

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

Assignment 6 - Due date 02/27/25

Jingze Dai

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
list_of_lib <- c(
  "forecast", "tseries", "dplyr", "sarima", "ggplot2", "cowplot")
for (i in list_of_lib){
  library(i, character.only = TRUE)
}
```

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: Autoregressive models have exponentially decaying ACF plots over time, while the PACF plot shows the order of the model. In this case, the AR model has an order of 2, thus the cut off point (when the PACF value becomes not significant) will be located at lag = 2.

- MA(1)

Answer: Moving average models have exponentially decaying PACF plots over time, while the ACF plot shows the order of the model. In this case, the MA model has an order of 1, thus the cut off point (when the ACF value becomes not significant) will be located at lag = 1.

Q2

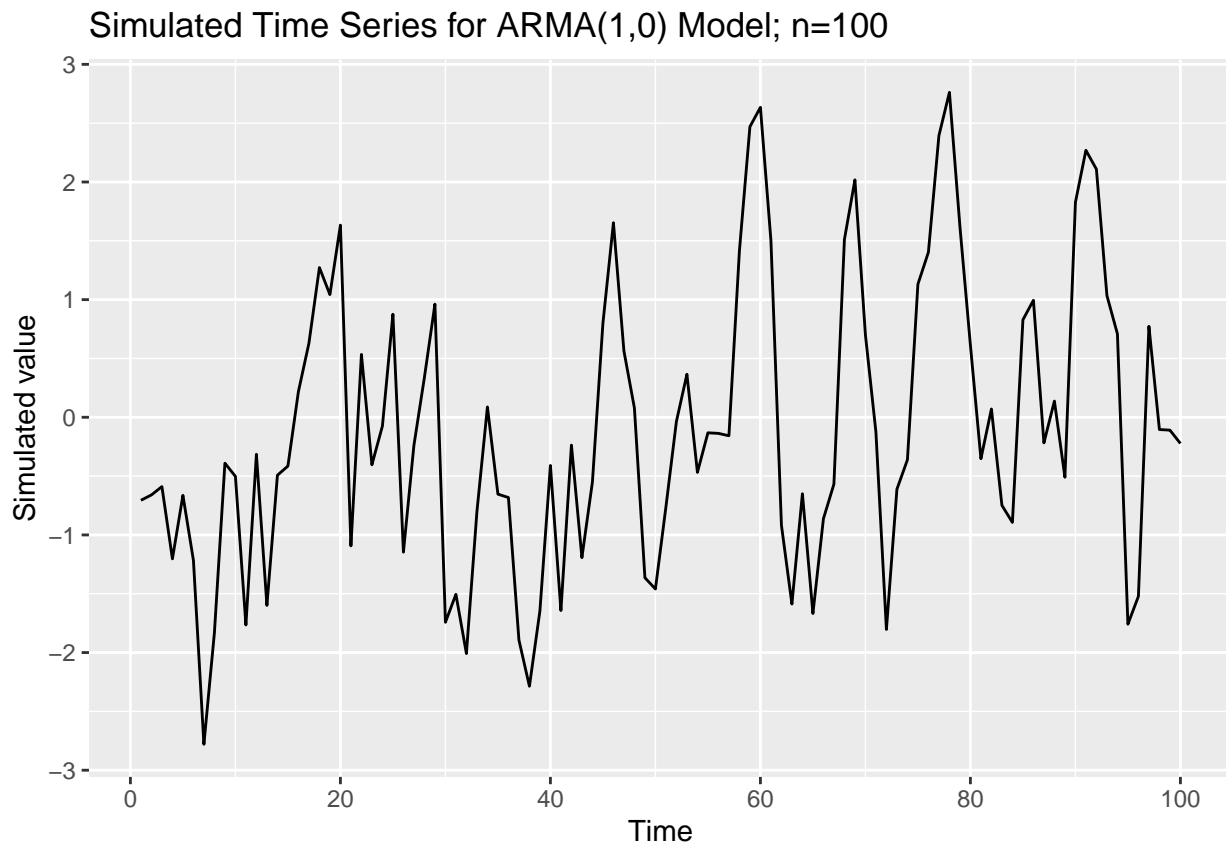
Recall that the non-seasonal ARIMA is described by three parameters $ARIMA(p, d, q)$ where p is the order of the autoregressive component, d is the number of times the series need to be differenced to obtain stationarity and q is the order of the moving average component. If we don't need to difference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

```
# setting seed for reproducibility
set.seed(101)

# ARMA(1,0)
model_1 <- arma.sim(list(order = c(1,0,0), ar = 0.6), n = 100)

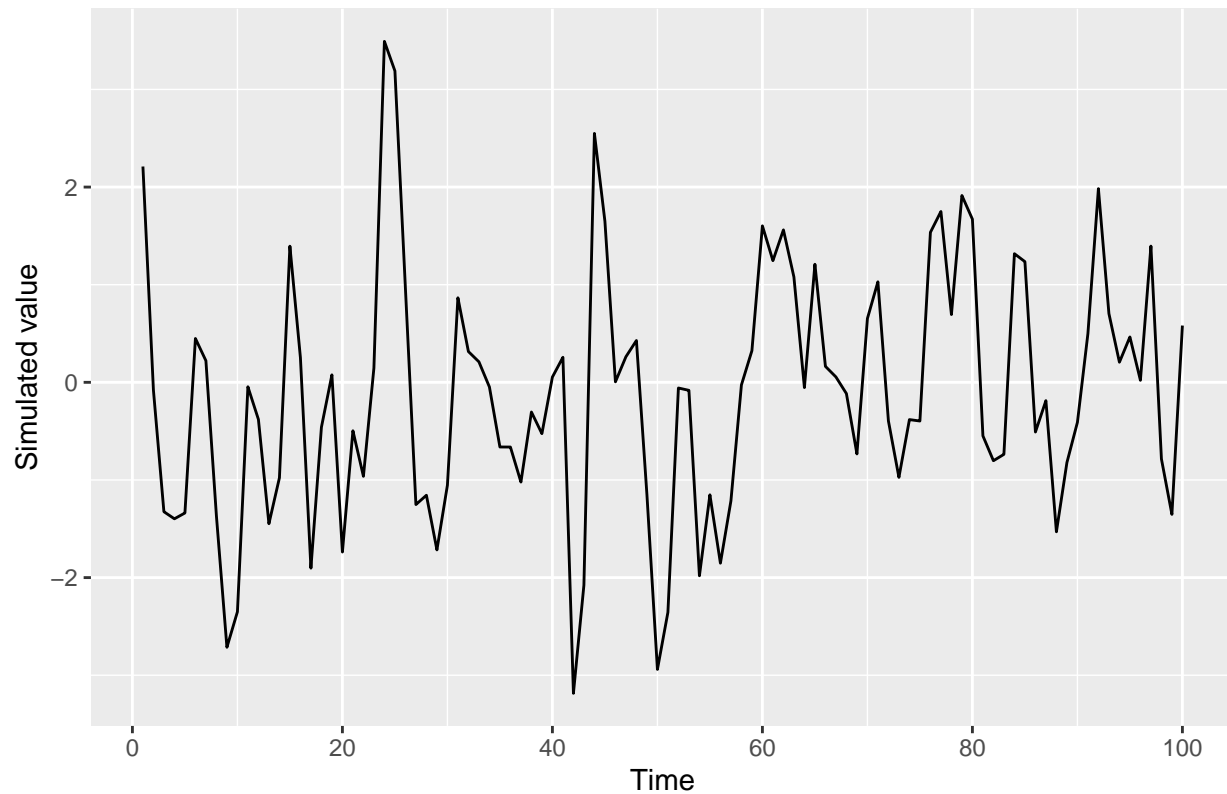
autoplot(model_1)+
  ylab("Simulated value") +
  ggtitle("Simulated Time Series for ARMA(1,0) Model; n=100")
```



```
# ARMA(0,1)
model_2 <- arima.sim(list(order = c(0,0,1), ma = 0.9), n = 100)

autoplot(model_2)+
  ylab("Simulated value") +
  ggtitle("Simulated Time Series for ARMA(0,1) Model; n=100")
```

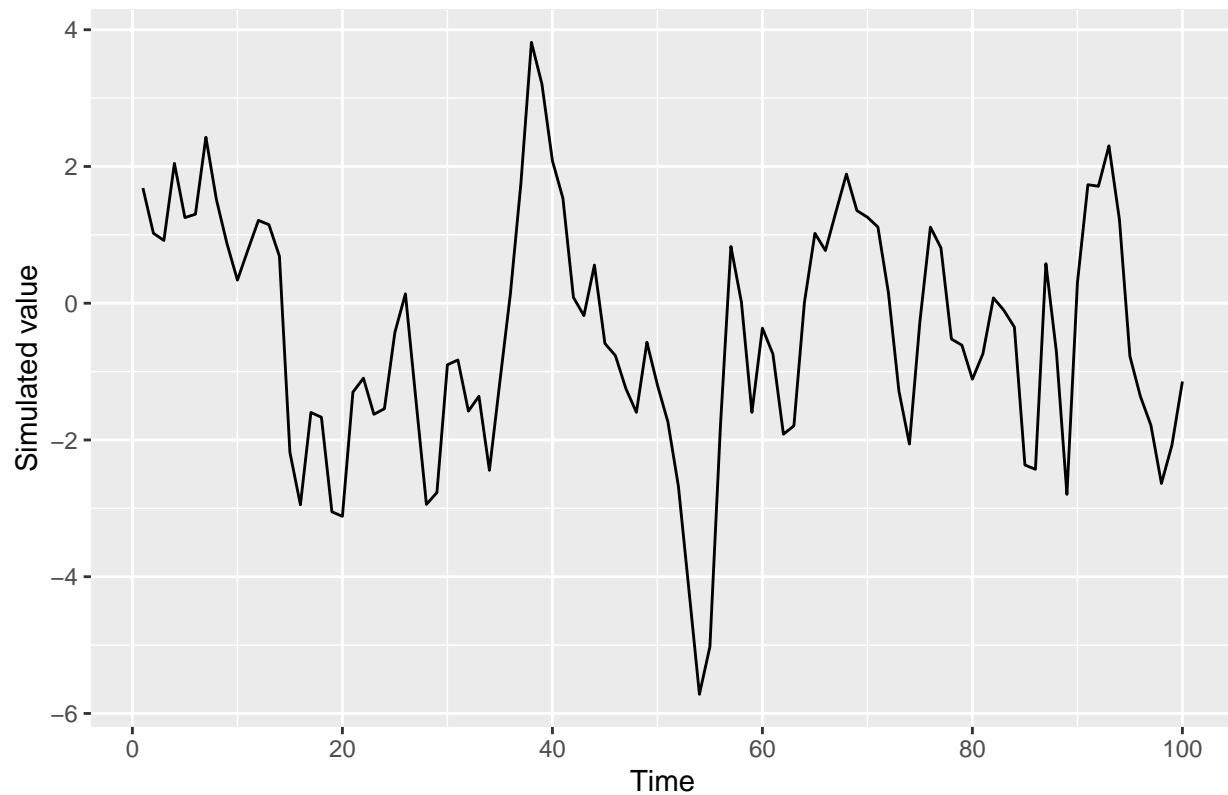
Simulated Time Series for ARMA(0,1) Model; n=100



```
# ARMA(1,1)
model_3 <- arima.sim(list(order = c(1,0,1), ar = 0.6, ma = 0.9), n = 100)

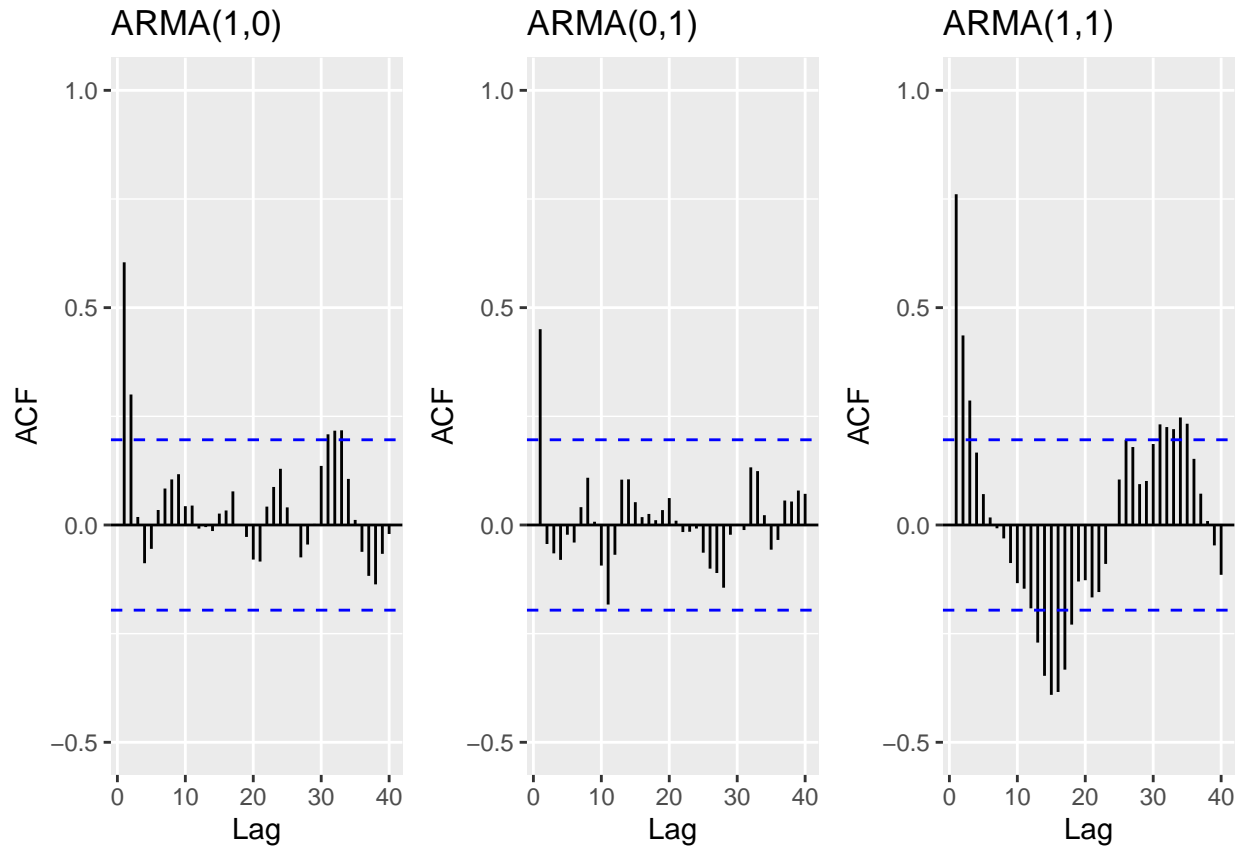
autoplot(model_3)+
  ylab("Simulated value") +
  ggtitle("Simulated Time Series for ARMA(1,1) Model; n=100")
```

Simulated Time Series for ARMA(1,1) Model; n=100



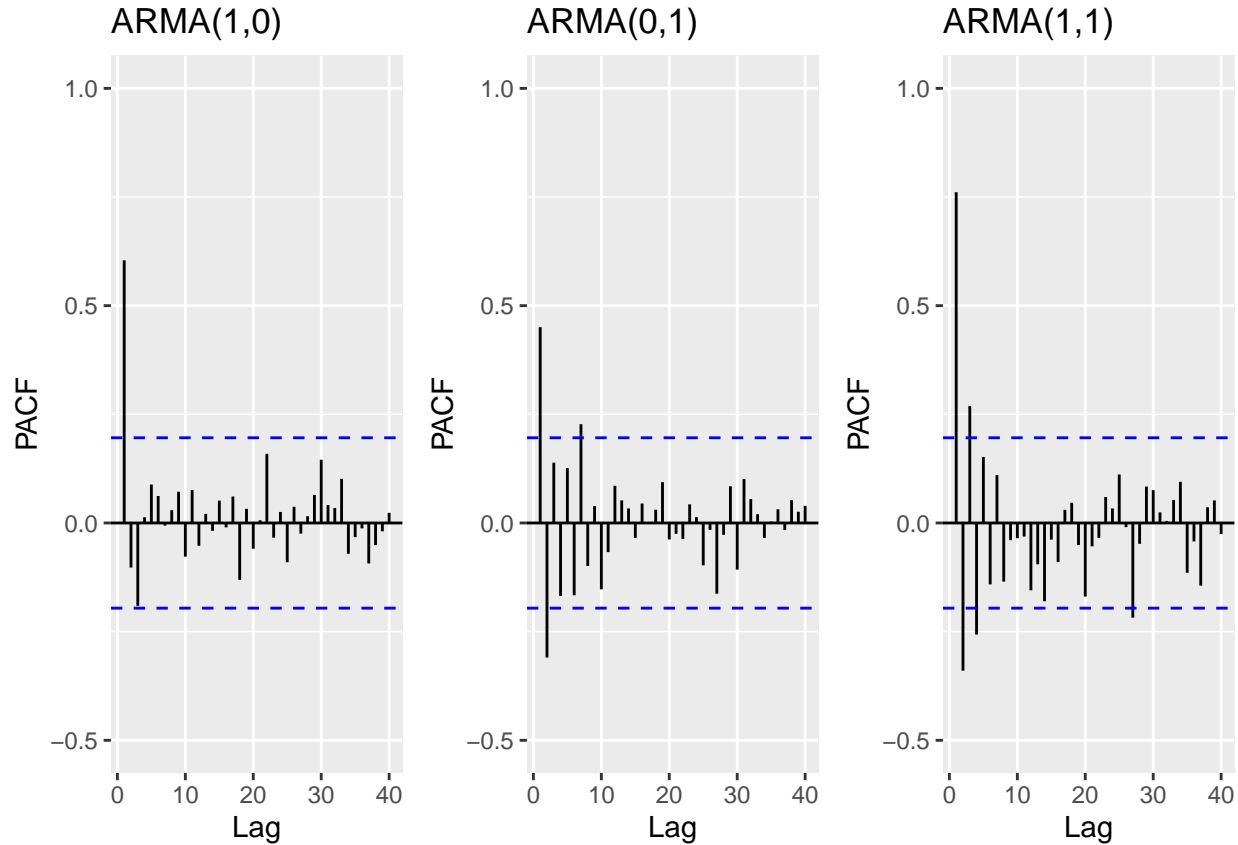
(b) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use `cowplot::plot_grid()`).

```
plot_grid(autoplot(Acf(model_1, lag=40, plot=FALSE),
  ylim=c(-0.5, 1), main="ARMA(1,0)"),
  autoplot(Acf(model_2, lag=40, plot=FALSE),
  ylim=c(-0.5, 1), main="ARMA(0,1)"),
  autoplot(Acf(model_3, lag=40, plot=FALSE),
  ylim=c(-0.5, 1), main="ARMA(1,1)"),
  nrow=1)
```



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

```
plot_grid(autoplot(pacf(model_1, lag=40, plot=FALSE),
  ylim=c(-0.5, 1), main="ARMA(1,0)"),
  autoplot(pacf(model_2, lag=40, plot=FALSE),
  ylim=c(-0.5, 1), main="ARMA(0,1)"),
  autoplot(pacf(model_3, lag=40, plot=FALSE),
  ylim=c(-0.5, 1), main="ARMA(1,1)"),
  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 to identify them correctly? Explain your answer.

Answer: If I were given the plots, I would predict the following: - For ARMA(1,0), since both ACF and PACF plots show a mix of cut-off point and exponentially decaying pattern (although it is not obvious in the PACF plot), I would identify them as ARMA(1,2) - For ARMA(0,1), the ACF plot has a cut-off point at lag=1, and PACF has exponential decay, thus I would identify it as MA(1). - For ARMA(1,1), it is obvious that it is ARMA due to the combination of the both models, however it is difficult to determine the order of the model, and by observing the cut-off points I would describe it as ARMA(1,2).

Thus, I only identified one out of the three models. This inaccuracy might be resulted from the limited sampling size of 100. The small sampling size, combined with the interaction of overlapping AR and MA models, makes it difficult to identify the cut-off point accurately.

- (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: For ARMA(1,0), the PACF at lag=1 is exactly 0.6. This matches with the ϕ value at 0.6 because it is only models an AR model with the coefficient being 0.6. For ARMA(1,1), the PACF at lag=1 is around 0.75, larger than the theoretical set value of ϕ at 0.6. This is because the ϕ value only controls the autoregressive model, but not the moving average model. Due to the overlapping interactions of both models, the PACF value for ARMA(1,1) may not match with the ϕ value.

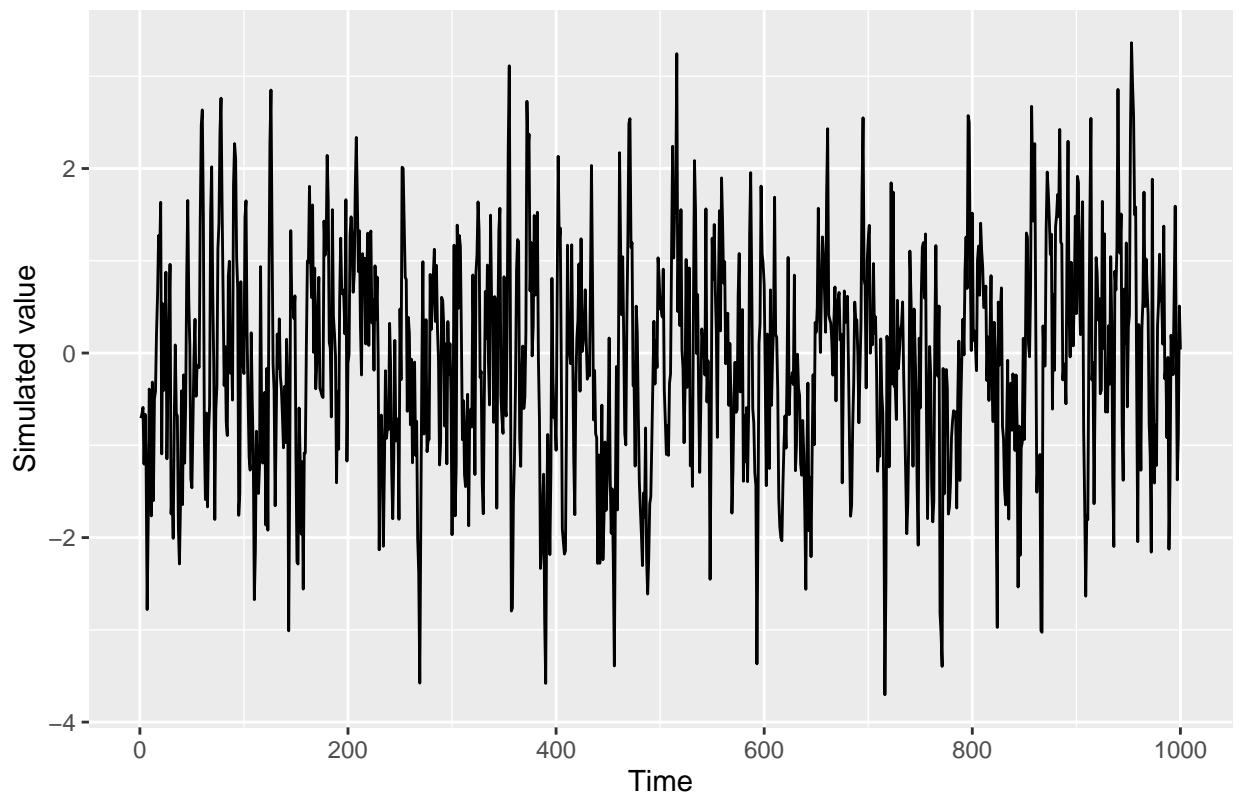
(f) Increase number of observations to $n = 1000$ and repeat parts (b)-(e).

```
# setting seed for reproducibility
set.seed(101)

# for n=1000
# ARMA(1,0)
model_1b <- arima.sim(list(order = c(1,0,0), ar = 0.6), n = 1000)

autoplot(model_1b)+
  ylab("Simulated value") +
  ggtitle("Simulated Time Series for ARMA(1,0) Model; n=1000")
```

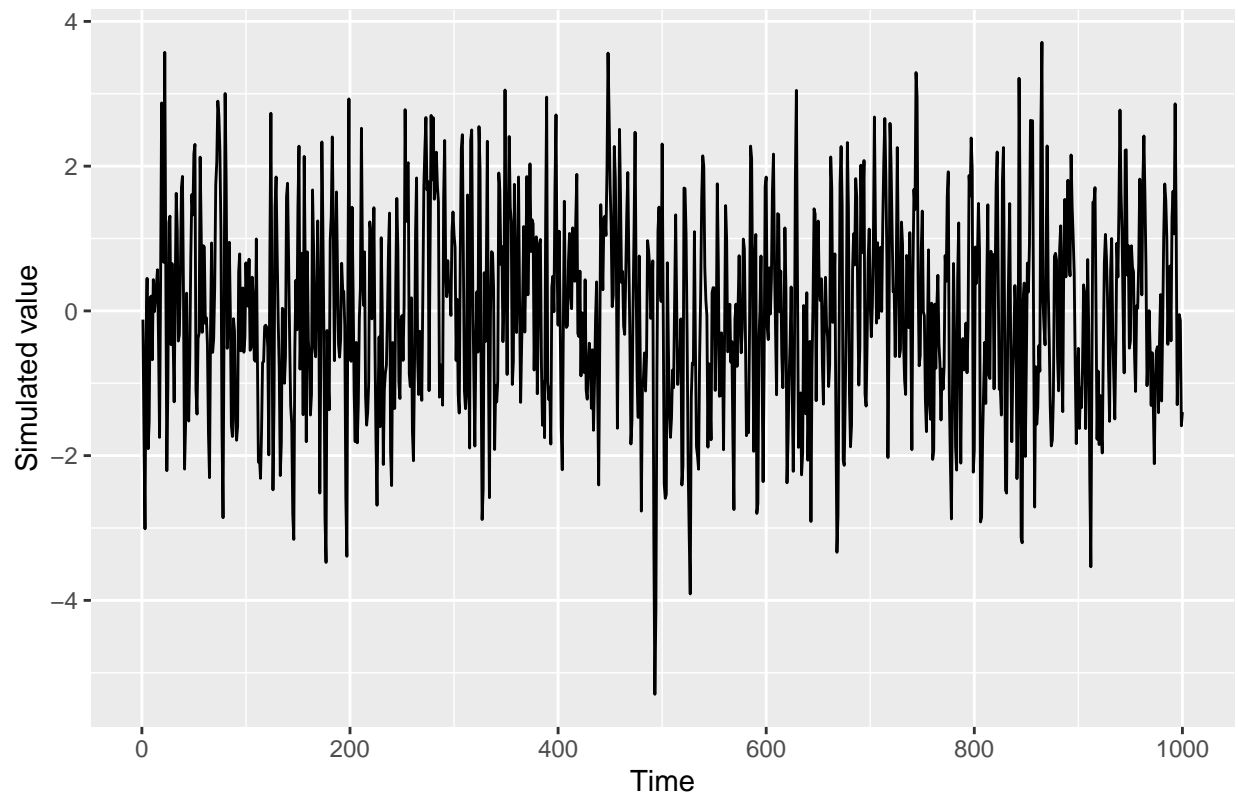
Simulated Time Series for ARMA(1,0) Model; n=1000



```
# ARMA(0,1)
model_2b <- arima.sim(list(order = c(0,0,1), ma = 0.9), n = 1000)

autoplot(model_2b)+
  ylab("Simulated value") +
  ggtitle("Simulated Time Series for ARMA(0,1) Model; n=1000")
```

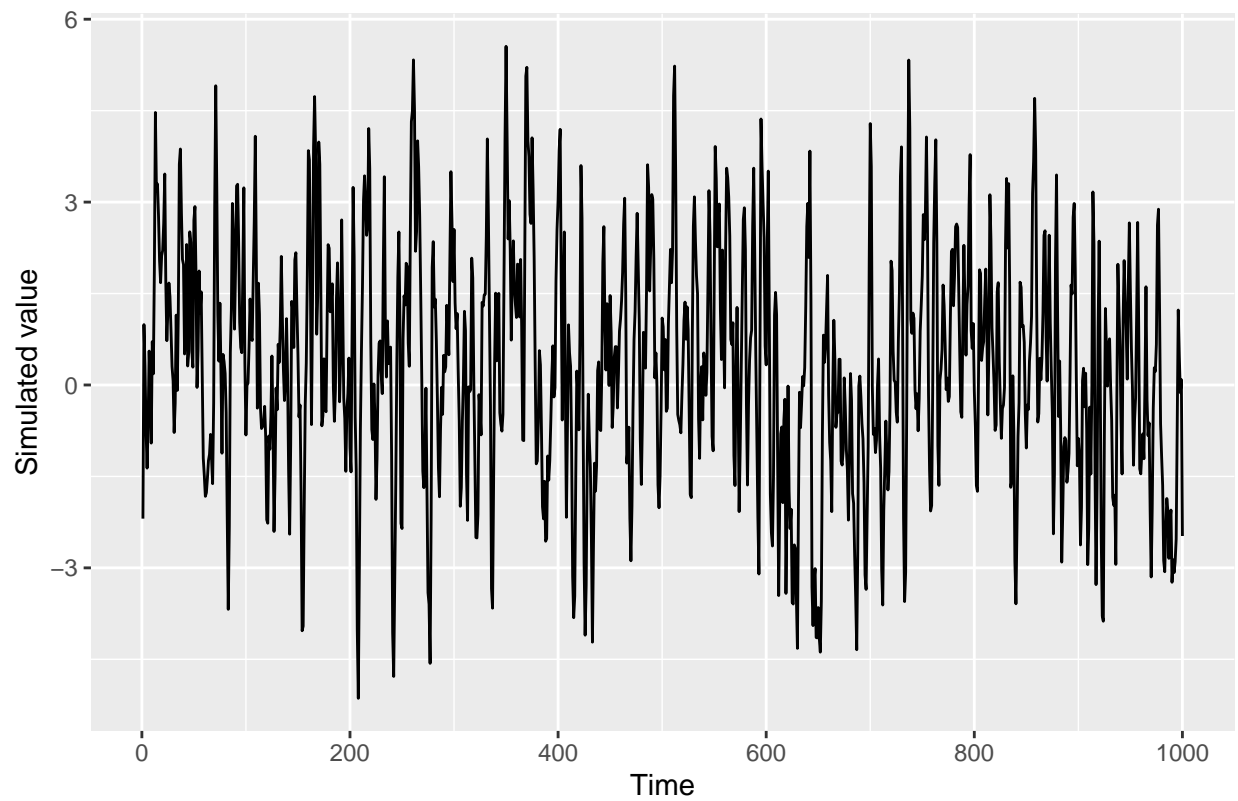
Simulated Time Series for ARMA(0,1) Model; n=1000



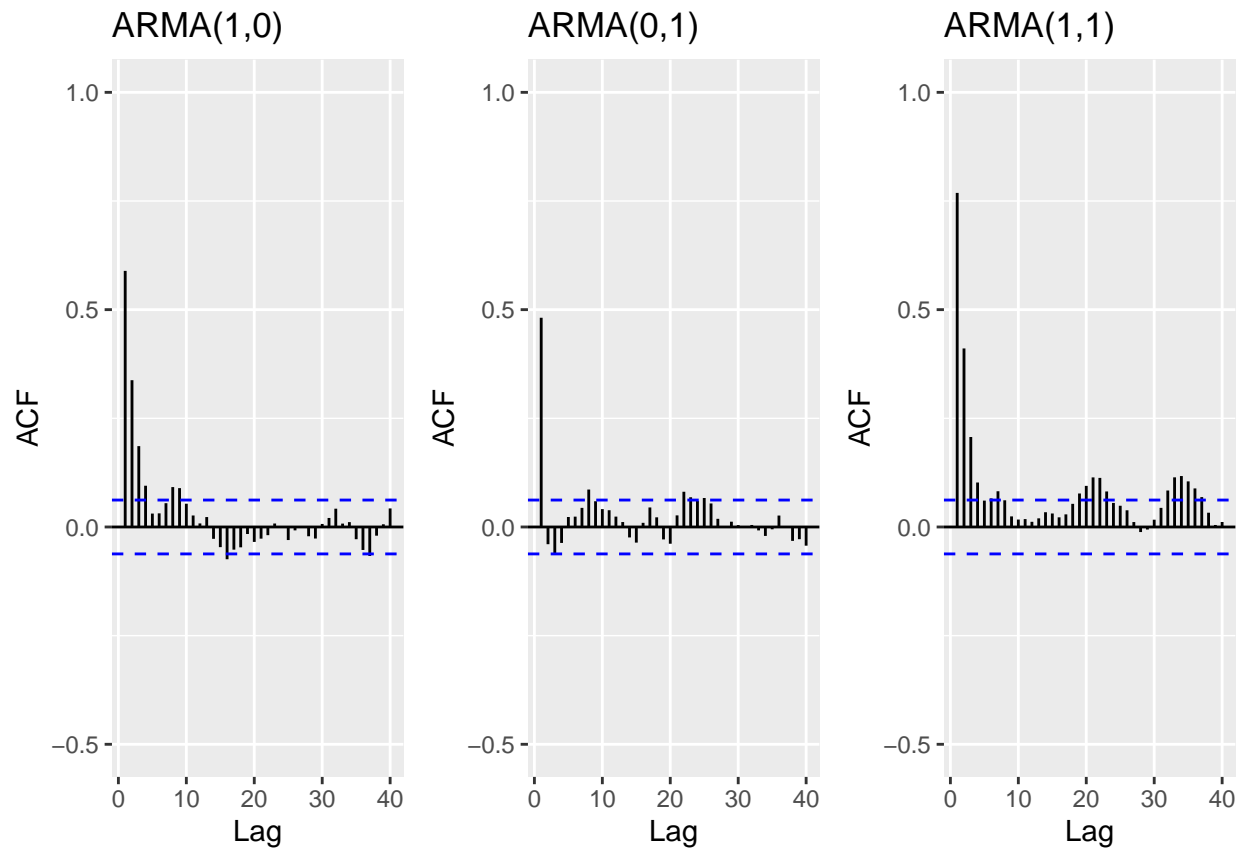
```
# ARMA(1,1)
model_3b <- arima.sim(list(order = c(1,0,1), ar = 0.6, ma = 0.9), n = 1000)

autoplot(model_3b) +
  ylab("Simulated value") +
  ggtitle("Simulated Time Series for ARMA(1,1) Model; n=1000")
```

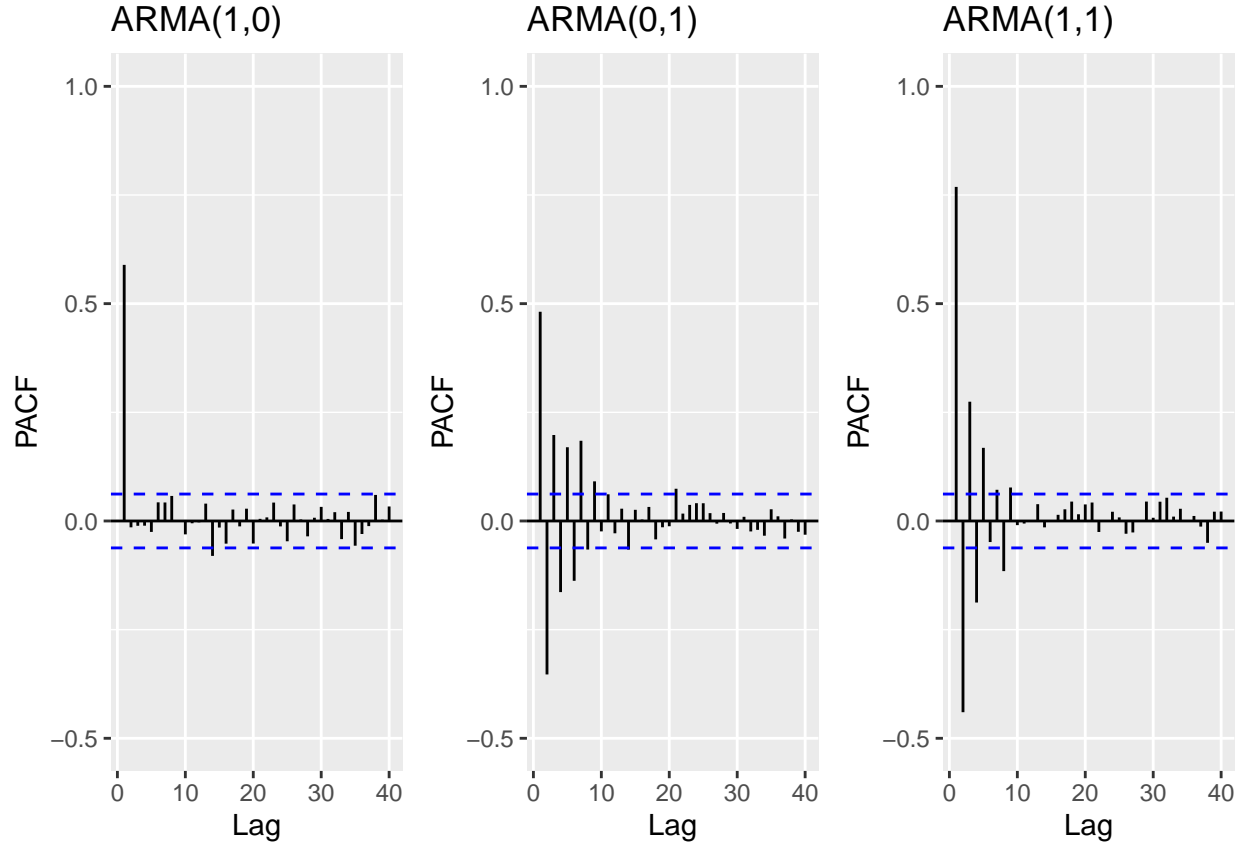

Simulated Time Series for ARMA(1,1) Model; n=1000



```
# part b, plotting ACF values
plot_grid(autoplot(Acf(model_1b, lag=40, plot=FALSE),
  ylim=c(-0.5, 1), main="ARMA(1,0)"),
  autoplot(Acf(model_2b, lag=40, plot=FALSE),
  ylim=c(-0.5, 1), main="ARMA(0,1)"),
  autoplot(Acf(model_3b, lag=40, plot=FALSE),
  ylim=c(-0.5, 1), main="ARMA(1,1)"),
  nrow=1)
```



```
# part c, plotting PACF values
plot_grid(autoplot(pacf(model_1b, lag=40, plot=FALSE),
  ylim=c(-0.5, 1), main="ARMA(1,0)"),
  autoplot(pacf(model_2b, lag=40, plot=FALSE),
  ylim=c(-0.5, 1), main="ARMA(0,1)"),
  autoplot(pacf(model_3b, lag=40, plot=FALSE),
  ylim=c(-0.5, 1), main="ARMA(1,1)"),
  nrow=1)
```



part d: For ARMA(1,0), I can observe slowly decaying ACF values and a clear cut-off point at lag=1 with PACF=0.6. It is clear that this is an AR(1) model. For ARMA(0,1), the cut-off point for ACF is at lag=1, with an ACF value of 0.5. The PACF shows a slowly decaying pattern. It is clear that this is an MA(1) model. For ARMA(1,1), I can identify this model as an ARMA model, but due to the interactions, it is difficult to determine the exact order. However, it is likely to be ARMA(1,1) because for both ACF and PACF plots, the drop in their respective values from lag=1 to lag=2 is significant.

Thus, I am able to identify all models correctly.

part e: For ARMA(1,0), the PACF at lag=1 is exactly 0.6. This matches with the phi value at 0.6 because it is only models an AR model with the coefficient being 0.6. For ARMA(1,1), the PACF at lag=1 is around 0.75 again, larger than the theoretical set value of phi at 0.6. This is because the phi value only controls the autoregressive model, but not the moving average model. Due to the overlapping interactions of both models, the PACF value for ARMA(1,1) may not match with the phi value.

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.

$0.7 * y_{t-1}$ is an AR term, and it only goes up to $t-1$, thus $p=1$. $-0.25 * y_{t-12}$ is an seasonal AR term, and it only goes to $t-12$, not $t-24$. Thus $P=1$. a_t is the error term $0.1 * a_{t-1}$ is the MA term, it only goes to $t-1$, thus $q=1$

Thus, the notation is $ARIMA(1, 0, 1)(1, 0, 0)_{12}$

- (b) Also from the equation what are the values of the parameters, i.e., model coefficients. $0.7 * y_{t-1}$ indicates that $\phi_1 = 0.7$ $-0.25 * y_{t-12}$ indicates that $\phi_{12} = -0.25$ $0.1 * a_{t-1}$ indicates that $\theta_1 = 0.1$ (the coefficient is -0.1, but theta value is 0.1 because the minus sign follows convention) All other model coefficients = 0

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?

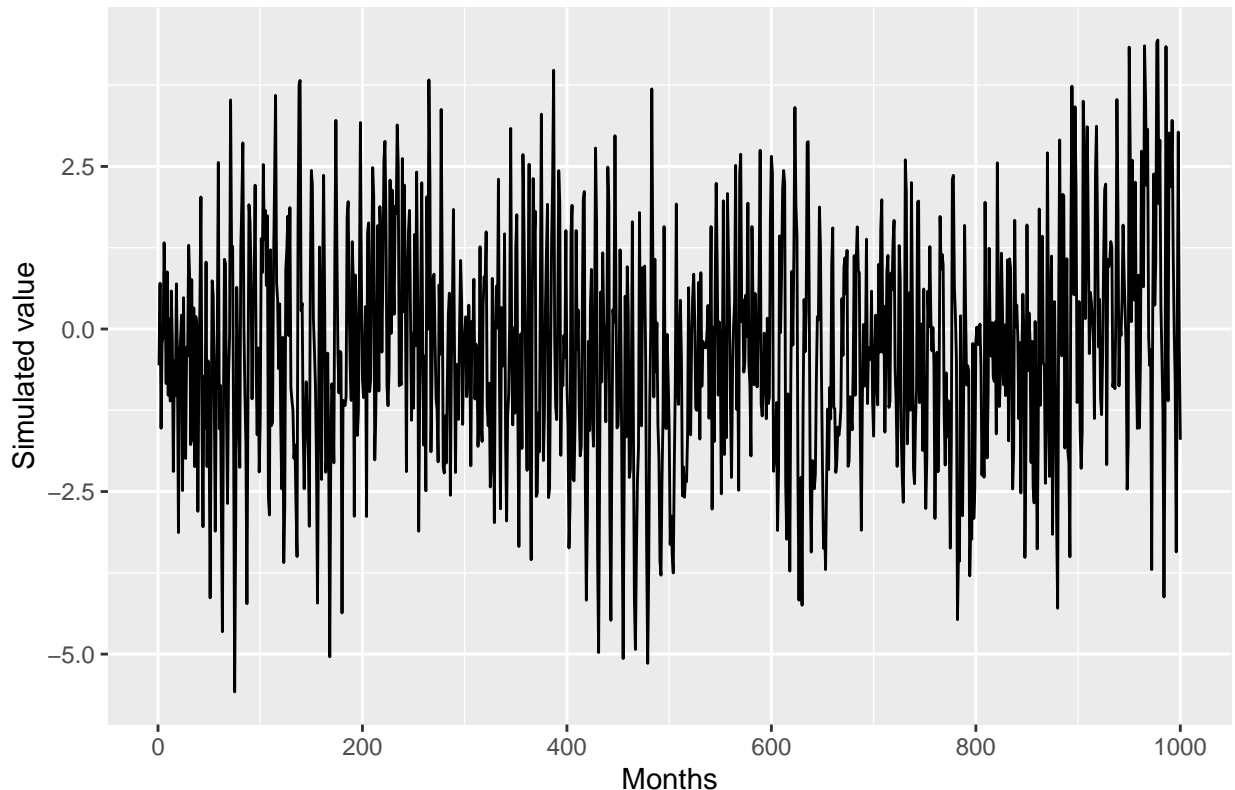
```
# setting seed for reproducibility
set.seed(101)

seasonalARIMA <- sim_sarima(n=1000, model = list(ma=0.5, sar=0.8, nseasons=12))

seasonalARIMA_ts <- ts(seasonalARIMA)

autoplot(seasonalARIMA_ts) +
  ylab("Simulated value") +
  xlab("Months") +
  ggtitle("Simulated Time Series for the given ARIMA model; n=1000 months")
```

Simulated Time Series for the given ARIMA model; n=1000 months

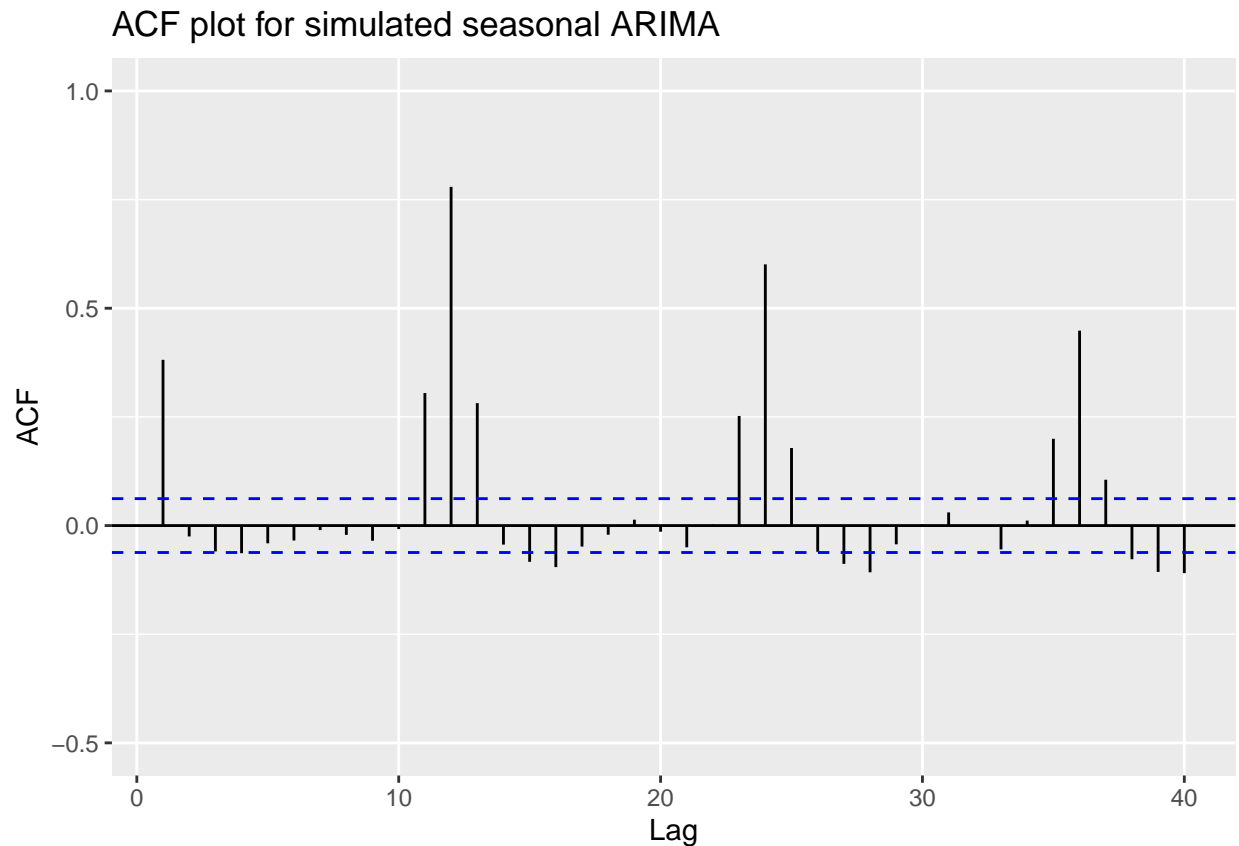


The plot indeed looks seasonal based on the overall sinusoidal shape. However, further analysis is needed to verify whether the period generated indeed=12 months, because there are so many cycles it is difficult to find the overall pattern.

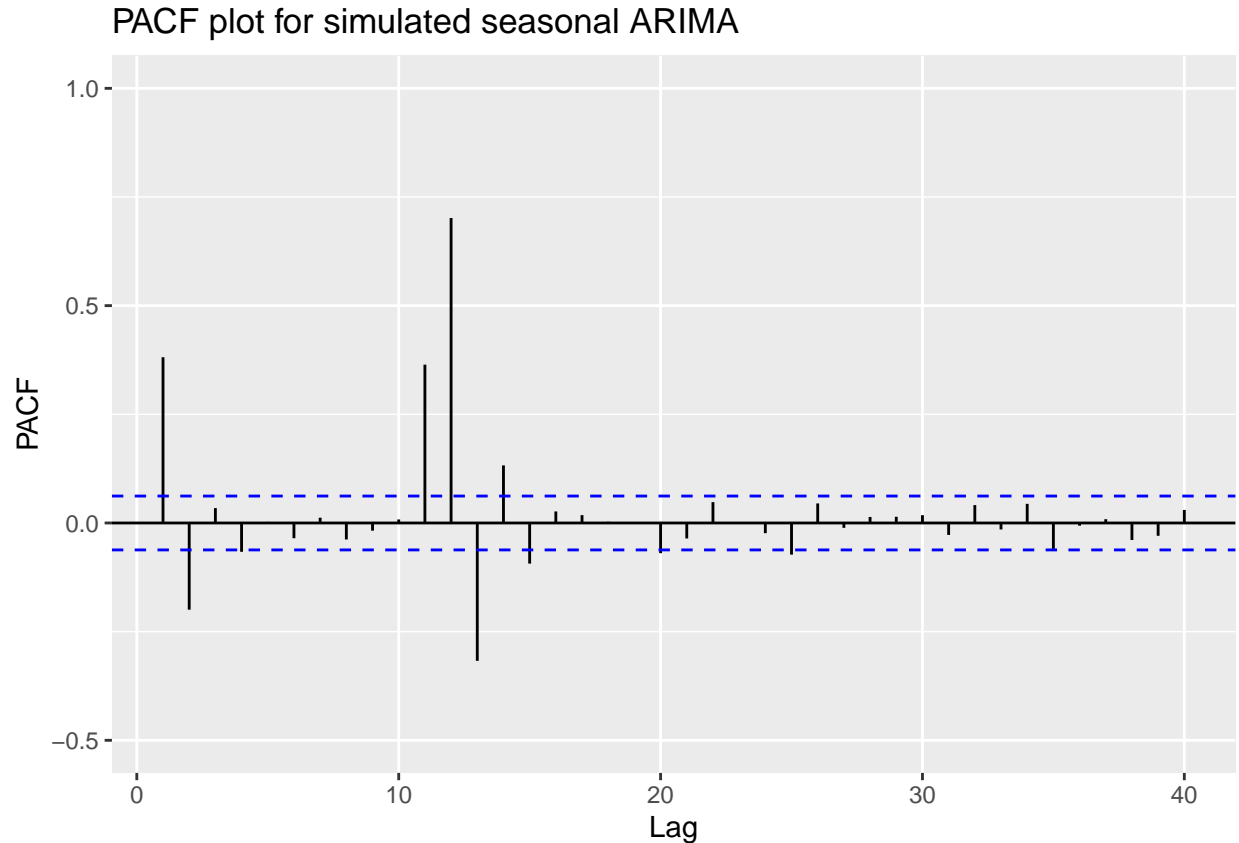
Q5

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(seasonalARIMA_ts,lag=40, plot=FALSE), ylim=c(-0.5,1),  
         main="ACF plot for simulated seasonal ARIMA")
```



```
autoplot(pacf(seasonalARIMA_ts,lag=40, plot=FALSE), ylim=c(-0.5,1),  
         main="PACF plot for simulated seasonal ARIMA")
```



The plots are somewhat-representing for the simulated ARIMA series. The seasonality is obvious, with a peaking value for every 12 months, thus the period is 12. Based on the PACF, there is no obvious cut-off point for the first few lags, thus the AR component is 0. However, the second season has a high PACF value that cuts off after lag=13. Thus, there is SAR component in the model and it is 1. However, from the graph, the deduced coefficient for the SAR component is 0.7, but the real value is 0.8. This may be resulted from the interactions of the seasonal AR component and the MA component.

The ACF plot has a cut-off at lag=1, with magnitude of 0.4. Thus from the graph it can be known that the model has MA(1) component with coefficient being around 0.4. However, this again differs from the real value at 0.5. This may again be resulted from the interactions of the different components in the model. The seasonality is obvious on the ACF plot, and from the second season there seems to be a cut-off point at lag=13, which misleads the audience to think there is a SMA(1) component while in fact there is not. This may be resulted from the strong seasonal AR component, which can be confirmed by the exponentially decaying ACF value from the second season onwards.