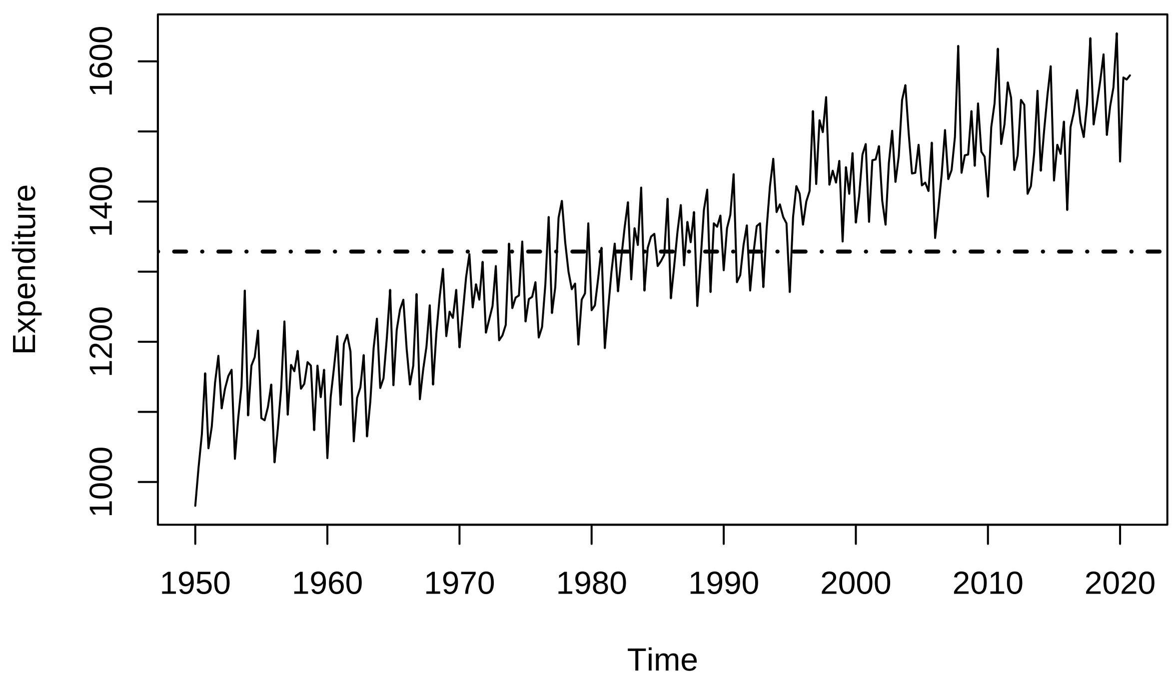
Time series decomposition

2024-04-01

This file can take in any consumption expenditure time series and decompose it into its four components. We will be using an artificial time series from 1950 to 2020.

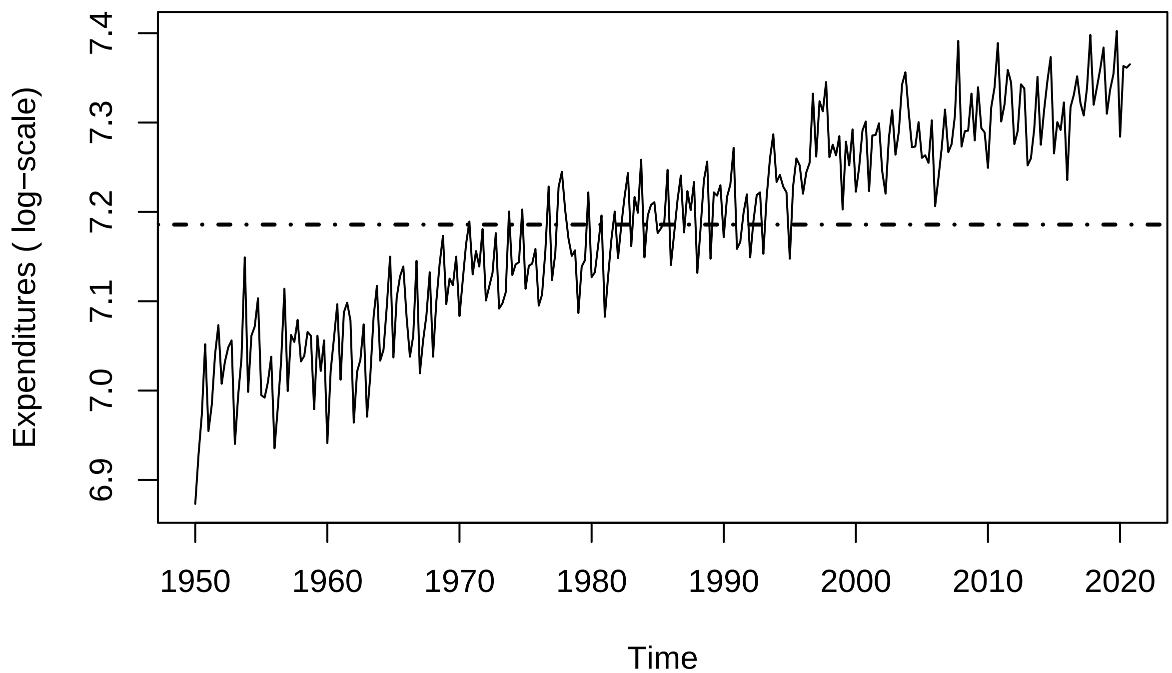
```
data <- read.csv("dat77.csv")
raw_data = data$Expenditure
ts_data <- ts(raw_data, frequency = 4, start = c(1950, 1))
plot(ts_data, ylab="Expenditure")
abline(h=mean(data$Expenditure),col=1,lwd = 2, lty=4)
```



After ploting the series, we can see that it is clearly a positive trend. We added a horizontal line that represents the average of the expenditures to have a better understanding. The trend seems to be slightly decreasing on average over the time period. In terms of short term fluctuation, we observe that most annual increases are followed by a decrease.

Now ploting the series using the log-scale.

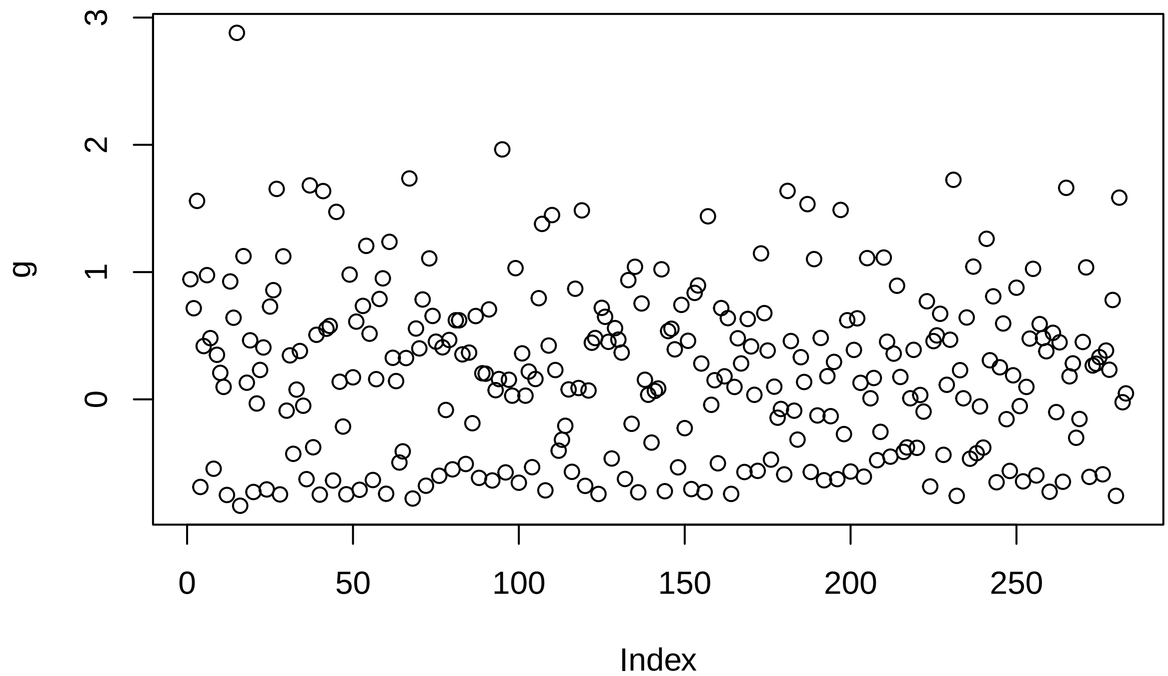
```
plot(log(ts_data), ylab="Expenditures ( log-scale)")
abline(h=mean(log(data$Expenditure)),col=1,lwd = 2, lty=4)
```



We know that the difference in logs is approximately equal to the growth rate. After plotting the line chart expressed in log scale, we see that the growth rate is slightly decreasing on average over the time period.

To better see how the growth rate evolves through time, we plot the annualized growth rate of consumption expenditure.

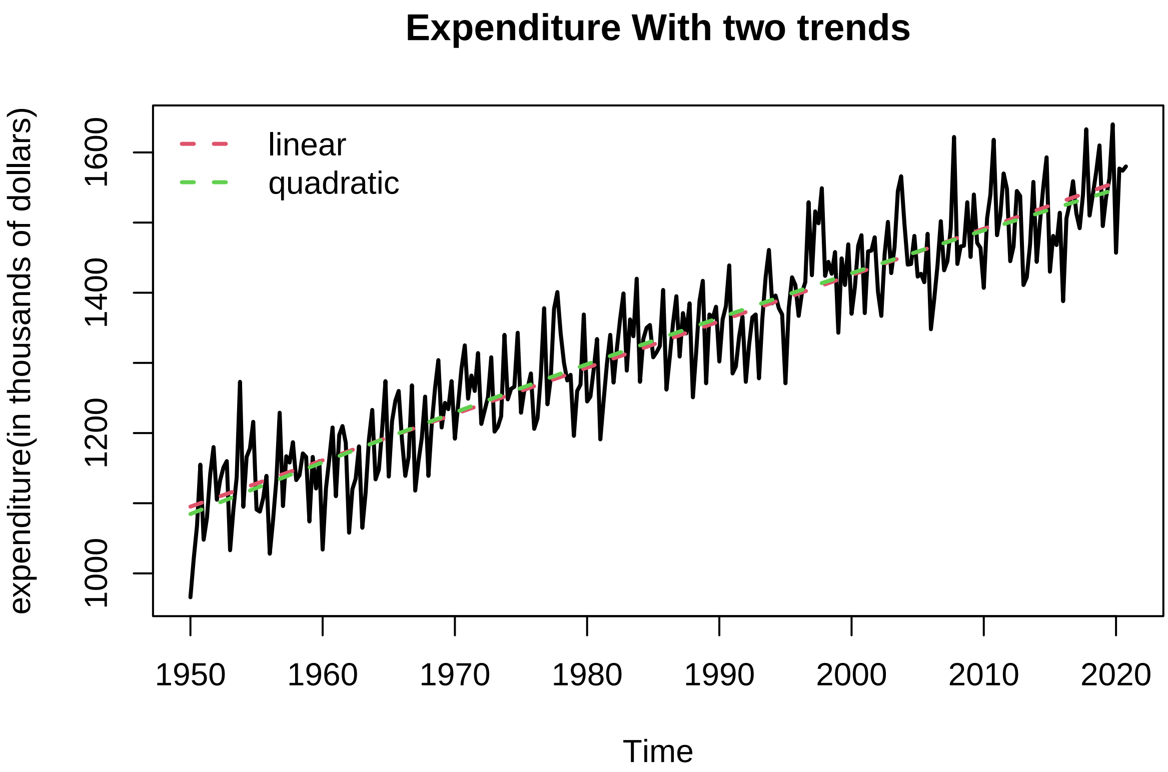
```
difference <- diff(raw_data)
new_values = head(raw_data, -1)
g <- difference/new_values
g <- (1+g)^12-1 ## exact formula
plot(g)
```



After plotting the annualized growth rate of consumption expenditure using a scatter plot, we see the data points spread across the graph. Therefore it is not constant on average.

We now fit a linear and quadratic trends to the series and exam which one fits better.

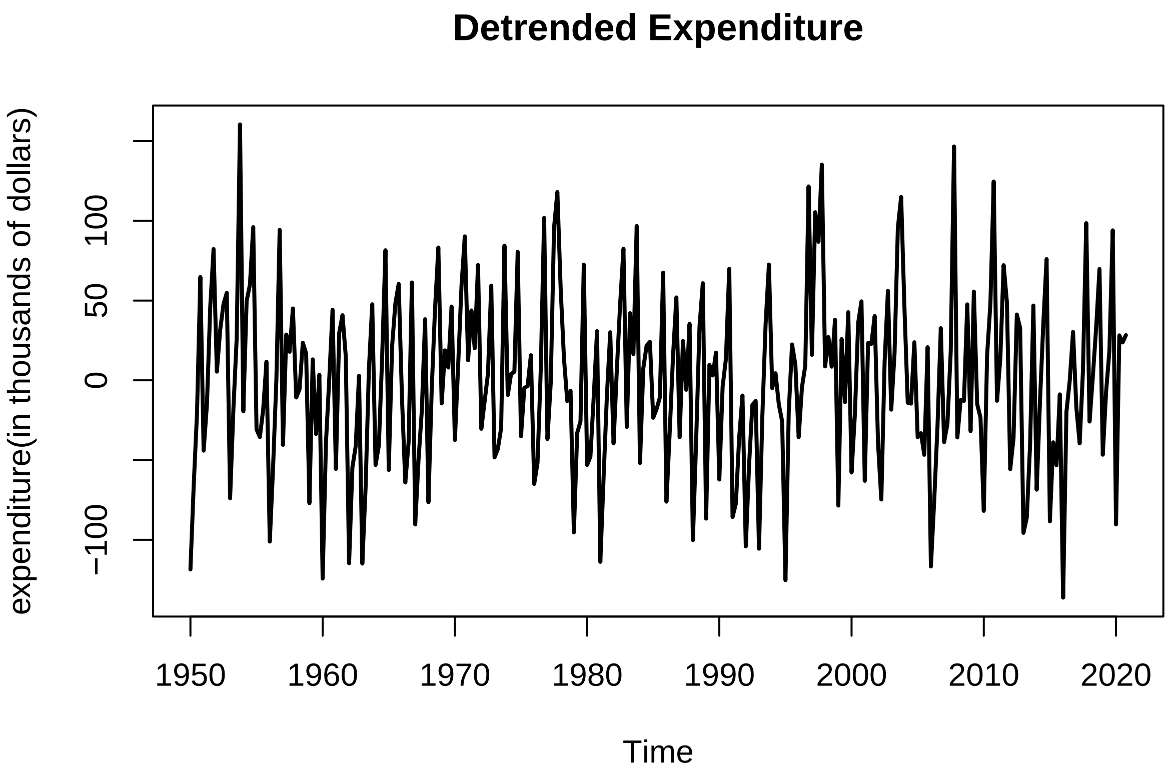
```
t <- time(ts_data)
coefT <- coef(lm(ts_data~t))
trend <- coefT[1] + coefT[2]*t
plot(ts_data, lwd=2, main="Expenditure With two trends",
     ylab="expenditure(in thousands of dollars)")
lines(trend, col=2, lty=2, lwd=2)
t2 <- t^2
fit2 <- lm(ts_data~t+t2)
coefT2 <- coef(fit2)
trend2 <- coefT2[1] + coefT2[2]*t + coefT2[3]*t2
lines(trend2, col=3, lty=2, lwd=2)
legend("topleft", c("linear", "quadratic"), col=2:3, lty=2:2,
       lwd=2, bty='n')
```



After creating a line chart with the original series and the two types of trends. We see that the linear and quadratic trends mostly overlap with each other. Therefore, I would say that they both fit the series. However, I think the quadratic trend best fit the series since it would be more accurate.

We now plot the detrended series using the trend that best fit the series.

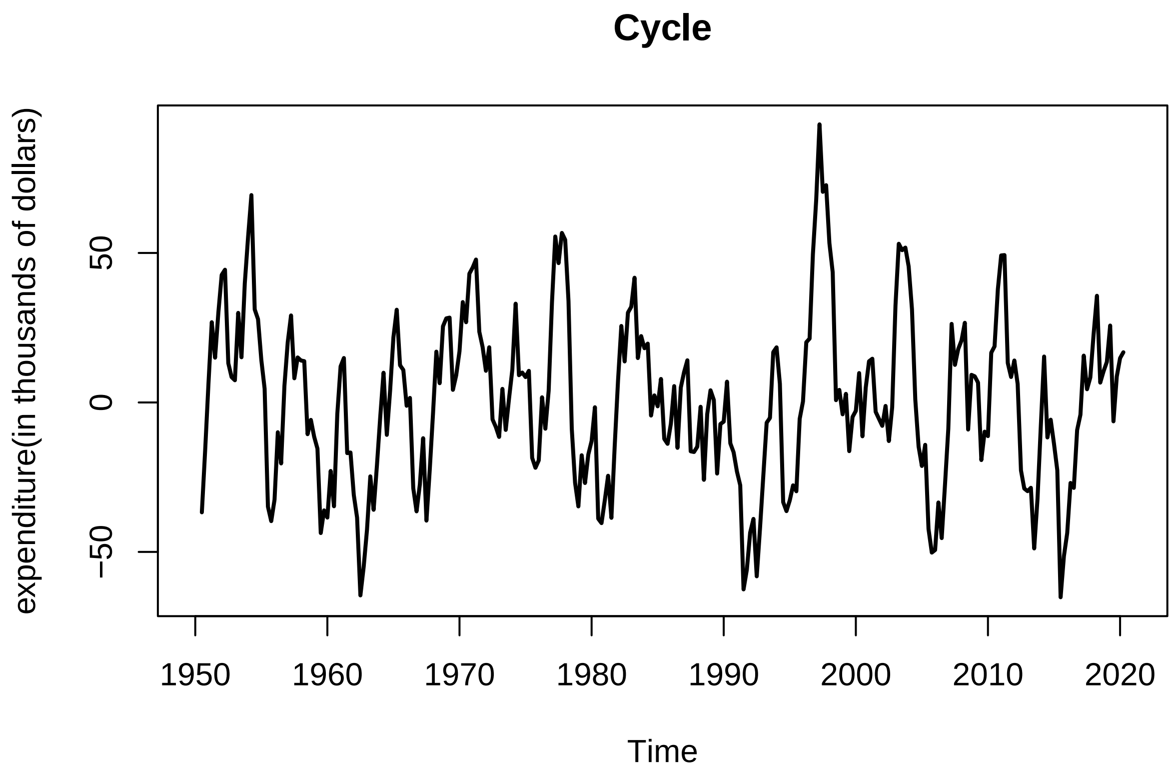
```
t <- time(ts_data)
coefT <- coef(lm(ts_data~t))
trend <- coefT[1] + coefT[2]*t
t2 <- t^2
fit2 <- lm(ts_data~t+t2)
coefT2 <- coef(fit2)
trend2 <- coefT2[1] + coefT2[2]*t + coefT2[3]*t2

CSI <- ts_data-trend2
plot(CSI,
     main="Detrended Expenditure",
     ylab="expenditure(in thousands of dollars)",
     lwd=2)
```



After plotting the detrended series, the short term fluctuation becomes more clear since the range of the y-axis becomes smaller after removing the trend.

Now, we use a moving average of order 5 to compute the cyclical component of the series.

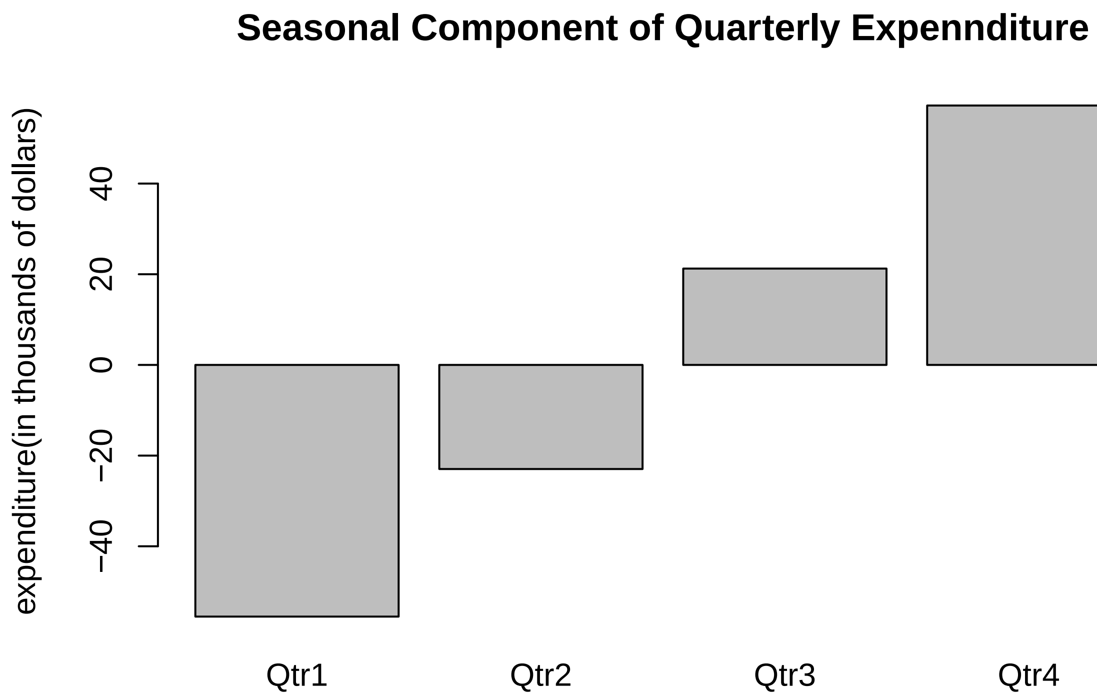
```
dec <- decompose(CSI, filter=rep(1/5,5))
C <- dec$trend

plot(C,
     main="Cycle",
     ylab="expenditure(in thousands of dollars)",
     lwd=2)
```



We see from the graph of the cycle that there are around 10 peaks and troughs in each decade. My interpretation is that the expenditure starts with a trough in a year, increases to a peak at the end of the year, and falls back to a trough at the beginning of the next year.

We now compute the seasonal component

```
barplot(dec$figure,names.arg=c("Qtr1","Qtr2", "Qtr3", "Qtr4"), main="Seasonal Component of Quarterly Exp
```



We see from the bar chart that the Expenditure is at his peak in the forth quarter, and at its trough in the first quarter