameters, i.e., model coefficients.

Answer: 0.7 is the ϕ_1, 0.1 is the θ_1, 12 is the seasonal period, -0.25 is the ϕ_{12}.

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?

```
SModel001_100_12 <- sim_sarima(n = 1000, model = list(ma = 0.5, sar = 0.8, nseasons=12))
#Weird, I tried many times but sim_arima only create a numeric data
SModel001_100_12 <- ts(SModel001_100_12, frequency = 12)
autoplot(SModel001_100_12)
```



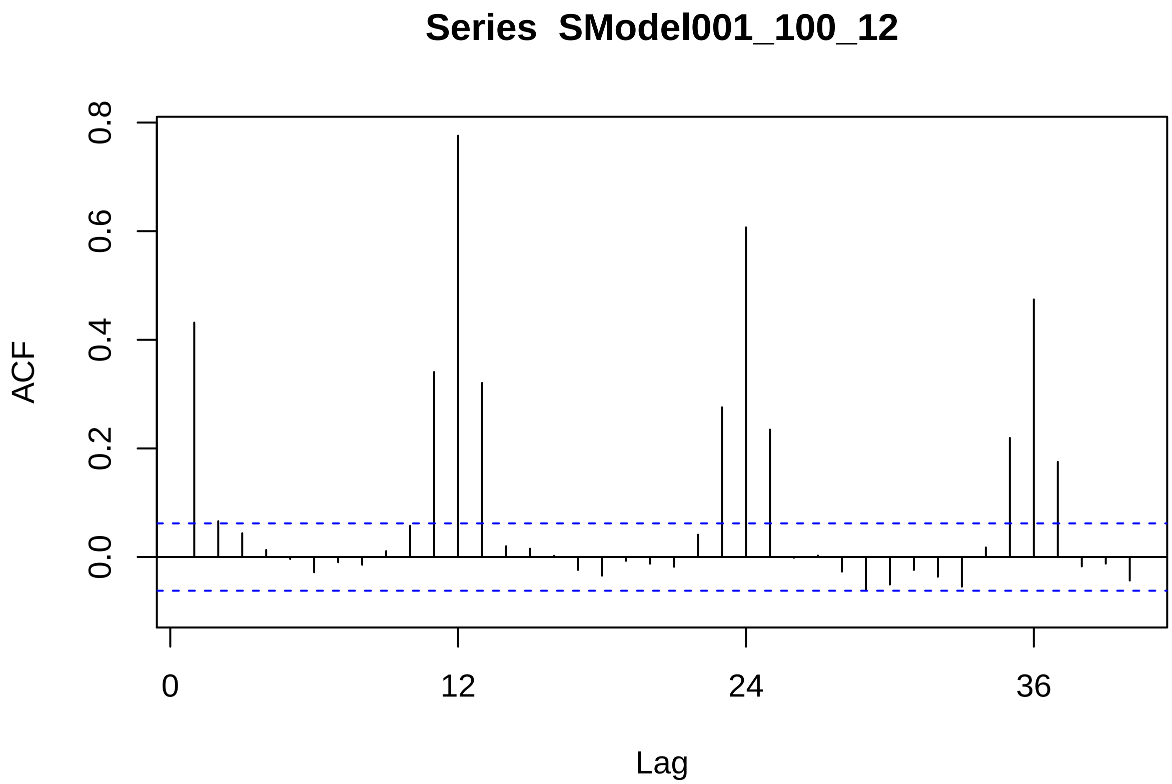
>

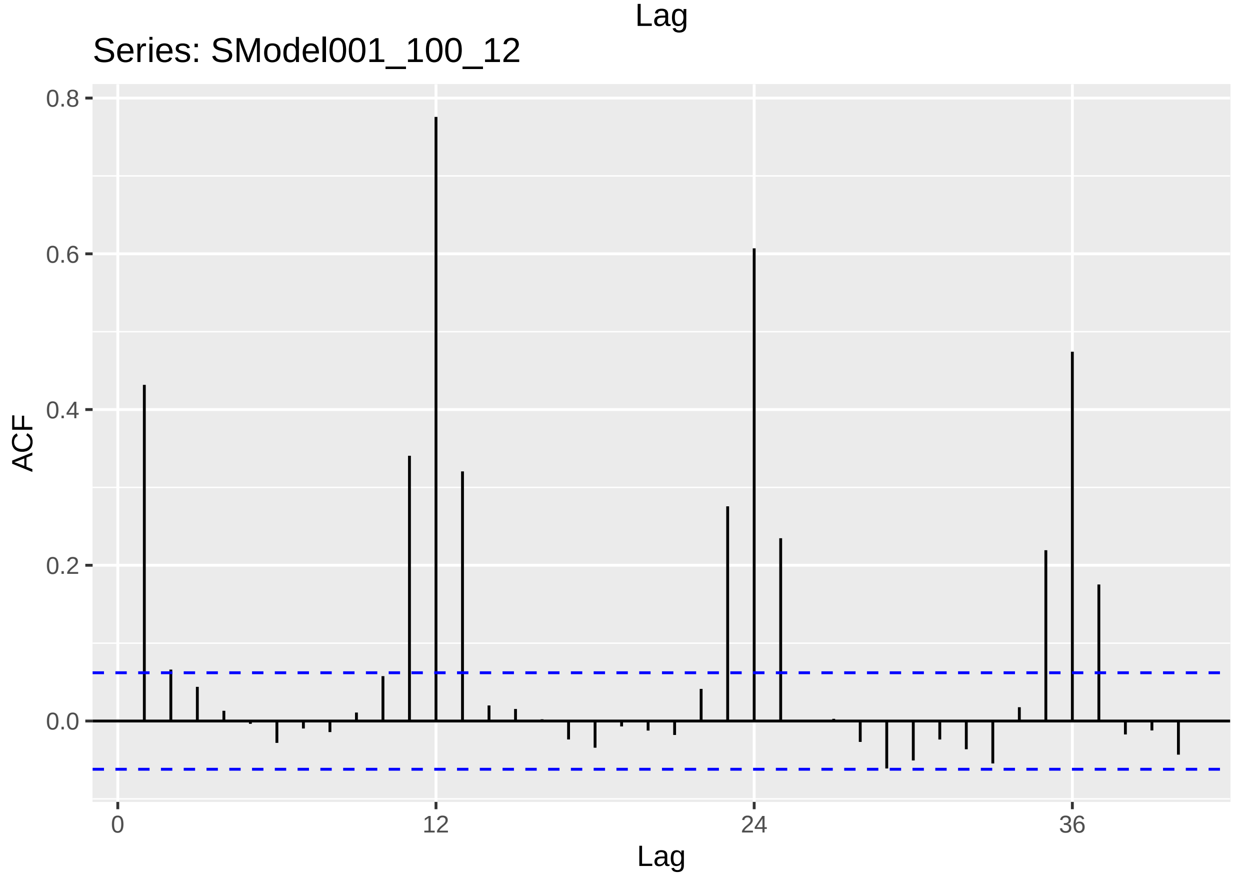
Answer: The plot didn't show seasonality.

Q5

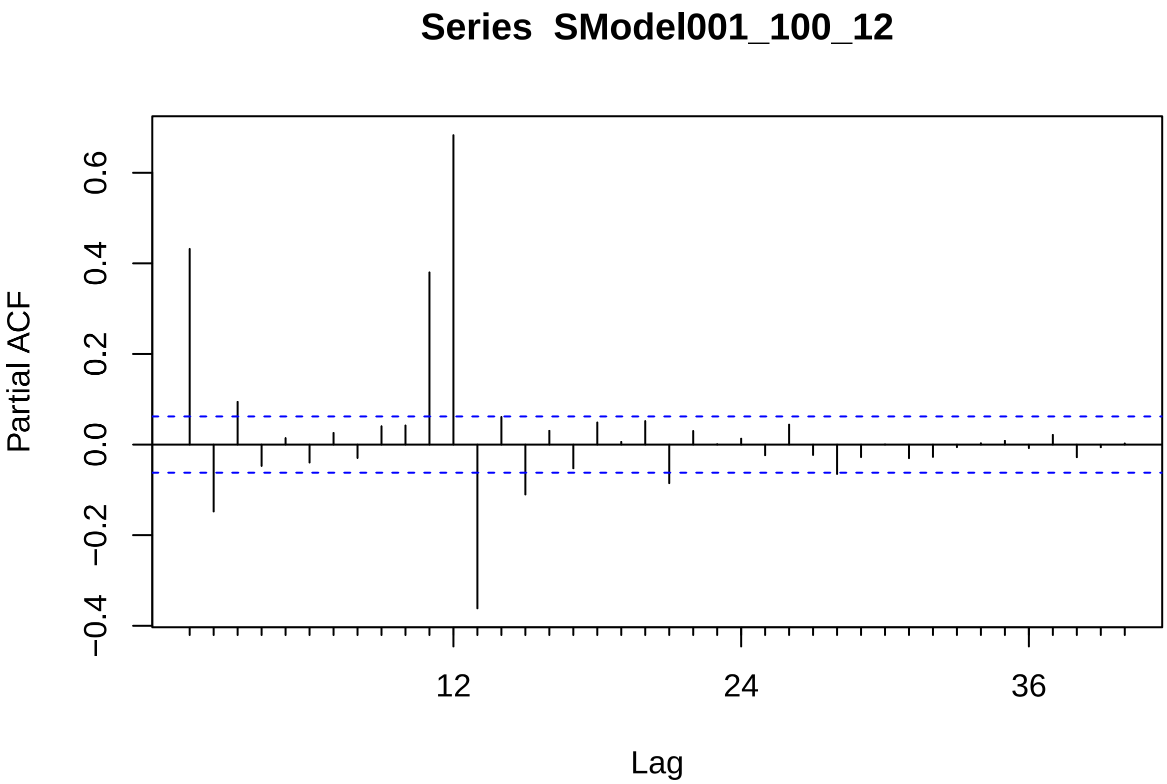
Plot ACF and PACF of the simulated series in Q4. Comment if the plots are well representing the model you simulated, i.e., would you be able to identify the order of both non-seasonal and seasonal components from the plots? Explain.

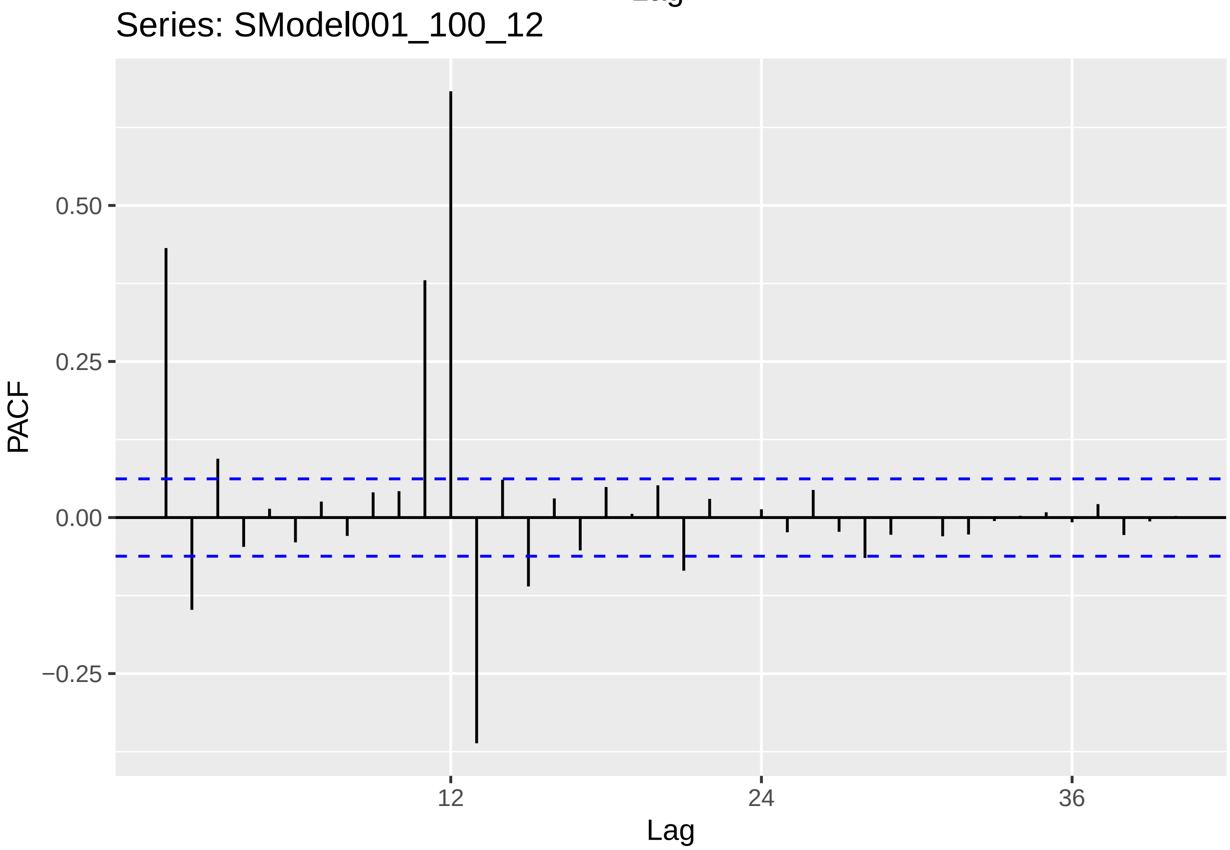
```
autoplot(Acf(SModel001_100_12, lag = 40, plot = TRUE))
```


Series SModel001_100_12
ACF
0.0
0.2
0.4
0.6
0.8
0
12
24
36
Lag


Series: SModel001_100_12
ACF
0.0
0.2
0.4
0.6
0.8
0
12
24
36
Lag

```r
autoplot(Pacf(SModel001_100_12, lag = 40, plot = TRUE))
```





> Answer: In acf, there is multiple spike after lag 12, also there is single spike after 12 in pacf, so it

has a SAR process. In acf non-seasonal part, there is a cut off but there isn't a obvious tail off in pacf non-seasonal part (since it gradually decreases and at lag 6 it starts increasing gradually), so it's little bit confusing when identifying the MA process.