**STAT 535: Forecasting Methods for Management**
Assignment - 2
**Renuka Chintalapati**

The file ConstructionSpending.txt gives monthly total US construction spending (as in millions of dollars). This dataset is for the years from 1994 to 2022. We have the spending data, log spending data, log return of the spending data and the variable Changepoint that is zero from 1994 to 2008 and 1 thereafter. There are several dummies as well.

rconstruct<-read.csv("/Users/renukachintalapati/Downloads/ConstructionSpending.txt")
attach(rconstruct)

Time<-as.numeric(Time)
fMonth<-as.factor(Month)

#Augmenting fMonth to rconstruct:
rconstruct<-data.frame(rconstruct,fMonth)

#Checking a sample of rconstruct dataframe:
head(rconstruct)

**OUTPUT:**

```
> head(rconstruct)
        Date Month Time Spending logSpending dlogSpending Changepoint        c348
1 1994-01-01     1    1    34917    10.46073  -0.12421254           0 -0.57757270
2 1994-02-01     2    2    33222    10.41097  -0.04976151           0 -0.33281954
3 1994-03-01     3    3    38215    10.55098   0.14001580           0  0.96202767
4 1994-04-01     4    4    41754    10.63955   0.08856715           0 -0.77846230
5 1994-05-01     5    5    45711    10.73009   0.09054371           0 -0.06279052
6 1994-06-01     6    6    49032    10.80023   0.07013418           0  0.85099448
        s348       c432       s432 obs169 obs205 obs316 fMonth
1  0.8163393 -0.9101060  0.4143756      0      0      0      1
2 -0.9429905  0.6565858 -0.7542514      0      0      0      2
3  0.2729519 -0.2850193  0.9585218      0      0      0      3
4  0.6276914 -0.1377903 -0.9904614      0      0      0      4
5 -0.9980267  0.5358268  0.8443279      0      0      0      5
6  0.5251746 -0.8375280 -0.5463943      0      0      0      6
```

tail(rconstruct)

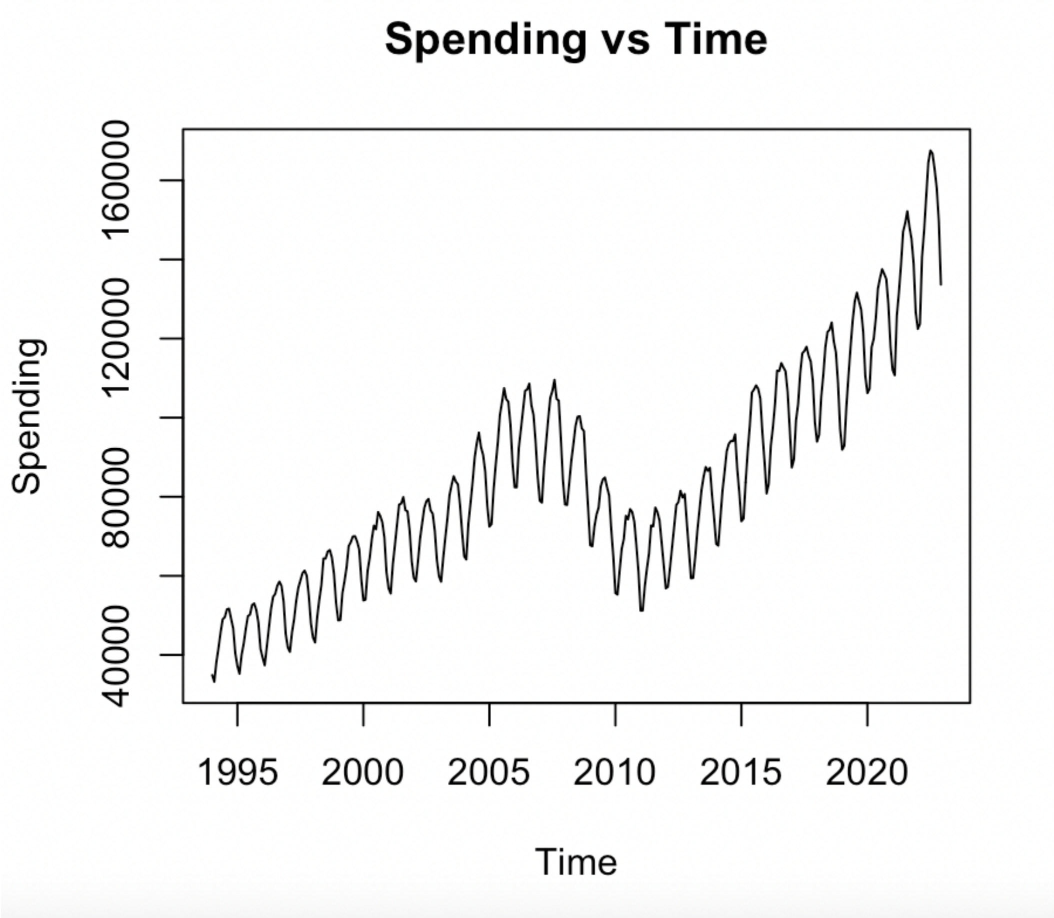
**OUTPUT:**

```
> tail(rconstruct)
          Date Month Time Spending logSpending dlogSpending Changepoint        c348
343 2022-07-01     7  343   167561    12.02910  0.018442973           1 -0.65658576
344 2022-08-01     8  344   166768    12.02436 -0.004743839           1 -0.23649900
345 2022-09-01     9  345   162810    12.00034 -0.024019748           1  0.92977649
346 2022-10-01    10  346   158075    11.97082 -0.029514273           1 -0.83752804
347 2022-11-01    11  347   149659    11.91611 -0.054710231           1  0.03769018
348 2022-12-01    12  348   133656    11.80302 -0.113090038           1  0.79399040
          s348       c432       s432 obs169 obs205 obs316 fMonth
343  0.7542514  0.4483832  0.8938414      0      0      0      7
344 -0.9716317 -0.7784623 -0.6276914      0      0      0      8
345  0.3681246  0.9685832  0.2486899      0      0      0      9
346  0.5463943 -0.9845643  0.1750231      0      0      0     10
347 -0.9992895  0.8235326 -0.5672689      0      0      0     11
348  0.6079303 -0.5144395  0.8575267      0      0      0     12
```

Q1. #Time series plot of Spending:

plot(ts(Spending,start=c(1994,1),freq=12),xlab="Time",ylab="Spending",main="Spending vs Time")

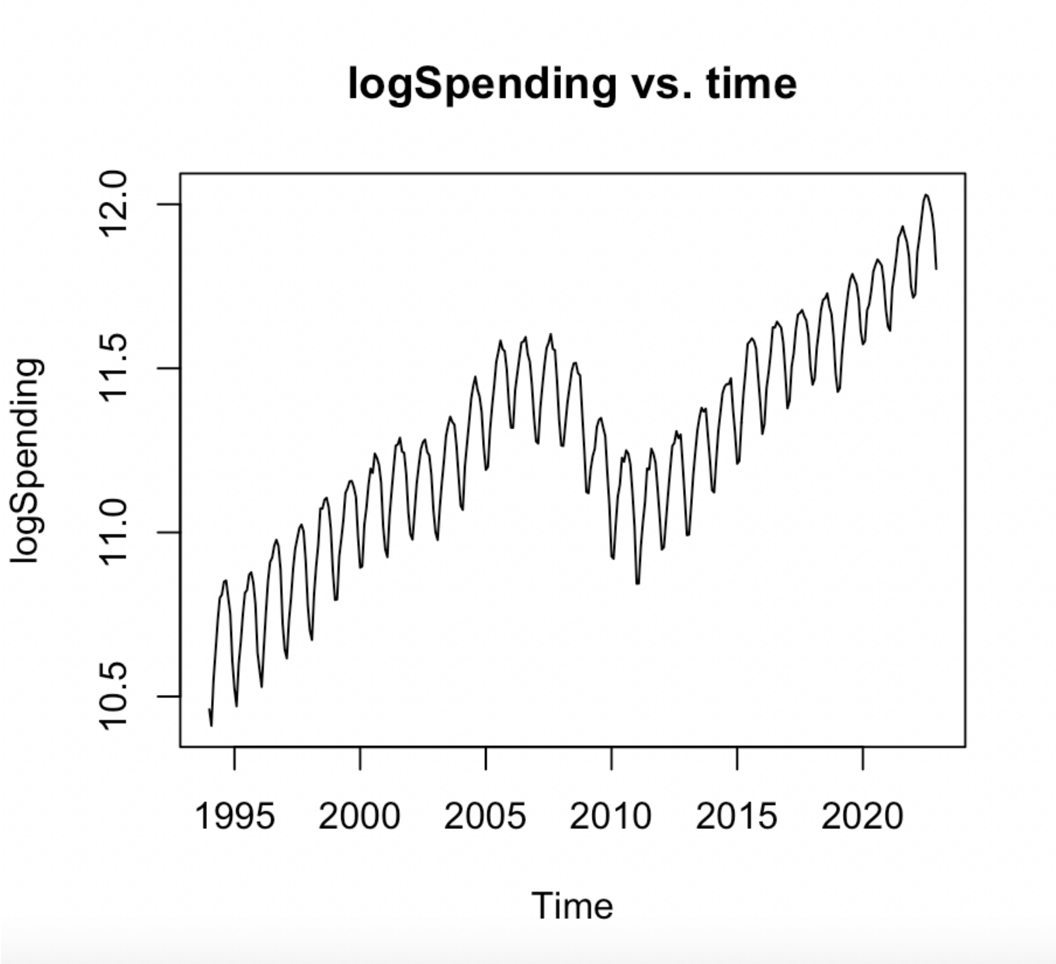
**OUTPUT:**



#Time series plot of logSpending:

```
plot(ts(logSpending,start=c(1994,1),freq=12),xlab="Time",ylab="logSpending",main="logSpending vs. time")
```

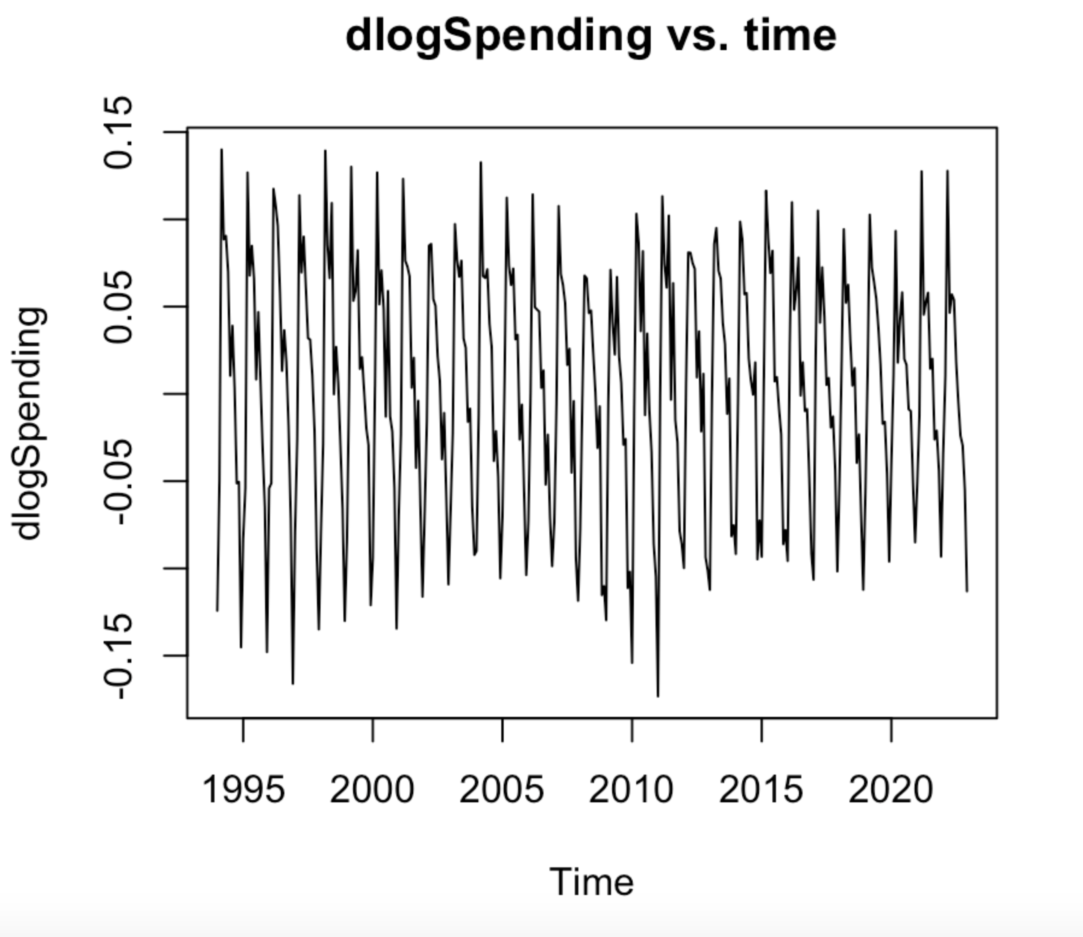
**OUTPUT:**




# Time series plot of log returns of Spending:

```
plot(ts(dlogSpending,start=c(1994,1),freq=12),xlab="Time",ylab="dlogSpending",main="dlogSpending vs. time")
```

**OUTPUT:**

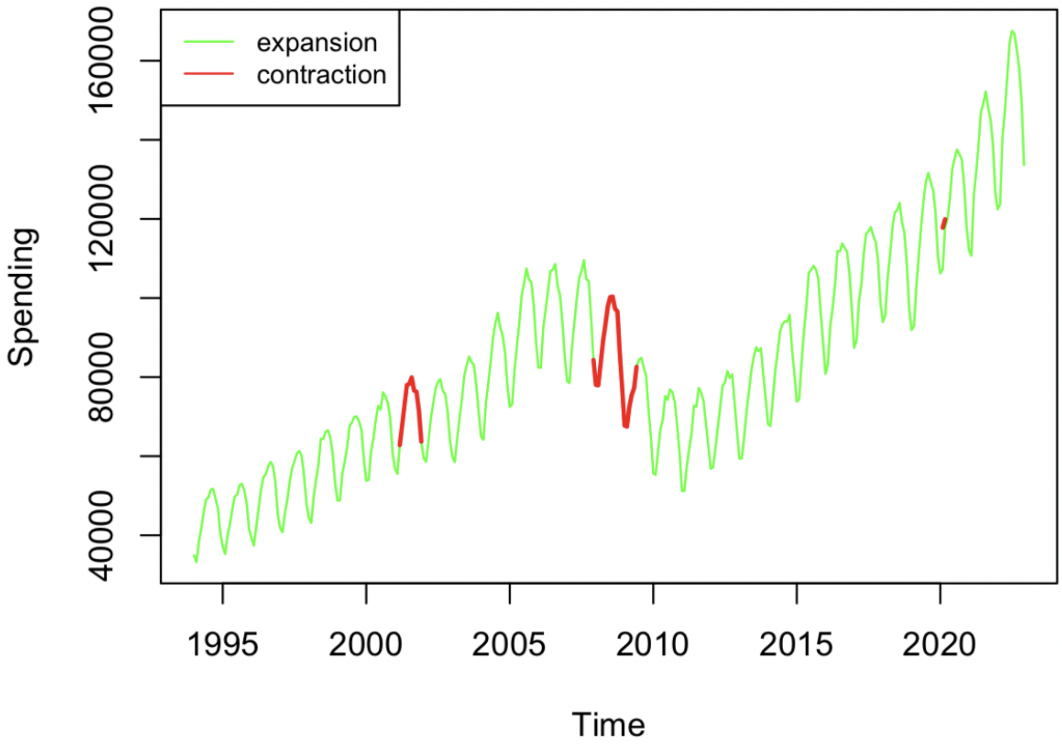



#List the periods of economic downturn in each of the plots:

#For Spending plot:

```
Spendingexpansion<-c(Spending[1:99],rep(NA,8),Spending[108:180],rep(NA,18),Spending[199:326],rep (NA,2),Spending[326:355])
Spendingcontraction<-c(rep(NA,99),Spending[100:107],rep(NA,73),Spending[181:198],rep(NA,128) ,Spending[327:328],rep(NA,27 ))
plot(ts(Spendingexpansion,start=c(1994,1),freq=12),ylab="Spending",main="Construction Spending ",col="green",lwd=2)
lines(ts(Spendingcontraction,start=c(1994,1),freq=12),col="red",lwd=2)
legend("topleft",legend=c("expansion","contraction"),col=c("green","red"),lty=1,cex=0.8 )
```

**OUTPUT:**



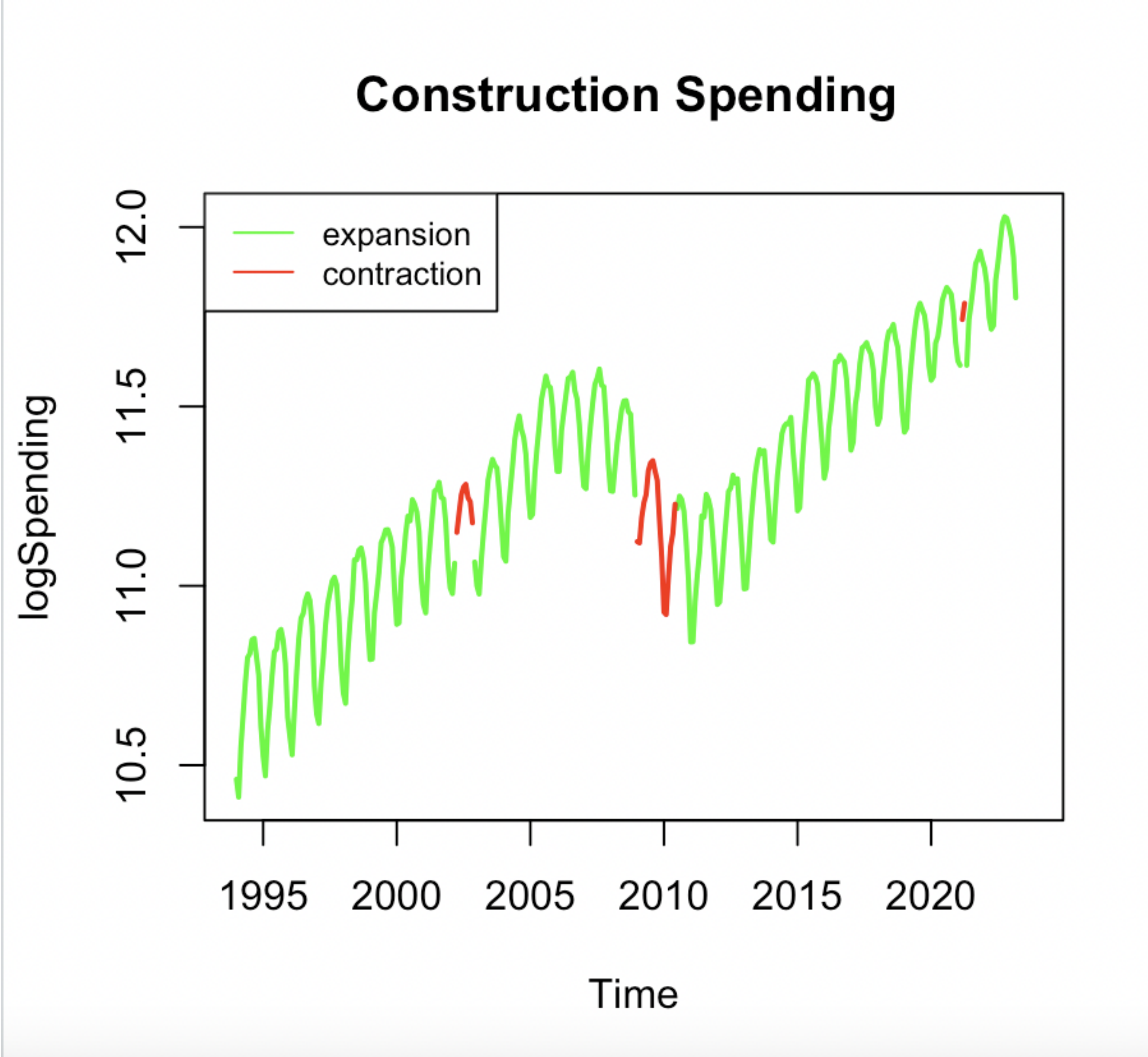

**OBSERVATION:** We observe that there is an increase in the value of trend from years 1995 to 2006 and then there is a sudden decrease and there is an increase again. We can say that due to better standard of living, many people started buying more and more houses and there was more development in the urban and rural areas due to which more money was spent on construction. We can also see that there are cases when there was economic downturn because of which (mostly in 2001 due to dot com bubble), 2008-2009 recession (the great recession) and in 2020 - Covid-19 outbreak. Volatility is high as the years increase. We can see that the spending is more volatile around the period 2020 due to covid outbreak.

Let us now talk about seasonality in the plot, we can see that seasonality decreases as the year begins and later on increases and then decreases. This implies that there was less amount of money spent in the beginning of the year and later on the money spent increases usually during the summer months and the fall months when construction was more active due to weather conditions possibly. Also, people might not have much money to spent on construction during the winter months as they go out on trips and thus, they spent money usually in the summer and the fall months.

```
#For logSpending plot:
logSpendingexpansion<-c(logSpending[1:99],rep(NA,8),logSpending[108:180],rep(NA,18),logSpending[199:326],rep (NA,2),logSpending[326:355])
logSpendingcontraction<-c(rep(NA,99),logSpending[100:107],rep(NA,73),logSpending[181:198],rep(NA,128) ,logSpending[327:328],rep(NA,27 ))
plot(ts(logSpendingexpansion,start=c(1994,1),freq=12),ylab="logSpending",main="Construction Spending ",col="green",lwd=2)
```

```
lines(ts(logSpendingcontraction,start=c(1994,1),freq=12),col="red",lwd=2)
legend("topleft",legend=c("expansion","contraction"),col=c("green","red"),lty=1,cex=0.8 )
```

**OUTPUT:**




**OBSERVATION:** The points of economic downturns have been marked on the plot respectively. We can see that the volatility is high in the initial years. This maybe due to overcorrection. The seasonal component is similar to the case of Spending. The trend is not uniform again. We can see that it increases till 2005, then is constant from 2005 to 2009. It then reduces till 2011 and then increases. We can see that the great recession of 2008-2009 has a profound impact on the trend. The construction log spending was constant for some years and then it decreased because of this. We can see that it increases later on.

```
#For log returns of Spending plot:
dlogSpendingexpansion<-c(dlogSpending[1:99],rep(NA,8),dlogSpending[108:180],rep(NA,18),dlogSpending[199:326],rep (NA,2),dlogSpending[326:355])
dlogSpendingcontraction<-c(rep(NA,99),dlogSpending[100:107],rep(NA,73),dlogSpending[181:198],rep(NA,128) ,dlogSpending[327:328],rep(NA,27 ))
```

```
plot(ts(dlogSpendingexpansion,start=c(1994,1),freq=12),ylab="dlogSpending",main="Constr
uction Spending ",col="green",lwd=2)
lines(ts(dlogSpendingcontraction,start=c(1994,1),freq=12),col="red",lwd=2)
legend("topleft",legend=c("expansion","contraction"),col=c("green","red"),lty=1,cex=0.8 )
```

**OUTPUT:**


**OBSERVATION:** We have marked the economic downturns in the log return plot as shown above. We can say that 2001 - 2003 has been marked due to dot cum bubble, 2008-2009 due to great recession and 2021 due to covid-19 outbreak. The trend has not been captured that well in the above plot. There is seasonality in the above plot. We can see that it is low at the beginning of the year and then gradually increases. The seasonal component then decreases. There is some volatility in the economic downturn periods, and we can say that this is mainly because of no log returns of spending of construction during dot com bubble, and the great recession.

**NOTE:** Also, as the rate of change of trend is almost proportional to the series level, we can say that a multiplicative model would be appropriate in this case.

b).

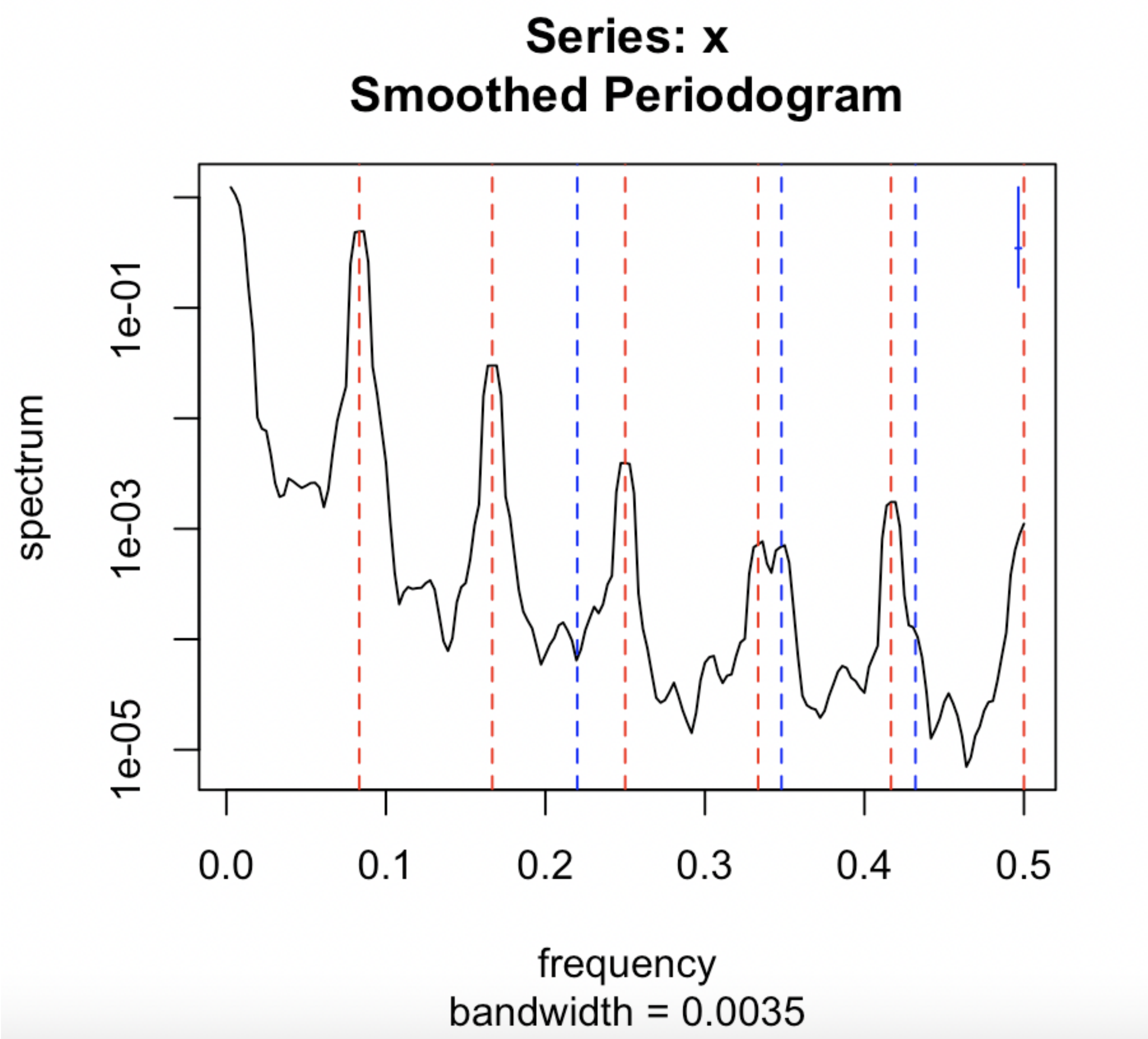
Q2. Spectral plots for

i). Log spending:

#Spectral plot for logSpending:

```
spectrum(logSpending,span=5)
abline(v=c(1/12,2/12,3/12,4/12,5/12,6/12),col="red",lty=2)
abline(v=c(0.220,0.348,0.432),col="blue",lty=2)
```

**OUTPUT:**



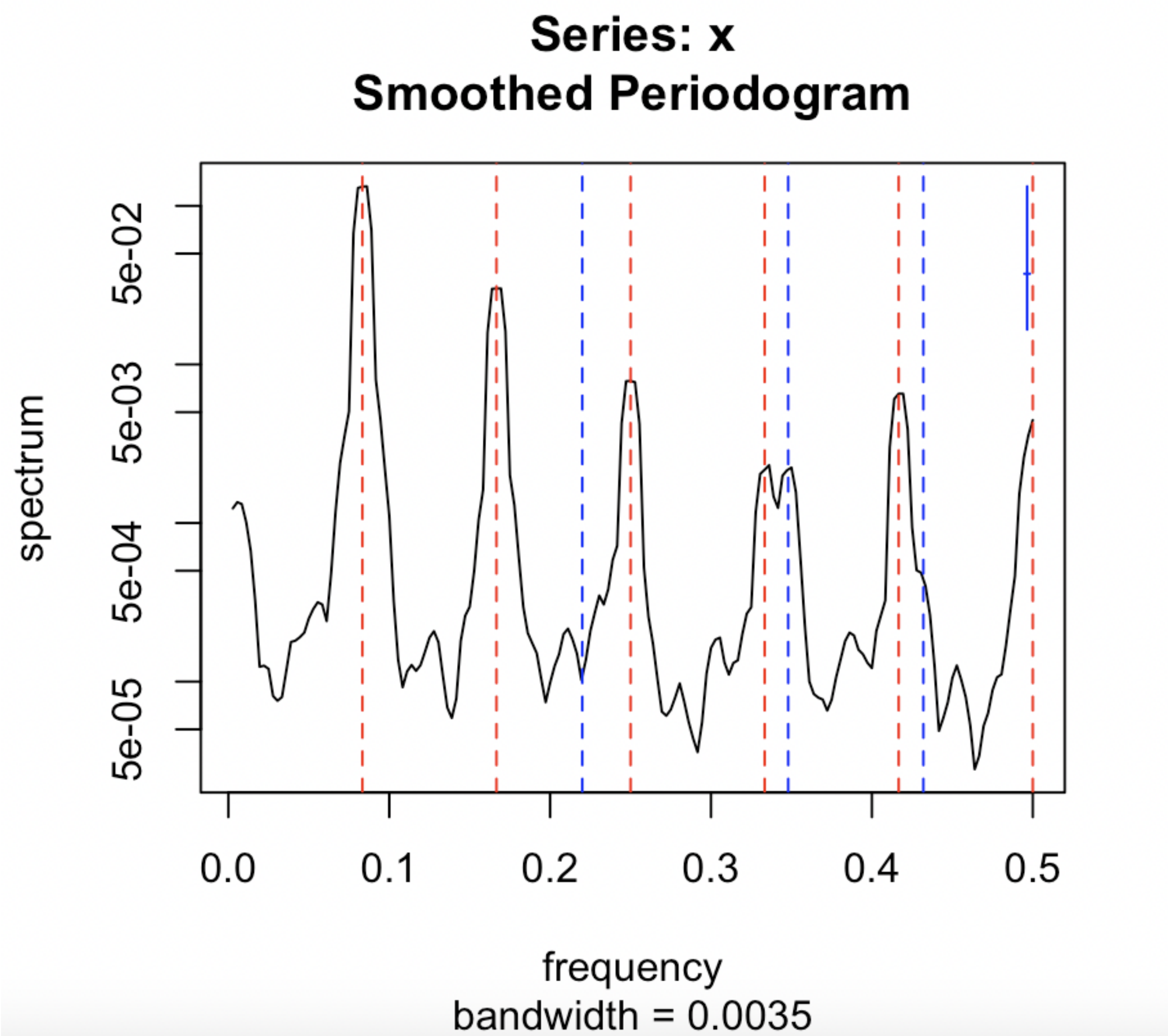
**OBSERVATION:** We can observe that there is a peak near the point of origin or 0. This indicates that there is some trend structure. Also, we can notice the dashed red lines at the seasonal frequencies of 1/12, 2/12, 3/12, 4/12, 5/12, and 6/12. There are higher peaks in the lower frequency range. Also, the blue lines indicate the calendar component. These are at the higher frequency range in the plot. We can clearly observe that the frequency 0.348 is prominently present as compared to frequencies 0.220 and 0.432. This plot has a strong seasonal component and there is little amount of calendar structure in this.

ii). Log return spending series:

#Spectral plot for dlogSpending:

```
spectrum(dlogSpending,span=5)
abline(v=c(1/12,2/12,3/12,4/12,5/12,6/12),col="red",lty=2)
abline(v=c(0.220,0.348,0.432),col="blue",lty=2)
```

**OUTPUT:**



**OBSERVATION:** We can notice that there is a trend component in the above plot as there is a peak at low frequency. The peaks have become extra prominent here. The red dashed lines are at the seasonal frequencies of 1/12, 2/12, 3/12, 4/12, 5/12, and 6/12. The blue dashed lines are at the calendar frequencies of 0.220, 0.348, and 0.432. We can again observe that the frequency of 0.348 is more prominent as compared to the frequencies of 0.220 and 0.432. The plot indicates the presence of strong seasonal component and almost very little calendar structure.

Q3. Fitting multiplicative model to the spending model:

```
Time2<-Time*Time
Time3<-Time2*Time
Time4<-Time3*Time
Time5<-Time4*Time

Changepoint2<-Changepoint*Time
Changepoint3<-Changepoint*Time2
Changepoint4<-Changepoint*Time3
Changepoint5<-Changepoint*Time4
Changepoint6<-Changepoint*Time5

model1<-lm(log(Spending)~ poly(Time,5) + Changepoint+ Changepoint2 + Changepoint3 + Changepoint4+ Changepoint5 + Changepoint6 + fMonth +c348 + s348 +c432 +s432+ obs169+ obs205 + obs316, data = rconstruct);summary(model1)
```

**OUTPUT:**

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)     -9.124e+00  2.142e+00  -4.261 2.69e-05 ***
poly(Time, 5)1 -5.178e+02  5.514e+01  -9.392  < 2e-16 ***
poly(Time, 5)2 -4.246e+02  4.447e+01  -9.548  < 2e-16 ***
poly(Time, 5)3 -2.352e+02  2.475e+01  -9.502  < 2e-16 ***
poly(Time, 5)4 -8.573e+01  9.089e+00  -9.432  < 2e-16 ***
poly(Time, 5)5 -1.626e+01  1.758e+00  -9.249  < 2e-16 ***
Changepoint      5.406e+01  1.126e+01   4.802 2.42e-06 ***
Changepoint2    -8.348e-01  2.285e-01  -3.653 0.000303 ***
Changepoint3     4.383e-03  1.832e-03   2.393 0.017285 *  
Changepoint4    -4.281e-06  7.303e-06  -0.586 0.558168    
Changepoint5    -4.491e-08  1.554e-08  -2.889 0.004128 ** 
Changepoint6     1.397e-10  1.884e-11   7.413 1.13e-12 ***
fMonth2         -1.572e-02  8.539e-03  -1.841 0.066525 .  
fMonth3          9.057e-02  8.533e-03  10.613  < 2e-16 ***
fMonth4          1.556e-01  8.615e-03  18.058  < 2e-16 ***
fMonth5          2.159e-01  8.536e-03  25.298  < 2e-16 ***
fMonth6          2.776e-01  8.537e-03  32.515  < 2e-16 ***
fMonth7          2.891e-01  8.542e-03  33.840  < 2e-16 ***
fMonth8          3.093e-01  8.543e-03  36.203  < 2e-16 ***
fMonth9          2.888e-01  8.552e-03  33.766  < 2e-16 ***
fMonth10         2.691e-01  8.553e-03  31.456  < 2e-16 ***
fMonth11         1.967e-01  8.565e-03  22.971  < 2e-16 ***
fMonth12         8.471e-02  8.571e-03   9.884  < 2e-16 ***
c348             3.992e-03  2.431e-03   1.642 0.101531    
s348            -2.330e-03  2.427e-03  -0.960 0.337736    
c432            -3.487e-04  2.442e-03  -0.143 0.886517    
s432            -1.810e-03  2.422e-03  -0.747 0.455527    
obs169          -1.087e-01  3.603e-02  -3.016 0.002764 ** 
obs205          -7.985e-02  3.304e-02  -2.417 0.016208 *  
obs316          -1.276e-03  3.309e-02  -0.039 0.969265    
---
```

**OBSERVATION:** It is to be noticed that the trigonometric pairs (c348,s348) and (c432,s432) are not significant above. Also, the outlier obs316 is not significant. The variable Changepoint4 is insignificant in nature. Let us now conduct a partial F-test on the model to see if the variable Changepoint4 is significant or not. We notice that the value of RSS is higher in the case when there is no term of Changepoint4. So, we keep the term as it is significant in nature. Also, the trigonometric pairs would be discarded as they are insignificant with p-values less than 0.05 each.

Let us refit the model without the above terms.

So,

#Refitting the model by removing insignificant terms:

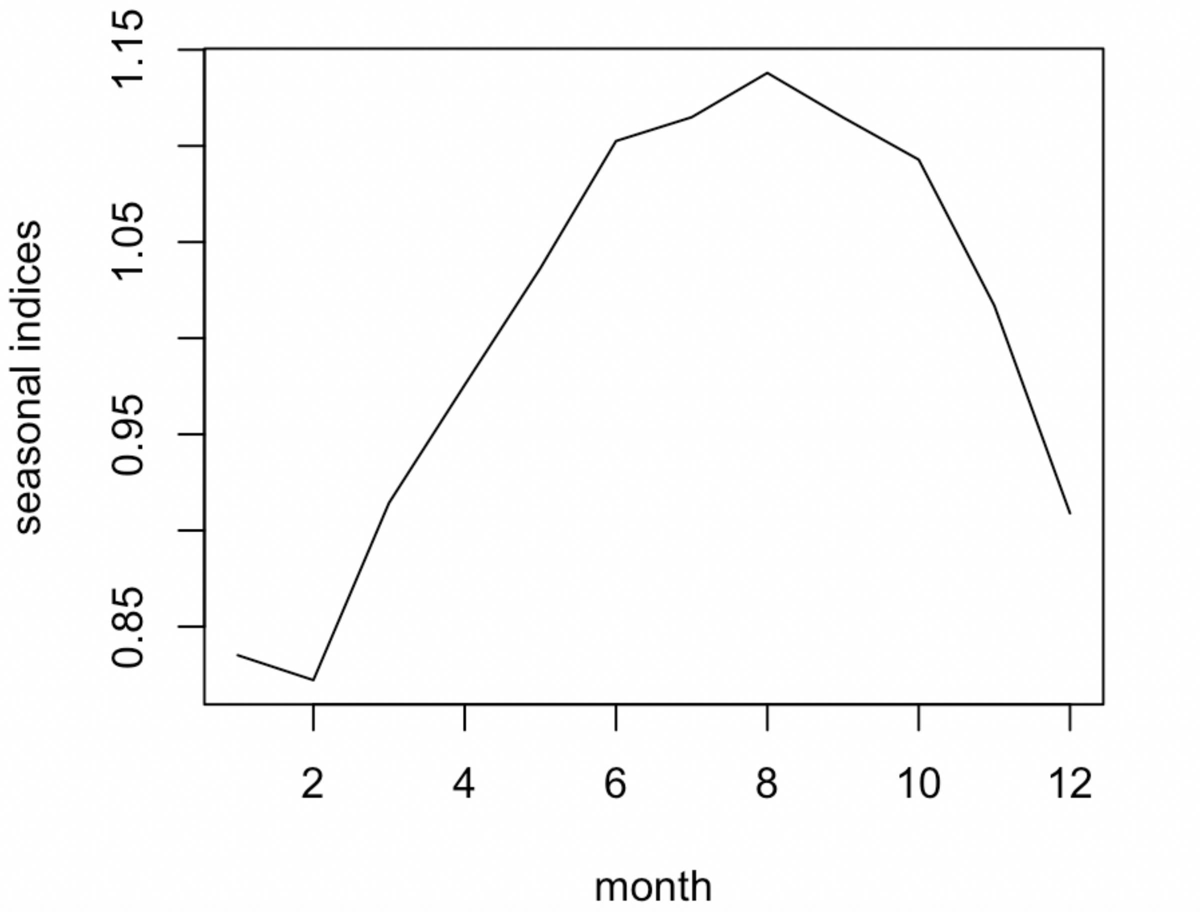
model2<-lm(log(Spending)~fMonth+ poly(Time, 5) + Changepoint+ Changepoint2 + Changepoint3 + Changepoint4 + Changepoint5 + Changepoint6 +obs169+ obs205, data = rconstruct);summary(model2)

**OUTPUT:**

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)    -9.131e+00  2.139e+00  -4.269 2.59e-05 ***
fMonth2        -1.564e-02  8.523e-03  -1.835 0.067392 .
fMonth3         9.065e-02  8.521e-03  10.638  < 2e-16 ***
fMonth4         1.556e-01  8.521e-03  18.257  < 2e-16 ***
fMonth5         2.159e-01  8.522e-03  25.337  < 2e-16 ***
fMonth6         2.778e-01  8.524e-03  32.587  < 2e-16 ***
fMonth7         2.889e-01  8.527e-03  33.886  < 2e-16 ***
fMonth8         3.094e-01  8.531e-03  36.267  < 2e-16 ***
fMonth9         2.888e-01  8.536e-03  33.838  < 2e-16 ***
fMonth10        2.690e-01  8.542e-03  31.486  < 2e-16 ***
fMonth11        1.969e-01  8.550e-03  23.034  < 2e-16 ***
fMonth12        8.468e-02  8.559e-03   9.894  < 2e-16 ***
poly(Time, 5)1 -5.180e+02  5.507e+01  -9.406  < 2e-16 ***
poly(Time, 5)2 -4.247e+02  4.441e+01  -9.562  < 2e-16 ***
poly(Time, 5)3 -2.352e+02  2.471e+01  -9.515  < 2e-16 ***
poly(Time, 5)4 -8.572e+01  9.078e+00  -9.443  < 2e-16 ***
poly(Time, 5)5 -1.626e+01  1.756e+00  -9.257  < 2e-16 ***
Changepoint     5.379e+01  1.123e+01   4.789 2.56e-06 ***
Changepoint2   -8.293e-01  2.279e-01  -3.639 0.000319 ***
Changepoint3    4.341e-03  1.827e-03   2.376 0.018069 *
Changepoint4   -4.123e-06  7.283e-06  -0.566 0.571678
Changepoint5   -4.519e-08  1.550e-08  -2.915 0.003805 **
Changepoint6    1.399e-10  1.881e-11   7.438 9.31e-13 ***
obs169         -1.050e-01  3.575e-02  -2.938 0.003545 **
obs205         -8.303e-02  3.281e-02  -2.531 0.011864 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

A). Tabulate and plot the estimated seasonal indices:

```
b1<-coef(model2)[1]
b2<-coef(model2)[2:12]+b1
b3<-c(b1,b2)
newintrcpt<-b1-(b3-mean(b3))[1]
newintrcpt
seas<-exp(b3-mean(b3))
seas
seas.ts<-ts(seas)
```



```
> seas
(Intercept)     fMonth2     fMonth3     fMonth4     fMonth5     fMonth6     fMonth7 
  0.8351272   0.8221664   0.9143693   0.9757027   1.0364023   1.1025322   1.1149156 
    fMonth8     fMonth9    fMonth10    fMonth11    fMonth12 
  1.1379529   1.1147980   1.0928500   1.0169026   0.9089218 
```

plot(seas.ts,ylab="seasonal indices",xlab="month")

**OBSERVATION:** We notice that the peak spending is in the month of August. Also, there is an increase in spending from February to August. There is a decrease in the spending from August onwards. We can also observe that the spending is above the trend level in the months of May, June, October, and November. The sales are lowest in the month of February. So, the spending of construction is highest in the fall months and lowest in the winter months (like December, January, and February). We can also notice that spending gradually increases in the summer months.

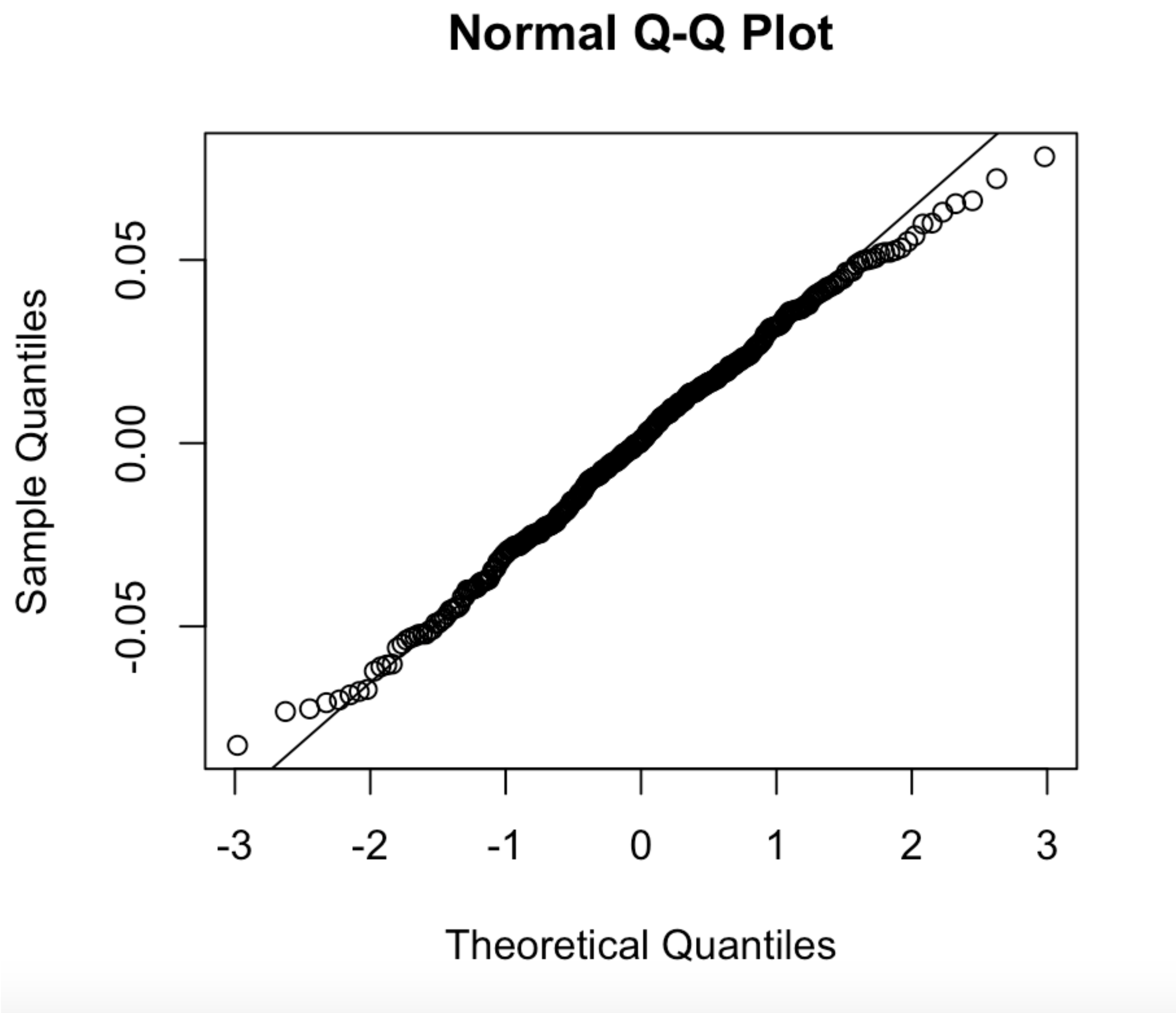
**CONCLUSION:** We can say that the money spent on construction is less in the winter months as compared to the fall and the summer months. This maybe due to the favourable weather conditions, where people were more comfortable to build in these months. Also, during the winter months, it is possible that people spent their money on winter vacation because of which they were not able to spend money on building stuff. Thus there is a decrease in the spending.

B).

#Residual analysis:

#Normal quantile plot:
qqnorm(resid(model2))
qqline(resid(model2))

**OUTPUT:**




**OBSERVATION:** It can be seen that the model is not able to fit all point on the same line above. There are outliers at both ends of the tail. Also, we would be conducting the Shapiro-Wilk test.

#Test for normality:

shapiro.test(resid(model2))

**OUTPUT:**

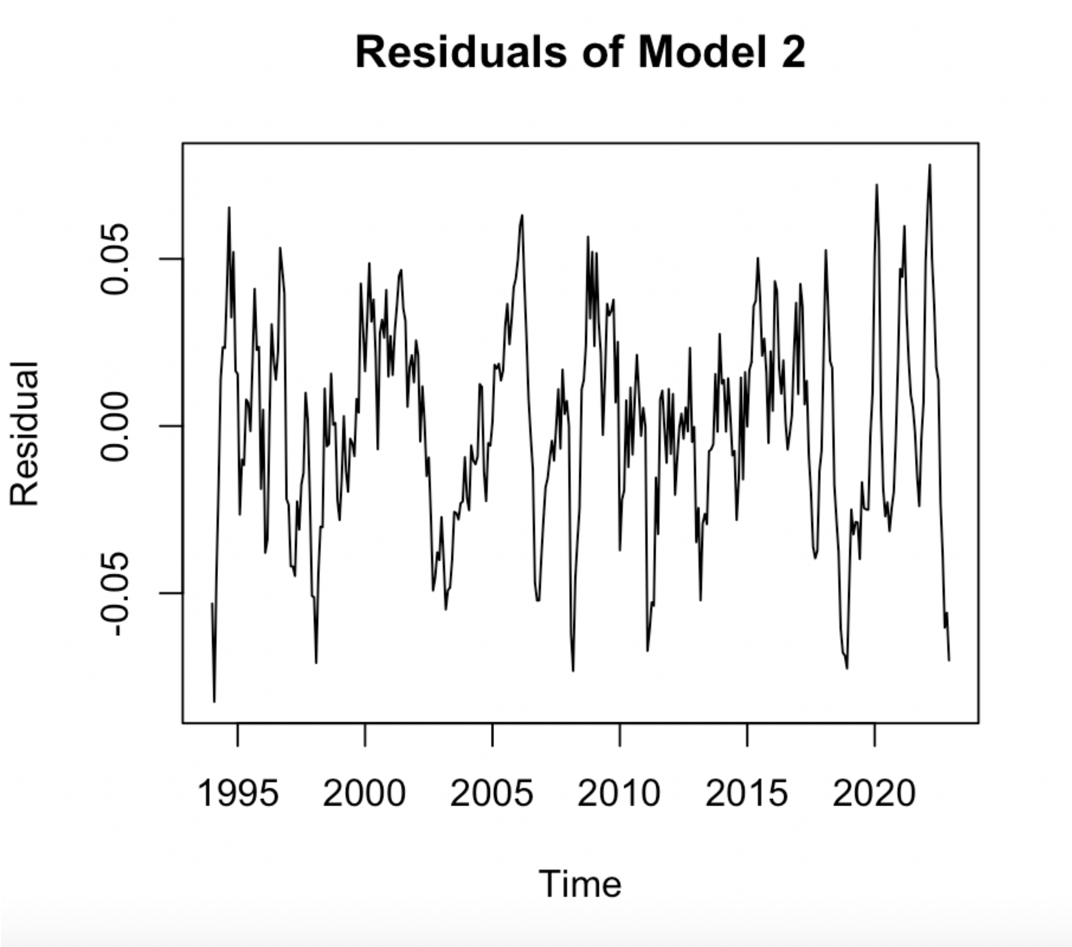
```
> shapiro.test(resid(model2))

        Shapiro-Wilk normality test

data:  resid(model2)
W = 0.99483, p-value = 0.2949
```

**OBSERVATION:** As the value of p is greater than 0.05, we can notice that the data is distributed normally.

#Plot residuals versus time:

plot(ts(resid(model2),start=c(1994,1),freq=12),xlab="Time",ylab="Residual",main="Residuals of Model 2")

**OUTPUT:**

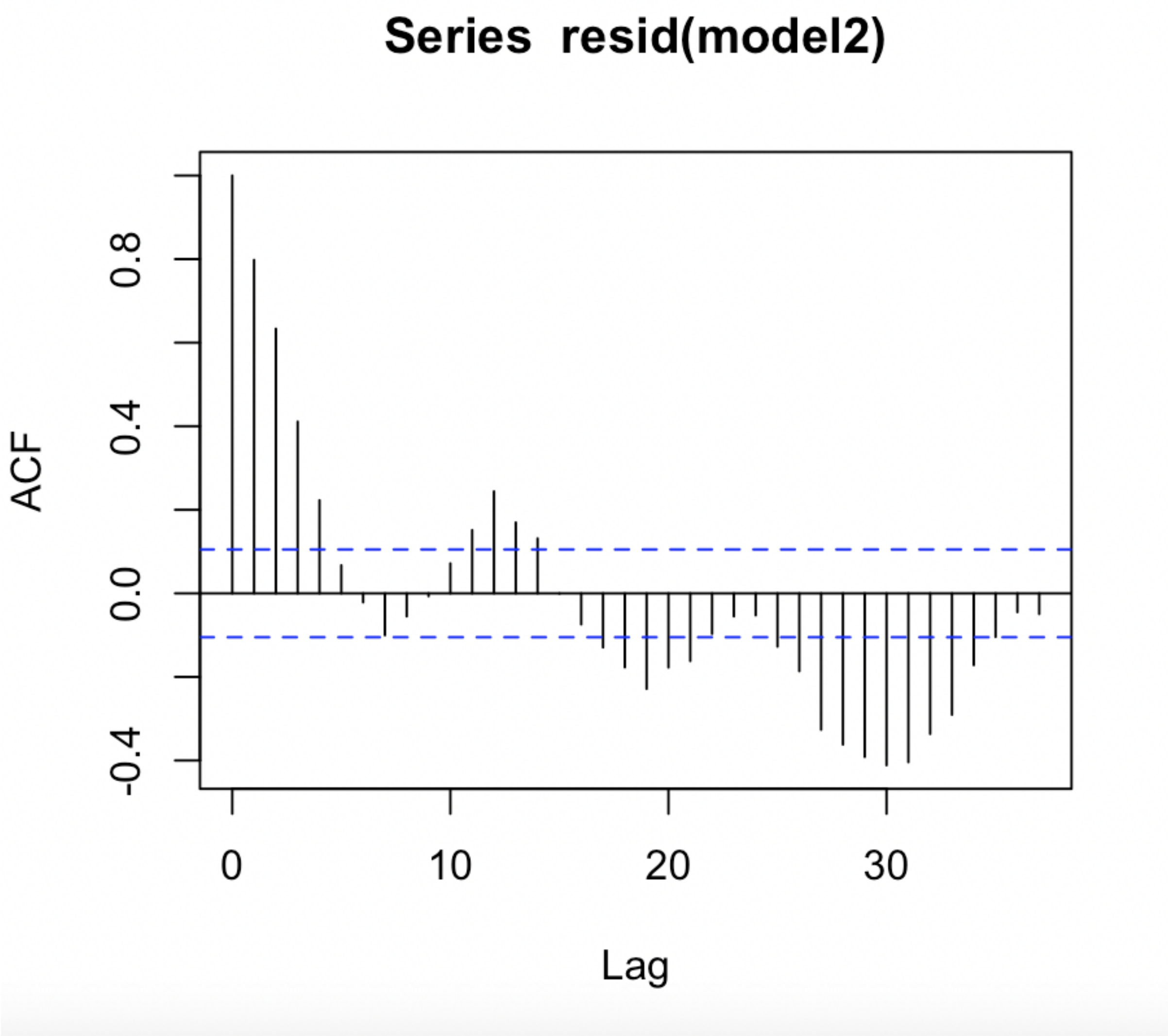



**OBSERVATION:** The above model has failed to capture the trend component. This can be majorly due to the dot com bubble burst, and the 2008-2009 recession period. These had a great impact years later as well. We can also see that the years from 2013 to 2017 have a dissimilar trend.

#Estimated auto-correlations:

acf(resid(model2),37)

**OUTPUT:**

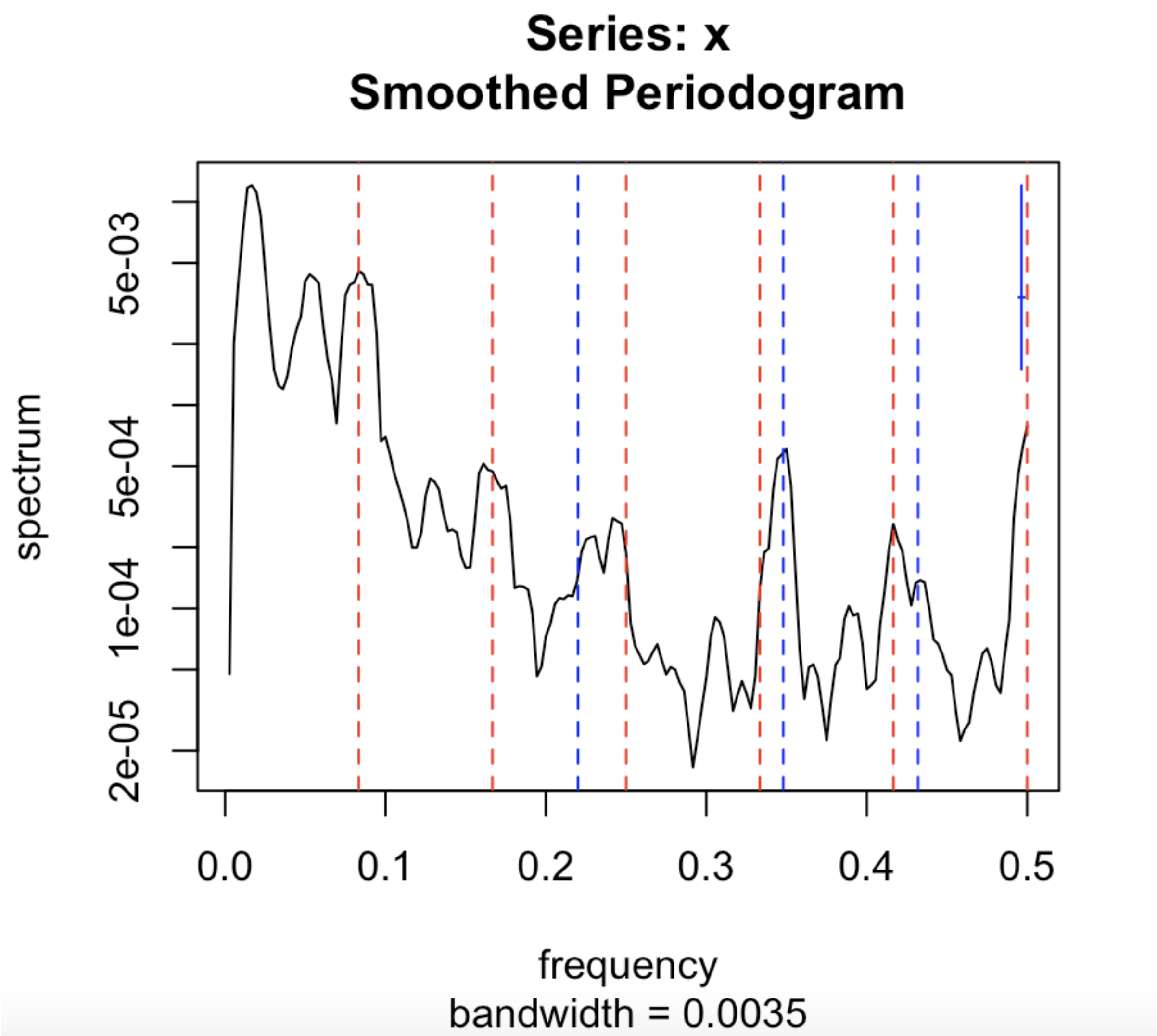



**OBSERVATION:** It can be clearly seen that the trend has been failed to captured by the above ACF plot. It has not been reduced to white noise. The blue dashed lines indicate the 95% confidence interval. There are spikes above the 95% confidence interval as well. This clearly says that there is some uncaptured trend and some parts of dynamic seasonal structure. We can also notice that lags 12 and 19 have considerable amounts of correlation.

iii). #Spectral densities for Model 2:

```
spectrum(resid(model2),span=5)
abline(v=c(1/12,2/12,3/12,4/12,5/12,6/12),col="red",lty=2)
abline(v=c(0.220,0.348,0.432),col="blue",lty=2)
```

**OUTPUT:**




**OBSERVATION:** There is a peak at the lower frequency and this signals that there is trend structure remaining. Also, we can observe peaks at seasonal frequencies of 1/12 and 5/12. Thus there is some seasonal component captured. Also, the calendar frequencies are situated at the higher frequencies range. The frequency of 0.348 is the most prominent. We can say that there is not much calendar component above in the spectral plot. Also, the trend component is seen to be less prominent here.

**CONCLUSION:** We can conclude that much trend has not been captured in the model. The reason is that there are outliers at the end of tails in the normal quantile plot. The trend has also been uncaptured in the plot of residual versus time. Also, we can observe in the spectral plot that there is some uncaptured trend close to lower frequencies (that is 0).

Q4. Form lag1, lag2, and lag3 residuals for model in part 3:

```
resid1 <- resid(model2)
lag1resid <- c(NA, resid1[1:(length(resid1) -1)])
lag2resid <- c(NA, NA, resid1[1:(length(resid1) -2)])
lag3resid <- c(NA, NA, NA, resid1[1:(length(resid1) -3)])
```

rconstruct <- data.frame(rconstruct, lag1resid, lag2resid, lag3resid)

model4<-lm(logSpending~ poly(Time,5) + Changepoint+ Changepoint2 + Changepoint3 + Changepoint4+ Changepoint5 + Changepoint6 + lag1resid + lag2resid + lag3resid + fMonth +c348 + s348 +c432 +s432+ obs169+ obs205 + obs316, data = rconstruct);summary(model4)

**OUTPUT:**

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)     -1.048e+01  1.183e+00  -8.853  < 2e-16 ***
poly(Time, 5)1 -5.529e+02  3.051e+01 -18.125  < 2e-16 ***
poly(Time, 5)2 -4.534e+02  2.466e+01 -18.386  < 2e-16 ***
poly(Time, 5)3 -2.517e+02  1.378e+01 -18.261  < 2e-16 ***
poly(Time, 5)4 -9.206e+01  5.095e+00 -18.070  < 2e-16 ***
poly(Time, 5)5 -1.762e+01  1.003e+00 -17.564  < 2e-16 ***
Changepoint      5.703e+01  5.823e+00   9.793  < 2e-16 ***
Changepoint2    -8.955e-01  1.182e-01  -7.574 4.14e-13 ***
Changepoint3     4.836e-03  9.483e-04   5.099 5.93e-07 ***
Changepoint4    -5.492e-06  3.793e-06  -1.448 0.148693    
Changepoint5    -4.607e-08  8.235e-09  -5.594 4.86e-08 ***
Changepoint6     1.481e-10  1.044e-11  14.187  < 2e-16 ***
lag1resid        8.406e-01  5.289e-02  15.892  < 2e-16 ***
lag2resid        2.436e-01  6.912e-02   3.524 0.000489 ***
lag3resid       -3.160e-01  5.247e-02  -6.022 4.84e-09 ***
fMonth2         -1.402e-02  4.475e-03  -3.133 0.001894 ** 
fMonth3          9.031e-02  4.474e-03  20.184  < 2e-16 ***
fMonth4          1.573e-01  4.485e-03  35.067  < 2e-16 ***
fMonth5          2.158e-01  4.440e-03  48.603  < 2e-16 ***
fMonth6          2.774e-01  4.438e-03  62.514  < 2e-16 ***
fMonth7          2.892e-01  4.439e-03  65.166  < 2e-16 ***
fMonth8          3.093e-01  4.437e-03  69.699  < 2e-16 ***
fMonth9          2.890e-01  4.440e-03  65.090  < 2e-16 ***
fMonth10         2.695e-01  4.441e-03  60.681  < 2e-16 ***
fMonth11         1.970e-01  4.444e-03  44.335  < 2e-16 ***
fMonth12         8.527e-02  4.447e-03  19.175  < 2e-16 ***
c348             6.942e-03  1.327e-03   5.233 3.07e-07 ***
s348            -5.988e-03  1.318e-03  -4.544 7.89e-06 ***
c432            -7.462e-04  1.278e-03  -0.584 0.559770    
s432            -2.776e-03  1.257e-03  -2.208 0.027969 *  
obs169          -1.159e-01  1.857e-02  -6.243 1.40e-09 ***
obs205          -7.626e-02  1.705e-02  -4.471 1.09e-05 ***
obs316          -5.688e-02  1.723e-02  -3.301 0.001077 ** 
---
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01641 on 312 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.9977,    Adjusted R-squared:  0.9975
F-statistic:  4261 on 32 and 312 DF,  p-value: < 2.2e-16
```

OBSERVATION: We notice that all the lag residuals are significant. The trigonometric pair (c432, s432) is insignificant. Let us now conduct partial F-test to decide if the pair will remain or not.
So,

model5<-lm(logSpending~ poly(Time,5) + Changepoint+ Changepoint2 + Changepoint3 + Changepoint4+ Changepoint5 + Changepoint6 + lag1resid + lag2resid + lag3resid + fMonth +c348 + s348 + obs169+ obs205 + obs316, data = rconstruct);summary(model5)

anova(model4, model5)

**OUTPUT:**

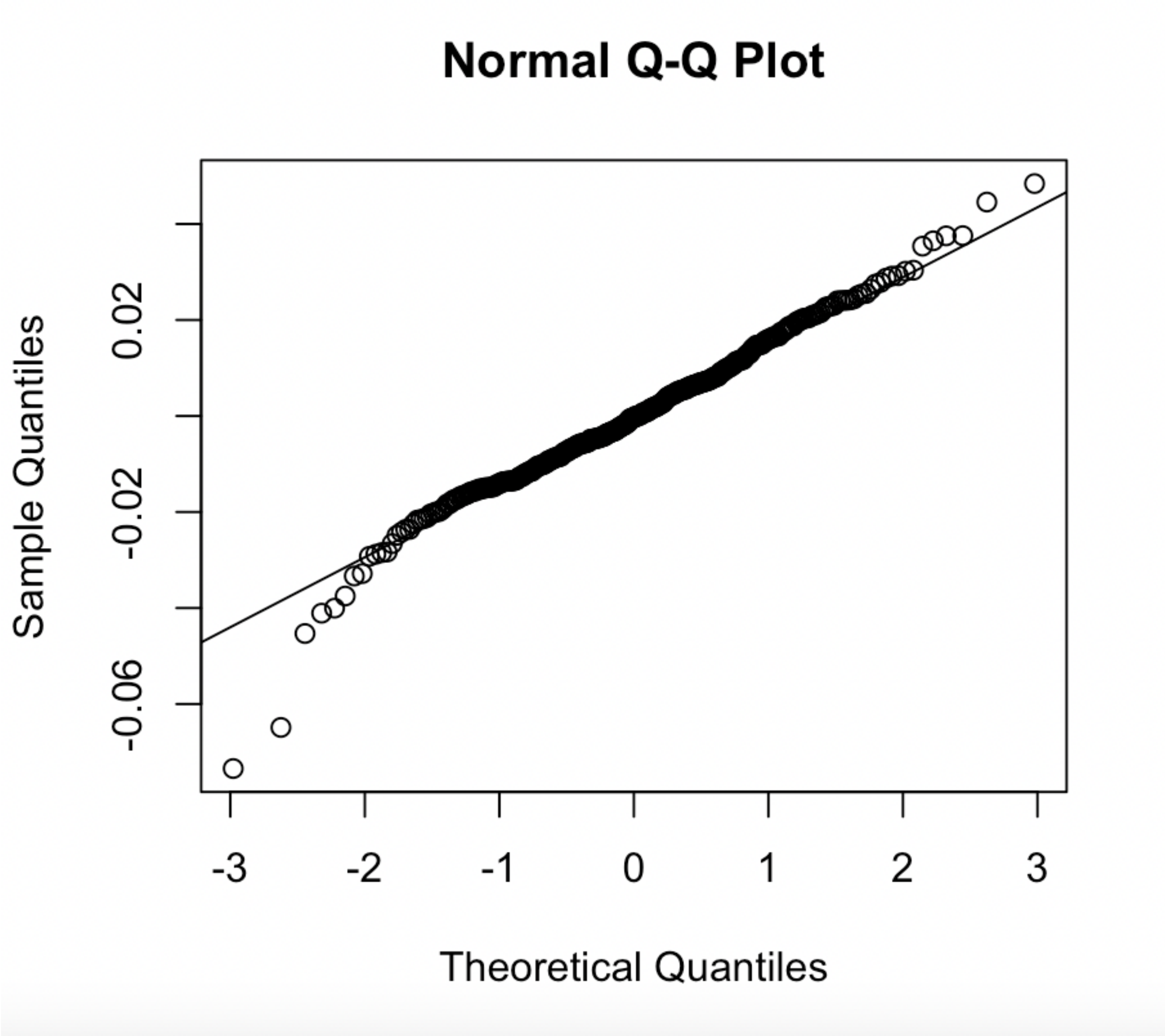
```
> anova(model4, model5)
Analysis of Variance Table

Model 1: logSpending ~ poly(Time, 5) + Changepoint + Changepoint2 + Changepoint3 +
    Changepoint4 + Changepoint5 + Changepoint6 + lag1resid +
    lag2resid + lag3resid + fMonth + c348 + s348 + c432 + s432 +
    obs169 + obs205 + obs316
Model 2: logSpending ~ poly(Time, 5) + Changepoint + Changepoint2 + Changepoint3 +
    Changepoint4 + Changepoint5 + Changepoint6 + lag1resid +
    lag2resid + lag3resid + fMonth + c348 + s348 + obs169 + obs205 +
    obs316
  Res.Df      RSS Df Sum of Sq      F Pr(>F)
1    312 0.083985
2    314 0.085381 -2 -0.001396 2.5931 0.0764 .
---
```

**OBSERVATION:** We see that the p-value in the above case is greater than 0.05. So, we would be preferring model 5 over model 4.
Thus, we would discard the trigonometric pair (c432, s432).

**Residual Analysis of Model 5:**

#Normal quantile plot:
qqnorm(resid(model5))
qqline(resid(model5))

**OUTPUT:**




**OBSERVATION:**

#Test for normality:

shapiro.test(resid(model5))

**OUTPUT:**

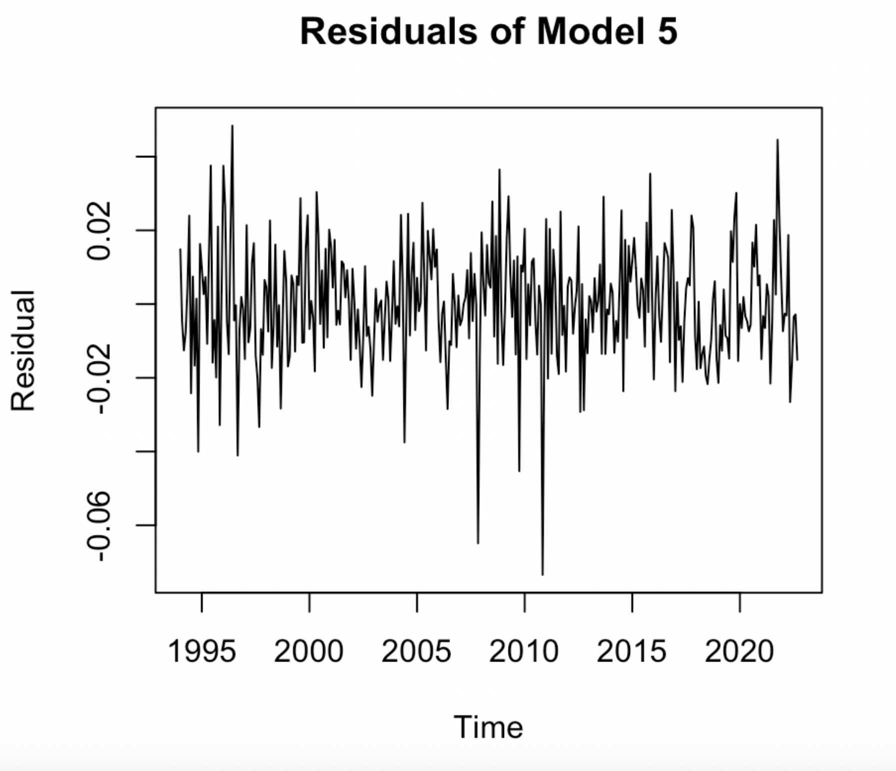
```
        Shapiro-Wilk normality test

data:  resid(model5)
W = 0.97938, p-value = 7.503e-05
```

**OBSERVATION:** We see that the p-value is less than 0.05, thus the data is not normally distributed. Also, this p-value clearly says that the outliers are insignificant (these are placed at the ends of the tails of the Normality quantile plot).

#Plot residuals versus time:

plot(ts(resid(model5),start=c(1994,1),freq=12),xlab="Time",ylab="Residual",main="Residuals of Model 5")

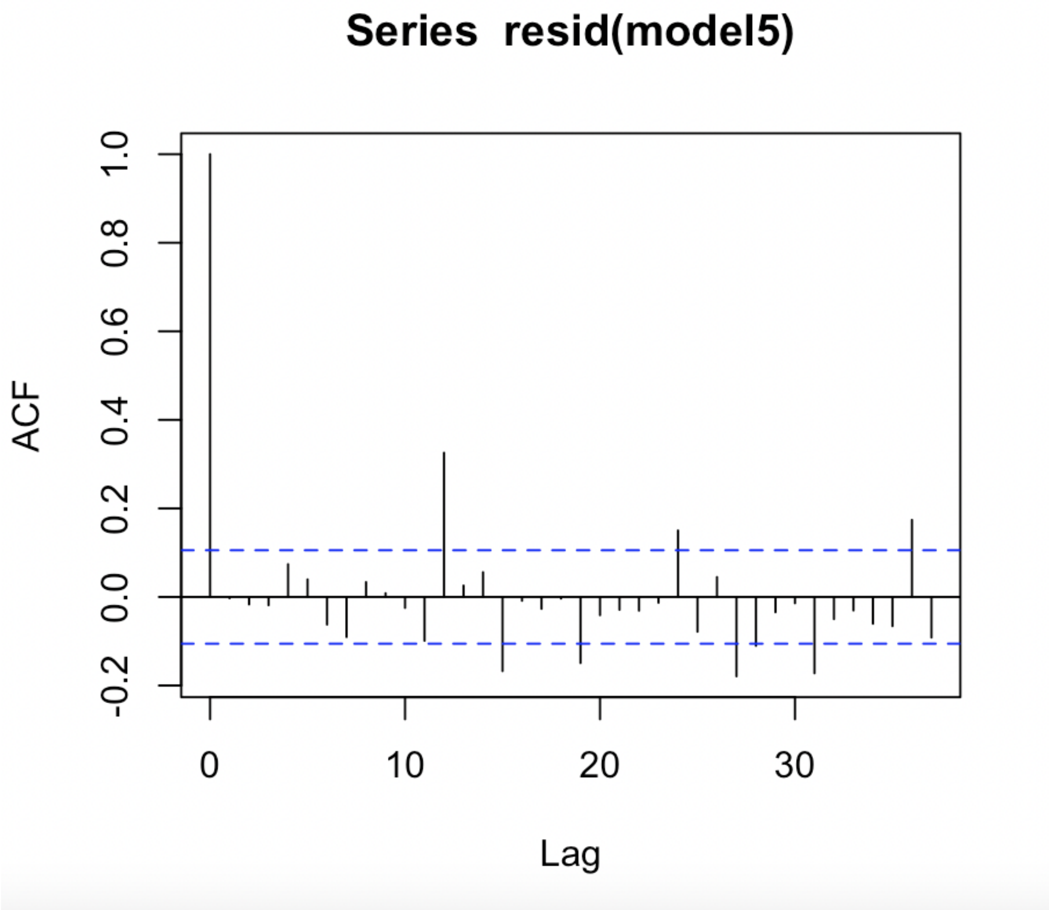
**OUTPUT:**




**OBSERVATION:** The trend has been captured well in the above plot by our model. The trend following the years of recession of 2008-2009 has also been captured well by this model of ours.
#Estimated auto correlations:

acf(resid(model5),37)

**OUTPUT:**

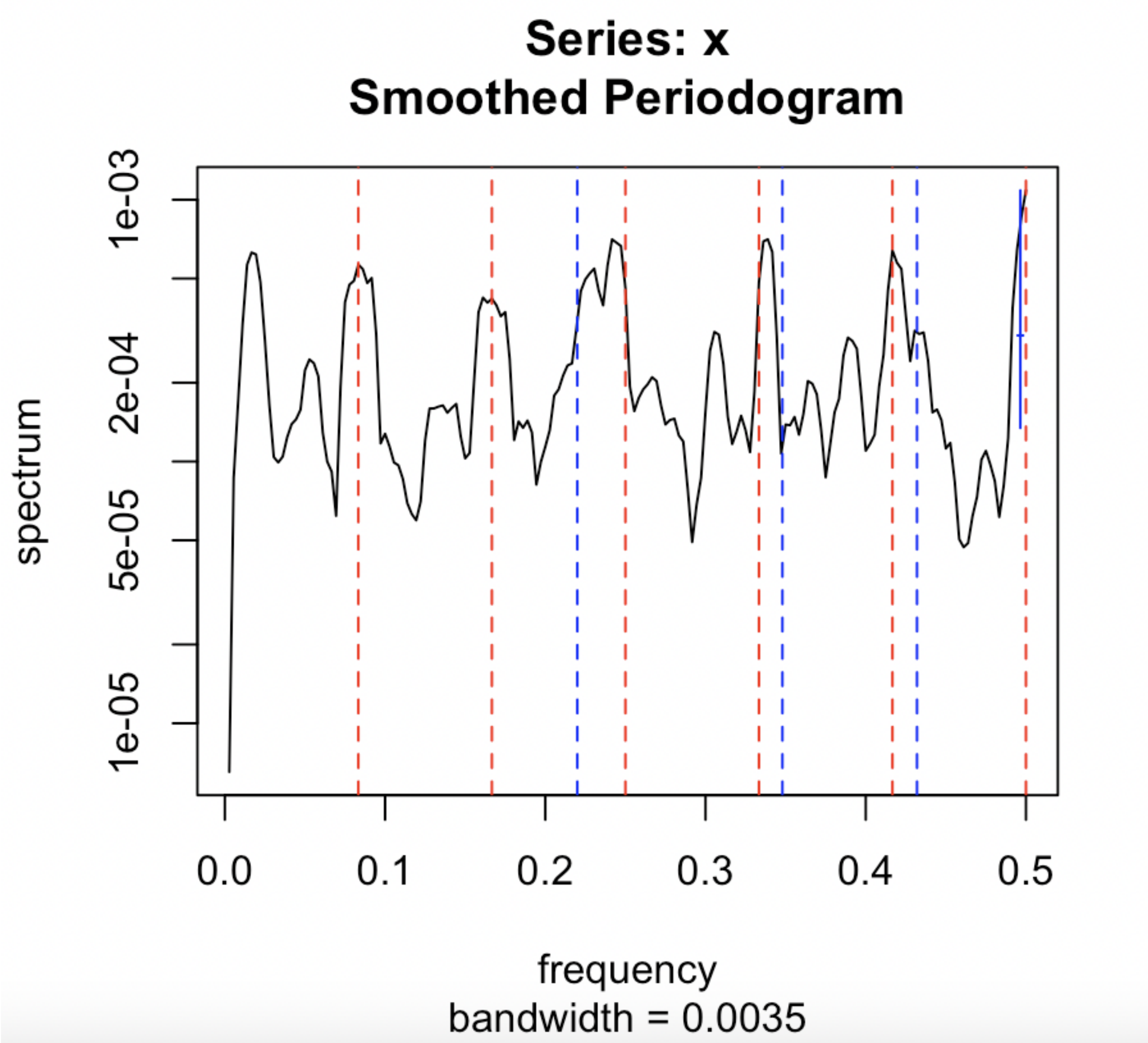



**OBSERVATION:** We observe that there are only a few significant structures left in the above ACF plot. The plot has almost come close to white noise with only a few spikes above the 95% confidence interval (blue-dashed lines).

#Spectral densities for Model 5:

```
spectrum(resid(model5),span=5)
abline(v=c(1/12,2/12,3/12,4/12,5/12,6/12),col="red",lty=2)
abline(v=c(0.220,0.348,0.432),col="blue",lty=2)
```

**OUTPUT:**




**OBSERVATION:** We can observe that there are peaks at low frequency and so we can conclude that there is some amount of trend structure present. The peaks at seasonal frequencies of 1/12, 2/12 and 5/12 indicate that dynamic seasonal component is present. Finally, we can say that there is no calendar component present in the plot above.

**CONCLUSION:** We can conclude by the residual analysis that there is a considerable amount of improvement in trend capture in comparison to model in question 3. This can be proven by the residual plot and the ACF. Also, there is some amount of trend structure that has not been captured as observed in the spectral plot. There is a peak near 0 and so there is some uncaptured trend.

Q5.

model6<-lm(dlogSpending~poly(Time, 4) + fMonth + c348 + s348 + c432 + s432 + obs169+ obs205+ obs316, data = rconstruct);summary(model6)

**OUTPUT:**

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)     -0.082422   0.003436 -23.985  < 2e-16 ***
poly(Time, 4)1   0.007558   0.017906   0.422 0.673234    
poly(Time, 4)2   0.031610   0.017933   1.763 0.078896 .  
poly(Time, 4)3   0.004074   0.017875   0.228 0.819876    
poly(Time, 4)4  -0.045399   0.017899  -2.536 0.011667 *  
fMonth2          0.075957   0.004774  15.910  < 2e-16 ***
fMonth3          0.191480   0.004772  40.129  < 2e-16 ***
fMonth4          0.152641   0.004817  31.690  < 2e-16 ***
fMonth5          0.145905   0.004773  30.568  < 2e-16 ***
fMonth6          0.147240   0.004773  30.851  < 2e-16 ***
fMonth7          0.097249   0.004774  20.369  < 2e-16 ***
fMonth8          0.105999   0.004773  22.208  < 2e-16 ***
fMonth9          0.065430   0.004776  13.701  < 2e-16 ***
fMonth10         0.066283   0.004774  13.885  < 2e-16 ***
fMonth11         0.013814   0.004777   2.892 0.004086 ** 
fMonth12        -0.025913   0.004776  -5.426 1.13e-07 ***
c348             0.005318   0.001360   3.911 0.000112 ***
s348            -0.007054   0.001358  -5.196 3.60e-07 ***
c432            -0.001978   0.001366  -1.448 0.148569    
s432            -0.002618   0.001355  -1.932 0.054267 .  
obs169           0.003108   0.018348   0.169 0.865590    
obs205          -0.081212   0.018319  -4.433 1.27e-05 ***
obs316          -0.064967   0.018369  -3.537 0.000464 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01784 on 325 degrees of freedom
Multiple R-squared:  0.9343,	Adjusted R-squared:  0.9299 
F-statistic: 210.1 on 22 and 325 DF,  p-value: < 2.2e-16
```

**OBSERVATION:** We can notice that the trigonometric pair (c432, s432) and outlier obs169 are not significant above. Hence we will refit another model removing these insignificant terms.

model7<-lm(dlogSpending~poly(Time, 4) + fMonth + c348 + s348 + obs205+ obs316, data = rconstruct);summary(model7)

**OUTPUT:**

```
Residuals:
     Min       1Q   Median       3Q      Max 
-0.06417 -0.01042  0.00020  0.01117  0.05565 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)     -0.082293   0.003387 -24.300  < 2e-16 ***
poly(Time, 4)1   0.007199   0.017981   0.400 0.689145    
poly(Time, 4)2   0.031414   0.017973   1.748 0.081437 .  
poly(Time, 4)3   0.003937   0.017950   0.219 0.826544    
poly(Time, 4)4  -0.045297   0.017938  -2.525 0.012033 *  
fMonth2          0.075778   0.004746  15.967  < 2e-16 ***
fMonth3          0.191322   0.004746  40.312  < 2e-16 ***
fMonth4          0.152535   0.004789  31.852  < 2e-16 ***
fMonth5          0.145622   0.004746  30.681  < 2e-16 ***
fMonth6          0.147290   0.004747  31.030  < 2e-16 ***
fMonth7          0.096948   0.004747  20.422  < 2e-16 ***
fMonth8          0.106005   0.004747  22.329  < 2e-16 ***
fMonth9          0.065229   0.004748  13.739  < 2e-16 ***
fMonth10         0.066153   0.004748  13.931  < 2e-16 ***
fMonth11         0.013761   0.004749   2.898 0.004011 ** 
fMonth12        -0.026177   0.004750  -5.511 7.21e-08 ***
c348             0.005310   0.001365   3.891 0.000121 ***
s348            -0.007083   0.001360  -5.208 3.37e-07 ***
obs205          -0.078533   0.018340  -4.282 2.43e-05 ***
obs316          -0.062775   0.018388  -3.414 0.000721 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01791 on 328 degrees of freedom
Multiple R-squared:  0.9331,	Adjusted R-squared:  0.9293 
F-statistic: 240.9 on 19 and 328 DF,  p-value: < 2.2e-16
```

Now, let us perform the partial F-test on the above two models.

anova(model6, model7)

**OUTPUT:**

```
> anova(model6, model7)
Analysis of Variance Table

Model 1: dlogSpending ~ poly(Time, 4) + fMonth + c348 + s348 + c432 +
    s432 + obs169 + obs205 + obs316
Model 2: dlogSpending ~ poly(Time, 4) + fMonth + c348 + s348 + obs205 +
    obs316
  Res.Df     RSS Df  Sum of Sq      F Pr(>F)
1    325 0.10339
2    328 0.10523 -3 -0.0018413 1.9293 0.1247
```

As the p-value is greater than 0.05 above, it is advisable to discard c432, s432, and obs 169.

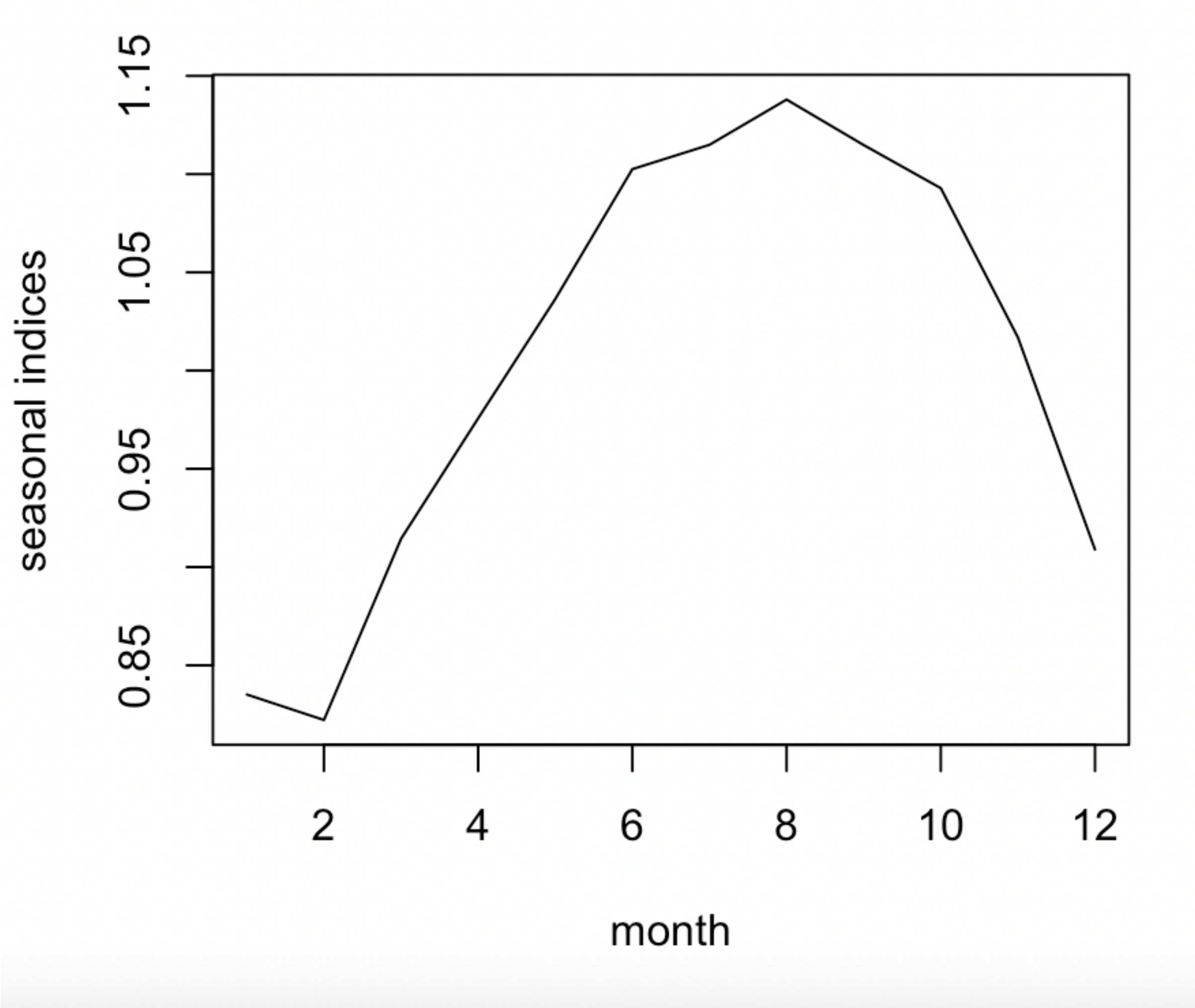
Q5A).
Hence, we will select model7 now.
So,

```
b1<-coef(model7)[1]
b2<-coef(model7)[6:16]+b1
b3<-c(b1,b2)
x<-b3-mean(b3)
s12<-0
for(j in 2:12){
  xsub<-x[j:12]
  s12<-s12+sum(xsub)
}
s12<-s12/12
s<-c(rep(0,times=12))
s[12]<-s12
for(j in 1:11){
  xsub<-x[1:j]
  s[j]<-s[12]+sum(xsub)
}
seas1<-exp(s)
cbind(seas,seas1)
```

**OUTPUT:**

```
> cbind(seas,seas1)
                 seas     seas1
(Intercept) 0.8351272 0.8325969
fMonth2     0.8221664 0.8239602
fMonth3     0.9143693 0.9152875
fMonth4     0.9757027 0.9780571
fMonth5     1.0364023 1.0379313
fMonth6     1.1025322 1.1033092
fMonth7     1.1149156 1.1152251
fMonth8     1.1379529 1.1375258
fMonth9     1.1147980 1.1139127
fMonth10    1.0928500 1.0917988
fMonth11    1.0169026 1.0155012
fMonth12    0.9089218 0.9075559
```

**PLOT:**



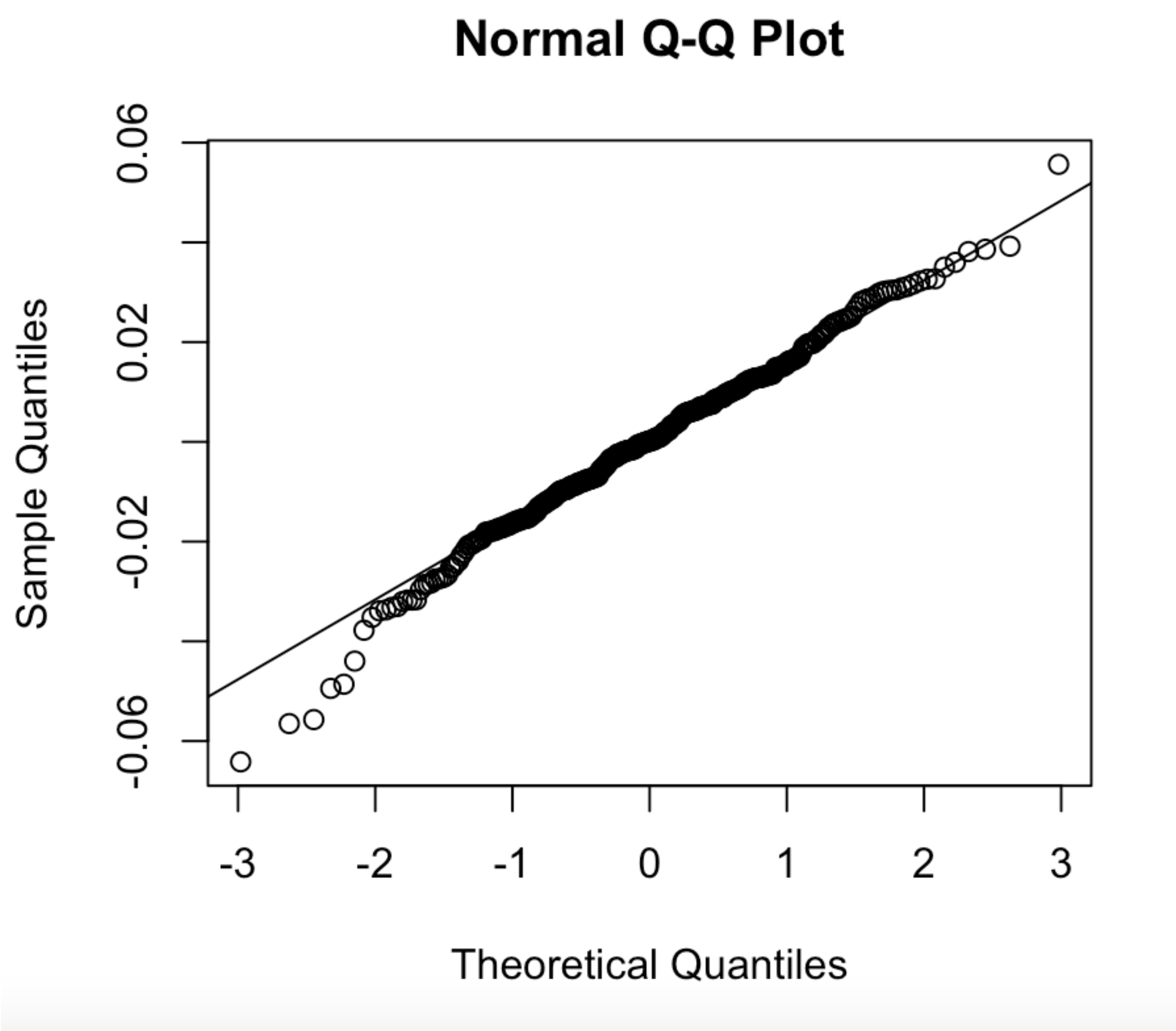

**OBSERVATION:**
We notice that the peak spending is in the month of August. Also, there is an increase in spending from February to August. There is a decrease in the spending from August onwards. We can also observe that the spending is above the trend level in the months of May, June, October, and November. The sales are lowest in the month of February. So, the spending of construction is highest in the fall months and lowest in the winter months (like December, January, and February). We can also notice that spending gradually increases in the summer months. It is similar to question 3.

**CONCLUSION:** We can say that the money spent on construction is less in the winter months as compared to the fall and the summer months. This maybe due to the favourable weather conditions, where people were more comfortable to build in these months. Also, during the winter months, it is possible that people spent their money on winter vacation because of which they were not able to spend money on building stuff. Thus there is a decrease in the spending.

Q5b). i).
#Normal quantile plot:

qqnorm(resid(model7))
qqline(resid(model7))

**OUTPUT:**



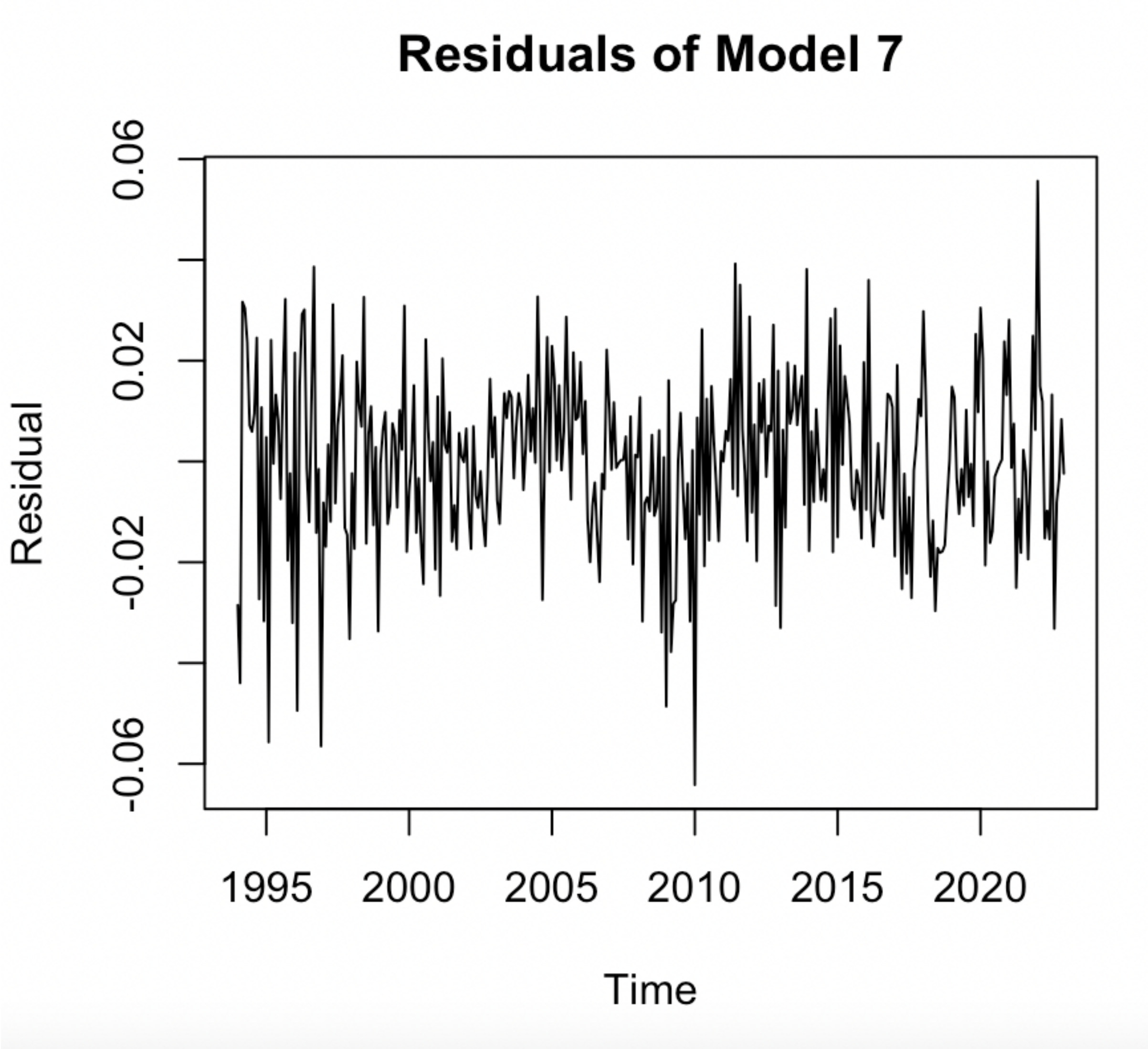

#Test for normality:

shapiro.test(resid(model7))

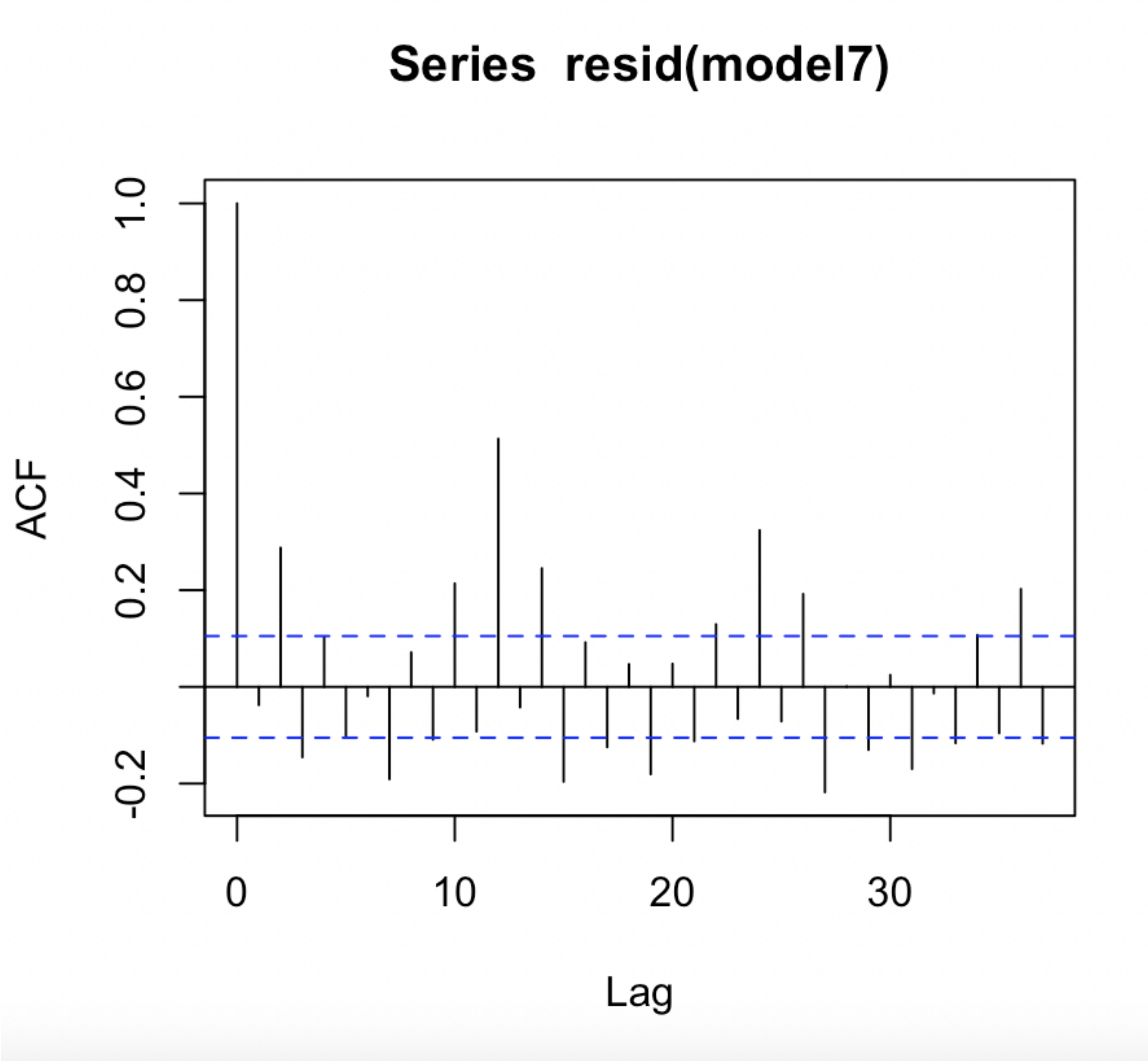
**OUTPUT:**

```
        Shapiro-Wilk normality test

data:  resid(model7)
W = 0.99026, p-value = 0.02074
```

**OBSERVATION:** We notice that the p-value is less than 0.05 here.



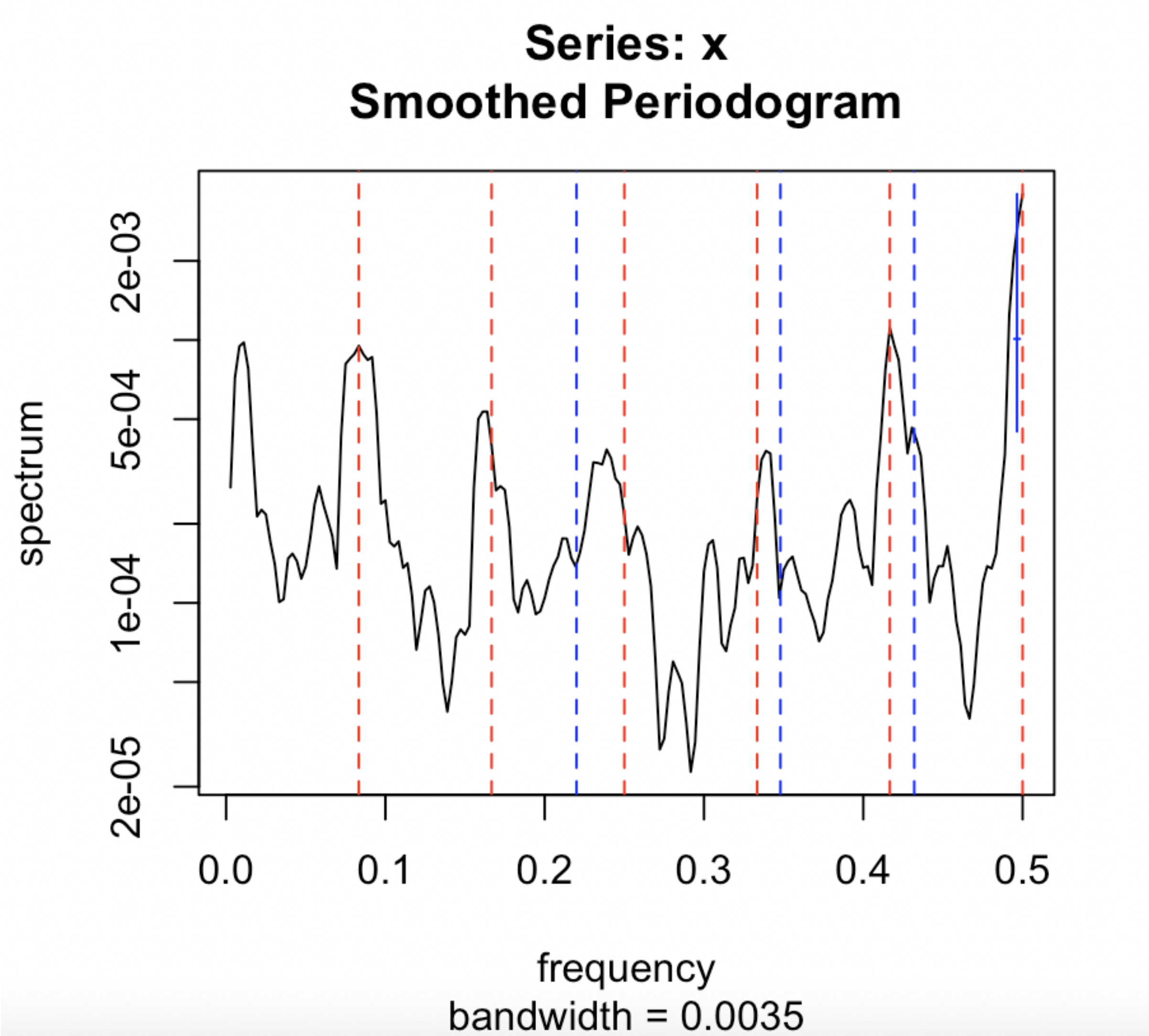

**OBSERVATION:** We can clearly identify shorter upper and lower tails. Also, we can say that there is some significant amount of trend in the residual plot (near 2008-2009 great recession) and let us now see tha ACF plot below,




**OBSERVATION:** We can notice that there is significant autocorrelation left in the autocorrelation plot. There is greater amount of spikes observed here as compared to the previous model.

```
#Spectral densities for Model 7:

spectrum(resid(model7),span=5)
abline(v=c(1/12,2/12,3/12,4/12,5/12,6/12),col="red",lty=2)
abline(v=c(0.220,0.348,0.432),col="blue",lty=2)
```

**OUTPUT:**




**OBSERVATION:** There is a peak near zero (at lower frequency) and thus, we can conclude that there is some amount of trend component in the above plot.
Also, the red dashed lines or the seasonal frequencies are at 1/12 and 5/12. The blue-dashed lines or the calendar frequencies are present at the higher frequencies. Also, the frequency, 0.432 is significant this time. The plot has captured a strong seasonal component and almost no calendar component.

Q5 C). We can include external variables such as inflation, interest rates etc to Spending. These can impact this considerably. Also, various model specifications can be explored to observe if they lead to better performance. Trend should be captured by the model. We can try adding Changepoint variable with lag residuals here.

Q6 A).

```
model8 <- lm(dlogSpending ~ poly(Time, 3)+ Changepoint + Changepoint*Time+ c348 + s348 + c432 + s432 + fMonth + obs169+ obs205, data = rconstruct[1:288,]); summary(model8)
```
**OUTPUT:**

```
Residuals:
      Min        1Q    Median        3Q       Max 
-0.056032 -0.010850  0.001082  0.009832  0.042018 

Coefficients: (1 not defined because of singularities)
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)      -0.1068880  0.0068459 -15.614  < 2e-16 ***
poly(Time, 3)1   -0.3729370  0.1206819  -3.090 0.002213 ** 
poly(Time, 3)2   -0.2356678  0.0752537  -3.132 0.001933 ** 
poly(Time, 3)3   -0.0622385  0.0278638  -2.234 0.026339 *  
Changepoint      -0.1678137  0.0410112  -4.092 5.68e-05 ***
Time                     NA         NA      NA       NA    
c348              0.0052855  0.0014451   3.658 0.000307 ***
s348             -0.0074553  0.0014460  -5.156 4.95e-07 ***
c432             -0.0016575  0.0014509  -1.142 0.254321    
s432             -0.0033918  0.0014443  -2.348 0.019591 *  
fMonth2           0.0815125  0.0051059  15.964  < 2e-16 ***
fMonth3           0.2001400  0.0051005  39.239  < 2e-16 ***
fMonth4           0.1635759  0.0051021  32.060  < 2e-16 ***
fMonth5           0.1558610  0.0051029  30.543  < 2e-16 ***
fMonth6           0.1586717  0.0051011  31.105  < 2e-16 ***
fMonth7           0.1050472  0.0051033  20.584  < 2e-16 ***
fMonth8           0.1163936  0.0051017  22.815  < 2e-16 ***
fMonth9           0.0753427  0.0051055  14.757  < 2e-16 ***
fMonth10          0.0755271  0.0051015  14.805  < 2e-16 ***
fMonth11          0.0185633  0.0051079   3.634 0.000335 ***
fMonth12         -0.0190613  0.0051052  -3.734 0.000231 ***
obs169            0.0276708  0.0183736   1.506 0.133256    
obs205           -0.0727875  0.0178536  -4.077 6.04e-05 ***
Changepoint:Time  0.0009019  0.0002329   3.873 0.000136 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01726 on 265 degrees of freedom
Multiple R-squared:  0.9431,	Adjusted R-squared:  0.9384 
F-statistic: 199.6 on 22 and 265 DF,  p-value: < 2.2e-16
```

**OBSERVATION:** We have only included obs169 and obs205 above. We observe that c432 is not significant here. Outlier obs 169 is also not significant here.
So,
Let us refit the model.

model9<-lm(dlogSpending~ poly(Time, 3) + Changepoint + Changepoint*Time + c348 +s348 + fMonth +obs205, data = rconstruct[1:288,]);summary(model9)

**OUTPUT:**

```
Residuals:
      Min        1Q    Median        3Q       Max
-0.054275 -0.011525  0.000651  0.010775  0.038919

Coefficients: (1 not defined because of singularities)
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)      -0.1044587  0.0067409 -15.496  < 2e-16 ***
poly(Time, 3)1   -0.3456665  0.1206220  -2.866 0.004491 **
poly(Time, 3)2   -0.2171893  0.0750287  -2.895 0.004107 **
poly(Time, 3)3   -0.0556330  0.0277649  -2.004 0.046107 *
Changepoint      -0.1549117  0.0404829  -3.827 0.000162 ***
Time                     NA         NA      NA       NA
c348              0.0053540  0.0014598   3.668 0.000295 ***
s348             -0.0076501  0.0014567  -5.252 3.07e-07 ***
fMonth2           0.0801302  0.0050910  15.739  < 2e-16 ***
fMonth3           0.1988339  0.0050914  39.053  < 2e-16 ***
fMonth4           0.1624468  0.0050910  31.909  < 2e-16 ***
fMonth5           0.1543504  0.0050916  30.315  < 2e-16 ***
fMonth6           0.1576771  0.0050922  30.965  < 2e-16 ***
fMonth7           0.1035138  0.0050922  20.328  < 2e-16 ***
fMonth8           0.1153345  0.0050933  22.644  < 2e-16 ***
fMonth9           0.0739548  0.0050941  14.518  < 2e-16 ***
fMonth10          0.0742800  0.0050947  14.580  < 2e-16 ***
fMonth11          0.0173907  0.0050963   3.412 0.000743 ***
fMonth12         -0.0205165  0.0050973  -4.025 7.42e-05 ***
obs205           -0.0712001  0.0179738  -3.961 9.56e-05 ***
Changepoint:Time  0.0008333  0.0002307   3.613 0.000362 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01744 on 268 degrees of freedom
Multiple R-squared:  0.9412,    Adjusted R-squared:  0.937
F-statistic: 225.8 on 19 and 268 DF,  p-value: < 2.2e-16
```

Let us perform partial F-test on the above two models.
So,

anova(model8,model9)

**OUTPUT:**

```
> anova(model8,model9)
Analysis of Variance Table

Model 1: dlogSpending ~ poly(Time, 3) + Changepoint + Changepoint * Time +
    c348 + s348 + c432 + s432 + fMonth + obs169 + obs205
Model 2: dlogSpending ~ poly(Time, 3) + Changepoint + Changepoint * Time +
    c348 + s348 + fMonth + obs205
  Res.Df      RSS Df  Sum of Sq      F  Pr(>F)
1    265 0.078940
2    268 0.081547 -3 -0.0026074 2.9176 0.03468 *
---
```

**OBSERVATION:** We notice that the p-value is less than 0.05 and thus we will consider model 8 here.

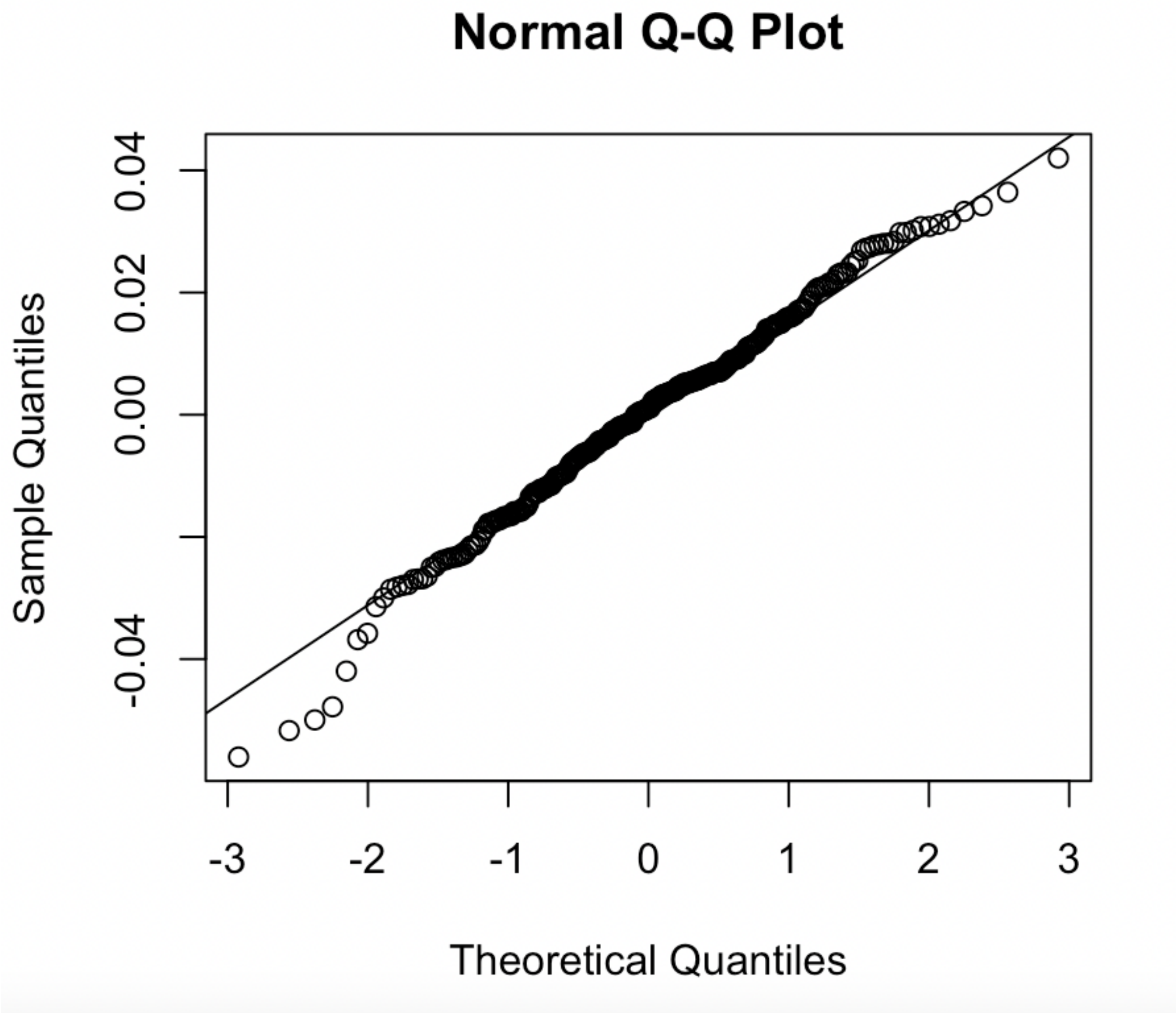
So,

b). Perform residual analysis of the above model:

#Normal quantile plot:

qqnorm(resid(model8))
qqline(resid(model8))

**OUTPUT:**




**OBSERVATION:** There are outliers on both the ends of the tails of the line above. Our model doesn't fit the line very well. Also, it looks normally distributed mostly but there are outliers. Also, let us perform the Shapiro-Wilk test now.

#Test for normality:

shapiro.test(resid(model8))

**OUTPUT:**

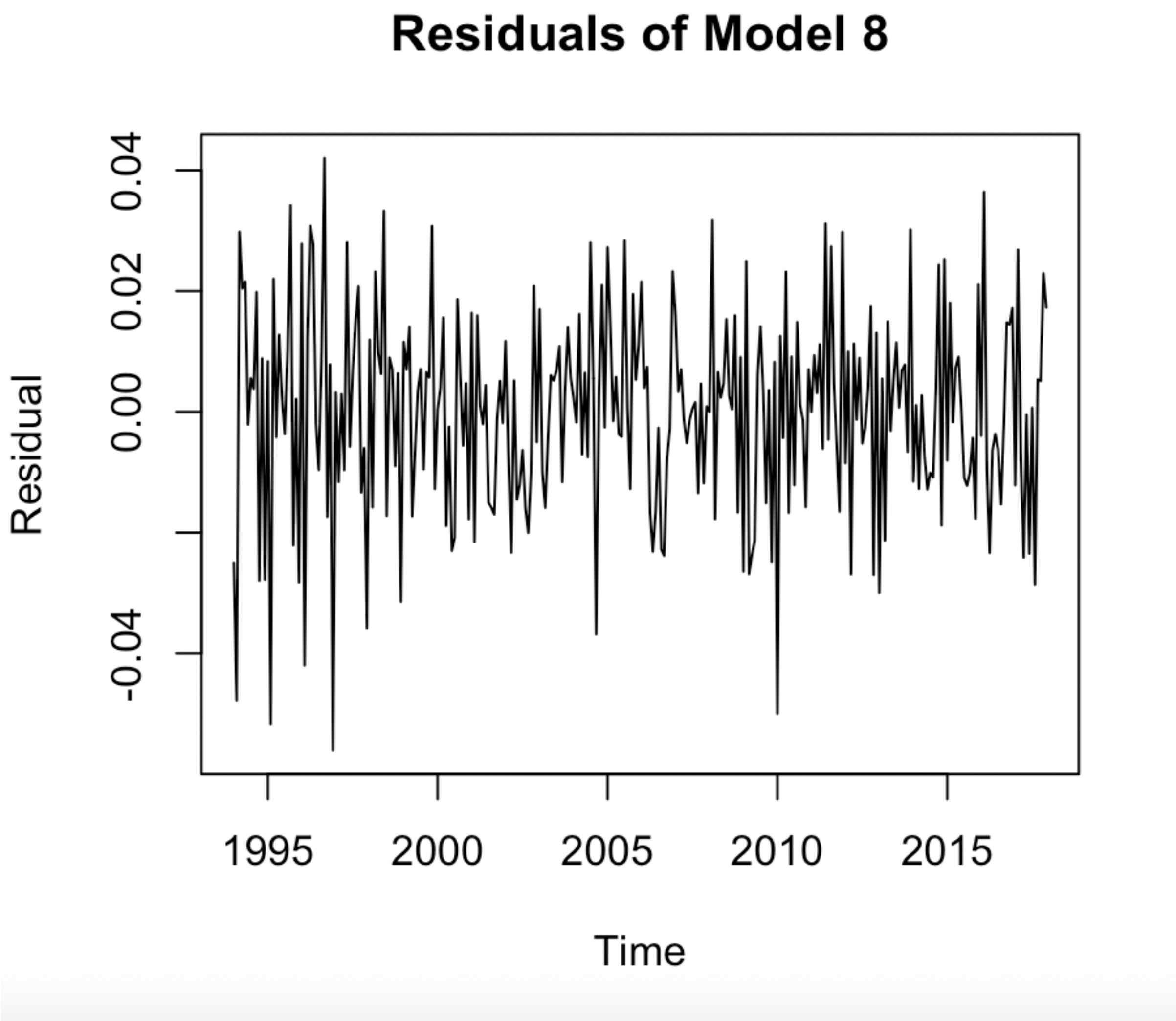
```
> shapiro.test(resid(model8))

	Shapiro-Wilk normality test

data:  resid(model8)
W = 0.98972, p-value = 0.03996
```

**OBSERVATION:** We see that the value of p is less than 0.05 and so it is not significant. Thus, we can say that there are outliers present in the Q-Q plot above.

#Plot residuals versus time:

plot(ts(resid(model8),start=c(1994,1),freq=12),xlab="Time",ylab="Residual",main="Residuals of Model 8")

**OUTPUT:**

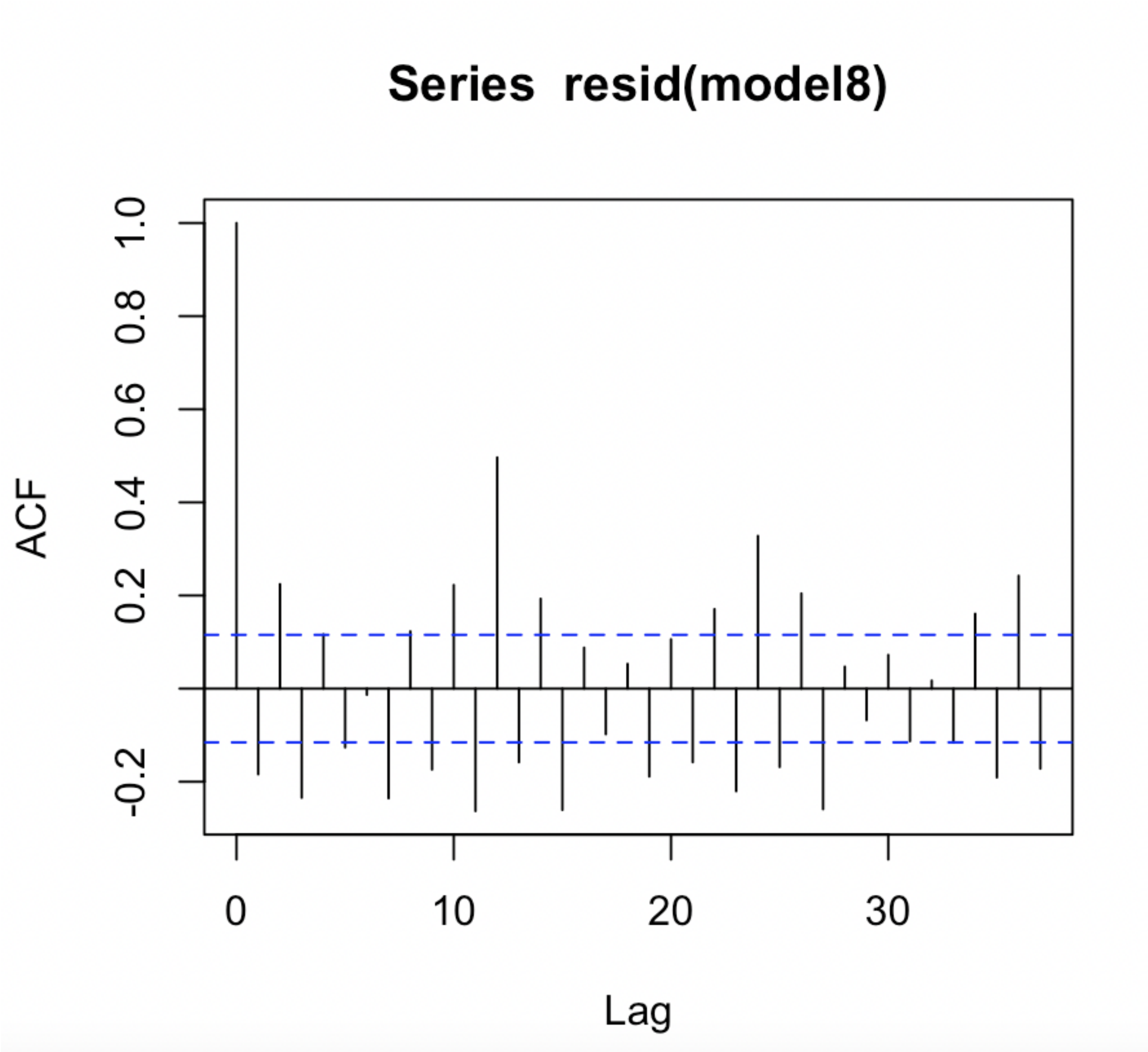



**OBSERVATION:**
Most of the trend has been captured by our model as shown in the above plot. We can say that following the years of recession (2008-2009), some amount of trend has not been captured.

#Estimated autocorrelations:

acf(resid(model8),37)

**OUTPUT:**




**OBSERVATION:** We can see that there are a few spikes that are above the blue-dashed lines (95% confidence interval) and so there is still some amount of correlation remaining. This is not close to white noise completely, but the model has made an attempt to come close to white noise.

Q6 C).

```
forecast<-predict(model8,newdata=rconstruct[289:348,])
forecast

actual_logret <- dlogSpending[289:348]
```

**OUTPUT:**

```
> forecast
          288           289           290           291           292           293 
-0.0932105977 -0.0072376147  0.1014342699  0.0660699760  0.0704170390  0.0521723474 
          294           295           296           297           298           299 
 0.0125490924  0.0258426412 -0.0336877756 -0.0130842951 -0.0808981160 -0.1232346832 
          300           301           302           303           304           305 
-0.0943182288 -0.0174818377  0.0928026978  0.0687954503  0.0550185948  0.0467153028 
          306           307           308           309           310           311 
 0.0147286300  0.0053403360 -0.0312384455 -0.0202892903 -0.0954534761 -0.1236905214 
          312           313           314           315           316           317 
-0.1025740176 -0.0306355252  0.0898293636  0.0629351149  0.0365489086  0.0504016919 
          318           319           320           321           322           323 
 0.0008132782 -0.0085242445 -0.0281882610 -0.0394008782 -0.1024183202 -0.1281089160 
          324           325           326           327           328           329 
-0.1172404816 -0.0411061436  0.0873184356  0.0455888607  0.0265607223  0.0496075970 
          330           331           332           333           334           335 
-0.0232479553 -0.0100466570 -0.0379484725 -0.0590107249 -0.1046524804 -0.1416012841 
          336           337           338           339           340           341 
-0.1329244941 -0.0478846564  0.0771945797  0.0246936259  0.0251580589  0.0334640029 
          342           343           344           345           346           347 
-0.0420978450 -0.0095905746 -0.0614682873 -0.0696160353 -0.1120728556 -0.1614749805 
```

**#Tabulate the plot:**

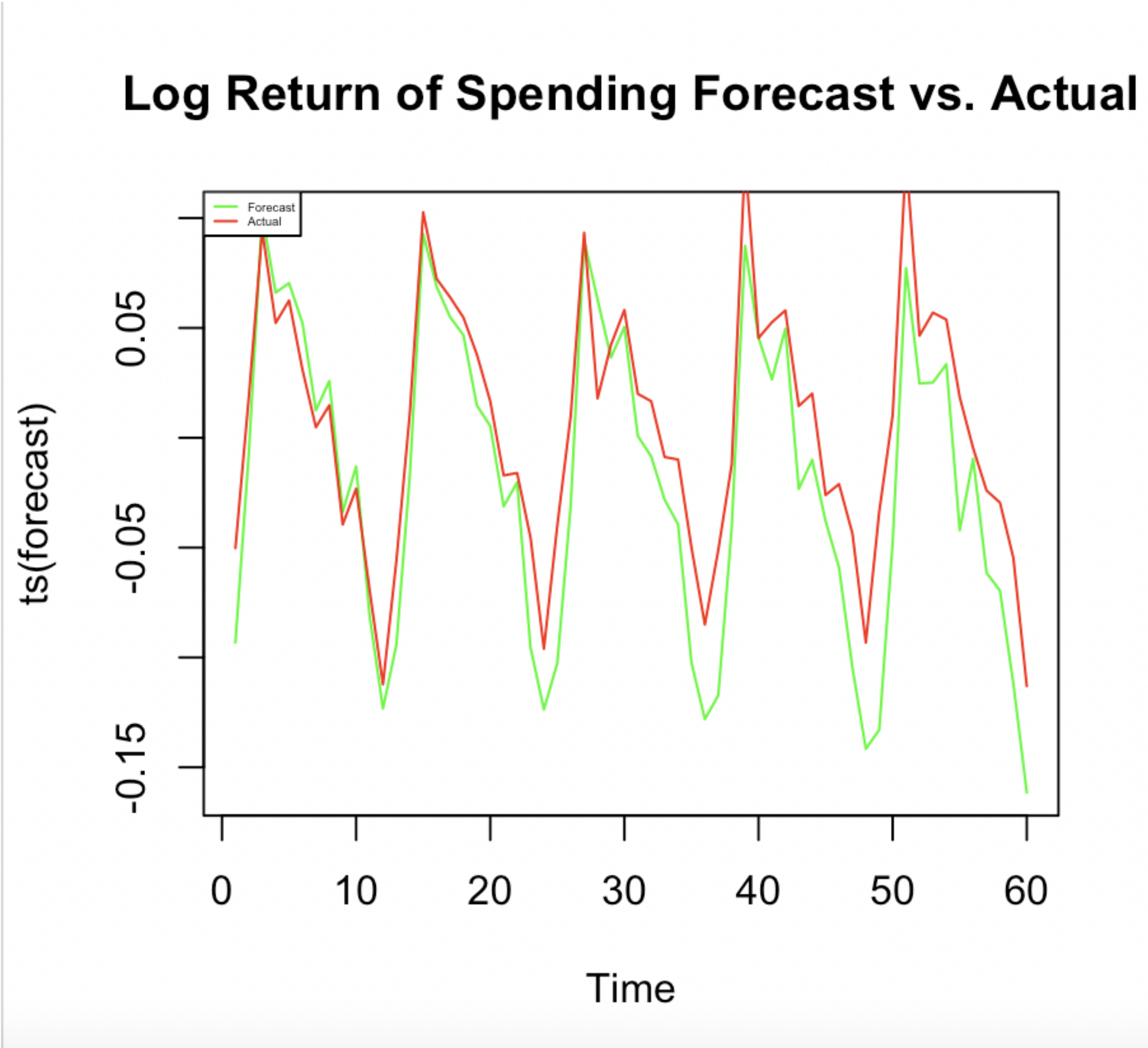
#Tabulating them:
cbind(forecast,actual_logret)

```
> cbind(forecast,actual_logret)
         forecast actual_logret
288 -0.0932105977  -0.050212153
289 -0.0072376147   0.018563388
290  0.1014342699   0.094328340
291  0.0660699760   0.052253284
292  0.0704170390   0.062483161
293  0.0521723474   0.030962794
294  0.0125490924   0.004780872
295  0.0258426412   0.014795596
296 -0.0336877756  -0.039493526
297 -0.0130842951  -0.023218487
298 -0.0808981160  -0.069077473
299 -0.1232346832  -0.112219865
300 -0.0943182288  -0.056136755
301 -0.0174818377   0.010963610
302  0.0928026978   0.102667018
303  0.0687954503   0.072400400
304  0.0550185948   0.064025081
305  0.0467153028   0.054722242
306  0.0147286300   0.037944163
307  0.0053403360   0.016585780
308 -0.0312384455  -0.017134388
309 -0.0202892903  -0.016026271
310 -0.0954534761  -0.045421827
311 -0.1236905214  -0.096041922
312 -0.1025740176  -0.039793391
313 -0.0306355252   0.010088464
314  0.0898293636   0.093346796
315  0.0629351149   0.017946039
316  0.0365489086   0.042350436
317  0.0504016919   0.058172086
318  0.0008132782   0.020001040
319 -0.0085242445   0.016625389
320 -0.0281882610  -0.008769392
321 -0.0394008782  -0.009890782
322 -0.1024183202  -0.049917366
323 -0.1281089160  -0.085015943
```

| | | |
|---|---|---|
| 324 | -0.1172404816 | -0.051101471 |
| 325 | -0.0411061436 | -0.012225047 |
| 326 | 0.0873184356 | 0.127494114 |
| 327 | 0.0455888607 | 0.045427394 |
| 328 | 0.0265607223 | 0.052605164 |
| 329 | 0.0496075970 | 0.057965330 |
| 330 | -0.0232479553 | 0.014450533 |
| 331 | -0.0100466570 | 0.020109651 |
| 332 | -0.0379484725 | -0.026087365 |
| 333 | -0.0590107249 | -0.021080530 |
| 334 | -0.1046524804 | -0.043564556 |
| 335 | -0.1416012841 | -0.093217302 |
| 336 | -0.1329244941 | -0.033625993 |
| 337 | -0.0478846564 | 0.010279003 |
| 338 | 0.0771945797 | 0.127775165 |
| 339 | 0.0246936259 | 0.046483796 |
| 340 | 0.0251580589 | 0.056987486 |
| 341 | 0.0334640029 | 0.053798765 |
| 342 | -0.0420978450 | 0.018442973 |
| 343 | -0.0095905746 | -0.004743839 |
| 344 | -0.0614682873 | -0.024019748 |
| 345 | -0.0696160353 | -0.029514273 |
| 346 | -0.1120728556 | -0.054710231 |
| 347 | -0.1614749805 | -0.113090038 |

```
# Plot the time series
plot(ts(forecast), col = "green", main = "Log Return of Spending Forecast vs. Actual")
lines(ts(actual_logret), col = "red")
```

**OUTPUT:**




Q6 D).

```
jan_spend <- Spending[288]
temp <- numeric()
for(i in 1:60) {
  if (i == 1) {
```

```
    temp <- c(temp, exp(forecast[i] + log(jan_spend, base = exp(1))))
  } else {
    temp <- c(temp, exp(forecast[i] + log(as.numeric(temp[i-1]), base = exp(1))))
  }
}
combined_actual_forecast <- cbind(Spending[289:348], temp, actual_logret, forecast)

Combined_actual_forecast
```

**OUTPUT:**

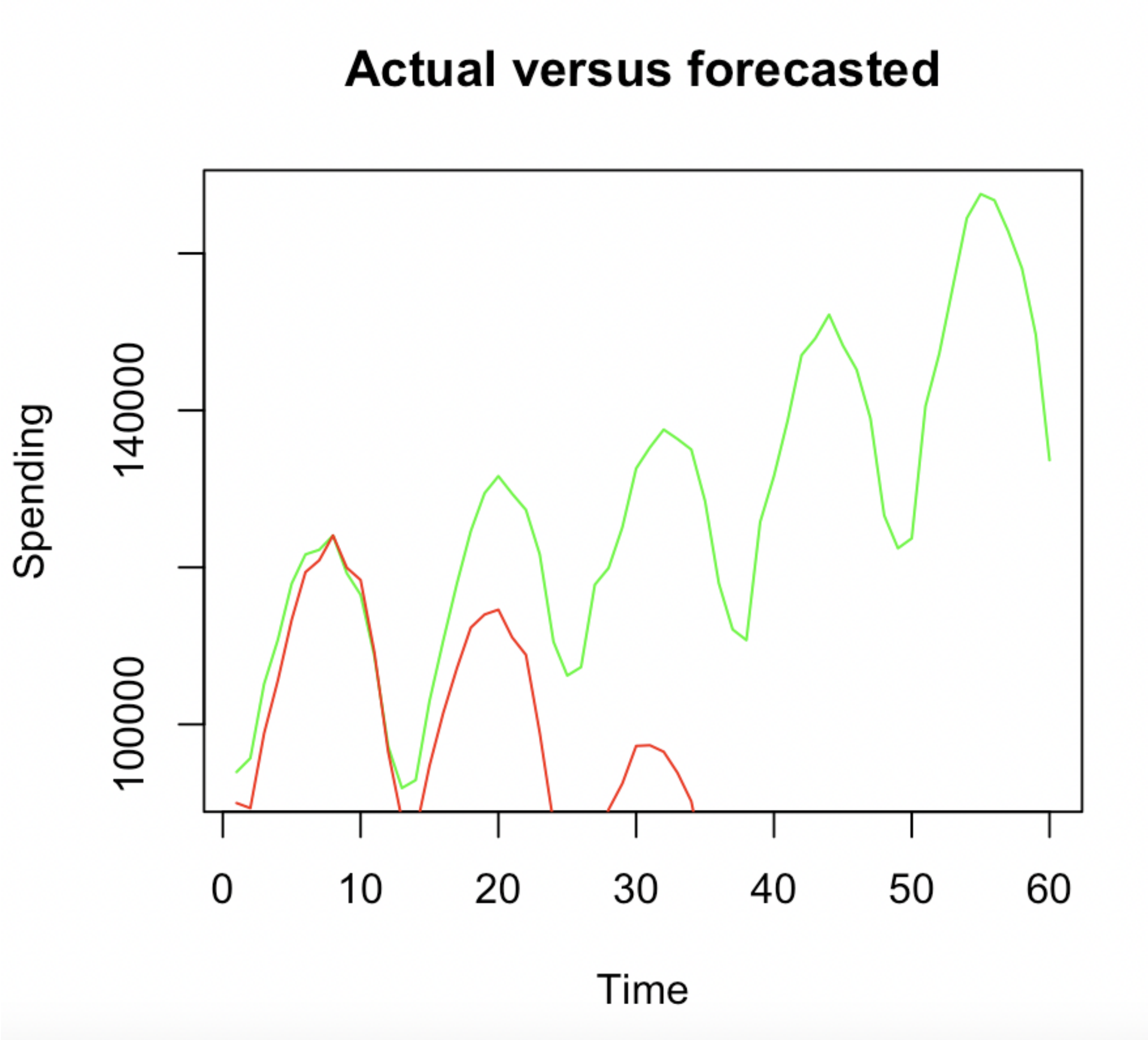
```
> combined_actual_forecast
                  temp actual_logret      forecast
288  93933  89979.63  -0.050212153 -0.0932105977
289  95693  89330.74   0.018563388 -0.0072376147
290 105159  98867.44   0.094328340  0.1014342699
291 110800 105620.23   0.052253284  0.0660699760
292 117944 113325.81   0.062483161  0.0704170390
293 121653 119395.24   0.030962794  0.0521723474
294 122236 120902.98   0.004780872  0.0125490924
295 124058 124068.16   0.014795596  0.0258426412
296 119254 119958.19  -0.039493526 -0.0336877756
297 116517 118398.85  -0.023218487 -0.0130842951
298 108740 109197.80  -0.069077473 -0.0808981160
299  97197  96536.98  -0.112219865 -0.1232346832
300  91891  87847.99  -0.056136755 -0.0943182288
301  92904  86325.59   0.010963610 -0.0174818377
302 102949  94720.35   0.102667018  0.0928026978
303 110679 101466.05   0.072400400  0.0687954503
304 117997 107205.00   0.064025081  0.0550185948
305 124634 112331.93   0.054722242  0.0467153028
306 129454 113998.67   0.037944163  0.0147286300
307 131619 114609.09   0.016585780  0.0053403360
308 129383 111084.23  -0.017134388 -0.0312384455
309 127326 108853.12  -0.016026271 -0.0202892903
310 121672  98943.20  -0.045421827 -0.0954534761
311 110530  87431.48  -0.096041922 -0.1236905214
312 106218  78907.90  -0.039793391 -0.1025740176
313 107295  76527.17   0.010088464 -0.0306355252
314 117793  83719.78   0.093346796  0.0898293636
315 119926  89158.02   0.017946039  0.0629351149
316 125114  92476.93   0.042350436  0.0365489086
317 132608  97257.39   0.058172086  0.0504016919
318 135287  97336.52   0.020001040  0.0008132782
319 137555  96510.32   0.016625389 -0.0085242445
320 136354  93827.85  -0.008769392 -0.0281882610
321 135012  90202.83  -0.009890782 -0.0394008782
322 128438  81421.76  -0.049917366 -0.1024183202
323 117070  71631.41  -0.085015943 -0.1381089160
```

| | | | | |
|---|---|---|---|---|
| 324 | 112093 | 63706.92 | -0.051101471 | -0.1172404816 |
| 325 | 110731 | 61141.26 | -0.012225047 | -0.0411061436 |
| 326 | 125788 | 66720.04 | 0.127494114 | 0.0873184356 |
| 327 | 131634 | 69832.13 | 0.045427394 | 0.0455888607 |
| 328 | 138744 | 71711.78 | 0.052605164 | 0.0265607223 |
| 329 | 147024 | 75358.94 | 0.057965330 | 0.0496075970 |
| 330 | 149164 | 73627.21 | 0.014450533 | -0.0232479553 |
| 331 | 152194 | 72891.20 | 0.020109651 | -0.0100466570 |
| 332 | 148275 | 70176.92 | -0.026087365 | -0.0379484725 |
| 333 | 145182 | 66155.55 | -0.021080530 | -0.0590107249 |
| 334 | 138993 | 59582.17 | -0.043564556 | -0.1046524804 |
| 335 | 126622 | 51715.37 | -0.093217302 | -0.1416012841 |
| 336 | 122435 | 45278.42 | -0.033625993 | -0.1329244941 |
| 337 | 123700 | 43161.37 | 0.010279003 | -0.0478846564 |
| 338 | 140560 | 46625.17 | 0.127775165 | 0.0771945797 |
| 339 | 147248 | 47790.84 | 0.046483796 | 0.0246936259 |
| 340 | 155883 | 49008.42 | 0.056987486 | 0.0251580589 |
| 341 | 164499 | 50676.19 | 0.053798765 | 0.0334640029 |
| 342 | 167561 | 48587.11 | 0.018442973 | -0.0420978450 |
| 343 | 166768 | 48123.36 | -0.004743839 | -0.0095905746 |
| 344 | 162810 | 45254.38 | -0.024019748 | -0.0614682873 |
| 345 | 158075 | 42211.11 | -0.029514273 | -0.0696160353 |
| 346 | 149659 | 37735.85 | -0.054710231 | -0.1120728556 |
| 347 | 133656 | 32108.97 | -0.113090038 | -0.1614749805 |

**PLOT:**

```
#Plot:
plot(Spending[289:348], col='green', type='l', xlab='Time',
     ylab ='Spending', main ='Actual versus forecasted')
lines(as.integer(temp), type='l', col='red')
```

**OUTPUT:**




**OBSERVATION:** The Actual data doesn't fit well with the forecasted data above. Hence, we observe that this model is not helping us fit the data well.

Q6 E).
model10 <- lm(dlogSpending ~ Changepoint + Changepoint*Time+ c348 + s348 + c432 + s432 + fMonth + obs205, data = rconstruct[1:288,]); summary(model10)

**OUTPUT:**

```
Residuals:
      Min        1Q    Median        3Q       Max 
-0.056339 -0.009627  0.000738  0.010747  0.041349 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)     -8.736e-02  4.376e-03 -19.963  < 2e-16 ***
Changepoint     -4.082e-02  1.105e-02  -3.694 0.000268 ***
Time            -1.554e-05  2.791e-05  -0.557 0.578082    
c348             5.338e-03  1.466e-03   3.642 0.000325 ***
s348            -7.668e-03  1.463e-03  -5.243 3.21e-07 ***
c432            -1.429e-03  1.466e-03  -0.974 0.330831    
s432            -3.402e-03  1.466e-03  -2.321 0.021014 *  
fMonth2          8.037e-02  5.114e-03  15.716  < 2e-16 ***
fMonth3          1.990e-01  5.113e-03  38.932  < 2e-16 ***
fMonth4          1.625e-01  5.113e-03  31.772  < 2e-16 ***
fMonth5          1.547e-01  5.114e-03  30.259  < 2e-16 ***
fMonth6          1.576e-01  5.114e-03  30.810  < 2e-16 ***
fMonth7          1.039e-01  5.115e-03  20.313  < 2e-16 ***
fMonth8          1.152e-01  5.114e-03  22.533  < 2e-16 ***
fMonth9          7.415e-02  5.117e-03  14.491  < 2e-16 ***
fMonth10         7.434e-02  5.116e-03  14.531  < 2e-16 ***
fMonth11         1.730e-02  5.119e-03   3.380 0.000832 ***
fMonth12        -2.032e-02  5.119e-03  -3.970 9.24e-05 ***
obs205          -7.143e-02  1.809e-02  -3.949 0.000100 ***
Changepoint:Time  1.768e-04  5.402e-05   3.273 0.001203 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01752 on 268 degrees of freedom
Multiple R-squared:  0.9407,	Adjusted R-squared:  0.9365 
F-statistic: 223.8 on 19 and 268 DF,  p-value: < 2.2e-16
```

**OBSERVATION:** We observe that the trigonometric pair (c432, s432) is not significant. Thus, we shall remove them and refit the model and perform anova test.

model11 <- lm(dlogSpending ~ Changepoint + Changepoint*Time+ c348 + s348 + fMonth + obs205, data = rconstruct[1:288,]); summary(model11)

anova(model10, model11)

**OUTPUT:**

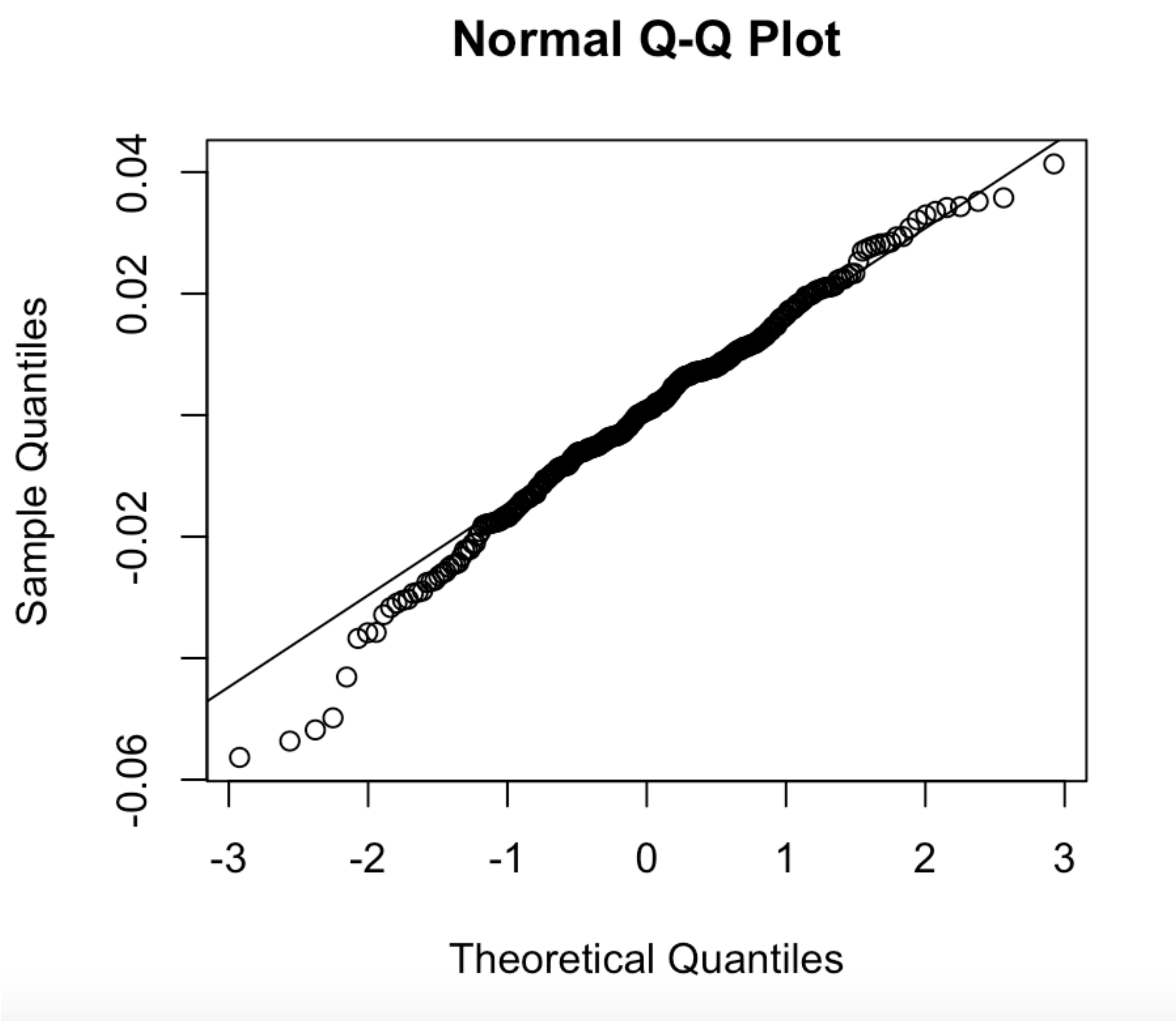
```
> anova(model10, model11)
Analysis of Variance Table

Model 1: dlogSpending ~ Changepoint + Changepoint * Time + c348 + s348 +
    c432 + s432 + fMonth + obs205
Model 2: dlogSpending ~ Changepoint + Changepoint * Time + c348 + s348 +
    fMonth + obs205
  Res.Df      RSS Df  Sum of Sq      F  Pr(>F)
1    268 0.082216
2    270 0.084154 -2 -0.0019377 3.1582 0.04409 *
```

**OBSERVATION:** We observe that p-value < 0.05 and thus it is significant and we retain our model 10. We donot discard the trigonometric pair.

So,

#Normal quantile plot:
qqnorm(resid(model10))
qqline(resid(model10))

**OUTPUT:**




**OBSERVATION:**

We notice that there are outliers at the ends of the two tails of the line above. Our model has thus not properly fit all points. We can also confirm this by Shapiro-Wilk test below:

#Test for normality:

shapiro.test(resid(model10))

**OUTPUT:**

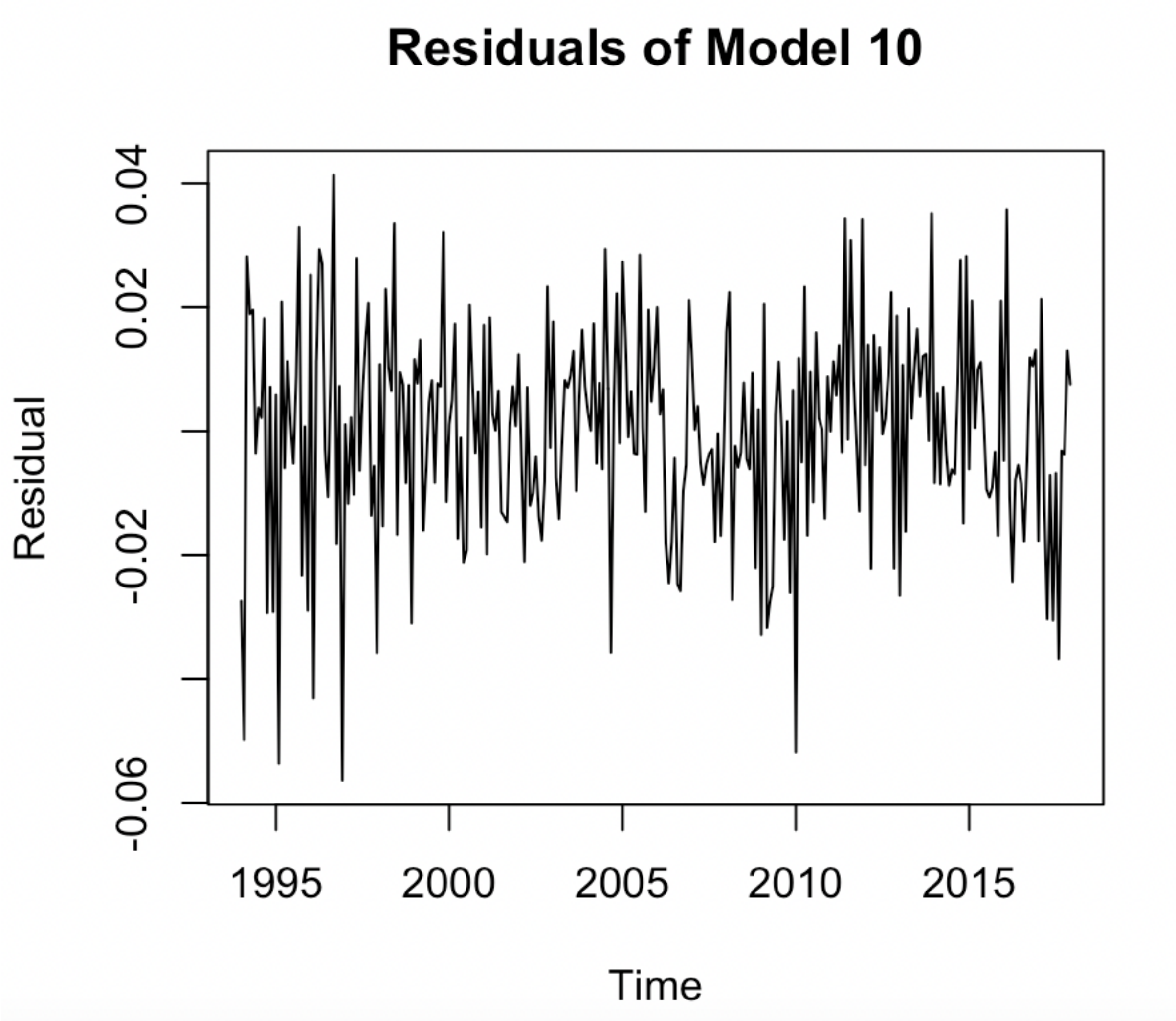
```
Shapiro-Wilk normality test

data:  resid(model10)
W = 0.98811, p-value = 0.01817
```

**OBSERVATION:**
Hence, we can say that the p-value is less than 0.05 and so, we can say that the outliers are present in our model in the normality quantile plot.

#Plot residuals versus time:

plot(ts(resid(model10),start=c(1994,1),freq=12),xlab="Time",ylab="Residual",main="Residuals of Model 10")

**OUTPUT:**

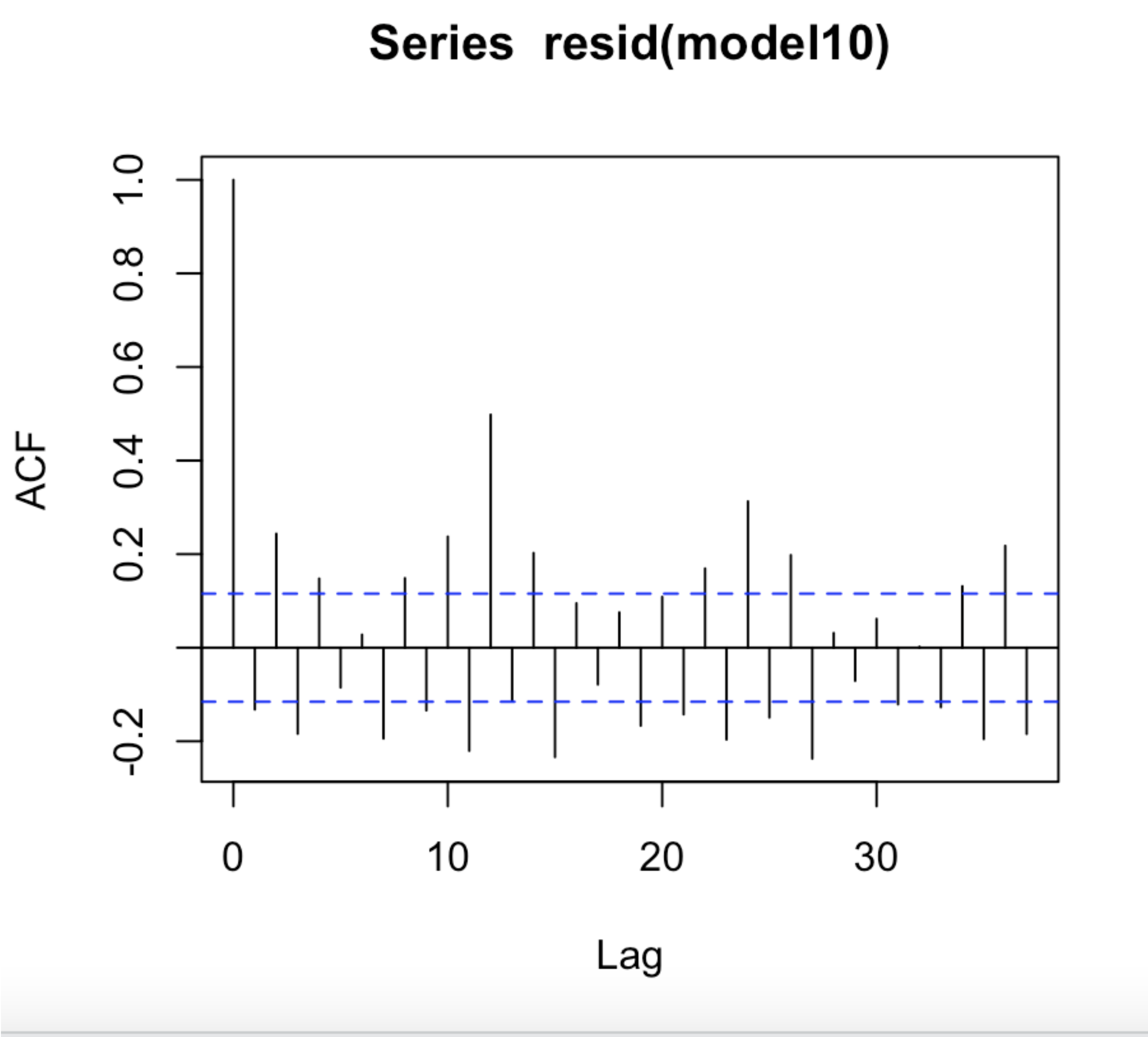



**OBSERVATION:** The trend has not been captured by our model completely. As it is log return data, we can say that the trend component will not be captured well.

#Estimated autocorrelations:

acf(resid(model10),37)

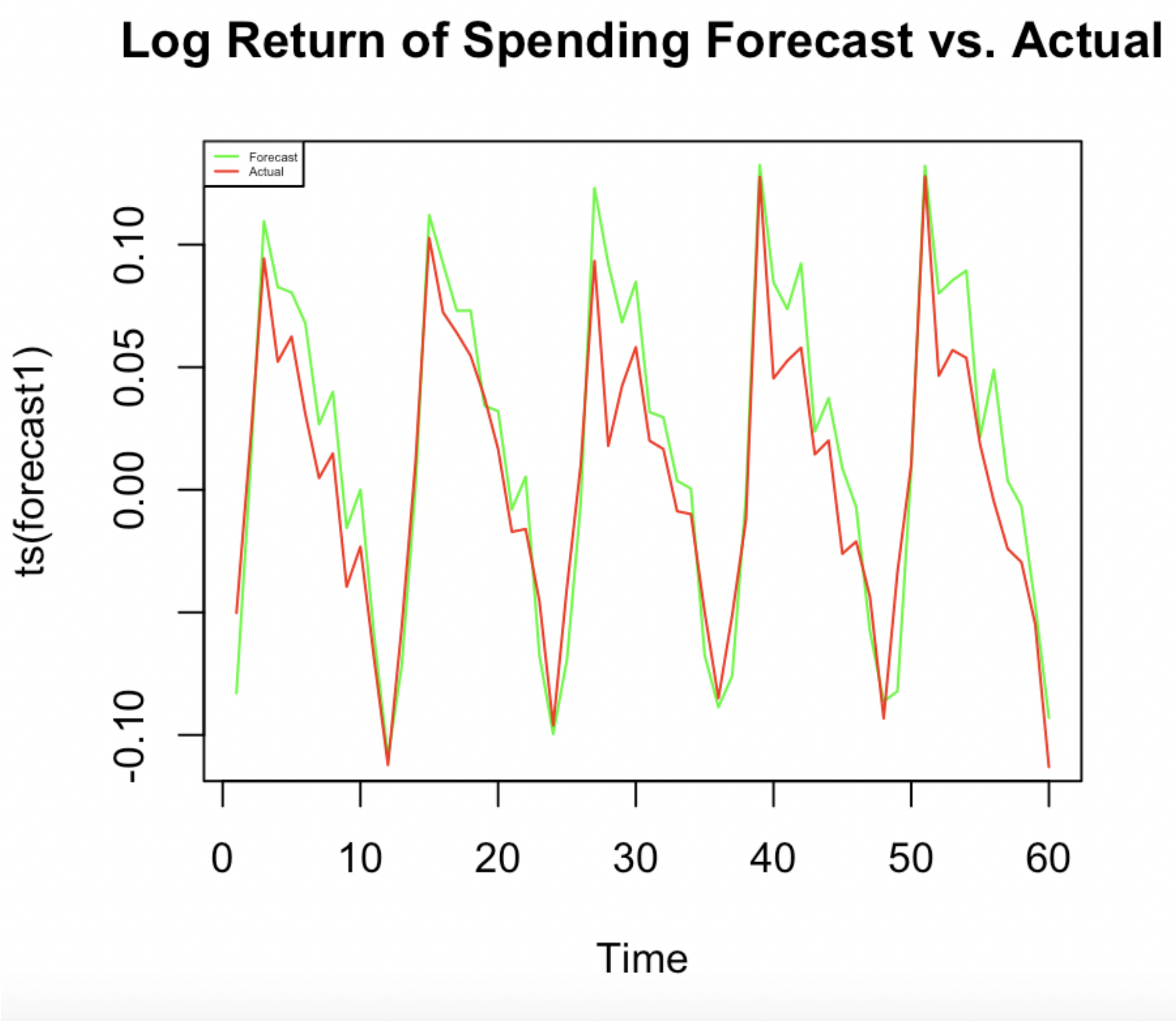
**OUTPUT:**




**OBSERVATION:** We see that the components have been reduced to white noise completely here. There is some significant structure remaining.

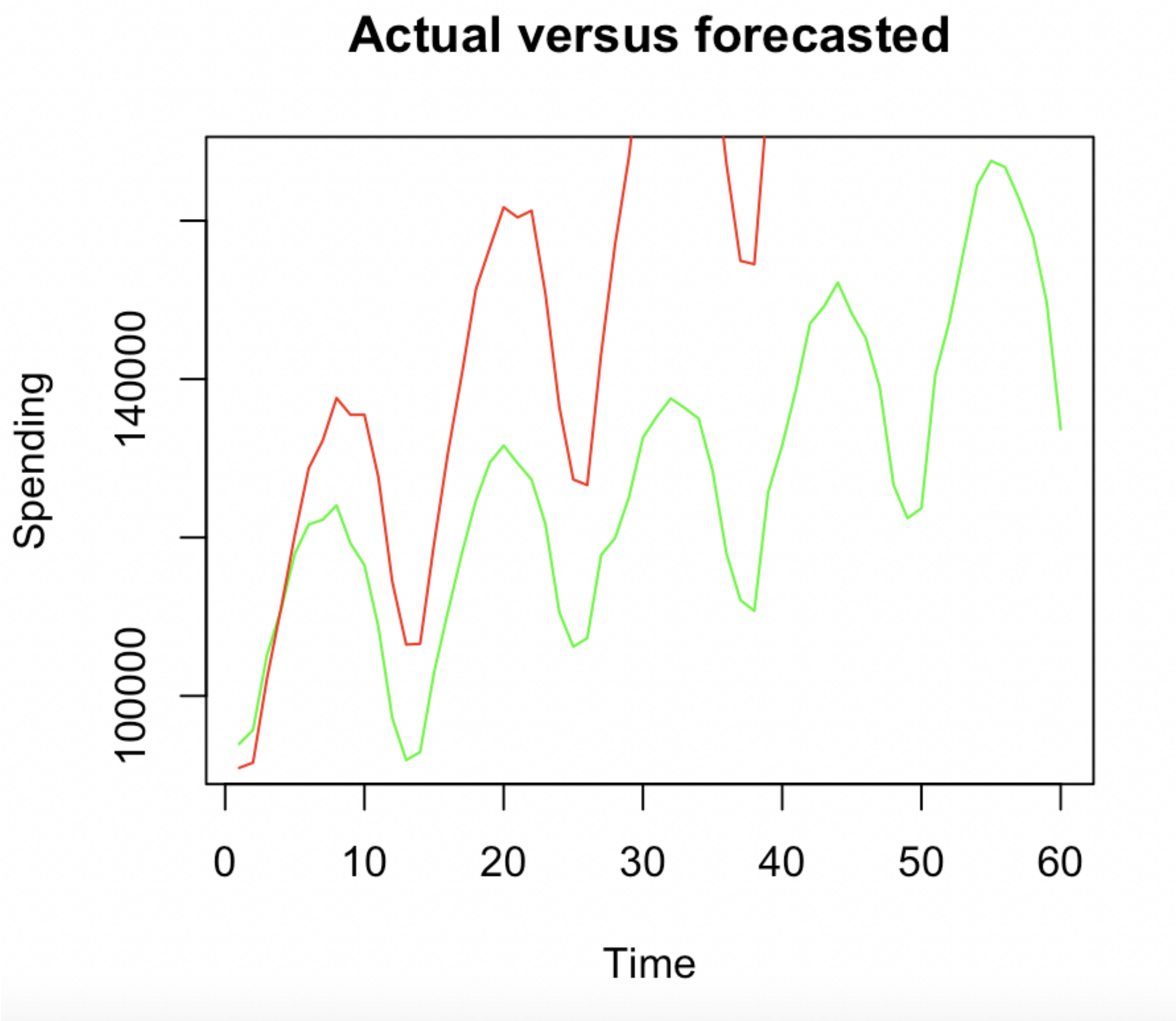
```
#Forecast:
forecast1<-predict(model10,newdata=rconstruct[289:348,])
forecast1
actual_logret1 <- dlogSpending[289:348]


plot(ts(forecast1), col="green", main = "Log Return of Spending Forecast vs. Actual")
lines(ts(actual_logret1), col="red")
legend("topleft", legend = c("Forecast", "Actual"), col = c("green", "red"), lty = 1, cex= 0.3)
```

**OUTPUT:**




**OBSERVATION:** Forecasted log returns are in line with the actual values. The model has captured the seasonal and the trend part well in this case.

```
#d).
jan_spend1 <- Spending[288]
temp1 <- numeric()
for(i in 1:60) {
  if (i == 1) {
    temp1 <- c(temp1, exp(forecast1[i] + log(jan_spend1, base = exp(1))))
  } else {
    temp1 <- c(temp1, exp(forecast1[i] + log(as.numeric(temp1[i-1]), base = exp(1))))
  }
}

#Plot:
```




**OBSERVATION:** We can see that this model looks better than the previous one. The forecasted spending is somewhat aligned with the actual values.

iii). We can see that the log return series has a pretty strong seasonal component. It has no proper trend component present. The calculation of log return involves a differencing and it eliminates the trend component usually. Thus, we can see that analysis of the log return series, leads to no proper estimation of the trend component.

We can also say that, fast movements are made better and the slow ones are attenuated with differencing time series.

We also see that for log return data, there is no proper significant polynomial trend. We thus donot enclose a trend while fitting our model with the log return ones.

We also observed the ACF plot, wherein there were few correlations at a few lags and it is also worthy to note that their magnitude is less compared to the model that had a trend component in it.

It is imperative to note that trend component is not required with a log return series. Results maybe obtained only with several dummy variables and the seasonal components.