ENV 790.30 - Time Series Analysis for Energy Data | Spring 2023

Assignment 6 - Due date 03/06/23

Katherine Burley

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 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_Sp23.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: "xlsx" or "readxl", "ggplot2", "forecast","tseries", and "Kendall". 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.

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.\

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

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

```
library(tseries)

library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2
## --
## v ggplot2 3.4.0     v purrr   1.0.1
## v tibble  3.1.8     v dplyr   1.1.0
## v tidyr   1.3.0     v stringr 1.5.0
## v readr   2.1.3     v forcats 1.0.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
library(readxl)
library(Kendall)
```

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 an auto-regressive order 2 model (AR(2)), the ACF plot should decay exponentially over the number of lags, and the PACF should have a sharp cutoff, where the values for lags 1, and 2 are significant and values for lags beyond that are generally much smaller and not significant.

- MA(1)

  Answer: In a moving average order 1 model (AR(2)), the PACF plot should decay exponentially over the number of lags with no clear cutoff, while the ACF will have a sharp cutoff, where the value for lag 1 is significant and values for lags beyond 1 are not significant.

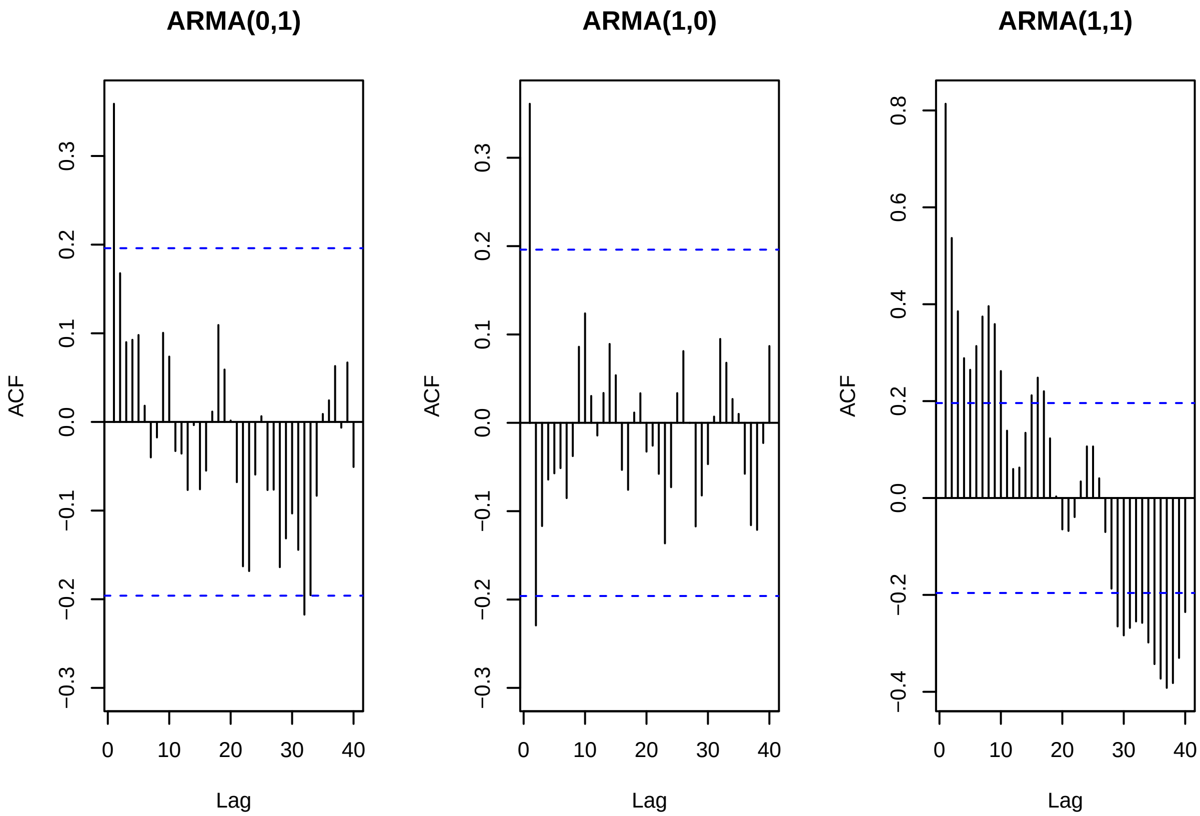
Q2

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

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

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

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

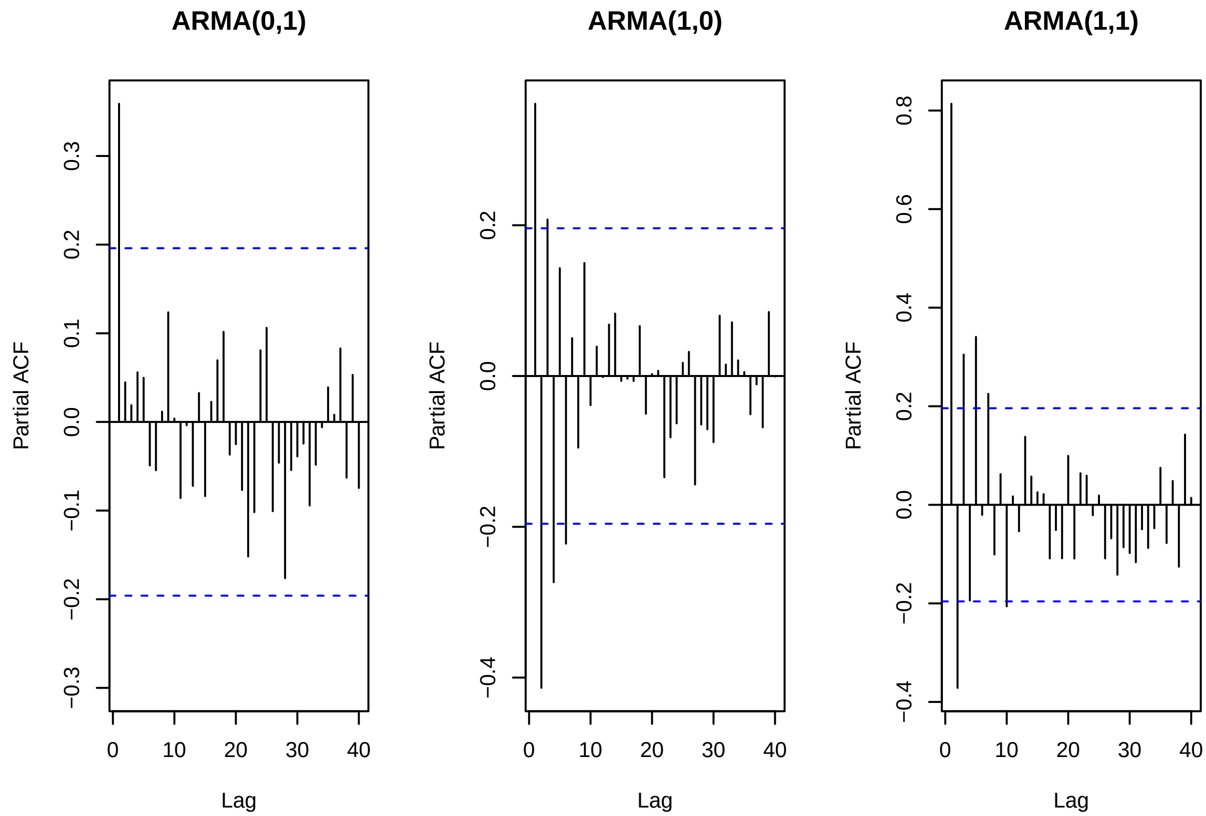
(a) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use command $par(mfrow = c(1, 3))$ that divides the plotting window in three columns).

```
# Plot 3 ACFs
par(mfrow=c(1,3))
Acf(ar_sim, lag=40, main = "ARMA(0,1)")
Acf(ma_sim, lag=40, main = "ARMA(1,0)")
Acf(arma_sim, lag=40, main = "ARMA(1,1)")
```



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

```
# Plot 3 PACFs
par(mfrow=c(1,3))
Pacf(ar_sim, lag=40, main = "ARMA(0,1)")
Pacf(ma_sim, lag=40, main = "ARMA(1,0)")
Pacf(arma_sim, lag=40, main = "ARMA(1,1)")
```



(c) 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: The only one I think I would definitely identify correctly is the ARMA(0,1), which has a fairly clear cutoff after lag 1 in the ACF and what looks like an exponential decay in the PACF. I might have identified the ARMA(1,0) since the ACF looks more like its decaying and the PACF looks like it cuts off at 1, although the values are high through the first few lags, which makes it seem like there is also some decay here. For the ARMA(1,1) both plots look like they are decaying to me, the ACF plot looks like its decaying similar to the ARMA(1,0), but the PACF plot also looks like its decaying, similar to the ARMA(0,1). I probably would have guessed ARMA(1,0) for this model.

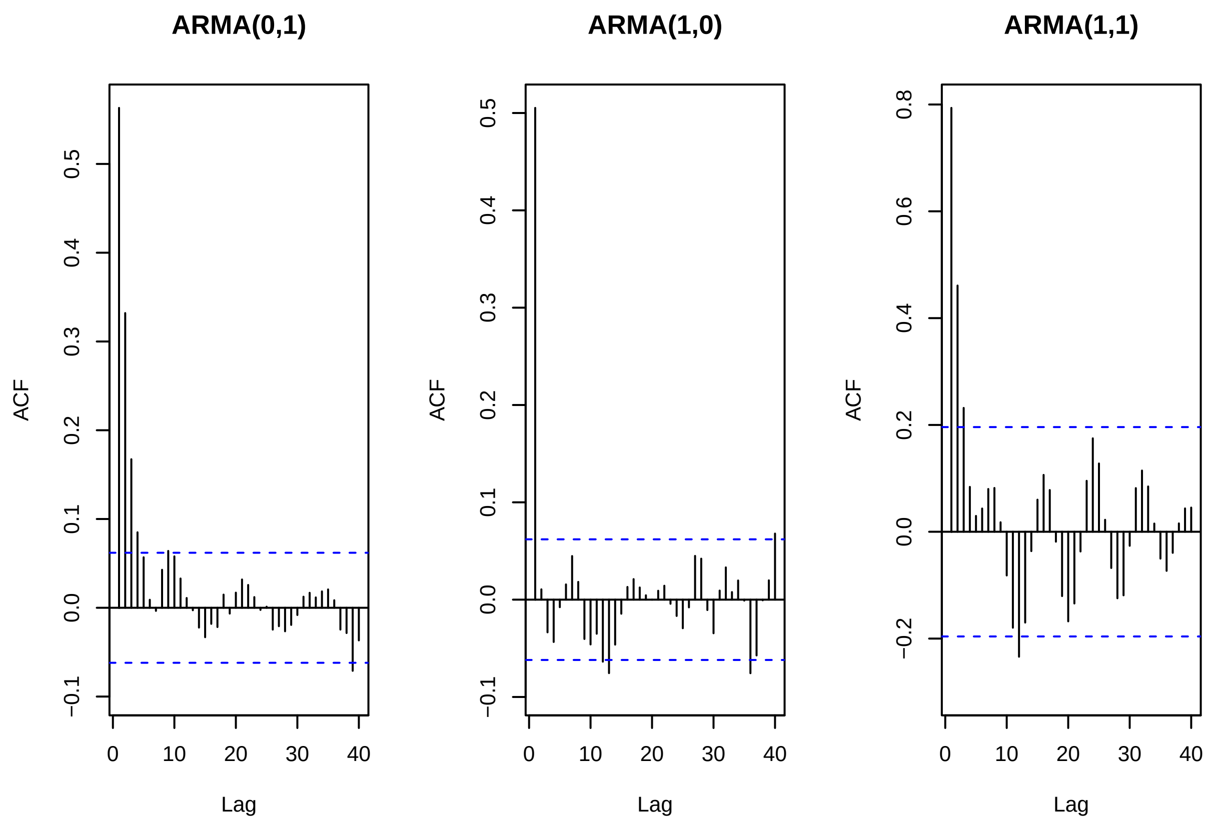
(d) Compare the ACF and PACF values R computed with the theoretical values you provided for the coefficients. Do they match? Explain your answer.

Answer: The coefficients don't match exactly. In the ACF plot for the ARMA(1,0), the coefficient on the first lag is slightly above our theoretical value of 0.6. In the PACF plot for the ARMA(0,1) we can only see that the coefficient on lag 1 is above 0.8 but it looks like it might be slightly below 0.9. On the ARMA(1,1) plot, both the coefficients on lag 1 for ACF and PACF appear to be above slightly 0.8, which is farther off from our theoretical values than the other two models.

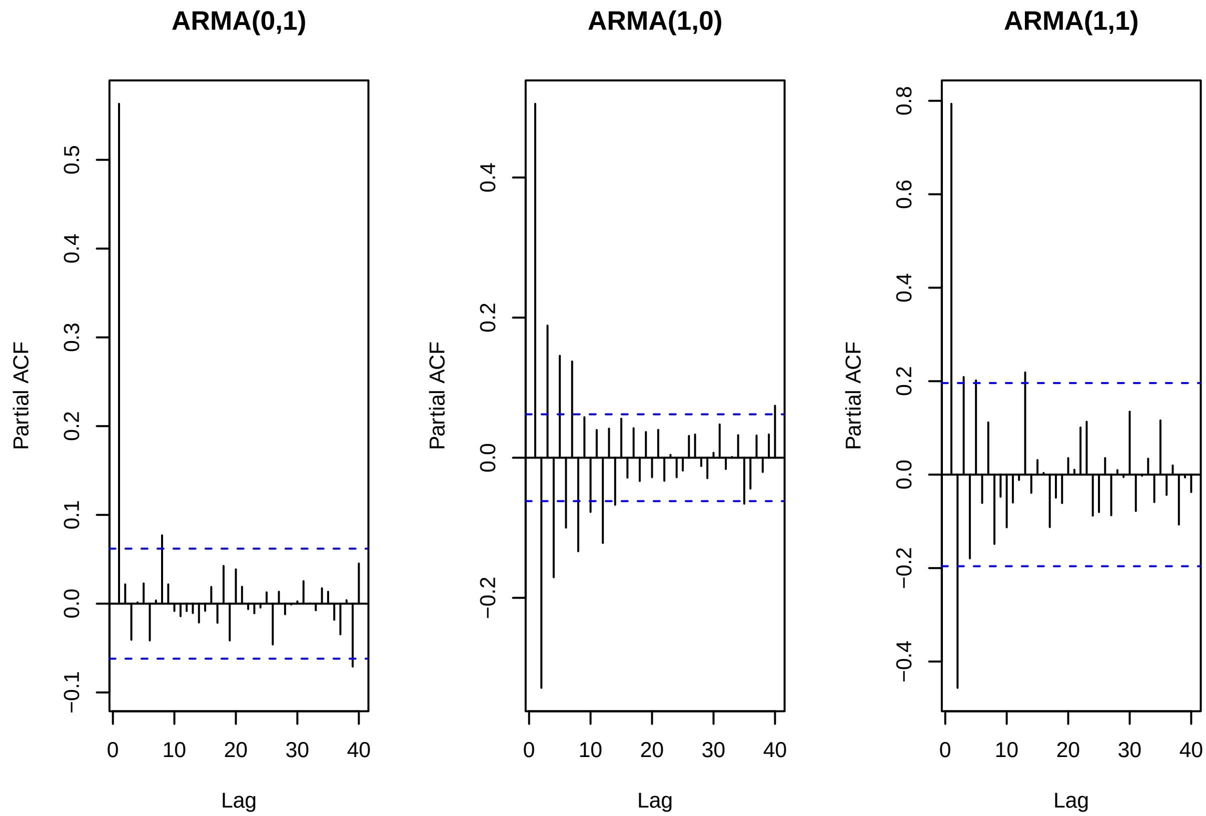
(e) Increase number of observations to $n = 1000$ and repeat parts (a)-(d).

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

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

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

# a. Plot 3 ACFs
par(mfrow=c(1,3))
Acf(ar_sim, lag=40, main = "ARMA(0,1)")
Acf(ma_sim, lag=40, main = "ARMA(1,0)")
Acf(arma_sim, lag=40, main = "ARMA(1,1)")
```




```r
# b. Plot 3 PACFs
par(mfrow=c(1,3))
Pacf(ar_sim, lag=40, main = "ARMA(0,1)")
Pacf(ma_sim, lag=40, main = "ARMA(1,0)")
Pacf(arma_sim, lag=40, main = "ARMA(1,1)")
```



(c). The decay in the ACF plot for ARMA(1,0) and the PACF plot for ARMA(0,1) and the sharp cutoffs at lag 1 in the PACF plot for ARMA(1,0) and ACF plot for ARMA(0,1) are much clearer with 1000 observations. When compared with the other two models, it's more difficult to identify the ARMA(1,1) because both ACF and PACF plots appear to decay and do not have sharp cutoffs.

(d). ARMA(1,0): The coefficient on lag 1 in the PACF plot appears to be slightly higher than our theoretical value of 0.6, similar to the first model with 100 observations. ARMA(0,1): The coefficient on lag 1 in the ACF appear to be around 0.5, which is even farther off than the first model with 100 observations. ARMA(1,1): In the ACF and PACF plot, the coefficients on lag 1 appear to be around 0.8, which is in between our two theoretical coefficients and similar to the first model with 100 observations.

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: ARIMA(1,0,1)(1,0,0)_12 p=1 because previous observations of y appear in the equation d=0 because there is no constant term appearing in the model q=1 because previous deviations from the mean appear in the model P=1 because seasonal autoregressive terms appear in the model D=0 becuse there is no constant term Q=0 because there are no seasonal moving average terms, and we know P+Q cannot be > 1 s=12 because this is the s in the seasonal AR term

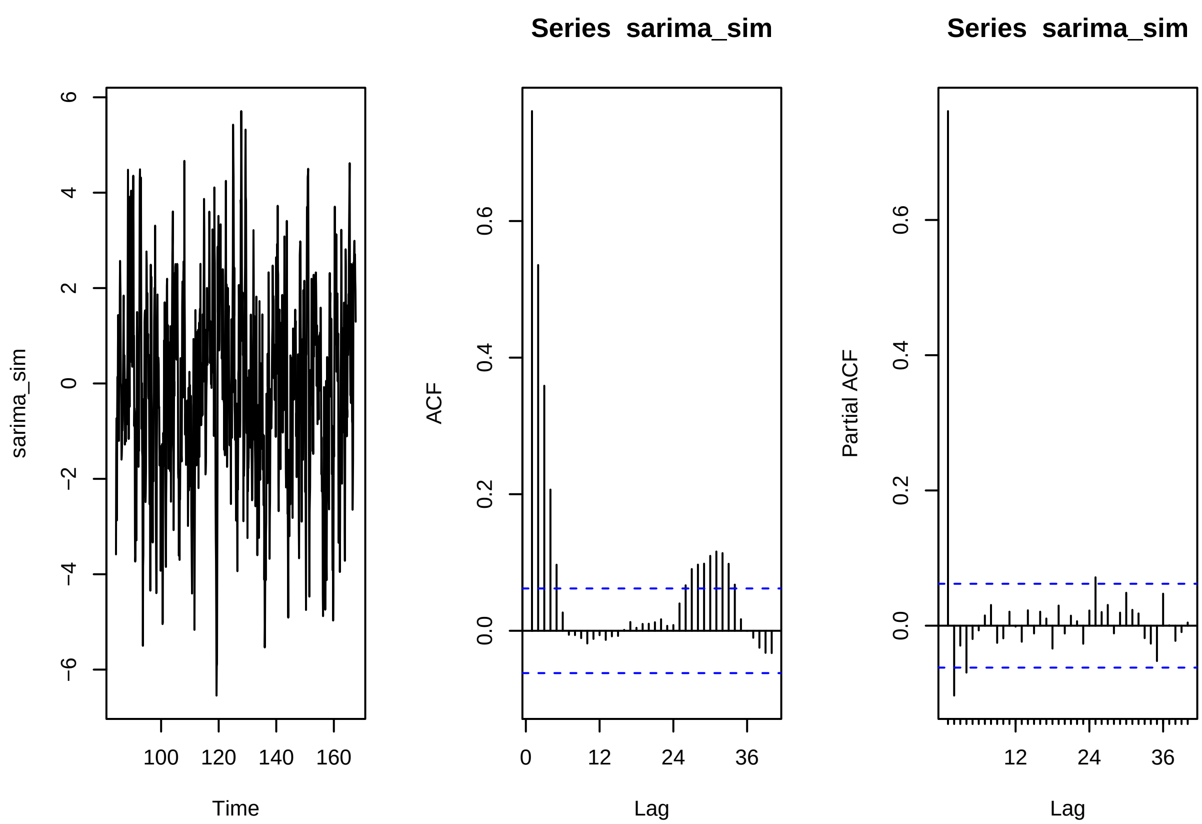
(b) Also from the equation what are the values of the parameters, i.e., model coefficients.

Answer: phi_1 = 0.7, the coefficient on the AR term theta = 0.1, the coefficient on the MA term, which is negative in the model due to the convention of how MA is specified phi_12 = -0.25, because this is the coefficient on the seasonal AR term

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.

```
model <- Arima(ts(rnorm(1000),freq=12), order=c(1,0,1), seasonal=c(1,0,0),
               fixed=c(phi=0.7, theta=0.1, Theta=0, Phi=-0.25))

sarima_sim <- simulate(model, nsim=1000)

# a. Plot 3 ACFs
par(mfrow=c(1,3))
plot(sarima_sim)
Acf(sarima_sim, lag=40)
Pacf(sarima_sim, lag=40)
```



> Answer: The plot represents the AR component well, but not the MA and seasonal AR components. The PACF plot does have a sharp cutoff around 0.7, which is the coefficient we identified in the model. It also looks like there is a negative spike in the ACF at lag 12 near our specified value of -0.25. However, the MA coefficient is not clear from the ACF plot which appears to decay. Perhaps since the coefficient on the MA term is so low (0.1), this plot more closely represents a seasonal AR model and it is more difficult to see the MA component visually.