Econ 144 Homework 1

Ian Cranfield UID 804-991-210

Due Tuesday 4/14/2020

Packages:

```r
library(dynlm)
library(pastecs)
library(openxlsx)
library(tidyverse)
library(data.table)
library(scales)
library(TSstudio)
library("readxl")
library(ggplot2)
require(cowplot)
library(psych)
library(corrplot)
library(gtable)
library(timeDate)
library(PerformanceAnalytics)
```

Problem 2.2:

```r
qgrowth_rates <- read.xlsx("Chapter2_exercises_data.xlsx", sheet = 1, detectDates =  TRUE)[ , 1:3]
summary(qgrowth_rates)
```

```
##      date               GRGDP            RETURN
##  Length:248         Min.   :-2.7075   Min.   :-26.937
##  Class :character   1st Qu.: 0.3273   1st Qu.: -0.915
##  Mode  :character   Median : 0.7679   Median :  2.023
##                     Mean   : 0.7958   Mean   :  1.962
##                     3rd Qu.: 1.3107   3rd Qu.:  5.753
##                     Max.   : 3.9343   Max.   : 20.117
```

```r
skewness(qgrowth_rates$GRGDP)
```

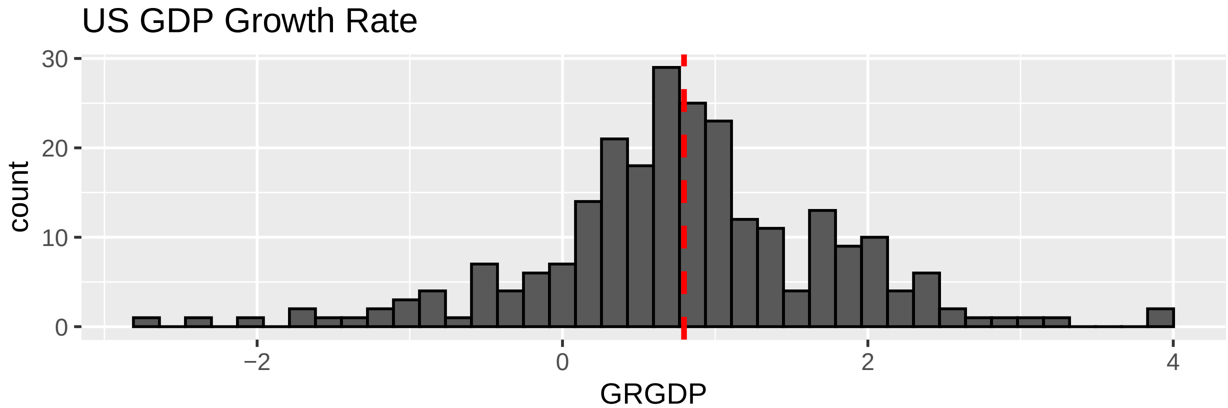
```
## [1] -0.1888557
```

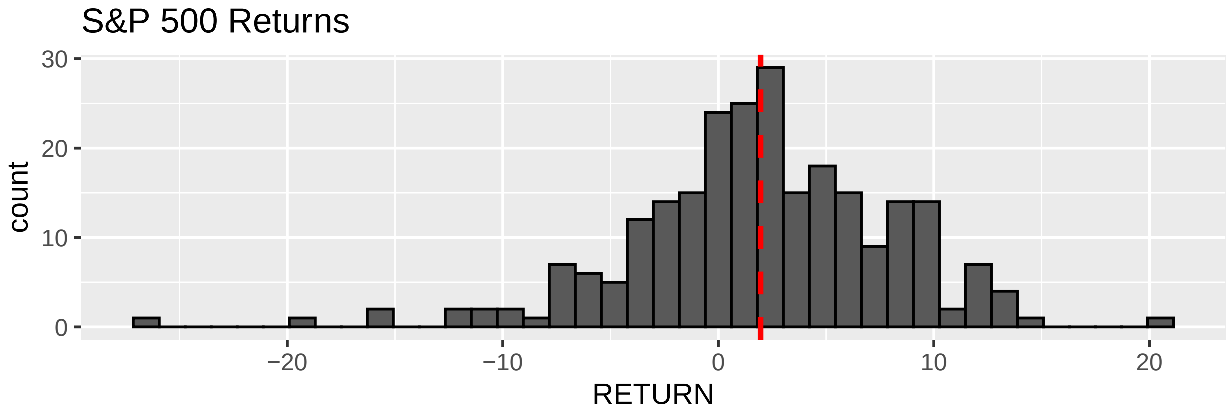
```r
skewness(qgrowth_rates$RETURN)
```

```
## [1] -0.6840202
```

From this summary, we can see that GRGDP mean is closer to zero than RETURN mean which is closer to 2. The range for GRGDP is [-2.7, 3.93] which is much smaller than the range for RETURN [-26.9, 20.1]. The skewness for both are negative so their distribution is skewed towards the left.

```r
x <- ggplot(qgrowth_rates, aes(x=GRGDP)) + geom_histogram(color = "black", bins=40) + geom_vline(aes(xi

y <- ggplot(qgrowth_rates, aes(x=RETURN)) + geom_histogram(color = "black", bins=40) + geom_vline(aes(x

plot_grid(x , y, ncol = 1)
```






```r
cor.test(qgrowth_rates$GRGDP,qgrowth_rates$RETURN)
```

```
##
##  Pearson's product-moment correlation
##
## data:  qgrowth_rates$GRGDP and qgrowth_rates$RETURN
## t = 4.4024, df = 246, p-value = 1.597e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1507504 0.3819521
## sample estimates:
##       cor
## 0.2702427
```

The contemporaneous sample correlation is **0.270** . The positive correlation represents similar movements between U.S. GDP Growth Rate and S&P 500 Returns.

Problem 2.3:

Convert the data to time series data.

```r
gdp = ts(qgrowth_rates$GRGDP,start=1950.25, freq=4)
returns = ts(qgrowth_rates$RETURN,start=1950.25, freq=4)
```

Run OLS for Following Models:

Model a.

```r
olsa <- lm(gdp ~ returns)
summary(olsa)
```

```
##
## Call:
## lm(formula = gdp ~ returns)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5068 -0.5271 -0.0500  0.5344  3.2189
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.710672   0.062819  11.313  < 2e-16 ***
## returns     0.043388   0.009856   4.402  1.6e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9412 on 246 degrees of freedom
## Multiple R-squared:  0.07303,    Adjusted R-squared:  0.06926
## F-statistic: 19.38 on 1 and 246 DF,  p-value: 1.597e-05
```

```r
r2a <- summary(olsa)$r.squared
ar2a <- summary(olsa)$adj.r.squared
data.frame(rSquared = r2a, adjRsquared = ar2a)
```

```
##     rSquared adjRsquared
## 1 0.07303114  0.06926298
```

Model b.

```r
olsb <- dynlm(gdp ~ L(returns,1))
summary(olsb)
```

```
##
## Time series regression with "ts" data:
## Start = 1950(3), End = 2012(1)
##
## Call:
## dynlm(formula = gdp ~ L(returns, 1))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.0231 -0.5339 -0.0346  0.4720  3.5844
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.661818   0.059134  11.192  < 2e-16 ***
## L(returns, 1) 0.064728   0.009305   6.956 3.17e-11 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8855 on 245 degrees of freedom
## Multiple R-squared:  0.1649, Adjusted R-squared:  0.1615
## F-statistic: 48.39 on 1 and 245 DF,  p-value: 3.169e-11
```

```
r2b <- summary(olsb)$r.squared
ar2b <- summary(olsb)$adj.r.squared
data.frame(rSquared = r2b, adjRsquared = ar2b)
```

```
##    rSquared adjRsquared
## 1 0.1649293   0.1615209
```

Model c.

```
olsc <- dynlm(gdp ~ L(returns, 1) + L(returns, 2) + L(returns, 3) + L(returns, 4))
summary(olsc)
```

```
##
## Time series regression with "ts" data:
## Start = 1951(2), End = 2012(1)
##
## Call:
## dynlm(formula = gdp ~ L(returns, 1) + L(returns, 2) + L(returns,
##     3) + L(returns, 4))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9786 -0.5113  0.0068  0.4842  3.7592
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.571948   0.061504   9.299  < 2e-16 ***
## L(returns, 1) 0.056594   0.009735   5.814 1.95e-08 ***
## L(returns, 2) 0.018011   0.010471   1.720   0.0867 .
## L(returns, 3) 0.015672   0.010493   1.494   0.1366
## L(returns, 4) 0.011948   0.009719   1.229   0.2201
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8526 on 239 degrees of freedom
## Multiple R-squared:  0.2066, Adjusted R-squared:  0.1933
## F-statistic: 15.56 on 4 and 239 DF,  p-value: 2.509e-11
```

```
r2c <- summary(olsc)$r.squared
ar2c <- summary(olsc)$adj.r.squared
data.frame(rSquared = r2c, adjRsquared = ar2c)
```

```
##    rSquared adjRsquared
## 1 0.2065937   0.1933149
```

Model d.

```
olsd <- dynlm(gdp ~ L(returns, 1) + L(returns, 2) + L(returns, 3) + L(returns, 4) + L(gdp, 1))
summary(olsd)
```

```
##
## Time series regression with "ts" data:
```

```
## Start = 1951(2), End = 2012(1)
##
## Call:
## dynlm(formula = gdp ~ L(returns, 1) + L(returns, 2) + L(returns,
##     3) + L(returns, 4) + L(gdp, 1))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8733 -0.4469 -0.0007  0.5037  3.7168
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.444490   0.069652   6.382 9.07e-10 ***
## L(returns, 1) 0.050535   0.009646   5.239 3.56e-07 ***
## L(returns, 2) 0.007459   0.010629   0.702 0.483503
## L(returns, 3) 0.011149   0.010316   1.081 0.280918
## L(returns, 4) 0.007133   0.009577   0.745 0.457103
## L(gdp, 1)     0.230396   0.063898   3.606 0.000379 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.832 on 238 degrees of freedom
## Multiple R-squared:  0.2477, Adjusted R-squared:  0.2319
## F-statistic: 15.67 on 5 and 238 DF,  p-value: 2.483e-13
```

```r
r2d <- summary(olsd)$r.squared
ar2d <- summary(olsd)$adj.r.squared
data.frame(rSquared = r2d, adjRsquared = ar2d)
```

```
##    rSquared adjRsquared
## 1 0.2476899    0.231885
```

We will run AIC and BIC tests for model selection

```r
aic23 <- AIC(olsa, olsb, olsc, olsd)
```

```
## Warning in AIC.default(olsa, olsb, olsc, olsd): models are not all fitted to the
## same number of observations
```

```r
bic23 <- BIC(olsa, olsb, olsc, olsd)
```

```
## Warning in BIC.default(olsa, olsb, olsc, olsd): models are not all fitted to the
## same number of observations
```

```r
data.frame(aic23, bic23)
```

```
##      df      AIC df.1      BIC
## olsa  3 677.7493    3 688.2896
## olsb  3 644.8691    3 655.3973
## olsc  6 621.5932    6 642.5762
## olsd  7 610.6156    7 635.0958
```

Model d is the best model. This is seen from the data table of AIC and BIC where model d has the lowest values. Also Model d has the higher R Squared and Adj R Squared. When a model has multiple regression, we will look at the Adjusted R Squared for goodness of fit.

Problem 2.7:

```
unemp_pov <- read.xlsx("Chapter2_exercises_data.xlsx", sheet = 2, detectDates= TRUE)[ ,1:3]
poverty = ts(unemp_pov$POV,start=1959, freq=1)
unemployment = ts(unemp_pov$UNEM, start=1959, freq=1)

mean_pov <- diff(poverty) / (poverty)[-length(poverty)] *100
summary(mean_pov)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -14.0877  -3.0790  -0.5596   0.4489   4.6262  12.2737
```

```
describe(mean_pov)
```

```
##    vars  n mean   sd median trimmed  mad    min   max range skew kurtosis   se
## X1    1 51 0.45 5.38  -0.56    0.39 5.55 -14.09 12.27 26.36 0.03    -0.23 0.75
```

```
mean_unem <- diff(unemployment) /(unemployment)[-length(unemployment)] * 100
summary(mean_unem)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -20.245  -7.125  -2.376   4.004   9.735  59.770
```

```
describe(mean_unem)
```

```
##    vars  n mean    sd median trimmed  mad    min   max range skew kurtosis   se
## X1    1 51    4 17.39  -2.38    1.54 9.21 -20.24 59.77 80.01 1.38     1.56 2.44
```

```
cor.test(mean_pov, mean_unem)
```

```
##
##  Pearson's product-moment correlation
##
## data:  mean_pov and mean_unem
## t = 7.0632, df = 49, p-value = 5.296e-09
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.5404862 0.8244756
## sample estimates:
##      cor
## 0.710275
```

The correlation coefficient for growth rates of unemployed persons and number of people in poverty is **0.710275**. This correlation coefficient is a large positive so this means that when unemployed persons increase, the number of people in poverty increase (vice versa).

The growth rate of unemployed persons has a larger range than people in poverty. Which results in a higher mean as well. This is a result of a possible recession where theres many people unemployed, but not as many go into poverty.

Problem 2.8

a.

```
olsPovA <- dynlm(mean_pov ~ L(mean_unem,1))
r28a <- summary(olsPovA)$r.squared
```

```r
ar28a <- summary(olsPovA)$adj.r.squared
data.frame(rSquared = r28a, adjRsquared = ar28a)
```

```
##    rSquared adjRsquared
## 1 0.1267487   0.1085559
```

b.

```r
olsPovB <- dynlm(mean_pov ~ L(mean_unem, 1) + L(mean_unem, 2) + L(mean_unem, 3) + L(mean_unem, 4) )
r28b <- summary(olsPovB)$r.squared
ar28b <- summary(olsPovB)$adj.r.squared
data.frame(rSquared = r28b, adjRsquared = ar28b)
```

```
##    rSquared adjRsquared
## 1 0.1596055  0.07956795
```

c.

```r
olsPovC <- dynlm(mean_pov ~ L(mean_unem, 1) + L(mean_unem, 2) + L(mean_unem, 3) + L(mean_unem, 4) + L(me
r28c <- summary(olsPovC)$r.squared
ar28c <- summary(olsPovC)$adj.r.squared
data.frame(rSquared = r28c, adjRsquared = ar28c)
```

```
##    rSquared adjRsquared
## 1 0.3694377   0.2925399
```

Following the similar models as in 2.3. I used the models b-d from 2.3 to investigate whether changes in unemployment is one of the causes of changes in poverty.

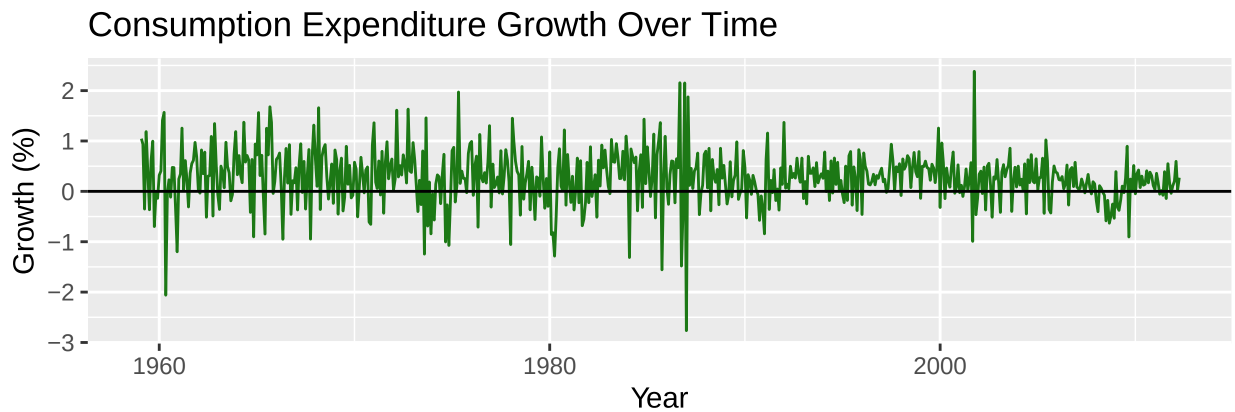
Examining Adjusted R Squared again for goodness of fit test for multiple regression, we prefer **Model c** to explain the data because it has a higher Adjusted R Squared.

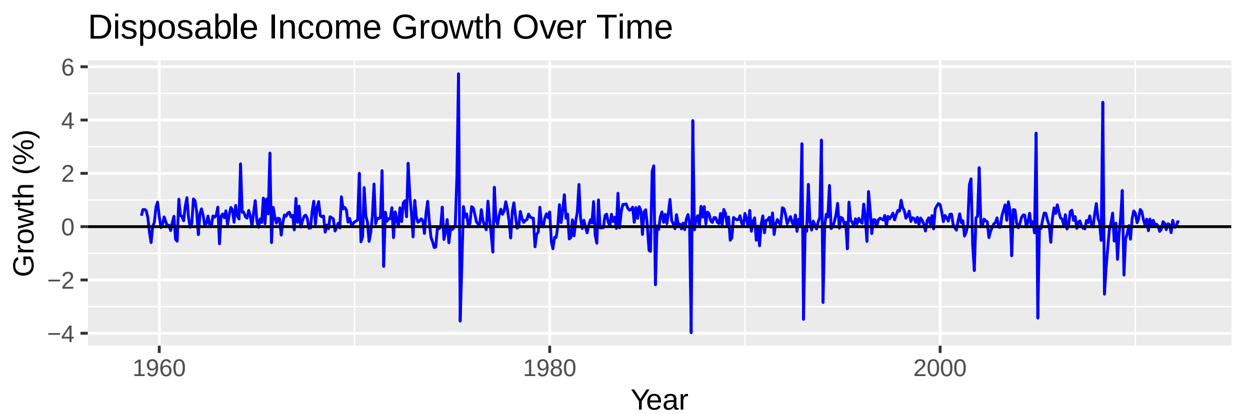
Problem 3.1:

```r
fred_mon <- read.xlsx("Chapter3_exercises_data.xlsx", sheet=1, detectDates=TRUE)[,1:3]
rpce = ts(fred_mon$rpce,start=1959-01-01, freq=12)
rdpi = ts(fred_mon$rdpi, start=1959-01-01, freq=12)
```

a. Calc and plot growth rates of expenditures and income

```r
fred_mon['rpce_lag'] <- lag(fred_mon$rpce,k=1)
fred_mon['rpce_growth'] <- 100 * (log(fred_mon$rpce) - log(fred_mon$rpce_lag) )


fred_mon['rdpi_lag'] <- lag(fred_mon$rdpi, k=1)
fred_mon['rdpi_growth'] <- 100 * (log(fred_mon$rdpi) - log(fred_mon$rdpi_lag))

plot_rpce <- ggplot(fred_mon) +geom_line(aes(x = date, y = rpce_growth), color = '#1C7815', na.rm = TRUE

plot_rdpi <- ggplot(fred_mon) + geom_line(aes(x = date, y = rdpi_growth), color = 'blue', na.rm = TRUE)

plot_grid(plot_rpce, plot_rdpi, ncol = 1)
```





The growth rate of consumption has lower volatility [+- 2 %] than the growth rate of personal income [+- 4 %]. This happens because of permanent income hypothesis which states that consumers will spend money at a level consistent with their expected long-term average income. So a more volatility in income will cause a lesser volatility in consumption expenditure.

b.

```
regress_growth <- lm(rpce_growth ~ rdpi_growth, data = fred_mon)
summary(regress_growth)
```

```
##
## Call:
## lm(formula = rpce_growth ~ rdpi_growth, data = fred_mon)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.04050 -0.29792  0.01606  0.30383  2.44504
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.22543    0.02242  10.056  < 2e-16 ***
## rdpi_growth  0.17452    0.02920   5.976  3.8e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5317 on 637 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.05309,    Adjusted R-squared:  0.05161
```

```
## F-statistic: 35.72 on 1 and 637 DF,  p-value: 3.799e-09
```

When we regress consumption growth on disposble income growth, both estimates of the intercept and the coefficient for disposable income growth are statistically significant. The R-Squared is 0.053 and the adjusted R-Squared is 0.051, which mean that about 5.1% of the total sample variation of consumption growth is explained by disposable income growth. This is a low value so this regression is not a good fit. The coefficient for disposable income growth is 0.17 meaning that a 1% increase in disposable income growth results in a 0.17% growth in consumption. Because 0.17% is less than 1%, this aligns with the permanent income hypthesis.

c.

```
rpce_growth = ts(fred_mon$rpce_growth, start=1959-01-01, freq=12)
rdpi_growth = ts(fred_mon$rdpi_growth, start=1959-01-01, freq=12)
regress_growth_lag <- dynlm(rpce_growth ~ rdpi_growth + L(rdpi_growth,1))
summary(regress_growth_lag)
```

```
##
## Time series regression with "ts" data:
## Start = 1957(3), End = 2010(4)
##
## Call:
## dynlm(formula = rpce_growth ~ rdpi_growth + L(rdpi_growth, 1))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.00818 -0.28874 -0.00051  0.29768  2.55088
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.19889    0.02405   8.269 7.90e-16 ***
## rdpi_growth        0.18722    0.02939   6.371 3.61e-10 ***
## L(rdpi_growth, 1)  0.08284    0.02939   2.819  0.00497 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5284 on 635 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.06476,    Adjusted R-squared:  0.06182
## F-statistic: 21.99 on 2 and 635 DF,  p-value: 5.855e-10
```

Adding a lag to the regression in b. For the added lagged coefficient for disposable income growth it is 0.08 so a 1% increase in disposable income growth results in a 0.08% growth in consumption. The adjusted R Squared increases slightly to 0.061 but still remains small. The intercept and the personal income growth remain statistically significant, but the coefficient for the lagged personal income growth rate just slightly passes the 95% significance level.

Problem 3.3:

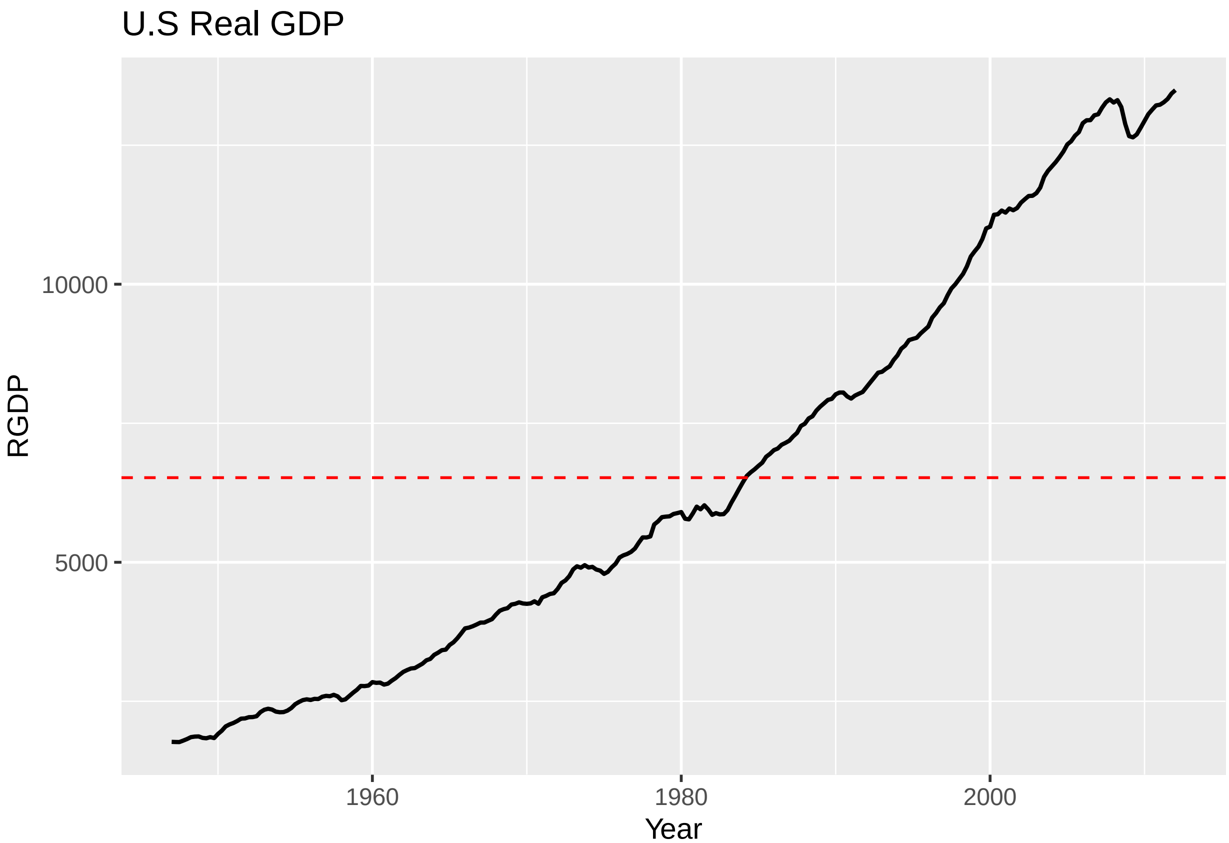
Load Data

```
usgdp <- read.xlsx("Chapter3_exercises_data.xlsx", sheet = 3, detectDates = TRUE)[ ,1:2]
exYen <- read.xlsx("Chapter3_exercises_data.xlsx", sheet = 4, detectDates = TRUE)[ ,1:2]
UStreasYield <- read.xlsx("Chapter3_exercises_data.xlsx", sheet = 5, detectDates = TRUE)[ ,1:2]
UnempRate <- read.xlsx("Chapter3_exercises_data.xlsx", sheet = 6, detectDates = TRUE)[ ,1:2]
```

Plot Time Series , Red lines indicate mean

a. U.S Real GDP

```
usgdp_mean <- mean(usgdp$rgdp)
ggplot(usgdp, aes(date, rgdp)) + geom_line(color= 'black', lwd = 0.75, na.rm = TRUE) + geom_hline(yinte
```



Definiton: Real GDP is the inflation adjusted value of goods and services produced in the US.

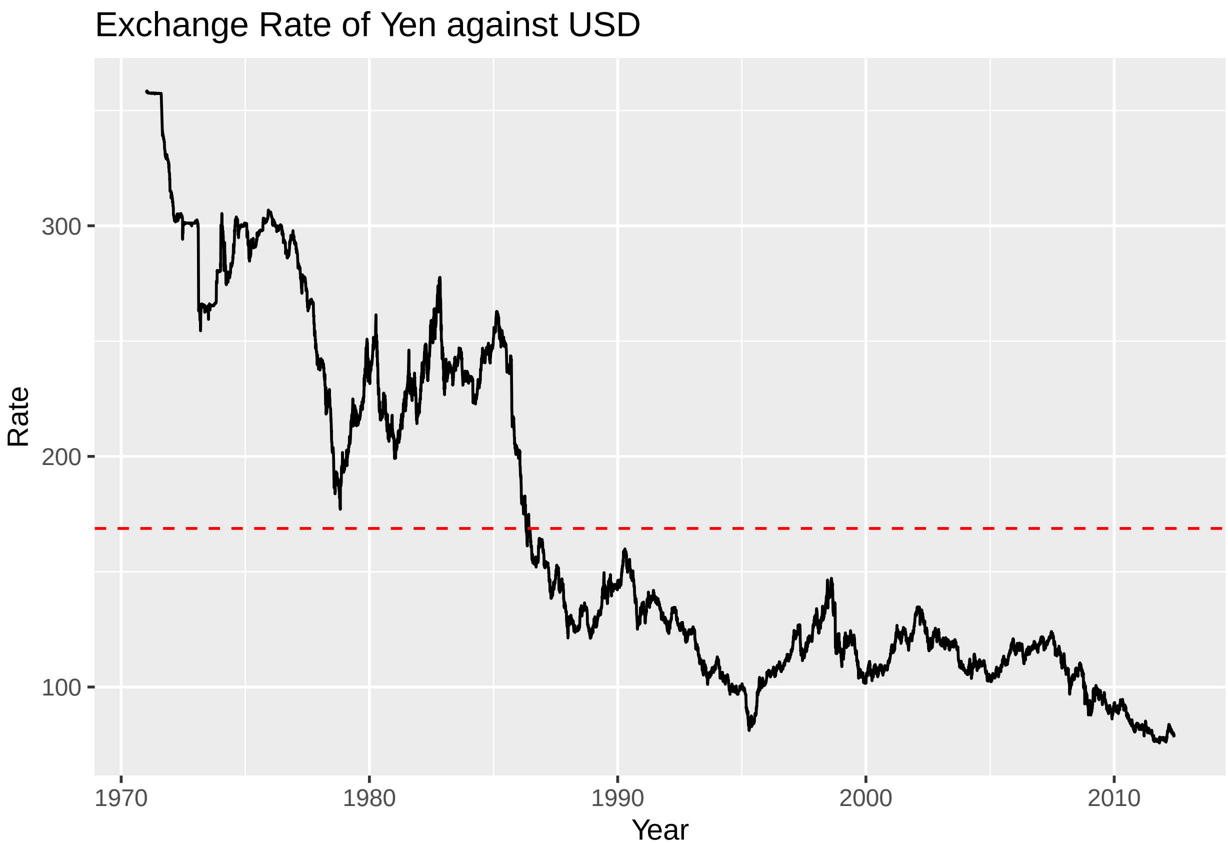
Periodicity: Quarterly, 1947Q1 - 2012Q1

Units: Billions Chained Weighted in US Dollars

Stationary: Plot exhibits upward trend with occasional local dips and peaks. This is not a first or second order stationary.

b. The Exchange Rate of the Japanese Yen against the U.S Dollar

```
exYen_mean <- mean(exYen$jpy_usd)
ggplot(exYen, aes(DATE, jpy_usd)) + geom_line(color= 'black', lwd = 0.5, na.rm = TRUE) + geom_hline(yin
```



Definiton: Japan/U.S foreign exchange rate for the value of Yen that equals 1 USD

Periodicity: Daily, 1971-01-04 to 2012-06-01

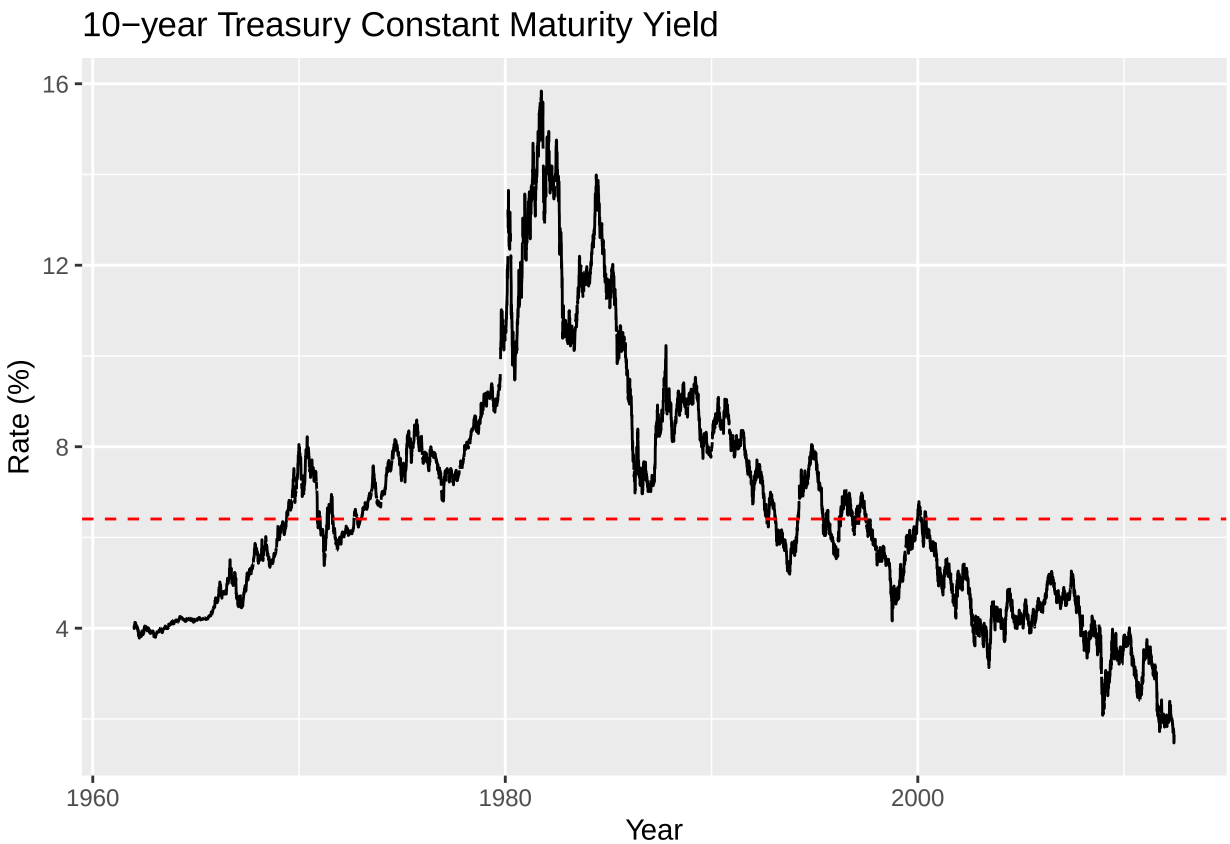
Units: Japanese Yen to 1 USD

Stationary: Downward trend in the plot with small and moderate local dips and peaks. This is not first order stationary.

c. The 10-year U.S Treasury Constant Maturity Yield

```
UStreasYield_mean <- mean(UStreasYield$CMRate10Yr)
UStreasYield[UStreasYield==0] <- NA

ggplot(UStreasYield, aes(DATE, CMRate10Yr)) + geom_line(color= 'black', lwd = 0.5) + geom_hline(yinterce
```

```
## Warning: Removed 1 row(s) containing missing values (geom_path).
```



Definiton: Yields on traded non-inflation indexed issues adjusted to constant maturities

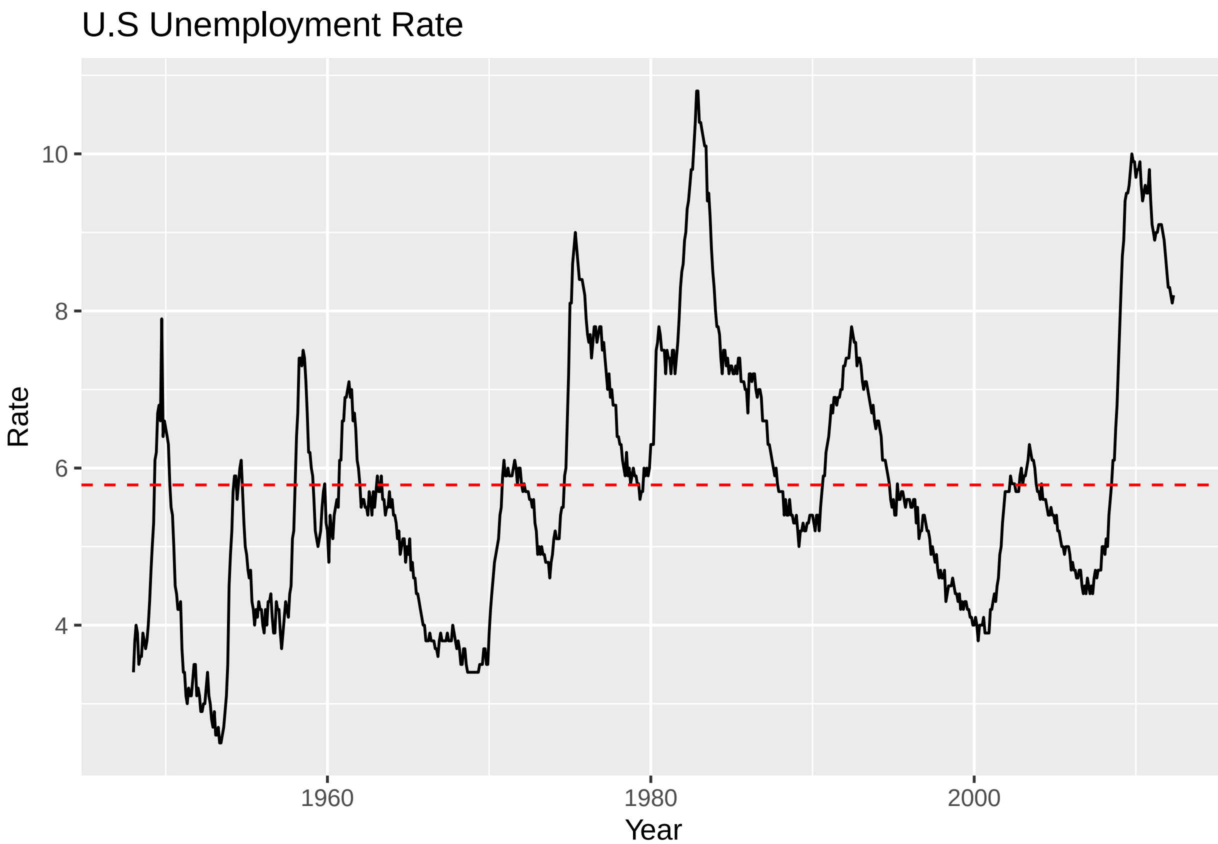
Periodicity: Daily, 1962-01-02 to 2012-06-07

Units: Rate in Percent

Stationary: Before mid 1980s there is an upward trend then a downward trend. The sample mean here again does not play a strong role in describing the centrality. And appears to not have a constant variance. So it is not clear to distinguish it a first order weakly stationary.

d. The U.S Unemployment Rate

```
UnempRate_mean <- mean(UnempRate$unemrate)
ggplot(UnempRate, aes(DATE, unemrate)) + geom_line(color= 'black', lwd = 0.5, na.rm = TRUE) + geom_hlin
```



Definiton: The percent of unemployed people over the labor force. The US Labor force includes people 16 years of age and older, not in institutions, not on active military duty, residing in the United States.

Periodicity: Monthly, 1948-01-01 to 2012-05-01

Units: Rate in Percent

Stationary: Different from the other three where it has persistent dips and peaks around the same area, with a sloght upward trend. It seems to be more stationary than the other three, but it is hard to be confirm it is first order weakly stationary. The variances do not seem constant so I am confident to say it is not second order weakly stationary.

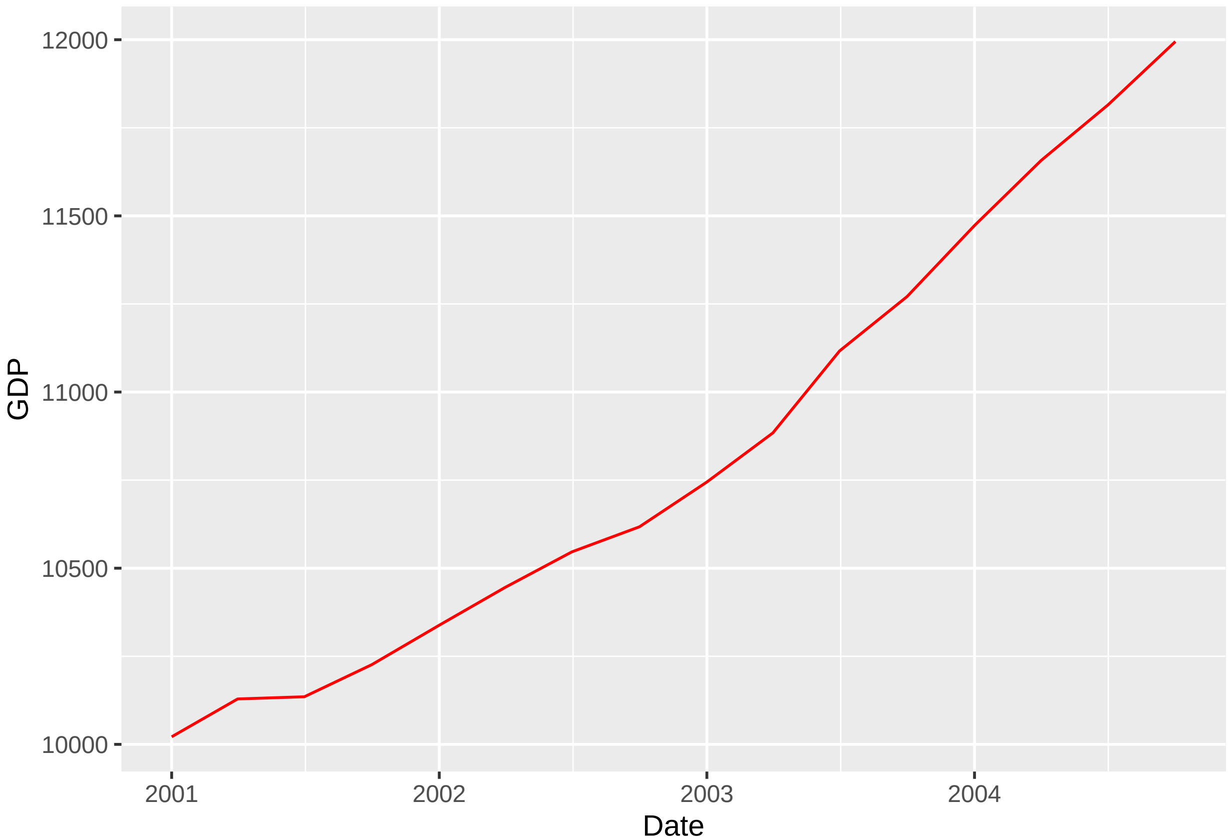
Problem 3.5:

Load Date

```
nUSGDP <- read.xlsx("Chapter3_exercises_data.xlsx", sheet = 7, detectDates=TRUE)[ , 1:2]
```

a.

```
ggplot(nUSGDP, aes(Date, GDP) ) + geom_line(color = "red", lwd = 0.5)
```



The plot shows an upward trend that is no weakly stationary. The trend indicates that it has different means in different periods of time.

b.

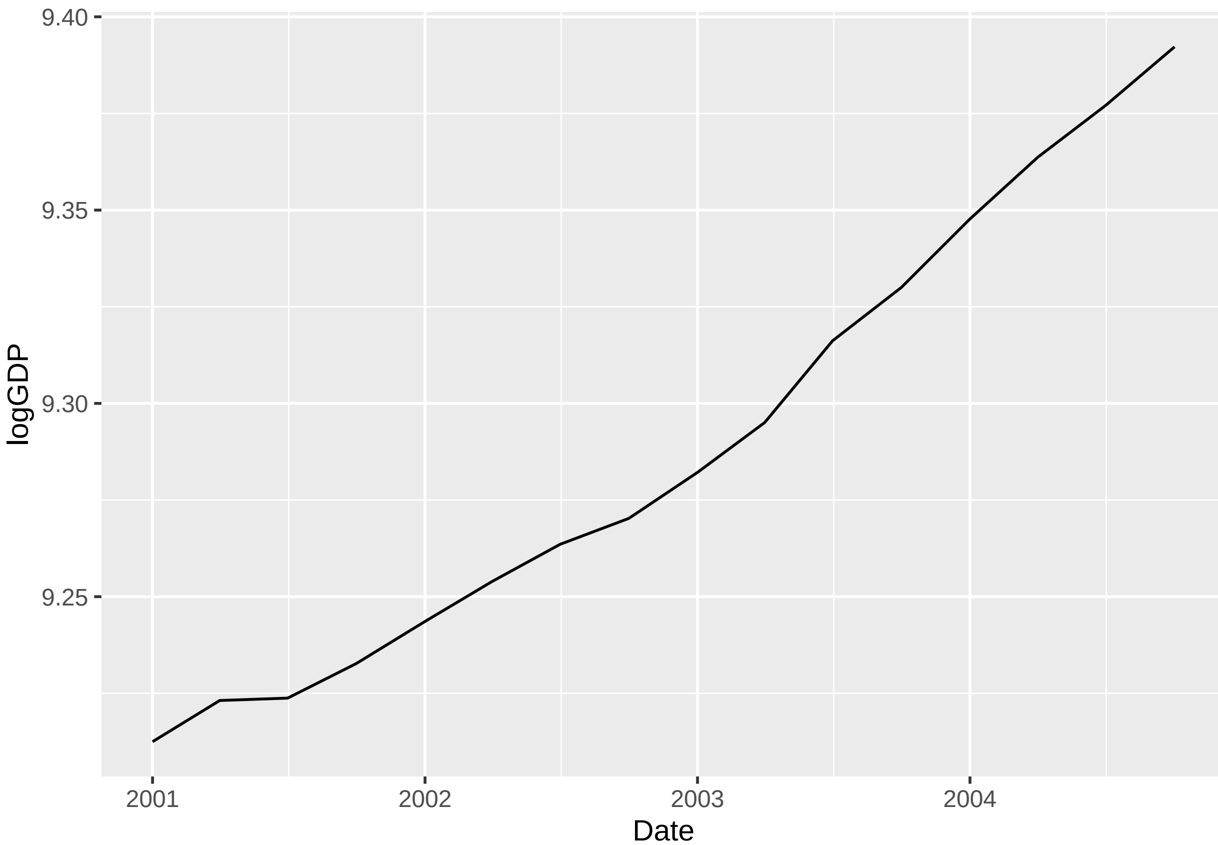
Growth rates of nominal GDP.

```
nUSGDP['gdp_lag'] <- lag(nUSGDP$GDP, k=1)
#nUSGDP = ts(nUSGDP, start=2001-01-01, freq=4)
nUSGDP['gdp_growth'] <- 100*(nUSGDP$GDP - nUSGDP$gdp_lag) / nUSGDP$gdp_lag
nUSGDP$gdp_growth
```

```
##  [1]         NA 1.07169585 0.06121099 0.89984312 1.09423741 1.03983285
##  [7] 0.96499038 0.67320912 1.19708029 1.29739590 2.13800074 1.38710229
## [13] 1.78956428 1.61166606 1.35020373 1.52265360
```

c.

```
nUSGDP['logGDP'] <- log(nUSGDP$GDP)

ggplot(nUSGDP, aes(Date, logGDP) ) + geom_line(color = "black", lwd = 0.5)
```



The logarithmic transformation helps stabilize the variance, but in the plot shows that it does not affect the trending behavior of the series. It is still not first order stationary, but it is smoother than the plot in a.

d.

First log-differences.

```
nUSGDP['lag_logGDP'] <- lag(nUSGDP$logGDP, k=1)
nUSGDP['gdp_growth2'] <- (nUSGDP$logGDP - nUSGDP$lag_logGDP) * 100
nUSGDP$gdp_growth2
```

```
##  [1]         NA 1.06599390 0.06119226 0.89581866 1.08829395 1.03446378
##  [7] 0.96036409 0.67095319 1.18997196 1.28905181 2.11546613 1.37757007
## [13] 1.77374009 1.59881660 1.34116971 1.51117758
```

e.

There is no significant difference between the growth rates of nominal GDP and the first log-differences computued in ii. and iv.