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

Assignment 6 - Due date 02/28/24

Jenn McNeill

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(cowplot)
#install.packages("sarima")
library(sarima)
```

```
## Loading required package: stats4

##
## Attaching package: 'sarima'

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

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 models, the ACF will decay exponentially with time, which means that what happened in the past is a good indication of what happens in the future. Significant magnitude values on the ACF at the beginning decay because the model is autoregressive. The PACF will identify the order of the AR model depending on how many significant lags it displays. An AR(2) model will display 2 significant lags in the PACF.

- MA(1)

  Answer: MA models show the opposite trends as AR models for ACF and PACF plots. The ACF will identify the order of the MA model depending on how many significant lags it displays, so an MA(1) will show one significant lag. The PACF lags will decay exponentially.

Q2

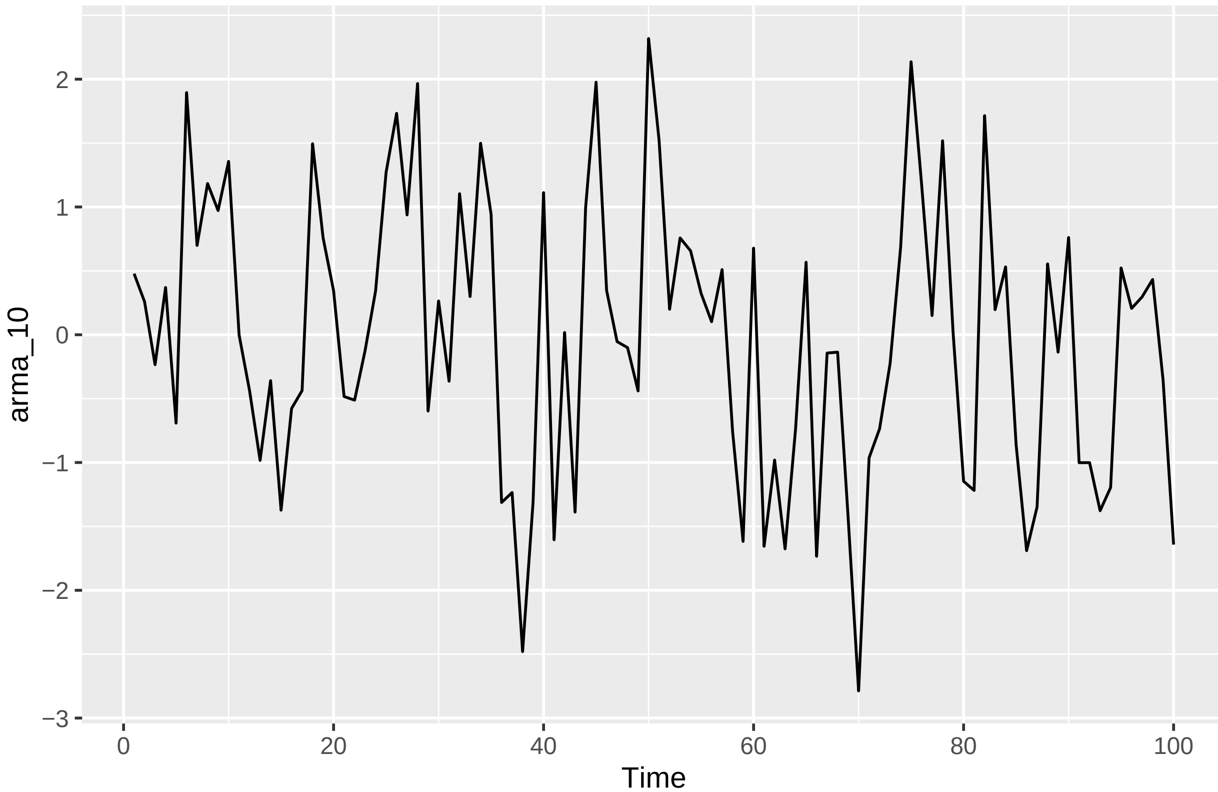
Recall that the non-seasonal ARIMA is described by three parameters ARIMA(p, d, q) where p is the order of the autoregressive component, d is the number of times the series need to be differenced to obtain stationarity and q is the order of the moving average component. If we don't need to difference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

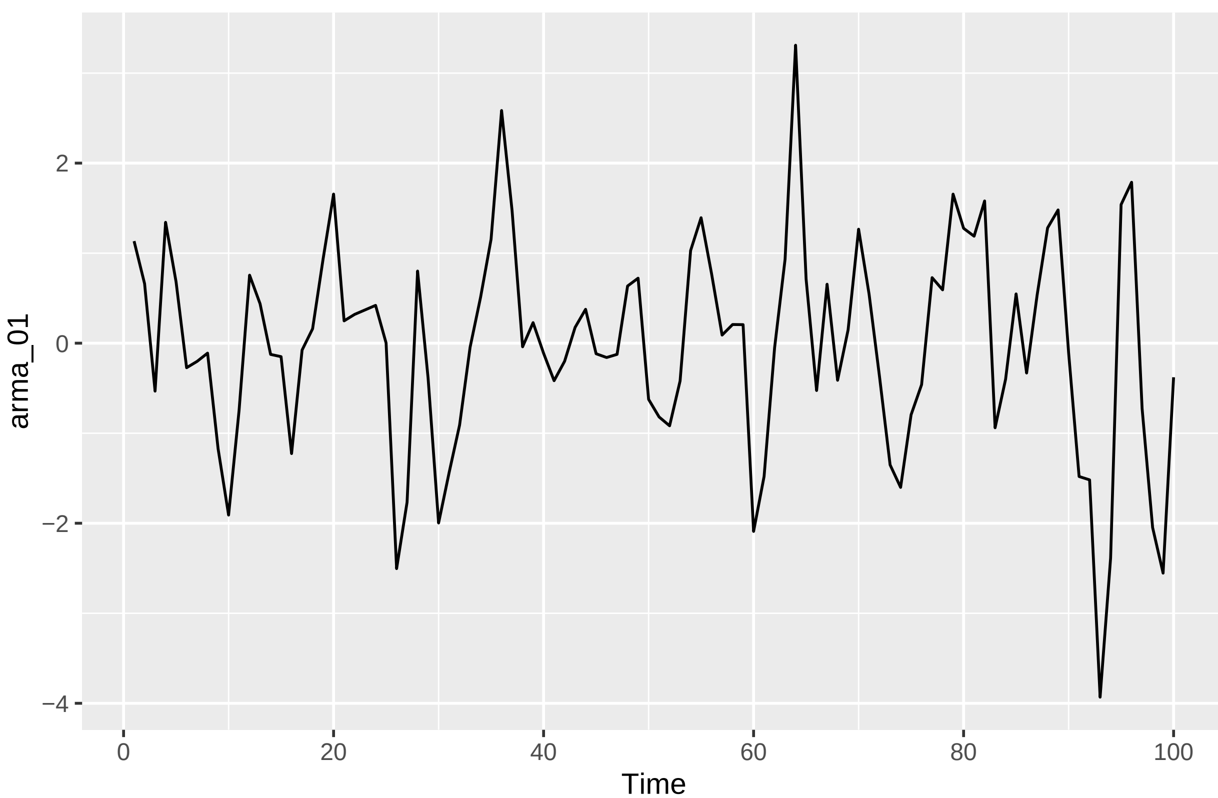
```r
set.seed(100)
arma_10 <- arima.sim(n=100, list(ar = 0.6, ma = 0))
arma_01 <- arima.sim(n=100, list(ar = 0, ma = 0.9))
```

```
## Warning in min(Mod(polyroot(c(1, -model$ar)))): no non-missing arguments to
## min; returning Inf
```

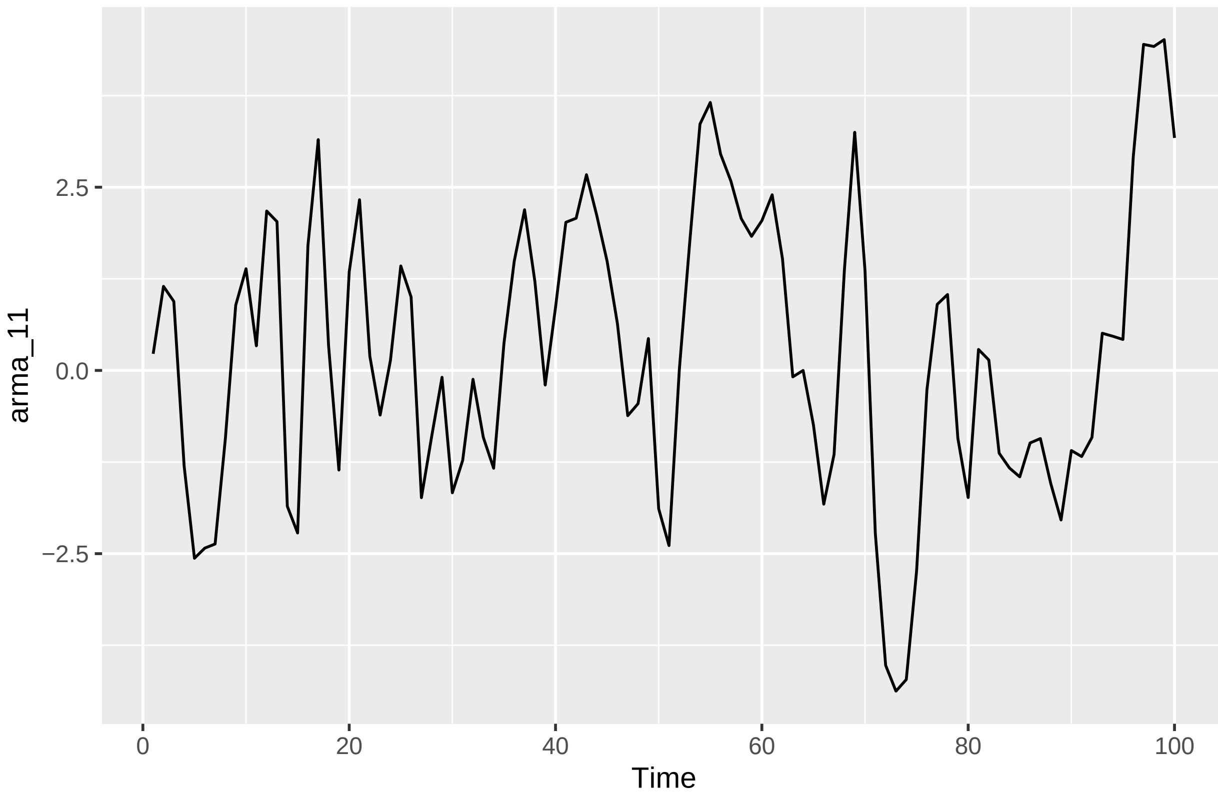
```r
arma_11 <- arima.sim(n=100, list(ar = 0.6, ma = 0.9))

autoplot(arma_10)
```




```r
autoplot(arma_01)
```



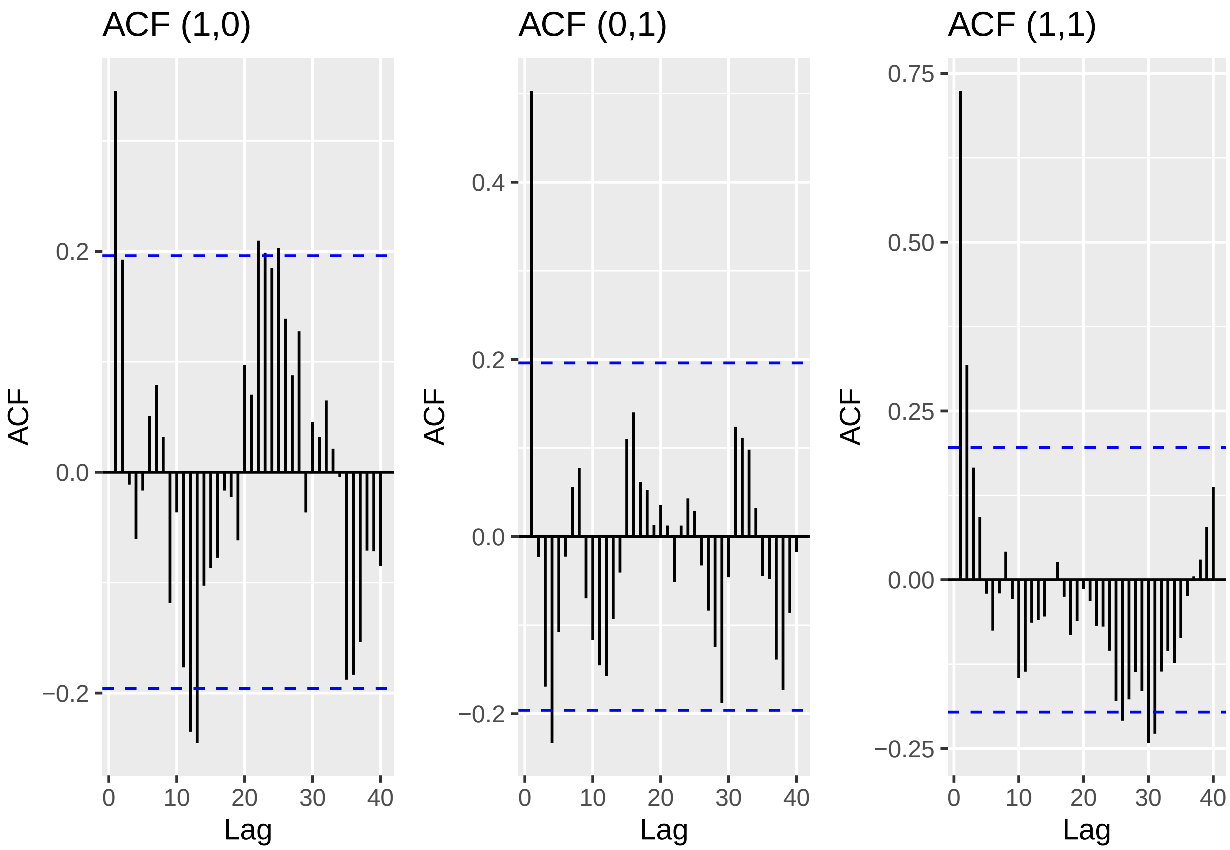
```r
autoplot(arma_11)
```



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

```r
plot_grid(
  autoplot(Acf(arma_10, lag = 40, plot=FALSE), main = "ACF (1,0)"),
  autoplot(Acf(arma_01, lag = 40, plot=FALSE), main = "ACF (0,1)"),
  autoplot(Acf(arma_11, lag = 40, plot=FALSE), main = "ACF (1,1)"),
  ncol=3
)
```

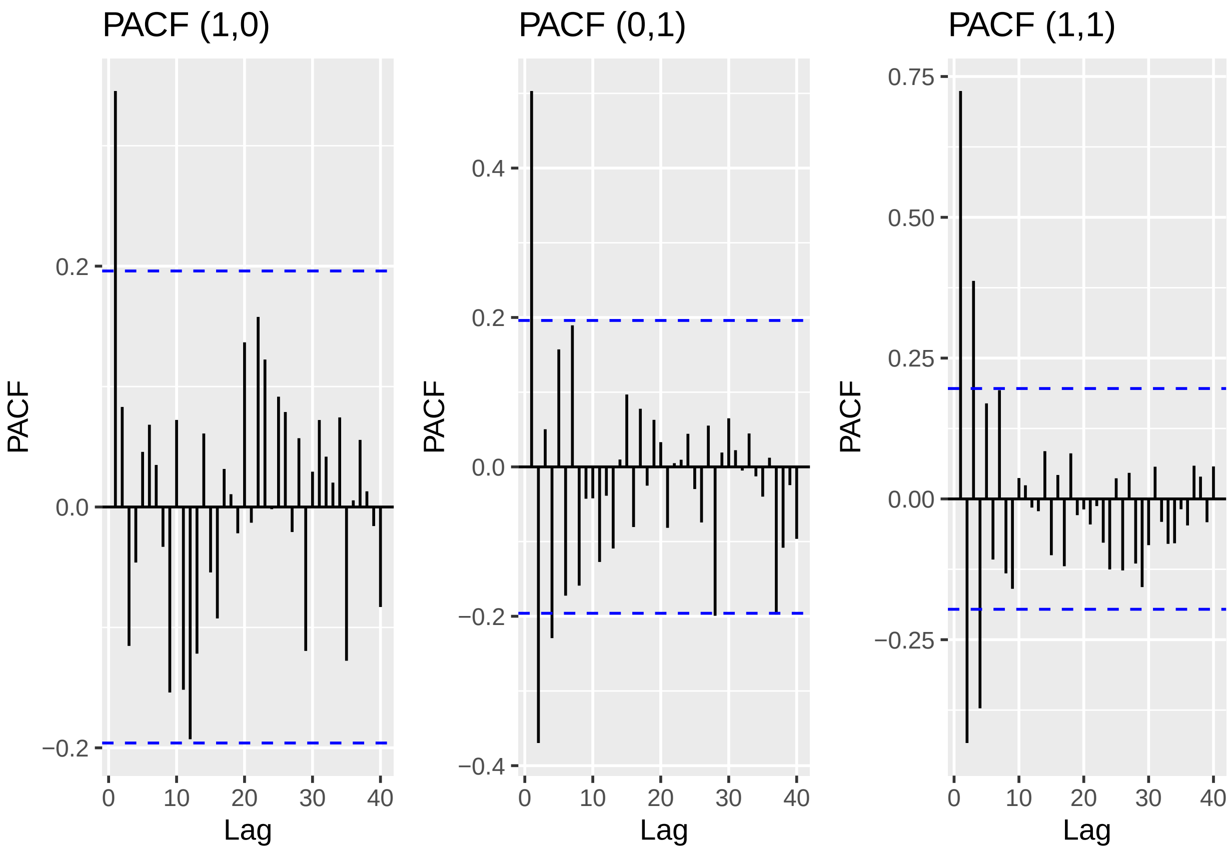
```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown parameters: ‘main‘
## Ignoring unknown parameters: ‘main‘
## Ignoring unknown parameters: ‘main‘
```



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

```
plot_grid(
  autoplot(Pacf(arma_10, lag = 40, plot=FALSE), main = "PACF (1,0)"),
  autoplot(Pacf(arma_01, lag = 40, plot=FALSE), main = "PACF (0,1)"),
  autoplot(Pacf(arma_11, lag = 40, plot=FALSE), main = "PACF (1,1)"),
  ncol=3
)
```

```
## 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: It is difficult to accurately identify the models just based on the looks of the ACFs and PACFs in this context with n = 100. The first order AR model should show an ACF that exponentially decays and a PACF that has one significant lag. If we consider the absolute value of the magnitude of the ACF and PACF, the two first order AR graphs generally follow this rule but have a number of extra "significant" lags that fall outside the dashed blue lines and alter our results. For the first order MA model, we should see an ACF with one significant lag and a PACF that exponentially decays. Again considering the absolute values of these plots, we see the same issue that there are a few too many significant lags even though the shape of the graphs is more or less correct. As we increase the number of observations, we should see ACFs and PACFs that are better representations of the orders of these models.

(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?

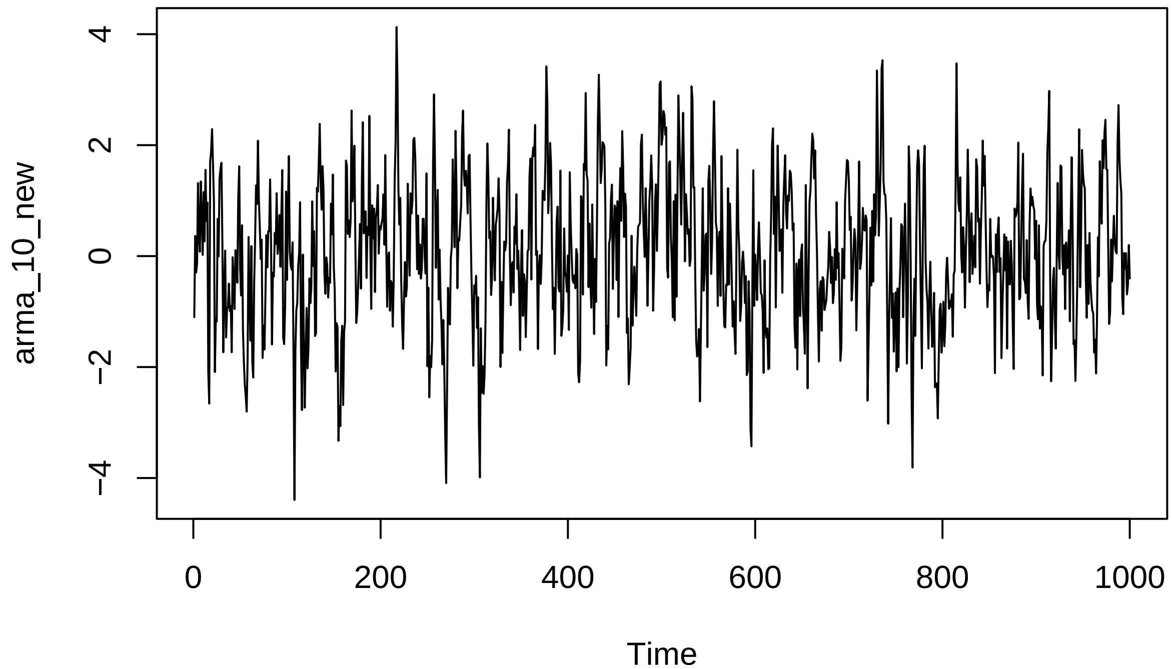
Answer: The phi = 0.6 is not represented on the PACF for ARMA(1,0). The significant lag has a magnitude of about 0.35 instead of 0.6. The PACF(1,1) has one significant lag at the 0.6 mark, but it also has other significant lags that go outside the blue dashed lines and are adding noise to the PACF that makes it appear as if it is higher than a first-order model.

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

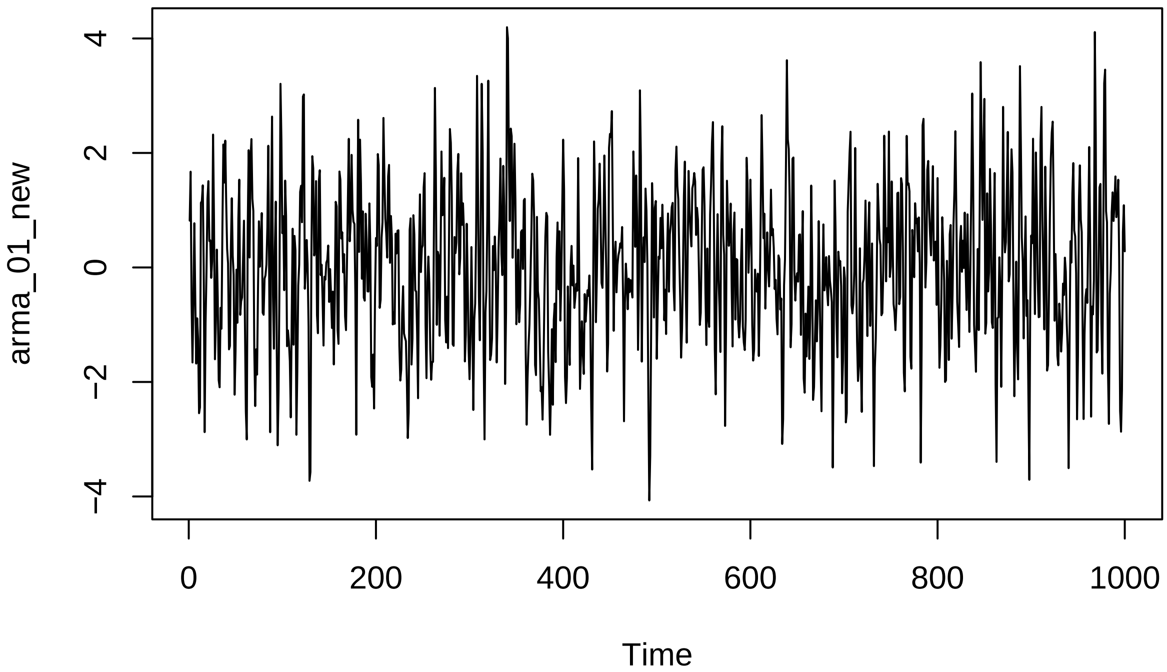
```r
arma_10_new <- arima.sim(n=1000, list(ar = 0.6, ma = 0))
arma_01_new <- arima.sim(n=1000, list(ar = 0, ma = 0.9))
```

```
## Warning in min(Mod(polyroot(c(1, -model$ar)))): no non-missing arguments to
## min; returning Inf
```

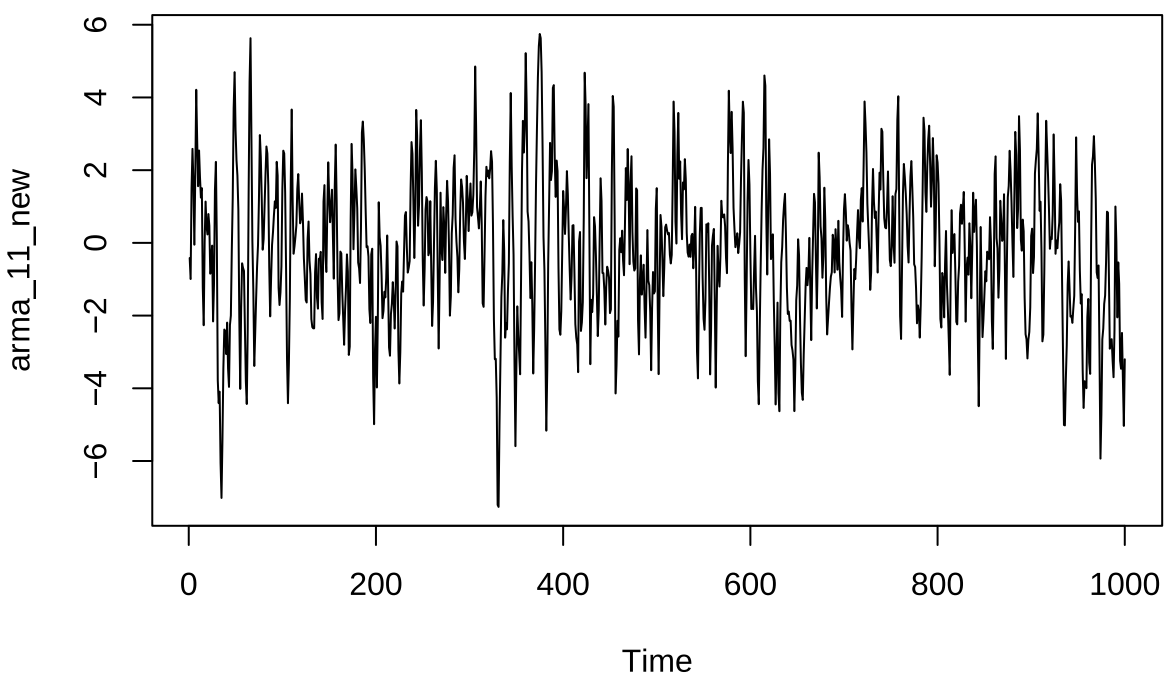
```r
arma_11_new <- arima.sim(n=1000, list(ar = 0.6, ma = 0.9))

plot(arma_10_new)
```




```r
plot(arma_01_new)
```



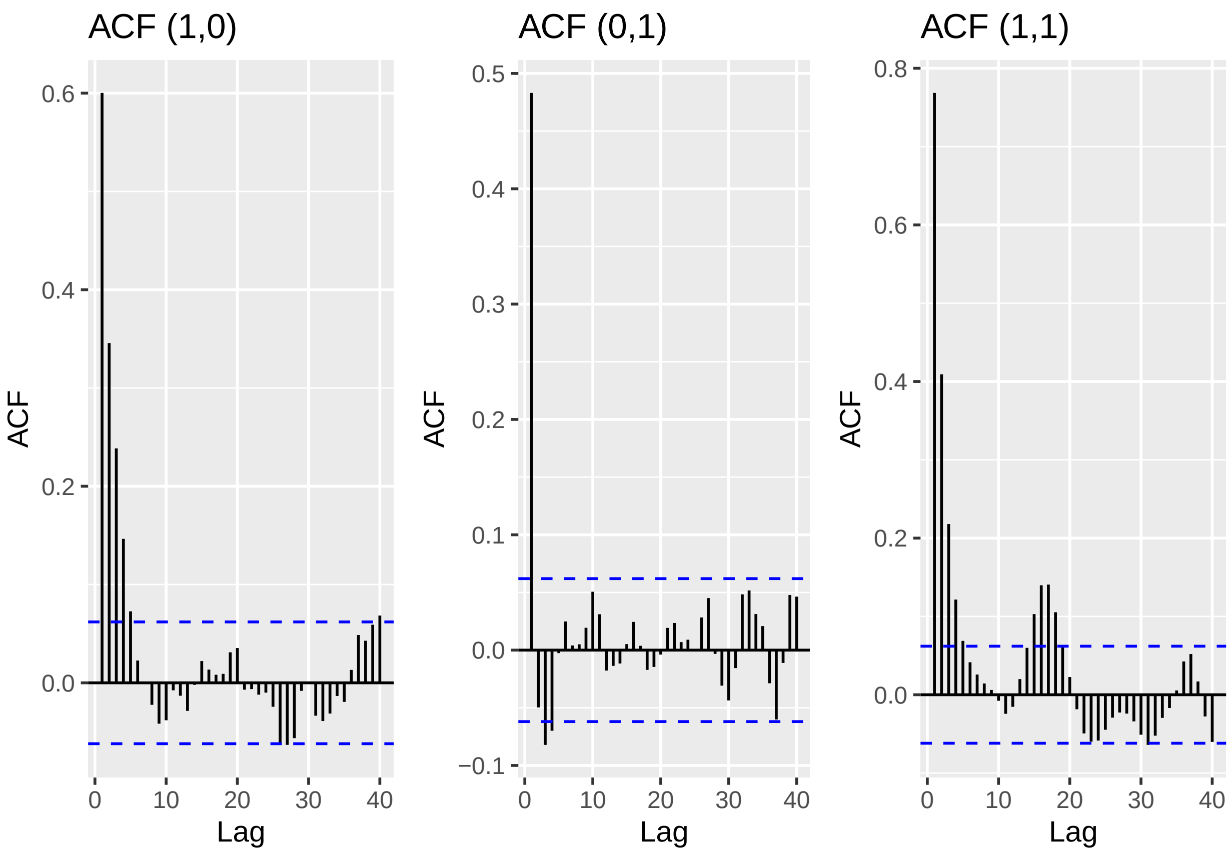

```
plot(arma_11_new)
```




```
plot_grid(
  autoplot(Acf(arma_10_new, lag = 40, plot=FALSE), main = "ACF (1,0)"),
  autoplot(Acf(arma_01_new, lag = 40, plot=FALSE), main = "ACF (0,1)"),
  autoplot(Acf(arma_11_new, lag = 40, plot=FALSE), main = "ACF (1,1)"),
  ncol=3
)
```

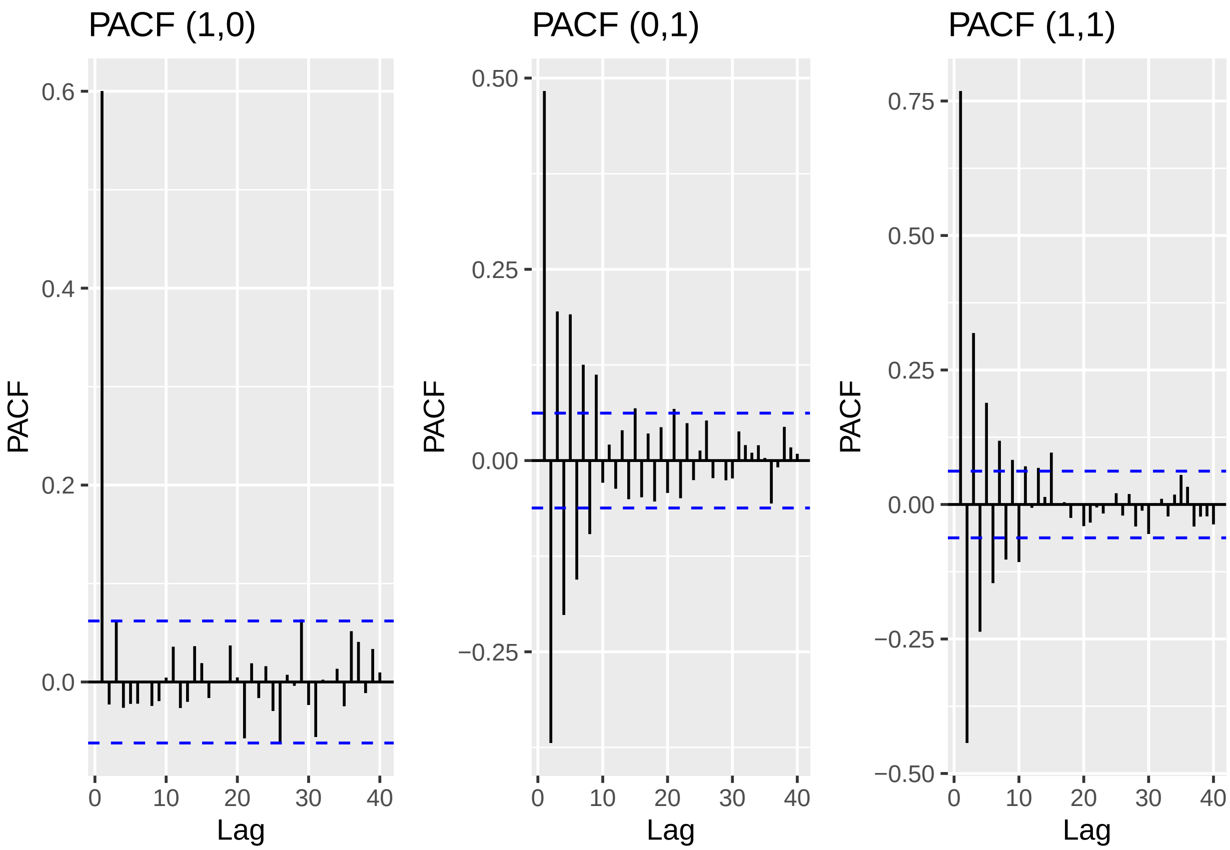
```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown
```

```
## parameters: ‘main‘

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




```r
plot_grid(
  autoplot(Pacf(arma_10_new, lag = 40, plot=FALSE), main = "PACF (1,0)"),
  autoplot(Pacf(arma_01_new, lag = 40, plot=FALSE), main = "PACF (0,1)"),
  autoplot(Pacf(arma_11_new, lag = 40, plot=FALSE), main = "PACF (1,1)"),
  ncol=3
)
```

```
## 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: Increasing n to 1000 definitely helps with the shapes of the ACF and PACF plots. The plots become easier to disginguish and become more useful in determining the orders of our models. For model ARMA(1,0), we see exponential decay in the ACF and one significant lag in the PACF. For model ARMA(0,1), we see exponential decay in the PACF and one significant lag in the ACF. These plots still are not perfect because they display excessive noise outside of the blue dashed lines that disrupts the orders, but they are an improvement from the n = 100 ACFs and PACFs.

(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: ARMA(1,1) model has a significant lag in the ACF that reaches 0.8 and a significant lag in the PACF that reaches 0.6, which closely correspond to our phi and alpha values. But these plots also have additional significant lags that fall outside the dashed blue lines, which means that the models are not effectively representing the (1,1) order of the model that we hoped to see.

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.

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

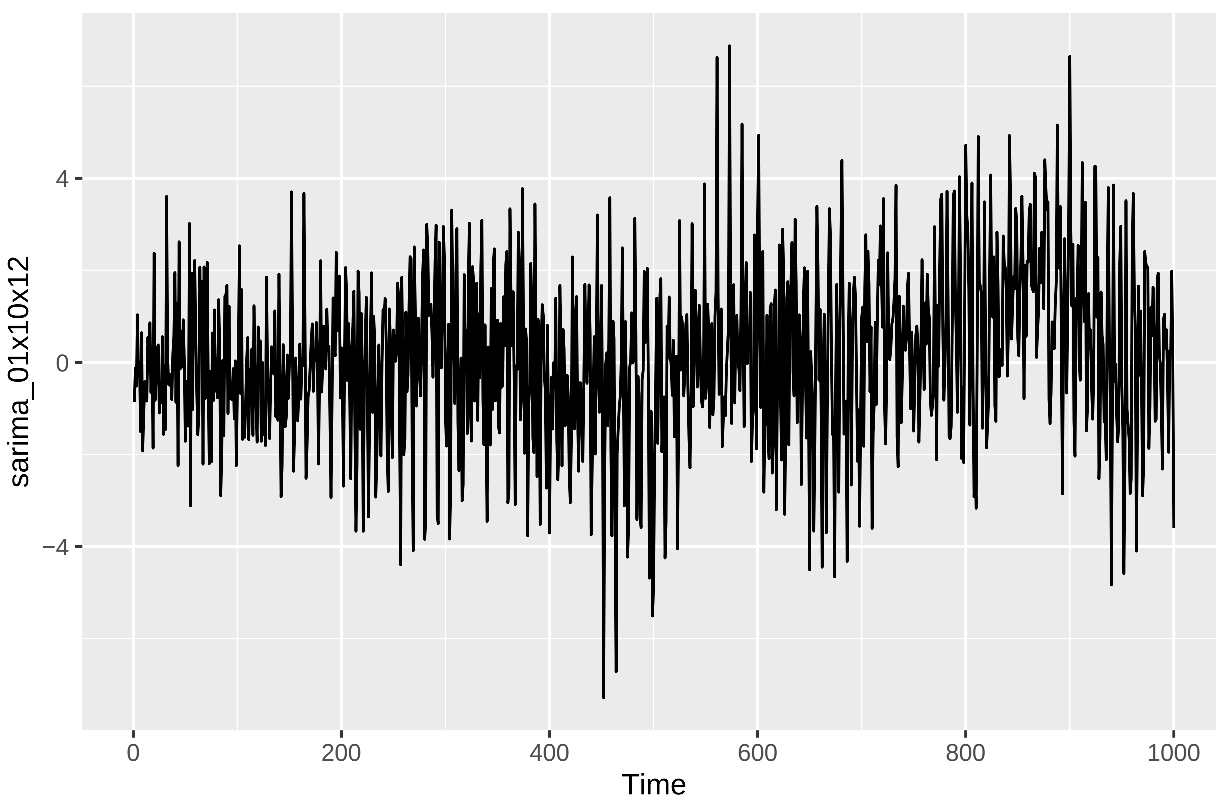
(b) Also from the equation what are the values of the parameters, i.e., model coefficients.

phi_1 = 0.7 $theta_1$ = -0.1 $phi_1 2$ = -0.25

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?

```
set.seed(100)
sarima_01x10x12 <- as.ts(sim_sarima(n=1000, list(ar = 0, ma = 0.5, sar = 0.8, sma = 0, nseasons = 12)))

autoplot(sarima_01x10x12)
```



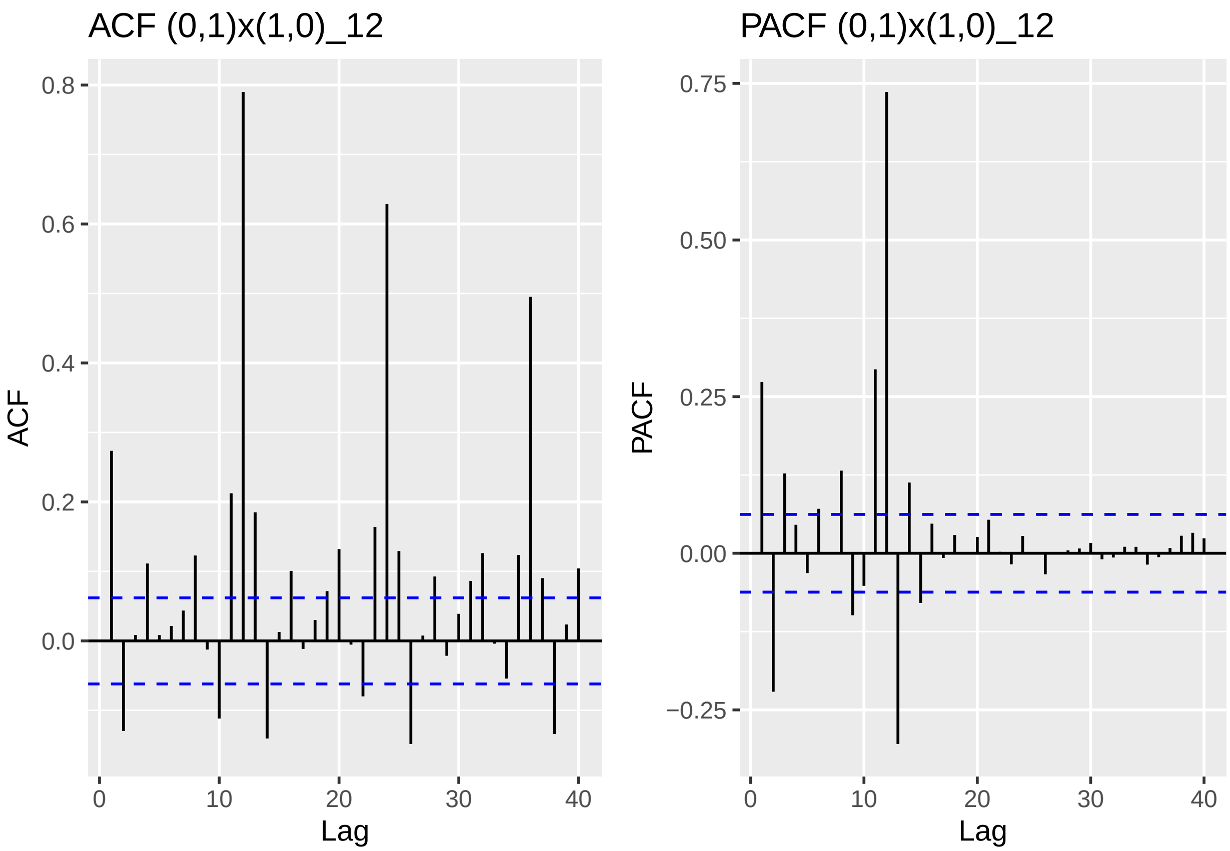
Answer: The plot shows some seasonality in the sense that there appears to be a yearly pattern with a maximum and minimum. The pattern is not extremely obvious nor repeatable in terms of magnitude. The seasonality honestly looks a bit random because our number of observations is so high. Seeing the ACF and PACF will be able to help with seeing seasonal lags and trends.

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(
  autoplot(Acf(sarima_01x10x12, lag = 40, plot=FALSE), main = "ACF (0,1)x(1,0)_12"),
  autoplot(Pacf(sarima_01x10x12, lag = 40, plot=FALSE), main = "PACF (0,1)x(1,0)_12"),
  ncol=2
)
```

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



A SARIMA model with P=1, as our model simulates, should display positive spikes in the ACF at lags 12, 24, 36, etc. and a single positive spike in PACF at lag 12. Our ACF and PACF match these descriptions, which means that the plots are a good representation of the seasonal components of the model. In terms of the non-seasonal components, we would expect the ACF to have a significant first lag with phi = 0.8 and the PACF to have a significant first lag with alpha = 0.5 and then exhibit exponential decay. The seasonality of the time series seems like it might be interfering with the ACFs and PACFs because they do not do a good job of determining the order of the AR and MA components of this model.