Homework6

Hsueh-Pin Liu

2023-04-12

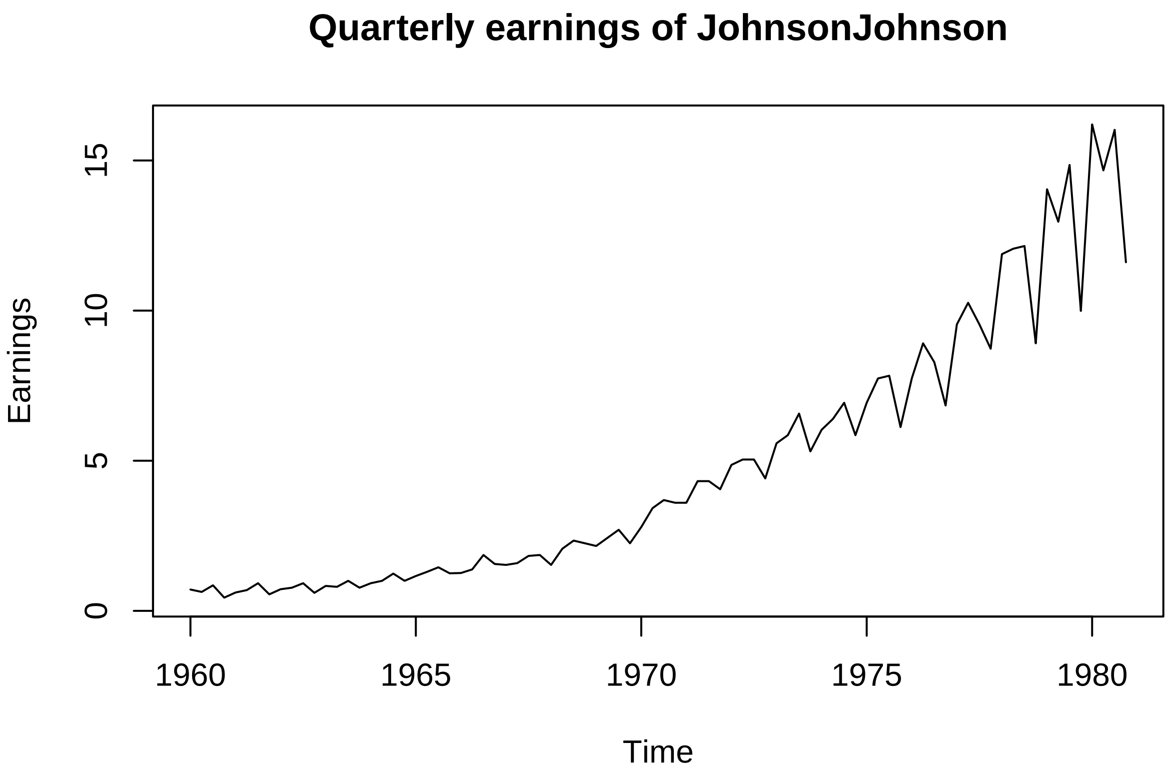
Problem 1

Consider the Johnson and Johnson quarterly earnings data from January 1960 to December 1980 (R data JohnsonJohnson)

a. Plot the data. Describe the features of the data. Do the data look stationary? Explain your answer.
b. Apply an appropriate variance stabilizing transformation, if necessary.
c. Carry out classical decomposition of the data, plot the transformed series along with the ACF and PACF.
d. Identify an ARMA model for the transformed data.
e. Repeat c and d, but instead of classical decomposition, use differencing to make the data stationary.

(a)

```
plot.ts(JohnsonJohnson,main="Quarterly earnings of JohnsonJohnson",ylab="Earnings")
```

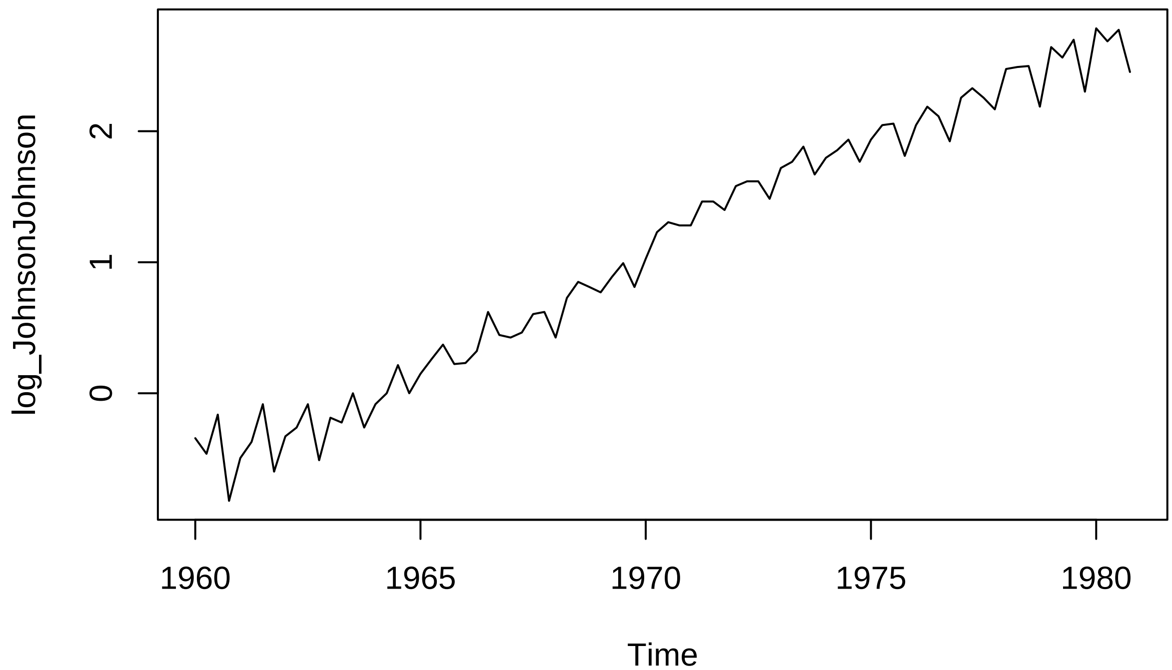


The plot has a increasing trend, so I think it's not stationary.

(b)

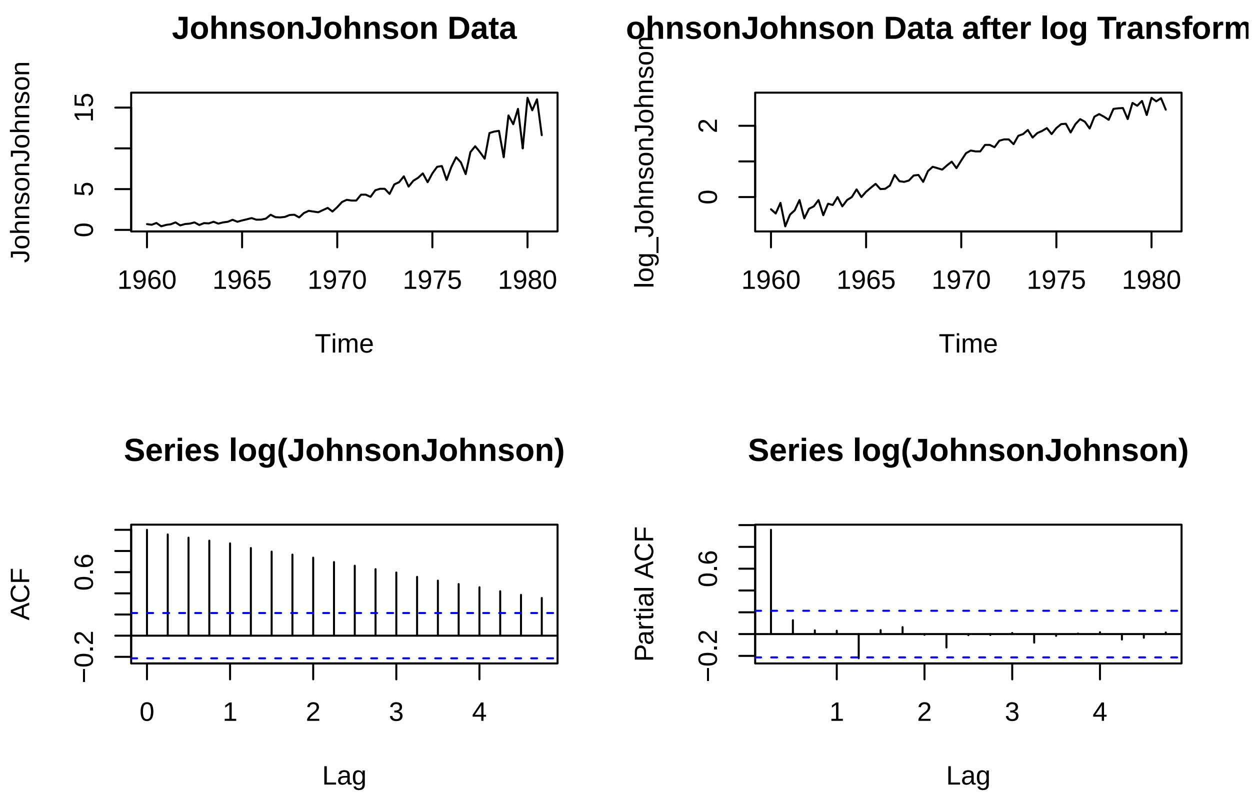
Since the data is not stationary, it's necessary to apply an appropriate variance stabilizing transformation, I use log transformation here.

```
log_JohnsonJohnson <- log(JohnsonJohnson)
plot.ts(log_JohnsonJohnson)
```



(c)

```
par(mfrow=c(2,2))
plot.ts(JohnsonJohnson,main="JohnsonJohnson Data")
plot.ts(log_JohnsonJohnson, main="JohnsonJohnson Data after log Transformation")
acf(log_JohnsonJohnson,main="Series log(JohnsonJohnson)")
pacf(log_JohnsonJohnson,main="Series log(JohnsonJohnson)")
```



ACF decays to zero very slowly and pacf is close to 1 at lag 1.

(d)

It's hard to judge the ARMA model by only the ACF and PACF plot, so I decide to check the model in R.

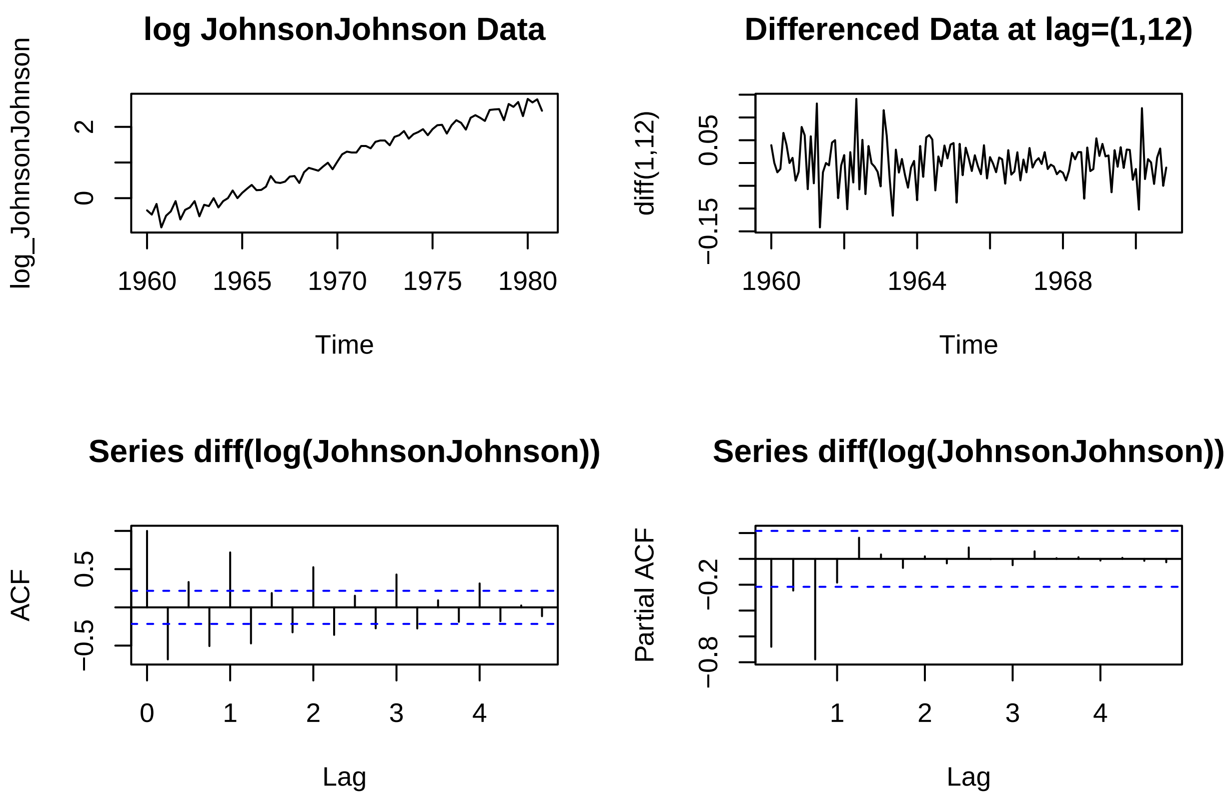
```
x <- decompose(log_JohnsonJohnson)$random
x <- x[!is.na(x)]
bestAICC <- Inf
best_pq <- c(0,0)
for(p in 0:4){
  for(q in 0:4){
    tempAICC <- arima(x, order=c(p,0,q), include.mean=F)$aic
    if(tempAICC < bestAICC){
      bestAICC <- tempAICC
      best_pq <- c(p,q)
    }
  }
}
print(paste0('The best AICC value of ', bestAICC, ' occurs for an ARMA(',best_pq[1],',',best_pq[2],') m
```

```
## [1] "The best AICC value of -275.585180950988 occurs for an ARMA(4,1) model."
```

So I guess it's an ARMA(4,1) model.

(e)

```r
par(mfrow=c(2,2))
plot(log_JohnsonJohnson,main="log JohnsonJohnson Data")
plot(ts(diff(diff(log(AirPassengers)),lag=12),frequency=12,start=c(1960,1)),
        ylab="diff(1,12)",main="Differenced Data at lag=(1,12)")
acf(diff(diff(log_JohnsonJohnson)),main="Series diff(log(JohnsonJohnson))")
pacf(diff(diff(log_JohnsonJohnson)),main="Series diff(log(JohnsonJohnson))")
```



According to the plots, I guess it's an AR(3) model.

Problem 2

Consider the time series of the numbers of users connected to the Internet through a server every minute (R data WWWusage). Carry out a test for unit root. Apply necessary transformation and identify plausible ARMA models.

```
library(tseries)
```

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

```
adf.test(WWWusage)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  WWWusage
## Dickey-Fuller = -2.6421, Lag order = 4, p-value = 0.3107
## alternative hypothesis: stationary
```

So we can't reject the null hypothesis, which means that the time series has a unit root, so we need to provide some transformation.

```
adf.test(diff(WWWusage))
```

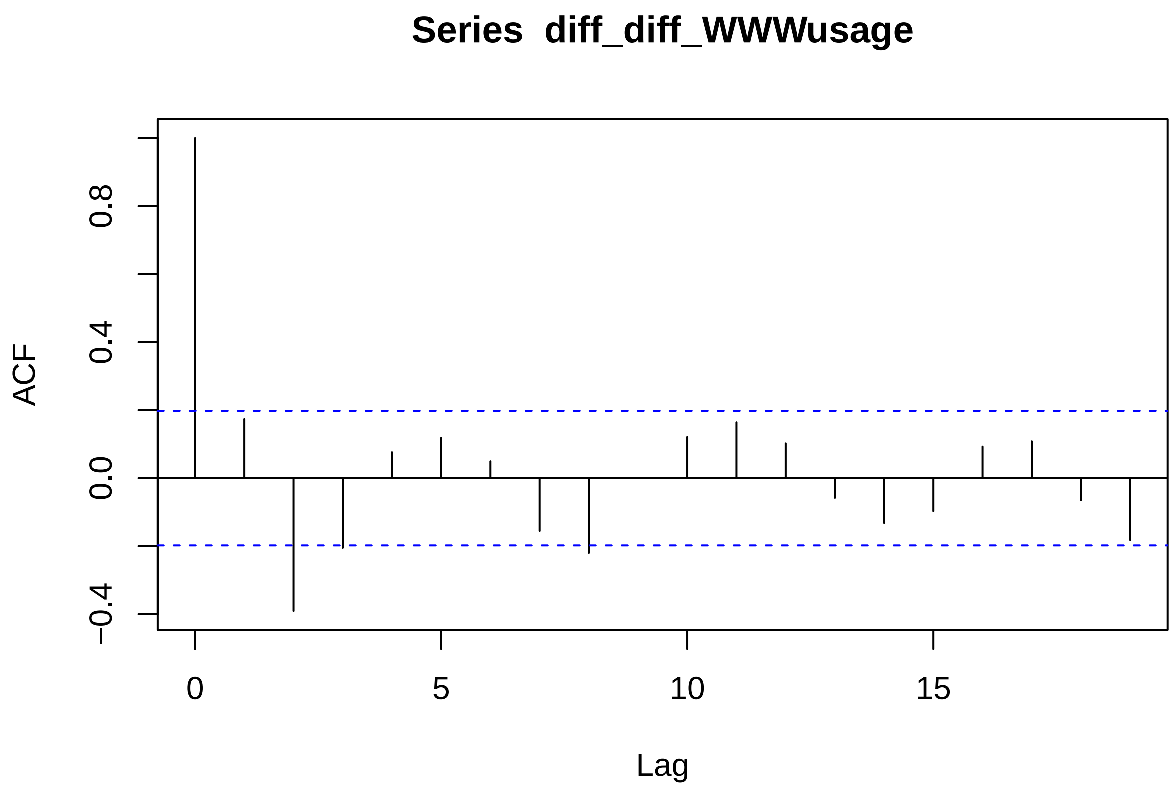
```
##
##  Augmented Dickey-Fuller Test
##
## data:  diff(WWWusage)
## Dickey-Fuller = -2.5459, Lag order = 4, p-value = 0.3506
## alternative hypothesis: stationary
```

So differencing for only one time is not enough, so I difference 2 times.

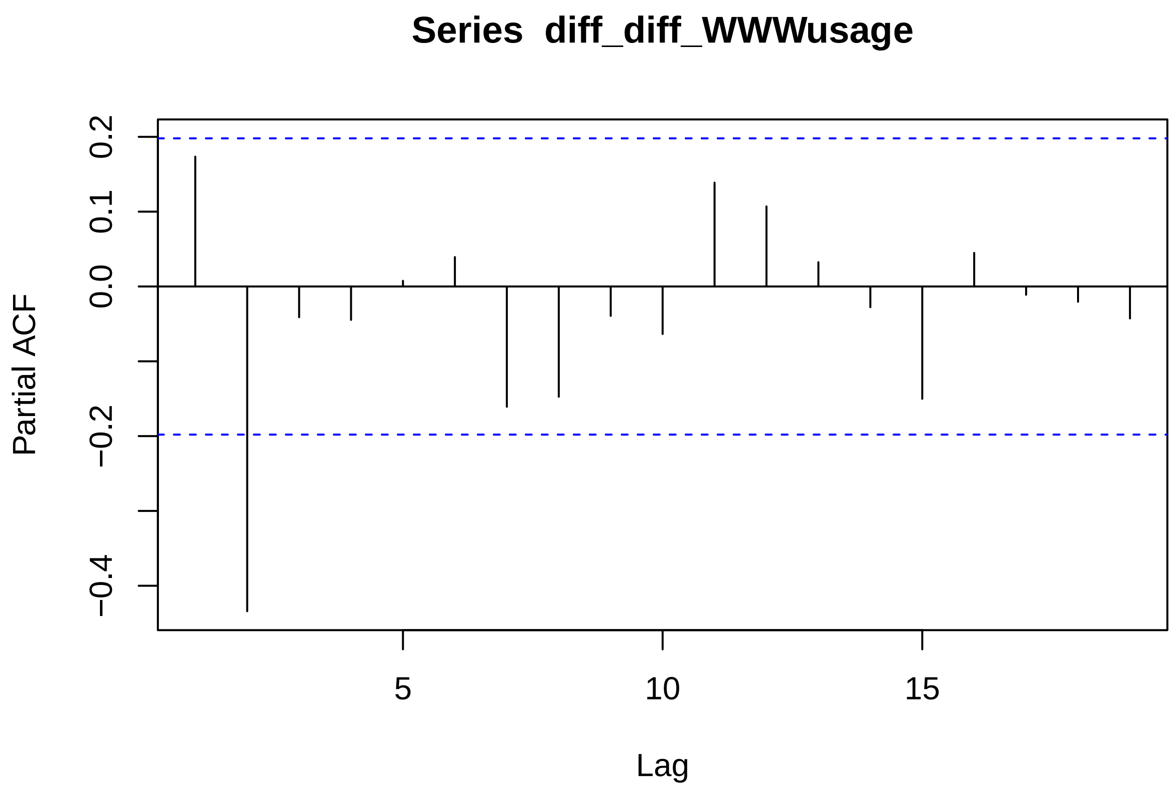
```
diff_diff_WWWusage <- diff(diff(WWWusage))
adf.test(diff_diff_WWWusage)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  diff_diff_WWWusage
## Dickey-Fuller = -4.828, Lag order = 4, p-value = 0.01
## alternative hypothesis: stationary
```

```
acf(diff_diff_WWWusage)
```




```
pacf(diff_diff_WWWusage)
```



Now we can reject the null hypothesis and the time series doesn't have a unit root now.

Problem 3

Identify each of the models below as ARIMA(p, d, q). Specify the order of the models (p, d, q) and the model parameters ϕ and θ:

$a.X_t = 10 + X_{t-1} + e_t + 0.6e_{t-1}$

$b.X_t = 3 + 1 : 25X_{t-1} - 0 : 25X_{t-2} + e_t - 0.2e_{t-2}$

$c.X_t = 1.7X_{t-1} + 0.7X_t$

3\. a. $X_t = 10 + X_{t-1} + e_t + 0.6e_{t-1}$

$(1-B)X_t = (1+0.6B)e_t + 10$

So it's an ARIMA (0,1,1) model, $\theta = 0.6$.

b. $X_t = 3 + 1.25X_{t-1} - 0.25X_{t-2} + e_t - 0.2e_{t-2}$

$(1-0.25B)(1-B)X_t = (1-0.2B^2)e_t + 3$

So it's an ARIMA(1,1,2) model, $\phi = 0.25$ and $\theta_1 = 0$, $\theta_2 = -0.2$.

c. $X_t - 1.7X_{t-1} + 0.7X_{t-2} = -8 + e_t$

$(1-0.7B)(1-B)X_t = e_t - 8$

So it's an ARIMA(1,1,0) model, $\phi = 0.7$.

Figure 1: Problem 3

Problem 4

Consider the two models $X_t = 0.9X_{t-1} + 0.09X_{t-2} + e_t$ and $X_t = X_{t-1} + e_t - 0.1e_{t-1}$

a. Identify both models as ARIMA(p, d, q). Specify (p, d, q) and ARMA parameters ϕ and θ.

b. In what way the two models are different?

c. In what way the two models are similar? What does this tell you about model selection for time series data?

4. (a) $X_t = 0.9X_{t-1} + 0.09X_{t-2} + e_t$ $\qquad$ $X_t = X_{t-1} + e_t - 0.1e_{t-1}$

$(1 - 0.9B - 0.09B^2)X_t = e_t$ $\qquad$ $(1-B)X_t = (1-0.1B)e_t$

So it's an ARIMA(2,0,0) model, $\qquad$ So it's an ARIMA(0,1,1) model,

and $\phi_1 = 0.9$, $\phi_2 = 0.09$. $\qquad$ and $\theta = -0.1$.

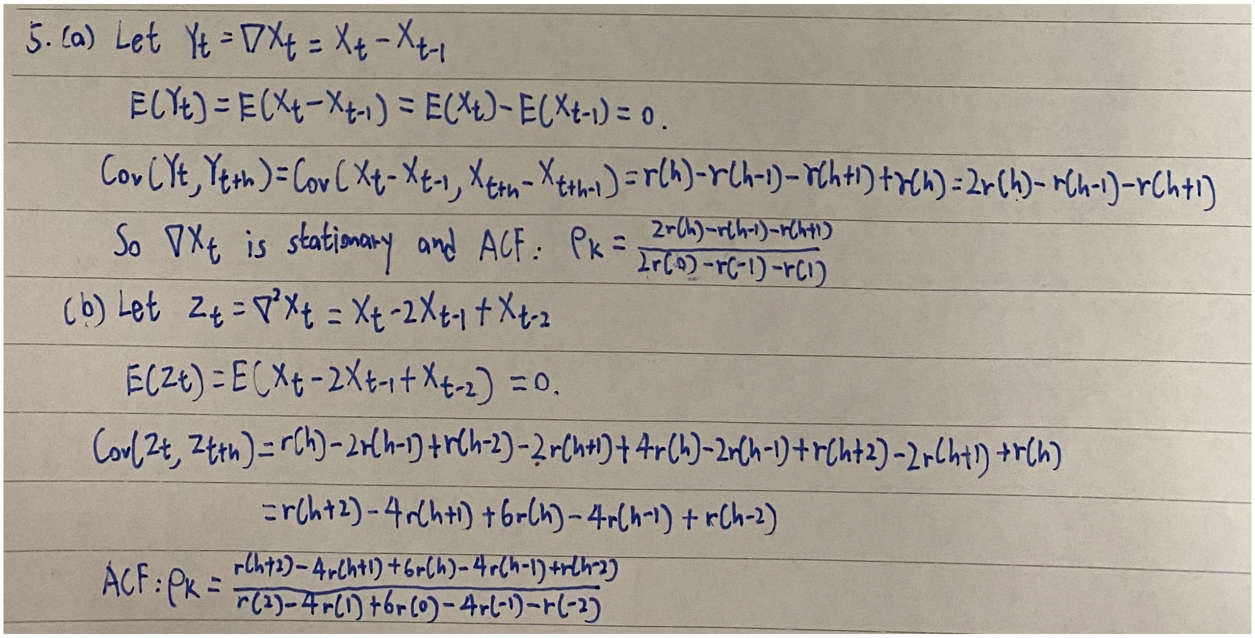
(b) The first one is causal while the second model is not causal.

(c) ②: $X_t = 0.9X_{t-1} + 0.09X_{t-2} + 0.01X_{t-3} + 0.01e_{t-2} - 0.001e_{t-3} + e_t - 0.1e_{t-1}$, almost the same as ①.

Figure 2: Problem 4

Problem 5

Let X_t be a stationary process with autocovariance function (h): a. Show that the process ∇X_t is stationary and find its autocovariance function. b. Show that the process $\nabla^2 X_t$ is also stationary.



5\. (a) Let $Y_t = \nabla X_t = X_t - X_{t-1}$

$E(Y_t) = E(X_t - X_{t-1}) = E(X_t) - E(X_{t-1}) = 0.$

$Cov(Y_t, Y_{t+h}) = Cov(X_t - X_{t-1}, X_{t+h} - X_{t+h-1}) = r(h) - r(h-1) - r(h+1) + r(h) = 2r(h) - r(h-1) - r(h+1)$

So ∇X_t is stationary and ACF: $\rho_k = \frac{2r(h) - r(h-1) - r(h+1)}{2r(0) - r(-1) - r(1)}$

(b) Let $Z_t = \nabla^2 X_t = X_t - 2X_{t-1} + X_{t-2}$

$E(Z_t) = E(X_t - 2X_{t-1} + X_{t-2}) = 0.$

$Cov(Z_t, Z_{t+h}) = r(h) - 2r(h-1) + r(h-2) - 2r(h+1) + 4r(h) - 2r(h-1) + r(h+2) - 2r(h+1) + r(h)$

$= r(h+2) - 4r(h+1) + 6r(h) - 4r(h-1) + r(h-2)$

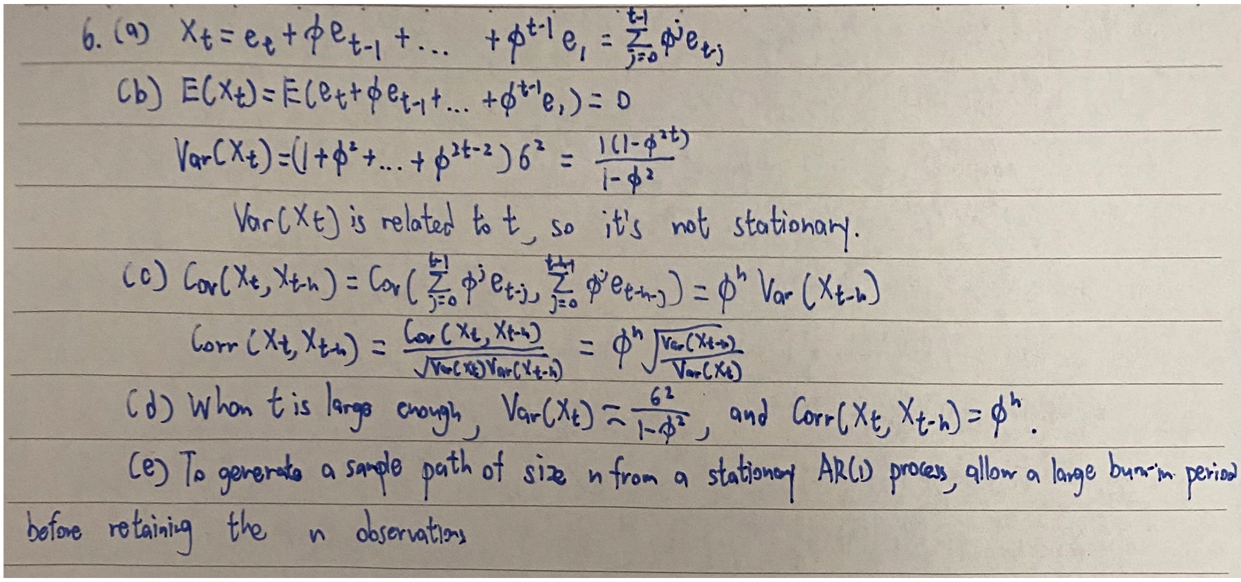
ACF: $\rho_k = \frac{r(h+2) - 4r(h+1) + 6r(h) - 4r(h-1) + r(h-2)}{r(2) - 4r(1) + 6r(0) - 4r(-1) - r(-2)}$

Figure 3: Problem 5

Problem 6

Let et be zero mean Gaussian white noise process with variance σ^2 and let $|\phi| < 1$ be a constant. Consider the process, starting at X_1, $X_1 = e_1$, $X_t = \phi X_{t-1} + e_t, t = 2, 3, ...$

a. Express X_t as a linear combination of the white noise process e_t.
b. Use the result in (a) to compute the mean and the variance of the process X_t. Is the process X_t stationary?
c. Show $Correlation(X_t, X_{t-h}) = \phi^h [\frac{Var(X_{t-h})}{Var(X_t)}]^{\frac{1}{2}}$ for $h \geq 0$
d. Argue that for large t, $Var(X_t) \approx \frac{\sigma^2}{1-\phi^2}$ and $Correlation(X_t, X_{t-h}) \approx \phi^h, h \geq 0$, so in a sense, X_t is "asymptotically stationary."
e. This result can be used to simulate observations from a stationary Gaussian AR(1) model. Explain how this can be done.
f. Write a R code generate a random sample of size 500 from the AR(1) process with $\phi = 0.6$ and $\sigma^2 = 0.8$. Plot the simulated series along with the sample ACF and PACF of the series. Is the sample ACF and PACF consistent with AR(1)?



6. (a) $X_t = e_t + \phi e_{t-1} + \dots + \phi^{t-1} e_1 = \sum_{j=0}^{t-1} \phi^j e_{t-j}$

(b) $E(X_t) = E(e_t + \phi e_{t-1} + \dots + \phi^{t-1} e_1) = 0$

$Var(X_t) = (1+\phi^2+\dots+\phi^{2t-2})\sigma^2 = \frac{1(1-\phi^{2t})}{1-\phi^2}$

$Var(X_t)$ is related to t, so it's not stationary.

(c) $Cov(X_t, X_{t-h}) = Cov\left(\sum_{j=0}^{t-1}\phi^j e_{t-j}, \sum_{j=0}^{t-h-1}\phi^j e_{t-h-j}\right) = \phi^h Var(X_{t-h})$

$Corr(X_t, X_{t-h}) = \frac{Cov(X_t, X_{t-h})}{\sqrt{Var(X_t)Var(X_{t-h})}} = \phi^h \sqrt{\frac{Var(X_{t-h})}{Var(X_t)}}$

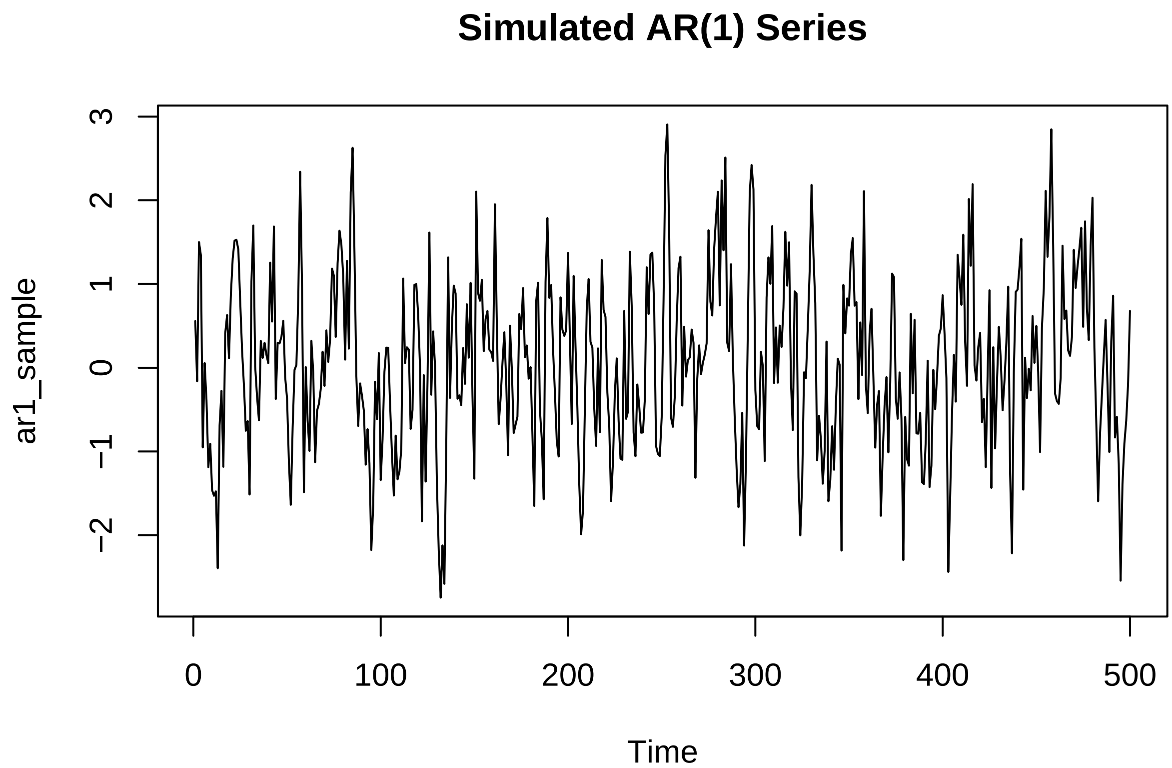
(d) When t is large enough, $Var(X_t) \approx \frac{\sigma^2}{1-\phi^2}$, and $Corr(X_t, X_{t-h}) = \phi^h$.

(e) To generate a sample path of size n from a stationary AR(1) process, allow a large burn-in period before retaining the n observations

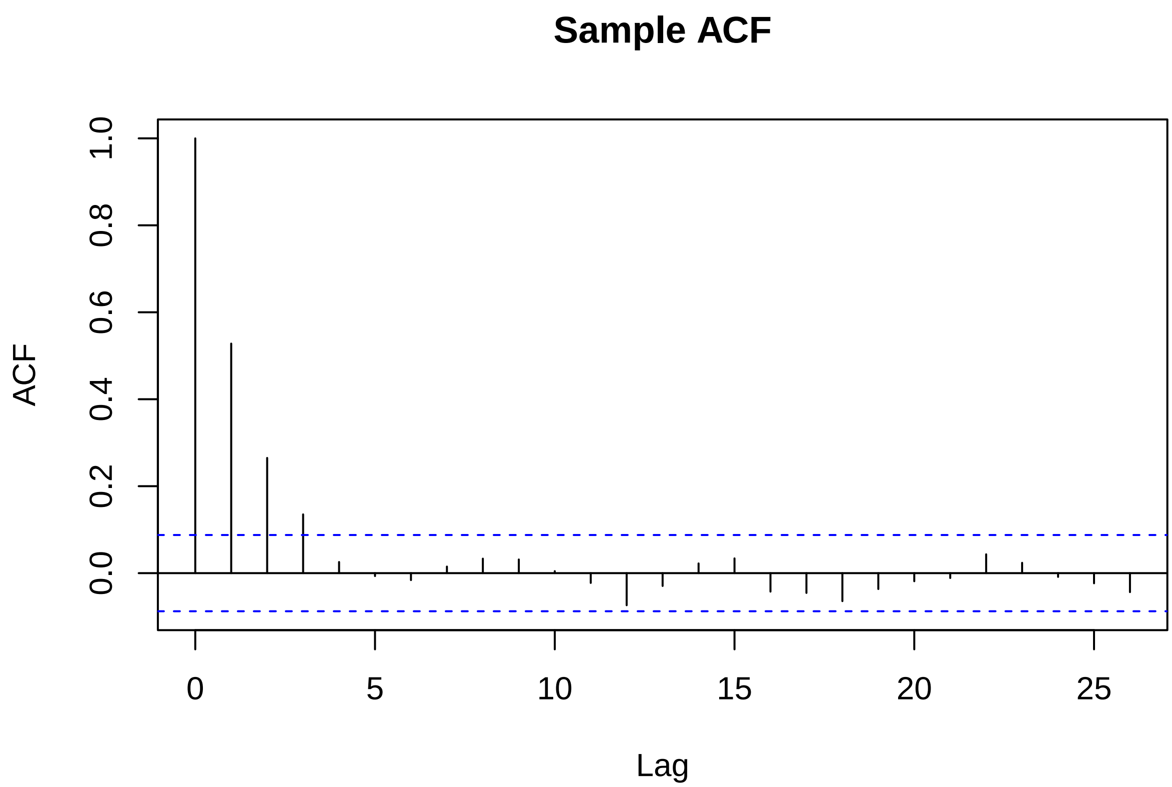
Figure 4: Problem 6

(f)

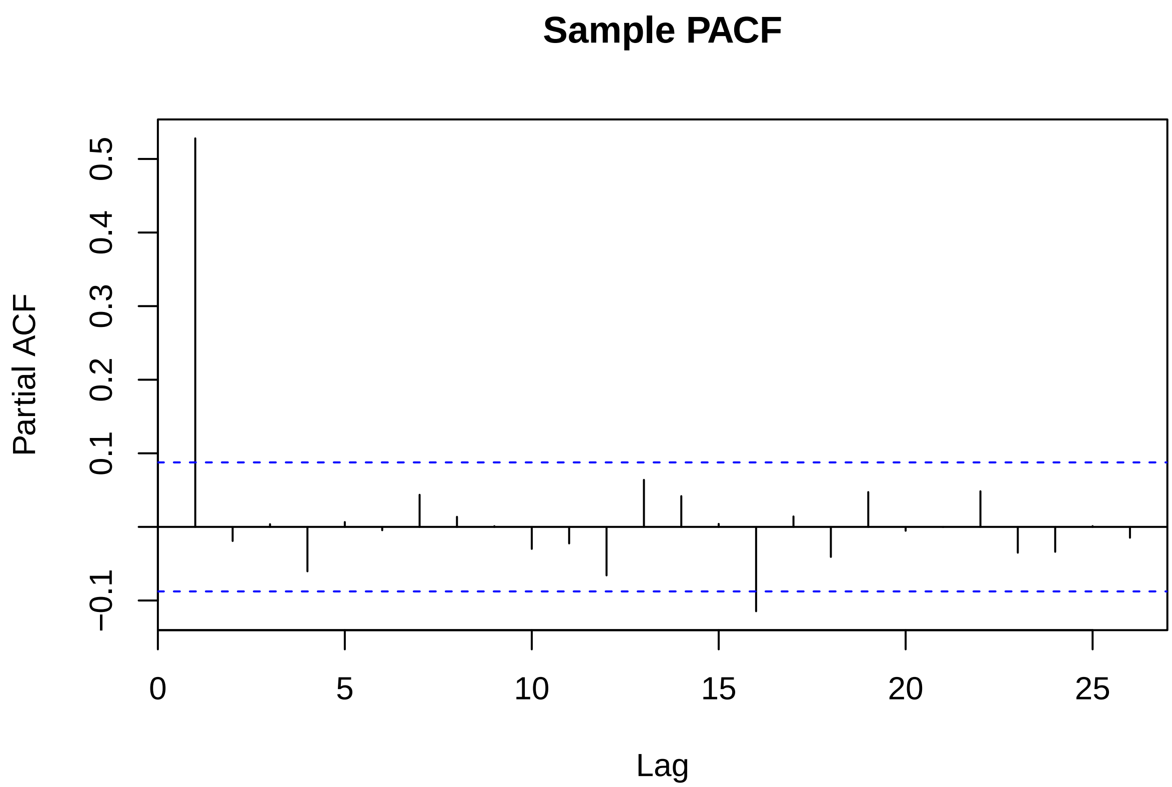
```r
set.seed(123) # for reproducibility
n <- 500
phi <- 0.6
sigma2 <- 0.8
ar1_sample <- arima.sim(model = list(ar = phi), n = n, sd = sqrt(sigma2))
plot(ar1_sample, type = "l", main = "Simulated AR(1) Series")
```




```
acf(ar1_sample, main = "Sample ACF")
```




```
pacf(ar1_sample, main = "Sample PACF")
```



According to the sample ACF and PACF plot, ACF decays to 0 while PACF is significant at lag 1, so it's consistent with AR(1).