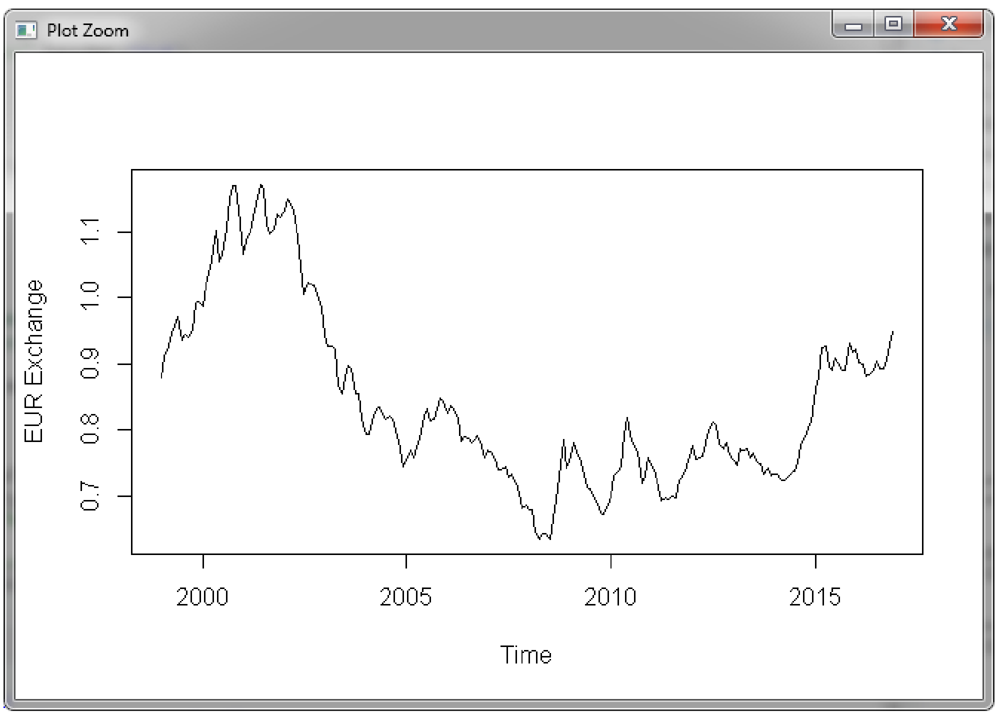
Exchange_rate_USD_EUR.csv

Question 1: Exploratory Data Analysis

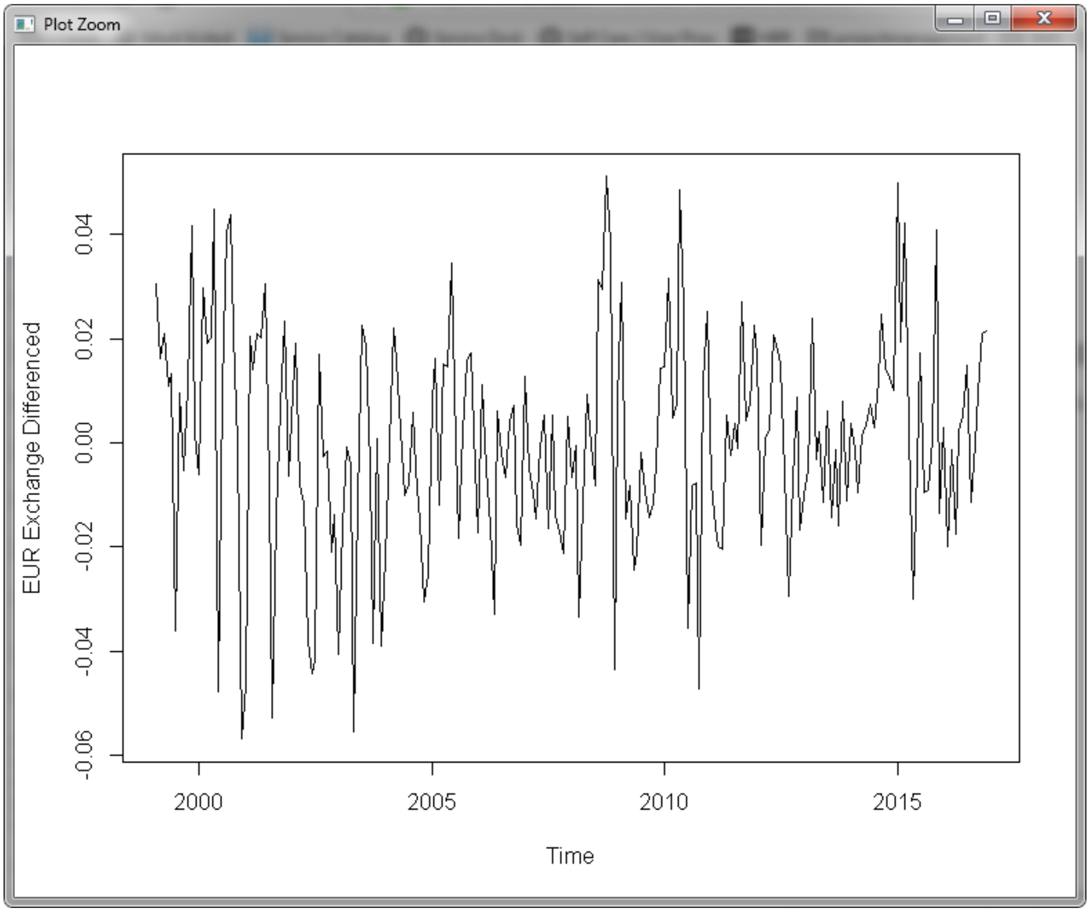
- Display and identify the main features of the time series plot of the data. Which assumptions of stationarity seem to be violated for these three time series?

  Response: the exchange rate over the years seems to be a stochastic process and there are variations in the realizations at different points in time. There is some correlation of future realizations to the realizations before that. We can also see there is a trend from the plot. Between 1999 and 2005 there is an increase in the exchange rate and after that the trend is downward. So the trend is fluctuating and it could change again. We see some cyclic patterns but no clear seasonality. There is not any heteroscedasticity observed. In the case of this time series all three conditions for stationary series – constant mean, finite variance and constant covariance are all violated.



- Display and identify the main features of the 1st order differenced time series plot of the data -- you may use the diff() R command. Which assumptions of stationarity seem to be violated for these three time series?

Response: applying difference=1 we difference the time series as such, x[(1+lag):n] - x[1:(n-lag)]. In the new plot we see that the series has a constant mean. There is no seasonality evidently seen, some periodicity can be observed. This new series is closer to a stationary time series as trend is now removed, seasonality could still be present and autocorrelation may not be constant across lags. This plot looks more like a random walk. An autocorrelation test can confirm the same.



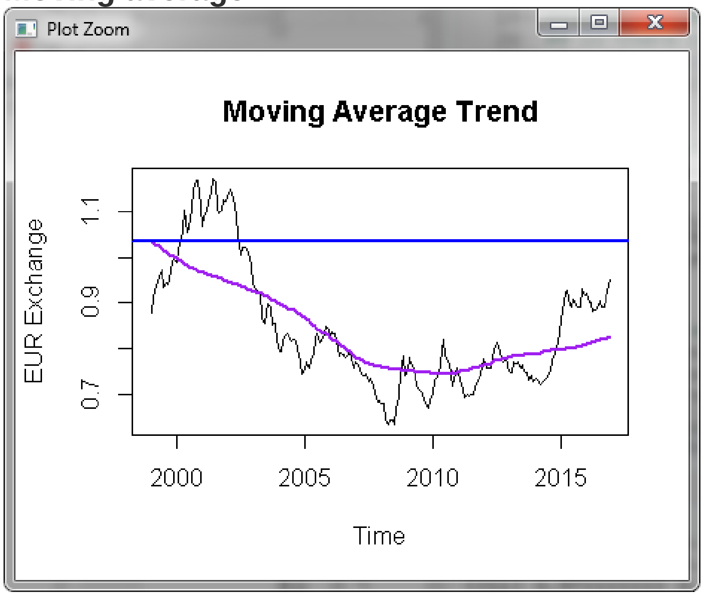
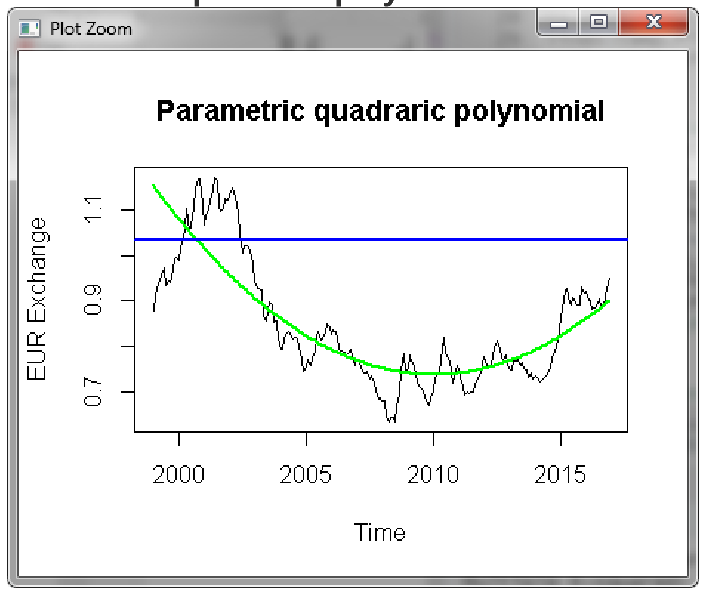
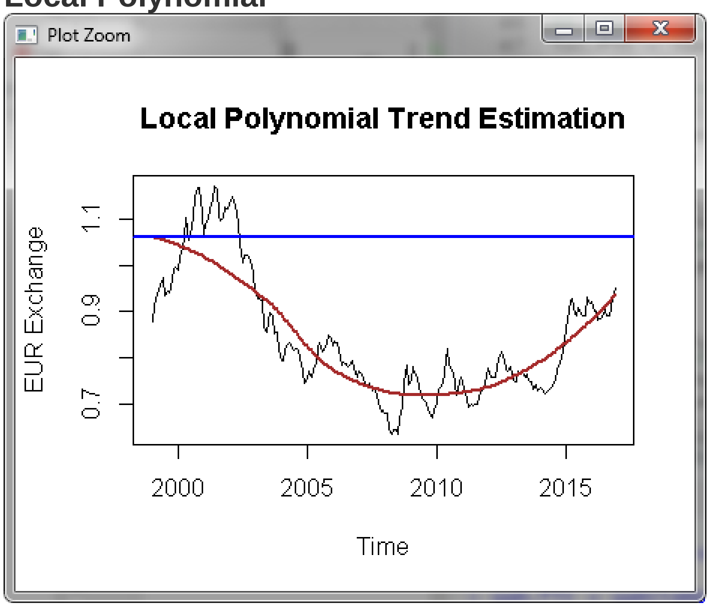
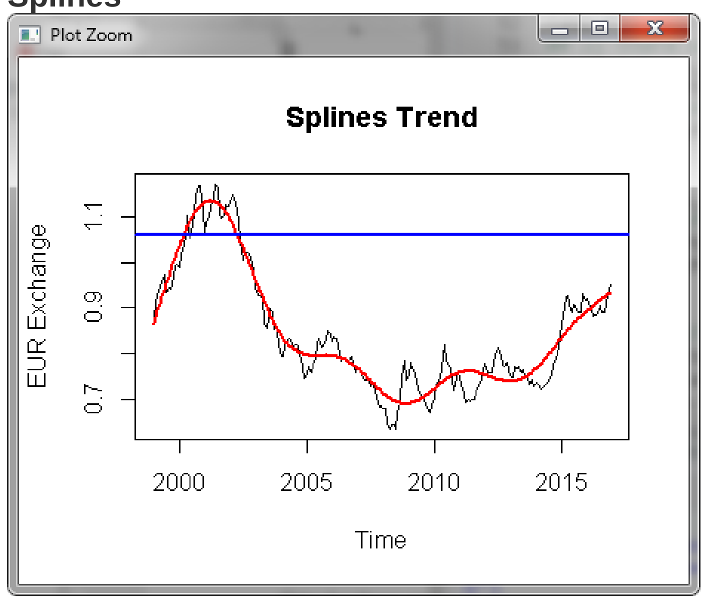
Question 2: Trend Estimation

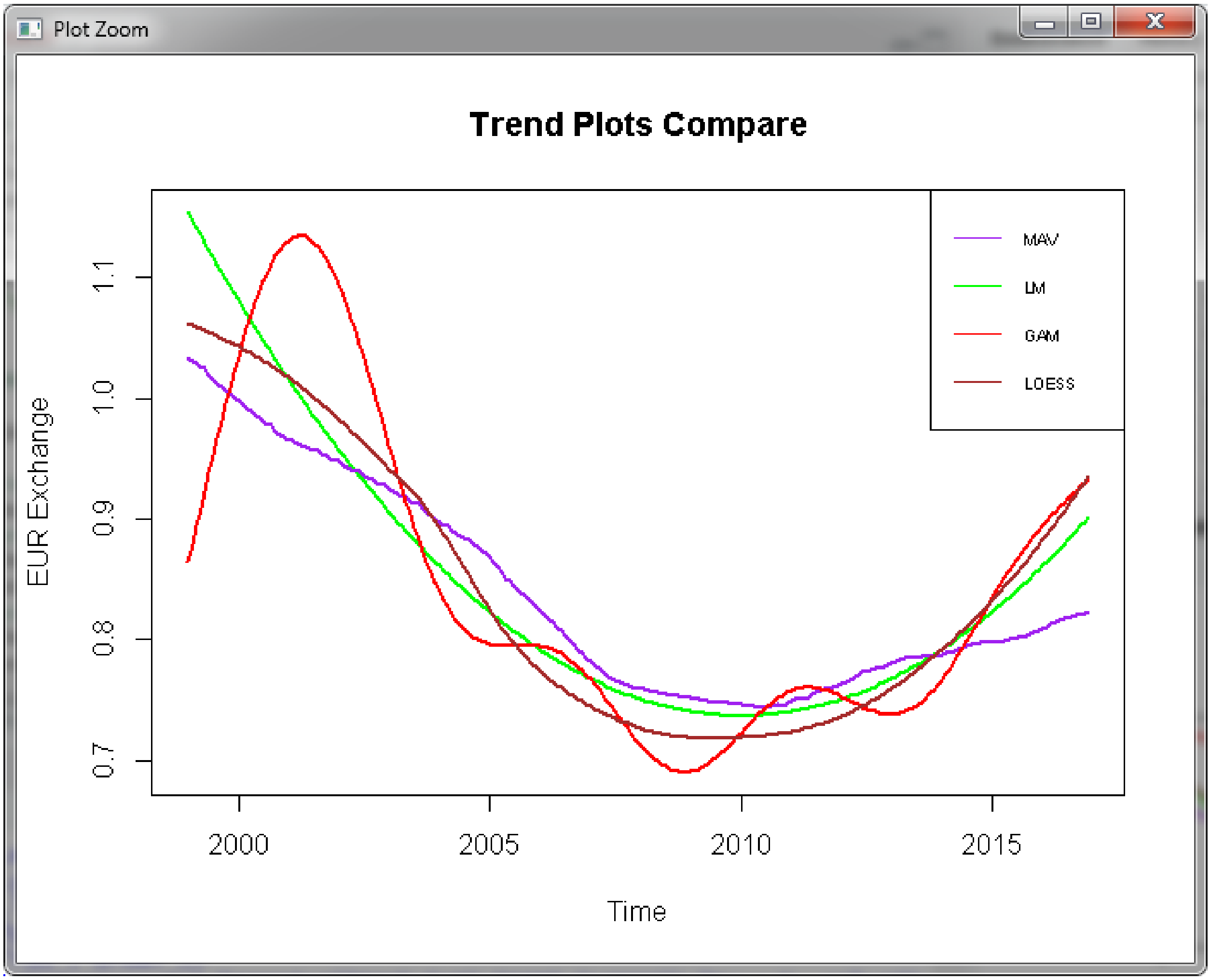
Fit the following models to the original time series to estimate the trend:

- Moving average
- Parametric quadratic polynomial
- Local Polynomial
- Splines

Plot the fitted values along with the original time series plot. Construct and plot the residuals vs fitted values and ACF plots. Comment on the four models fit and on the appropriateness of the stationarity assumption of the residuals.

Response:

Moving average	Parametric quadratic polynomial
	
Local Polynomial	**Splines**
	



Response: The overall trend is downward after 2000 until 2015 and it seems to be changing to an upward trend after that. The latest exchange rate after 2015 is higher than 1999. As expected from the trend plots comparison, we observe that the moving average (MAV) plot captures the trend quite weakly and not a great estimate. The trends generated using linear regression and local polynomial regression are comparable. The splines plot captures the trend more clearly when comparing with the series plot.

Question 3: Seasonality Estimation

Fit the following models to the original time series to estimate monthly seasonality:

- ANOVA approach
- cos-sin model

Plot the fitted values along with the original time series plot. Construct and plot the residuals vs fitted values and ACF plots. Comment on the two models fit and on the appropriateness of the stationarity assumption of the residuals.

Response:-

- **ANOVA**

With intercept:-

```
> model1 = lm(temp~month)
> summary(model1)

Call:
lm(formula = temp ~ month)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.21160 -0.09704 -0.04612  0.07679  0.32598 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)    0.837348   0.032496  25.768   <2e-16 ***
monthFebruary  0.006980   0.045956   0.152    0.879    
monthMarch     0.008564   0.045956   0.186    0.852    
monthApril     0.008492   0.045956   0.185    0.854    
monthMay       0.008665   0.045956   0.189    0.851    
monthJune      0.008252   0.045956   0.180    0.858    
monthJuly      0.005226   0.045956   0.114    0.910    
monthAugust    0.005368   0.045956   0.117    0.907    
monthSeptember 0.006677   0.045956   0.145    0.885    
monthOctober   0.006224   0.045956   0.135    0.892    
monthNovember  0.013960   0.045956   0.304    0.762    
monthDecember  0.005592   0.045956   0.122    0.903    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1379 on 204 degrees of freedom
Multiple R-squared:  0.0005333,	Adjusted R-squared:  -0.05336 
F-statistic: 0.009896 on 11 and 204 DF,  p-value: 1
```

Without intercept:-

```
> model2 = lm(temp~month-1)
> summary(model2)

Call:
lm(formula = temp ~ month - 1)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.21160 -0.09704 -0.04612  0.07679  0.32598 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
monthJanuary     0.8374     0.0325   25.77   <2e-16 ***
monthFebruary    0.8443     0.0325   25.98   <2e-16 ***
monthMarch       0.8459     0.0325   26.03   <2e-16 ***
monthApril       0.8458     0.0325   26.03   <2e-16 ***
monthMay         0.8460     0.0325   26.03   <2e-16 ***
```

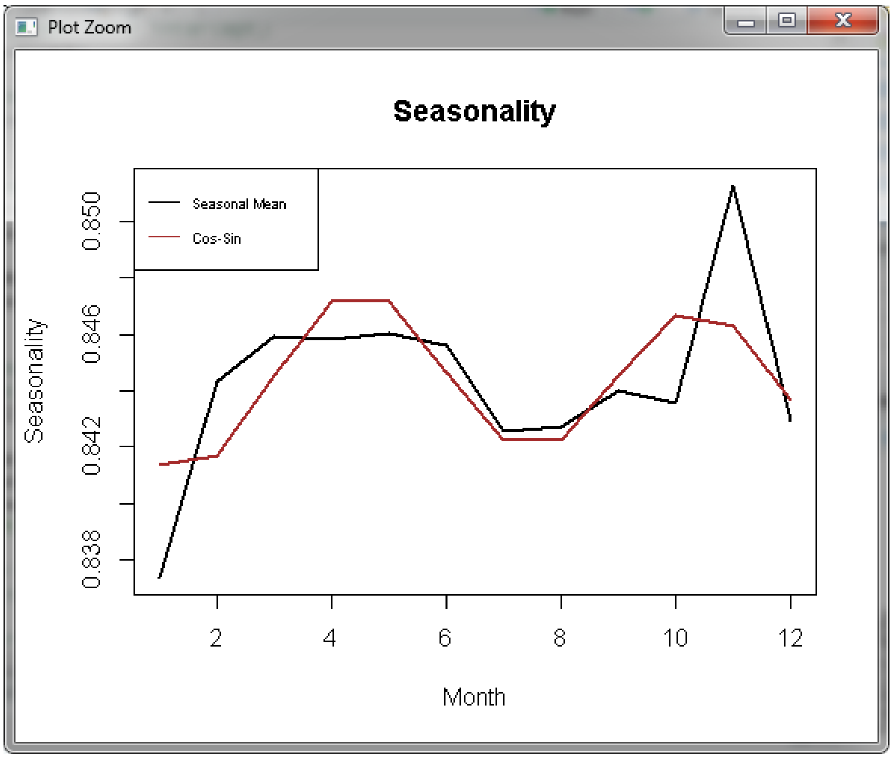
```
monthJune          0.8456     0.0325   26.02   <2e-16 ***
monthJuly          0.8426     0.0325   25.93   <2e-16 ***
monthAugust        0.8427     0.0325   25.93   <2e-16 ***
monthSeptember     0.8440     0.0325   25.97   <2e-16 ***
monthOctober       0.8436     0.0325   25.96   <2e-16 ***
monthNovember      0.8513     0.0325   26.20   <2e-16 ***
monthDecember      0.8429     0.0325   25.94   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1379 on 204 degrees of freedom
Multiple R-squared:  0.9754,   Adjusted R-squared:  0.974
F-statistic: 675.1 on 12 and 204 DF,  p-value: < 2.2e-16
```

- **cos-sin model**

```
> har=harmonic(temp,1)
> model3=lm(temp~har)
> summary(model3)

Call:
lm(formula = temp ~ har)

Residuals:
     Min       1Q   Median       3Q      Max
-0.21035 -0.09759 -0.04530  0.07811  0.32675

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     0.8443477  0.0091829  91.948   <2e-16 ***
harcos(2*pi*t) -0.0004237  0.0129866  -0.033    0.974
harsin(2*pi*t)  0.0002418  0.0129866   0.019    0.985
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.135 on 213 degrees of freedom
Multiple R-squared:  6.625e-06,    Adjusted R-squared:  -0.009383
F-statistic: 0.0007056 on 2 and 213 DF,  p-value: 0.9993
```

```
> har2=harmonic(temp,2)
> model4=lm(temp~har2)
> summary(model4)

Call:
lm(formula = temp ~ har2)

Residuals:
     Min       1Q   Median       3Q      Max
-0.21290 -0.09873 -0.04649  0.07845  0.32692

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      0.8443477  0.0092253  91.526   <2e-16 ***
har2cos(2*pi*t) -0.0004237  0.0130465  -0.032    0.974
har2cos(4*pi*t) -0.0025546  0.0130465  -0.196    0.845
```

```
har2sin(2*pi*t)  0.0002418  0.0130465   0.019    0.985
har2sin(4*pi*t) -0.0012818  0.0130465  -0.098    0.922
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1356 on 211 degrees of freedom
Multiple R-squared:  0.000234, Adjusted R-squared:  -0.01872
F-statistic: 0.01235 on 4 and 211 DF,  p-value: 0.9997
```



We see a peak at November in both cases with Seasonal Mean and Cos-Sin models which indicates a seasonality. The R squared values are very small in the LM model with intercept, indicates Seasonality does not explain much the variance in the monthly average. In the model which does not have intercept the multiple r squared value is very high with a low p-value and this explains that seasonality has a good influence on the model output.

Question 4: Trend and Seasonality Estimation

Fit the following models to the original time series to estimate the trend and seasonality:

- Linear regression
- Non-parametric model

Plot the fitted values along with the original time series plot. Construct and plot the residuals vs fitted and ACF plots. Comment on the two models fit and on the appropriateness of the stationarity assumption of the r

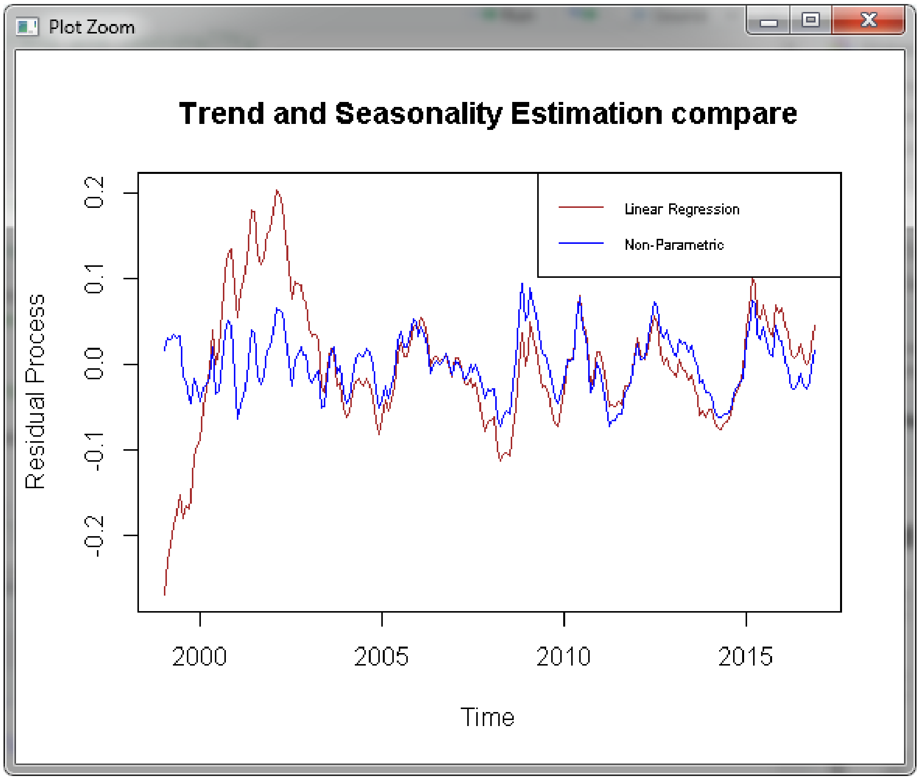
Response:

Linear Regression

```
> x1 = time.pts
> x2 = time.pts^2
> har2=harmonic(temp,2)
> lm.fit = lm(temp~x1+x2+har2)
> summary(lm.fit)

Call:
lm(formula = temp ~ x1 + x2 + har2)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.269429 -0.042119 -0.000513  0.040098  0.204879 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)      1.154020   0.015514  74.385   <2e-16 ***
x1              -1.364160   0.071987 -18.950   <2e-16 ***
x2               1.115483   0.070000  15.935   <2e-16 ***
har2cos(2*pi*t) -0.001932   0.007379  -0.262    0.794    
har2cos(4*pi*t) -0.003801   0.007379  -0.515    0.607    
har2sin(2*pi*t) -0.004055   0.007385  -0.549    0.584    
har2sin(4*pi*t) -0.003276   0.007380  -0.444    0.658    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07668 on 209 degrees of freedom
Multiple R-squared:  0.6833,	Adjusted R-squared:  0.6742 
F-statistic: 75.15 on 6 and 209 DF,  p-value: < 2.2e-16
```

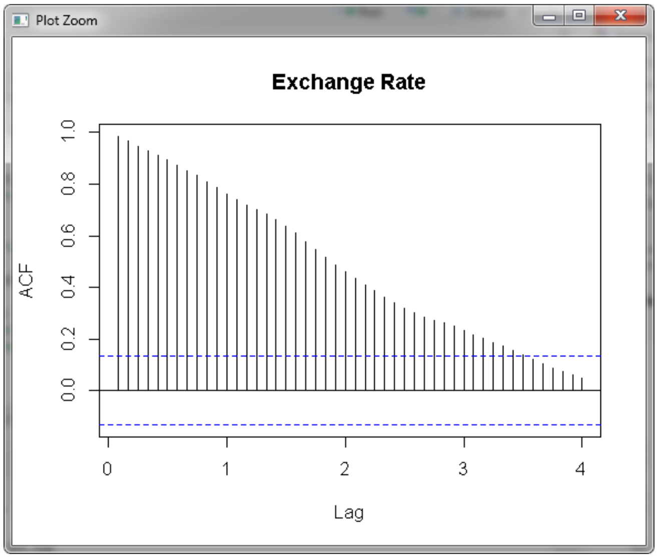
```
> summary(gam.fit)

Family: gaussian 
Link function: identity 

Formula:
temp ~ s(time.pts) + har2

Parametric coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)      0.8443477  0.0024808 340.357   <2e-16 ***
har2cos(2*pi*t)  0.0003117  0.0035101   0.089    0.929    
har2cos(4*pi*t) -0.0021118  0.0035098  -0.602    0.548    
har2sin(2*pi*t)  0.0014972  0.0035284   0.424    0.672    
har2sin(4*pi*t) -0.0006996  0.0035127  -0.199    0.842    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
              edf Ref.df   F p-value    
s(time.pts) 8.885  8.996 301  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

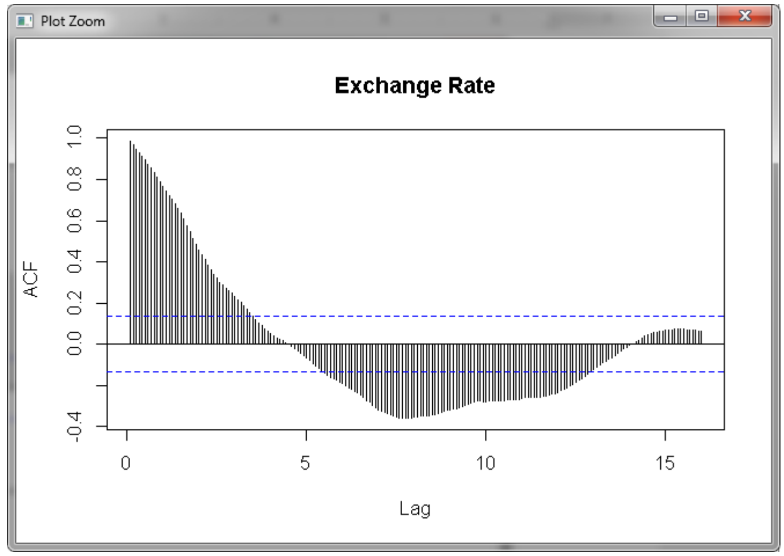
```
R-sq.(adj) =  0.926   Deviance explained = 93.1%
GCV = 0.0014206  Scale est. = 0.0013293  n = 216
```



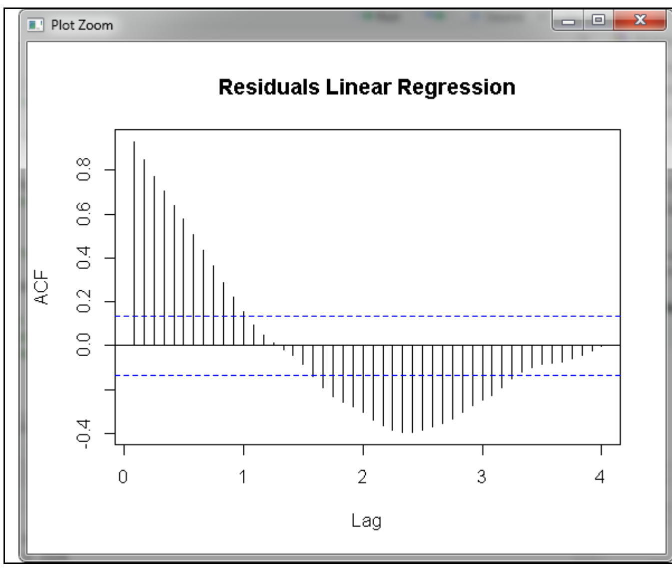
Response: Seasonality, as per the Seasonal Means and Cosine Sine model output, is already in the model, we will add trend to test whether trend adds to the explanatory power of the model. We do see that the R squared values are reasonably higher which indicates seasonality and trend influence the model output. This indicates that trend adds to the explanatory power of the model. The difference between linear and non-parametric model is negligible in most cases.

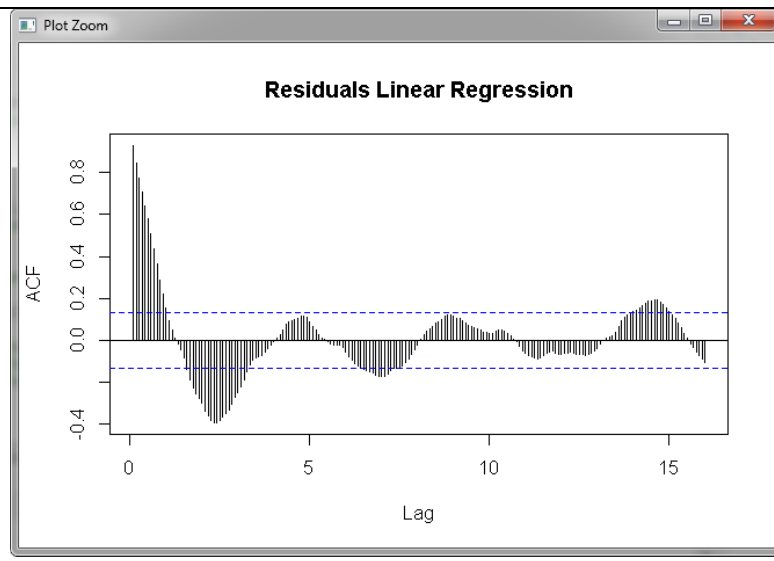
ACF:-



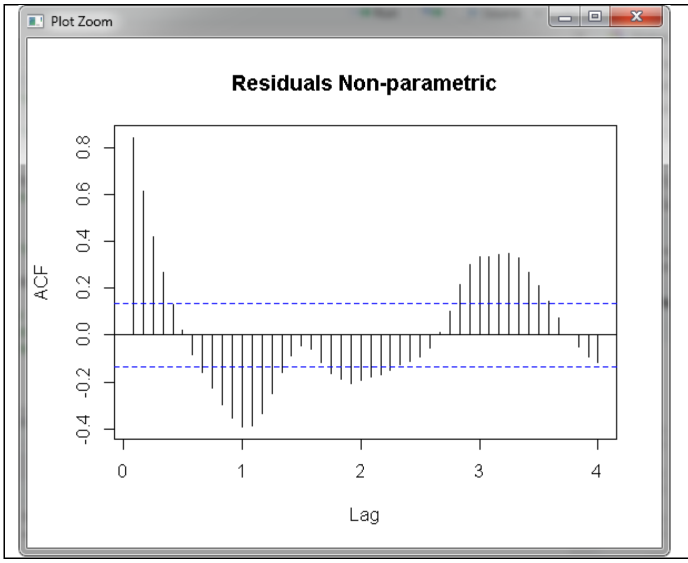


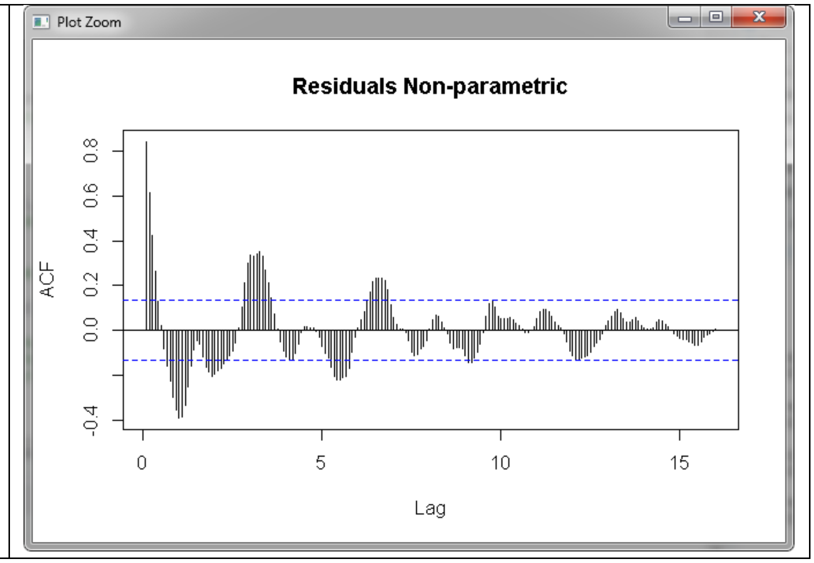
Response: we don't see any seasonality in the ACF plot as it is decaying to 0 pretty fast, there is some trend and the trend is changing over longer period of time.





Response: some cyclical patterns are present when more lags are added, there is no clear indication of seasonality.





Response: Some cyclic patterns are present in the case of residual acf plot also. This plot does have an indication of stationary series.

R Code:-

```
rm(list=ls())
setwd("C:\\RProgs")

library(data.table)
library(xts)
#Load USD/EUR data
data=read.csv("exchange_rate_USD_EUR.csv",header=TRUE)
data
temp = as.vector(t(data[,-1]))
temp
startyear=1999
temp = ts(temp,start=startyear,frequency=12)
temp
names(data)

ts.plot(temp,ylab="EUR Exchange")

#difference first order
xts = diff(temp, lag=1,differences = 1)
ts.plot(xts,ylab="EUR Exchange Differenced")

############### TREND ESTIMATION ###################
## Is there a trend in the average temperature?
time.pts = c(1:length(temp))
time.pts = c(time.pts - min(time.pts))/max(time.pts)
time.pts

## Fit a moving average
mav.fit = ksmooth(time.pts, temp, kernel = "box")
temp.fit.mav = ts(mav.fit$y,start=startyear,frequency=12)
ts.plot(temp,ylab="EUR Exchange",main="Moving Average Trend")
lines(temp.fit.mav,lwd=2,col="purple")
abline(temp.fit.mav[1],0,lwd=2,col="blue")
```

```
## Fit a parametric quadraric polynomial
x1 = time.pts
x2 = time.pts^2
lm.fit = lm(temp~x1+x2)
summary(lm.fit)
temp.fit.lm = ts(fitted(lm.fit),start=startyear,frequency=12)
ts.plot(temp,ylab="EUR Exchange",main="Parametric quadraric polynomial")
lines(temp.fit.lm,lwd=2,col="green")
abline(temp.fit.mav[1],0,lwd=2,col="blue")
#lines(temp.fit.mav,lwd=2,col="purple")  #mav line


## Fit a trend using non-parametric regression
## Local Polynomial Trend Estimation
loc.fit = loess(temp~time.pts)
temp.fit.loc = ts(fitted(loc.fit),start=startyear,frequency=12)
## Splines Trend Estimation
library(mgcv)
gam.fit = gam(temp~s(time.pts))
temp.fit.gam = ts(fitted(gam.fit),start=startyear,frequency=12)
## Is there a trend?
ts.plot(temp,ylab="EUR Exchange",main="Local Polynomial Trend Estimation")
lines(temp.fit.loc,lwd=2,col="brown")
#lines(temp.fit.gam,lwd=2,col="red")
abline(temp.fit.loc[1],0,lwd=2,col="blue")


ts.plot(temp,ylab="EUR Exchange",main="Splines Trend")
#lines(temp.fit.loc,lwd=2,col="brown")
lines(temp.fit.gam,lwd=2,col="red")
abline(temp.fit.loc[1],0,lwd=2,col="blue")


## Compare all estimated trends
all.val = c(temp.fit.mav,temp.fit.lm,temp.fit.gam,temp.fit.loc)
ylim= c(min(all.val),max(all.val))
ts.plot(temp.fit.lm,lwd=2,col="green",ylim=ylim,ylab="EUR Exchange",main="Trend Plots Compare")
lines(temp.fit.mav,lwd=2,col="purple")
lines(temp.fit.gam,lwd=2,col="red")
lines(temp.fit.loc,lwd=2,col="brown")
legend(x="topright",cex = 0.55,legend=c("MAV","LM","GAM","LOESS"),lty = 1, col=c("purple","green","red","brown"))

################ SEASONALITY ESTIMATION #########################
library(TSA)
## Estimate seasonality using ANOVA approach
month = season(temp)
month
## Drop January (model with intercept)
model1 = lm(temp~month)
summary(model1)
## All seasonal mean effects (model without intercept)
model2 = lm(temp~month-1)
summary(model2)

## Estimate seasonality using cos-sin model
har=harmonic(temp,1)
model3=lm(temp~har)
```

```
summary(model3)
har2=harmonic(temp,2)
model4=lm(temp~har2)
summary(model4)

## Compare Seasonality Estimates
## Seasonal Means Model
st1 = coef(model2)
## Cos-Sin Model
st2 = fitted(model4)[1:12]
plot(1:12,st1,lwd=2,type="l",xlab="Month",ylab="Seasonality",main="Seasonality")
lines(1:12,st2,lwd=2, col="brown")
legend(x="topleft",cex = 0.55,legend=c("Seasonal Mean","Cos-Sin"),lty = 1, col=c("black","brown"))

################ TREND AND SEASONALITY ESTIMATION #########################
## Using linear regression

## Fit a parametric model for both trend and seasonality
x1 = time.pts
x2 = time.pts^2
har2=harmonic(temp,2)
lm.fit = lm(temp~x1+x2+har2)
summary(lm.fit)
dif.fit.lm = ts((temp-fitted(lm.fit)),start=startyear,frequency=12)
ts.plot(dif.fit.lm,ylab="Residual Process")

## Fit a non-parametric model for trend and linear model for seasonality
gam.fit = gam(temp~s(time.pts)+har2)
summary(gam.fit)
dif.fit.gam = ts((temp-fitted(gam.fit)),start=startyear,frequency=12)
ts.plot(dif.fit.gam,ylab="Residual Process")

## Compare approaches
ts.plot(dif.fit.lm,ylab="Residual Process",col="brown",main="Trend and Seasonality Estimation compare")
lines(dif.fit.gam,col="blue")
legend(x="topright",cex = 0.55,legend=c("Linear Regression","Non-Parametric"),lty = 1, col=c("brown","blue"))

acf(temp,lag.max=12*16,main="Exchange Rate")
acf(dif.fit.lm,lag.max=12*16,main="Residuals Linear Regression")
acf(dif.fit.gam,lag.max=12*16,main="Residuals Non-parametric")
```