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

Assignment 6 - Due date 02/27/25

Daniel Whitehead

Directions

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.

Setting initial options

```
#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(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v lubridate 1.9.3     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(sarima)
```

```
## Loading required package: stats4
##
## Attaching package: 'sarima'
```

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

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: ◦ ACF: Because this is an AR model, and the ar order is 2, the ACF plot will show an exponential decay over time. ◦ PACF: The PACF plot will cut off at the lag 2 because the order of the AR is 2. For an AR(2) model, the PACF will have large spikes at lags 1 and 2 and then cut off to regular values for higher lags. The PACF identifies the order of the model.

- MA(1)

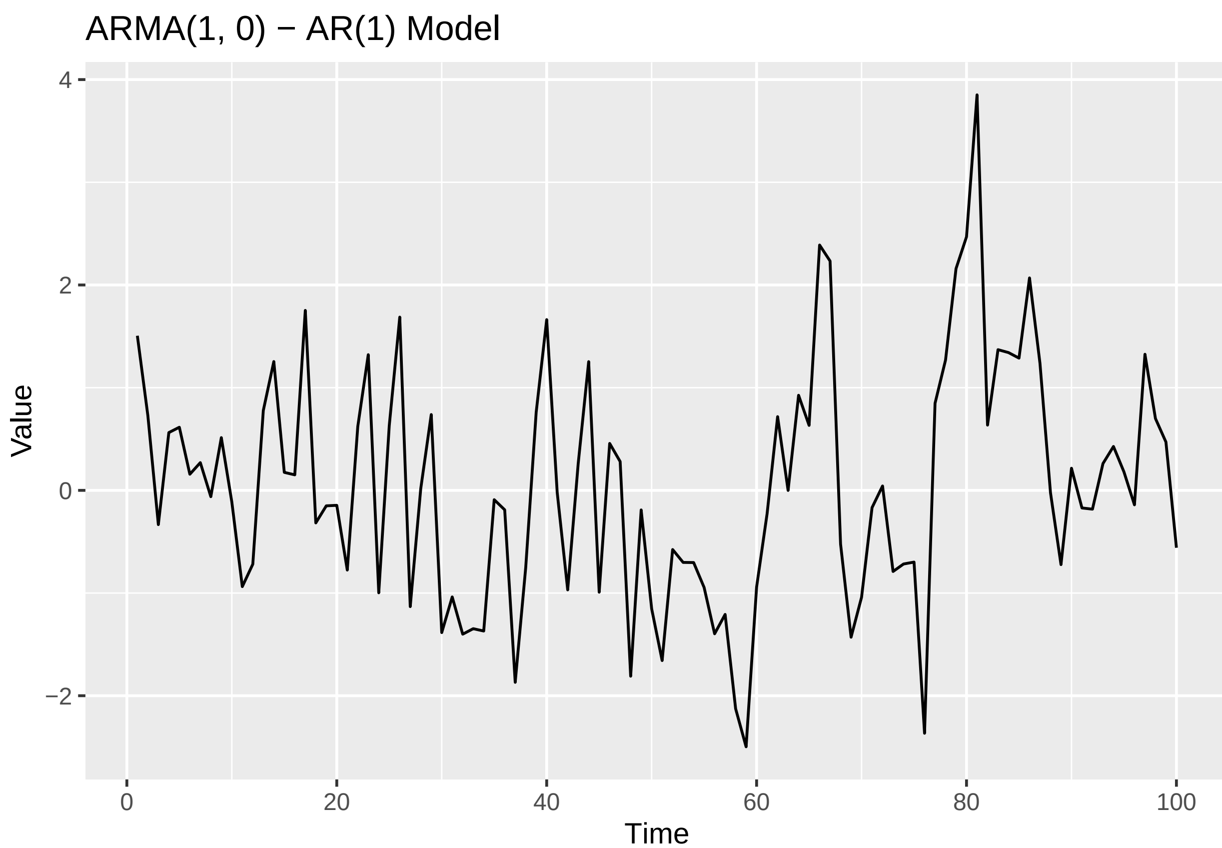
  Answer:◦ ACF: Because this is an MA model, we can look at where the lag cut off is in the acf plot. The ACF plot for MA(1) will cut off after lag 1. In an MA model the ACF identifies the order of the MA model. ◦ PACF: The PACF plot for an MA model will show an exponential decay over time.

Q2

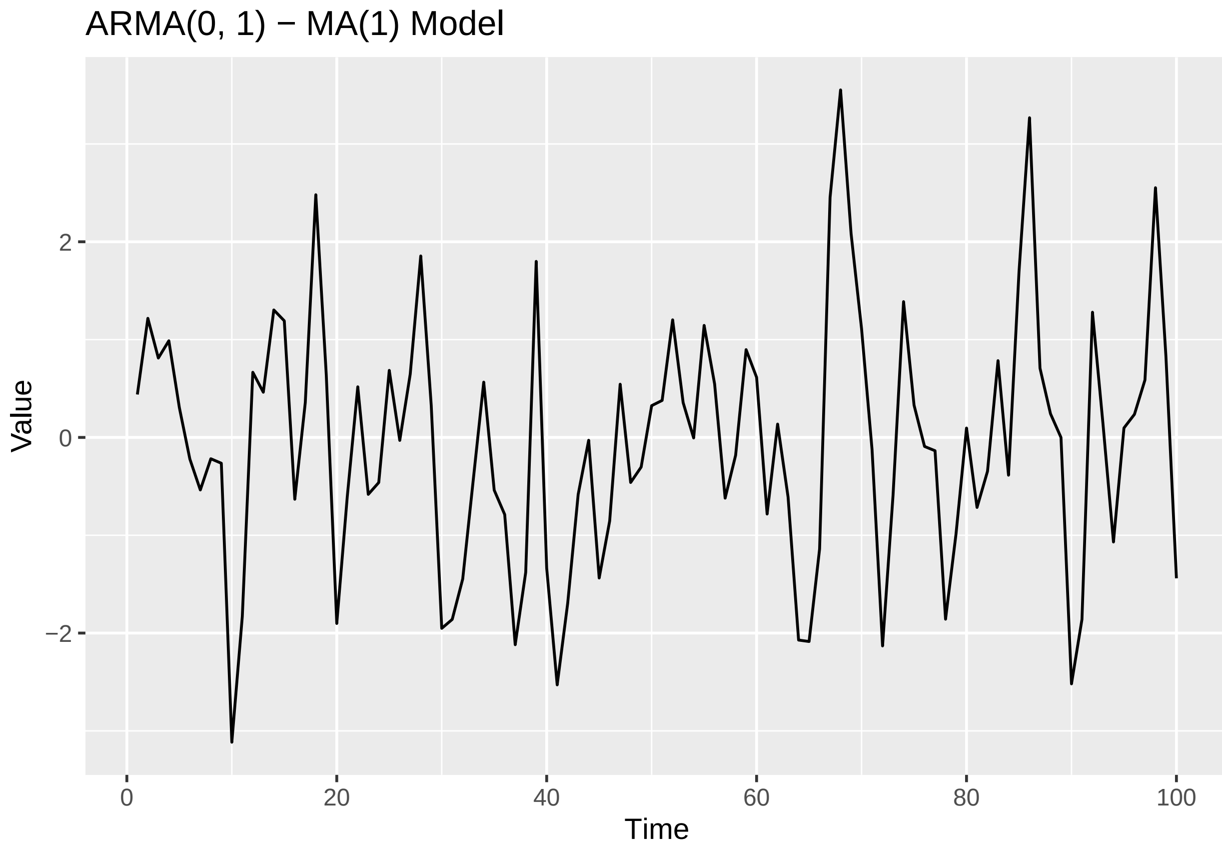
Recall that the non-seasonal ARIMA is described by three parameters ARIMA(p, d, q) where p is the order of the autoregressive component, d is the number of times the series need to be differenced to obtain stationarity and q is the order of the moving average component. If we don't need to difference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

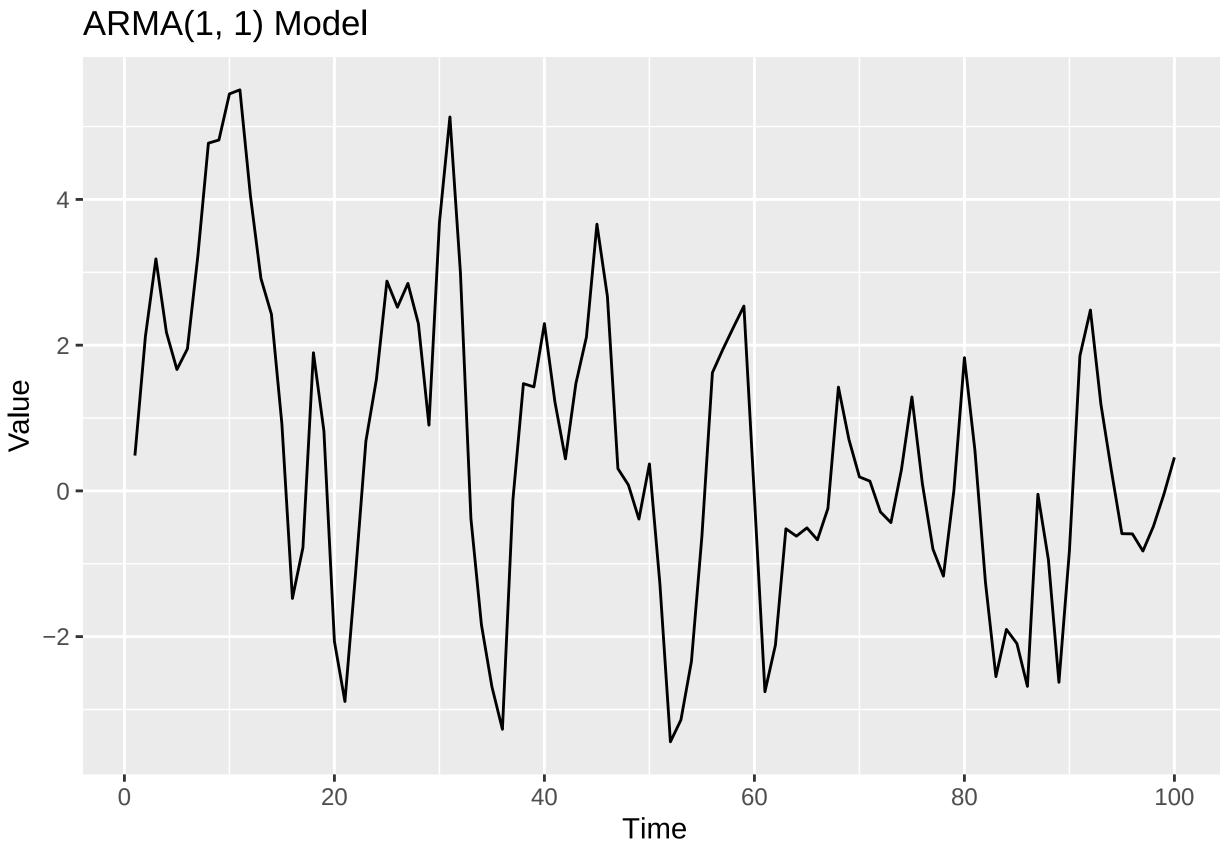
```r
# ARMA(1,0) series (AR(1) model)
arma10 <- arima.sim(n = 100, list(order = c(1, 0, 0), ar = 0.6))

# ARMA(0,1) series (MA(1) model)
arma01 <- arima.sim(n = 100, list(order = c(0, 0, 1), ma = 0.9))

# ARMA(1,1) series (ARMA(1,1) model)
arma11 <- arima.sim(n = 100, list(order = c(1, 0, 1), ar = 0.6, ma = 0.9))

# Create the plots using autoplot from the forecast package
autoplot(arma10) + ggtitle("ARMA(1, 0) - AR(1) Model") + xlab("Time") + ylab("Value")
```




```
autoplot(arma01) + ggtitle("ARMA(0, 1) - MA(1) Model") + xlab("Time") + ylab("Value")
```




```
autoplot(arma11) + ggtitle("ARMA(1, 1) Model") + xlab("Time") + ylab("Value")
```



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

```
# ACFs
acf_10 <- Acf(arma10, plot = FALSE, lag.max = 40)
acf_01 <- Acf(arma01, plot = FALSE, lag.max = 40)
acf_11 <- Acf(arma11, plot = FALSE, lag.max = 40)

# plot acf
plot_acf_10 <- autoplot(acf_10, main = "ACF of ARMA(1,0)")
```

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

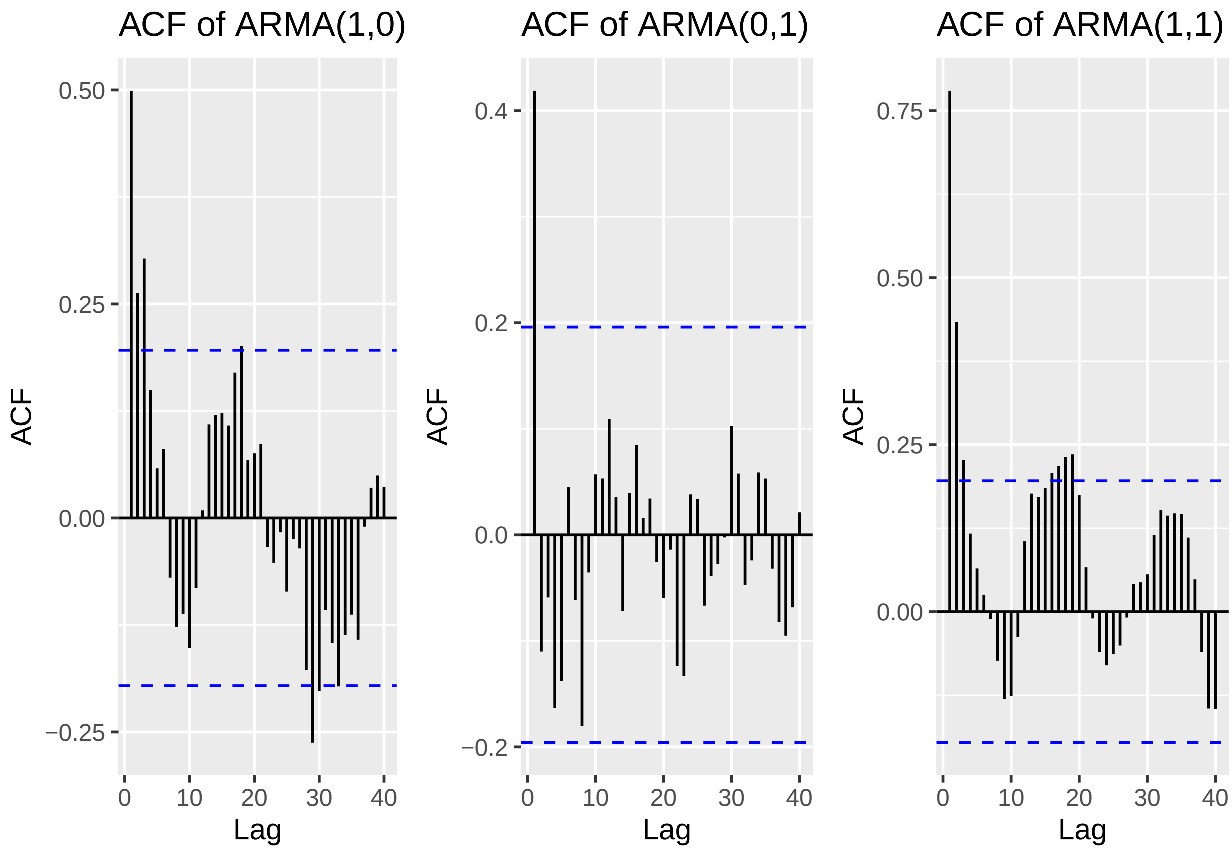
```
plot_acf_01 <- autoplot(acf_01, main = "ACF of ARMA(0,1)")
```

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

```
plot_acf_11 <- autoplot(acf_11, main = "ACF of ARMA(1,1)")
```

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

```r
# cowplot acf
cowplot::plot_grid(plot_acf_10, plot_acf_01, plot_acf_11, nrow = 1)
```



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

```r
# pacfs
pacf_10 <- Pacf(arma10, plot = FALSE, lag.max = 40)
pacf_01 <- Pacf(arma01, plot = FALSE, lag.max = 40)
pacf_11 <- Pacf(arma11, plot = FALSE, lag.max = 40)

# pacf plots
plot_pacf_10 <- autoplot(pacf_10, main = "PACF of ARMA(1,0)")
```

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

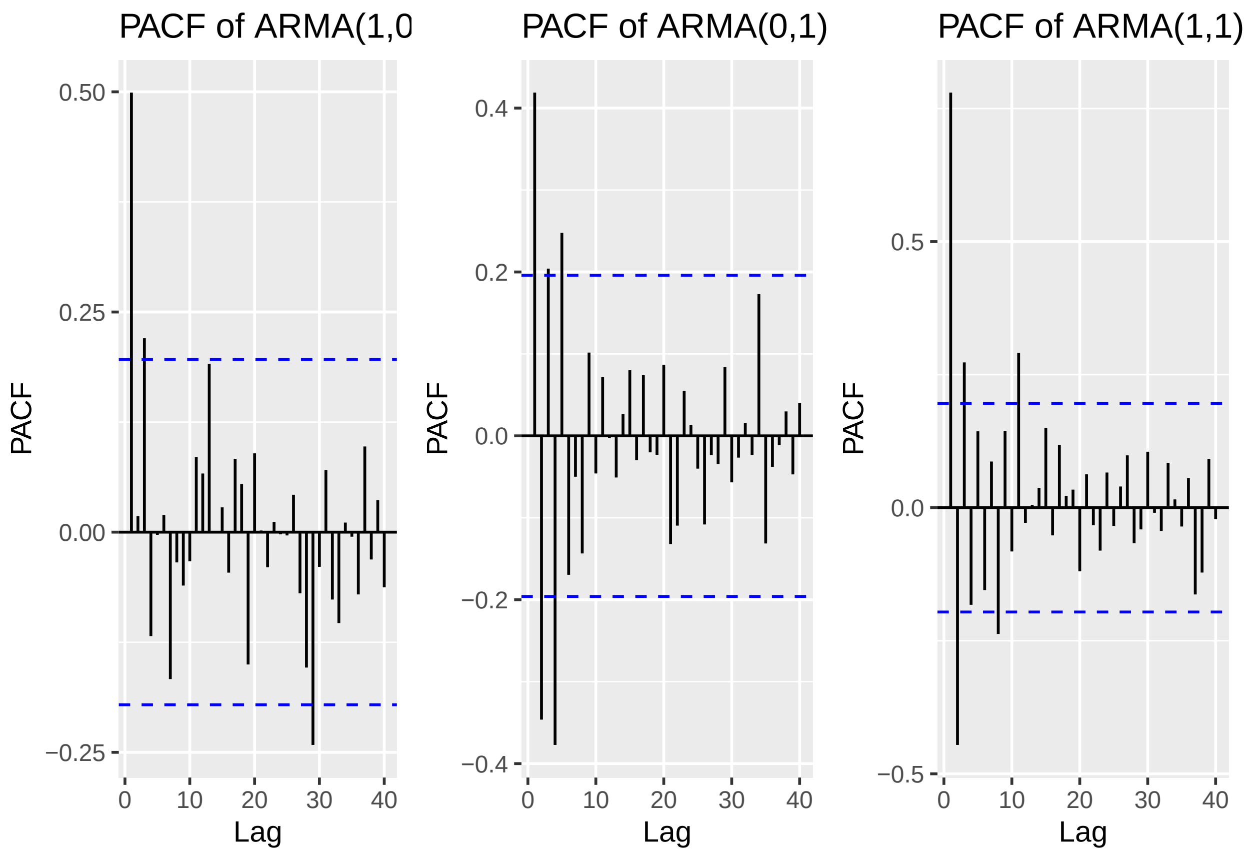
```r
plot_pacf_01 <- autoplot(pacf_01, main = "PACF of ARMA(0,1)")
```

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

```r
plot_pacf_11 <- autoplot(pacf_11, main = "PACF of ARMA(1,1)")
```

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

```r
# plot pacfs
cowplot::plot_grid(plot_pacf_10, plot_pacf_01, plot_pacf_11, 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:I would be able to determine the ARMA 1,0 from the pacf plot because there is a spike at lag 1 that cuts off after. That would mean the AR(1). There looks to be decay in the acf of the ar 1 plot as well, so I would think that p =1, d = 0, and q = 0. For the ARMA (0,1), I would be able to tell it is MA(1) because the acf cuts off after 1. For the MA 1 model, it would be p = 0, d = 0 , and q = 1. For the ARMA 1,1, I would be able to determine that this has both components p = 1 and q = 1 (d= 0) because both the acf plot and the pacf plot have decay. They both also have a spike at 1 then it drops off.

(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: Honestly, the plots do not look to match the values that I thought I specified earlier. While I thought that they should match, maybe they do not specifically need to match as long as the lag cutoffs are correct (they cutoff after the correct lag).

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

```
# ARMA(1,0)  phi = 0.6
arma_10_1000 <- arima.sim(n = 1000, list(order = c(1, 0, 0), ar = 0.6))

# ARMA(0,1), theta = 0.9
arma_01_1000 <- arima.sim(n = 1000, list(order = c(0, 0, 1), ma = 0.9))

# ARMA(1,1) phi = 0.6 and theta = 0.9
arma_11_1000 <- arima.sim(n = 1000, list(order = c(1, 0, 1), ar = 0.6, ma = 0.9))

# make acfs
acf_10_1000 <- Acf(arma_10_1000, plot = FALSE, lag.max = 40)
acf_01_1000 <- Acf(arma_01_1000, plot = FALSE, lag.max = 40)
acf_11_1000 <- Acf(arma_11_1000, plot = FALSE, lag.max = 40)

# plot acfs
plot_acf_10_1000 <- autoplot(acf_10_1000, main = "ACF of ARMA(1,0)")
```

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

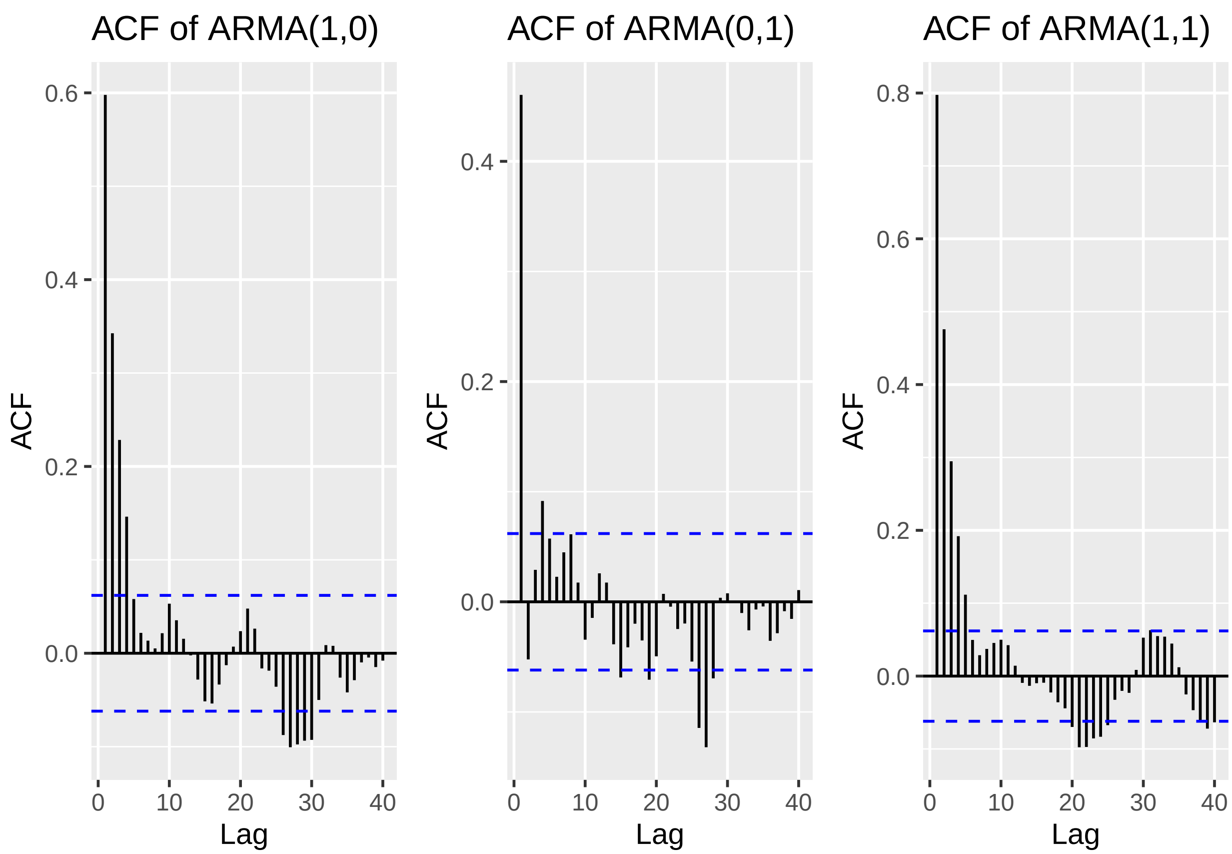
```
plot_acf_01_1000 <- autoplot(acf_01_1000, main = "ACF of ARMA(0,1)")
```

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

```
plot_acf_11_1000 <- autoplot(acf_11_1000, main = "ACF of ARMA(1,1)")
```

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

```
# cowplot acfs
cowplot::plot_grid(plot_acf_10_1000, plot_acf_01_1000, plot_acf_11_1000, nrow = 1)
```




```r
# make pacfs
pacf_10_1000 <- Pacf(arma_10_1000, plot = FALSE, lag.max = 40)
pacf_01_1000 <- Pacf(arma_01_1000, plot = FALSE, lag.max = 40)
pacf_11_1000 <- Pacf(arma_11_1000, plot = FALSE, lag.max = 40)

# plot pacfs
plot_pacf_10_1000 <- autoplot(pacf_10_1000, main = "PACF of ARMA(1,0)")
```

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

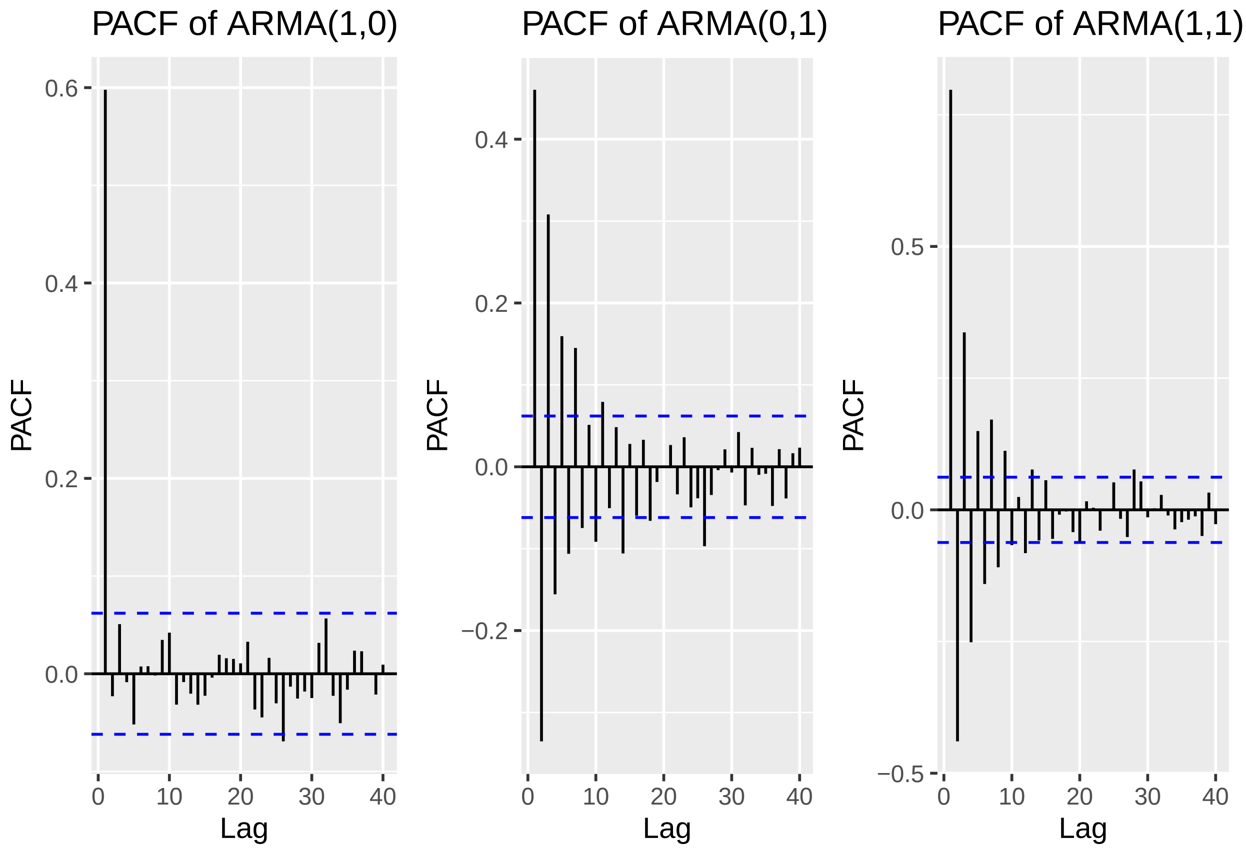
```r
plot_pacf_01_1000 <- autoplot(pacf_01_1000, main = "PACF of ARMA(0,1)")
```

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

```r
plot_pacf_11_1000 <- autoplot(pacf_11_1000, main = "PACF of ARMA(1,1)")
```

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

```r
# cowplot pacfs
cowplot::plot_grid(plot_pacf_10_1000, plot_pacf_01_1000, plot_pacf_11_1000, nrow = 1)
```



Answer:Honestly, with more observations, it makes it easier to recognize. I would be able to determine the ARMA 1,0 from the pacf plot because there is a spike at lag 1 that cuts off after. That would mean the AR(1). There looks to be decay in the acf of the ar 1 plot as well, so I would think that p =1, d = 0, and q = 0. For the ARMA (0,1), I would be able to tell it is MA(1) because the acf cuts off after 1. It also has decay in the pacf plot. For the MA 1 model, it would be p = 0, d = 0 , and q = 1. For the ARMA 1,1, I would be able to determine that this has both components p = 1 and q = 1 (d= 0) because both the acf plot and the pacf plot have decay. They both also have a spike at 1 then it drops off. Once again, the values do not look quite right in terms of meeting the 0.6 and 0.9. I think the ARMA 1,1 model may look correct. I feel like they should match but, maybe in this case it is okay.

Q3

Consider the ARIMA model $y_t = 0.7 * y_{t-1} - 0.25 * y_{t-12} + a_t - 0.1 * a_{t-1}$

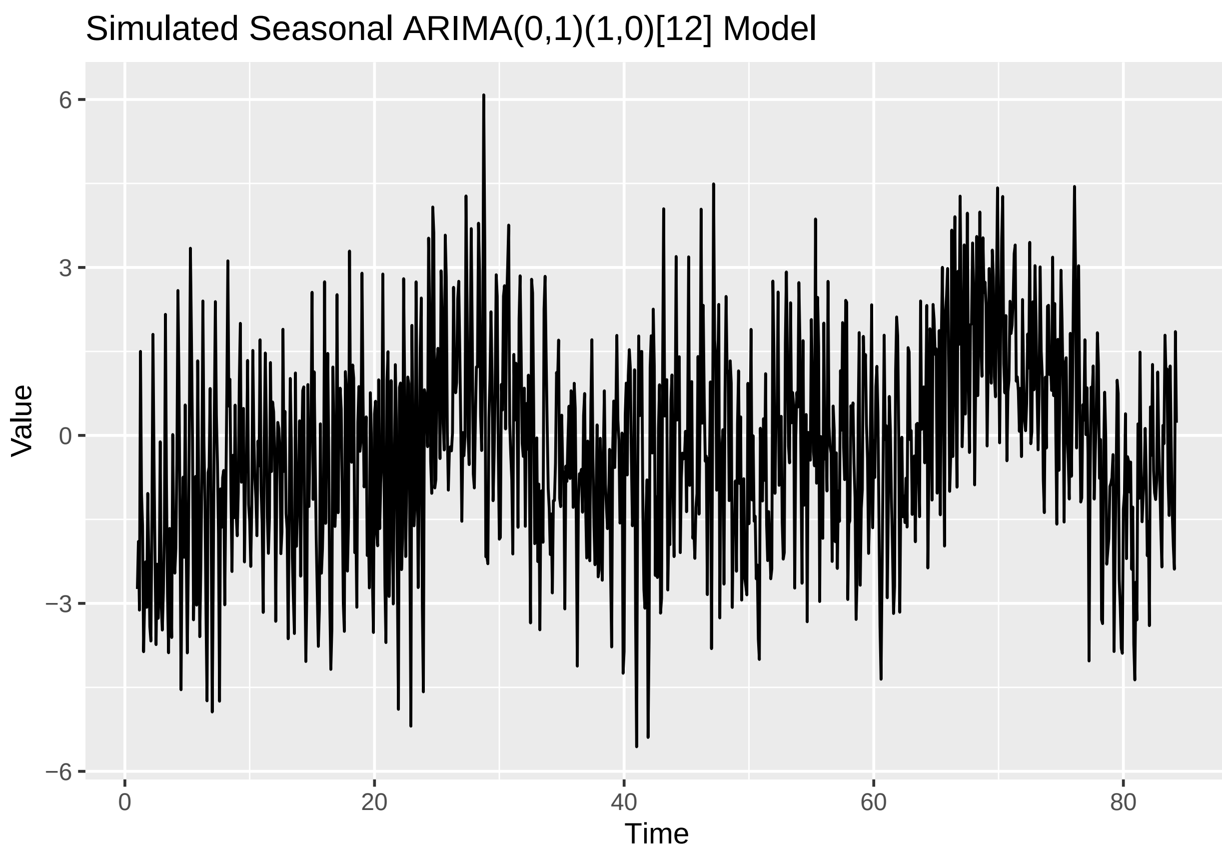
(a) Identify the model using the notation ARIMA$(p, d, q)(P, D, Q)_s$, i.e., identify the integers p, d, q, P, D, Q, s (if possible) from the equation. > p = 1, d = 0, q = 1 > P = 1, D = 0, Q = 0 > s would be 12

(b) Also from the equation what are the values of the parameters, i.e., model coefficients. > AR phi would be 0.7, MA theta would be -0.1, and the sar phi would be -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?

```
# simulate with sim_sarima
sarima_sim <- sim_sarima(n = 1000, model = list(ar = c(rep(0, 11),0.8), ma = 0.5))

ts_sarima <- ts(sarima_sim, frequency = 12)
autoplot(ts_sarima) +
  ggtitle("Simulated Seasonal ARIMA(0,1)(1,0)[12] Model") +
  xlab("Time") +
  ylab("Value")
```

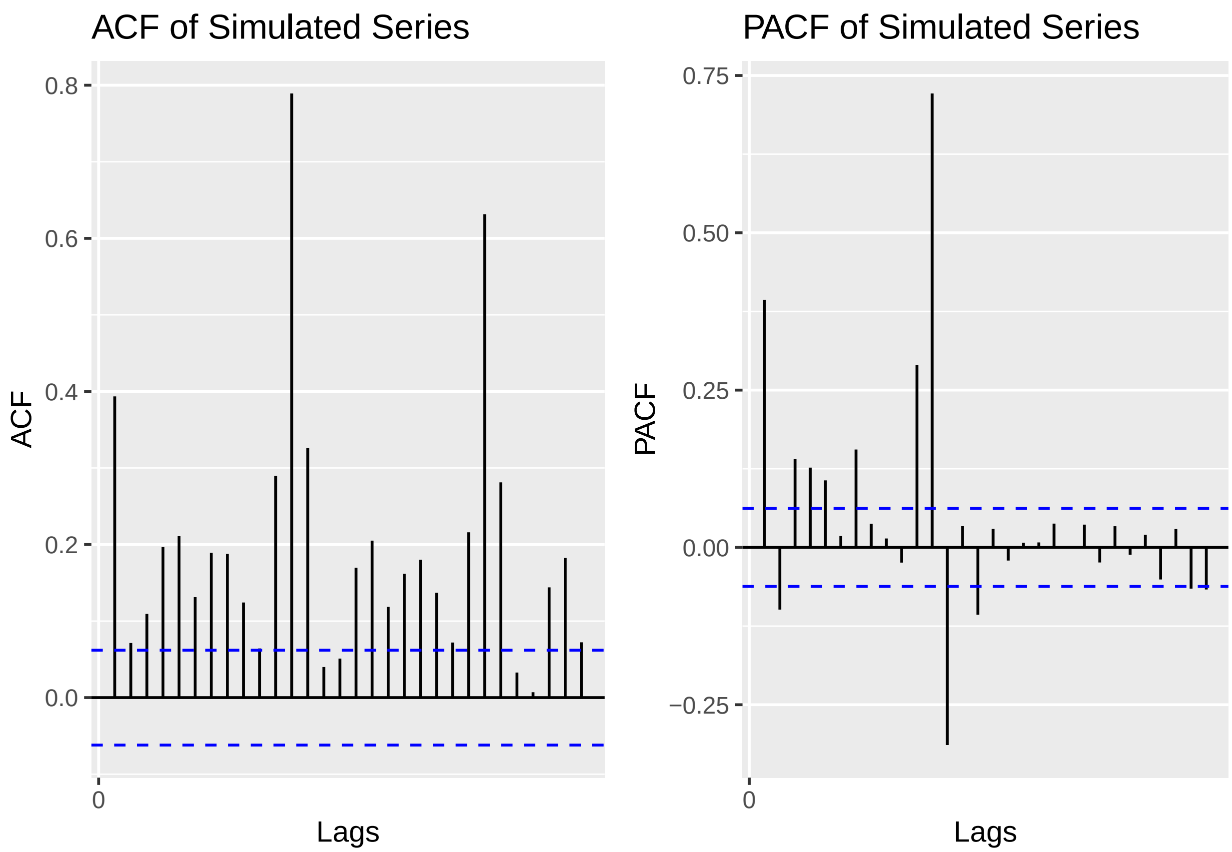


> Yes, there does look to be a seasonal component, as there seems to be somehwat a series of recurring spikes. I don't think it is enough of a trend to say it is explicitly an AR model, because the data looks stagnant at some points (plus I know that this has an MA component). There are some high and low points that seem to recur after certain time intervals though, so this is why I would think this has some 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.

```
acf_plot <- autoplot(acf(ts_sarima, plot = FALSE)) +
  ggtitle("ACF of Simulated Series") +
  xlab("Lags") +
  ylab("ACF")

pacf_plot <- autoplot(pacf(ts_sarima, plot = FALSE)) +
  ggtitle("PACF of Simulated Series") +
  xlab("Lags") +
  ylab("PACF")

sarima_plots <- cowplot::plot_grid(acf_plot, pacf_plot, ncol = 2)

# Display the combined plots
sarima_plots
```



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 the plots do represent the components pretty well. From the acf plot, it shows a large spike every 12 lags, which would demonstrate the seasonal ar component. Also, in the acf plot, it shows a spike at lag 1, which would be the MA(1) component. I am a little confused about the pacf plot, but I think there may be an unseeable lag that would be the 7th lag, and that would make more sense to me because there is a big spike at 1 and then another at what would be the 12th (if that invisible lag is real). I think I would be able to tell that the q = 1 because of the cut off in the acf, and I would be able to tell d = 0. Knowing this is an MA model and because I don't see decay in the pacf, I would say that the p is 0. Then I would see the 12 lag seasonal component in the ACF and be able to determine that Q = 0 and P = 1 (and D is still 0).