ISYE6402 HW1

Question 1: Temperature Analysis.

```
rm(list=ls())
library(TSA)
```

```
##
## Attaching package: 'TSA'
```

```
## The following objects are masked from 'package:stats':
##
##     acf, arima
```

```
## The following object is masked from 'package:utils':
##
##     tar
```

```
library(mgcv)
```

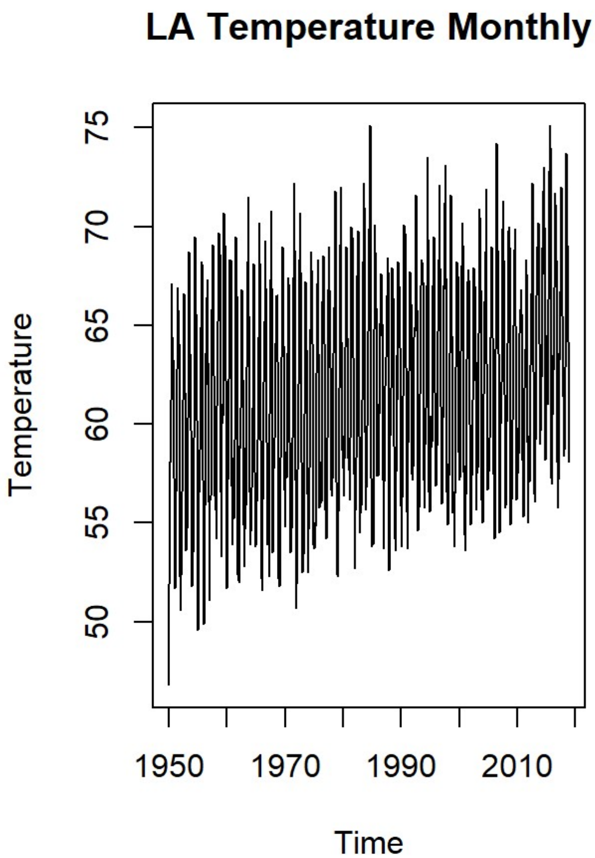
```
## Loading required package: nlme
```

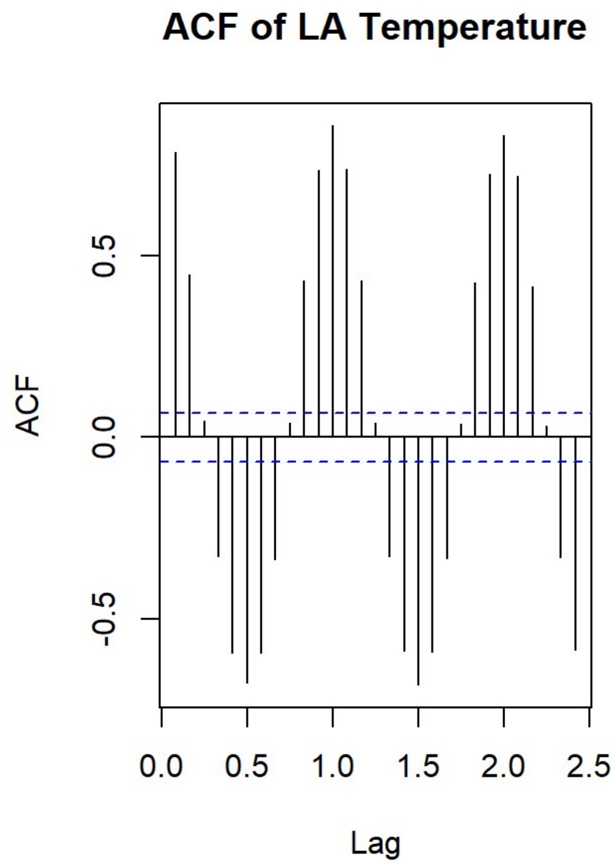
```
## This is mgcv 1.8-28. For overview type 'help("mgcv-package")'.
```

```
fname <- file.choose()
# Load Data
data <- read.csv(fname)
data <- data[,2]
# Convert to TS data in proper frame
temp <- ts(data, start = c(1950,1), freq = 12)
```

Question 1a: Exporatory Data Analysis.

```
par(mfrow=c(1,2))
ts.plot(temp, xlab = 'Time', ylab = 'Temperature', main = 'LA Temperature Monthly')
acf(temp, main = "ACF of LA Temperature")
```





Answer: For assumptions of stationarity, there includes (i) Constant mean ; (ii). finite variance; (iii). covariance does not change when time shifts.

In the plot, the mean temperature seems to increase slightly, and the variances seems to be obviously constant and finite. Therefore the trend violates the assuption of stationarity(i). Due to very clear seasonality in the plot, so we can use seasonlity model but not trend model to fit the data, because trend is very slight increase (looks very weak).

Question 1b: Trend Estimation.

(1). Fit a Moving averaging model.

```
time.pts = c(1:length(temp))
time.pts = c(time.pts - min(time.pts))/max(time.pts)       # range (0, 1)

mav.fit = ksmooth(time.pts, temp, kernel = "box")          # kernel regression
temp.fit.mav = ts(mav.fit$y, start = 1950, frequency = 12)
```

(2). Fit a parametric quadraric polynomial model.

```
x1 = time.pts
x2 = time.pts^2
lm.fit = lm(temp ~ x1+x2)
temp.fit.lm = ts(fitted(lm.fit), start = 1950, frequency = 12)
```

(3). Fit local polynomial (non-parametric) model.

```
loc.fit = loess(temp ~ time.pts)
temp.fit.loc = ts(fitted(loc.fit), start = 1950, frequency = 12)
```

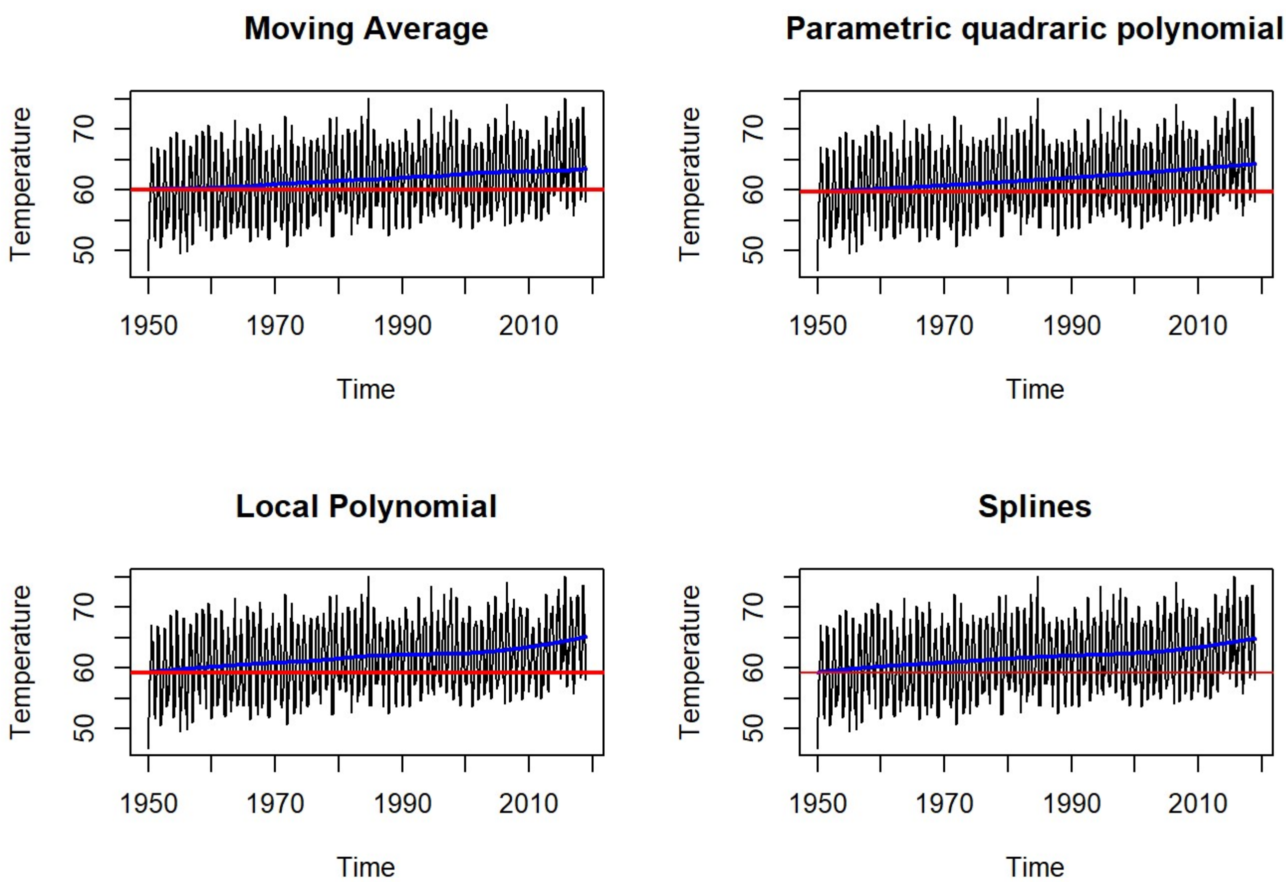
(4). Fit Splines model.

```
gam.fit = gam(temp ~ s(time.pts))
temp.fit.gam = ts(fitted(gam.fit), start = 1950, frequency = 12)
```

a) Plots of models above.

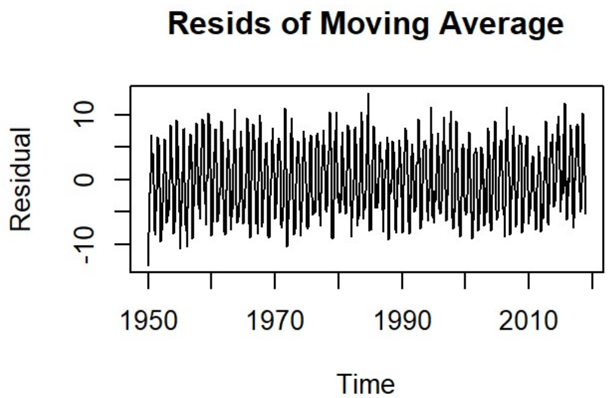
```
par(mfrow = c(2,2))
ts.plot(temp, xlab = 'Time', ylab = 'Temperature', main = 'Moving Average')
lines(temp.fit.mav, lwd=2, col='blue')
abline(temp.fit.mav[1], 0, lwd=2, col = 'red')

ts.plot(temp, xlab = 'Time', ylab = 'Temperature', main = 'Parametric quadraric polyno
mial')
lines(temp.fit.lm, lwd = 2, col = 'blue')
abline(temp.fit.lm[1], 0, lwd = 2, col = 'red')

ts.plot(temp, xlab = 'Time', ylab = 'Temperature', main = 'Local Polynomial')
lines(temp.fit.loc, lwd = 2, col = 'blue')
abline(temp.fit.loc[1], 0, lwd = 2, col = 'red')

ts.plot(temp, xlab = 'Time', ylab = 'Temperature', main = 'Splines')
lines(temp.fit.gam, lwd = 2, col = 'blue')
abline(temp.fit.gam[1], 0, low = 2, col = 'red')
```

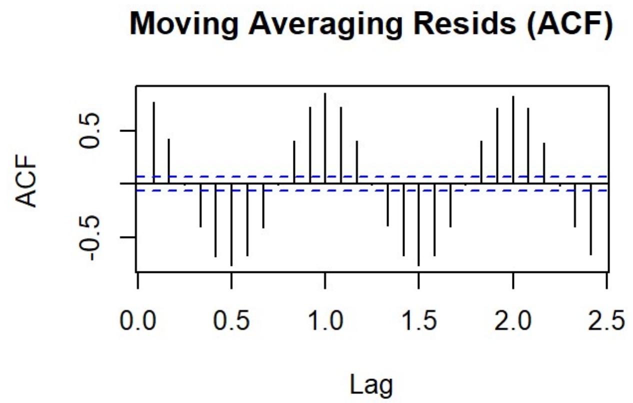
```
## Warning in int_abline(a = a, b = b, h = h, v = v, untf = untf, ...): "low"
## is not a graphical parameter
```

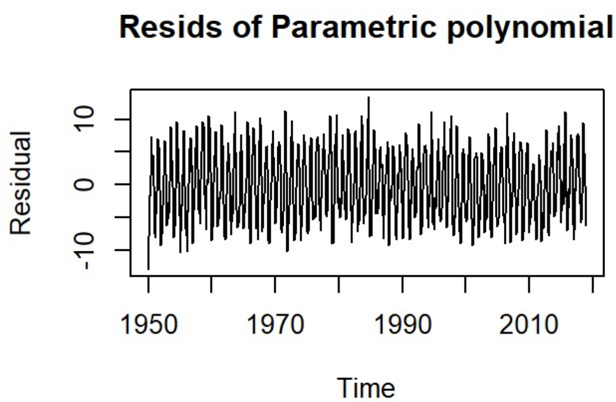


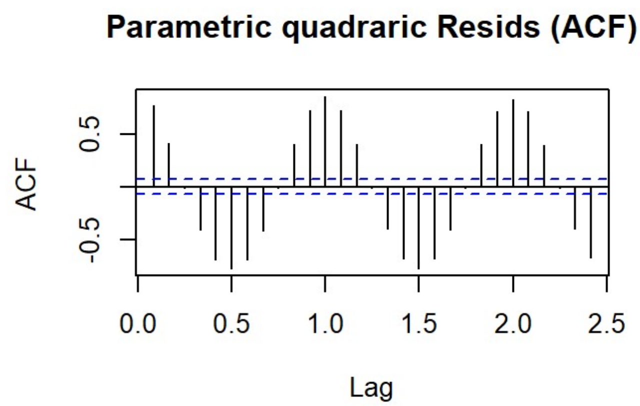
b) Plots of residuals of model above.

```
par(mfrow = c(2,2))
dif.fit.mav = ts(temp - mav.fit$y, start = 1950, frequency = 12)
ts.plot(dif.fit.mav, ylab = 'Residual', main = 'Resids of Moving Average')
acf(dif.fit.mav, main = 'Moving Averaging Resids (ACF)')

dif.fit.lm = ts(temp - fitted(lm.fit), start = 1950, frequency = 12)
ts.plot(dif.fit.lm, ylab = 'Residual', main = 'Resids of Parametric polynomial')
acf(dif.fit.lm, main = 'Parametric quadraric Resids (ACF)')
```

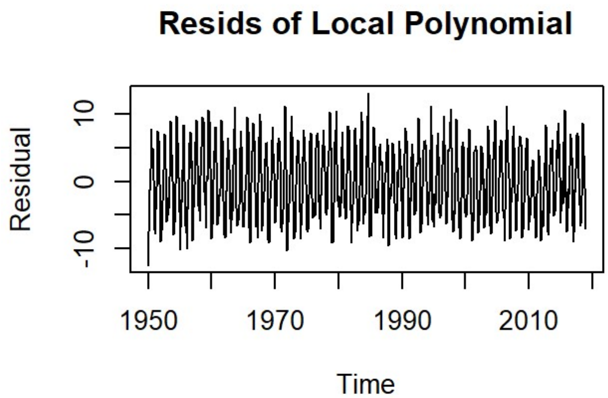


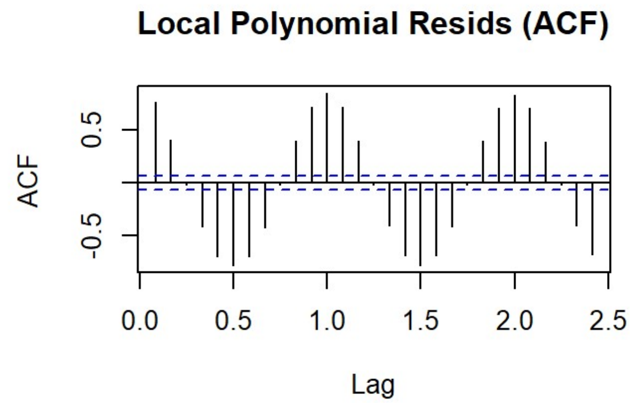


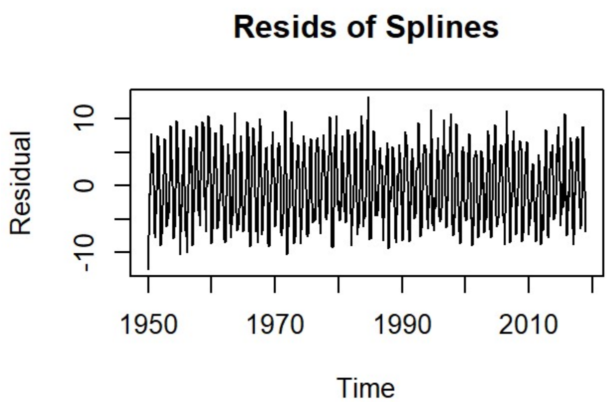


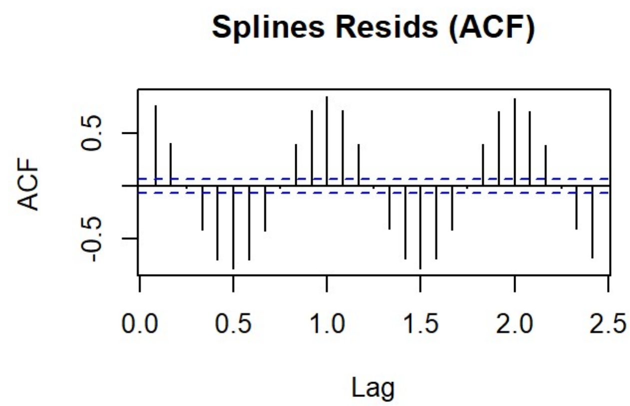



```
dif.fit.loc = ts(temp - fitted(loc.fit), start = 1950, frequency = 12)
ts.plot(dif.fit.loc, ylab = 'Residual', main = 'Resids of Local Polynomial')
acf(dif.fit.loc, main = 'Local Polynomial Resids (ACF)')

dif.fit.gam = ts(temp - fitted(gam.fit), start = 1950, frequency = 12)
ts.plot(dif.fit.gam, ylab = 'Residual', main = 'Resids of Splines')
acf(dif.fit.gam, main = 'Splines Resids (ACF)')
```







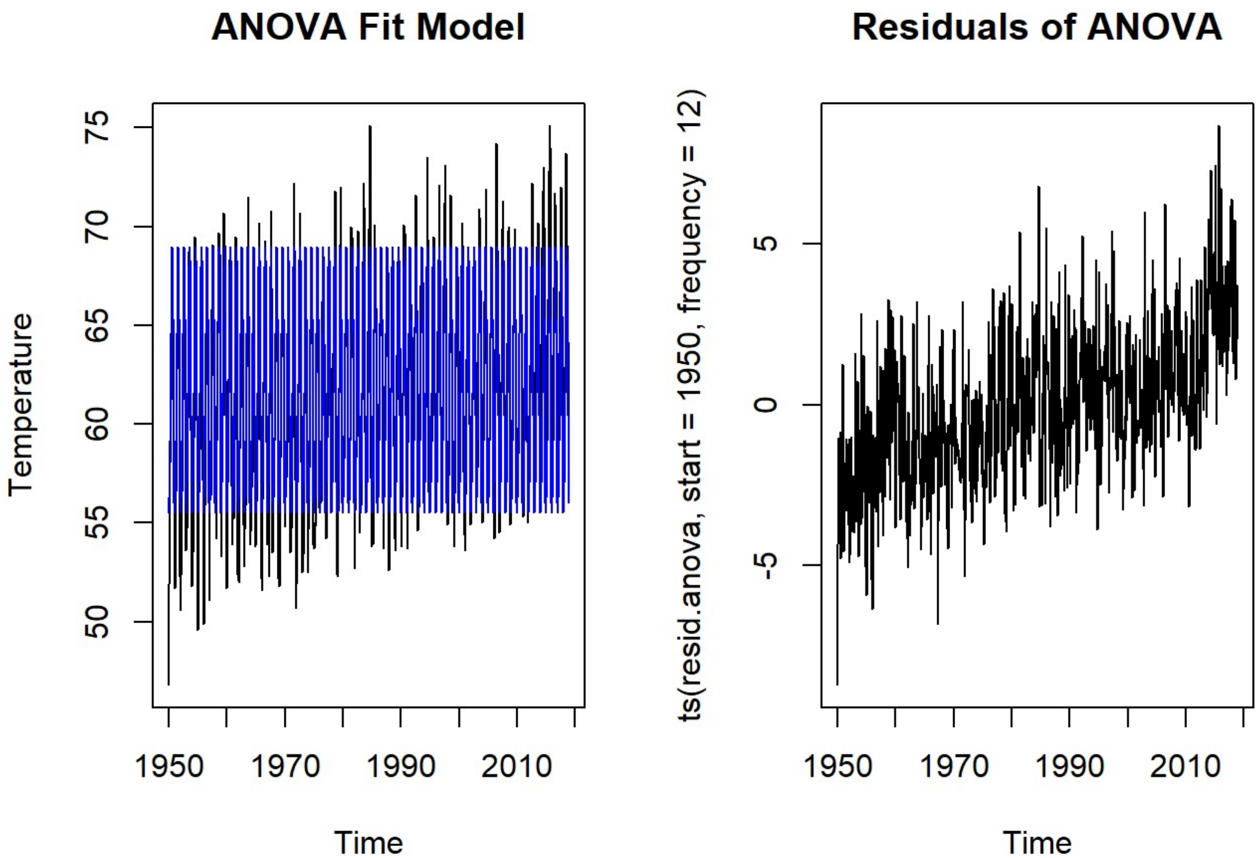


Answer for 1b: (1) In the fitted plots above, they apprear to have slightly increasing trend, and fitted patterns of four models look similar among the 4 different types of trend estimations. (2). In the residuals plot, they appear not to be obvoius trend, and the residuals values are stable with finite variance. All the residuals values look similar in the plots. (3). In ACF of residuals plots, all the four plots also looks similar, which seem to have obvious seasonality patterns and finite variance. Therefore, after taking out the trend from the orginal time series, the residuals violates the stationary assumption (iii), which does not match the assumption of covariance between an two observation depending only on the time lag between them.

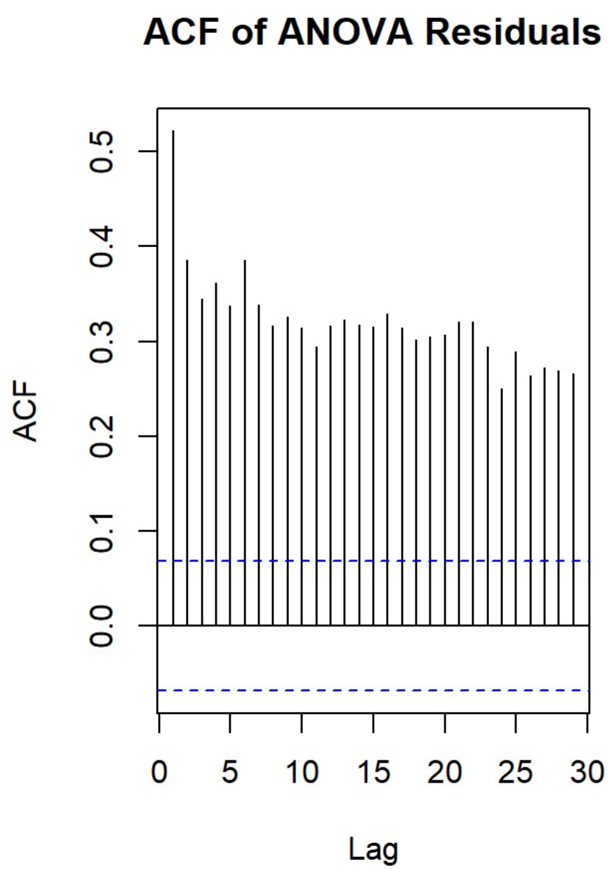
Question 1c: Seasonality estimation.

i). Categorical Linear Regression (ANOVA)

```
par(mfrow = c(1,2))
month = season(temp)
model1 = lm(temp ~ month)                          ## Model with intercept
model1.anova = ts(fitted(model1), start = 1950, frequency = 12)

ts.plot(temp, xlab = 'Time', ylab = 'Temperature', main = 'ANOVA Fit Model')
lines(model1.anova, col = 'blue')

resid.anova = residuals(model1)
ts.plot(ts(resid.anova, start = 1950, frequency = 12), main = 'Residuals of ANOVA')
```

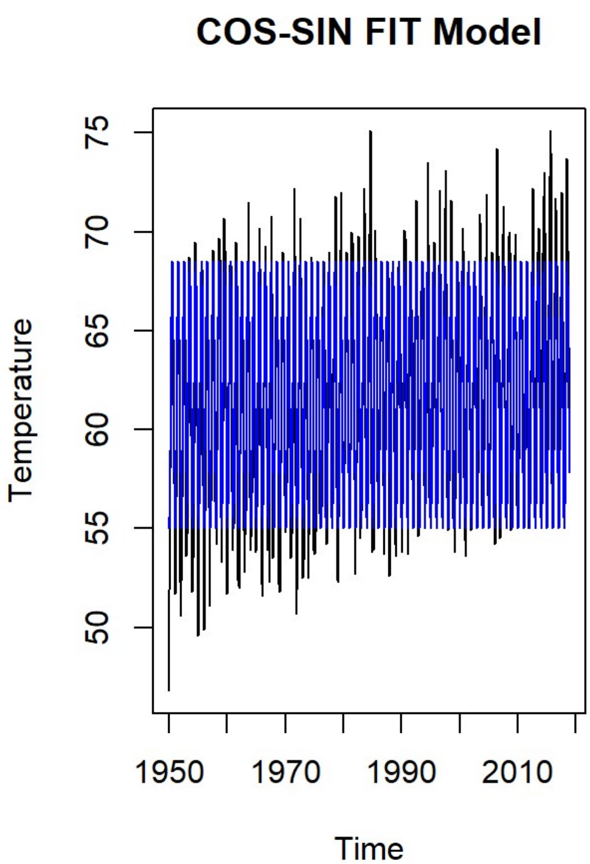


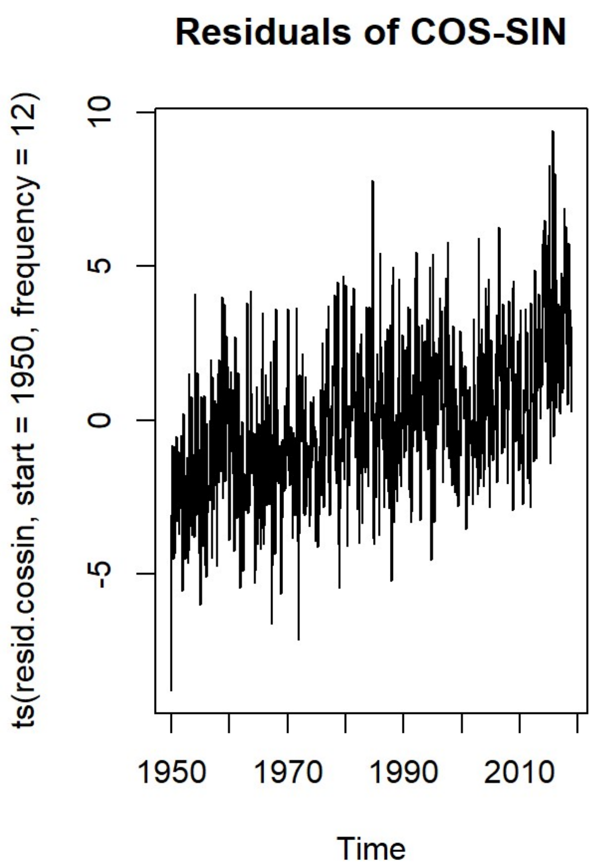

```
acf(resid.anova, main = 'ACF of ANOVA Residuals')
```



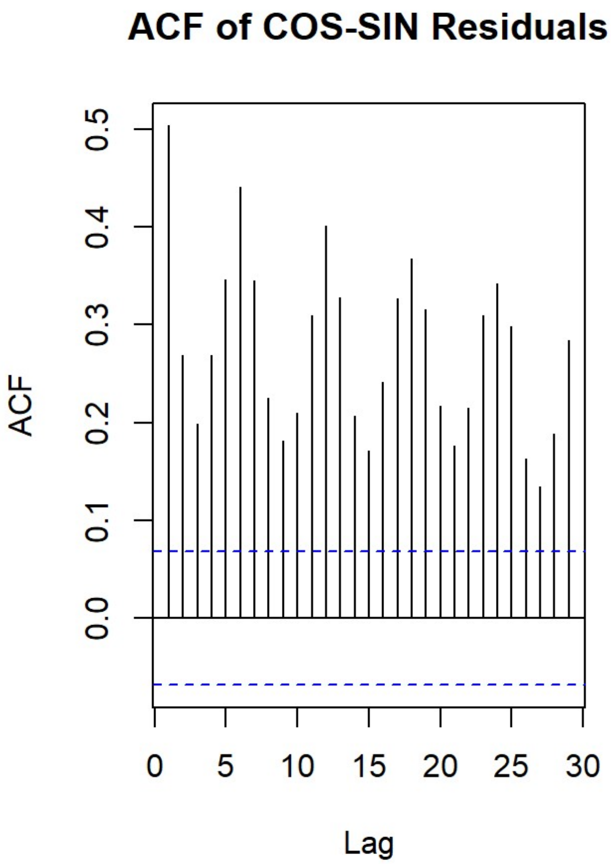
ii). COS-SIN

```
par(mfrow = c(1,2))
har = harmonic(temp,1)
model2 = lm(temp ~ har)
model2.cossin = ts(fitted(model2), start = 1950, frequency = 12)

ts.plot(temp, xlab = 'Time', ylab = 'Temperature', main = 'COS-SIN FIT Model')
lines(model2.cossin, col = 'blue')

resid.cossin = residuals(model2)
ts.plot(ts(resid.cossin, start = 1950, frequency = 12), main = 'Residuals of COS-SIN')
```






```
acf(resid.cossin, main = 'ACF of COS-SIN Residuals')
```



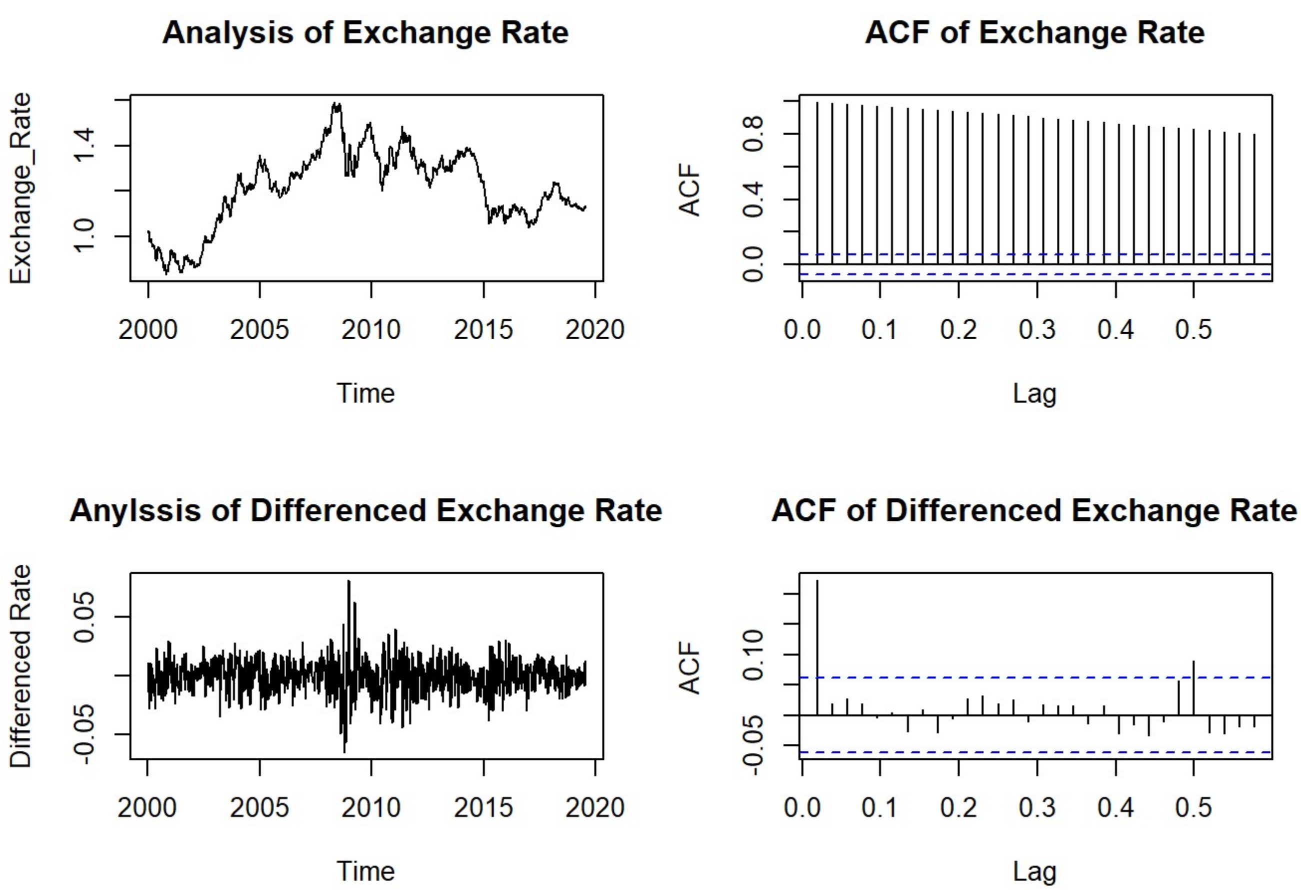
Answer: Our ANOVA models appear to be better fit comparing to our earlier origin models. The COS-SIN model also indicates the clear seasonality, which has been similarly plotted as ANOVA in fit plot. Both residuals plots (rediduals of ANOVA and COS-SIN) appear to have the increasing trend, it violates the assumption of stationarity (i) although their variances of their residuals seem to be constant and finite. However, both ACF of residuals have obvious autocorrelation related to time, which are also significant (above the significant level). So we can found some seasaonlity in the plots, but trend is not clear. Therefore seasonality is appropriate for the temp data, especially when using COS-SIN approach.

Question 2: Current Conversion Analysis.

```
rm(list=ls())
library(TSA)
library(mgcv)
fname <- file.choose()

data <- read.csv(fname)    # Load data
data <- data[,2]

# Convert to TS data in proper frame
rate<- ts(data, start = c(2000, 1), freq = 52)

# Generage differend data
rate.dif <- diff(rate)
```

Question 2a: Exploratory Data Analysis.

```
par(mfrow=c(2,2))
ts.plot(rate, ylab = 'Exchange_Rate', main = 'Analysis of Exchange Rate')
acf(rate, main = 'ACF of Exchange Rate')

ts.plot(rate.dif, ylab = 'Differenced Rate', main = 'Anylssis of Differenced Exchange
Rate')
acf(rate.dif, main = 'ACF of Differenced Exchange Rate')
```



Answer: For assumptions of stationarity, there includes (i) Constant mean; (ii). finite variance; (iii). Covariance function does not change when shifted in time.

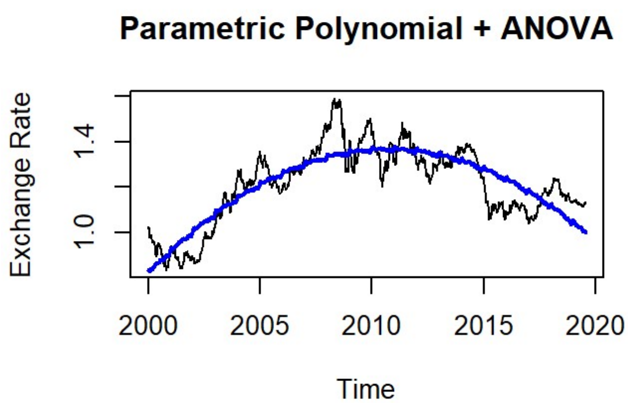
In the plot of Exchange Rate, there seems to have no constant mean, the trend goes up firstly and down later. Variance appear to fluctuate which has few outliers (around 2008). ACF plot shows obvious autocorrelation in the ACF (exchange rate). Therefore, the data is not consistant with assumption of stationary(violate stationary assumption)).

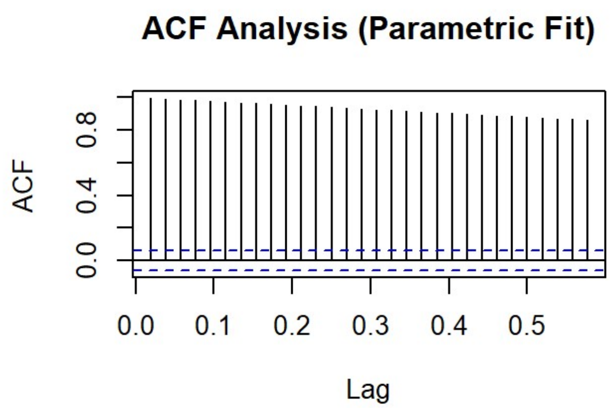
However, in the plot of Differenced Exchange Rate, there appear to have constant mean except few outlier. Variance appears to have evenly distribution. Plot of differenced ACF has no obvious autocorrelaton or seasonality. Therefore, differenced exchange rate seems to have weak stationary.

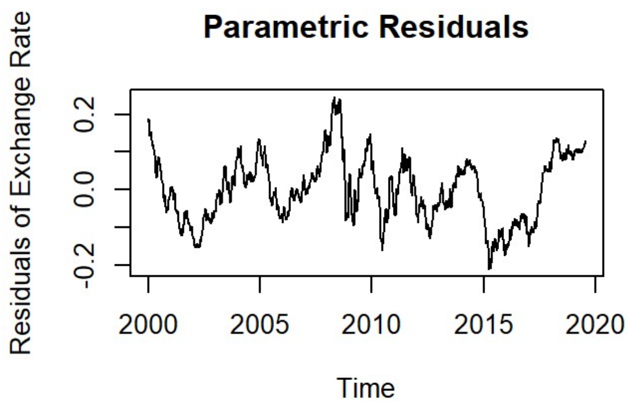
Question 2b: Trend-Seasonality Estimation.

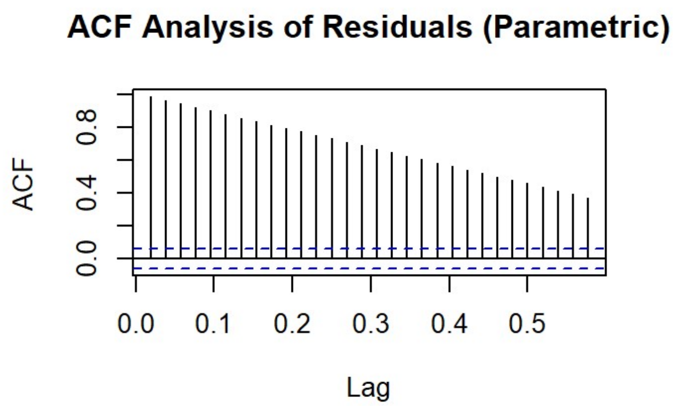
(1). Fit a Parametric Polynomial Regression + ANOVA

```
par(mfrow = c(2,2))
week = season(rate)

time.pts = c(1:length(rate))
time.pts = c(time.pts - min(time.pts))/max(time.pts)          # range(0,1)
x1 = time.pts
x2 = time.pts^2
lm.fit = lm(rate ~ x1+x2 + week)
rate.fit.lm = ts(fitted(lm.fit), start=c(2000,1),freq=52)

ts.plot(rate, ylab ="Exchange Rate", main = 'Parametric Polynomial + ANOVA')
lines(rate.fit.lm, lwd=2, col = "blue")
acf(rate.fit.lm, main = "ACF Analysis (Parametric Fit)")

rate.resid.lm = ts(residuals(lm.fit), start=c(2000,1),freq=52)
ts.plot(rate.resid.lm, ylab = "Residuals of Exchange Rate", main = "Parametric Residua
ls")
acf(rate.resid.lm, main = "ACF Analysis of Residuals (Parametric)")
```





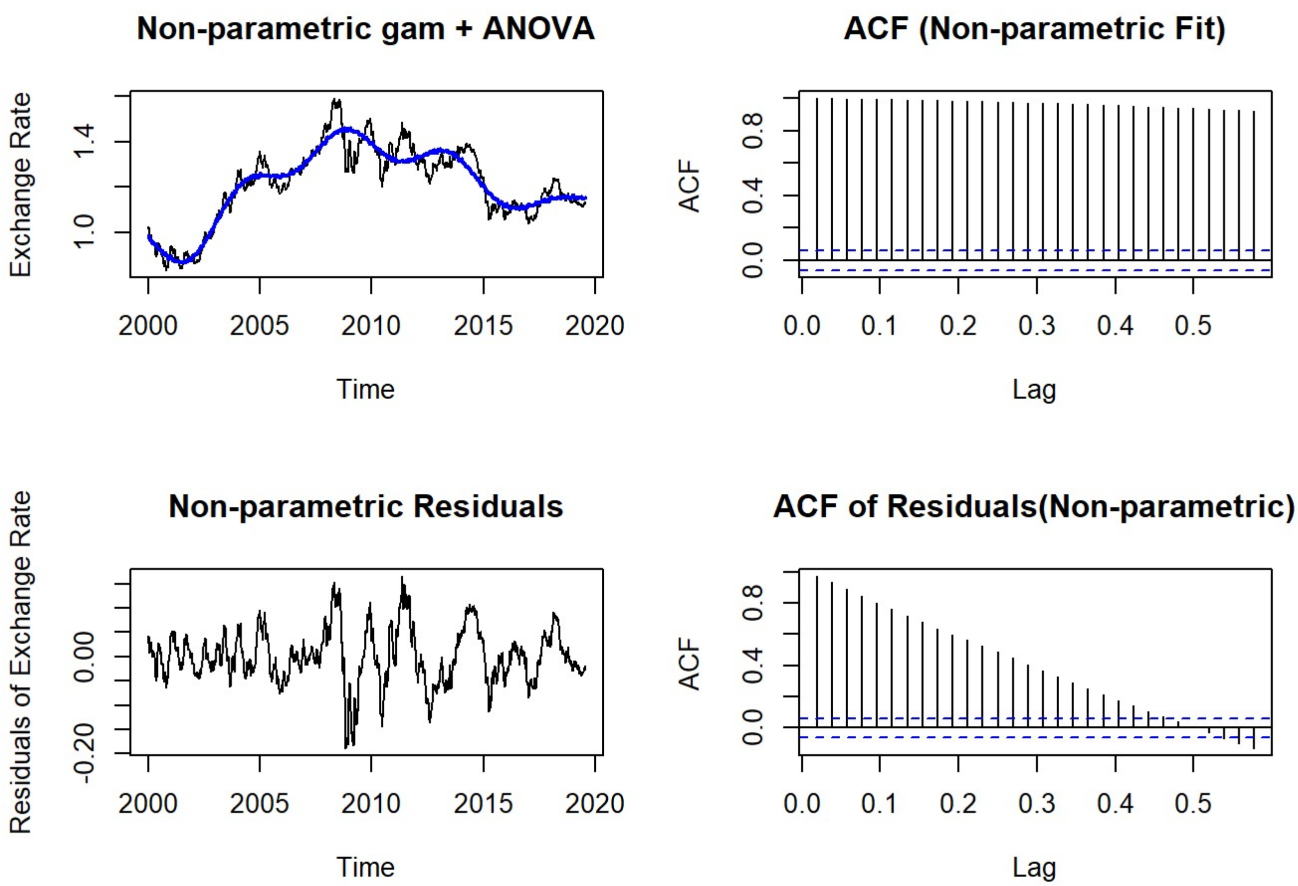




(2) Fit Non-parametric model

Splines Trend Estimation

```
library(mgcv)
gam.fit = gam(rate~s(time.pts) + week)
rate.fit.gam = ts(fitted(gam.fit), start=c(2000,1),freq=52)

par(mfrow=c(2,2))
ts.plot(rate, ylab ="Exchange Rate", main = "Non-parametric gam + ANOVA")
lines(rate.fit.gam, lwd=2, col = "blue")
acf(rate.fit.gam, main = "ACF (Non-parametric Fit)")

rate.resid.gam = ts(residuals(gam.fit), start=c(2000,1),freq=52)
ts.plot(rate.resid.gam, ylab = "Residuals of Exchange Rate", main = "Non-parametric Re
siduals")
acf(rate.resid.gam, main = "ACF of Residuals(Non-parametric)")
```

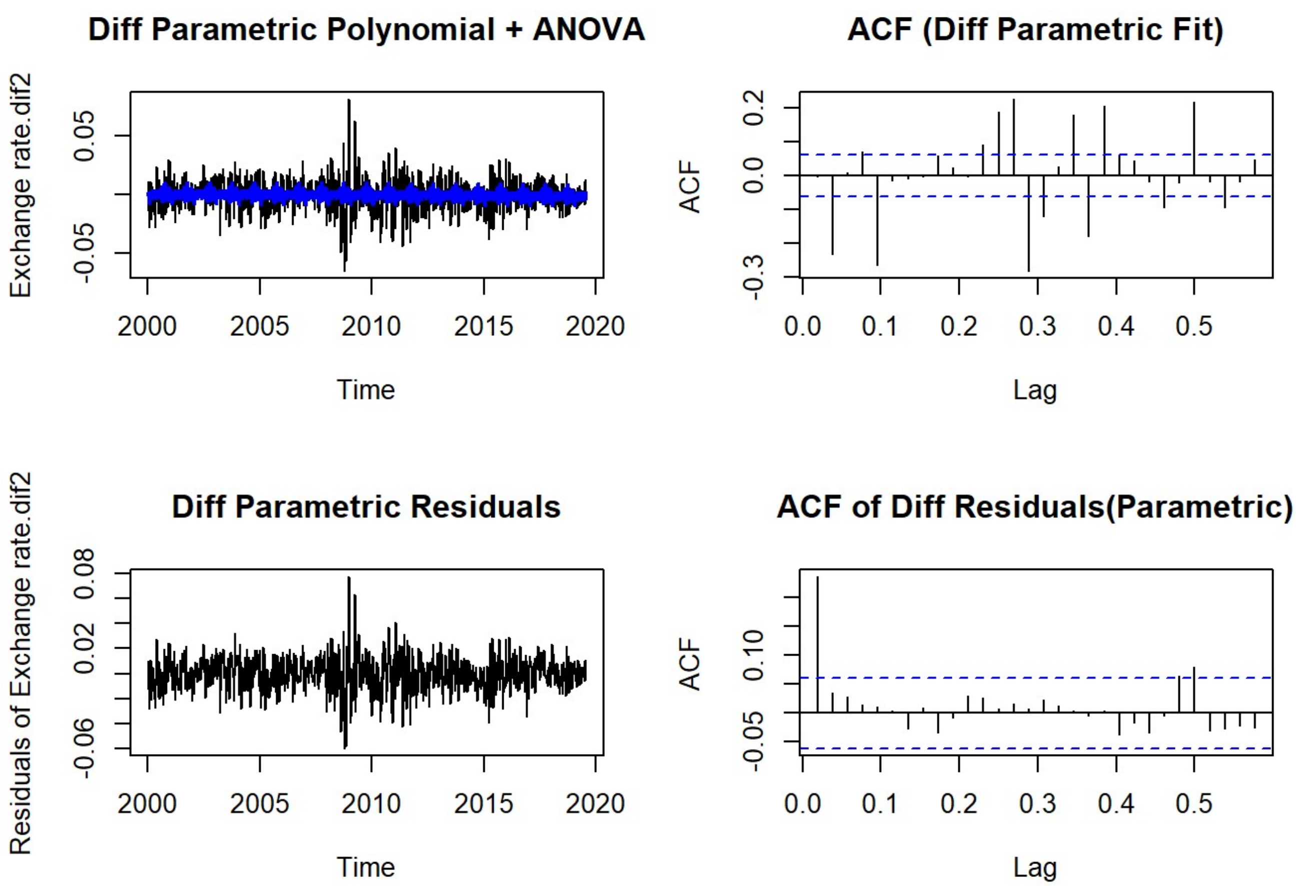


Answer: In both fitted plot, there indicates little fits, the plot of Parametric (Polynomial + ANOVA) fits the data less than No-Parametric(gam + ANOVA). The both residuals plots seem to have some constant mean, but variances flunctuates unevenly. In addition, both ACF (residuals) seems have obvious autocorrelation in ACF plots of residuals of parametric and non-parametric, which drop slowly, thus both are not stationarity.

Question 2c: Trend_Seasonality Estimation with Differenced Data.

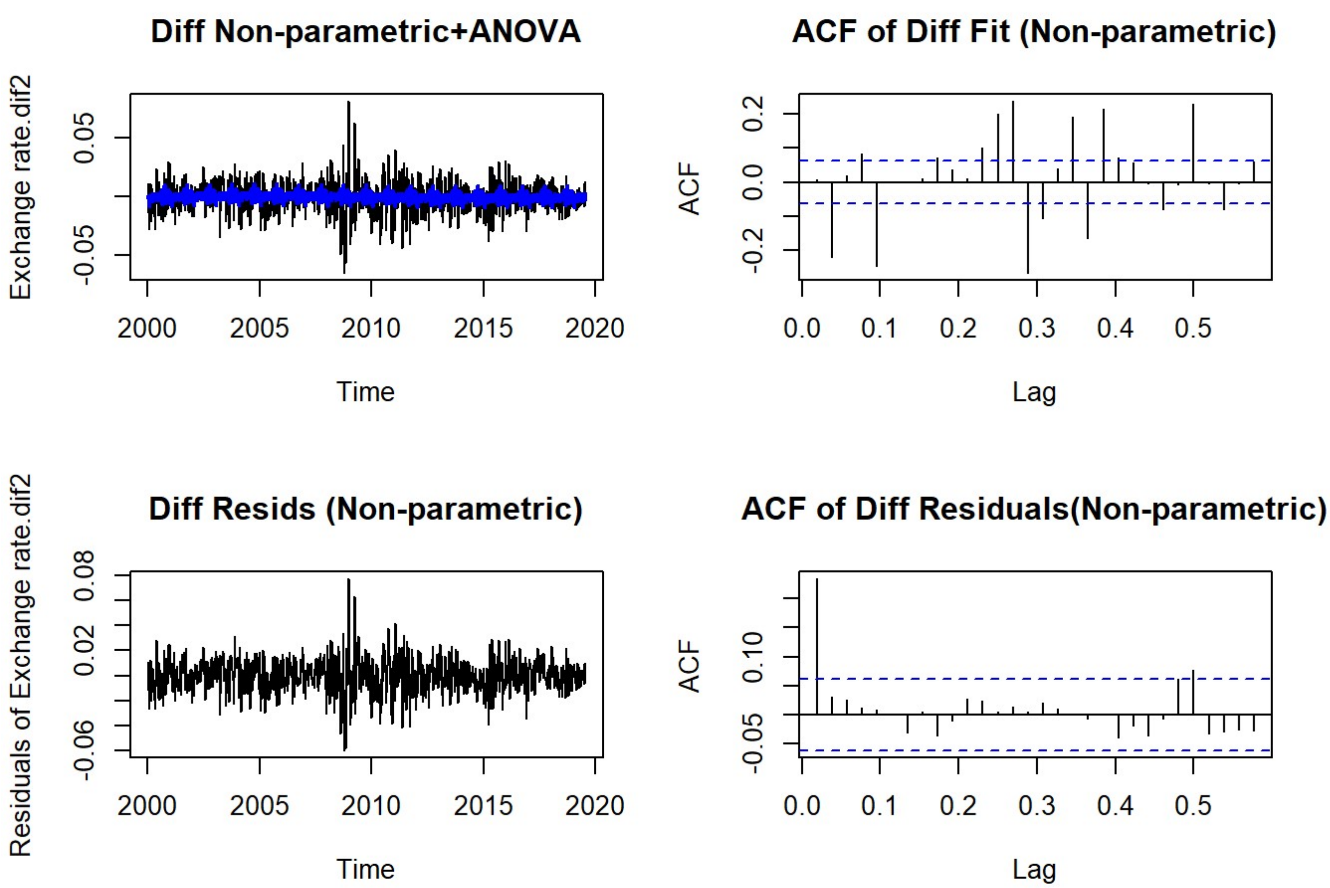
(1). Fit a Parametric (Polynomial Regression + ANOVA) with differenced data

```
rate.dif2 <- diff(rate)
time.pts2 = c(1:length(rate.dif2))
time.pts2 = c(time.pts2 - min(time.pts2))/max(time.pts2)

week = season(ts(rate.dif2, start=c(2000,2),freq=52))       # shift to c(2000,2) when
using diff

x1 = time.pts2
x2 = time.pts2^2
lm.fit2 = lm(rate.dif2 ~ x1+x2 + week)
rate.dif2.fit.lm = ts(fitted(lm.fit2), start=c(2000,2),freq=52)

par(mfrow=c(2,2))
ts.plot(rate.dif2, ylab ="Exchange rate.dif2", main = "Diff Parametric Polynomial + AN
OVA")
lines(rate.dif2.fit.lm, lwd=2, col = "blue")
acf(rate.dif2.fit.lm, main = "ACF (Diff Parametric Fit)")

rate.dif2.resid.lm = ts(residuals(lm.fit2), start=c(2000,2),freq=52)
ts.plot(rate.dif2.resid.lm, ylab = "Residuals of Exchange rate.dif2", main = "Diff Par
ametric Residuals")
acf(rate.dif2.resid.lm, main = "ACF of Diff Residuals(Parametric)")
```



(2) Fit Non-parametric model (Spline Trend + ANOVA) with difference data.

```
library(mgcv)
gam.fit2 = gam(rate.dif2~s(time.pts2) + week)
rate.dif2.fit.gam = ts(fitted(gam.fit2), start=c(2000,2),freq=52)

par(mfrow=c(2,2))
ts.plot(ts(rate.dif2, start = c(2000,2), freq=52), ylab ="Exchange rate.dif2", main =
"Diff Non-parametric+ANOVA")
lines(rate.dif2.fit.gam, lwd=2, col = "blue")
acf(rate.dif2.fit.gam, main = "ACF of Diff Fit (Non-parametric)")

rate.dif2.resid.gam = ts(residuals(gam.fit2), start=c(2000,2),freq=52)
ts.plot(rate.dif2.resid.gam, ylab = "Residuals of Exchange rate.dif2", main = "Diff Re
sids (Non-parametric)")
acf(rate.dif2.resid.gam, main = "ACF of Diff Residuals(Non-parametric)")
```



Answer: For both differenced data, it is difficult to see the pattern in fitting plots, they are no stationary in differenced ACF plots . However, in the residuals plot, both differenced data seems to have constant mean except very few outlier (about 2008), and variance of residuals are constant and finite. In both ACF plots of residuals, there have obvious stationarity, no autocorrelation in the ACF plot, also correlation are not significant.

[Summary], therefore, the models seems to perfectly fit with differenced residual data (both parametric and non-parametric).