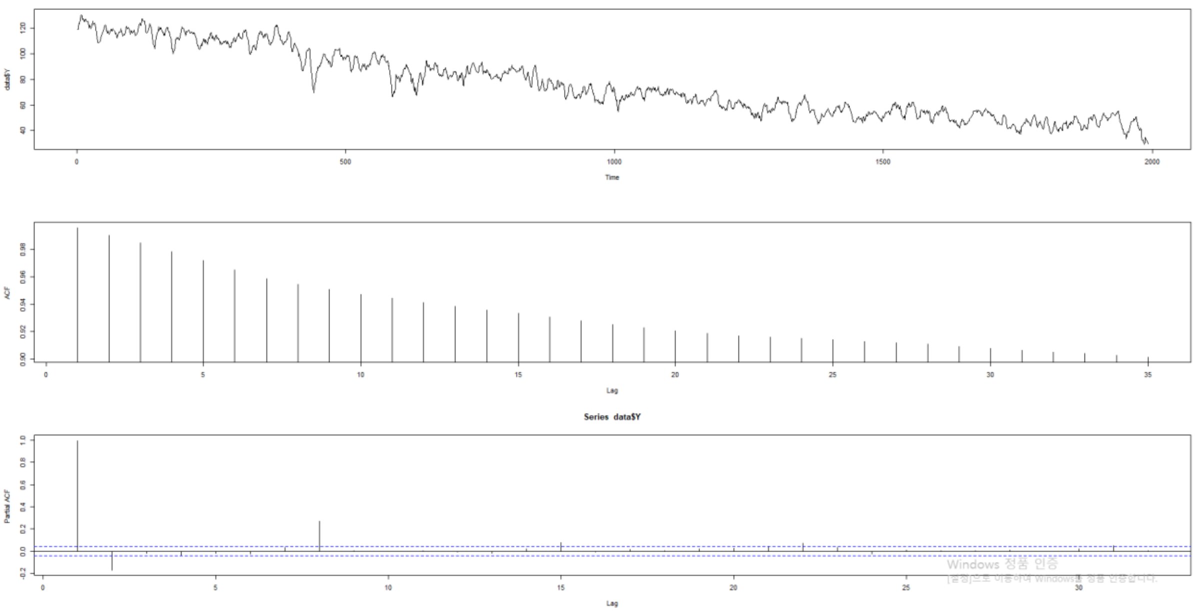
Name : Hee Chul Jeong

Student ID : 2016314895

Assignment Title : Introduction to Time Series Analysis Finals 1

1 – a)

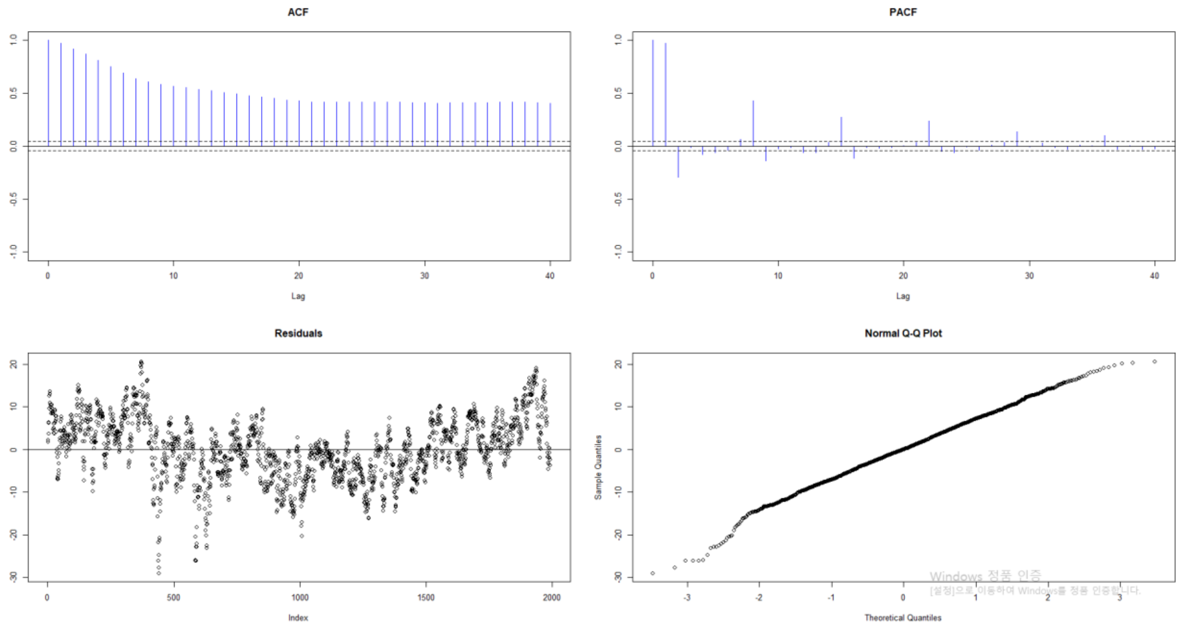
The time plot clearly shows that there is a linearly decreasing trend so detrending is needed in the analysis. The plot itself does not clearly indicate any seasonality but PACF diagram shows that the lag 1, 8, and 15 is not negligible, and this implies there may be a seasonality. Lastly, looking at the ACF diagram, it is slowly decaying, which indicates that differencing may be needed.



1 – b)

No, it is not stationary. The PACF shows that there are strong correlations between residuals at lag 1, 8, 15, and so on, and ACF is slowly decaying but stabilize at around 0.5. In addition, by the residual plot, it is clear that there is a quadratic trend in residuals, indicating that the mean may be 0 but not constant. Therefore, because the conditions of stationarity are violated for the above 3 reasons, it is not stationary.

```
> test(resi)
Null hypothesis: Residuals are iid noise.
Test                          Distribution Statistic   p-value
Ljung-Box Q                  Q ~ chisq(20)  16132.24         0 *
McLeod-Li Q                  Q ~ chisq(20)    6057.8         0 *
Turning points T(T-1326.7)/18.8 ~ N(0,1)       837         0 *
Diff signs S      (S-995.5)/12.9 ~ N(0,1)      1005     0.461
Rank P       (P-991518)/14823.3 ~ N(0,1)    982670    0.5506
```



1 – c)

Model : Y = 117.0267 - 0.0417 x + stationary errors

I have used ARIMA function to get the model estimation. The OLS would give the same parameters; however, the standard errors are usually much larger in OLS estimations. Surprisingly, though, the standard errors of the OLS method in this case are slightly smaller than using ARIMA. However, there is an advantage of using ARIMA over OLS in forecasting since there is a sign of seasonality in the data.

```
> summary(lmod)

Call:
lm(formula = vecdata ~ x)

Residuals:
     Min       1Q   Median       3Q      Max 
-29.0300  -4.6424   0.0021   4.9064  20.5832 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.170e+02  3.208e-01   364.8   <2e-16 ***
x           -4.169e-02  2.788e-04  -149.5   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.155 on 1990 degrees of freedom
Multiple R-squared:  0.9183,	Adjusted R-squared:  0.9182 
F-statistic: 2.236e+04 on 1 and 1990 DF,  p-value: < 2.2e-16

> fit.reg1

Call:
arima(x = vecdata, order = c(2, 0, 0), xreg = X1, include.mean = FALSE)

Coefficients:
         ar1      ar2     const        x
      1.2565  -0.2971  117.0267  -0.0417
s.e.  0.0214   0.0214    1.8269   0.0016

sigma^2 estimated as 2.858:  log likelihood = -3873.89,  aic = 7757.78
```

1 – d)

Model : SARIMA(4,0,1) x (2,1,0)

Because it has been previously noted that the data does not have stationary errors, I first transformed the data to stabilize the errors. The best lambda was -0.02020202 so all the data has been transformed by the power of -0.02020202, and it can seen by the residual plot and QQ plot that the errors have been stabilized.

As for seasonal ARMA part, I chose 2 because there are two distinct seasonality in the PACF, both at every 7th lag starting from at lag 1 and 2. I fixed D = 1 because the ACF is slowly decaying, which indicates the seasonal differencing is needed. As for temporal ARMA part, I chose p = 4 because within the first 7 lag in the PACF, there are 4 lags stood out from the line. Lastly, q = 1 because there is a decreasing trend. The AIC of the given model is -24314.05, which is low enough to be recognized.

```
> fit.2

Call:
arima(x = tran.data, order = c(4, 0, 1), seasonal = list(order = c(2, 1, 0),
    period = 7))

Coefficients:
         ar1      ar2     ar3      ar4      ma1      sar1     sar2
      2.0714  -1.3439  0.3426  -0.0885  -0.8678  -0.9587  -0.4695
s.e.  0.0358   0.0601  0.0514   0.0249   0.0289   0.0207   0.0204

sigma^2 estimated as 2.769e-07:  log likelihood = 12165.02,  aic = -24314.05
```

1 – e)

	June 17, 2010	**June 18, 2010**	**June 19, 2010**	**June 20, 2010**
Model (c) **Point Forecast** **95% PI**	Point Forecast : **33.92** Low : **27.67** High : **40.17**	Point Forecast : **33.88** Low : **27.62** High : **40.13**	Point Forecast : **33.83** Low : **27.58** High : **40.09**	Point Forecast : **33.79** Low : **27.53** High : **40.05**
Model (d) **Point Forecast** **95% PI**	Point Forecast : **0.93** Low : **0.93** High : **0.93**	Point Forecast : **0.93** Low : **0.93** High : **0.94**	Point Forecast : **0.93** Low : **0.93** High : **0.94**	Point Forecast : **0.93** Low : **0.93** High : **0.94**

*****The transformation was applied for Model (d)**

1 – f)

I would prefer to use (d) because that the (c) was not transformed, which may cause the unstable variance and that the forecasts of (c) would only be on a straight line, which make it easier to interpret the model, but it would cause the model not to reflect the seasonality and the variation.

1 – g)

The data has a decreasing trend and seasonality by the time plot, which make the data non-stationary. This non-stationarity can further be seen from the residual plot and QQ plot given above, and the 5 tests of normality indicate that the residuals are not IID. Therefore, before modeling, the Box-cox transformation was applied with Lambda = -0.02020202 in purpose of stabilizing the errors. Thus all the data has been transformed by the power of -0.02020202. Looking at various diagrams and comparing different models, SARIMA(4,0,1) x (2,1,0) seemed to be best-fitted to handle the data because of the fact that ACF was slowly decaying, that the PACF had 2 seasonal peaks and 4 significant lags within the first season, and that the data has a linearly decreasing trend. The coefficients of the model have small standard errors, and the AIC is very small so that it is plausible to be selected.

R-codes used

library(itsmr)

library(MASS)

library(forecast)

 # 1-a

data <- read.csv('G:/내 드라이브/skku class files/2021년도 1학기/시계열분석입문/Final 1 - 2016314895 - Hee Chul Jeong/practice2-2021sp.csv')

data <- list(Dates = data[,1], Y = data[,2])

par(mfrow=c(3,1)); plot.ts(data$Y); acf2(data$Y); pacf(data$Y); par(mfrow=c(1,1))

 # 1-b

vecdata <- as.vector(data$Y)

x <- 1:length(vecdata)

const <- rep(1,length(vecdata))

lmod <- lm(vecdata~x)

resi <- vecdata - lmod$fitted.values

test(resi)

 # 1-c

x <- 1:length(vecdata)

const <- rep(1,length(vecdata))

using OLS

lmod <- lm(vecdata~x)

summary(lmod)

using ARIMA

X1 <- cbind(const, x)

fit.reg1 <- arima(vecdata, order=c(2,0,0), xreg = X1, include.mean = FALSE)

fit.reg1

 # 1-d

library(MASS)

```
fit = boxcox(Mod(vecdata)~1, plotit=TRUE)

lambda = fit$x[which.max(fit$y)]

lambda

tran.data <- (vecdata^lambda)


lmod <- lm(tran.data~x)

resi <- tran.data - lmod$fitted.values

test(resi)


fit.1 = arima(tran.data, order = c(2,0,0), seasonal=list(order=c(1,1,0), period = 7))

fit.1


fit.2 = arima(tran.data, order = c(4,0,1), seasonal=list(order=c(2,1,0), period = 7))

fit.2
```

 # 1-e

```
detach("package:itsmr")

library(forecast)

forecast(fit.2,4)

plot(forecast(fit.2,4))


newx <- 1993:1996

point.pred <- (newx * (-0.0417)) + 117.0267


lower.beta <- -0.0417 - (1.96)*0.0016

high.beta <- -0.0417 + (1.96)*0.0016


lower.pred <- (newx * (lower.beta)) + 117.0267

high.pred <- (newx * (high.beta)) + 117.0267
```