Time Series Homework(附R代码实现)

杨宸宇 2016301550186

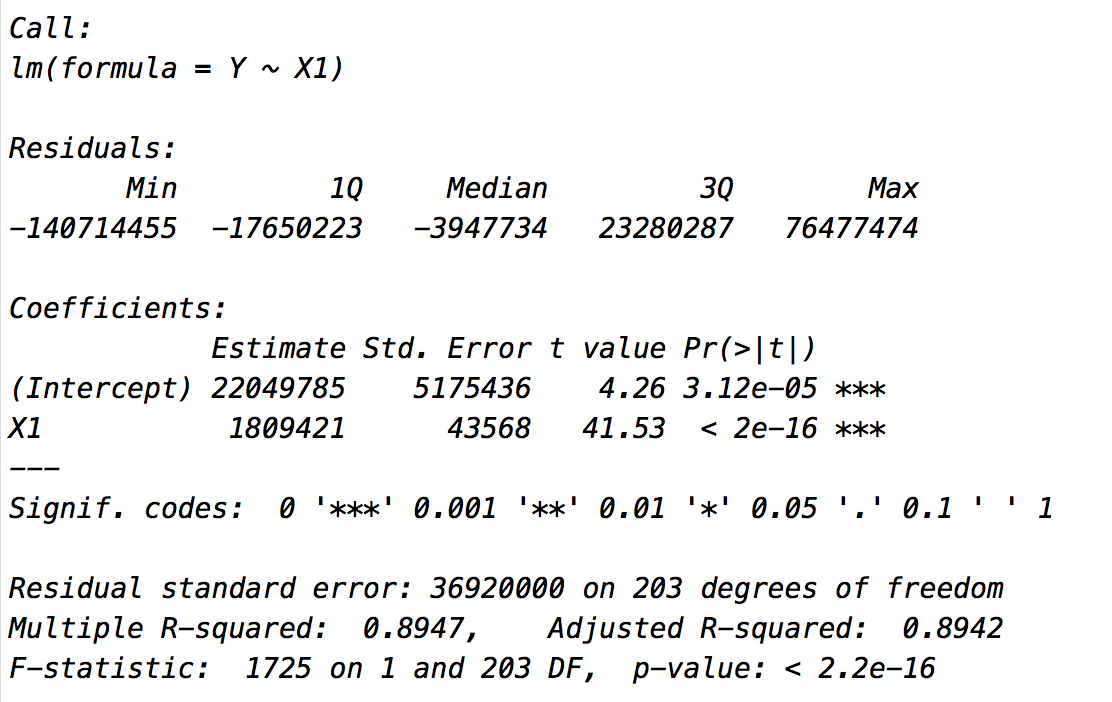
# 1.1 Trends

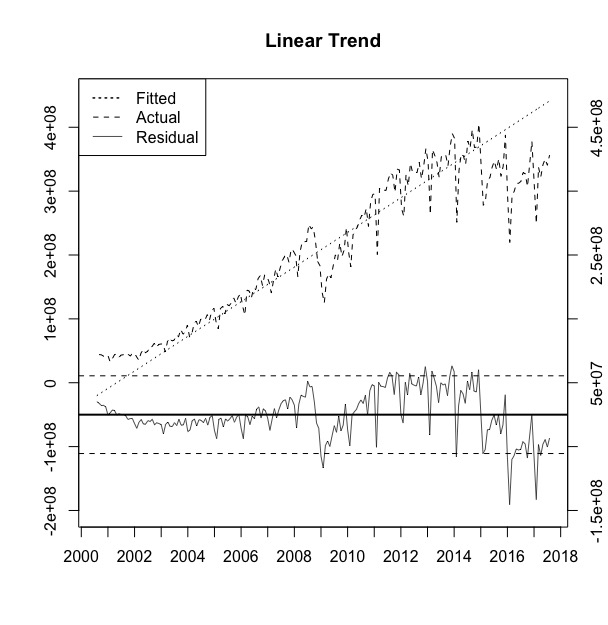
Initially, I modify the dataframe into the standard way, and I try to use different models to judge whether there are trends in the data.

Firstly, I use linear model to fit in the row data, and the results and plot show as followings:

MES = 1.349423e+15

AIC = 7729.648

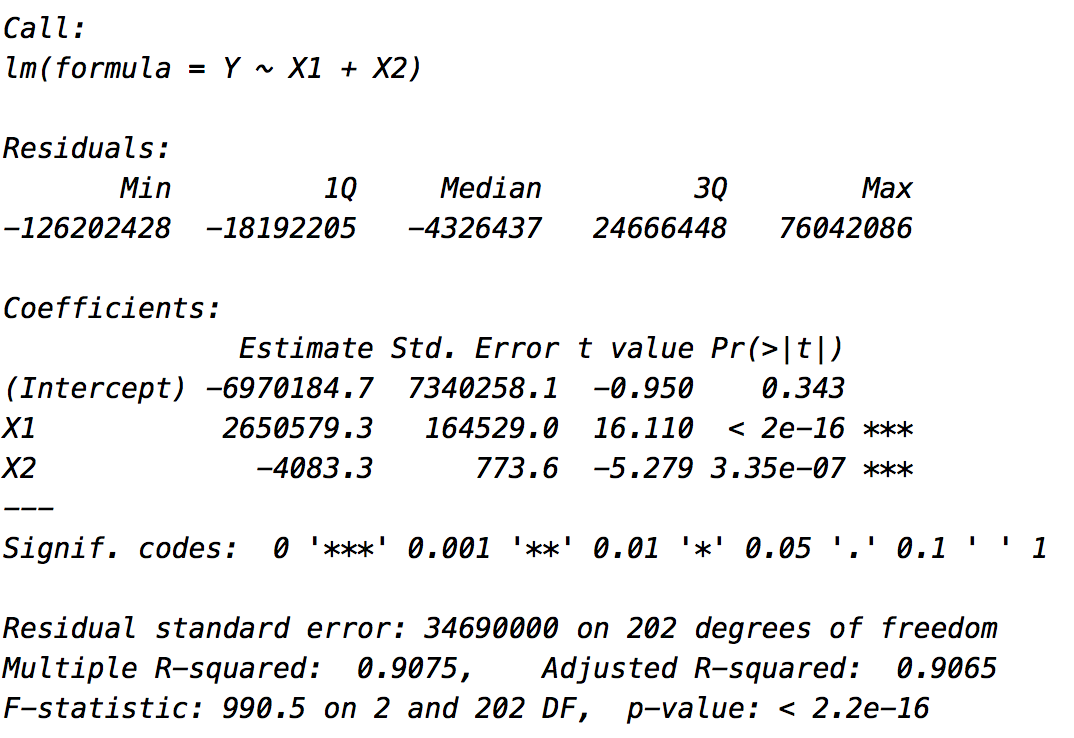


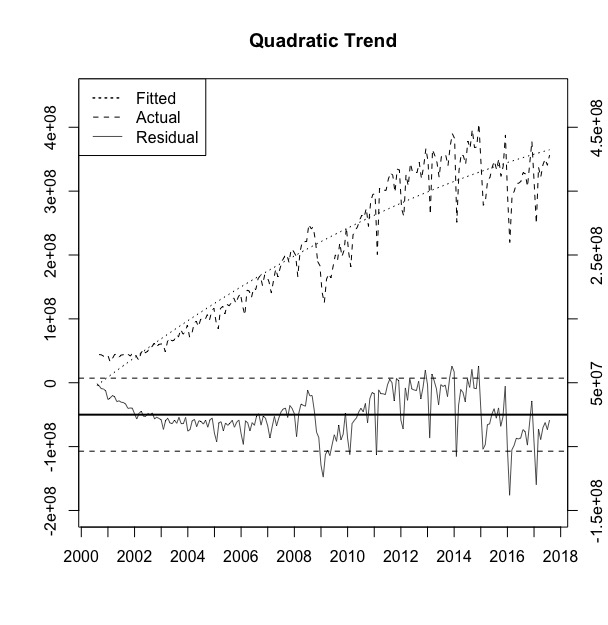


Then I use the quadratic model to fit in the row data, and the results and plot show as followings:

MSE = 1.18585e+15

AIC = 7705.158

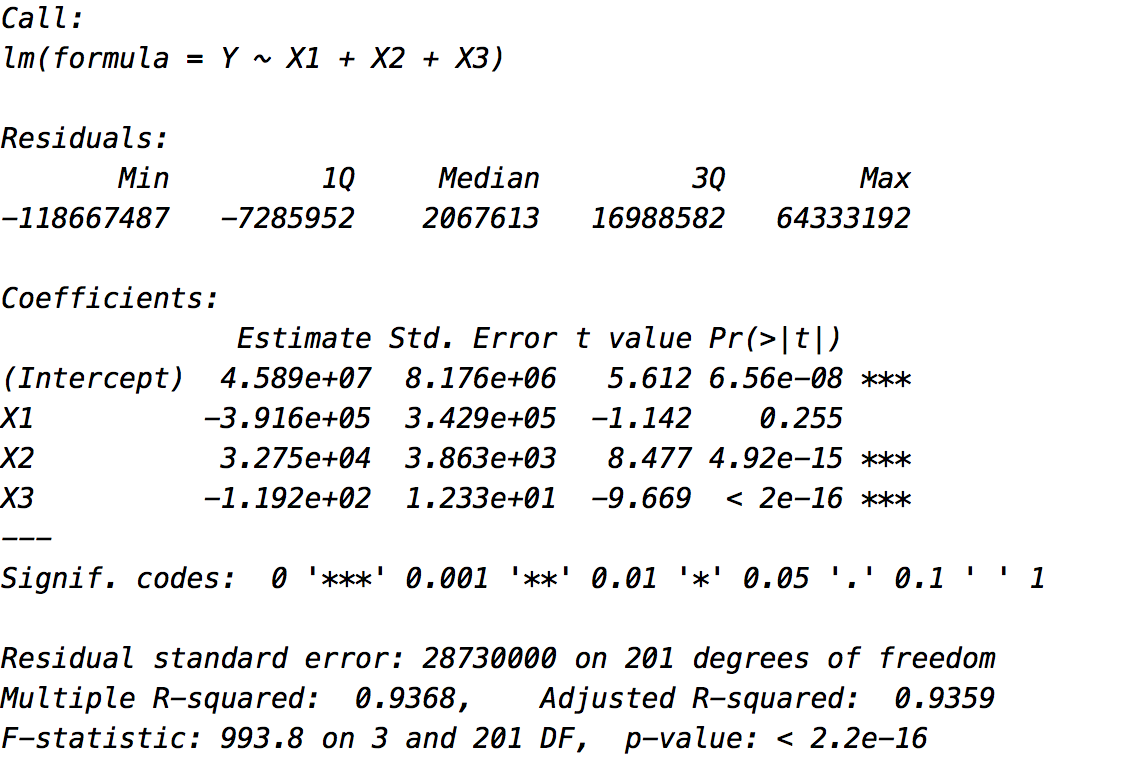


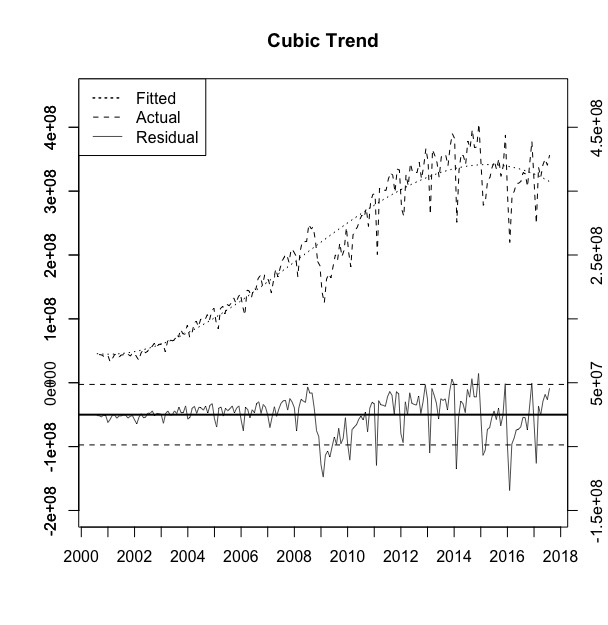


After that, I use the cubic model and even more complex model to fit in the row data, and the results and plot for cubic model are followings:

MSE = 8.094123e+14

AIC = 7645.482

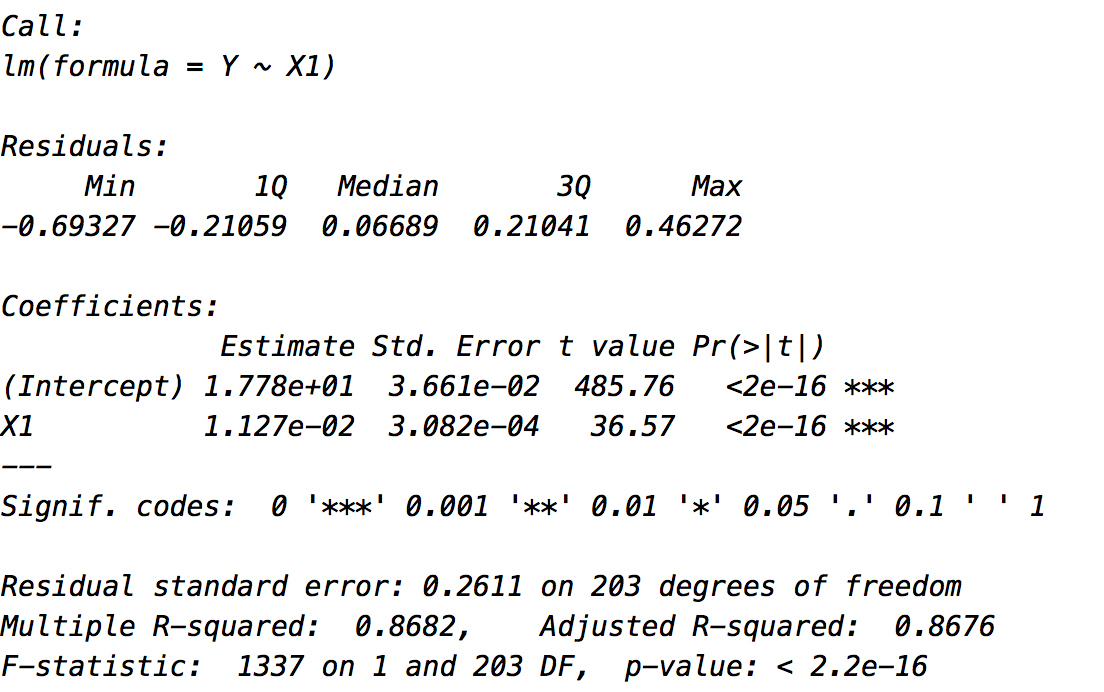


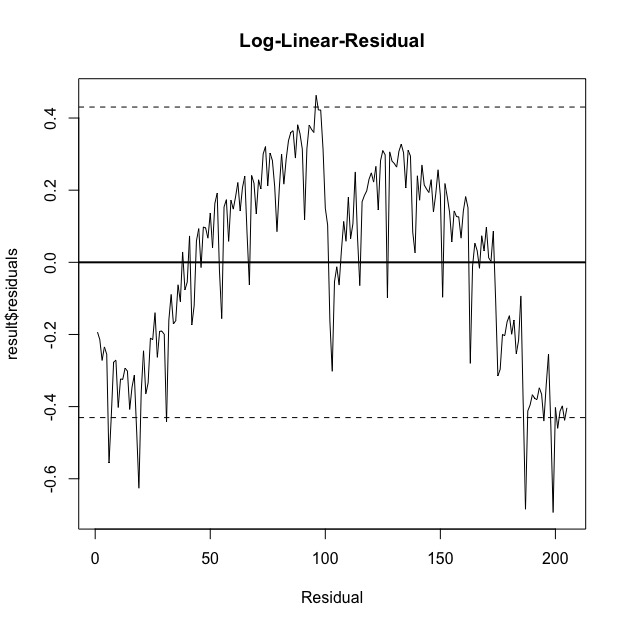
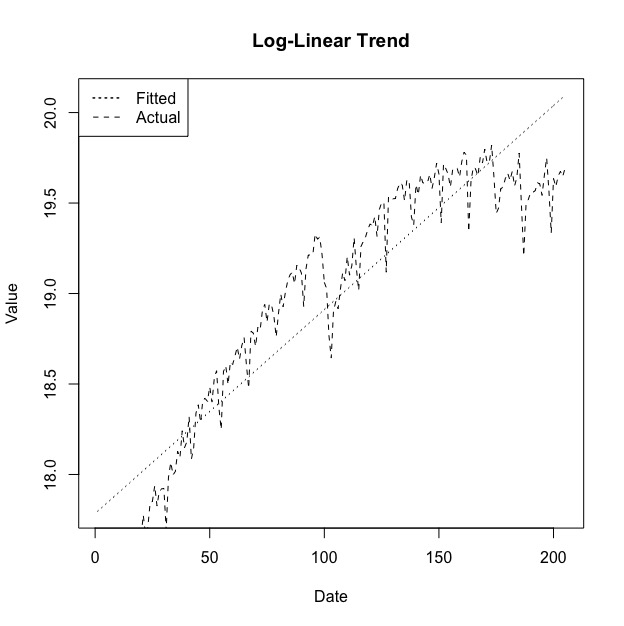


Finally, I try the log model, and I have found the results and plot for it, and them show as followings:

MSE = 0.06752925

AIC = 35.24995





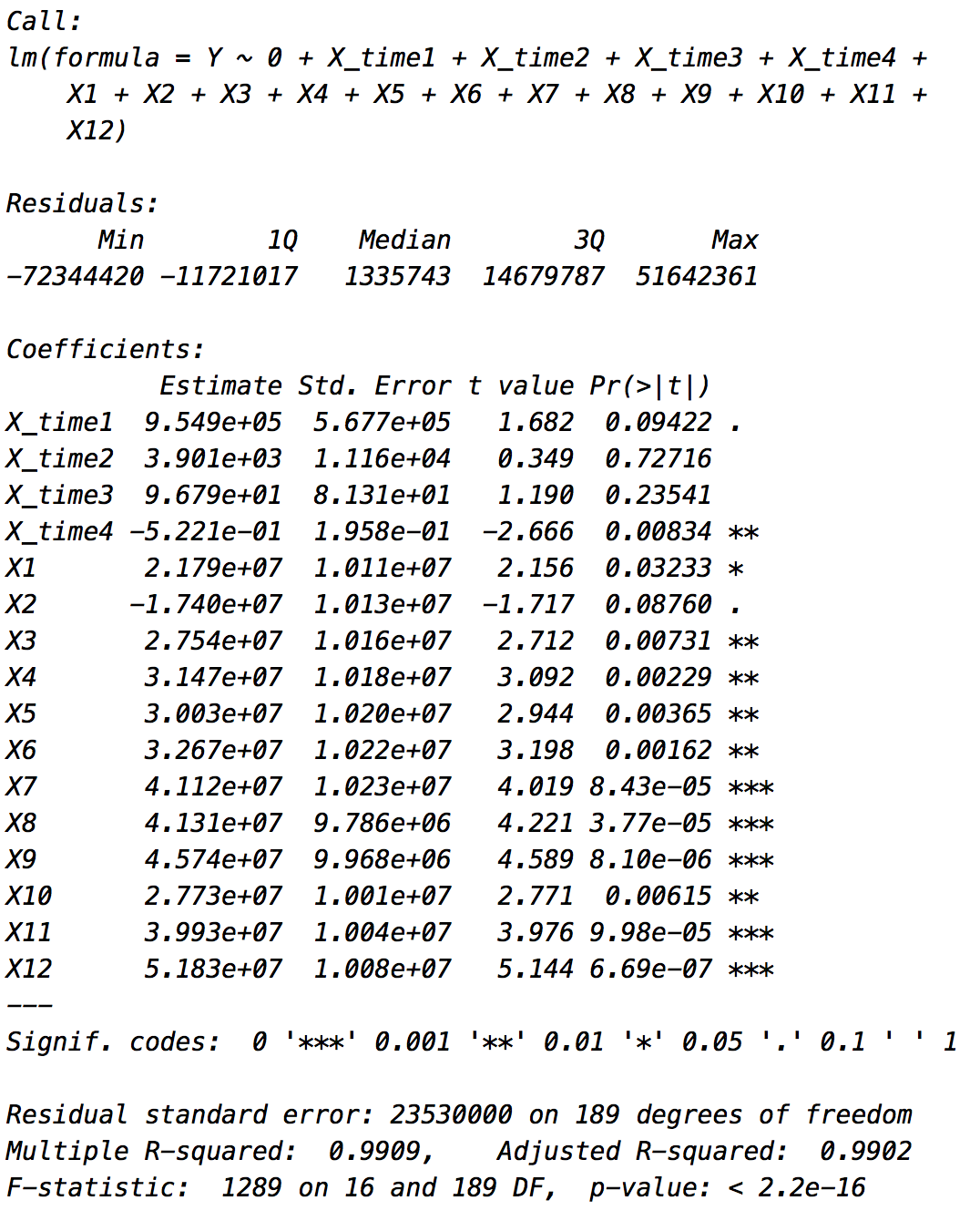
In the end, by analysis all these models, I find the row data truly has trend, and the 4th square can fit it well. In another word, I use this model to take apart the trend in the further process.

# 1.2 Seasonality

In this question, I set 12 dummy variables to judge whether there are seasonal trend in the row data, and the results are as followings:

MSE = 5.106391e+14

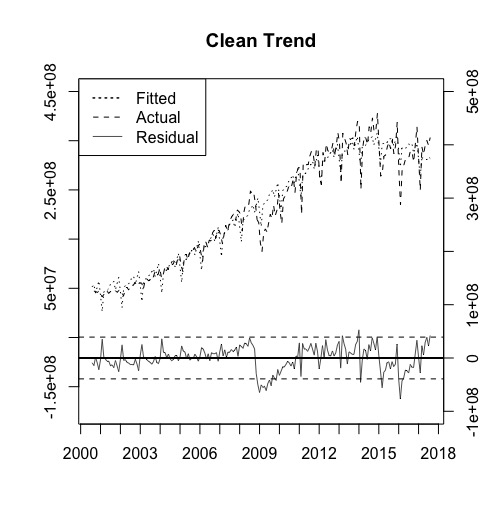
AIC = 7558.435

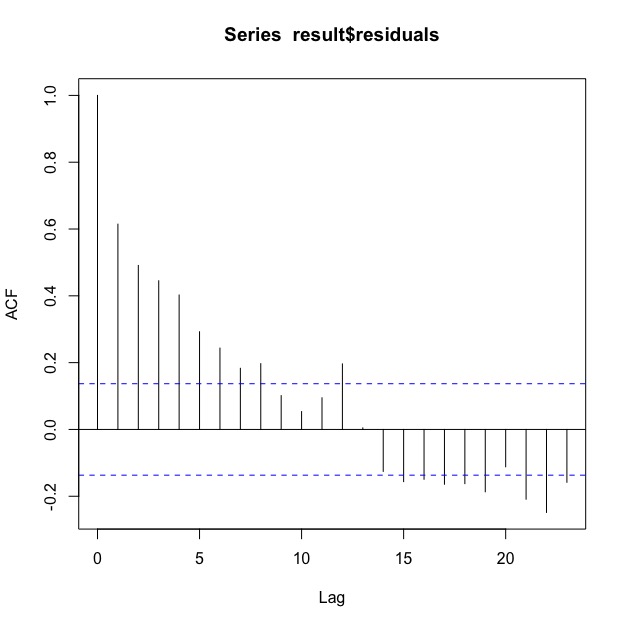


Through the results we can easily find there are great seasonal trends in the data(except February), especially in July, August, September, November and December. And we can draw a conclusion that the seasonality truly exists.

# 1.3 Analysis and plot data without trends and seasonality

Cleaned data plot & Residual ACF plot:





Through the plot, we can see the trend is departed most after my operations on it. However, if we focus on the ACF plot, we can find that there may be still some dependence structures in the cleaned data.

附录

第一题的R代码实现

# Topic: The Second Homework in Analysis of Time Series

# Name: YANG CHENYU

# Class: Financial Engineering 2

# Student ID:2016301550186

# load package----------------------------------------------

library(tseries)

library(lmtest)

library(tidyverse)

# set workspace----------------------------------------------

setwd('/Users/mac/Desktop/R\_Time\_Analysis')

# input data----------------------------------------------

data <- read.csv('hw2\_data.csv', head = T, sep = ',')

row=nrow(data)

# prepare and integrate date----------------------------------------------

data <- data %>%

# divide dates into year & month

separate(Year, into = c("Year", "Mon"), sep = '/')

data$Year <- as.double(data$Year)

data$Mon <- as.double(data$Mon)

# let the Month column maintain double and change Date into character

data["Month"] <- data$Mon

# add day as 1 in order to change Date from character to date(double in R)

data["add"] = 1

data <- data %>%

arrange(Year, Mon) %>%

unite(Date,c("Year", "Mon", "add"), sep = '-')

data$Date = as.Date(data$Date)

data = data[c(1,3,2)]

# add the dummy variables for each month

data <- cbind(data[1:2],1,2,3,4,5,6,7,8,9,10,11,12,data[3])

data$index = 0

data[3:14] = 0

data = data[c(16,1:15)]

for(i in seq\_along(data$Month)){

data[i, 1] = i

a = data[i, 3]%%12

if(a == 0){

data[i, 15] = 1

}

else{

data[i, 3 + a] = 1

}

}

# have a simple look at data

plot(x = data$Date, y = data$Value, xlab = 'Date', ylab = 'Value', main = 'First Look')

# regression analysis----------------------------------------------

Y = data$Value

# Problem\_1

# estimate the trends in the dataset

# 1.Linear Trend

X1 = data$index

result=lm(Y~X1)

MSE = sum(result$residuals^2)/row

SE\_2 = sum(result$residuals^2)/(row-1)

SE = sqrt(SE\_2)

logLik(result)

AIC(result)

AIC(result, k=log(row))

# 0.95 t\_text

t\_value\_up = qt(0.95, row - 1, lower.tail = T)

t\_value\_down = qt(0.05, row - 1, lower.tail = T)

up = SE \* t\_value\_up

down = SE \* t\_value\_down

# plot

plot.new()

plot(x = data$index, y = result$fitted.values,axes = F, type = 'l', ylim = c(-120000000,400000000), xlab = '', ylab = '', lty = 3, ldw = 1.5)

par(new = TRUE)

plot(x = data$index, y = data$Value, xlim = c(0,205), ylim = c(-200000000,450000000),type = 'l', lty = 2, lwd = 1, xlab = '', ylab = '', axes = F)

axis(side = 2, at = c(seq(-200000000,450000000,by=100000000)), lab = c(seq(-200000000,450000000,by=100000000)))

par(new=TRUE)

plot(x = data$index, y = result$residuals, main = 'Linear Trend', ylim = c(-150000000,500000000),type = 'l', lty = 1, lwd = 0.7, xlab = '', ylab = '', axes = F)

abline(h = 0, lwd = 2)

abline(h = up, lty= 2)

abline(h = down, lty= 2)

axis(side = 4, at = c(seq(-150000000,500000000,by=100000000)), lab = c(seq(-150000000,500000000,by=100000000)))

axis(side = 1, at = c(seq(6-12, 198+12, by = 12)), lab = c(seq(2000, 2018, by = 1)))

legend('topleft', c('Fitted', 'Actual', 'Residual'), lty=c(3, 2, 1), lwd = c(1.5,1, 0.7))

box()

# 2. Quadratic Trend

X1 = data$index

X2 = X\*\*2

result=lm(Y~X1+X2)

MSE = sum(result$residuals^2)/row

SE\_2 = sum(result$residuals^2)/(row-2)

SE = sqrt(SE\_2)

logLik(result)

AIC(result)

AIC(result, k=log(row))

# 0.95 t\_text

t\_value\_up = qt(0.95, row - 1, lower.tail = T)

t\_value\_down = qt(0.05, row - 1, lower.tail = T)

up = SE \* t\_value\_up

down = SE \* t\_value\_down

# plot

plot.new()

plot(x = data$index, y = result$fitted.values,axes = F, type = 'l', ylim = c(-200000000,450000000), xlab = '', ylab = '', lty = 3, ldw = 1.5)

axis(side = 2, at = c(seq(-200000000,450000000,by=100000000)))

par(new = TRUE)

plot(x = data$index, y = data$Value, xlim = c(0,205), ylim = c(-200000000,450000000),type = 'l', lty = 2, lwd = 1, xlab = '', ylab = '', axes = F)

axis(side = 2, at = c(seq(-200000000,450000000,by=100000000)), lab = c(seq(-200000000,450000000,by=100000000)))

par(new=TRUE)

plot(x = data$index, y = result$residuals, main = 'Quadratic Trend', ylim = c(-150000000,500000000),type = 'l', lty = 1, lwd = 0.7, xlab = '', ylab = '', axes = F)

abline(h = 0, lwd = 2)

abline(h = up, lty= 2)

abline(h = down, lty= 2)

axis(side = 4, at = c(seq(-150000000,500000000,by=100000000)), lab = c(seq(-150000000,500000000,by=100000000)))

axis(side = 1, at = c(seq(6-12, 198+12, by = 12)), lab = c(seq(2000, 2018, by = 1)))

legend('topleft', c('Fitted', 'Actual', 'Residual'), lty=c(3, 2, 1), lwd = c(1.5,1, 0.7))

box()

# 3.Cubic Trend

X1 = data$index

X2 = X1\*\*2

X3 = X1\*\*3

result=lm(Y~X1+X2+X3)

MSE = sum(result$residuals^2)/row

SE\_2 = sum(result$residuals^2)/(row-3)

SE = sqrt(SE\_2)

logLik(result)

AIC(result)

AIC(result, k=log(row))

# 0.95 t\_text

t\_value\_up = qt(0.95, row - 1, lower.tail = T)

t\_value\_down = qt(0.05, row - 1, lower.tail = T)

up = SE \* t\_value\_up

down = SE \* t\_value\_down

# plot

plot.new()

plot(x = data$index, y = result$fitted.values,axes = F, type = 'l', ylim = c(-200000000,450000000), xlab = '', ylab = '', lty = 3, ldw = 1.5)

axis(side = 2, at = c(seq(-200000000,450000000,by=100000000)))

par(new = TRUE)

plot(x = data$index, y = data$Value, xlim = c(0,205), ylim = c(-200000000,450000000),type = 'l', lty = 2, lwd = 1, xlab = '', ylab = '', axes = F)

axis(side = 2, at = c(seq(-200000000,450000000,by=100000000)), lab = c(seq(-200000000,450000000,by=100000000)))

par(new=TRUE)

plot(x = data$index, y = result$residuals, main = 'Cubic Trend', ylim = c(-150000000,500000000),type = 'l', lty = 1, lwd = 0.7, xlab = '', ylab = '', axes = F)

abline(h = 0, lwd = 2)

abline(h = up, lty= 2)

abline(h = down, lty= 2)

axis(side = 4, at = c(seq(-150000000,500000000,by=100000000)), lab = c(seq(-150000000,500000000,by=100000000)))

axis(side = 1, at = c(seq(6-12, 198+12, by = 12)), lab = c(seq(2000, 2018, by = 1)))

legend('topleft', c('Fitted', 'Actual', 'Residual'), lty=c(3, 2, 1), lwd = c(1.5,1, 0.7))

box()

# 4.Log-Linear Trend

Y = log(data$Value)

X1 = data$index

result=lm(Y~X1)

MSE = sum(result$residuals^2)/row

SE\_2 = sum(result$residuals^2)/(row-1)

SE = sqrt(SE\_2)

logLik(result)

AIC(result)

AIC(result, k=log(row))

# 0.95 t\_text

t\_value\_up = qt(0.95, row - 1, lower.tail = T)

t\_value\_down = qt(0.05, row - 1, lower.tail = T)

up = SE \* t\_value\_up

down = SE \* t\_value\_down

# plot

plot.new()

plot(x = data$index, y = result$fitted.values, main = 'Log-Linear Trend', type = 'l', xlab = 'Date', ylab = 'Value', lty = 3)

lines(x = data$index, y = Y, lty = 2, lwd = 1)

legend('topleft', c('Fitted', 'Actual'), lty=c(3, 2), lwd = c(1.5,1))

plot(x = data$index, result$residuals, type = 'l', xlab = 'Residual', main = 'Log-Linear-Residual')

abline(h = 0, lwd = 2)

abline(h = up, lty= 2)

abline(h = down, lty= 2)

# plot in one pic, but it doesn't seem good

plot.new()

plot(x = data$index, y = result$fitted.values,axes = F, type = 'l', ylim = c(-1, 21), xlab = '', ylab = '', lty = 3, ldw = 1.5)

plot(x = data$index, y = Y, xlim = c(0,205), ylim = c(-1, 21),type = 'l', lty = 2, lwd = 1, xlab = '', ylab = '', axes = F)

axis(side = 2, at = c(seq(-1,21,by=5)), lab = c(seq(-1,21,by=5)))

par(new=TRUE)

plot(x = data$index, y = result$residuals, main = 'Log-Linear Trend', ylim = c(-1,21),type = 'l', lty = 1, lwd = 0.7, xlab = '', ylab = '', axes = F)

abline(h = 0, lwd = 2)

abline(h = up, lty= 2)

abline(h = down, lty= 2)

axis(side = 4, at = c(seq(-1,1,by=0.1)), lab = c(seq(-1,1,by=0.1)))

axis(side = 1, at = c(seq(6-12, 198+12, by = 12)), lab = c(seq(2000, 2018, by = 1)))

legend('topleft', c('Fitted', 'Actual', 'Residual'), lty=c(3, 2, 1), lwd = c(1.5,1, 0.7))

# Problem\_2

X1 = data$`1`

X2 = data$`2`

X3 = data$`3`

X4 = data$`4`

X5 = data$`5`

X6 = data$`6`

X7 = data$`7`

X8 = data$`8`

X9 = data$`9`

X10 = data$`10`

X11 = data$`11`

X12 = data$`12`

Y = data$Value

result = lm(Y~0+X1+X2+X3+X4+X5+X6+X7+X8+X9+X10+X11+X12)

summary(result)

sum(result$residuals^2)

MSE = sum(result$residuals^2)/row

SE\_2 = sum(result$residuals^2)/(row-12)

SE = sqrt(SE\_2)

logLik(result)

AIC(result)

AIC(result, k=log(row))

# 0.95 t\_text

t\_value\_up = qt(0.95, row - 1, lower.tail = T)

t\_value\_down = qt(0.05, row - 1, lower.tail = T)

up = SE \* t\_value\_up

down = SE \* t\_value\_down

# plot

plot.new()

plot(x = data$Date, y = result$fitted.values, main = 'Seasonality', type = 'l', ylim = c(-190000000,400000000), xlab = 'Date', ylab = 'Value', lty = 3, ldw = 1.5)

lines(x = data$Date, y = data$Value, lty = 2, lwd = 1)

lines(x = data$Date, y = result$residuals, lty = 1, lwd = 0.7)

legend('topleft', c('Fitted', 'Actual', 'Residual'), lty=c(3, 2, 1), lwd = c(1.5,1, 0.7))

abline(h = 0, lwd = 2)

abline(h = up, lty= 2)

abline(h = down, lty= 2)

# Problem\_3

X\_time1 = data$index

X\_time2 = X\_time1\*\*2

X\_time3 = X\_time1\*\*3

X\_time4 = X\_time1\*\*4

X1 = data$`1`

X2 = data$`2`

X3 = data$`3`

X4 = data$`4`

X5 = data$`5`

X6 = data$`6`

X7 = data$`7`

X8 = data$`8`

X9 = data$`9`

X10 = data$`10`

X11 = data$`11`

X12 = data$`12`

Y = data$Value

result = lm(Y~0+X\_time1+X\_time2+X\_time3+X\_time4+X1+X2+X3+X4+X5+X6+X7+X8+X9+X10+X11+X12)

summary(result)

sum(result$residuals^2)

MSE = sum(result$residuals^2)/row

SE\_2 = sum(result$residuals^2)/(row-16)

SE = sqrt(SE\_2)

logLik(result)

AIC(result)

AIC(result, k=log(row))

# 0.95 t\_text

t\_value\_up = qt(0.95, row - 1, lower.tail = T)

t\_value\_down = qt(0.05, row - 1, lower.tail = T)

up = SE \* t\_value\_up

down = SE \* t\_value\_down

# ACF

acf(data$Value)

acf(result$residuals)

# plot

plot.new()

plot(x = data$index, y = result$fitted.values,axes = F, type = 'l', ylim = c(-120000000,400000000), xlab = '', ylab = '', lty = 3, ldw = 1.5)

axis(side = 2, at = c(seq(-150000000,450000000,by=100000000)))

par(new = TRUE)

plot(x = data$index, y = data$Value, xlim = c(0,205), ylim = c(-200000000,450000000),type = 'l', lty = 2, lwd = 1, xlab = '', ylab = '', axes = F)

axis(side = 2, at = c(seq(-150000000,450000000,by=100000000)), lab = c(seq(-150000000,450000000,by=100000000)))

par(new=TRUE)

plot(x = data$index, y = result$residuals, main = 'Clean Trend', ylim = c(-100000000,500000000),type = 'l', lty = 1, lwd = 0.7, xlab = '', ylab = '', axes = F)

abline(h = 0, lwd = 2)

abline(h = up, lty= 2)

abline(h = down, lty= 2)

axis(side = 4, at = c(seq(-100000000,500000000,by=100000000)), lab = c(seq(-100000000,500000000,by=100000000)))

axis(side = 1, at = c(seq(6-12, 198+12, by = 12)), lab = c(seq(2000, 2018, by = 1)))

legend('topleft', c('Fitted', 'Actual', 'Residual'), lty=c(3, 2, 1), lwd = c(1.5,1,0.7))

box()

# Problem\_3

X\_time1 = data$index

X\_time2 = X\_time1\*\*2

X\_time3 = X\_time1\*\*3

X\_time4 = X\_time1\*\*4

X1 = data$`1`

X3 = data$`3`

X4 = data$`4`

X5 = data$`5`

X6 = data$`6`

X7 = data$`7`

X8 = data$`8`

X9 = data$`9`

X10 = data$`10`

X11 = data$`11`

X12 = data$`12`

Y = data$Value

result = lm(Y~0+X\_time1+X\_time2+X\_time3+X\_time4+X1+X3+X4+X5+X6+X7+X8+X9+X10+X11+X12)

summary(result)

sum(result$residuals^2)

MSE = sum(result$residuals^2)/row

SE\_2 = sum(result$residuals^2)/(row-16)

SE = sqrt(SE\_2)

logLik(result)

AIC(result)

AIC(result, k=log(row))

# 0.95 t\_text

t\_value\_up = qt(0.95, row - 1, lower.tail = T)

t\_value\_down = qt(0.05, row - 1, lower.tail = T)

up = SE \* t\_value\_up

down = SE \* t\_value\_down

# ACF

acf(data$Value)

acf(result$residuals)

# plot

plot.new()

plot(x = data$index, y = result$fitted.values,axes = F, type = 'l', ylim = c(-200000000,450000000), xlab = '', ylab = '', lty = 3, ldw = 1.5)

par(new = TRUE)

plot(x = data$index, y = data$Value, xlim = c(0,205), ylim = c(-200000000,450000000),type = 'l', lty = 2, lwd = 1, xlab = '', ylab = '', axes = F)

axis(side = 2, at = c(seq(-150000000,450000000,by=100000000)), lab = c(seq(-150000000,450000000,by=100000000)))

par(new=TRUE)

plot(x = data$index, y = result$residuals, main = 'Clean Trend', ylim = c(-100000000,500000000),type = 'l', lty = 1, lwd = 0.7, xlab = '', ylab = '', axes = F)

abline(h = 0, lwd = 2)

abline(h = up, lty= 2)

abline(h = down, lty= 2)

axis(side = 4, at = c(seq(-100000000,500000000,by=100000000)), lab = c(seq(-100000000,500000000,by=100000000)))

axis(side = 1, at = c(seq(6-12, 198+12, by = 12)), lab = c(seq(2000, 2018, by = 1)))

legend('topleft', c('Fitted', 'Actual', 'Residual'), lty=c(3, 2, 1), lwd = c(1.5,1,0.7))

box()