ENV 790.30 - Time Series Analysis for Energy Data | Spring 2021

Assignment 6 - Due date 03/16/22

Rob Kravec

```
library(tidyverse)
library(forecast)
library(tseries)
```

Q1

Describe the important characteristics of the sample autocorrelation function (ACF) plot and the partial sample autocorrelation function (PACF) plot for the following models:

(a) AR(2)

> Answer: The AR(2) model will show a slow decay for its ACF plot and have a cutoff on the PACF plot after two lags. By "cutoff," I mean that the first two lags on the PACF plot will have PACF values (much) greater than the dashed lines representing a cutoff based on standard error values.

(b) MA(1)

> Answer: The MA(1) model will have a cutoff on the ACF plot after one lag (which is typically negative) and will show a slow decay for its PACF plot. If the ACF plot shows a value at lag = 0 (with value equal to 1), it is important not to count that lag.

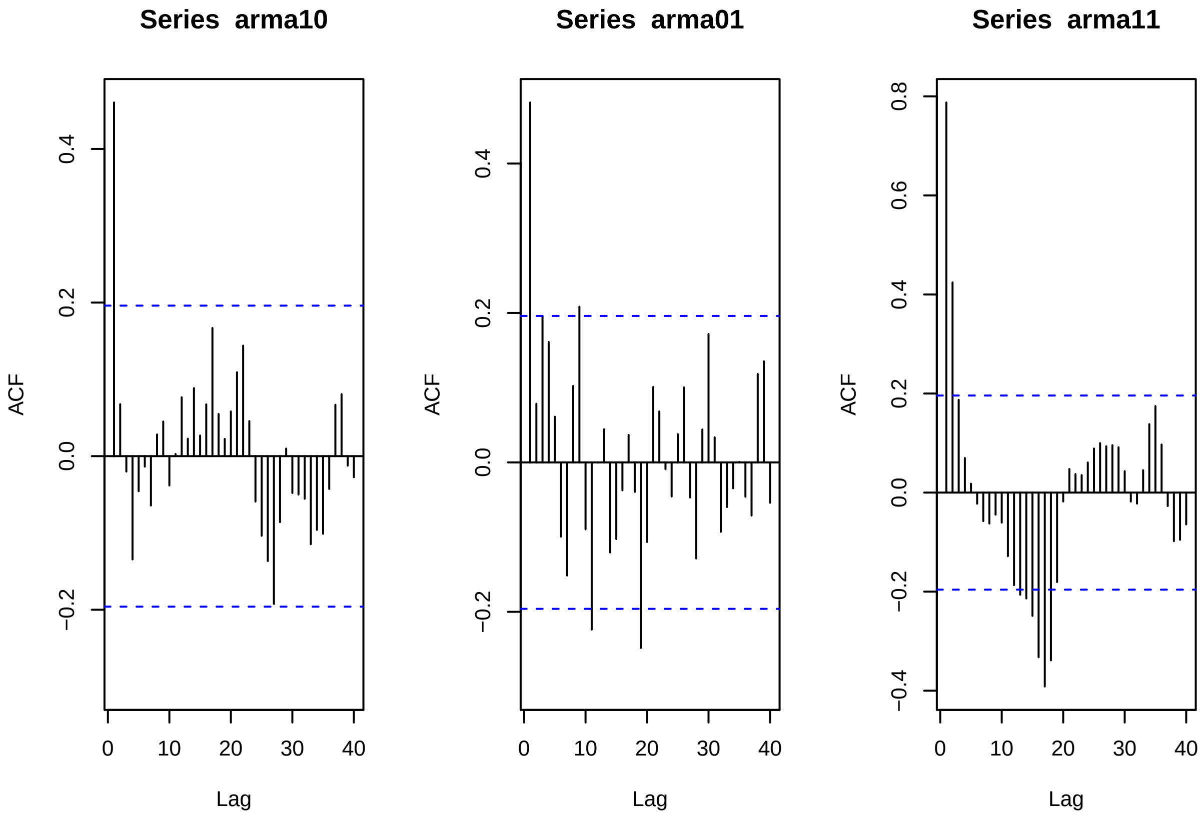
Q2

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

```
# Specify constants
n <- 100
phi <- 0.6
theta <- 0.9

# Simulate data
arma10 <- arima.sim(list(order = c(1, 0, 0), ar = phi), n = n)
arma01 <- arima.sim(list(order = c(0, 0, 1), ma = theta), n = n)
arma11 <- arima.sim(list(order = c(1, 0, 1), ar = phi, ma = theta), n = n)
```

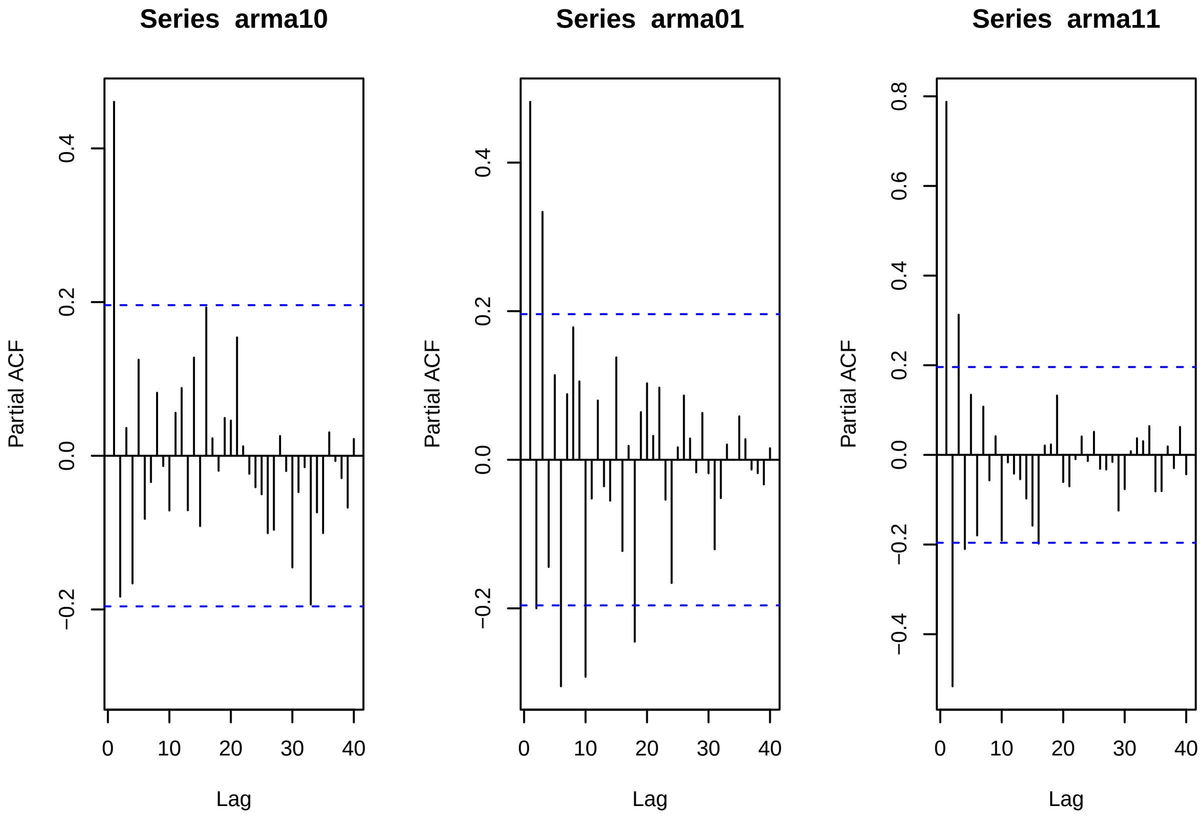
a) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use command $par(mfrow = c(1, 3))$ that divides the plotting window in three columns).

```
par(mfrow = c(1, 3))
acf10 <- Acf(x = arma10, lag.max = 40)
acf01 <- Acf(x = arma01, lag.max = 40)
acf11 <- Acf(x = arma11, lag.max = 40)
```



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

```
par(mfrow = c(1, 3))
pacf10 <- Pacf(arma10, lag.max = 40)
pacf01 <- Pacf(arma01, lag.max = 40)
pacf11 <- Pacf(arma11, lag.max = 40)
```



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: I think I'd be able to identify the ARMA(1, 0) and ARMA(0, 1) correctly. The ARMA(1, 0) plot has a clear PACF cutoff at lag 1, while the ARMA(0, 1) plot has a (relatively) clear ACF cutoff at lag 1. Perhaps the seasonality present in the ARMA(0, 1) ACF plot would have thrown me off. The ARMA(1, 1) is tricky because it contains both AR and MA behavior, and I don't know that I would have been able to identify the order of each component by sight.

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

```r
rho <- acf01$acf[2]
theta <- (1 - sqrt(1 - 4 * rho ^ 2)) / (2 * rho)

# Print computed values
print(paste0("Computed phi: ", round(pacf10$acf[1], 4),
             " ... Computed theta: ", round(theta, 4)))
```

```
## [1] "Computed phi: 0.4606 ... Computed theta: 0.7602"
```

Answer: We would expect the ARMA(1, 0) PACF value at lag 1 to be equal to ϕ (0.6). For the ARMA(0, 1) data, we would expect the ACF at lag 1 to be equal to $\frac{\theta}{1+\theta^2}$. From there, we can solve for θ using the quadratic formula: $\theta = \frac{1\pm\sqrt{1-4\rho^2}}{2\rho}, andweexpectavalueof 0.9$

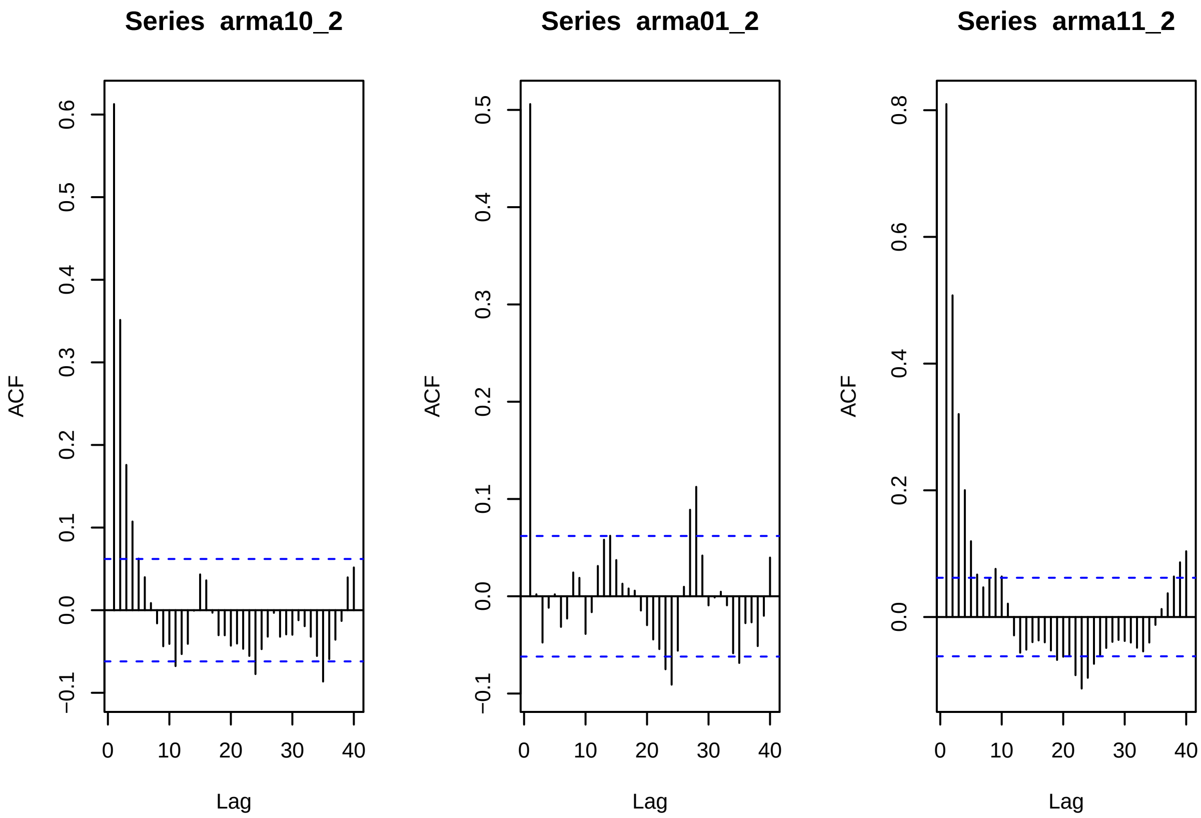
Citations:

- https://mcs.utm.utoronto.ca/~nosedal/sta457/pacf.pdf

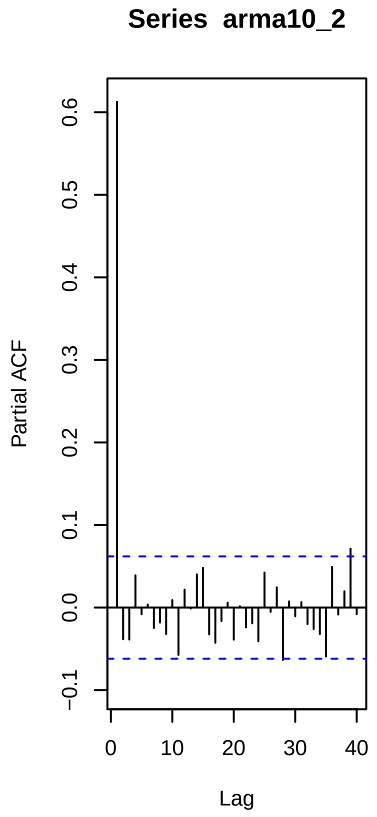
- https://online.stat.psu.edu/stat510/book/export/html/662

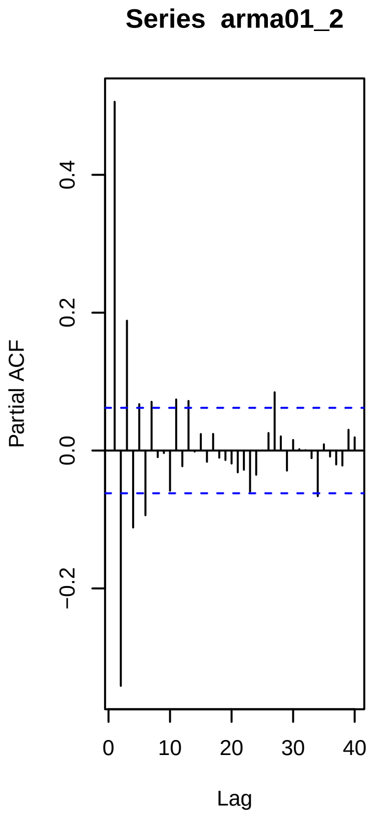
e) Increase number of observations to $n = 1000$ and repeat parts (a)-(d).

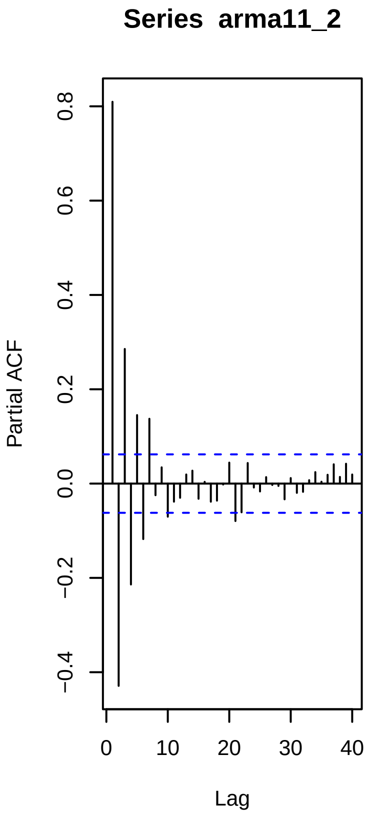
```
# Change n
n <- 1000

# Simulate data
arma10_2 <- arima.sim(list(order = c(1, 0, 0), ar = phi), n = n)
arma01_2 <- arima.sim(list(order = c(0, 0, 1), ma = theta), n = n)
arma11_2 <- arima.sim(list(order = c(1, 0, 1), ar = phi, ma = theta), n = n)

# Plot ACF
par(mfrow = c(1, 3))
acf10_2 <- Acf(x = arma10_2, lag.max = 40)
acf01_2 <- Acf(x = arma01_2, lag.max = 40)
acf11_2 <- Acf(x = arma11_2, lag.max = 40)
```



```
# Plot PACF
par(mfrow = c(1, 3))
pacf10_2 <- Pacf(arma10_2, lag.max = 40)
pacf01_2 <- Pacf(arma01_2, lag.max = 40)
pacf11_2 <- Pacf(arma11_2, lag.max = 40)
```







```
# Print computed values
rho_2 <- acf01_2$acf[2]
theta_2 <- (1 - sqrt(1 - 4 * rho_2 ^ 2)) / (2 * rho_2)
print(paste0("Computed phi: ", round(pacf10_2$acf[1], 4),
             " ... Computed theta: ", round(theta_2, 4)))
```

```
## [1] "Computed phi: 0.6127 ... Computed theta: NaN"
```

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.

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

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.