Name: Trần Thanh Phúc

Student ID: K194141740

PROGRAM PACKAGE IN FINANCE 2

LAST-TERM TEST

TOPIC: Leverage (total debt/total assets)

1. **Literature review:**
    a. **Leverage**

Frank and Goyal (2009) used the financial leverage variable like total debt to total assets and total debt to equity. Pandey (2002) used only the ratio of total debt to total assets at book value to identify the firms' financial leverage. In some of the other studies such as by Sheikh and Wang (2011), used the total liabilities ratio as the main measure of financial leverage. Frank and Goyal (2009) still suggested the use of the market value of debt as a suitable measurement for financial leverage as it reflects in a better way the current position and future position of the firms. However, there are many other studies using accounting measures to decide financial leverage like studies of Chen (2003), Keister (2004), Bhabra, Liu, and Tirtiroglu (2008), Frank and Goyal (2009), Hossain (2012), Dao and Lai (2018). As a result, this study used book value instead of market value to assess the impact of determinants on firms' financial leverage. The study has used the following measure as a proxy of firms' financial leverage.

Leverage = Total Debts / Total Assets

    b. **ROA and Leverage of the company**

Naveed, Zulfqar, and Ishfaq (2010) indicated the negative relationship between profitability and financial leverage. The study confirmed that Pakistan Insurance companies prefer to use internal financing rather than debts. The more profitable the firms are, the more internal financing they are going to have. These findings were also supported by the research results of Harris, and Raviv (1991), Booth, Aivazian, Demirguc-Kunt, and Maksimovic (2001), and Kajananthan (2012). On the other hand, Dilek, Ozlem, and

Ayca (2009) found that there is a positive relationship between profitability and debt ratios. They explained that firms with higher profitability can easily attract more debts from banks as well as from the financial market. Moreover, these firms also prefer debts in order to decrease tax charges. All studies measure profitability as earnings after interest and taxes divided by total assets.

Theoretically, as the degree of FL increases, the ratio of paying debts of the company in comparison with total equity also increases, which directly affects and results on a rise in financial costs occurred during the period. If the increase in profit is not enough to offset the corresponding borrowing costs, ROA will decrease. Empirically, Zeitun et al. (2007); Muritala (20 12); Sheikh and Wang (2013), and Pouraghajan et al.(2012) concluded that FL had a negative impact on ROA.

ROA = Net Income / Current Assets

c. Asset Structure ratio and Leverage of the company

Asset structure or tangibility is measured by the ratio of fixed assets to total assets. Many studies have proven that there is an influence of asset structure on financial leverage. Myers (1984), Bradley, Jarrell, and Kim (1984) and Daskalakis and Psillaki (2008) found that the firms with higher tangibility are often easier to access debts because of the availability of collateral. Their results showed the positive relationship between asset structure and debts. In contrast to these findings, Sheikh and Wang (2011) found that there is still a negative relationship between asset structures with debts of Pakistani firms.

Asset Structure = Fixed Assets / Current Assets

d. Firm Size and Leverage of the company

The assessment of firm size is performed by considering two benchmarks: the number of employees (Bas et al., 2009) or the sales logarithm (Kayo and Kimura, 2011). In the earliest research (Rajan and Zingales, 1995), the explanation of the positive relationship between leverage and firm size was based on the financial distress costs, recognising that larger firms face lower bankruptcy costs and have more diversified portfolios (with a lower bankruptcy probability)

As regards the theoretical grounding, there are common views on the relationship between firm size and leverage. Trade-off theory admits that firm size has a positive capitalisation effect. In the same spirit, pecking order theory postulates that since large firms are more diversified and have less volatile earnings, problems arising due to information asymmetry can be mitigated. The generally accepted assumption is that, in the case of large firms, firm size is positively correlated with leverage Accepting the previously formulated hypothesis, which states that large firms may have a greater debt carrying capacity, Byoun (2008) argues that large firms, being generally more transparent, tend to have higher levels of leverage and the diversification of leverage alternatives may lead to savings in debt issuance costs. In contrast, for smaller firms, financial institutions must allocate more resources to monitor and are likely to "punish" them, by imposing higher interest rates (Alves Pereira and Ferreira, 2011).

Some findings indicated that there is a positive relation between the firm size and the financial leverage. While some others considered that there is a negative relation. Some other findings even suggested it as a statistically insignificant relationship. According to Rajan and Zingales (1995), Booth, Aivazian, Demirguc-Kunt, and Maksimovic (2001), firm size is positively correlated with firms' financial leverage. The reason can be given as the fact that larger firms can easily approach debt at lower costs than smaller firms. A study of Hossain (2012) revealed a negative relationship between firm size and financial leverage. While, findings of Dilek, Ozlem, and Ayca (2009) reported that the impact of firm size on financial leverage is insignificant. However, all these studies still used the natural logarithm of the total asset as a proxy for firm size.

$$\text{Firm Size} = \ln(\text{Total Asset})$$

e. Liquidity

The liquidity of a firm is measured by the ratio of total assets to short term debts at year-end. Firms with large current assets often prefer to use internal funds to finance business activities instead of using external sources. Therefore, these firms often use fewer debts, and the study results revealed that there is a negative relationship between firms' liquidity and financial leverage (Myers, 1984). It is right with the identification of the pecking order theory. However, the trade-off theory suggested that there is a positive relationship between liquidity and the ability of a firm to pay a debt obligation on time. It means that higher liquidity expresses firms having the ability to pay their debt (Ozkan, 2002).

$$\text{Current Ratio} = \text{Current Assets} / \text{Current Liabilities}$$

2. Data collection and input

2.1. The reason for choosing the company

The company I chose is PetroVietnam Gas Joint Stock Corporation (PV Gas) with the stock code GAS. Over the years of operation, PV Gas has continuously developed and expanded and is considered as the leading enterprise in the gas industry in Vietnam. PV GAS brand has become familiar, becoming a strategic energy supplier for almost all households in electricity, fertilizer, steel and many industries, transportation, food, commercial and civil. with reputation and guaranteed quality. For many years in a row, PV GAS has always exceeded production and business plans and the obligation to collect and pay the State budget. In 2020, PV Gas achieved revenue of 64,150 billion VND; profit after tax reached 7,928 billion dong. In the coming time, PV Gas strives to maintain 100% of the dry gas market share and 50-60% of the national LPG market share. This shows that PV Gas greatly affects the gas industry and the stability of the Vietnamese economy. That's why I chose PV Gas for this final exam.

2.2 Process of getting and inputing data:

I have taken PV Gas' data from VietstockFinance website (finance.vietstock.vn) - A system of securities investment tools, a macro-financial-stock database developed under the motto "comprehensive platform economic - all business angles". Not only providing investors with the big picture, but also serving closely the unique needs of each individual. VietstockFinance is confident to be a valuable source of reference information with two versions in Vietnamese and English.

For the variables that affect leverage that I mentioned above, only ROA is available for data collection. Asset Structure, Firm Size, Current Ratio have no data available. Therefore, I am forced to take the data of the variables from the formula to calculate them. The variables that I get include Fix Assets, Current Assets, Total Assets, Current Liabilities, Total Debts, ROA. According to the request, the data I retrieved includes 42 quarters, from the fourth quarter of 2011 to the first quarter of 2022 and saved them to an excel file with the name is my student ID.

I have attached the data to this report or you can also access the data at the following link: https://docs.google.com/spreadsheets/d/1QGKv8YwVVgs-ybvUvo0LlokRRjnAYLjI/edit?usp=sharing&ouid=115863812061934169103&rtpof=true&sd=true

Then, I run a R programming IDE and I create a R script file to input the data into its.

I input the necessary libraries for use later in the exercise.

```
# Import library
library(readxl)
library(ggplot2) # for plotting
library(tidyverse)  # for dataframe manipulation
#install.packages("zoo")  # Install zoo package
library("zoo")  # dealing with missing values
#install.packages("SciViews")  # Install SciViews package
library("SciViews")
#install.packages("cowplot")  Install cowplot package
library("cowplot")
#install.packages("Hmisc")
library("Hmisc")
```

I input the data that I collected.

```
# Get data
data = read_excel('K194141740.xlsx', sheet = 'Data')
View(data)
```

	Date	FixedAssets	CurrentAssets	TotalAssets	CurrentLiabilites	TotalDebts	ROA
1	2011-12-31	17643133	28734335	54877735	12268850	28185322	16.51
2	2012-03-31	21190442	26601437	52252871	14382938	24728434	4.71
3	2012-06-30	21140793	25301160	50694528	13911067	23294864	4.71
4	2012-09-30	20693403	25071807	49737343	12359596	21765191	4.95
5	2012-12-31	20987959	20447784	45108421	7896823	16353385	5.27
6	2013-03-31	20634120	23936352	48101969	8251011	14937904	9.21
7	2013-06-30	20029617	25874986	49122286	9246511	15133789	5.98
8	2013-09-30	19631572	25938889	48356717	7211852	13292155	6.13
9	2013-12-31	19944853	28261081	50503546	9803611	15369123	4.44
10	2014-03-31	19083362	30397021	51487899	8721068	14384828	6.20
11	2014-06-30	18490906	29033428	49376259	7814424	12428786	6.41
12	2014-09-30	18839796	32563895	53208320	10571263	16098037	5.75
13	2014-12-31	19009760	33259312	53762672	10560575	16079060	8.94
14	2015-03-31	14402601	32882300	52968690	6847992	12595417	4.84
15	2015-06-30	13739187	31780593	52992179	6872786	11460704	4.88
16	2015-09-30	13753747	32576939	54878246	5597451	10994954	4.30
17	2015-12-31	14843716	33833057	56724385	8982041	13830394	1.87

I perform data processing such as checking data for blanks and looking at correlations between variables.

```
# Preprocessing data
# Replace missing values by median values
print(paste('There are/is ', sum(is.na(data)), 'missing values'))
for (i in names(data)){
  data[i] = na.locf(data[i])
}
print(paste('There are/is ', sum(is.na(data)), 'missing values')) # check missing values
# Correlation
corr <- rcorr(as.matrix(data[,2:ncol(data)]))
print(corr)
```

We see no missing values in the data.

```
 check misssing values
[1] "There are/is  0 missing values"
```

```
                  FixedAssets CurrentAssets TotalAssets CurrentLiabilites TotalDebts   ROA
FixedAssets              1.00         -0.27       -0.18              0.29       0.39  0.23
CurrentAssets           -0.27          1.00        0.97              0.65       0.33 -0.41
TotalAssets             -0.18          0.97        1.00              0.73       0.49 -0.39
CurrentLiabilites        0.29          0.65        0.73              1.00       0.84 -0.13
TotalDebts               0.39          0.33        0.49              0.84       1.00  0.15
ROA                      0.23         -0.41       -0.39             -0.13       0.15  1.00

n= 42


P
                  FixedAssets CurrentAssets TotalAssets CurrentLiabilites TotalDebts ROA
FixedAssets                   0.0782        0.2529      0.0606            0.0112     0.1509
CurrentAssets     0.0782                    0.0000      0.0000            0.0330     0.0064
TotalAssets       0.2529      0.0000                    0.0000            0.0009     0.0103
CurrentLiabilites 0.0606      0.0000        0.0000                        0.0000     0.4197
TotalDebts        0.0112      0.0330        0.0009      0.0000                       0.3361
ROA               0.1509      0.0064        0.0103      0.4197            0.3361
```

Then I do calculations with the variables in the data to figure out the variables that affect Leverage that I mentioned above and Leverage. I also create a Period column to prepare for categorizing the observations in the following tasks.

```
# Calculate Asset Structure, Firm Size, Current Ratio, Leverage
df = data.frame(data$Date)
colnames(df)[1] = 'Date'
df$ROA = data$ROA
df$AssetStructure = data$FixedAssets/data$CurrentAssets
df$FirmSize = ln(data$TotalAssets)
df$CurrentRatio = data$CurrentAssets/data$CurrentLiabilites
df$Leverage = data$TotalDebts/data$TotalAssets
df = mutate(df, Period = NA)
print(head(df))
```

```
        Date   ROA AssetStructure FirmSize CurrentRatio  Leverage Period
1 2011-12-31 16.51      0.6140087 17.82062     2.342056 0.5136021     NA
2 2012-03-31  4.71      0.7965901 17.77161     1.849513 0.4732455     NA
3 2012-06-30  4.71      0.8355662 17.74133     1.818779 0.4595144     NA
4 2012-09-30  4.95      0.8253654 17.72227     2.028530 0.4376026     NA
5 2012-12-31  5.27      1.0264173 17.62458     2.589368 0.3625351     NA
6 2013-03-31  9.21      0.8620411 17.68883     2.901020 0.3105466     NA
>
```

3. Provide descriptive statistics of all the variables for BEFORE and AFTER periods

I set a value for the Period variable with the condition:

Before 2020, assign the value 'before Covid-19 pandemic'.

Since 2020, assign to the value 'after Covid-19 pandemic'.

```
# Create Period column to verify data before 2020 and after 2020
for (i in 1:nrow(df)){
  if (i <= 33){
    df[i,'Period'] = 'before Covid-19 pandemic'
  } else {
    df[i,'Period'] = 'after Covid-19 pandemic'
  }
}
```

I perform descriptive statistics for the variables.

ROA

```
# Descriptive statistics of ROA
ROA_statistics = df %>%
  group_by(Period) %>%
  summarise(Median = median(ROA),
            Mean = mean(ROA),
            Min = min(ROA),
            Max = max(ROA),
            Standard_Deviation = sd(ROA))
View(ROA_statistics, 'ROA')
```

	Period	Median	Mean	Min	Max	Standard_Deviation
1	after Covid-19 pandemic	3.14	3.148889	2.58	4.18	0.5235323
2	before Covid-19 pandemic	4.84	5.176364	1.64	16.51	2.6039655

It is easy to see that PV Gas' ROA after the Covid-19 epidemic has decreased significantly with the mean value from 5.18 to 3.14 and the standard deviation also decreased.

AssetStructure

```
# Descriptive statistics of AssetStructure
AssetStructure_statistics = df %>%
  group_by(Period) %>%
  summarise(Median = median(AssetStructure),
            Mean = mean(AssetStructure),
            Min = min(AssetStructure),
            Max = max(AssetStructure),
            Standard_Deviation = sd(AssetStructure))
View(AssetStructure_statistics, 'AssetStructure')
```

	Period	Median	Mean	Min	Max	Standard_Deviation
1	after Covid-19 pandemic	0.3537459	0.3758224	0.3142525	0.5024493	0.06084234
2	before Covid-19 pandemic	0.4918572	0.5590057	0.3770714	1.0264173	0.17186490

Similar to the ROA variable, PV Gas' Asset Structure also decreased sharply with the mean value from 0.56 to 0.38 and the standard deviation also decreased.

Firm Size

```
# Descriptive statistics of FirmSize
FirmSize_statistics = df %>%
  group_by(Period) %>%
  summarise(Median = median(FirmSize),
            Mean = mean(FirmSize),
            Min = min(FirmSize),
            Max = max(FirmSize),
            Standard_Deviation = sd(FirmSize))
View(FirmSize_statistics, 'FirmSize')
```

	Period	Median	Mean	Min	Max	Standard_Deviation
1	after Covid-19 pandemic	18.04419	18.07157	17.93786	18.26016	0.1070204
2	before Covid-19 pandemic	17.85371	17.84288	17.62458	18.03507	0.1064481

Unlike ROA and Asset Structure, PV Gas's Firm Size has not been affected much by the Covid 19 pandemic. Firm Size continues to grow and the standard deviation is almost unchanged.

Current Ratio

```
# Descriptive statistics of CurrentRatio
CurrentRatio_statistics = df %>%
  group_by(Period) %>%
  summarise(Median = median(CurrentRatio),
            Mean = mean(CurrentRatio),
            Min = min(CurrentRatio),
            Max = max(CurrentRatio),
            Standard_Deviation = sd(CurrentRatio))
View(CurrentRatio_statistics, 'CurrentRatio')
```

	Period	Median	Mean	Min	Max	Standard_Deviation
1	after Covid-19 pandemic	3.658860	3.544925	2.505744	4.164975	0.5667204
2	before Covid-19 pandemic	3.329722	3.321421	1.818779	5.819960	0.8066590

The Current Ratio has also been largely unaffected by the pandemic. The evidence is that the mean values before and after the pandemic do not change much. However, the post-pandemic standard deviation has decreased significantly from 0.8 to 0.56.

Leverage

```
# Descriptive statistics of Leverage
Leverage_statistics = df %>%
  group_by(Period) %>%
  summarise(Median = median(Leverage),
            Mean = mean(Leverage),
            Min = min(Leverage),
            Max = max(Leverage),
            Standard_Deviation = sd(Leverage))
View(Leverage_statistics, 'Leverage')
```

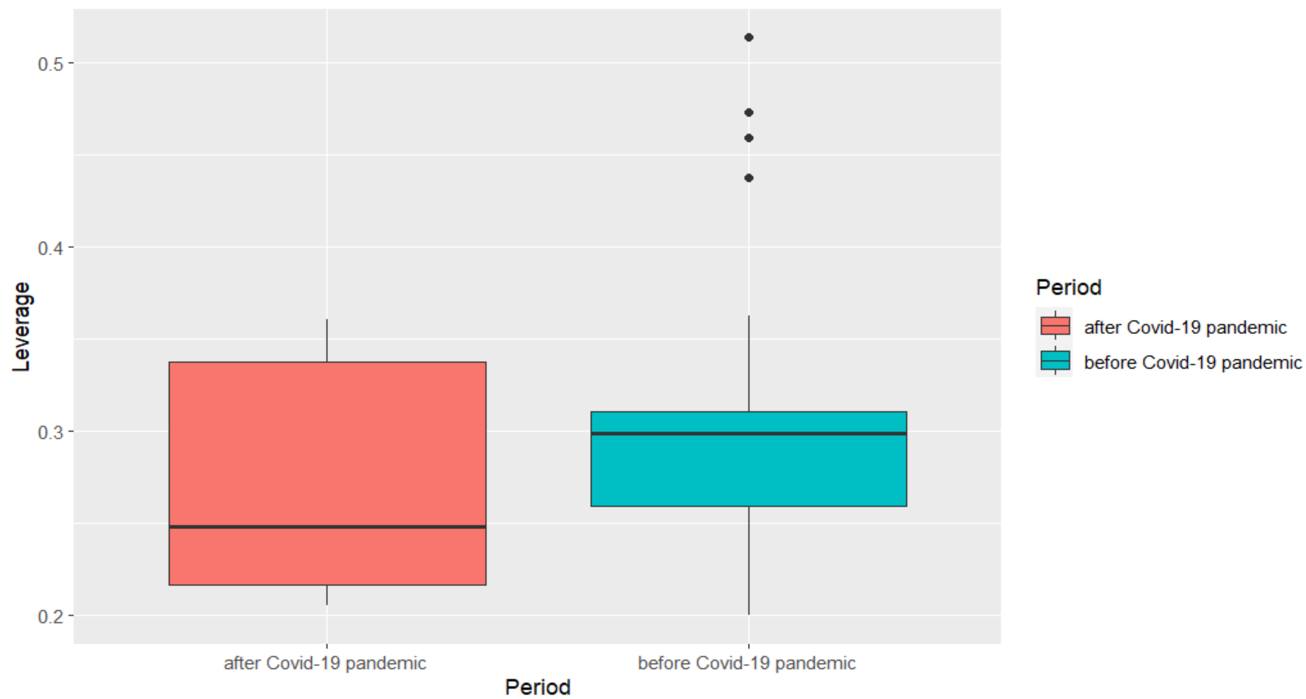
	Period	Median	Mean	Min	Max	Standard_Deviation
1	after Covid-19 pandemic	0.2484431	0.2733879	0.2053124	0.3605829	0.06512061
2	before Covid-19 pandemic	0.2990748	0.3030650	0.2003518	0.5136021	0.07369300

PV Gas's Leverage from 0.3 before the pandemic to 0.27 after the pandemic and the standard deviation also decreased from 0.07 to 0.06. This shows that the company's Leverage has not been affected much by the pandemic.

4. Provide box & whisker plot and histogram of Leverage

Box & Whisker plot of Leverage for the entire period

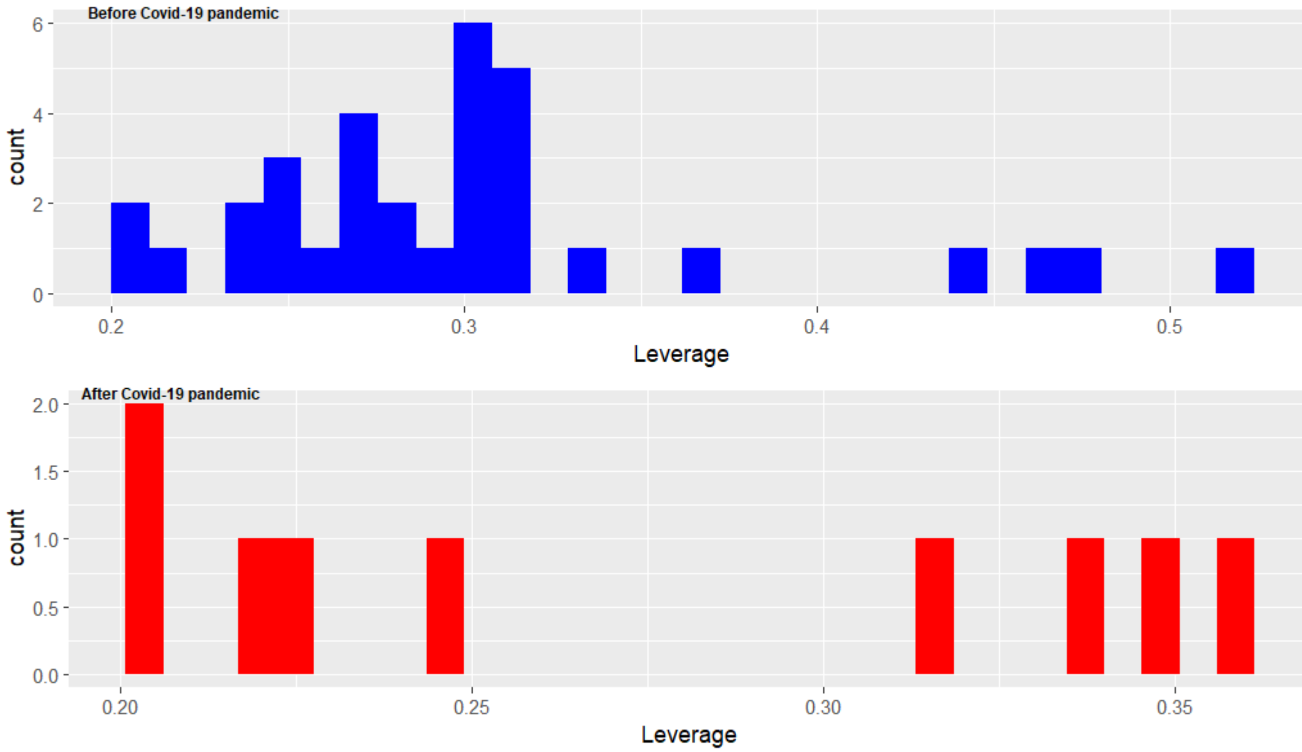
```
# Box & Whisker plot of Leverage for the entire period
df %>%
  ggplot(aes(x = Leverage, y = Period, fill = Period)) +
  geom_boxplot() +
  coord_flip()
```



Looking at the graph, we can see that after the pandemic Leverage has a smaller standard deviation because before the epidemic there are some observations with very high values up to 0.5. In general, Leverage before the pandemic is still higher. This shows that the pandemic has had a negative effect on the company's leverage ratio.

Histogram plot of Leverage for the entire period

```
# Histogram plot of Leverage for the entire period
htg_before = df %>%
  filter(Period == 'before Covid-19 pandemic') %>%
  ggplot(aes(x = Leverage)) +
  geom_histogram(fill = 'blue')
htg_after = df %>%
  filter(Period == 'after Covid-19 pandemic') %>%
  ggplot(aes(x = Leverage)) +
  geom_histogram(fill = 'red')
plot_grid(htg_before, htg_after, labels=c("Before Covid-19 pandemic", "After Covid-19 pandemic"),
          label_size = 7, ncol = 1, nrow = 2)
```



Similar to the previous chart, from this chart we also see that leverage after the pandemic is somewhat lower. However, we see that leverage after the pandemic is distributed fairly evenly while leverage is mainly at 0.3 before the pandemic.

5. **Perform multiple regression to determine the significant determinants of Leverage**

   5.1 With the usual individual variables (model 1)

```
# task 1
model_1<-lm(Leverage ~ ROA  + AssetStructure + FirmSize + CurrentRatio, data = df)
summary(model_1)
```

```
Call:
lm(formula = Leverage ~ ROA + AssetStructure + FirmSize + CurrentRatio,
    data = df)

Residuals:
      Min        1Q    Median        3Q       Max
-0.059142 -0.026330  0.000534  0.023327  0.083534

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -1.297576   1.687315  -0.769   0.4468
ROA             0.006854   0.002596   2.640   0.0121
AssetStructure  0.091407   0.084880   1.077   0.2885
FirmSize        0.097009   0.090486   1.072   0.2906
CurrentRatio   -0.065709   0.011662  -5.634 1.96e-06

(Intercept)
ROA            *
AssetStructure
FirmSize
CurrentRatio   ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03609 on 37 degrees of freedom
Multiple R-squared:  0.7747,    Adjusted R-squared:  0.7504
F-statistic: 31.81 on 4 and 37 DF,  p-value: 1.627e-11
```

We see that the output multiple R square gives a high difference of up to 0.77. However, provided that the significance level is 10%, there are many variables that are not statistically significant such as ROA, Asset Structure, Firm Size. The Current Ratio variable gives good results.

5.2 With the usual individual variables and the interaction between Covid-19 dummy variable and the independent variables (model 2)

```
# task 2
colnames(df)[ncol(df)] = 'Covid'
for (i in 1:nrow(df)){
  if (df[i, 'Covid'] == 'before Covid-19 pandemic'){
    df[i, 'Covid'] = 0
  } else {
    df[i, 'Covid'] = 1
  }
}
print(head(df))
model_2<-lm(Leverage ~ ROA  + AssetStructure + FirmSize + CurrentRatio + Covid, data = df)
summary(model_2)
```

```
Call:
lm(formula = Leverage ~ ROA + AssetStructure + FirmSize + CurrentRatio +
    Covid, data = df)

Residuals:
      Min        1Q    Median        3Q       Max
-0.058993 -0.021994 -0.003674  0.023482  0.079272

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -2.260264   2.171583  -1.041   0.3049
ROA             0.006734   0.002619   2.571   0.0144
AssetStructure  0.118904   0.093783   1.268   0.2130
FirmSize        0.149641   0.117342   1.275   0.2104
CurrentRatio   -0.062631   0.012512  -5.006 1.48e-05
Covid1         -0.014465   0.020327  -0.712   0.4813

(Intercept)
ROA            *
AssetStructure
FirmSize
CurrentRatio   ***
Covid1
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03634 on 36 degrees of freedom
Multiple R-squared:  0.7778,    Adjusted R-squared:  0.747
F-statistic: 25.21 on 5 and 36 DF,  p-value: 7.748e-11
```

Similar to the previous task, the results show that Multiple R squared is still quite high but many variables are not statistically significant. In which, the Covid variable has a very large P value of up to 0.48.

5.3 Predict the value of the variable of assigned topic for all the quarters of the sample using Model 1

```
# task 3
df_new = df
df_new$Date = NULL
df_new$Leverage = NULL
df_new$Covid = NULL
df$Predict = predict(model_1, newdata = df_new)
View(df[, c('Leverage', 'Predict')], 'Predict')
```

	Leverage	Predict
1	0.5136021	0.4465605
2	0.4732455	0.4099879
3	0.4595144	0.4126330
4	0.4376026	0.3977138
5	0.3625351	0.3719561
6	0.3105466	0.3696887
7	0.3080840	0.3482950

```r
# RMSE, MSE
#install.packages('Metrics')
library("Metrics")
rmse(df$Leverage , df$Predict)
mse(df$Leverage , df$Predict)
```

```
> rmse(df$Leverage , df$Predict)
[1] 0.03387558
> mse(df$Leverage , df$Predict)
[1] 0.001147555
```

We see that MSE and RMSE both give very low results. This shows that the predict model gives good and acceptable results. If you do not know what MSE and RMSE are, I will briefly explain that it stands for Mean Squared Erorr and Root mean squared error. They are used to evaluate the performance of your model. If you want to learn more details and how they work, you can find out more online. It is not convenient for me to elaborate here because it is very verbose and distracting.

6. Perform ARIMA model to predict the variable of interest for the 4 quarters in 2022

```
########## 6. Perform ARIMA model to predict the variable of interest for the 4 quarters in 2022

#install.packages('forecast')
library(forecast) #forecast, accuracy
#install.packages('quantmod')
library(quantmod) #getSymbols
#install.packages('tseries')
library(tseries) #adf.test
#install.packages('lmtest')
library(lmtest) #coeftest
#install.packages('stats')
library(stats) #Box.test

#adjust plot margins to avoid the error in plot.new() figure margins too large
par(mar = c(1, 1, 1, 1))
#create scatterplot
plot(1:30)

#use auto.arima function to determine best P, D, Q
dif = data.frame(diff(df$Leverage, differences = 2))
auto=auto.arima(dif,seasonal=F,trace = T,max.order=4,ic='aic')
coeftest(auto.arima(dif,seasonal=F))
acf(auto$residuals)
pacf(auto$residuals)
Box.test(auto$residuals,lag=20,type='Ljung-Box')
```

```
 ARIMA(2,0,2) with non-zero mean : Inf
 ARIMA(0,0,0) with non-zero mean : -122.0055
 ARIMA(1,0,0) with non-zero mean : -127.3088
 ARIMA(0,0,1) with non-zero mean : Inf
 ARIMA(0,0,0) with zero mean     : -123.9817
 ARIMA(2,0,0) with non-zero mean : -134.3133
 ARIMA(3,0,0) with non-zero mean : -136.4393
 ARIMA(4,0,0) with non-zero mean : -139.4403
 ARIMA(5,0,0) with non-zero mean : -137.4408
 ARIMA(4,0,1) with non-zero mean : -137.441
 ARIMA(3,0,1) with non-zero mean : Inf
 ARIMA(5,0,1) with non-zero mean : Inf
 ARIMA(4,0,0) with zero mean     : -140.5828
 ARIMA(3,0,0) with zero mean     : -138.1443
 ARIMA(5,0,0) with zero mean     : -138.6153
 ARIMA(4,0,1) with zero mean     : -138.6151
 ARIMA(3,0,1) with zero mean     : -139.7627
 ARIMA(5,0,1) with zero mean     : -137.1641

 Best model: ARIMA(4,0,0) with zero mean
```

The results show that the best model is ARIMA(4,0,0) with zero mean

```
z test of coefficients:

    Estimate Std. Error z value  Pr(>|z|)    
ar1 -0.84134    0.15028 -5.5983 2.164e-08 ***
ar2 -0.84752    0.17979 -4.7139 2.430e-06 ***
ar3 -0.57354    0.18986 -3.0208  0.002521 ** 
ar4 -0.37013    0.16746 -2.2103  0.027086 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

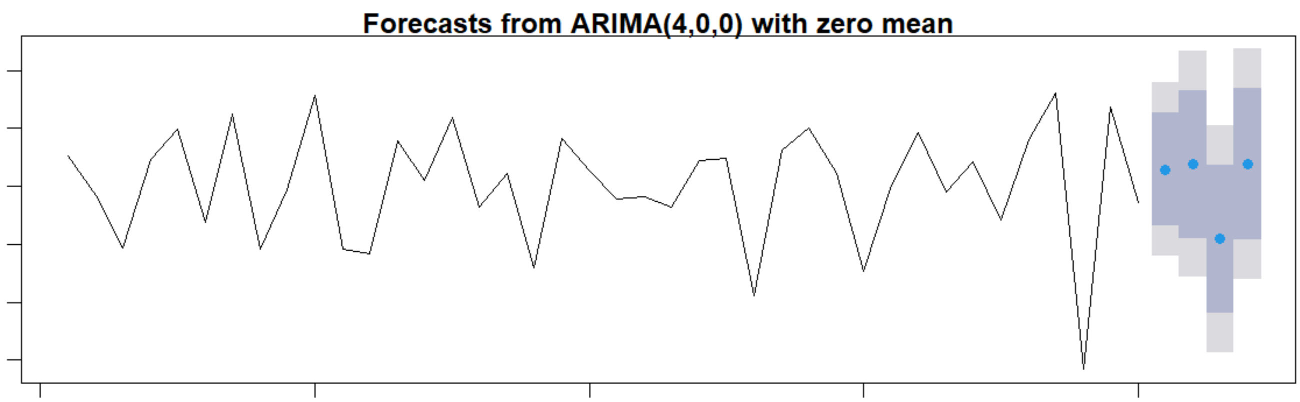
```
        Box-Ljung test

data:  auto$residuals
X-squared = 14.942, df = 20, p-value = 0.7797
```

Prediction

```
185  #prediction
186  term=4
187  fcastauto=forecast(auto,h=term)
188  fcastauto #fcastauto is the predicted values for 44 terms
189  plot(fcastauto)
```

```
   Point Forecast       Lo 80      Hi 80       Lo 95      Hi 95
41      0.01455965 -0.03428364 0.06340294 -0.06013971 0.08925902
42      0.01893555 -0.04489531 0.08276641 -0.07868532 0.11655643
43     -0.04556213 -0.10975647 0.01863221 -0.14373890 0.05261465
44      0.01926313 -0.04614716 0.08467342 -0.08077327 0.11929953
```



The chart above visualizes the model's four-quarter 2022 leverage forecast results.

Check Accurency

```
190  # check accuracy
191  accuracy(fcastauto)
```

```
                      ME       RMSE        MAE      MPE     MAPE      MASE        ACF1
Training set 0.004750468 0.03615681 0.02703245 25.11135 107.0847 0.4016127 -0.02694473
```

7. Explain in fewer than 150 words how Random forest can be used in this case to predict the variable of interest for the 4 quarters in 2022.

Because of the number of decision trees involved in this process, random forests is considered an accurate and powerful method. The model results are explanatory, minimizing the overfitting problem used in classification and regression, and are automatically generated. We can combine many variables together, and Random Forest will choose the variables that have the greatest influence on the dependent variable by calculating the Feature Importance of each variable in the training set (train set), making the model more efficient.

To forecast the value of the leverage ratio in the four quarters of 2022, we should use Random Forest in the form of a regression, giving the model the influence of the independent variables on the leverage ratio in the quarters using the input data. The trained model can then use the existing data to predict the future cash holding value.