STA457 A1

Jiayue Wu

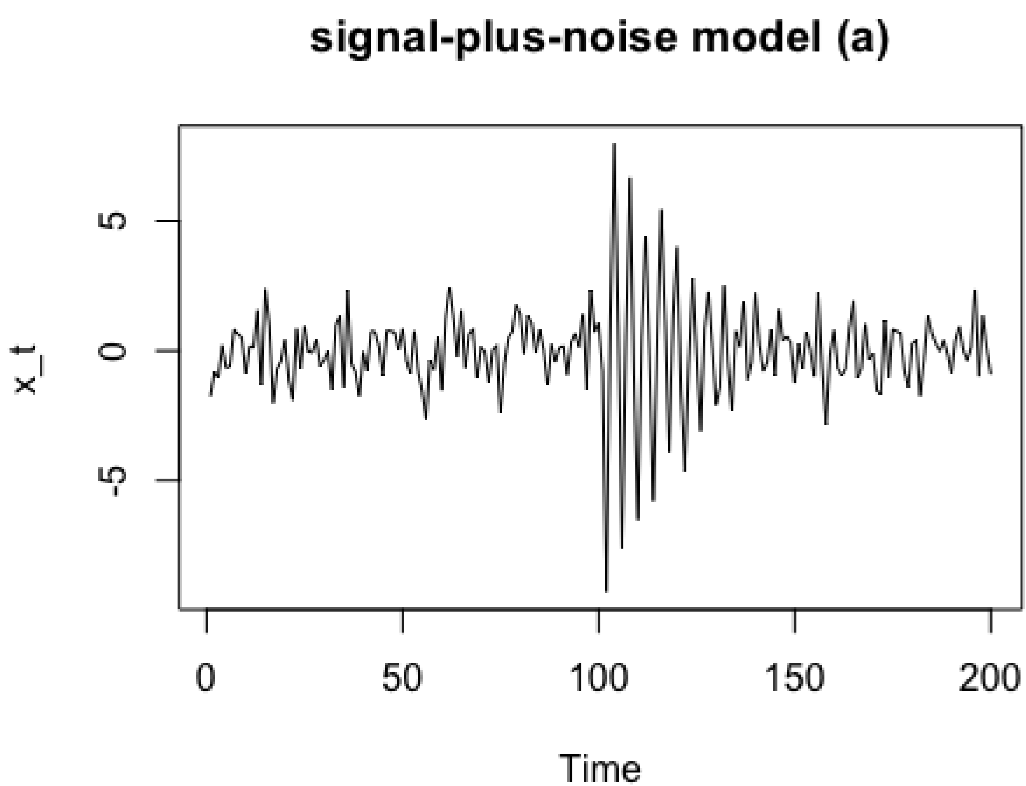
1/25/2021

```
library(astsa)
```

Question 1

model (a)

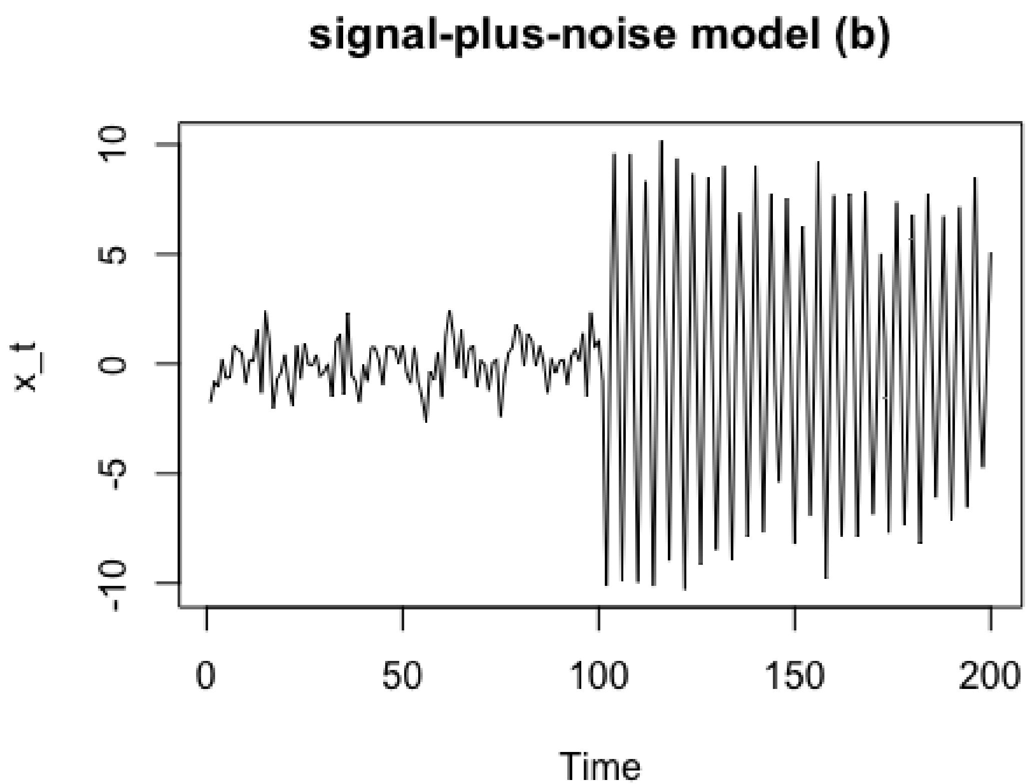
```
set.seed(1004768165) #set my student number to be the seed to ensure get the s
ame result every time I run
w_t <- rnorm(200) # get 200 Gaussian white noise
s_t <- c(rep(0,100), 10*exp(-(1:100)/20)*cos(2*pi*(101:200)/4)) # get s_t as q
uestion needed
x_t <- s_t + w_t # add s_t and w_t together to get x_t
plot.ts(x_t, main = "signal-plus-noise model (a)") # plot the graph(a)
```



(b)

model (b)

```
set.seed(1004768165) # set seed to get same result every time I run
w_t <- rnorm(200) # get 200 Gaussian white noise
s_t <- c(rep(0,100), 10*exp(-(1:100)/200)*cos(2*pi*(101:200)/4)) # get s_t wit
h denominator 200
x_t <- s_t + w_t # add s_t and w_t together to get x_t
plot.ts(x_t, main = "signal-plus-noise model (b)") # plot the graph(b)
```

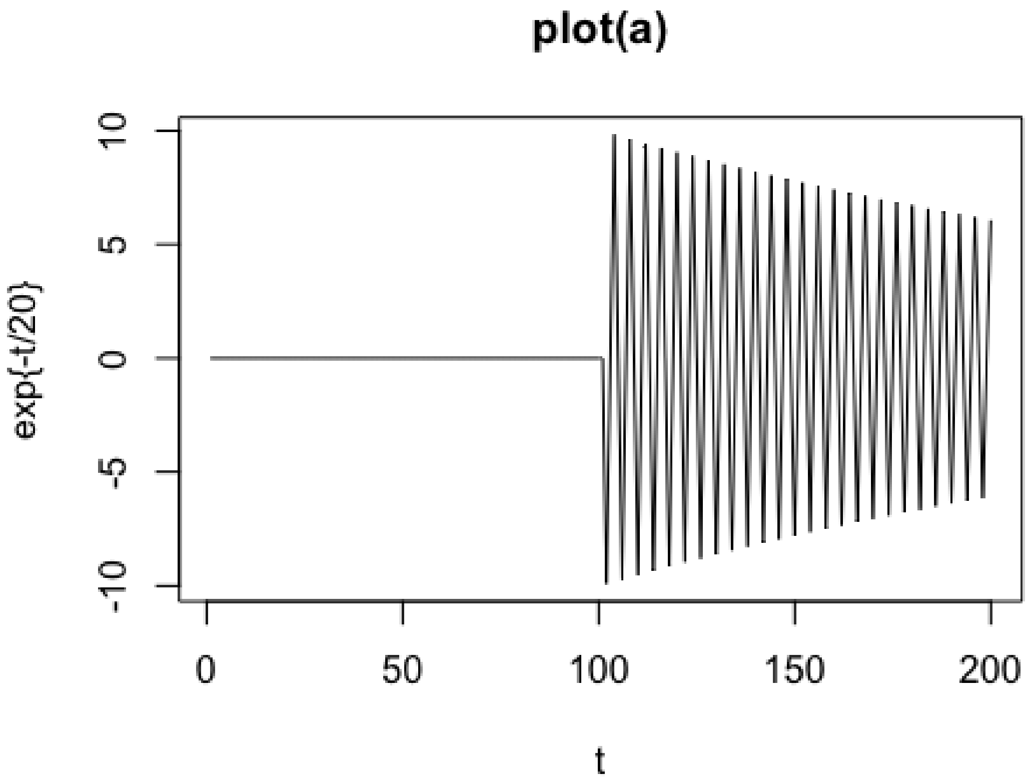


(c)
As we can see, the plots of model (a) and model (b) have similar behavior on the first half, their amplitude are relatively smooth, that should be due to they have same randomly distribution for t = 0:100, i.e. they all follow normal standard normal distribution for the first half time.

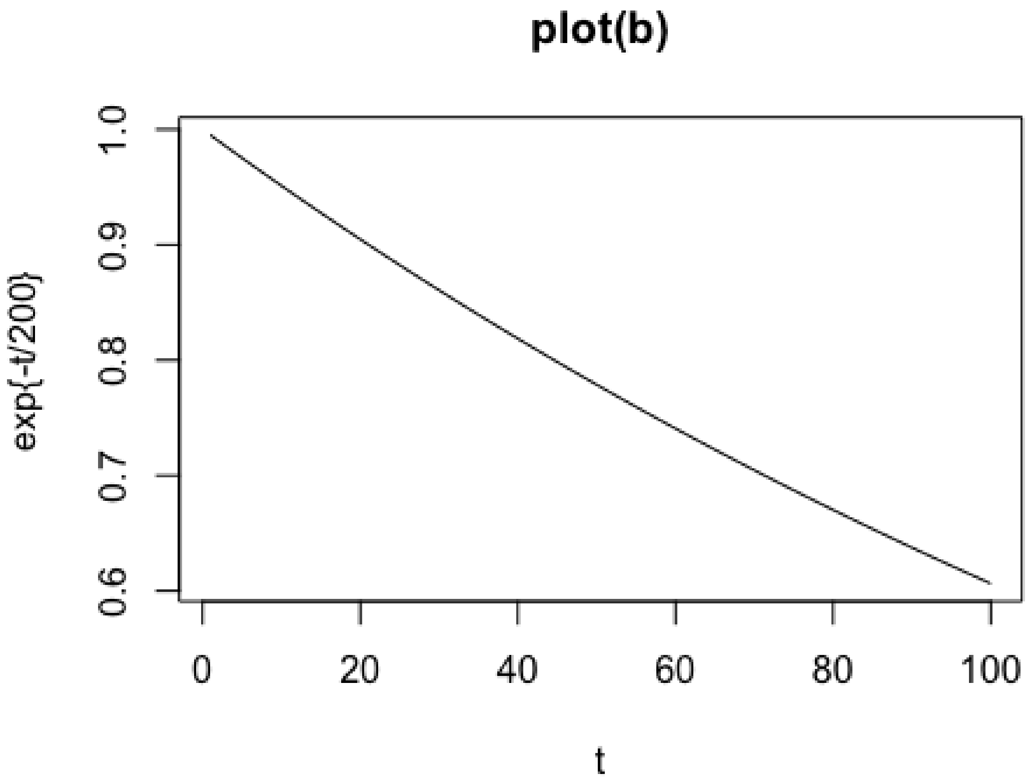
However, after t = 100, we can find that the model (a) has rather large amplitude when t = 100 begin, and gradually decrease to as the amplitude before t = 100, and it is noticeable that explosion has similar behavior, i.e. amplitude greatly reduced compared to peak.

Furthermore, we can find that the largest difference between model(a) and model (b) is that the amplitude of model (b) has changed relatively small at the end compared to the peak. I conclude that model (b) is more like earthquake, since the amplitude do not change too much for the later half time(i.e. after t = 100).

```
sig_t <- c(exp(-(1:100)/20)) # signal signal modulators for (a)
plot.ts(s_t, xlab = "t", ylab = "exp{-t/20}",main = "plot(a)") # plot the sign
al modulators
```




```
s_t <- c(exp(-(1:100)/200)) # signal modulators for (b)
plot.ts(s_t, xlab = "t", ylab = "exp{-t/200}",main = "plot(b)") # plot the sig
nal modulators
```

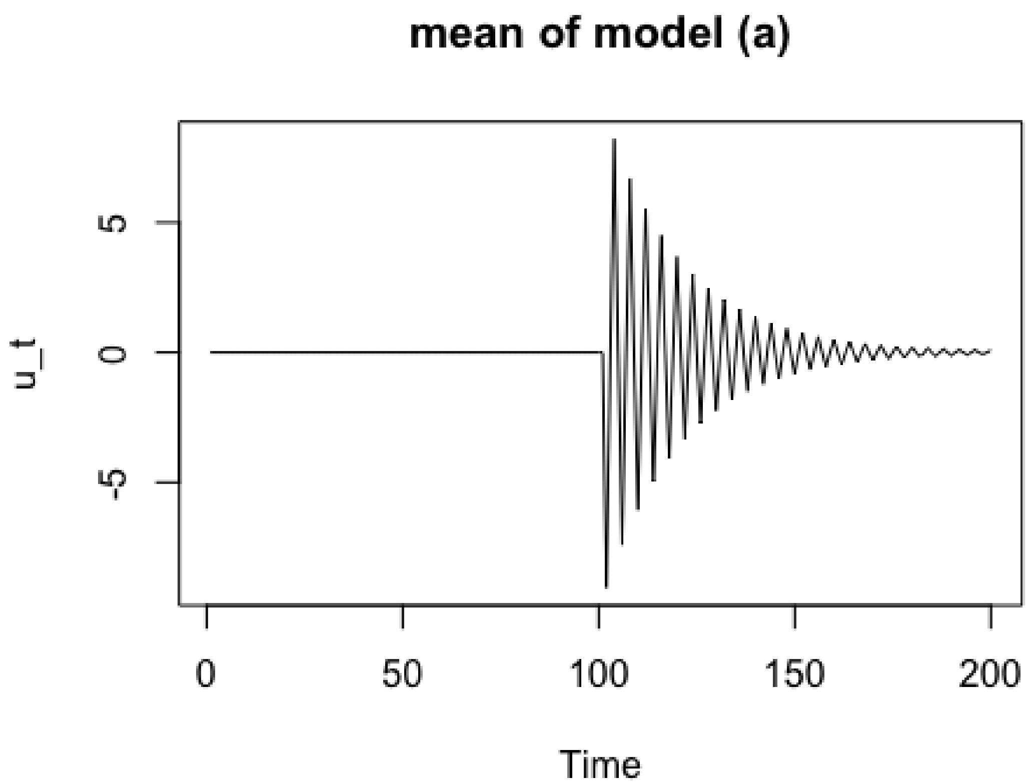


As we can see from the signal modulators plot(a) and plot(b), when the denominator of the exponential function is large, i.e. denominator = 200, the signal modulator plot becomes rather linearly decreasing. While if the denominator of the exponential function is small, i.e. denominator = 20, the signal modulator plot becomes rather decreasing in bending.

$E(x_t) = E(s_t + w_t) = E(s_t)$ since $E(w_t) = 0$ by definition.

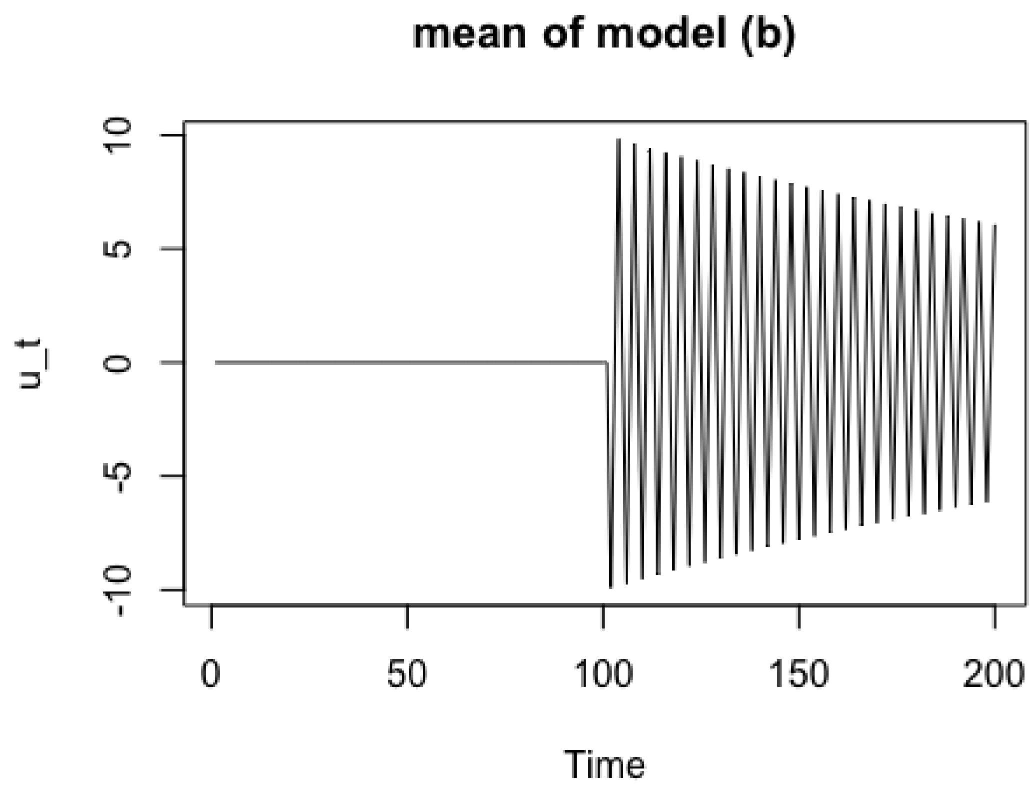
For model (a),$E(x_t) = E(s_t) = 0$ for t = 1:100; $E(x_t) = E(s_t) = E[10exp\frac{-(t-100)}{20}cos\frac{2\pi t}{4}] = 10exp\frac{-(t-100)}{20}cos\frac{2\pi t}{4}$ for t = 101:200

```
set.seed(1004768165) # set seed to get same result every time I run
u_t <- (c(rep(0,100), 10*exp(-(1:100)/20)*cos(2*pi*(101:200)/4))) # set up mea
n function, while E(w_t)= 0, and for the first 100 t, numerator = 0, hence the
 mean also = 0
plot.ts(u_t, main = "mean of model (a)")
```



For model (b), $E(x_t) = E(s_t) = 0$ for t = 1:100; $E(x_t) = E(s_t) = E[10exp\frac{-(t-100)}{200}cos\frac{2\pi t}{4}] = 10exp\frac{-(t-100)}{200}cos\frac{2\pi t}{4}$ for t = 101:200

```
set.seed(1004768165) # set seed to get same result every time I run
u_t <- (c(rep(0,100), 10*exp(-(1:100)/200)*cos(2*pi*(101:200)/4))) # set up me
an function, while E(w_t)= 0, and for the first 100 t, numerator = 0, hence th
e mean also = 0
plot.ts(u_t, main = "mean of model (b)")
```



Question 2

According to the Hint provided by textbooks:

```
trend <- time(jj)   # get the trend
Q <- factor(cycle(jj))    # make (Q)uarter factors
reg <- lm(log(jj) ~ 0 + trend + Q, na.action = NULL)   # no intercept
summary(reg) # get information of the model

##
## Call:
## lm(formula = log(jj) ~ 0 + trend + Q, na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29318 -0.09062 -0.01180  0.08460  0.27644
##
## Coefficients:
##          Estimate Std. Error t value Pr(>|t|)
## trend  1.672e-01  2.259e-03   74.00   <2e-16 ***
## Q1    -3.283e+02  4.451e+00  -73.76   <2e-16 ***
## Q2    -3.282e+02  4.451e+00  -73.75   <2e-16 ***
## Q3    -3.282e+02  4.452e+00  -73.72   <2e-16 ***
## Q4    -3.284e+02  4.452e+00  -73.77   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.1254 on 79 degrees of freedom
## Multiple R-squared:  0.9935, Adjusted R-squared:  0.9931
## F-statistic:  2407 on 5 and 79 DF,  p-value: < 2.2e-16
```

After fitting the regression model, the estimated $\beta = 0.1672$, $\alpha_1 = -328.3$, $\alpha_2 = -328.2$, $\alpha_3 = -328.2$, $\alpha_4 = -328.4$. Our model should be: $log(y_t) = x_t = 0.1672t - 328.3Q_1(t) - 328.2Q_2(t) - 328.2Q_3(t) - 328.4Q_4(t)$

The p-value of these four parameters all <2e-16, i.e. p-value all less than 0.05 = significant level, which indicates strong evidence against the null hypothesis that these parameters = 0.

(b)
If the model is correct, the estimated average annual increase in the logged earnings per share is:

$$E(x_{t+1} - x_t) = E\left[\beta[(t+1) - t] + \sum_{i=1,2,3,4} \alpha_i\,[Q_i(t+1) - Q_i(t)] + [w_{t+1} - w_t]\right]$$

since what the indicator variables care is which season it belongs with, $Q_i(t+1) = Q_i(t)$.

Moreover, $E(w_t) = 0$ since w_t is a Gaussian white noise sequence.

Hence, $E(x_{t+1} - x_t) = E[\beta[(t+1) - t]] = \beta = 0.1672$, i.e. the estimated average annual increase in the logged earnings per share is 0.1672.

(c)

The average logged earnings in Q3 is: $E(x_t) = E(\beta t - \alpha_3) = \beta t - \alpha_3 = 0.1672t - 328.2$
since $Q_i(t)$ is an indicator function, when is in the third quarter, $Q_3(t) = 1, Q_1(t) = Q_2(t) = Q_4(t) = 0$.
Similarly, The average logged earnings in Q4 is:$E(x_t) = \beta t - \alpha_4 = 0.1672t - 328.4$. The average difference from the third quarter to the fourth quarter is:

$[0.1672t - 328.4] - [0.1672t - 328.3] = -0.2.$

Hence, the average logged earnings rate decrease from the third quarter to the fourth quarter is:$\frac{0.2}{328.2} \approx 0.061\%$.

(d)

if we include an intercept term, the linear regression will be:

```
trend <- time(jj)   # get the trend
Q <- factor(cycle(jj))    # make (Q)uarter factors
reg_intercept <- lm(log(jj) ~  trend + Q, na.action = NULL)  # With intercept
summary(reg_intercept) # get information of the model

##
## Call:
## lm(formula = log(jj) ~ trend + Q, na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29318 -0.09062 -0.01180  0.08460  0.27644
##
## Coefficients:
```

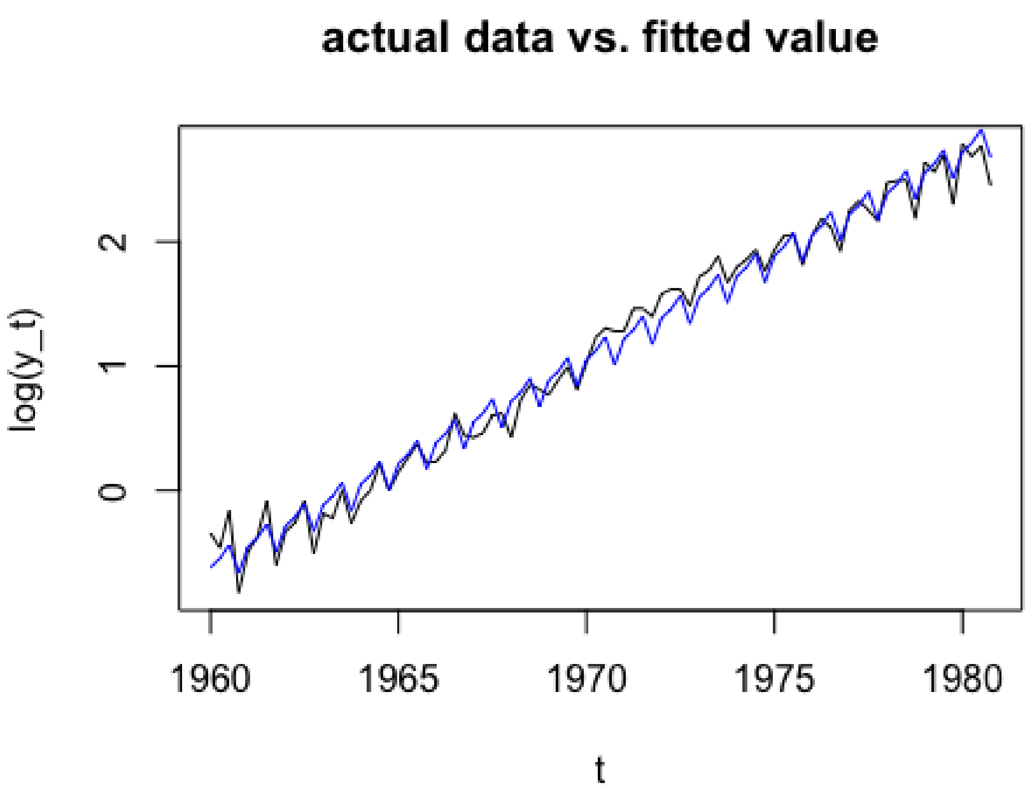
```
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.283e+02  4.451e+00 -73.761  < 2e-16 ***
## trend        1.672e-01  2.259e-03  73.999  < 2e-16 ***
## Q2           2.812e-02  3.870e-02   0.727   0.4695
## Q3           9.823e-02  3.871e-02   2.538   0.0131 *
## Q4          -1.705e-01  3.873e-02  -4.403 3.31e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1254 on 79 degrees of freedom
## Multiple R-squared:  0.9859, Adjusted R-squared:  0.9852
## F-statistic:  1379 on 4 and 79 DF,  p-value: < 2.2e-16
```

We can find that the Q1 term is missing if we include the intercept term. Moreover, the significant level of α_2 is not significant at all, with p-value relatively large = 0.4695. As for the Q3 term, if we set the significant level is 0.01, the α_3 is no longer significant.

(e)

The plot of actual data vs. fitted value:

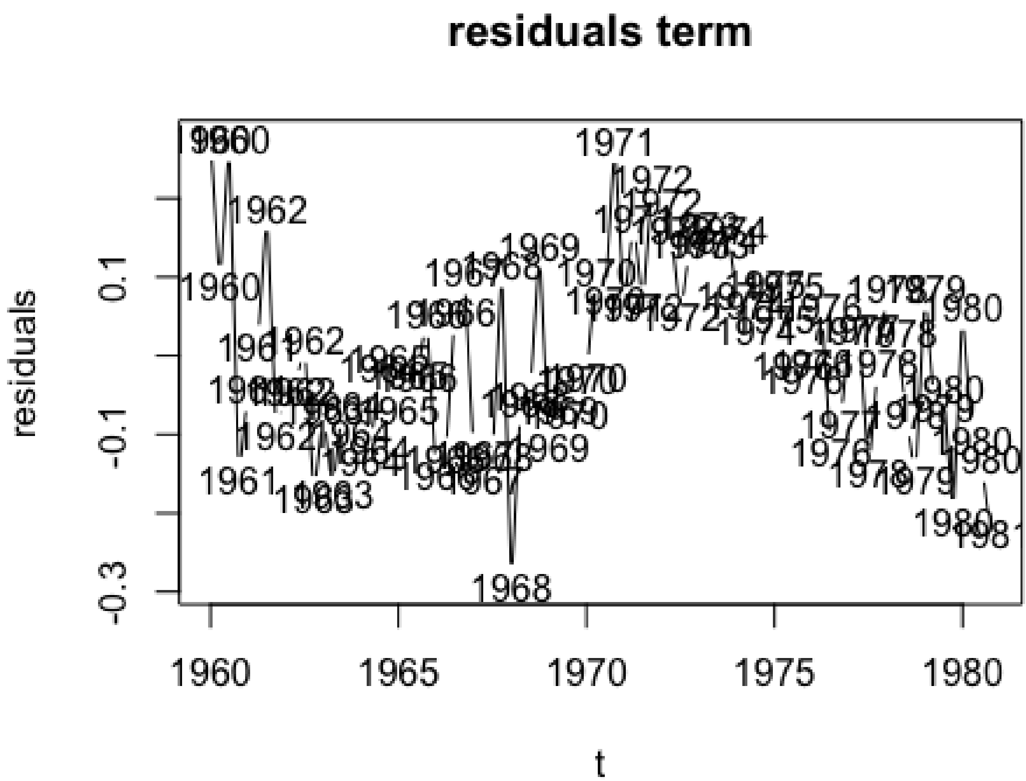
```
plot(log(jj), xlab = "t", ylab = "log(y_t)", main = "actual data vs. fitted va
lue") # plot the actual data graph
lines(fitted.values(reg), col="blue") # add the fitted value inside, with line
 blue.
```



Here, the black line denotes the actual data, the blue line denotes the fitted value. We can find that they have very similar behavior.

As for the residuals term:

```r
reg_ress <- resid(reg) # get the residual of the fitted model.
plot(time(jj), reg_ress, xlab="t", ylab = "residuals", main = "residuals term"
) # plot the residual graph
```



By observing the residual plot, we can find that after t = 1970, the residuals shows a decreasing trend, i.e. it does not shows random pattern, which indicates the fitted model may violate one of white noise's assumption - Homoscedasticity. Hence, by the residuals plot, I conclude that the fitted model may not fit the data that well.

Question 3

First, I will calculate the autocovariance: xt = wt−1 + 1.2wt + wt+1,
$E(x_t) = E(w_{t-1} + 1.2w_t + w_{t+1}) = E(w_{t-1}) + 1.2E(w_t) + E(w_{t+1}) = 0$
since $E(w_t) = 0$ for all t = 1,2, ….

Case (I). h = 0, i.e. s = t:

$\gamma(t,t) = var(x_t) = cov(w_{t-1}, w_{t-1}) + cov(1.2w_t, 1.2w_t) + cov(w_{t+1}, w_{t+1})$
$= var(w_{t-1}) + 1.2^2 var(w_t) + var(w_{t+1}) = 3.44{\sigma_w}^2$

since w_t are independent, i.e. $cov(x_t, x_s) = 0$ for $t \neq s$.

Case (II). h = 1, i.e. s = t+1:

$\gamma(t, t+1) = cov(x_t, x_{t+1}) = cov(w_{t-1} + 1.2w_t + w_{t+1}, w_t + 1.2w_{t+1} + w_{t+2})$
$= 1.2var(w_t) + 1.2var(w_{t+1}) = 2.4{\sigma_w}^2$

since w_t are independent, i.e. $cov(x_t, x_s) = 0$ for $t \neq s$.

Case (III). h = 2, i.e. s = t+2:

$\gamma(t, t+2) = cov(x_t, x_{t+2}) = cov(w_{t-1} + 1.2w_t + w_{t+1}, w_{t+1} + 1.2w_{t+2} + w_{t+3}) = var(w_{t+1}) = {\sigma_w}^2$

since w_t are independent, i.e. $cov(x_t, x_s) = 0$ for $t \neq s$.

Case(IV). h = 3, i.e. s = t+3:

$\gamma(t, t+3) = cov(x_t, x_{t+3}) = cov(w_{t-1} + 1.2w_t + w_{t+1}, w_{t+2} + 1.2w_{t+3} + w_{t+4}) = 0$

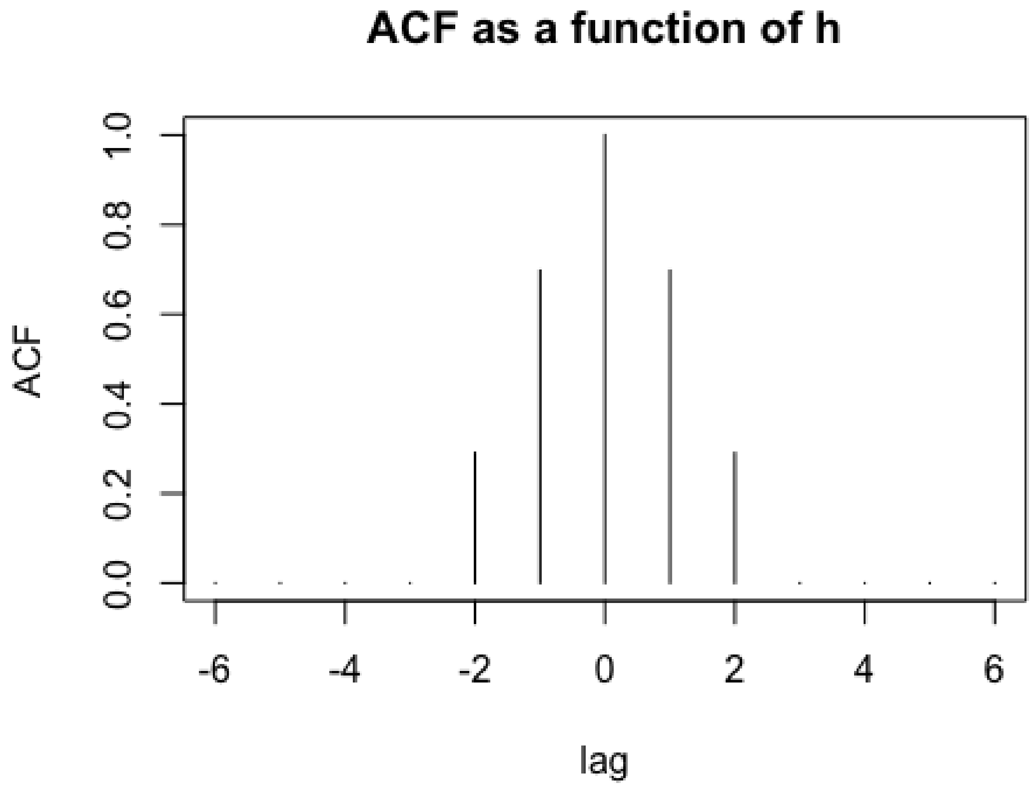
since w_t are independent, i.e. $cov(x_t, x_s) = 0$ for $t \neq s$.

For $h \geq 3, \gamma(t, s) = 0$

In conclusion, the autocovariance function is :
when h = 0, i.e. |s-t| = 0: $\gamma(0) = 3.44{\sigma_w}^2$
when h = 1, i.e. |s-t| = 1: $\gamma(1) = 2.4{\sigma_w}^2$
when h = 2, i.e. |s-t| = 2: $\gamma(2) = {\sigma_w}^2$
when $h \geq 3$, i.e. $|s-t| \geq 3: \gamma(h) = 0$

Next, I will calculate the autocorrelation function:
$\rho(h) = \dfrac{\gamma(h)}{\gamma(0)} = \dfrac{\gamma(h)}{3.44{\sigma_w}^2}$
when h = 0, i.e. |s-t| = 0: $\rho(0) = \frac{\gamma(0)}{\gamma(0)} = 1$
when h = 1, i.e. |s-t| = 1: $\rho(1) = \frac{2.4{\sigma_w}^2}{3.44{\sigma_w}^2} = \frac{30}{43}$
when h = 2, i.e. |s-t| = 2: $\rho(2) = \frac{{\sigma_w}^2}{3.44{\sigma_w}^2} = \frac{25}{86}$
when $h \geq 3$, i.e. $|s-t| \geq 3: \rho(h) = 0$

```
x <- c(-6:6) # set the plot x-axis range h(-6:6)
y <- c(0,0,0,0,25/86, 30/43, 1, 30/43, 25/86, 0,0,0,0) # label each lag
plot(x,y, type = 'h', xlab = 'lag', ylab = 'ACF', main = 'ACF as a function of
 h') # plot the graph
```


ACF as a function of h
ACF
0.0
0.2
0.4
0.6
0.8
1.0
-6
-4
-2
0
2
4
6
lag