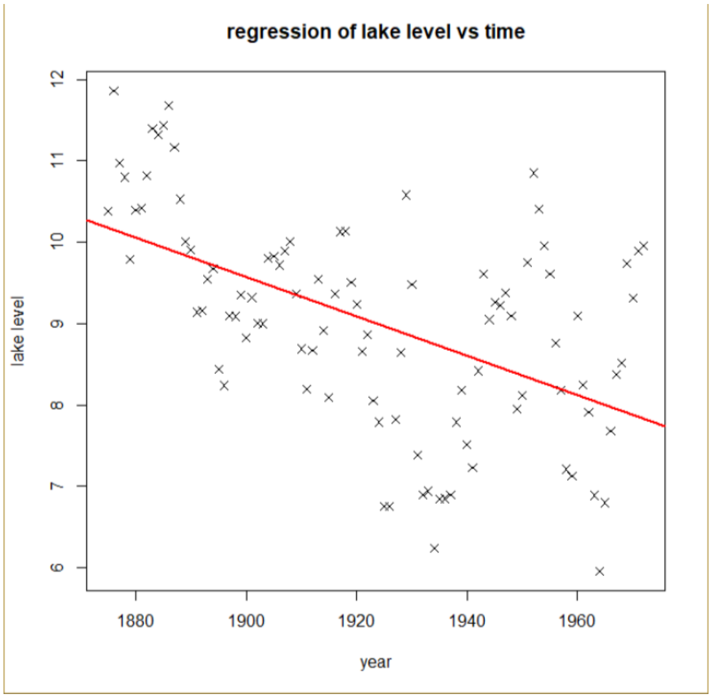
Econometrics and time series Homework1

N10528700 Yun-Feng(Kyle) Lee

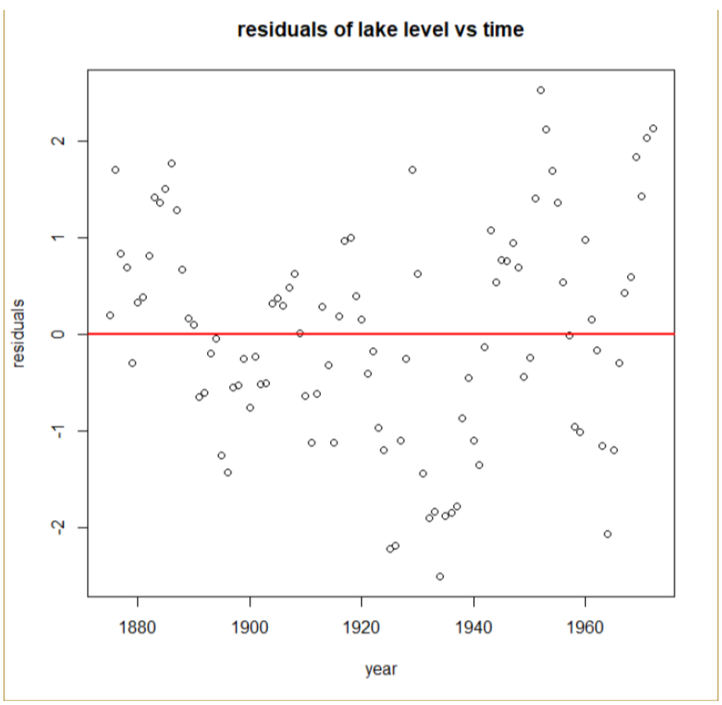
1. Huron Lake Data

a) Fit a regression line of lake level vs time, plot the residuals



In the linear regression, I used the lake level data as the dependent variable(Y), and the years as the dependent variable(X).

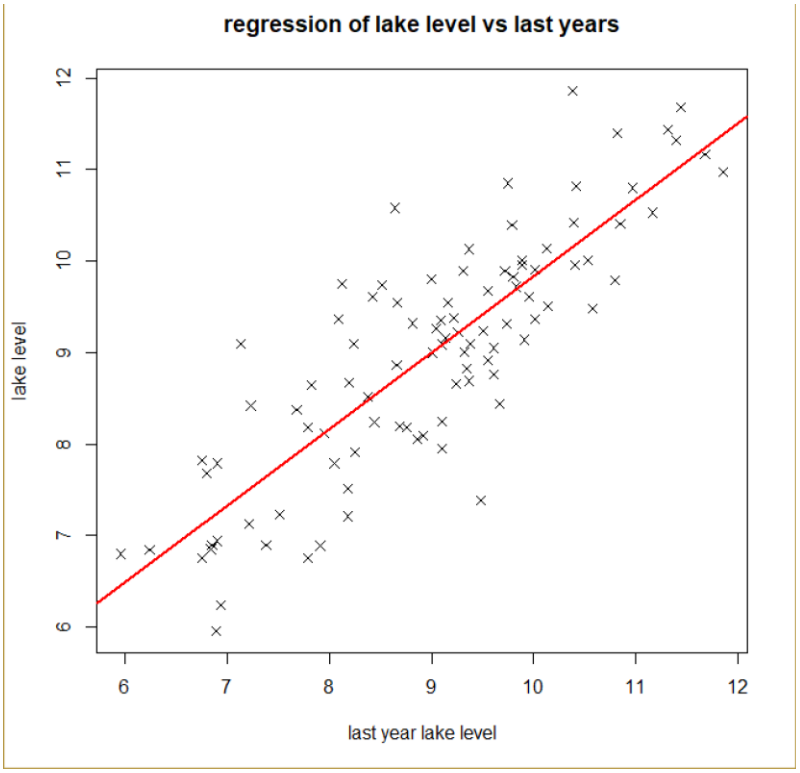
From this graph, we can see that as the year passes, the lake level tends to decrease.



This is the residual of the regression on time.

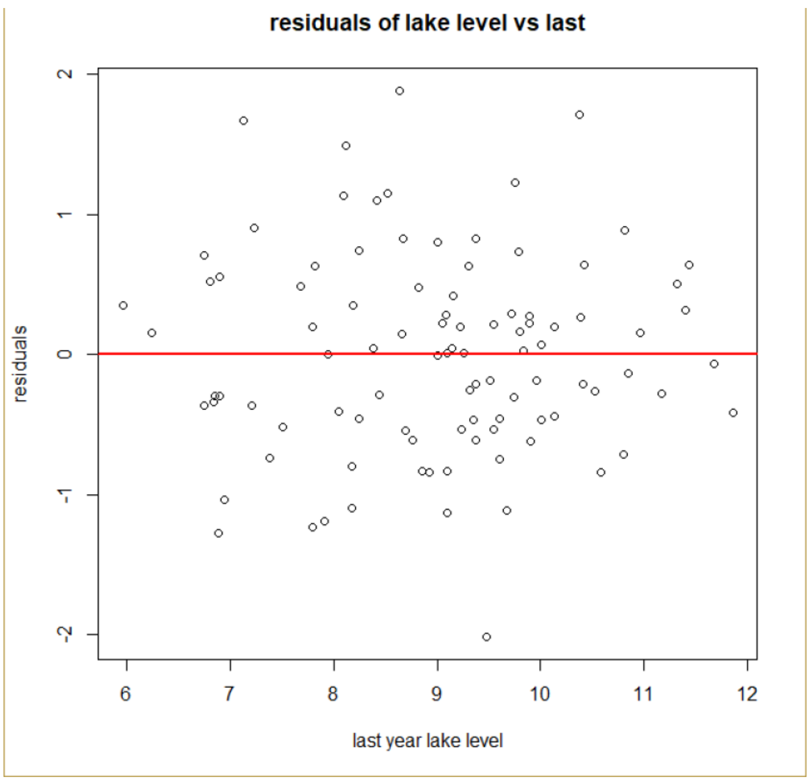
As shown in the graph, there seems to be a smile curve for the residuals.

b) Fit a regression line of lake level vs last year's level, plot the residuals



In the linear regression, I used the lake level data as the dependent variable(Y), and the last year lake level as the dependent variable(X).

From this graph, we can see that as last year's lake level increases, the lake level for this year also increases.



This is the graph of residual of last year's lake level.

It looks more evenly distributed compared to the residual graph when doing the regression on time.

c) Which is a better fit and why

The first regression(vs time) summary:

```
Call:
lm(formula = reg_formula)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.50997 -0.72726  0.00083  0.74402  2.53565 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 55.554918   7.764293   7.155 1.66e-10 ***
ti_me       -0.024201   0.004036  -5.996 3.55e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.13 on 96 degrees of freedom
Multiple R-squared:  0.2725,	Adjusted R-squared:  0.2649 
F-statistic: 35.95 on 1 and 96 DF,  p-value: 3.545e-08
```

The second regression(vs last years lake level) summary:

```
Call:
lm(formula = reg_formula2)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.01620 -0.46747  0.00781  0.47583  1.88638 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.46702    0.50611   2.899  0.00465 ** 
last_lake    0.83641    0.05568  15.022  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7209 on 95 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.7037,	Adjusted R-squared:  0.7006 
F-statistic: 225.7 on 1 and 95 DF,  p-value: < 2.2e-16
```

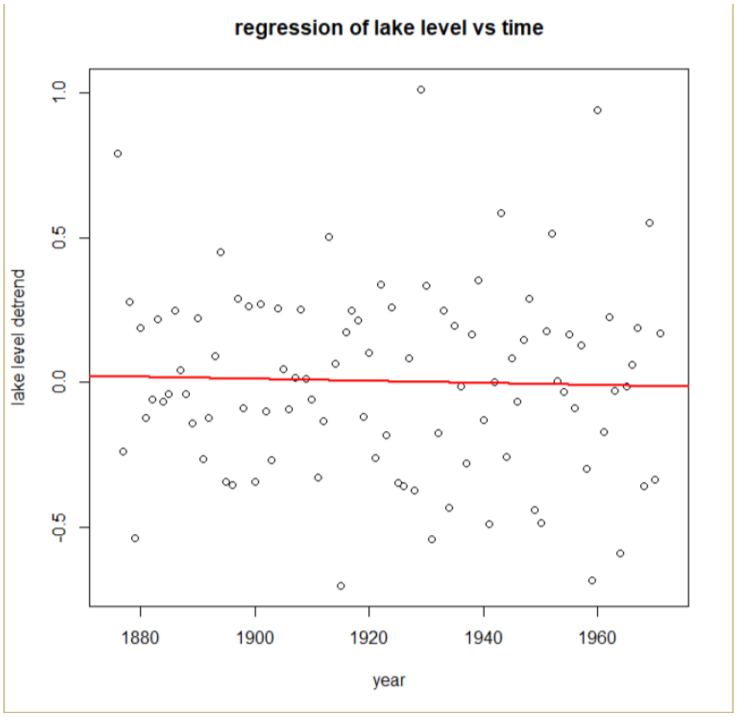
We can see that adjusted R^2 of the second regression (0.7006) is higher than the first(0.2649).

Which means the second regression has a better explanatory power than the first.

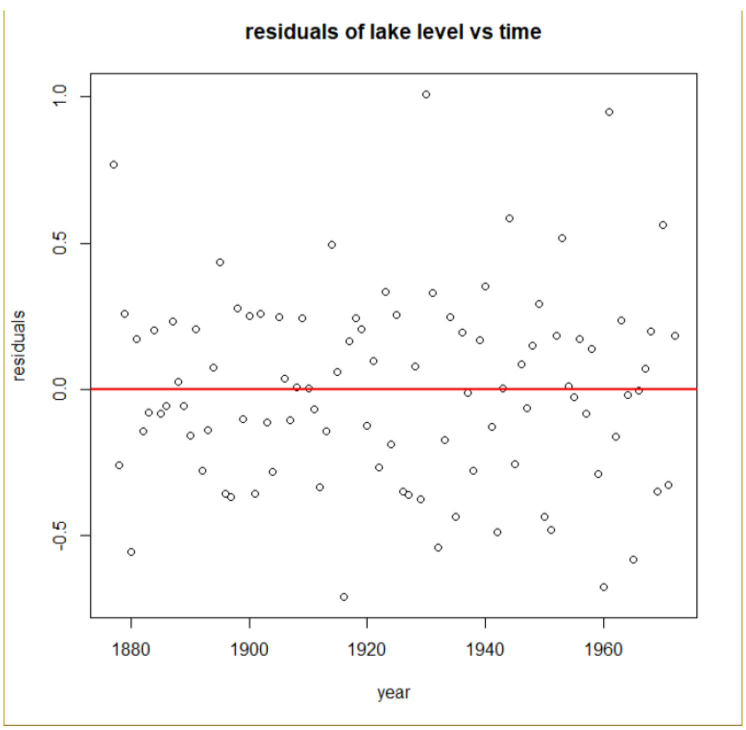
Also from the standard error, we can see that the first has 1.13, and the second has 0.7609.

From this we can also see that the second regression is a better fit. The lower standard error could make sure that the accuracy of the prediction is decent enough.

d) Detrend the data with a moving average, how do the residuals look?

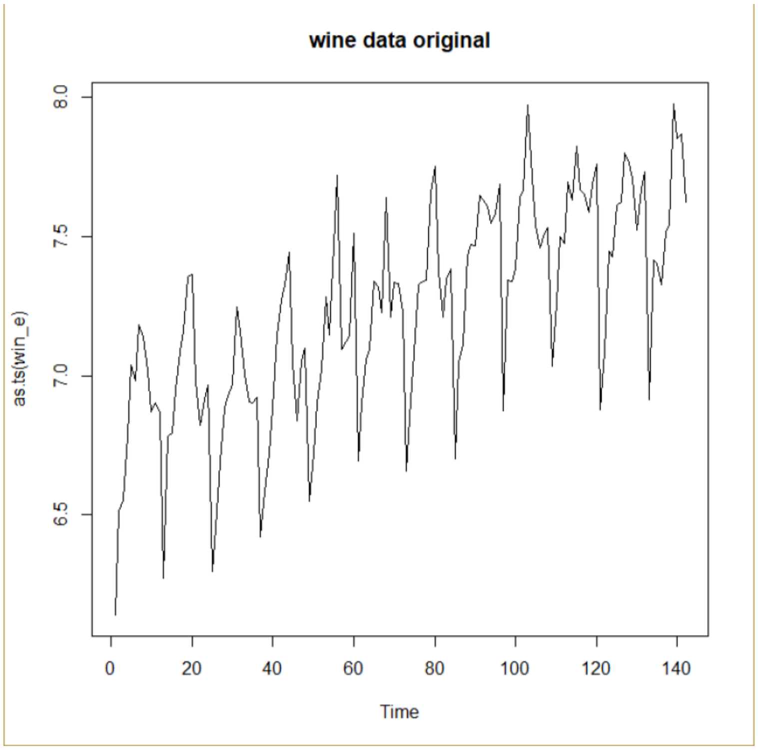


After subtracting the moving average of 3 years, we can see that the trend in a) has disappeared and the regression line is close to a flat line.



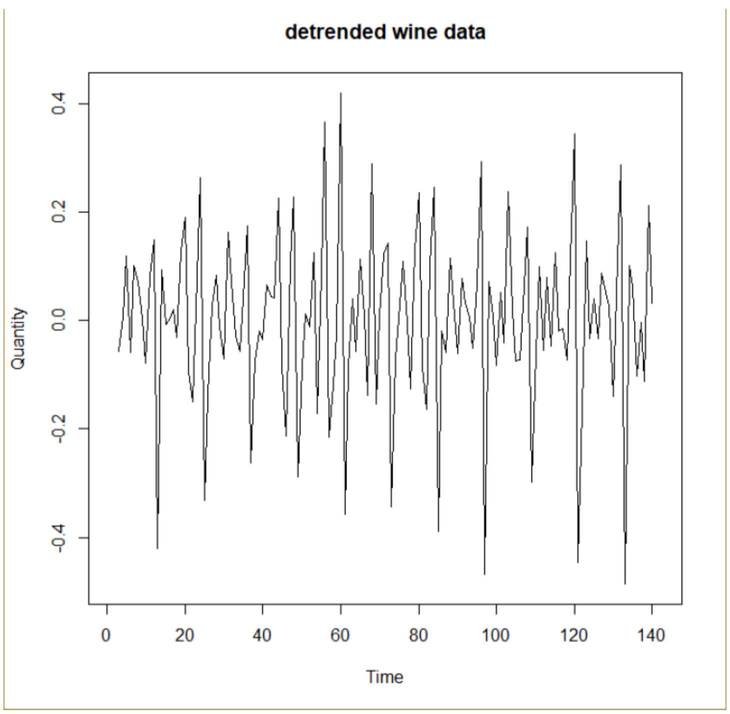
The residual of the lake level vs time looks a lot more symmetrical compared to the one in a)

2. Wine Sales, I use ln(wine) in this case



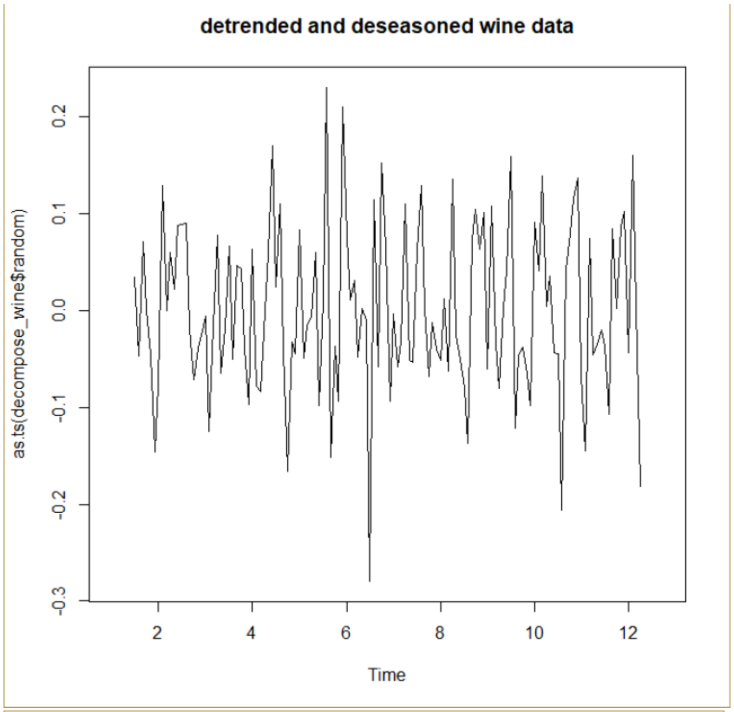
Before doing anything, I graph the ln(wine), we can see that there is a vaguely upward slanting trend and a mountain like seasonality for the log of sales for Australian wines. Then we proceed to do the followings.

a) Detrend the data

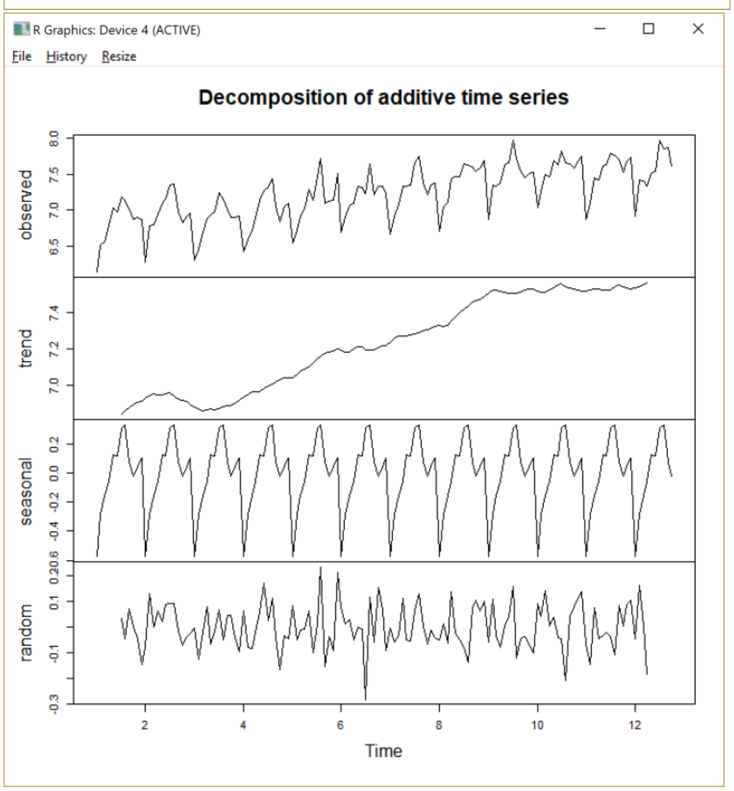


I detrended the data using a 4 month moving average. As the graph shows, it doesn't have a upward trend now

b) remove seasonality



Using the decompose function in R, I removed the trend and seasonality using 12 months.



The decomposition is as follows.

3.

① $X_t = a + bZ_t + cZ_{t-1}$

$E(X_t) = a$

$cov(X_t, X_{t+h}) = cov(bZ_t + cZ_{t-1}, bZ_{t+h} + cZ_{t+h-1})$

$= b^2 cov(Z_t, Z_{t+h}) + bc\, cov(Z_t, Z_{t+h-1}) + cb\, cov(Z_{t-1}, Z_{t+h})$
$+ c^2 cov(Z_{t-1}, Z_{t+h-1})$

$\Rightarrow \begin{cases} (b^2+c^2)\sigma^2 & \text{if } h=0 \\ bc\,\sigma^2 & \text{if } h=\pm 1 \\ 0 & \text{else} \end{cases}$

$E(X_t)$, $cov(X_t, X_{t+h})$ $\Rightarrow$ doesn't depend on t

$\Rightarrow$ Weakly Stationary

② $X_t = Z_1 \cos(ct) + Z_2 \sin(ct)$

$E(X_t) = 0 + 0 = 0$

$cov(X_t, X_{t+h}) = cov(Z_1 \cos(ct) + Z_2 \sin(ct),\ Z_1 \cos(c(t+h)) + Z_2 \sin(c(t+h)))$

$= \cos(ct)\cos(c(t+h))\, cov(Z_1, Z_1)$
$+ \cos(ct)\sin(c(t+h))\, cov(Z_1, Z_2)$
$+ \sin(ct)\cos(c(t+h))\, cov(Z_2, Z_1)$
$+ \sin(ct)\sin(c(t+h))\, cov(Z_2, Z_2)$

$= (\cos(ct)\cos(c(t+h)) + \sin(ct)\sin(c(t+h)))\sigma^2$

$= \cos(c(t+h) - ct)\sigma^2 = \cos(ch)\sigma^2$

$\Rightarrow$ both $E(X_t)$, $cov(X_t, X_{t+h})$ doesn't depend on t

$\Rightarrow$ Weakly Stationary

③ $X_t = Z_t \cos(ct) + Z_{t-1} \sin(ct)$

$\mathbb{E}(X_t) = 0$

$cov(X_t, X_{t+h}) = cov(Z_t \cos(ct) + Z_{t-1} \sin(ct), Z_{t+h} \cos(c(t+h)) + Z_{t+h-1} \sin(c(t+$

$$
\begin{aligned}
&= \cos(ct)\cos(c(t+h))\, cov(Z_t, Z_{t+h}) \\
&+ \cos(ct)\sin(c(t+h))\, cov(Z_t, Z_{t+h-1}) \\
&+ \sin(ct)\cos(c(t+h))\, cov(Z_{t-1}, Z_{t+h}) \\
&+ \sin(ct)\sin(c(t+h))\, cov(Z_{t-1}, Z_{t+h-1})
\end{aligned}
$$

$$
= \begin{cases} \cos(ct)\sigma^2 & \text{if } h=0 \\ \cos(ct)\sin(c(t+1))\sigma^2 & \text{if } h=1 \\ \sin(ct)\cos(c(t-1))\sigma^2 & \text{if } h=-1 \\ 0 & \text{else} \end{cases}
$$

→ depends on t

⟹ Not stationary

④ $X_t = a + bZ_0$

$\mathbb{E}(X_t) = a$

$cov(X_t, X_{t+h}) = cov(bZ_0, bZ_0) = b^2\sigma^2$

→ both doesn't depend on t ⟹ Weakly Stationary

⑤ $X_t = Z_0 \cos(ct)$

$\mathbb{E}(X_t) = \cos(ct) \cdot 0 = 0$

$cov(X_t, X_{t+h}) = cov(Z_0 \cos(ct), Z_0 \cos(c(t+h)))$

$= \cos(ct)\cos(c(t+h))\sigma^2$

since $cov(X_t, X_{t+h})$ depends on t ⟹ Not Stationary

⑥ $X_t = Z_t Z_{t-1}$

$\mathbb{E}(X_t) = \mathbb{E}(Z_t Z_{t-1}) = \mathbb{E}(Z_t)\mathbb{E}(Z_{t-1}) = 0$

$\text{cov}(X_t, X_{t+h}) = \text{cov}(Z_t Z_{t-1}, Z_{t+h} Z_{t+h-1})$

$= \mathbb{E}(Z_t Z_{t-1} Z_{t+h} Z_{t+h-1})$

$$= \begin{cases} \mathbb{E}[Z_{t-1} Z_{t-1}]\,\sigma^2 = \sigma^4 & \text{if } h = 0 \\ \mathbb{E}[Z_{t-1} Z_{t+1}]\,\sigma^2 = 0 & \text{if } h = 1 \\ \mathbb{E}[Z_t \; Z_{t-2}]\,\sigma^2 = 0 & \text{if } h = -1 \\ 0 & \text{else} \end{cases}$$

$\Rightarrow$ Neither depends on t, $\Rightarrow$ Weakly Stationary

⑦ $X_t = a + bZ_t + ct$

$\mathbb{E}(X_t) = a + ct$

$\rightarrow$ the mean depends on $t \Rightarrow$ Not stationary

4. If we take out the (ct) term in ⑦, it would be stationary, as follows

$X_t = a + bZ_t$

$\mathbb{E}(X_t) = a$

$\text{cov}(X_t, X_{t+h}) = \text{cov}(bZ_t, bZ_{t+h})$

$= b^2 \text{cov}(Z_t, Z_{t+h})$

$$= \begin{cases} b^2\sigma^2 & \text{if } h = 0 \\ 0 & \text{else} \end{cases}$$

$\Rightarrow$ It would become weakly Stationary.

Codes used in assignment: (Language:R)

```
# Fit a regression line of lake level vs time
da_ta <- read.table("lake.txt")
la_ke <- unname(unlist(da_ta))
ti_me <- 1875:1972
x11()
plot(ti_me,la_ke, xlab="year",ylab="lake level", pch=4,main ="regression of lake level vs time")
reg_formula <- la_ke ~ ti_me
reg_model <- lm(reg_formula)
abline(reg_model,lwd=2,col="red")

# plot the residuals
x11()
plot(ti_me,reg_model$residuals,xlab="year",ylab="residuals", main ="residuals of lake level vs time")
abline(h=0,lwd=2,col="red")


# Fit a regression line of lake level vs last years level
x11()
last_lake <- c(NA,la_ke[1:97])
plot(last_lake,la_ke,xlab="last year lake level",ylab="lake level", main="regression of lake level vs last
years",pch=4)
reg_formula2 <- la_ke ~ last_lake
reg_model2 <- lm(reg_formula2)
abline(reg_model2,lwd=2,col="red")

# plot the residuals
last_lake_plot <- last_lake[2:98]
plot(last_lake_plot,reg_model2$residuals,xlab="last year lake level",ylab="residuals", main ="residuals
of lake level vs last")
abline(h=0,lwd=2,col="red")


#better fit comparison(compare adjusted R^2(larger) and residual standard error(smaller) better)
summary(reg_model)
summary(reg_model2)

#moving average
movavg <- function(x,n=3){stats::filter(x,rep(1/n,n), sides=2)}
lake_ma <- movavg(la_ke)
lake_ma2 <- la_ke-lake_ma
plot(lake_ma2)
plot(ti_me,lake_ma2, xlab="year",ylab="lake level detrend",main ="regression of lake level vs time")
reg_formula3 <- lake_ma2 ~ ti_me
reg_model3 <- lm(reg_formula3)
abline(reg_model3,lwd=2,col="red")
```

```
time_plot <- ti_me[3:98]
plot(time_plot,reg_model3$residuals,xlab="year",ylab="residuals", main ="residuals of lake level vs
time")
abline(h=0,lwd=2,col="red")



#2 import wine data
wi_ne <- read.table("wine.txt")
colnames(wi_ne) <- c("year", "month","sales")
win_e <- log(wi_ne$sales)
library(zoo)
tim_e <- as.yearmon(paste(wi_ne$year, wi_ne$month),"%Y %m")
plot(tim_e, win_e,main="Wine data")
plot(as.ts(win_e),main="wine data original")
library("forecast")
trend_wine <- ma(win_e,order=4,centre=T)
detrend_wine <- win_e-trend_wine
plot(as.ts(detrend_wine),main="detrended wine data", xlab="Time",ylab="Quantity")

ts_wine <- ts(win_e,frequency=12)
decompose_wine <- decompose(ts_wine,"additive")
plot(as.ts(decompose_wine$seasonal),main="seasonality of wine data")
plot(as.ts(decompose_wine$trend))
plot(as.ts(decompose_wine$random),main="detrended and deseasoned wine data")
plot(decompose_wine)
```