VIETNAM NATIONAL UNIVERSITY HO CHI MINH CITY
UNIVERSITY OF ECONOMICS AND LAW

FINAL REPORT

Subject: PROGRAM PACKAGE IN FINANCE 2

LECTURER: PhD. Nguyen Thanh Liem

Submitted by: Pham Quynh Huong
ID: K194141724

Ho Chi Minh City, June 15th, 2022

TABLE OF CONTENTS

1. Introduction

1.1. Definition of leverage

A leverage ratio is one of the important financial measurements that looks at how much capital comes in the form of debt (loans) or assesses the ability of a company to meet its financial obligations. The leverage ratio is necessary because companies rely on a mixture of equity and debt to finance their operations, and knowing the amount of debt held by a company is useful in evaluating whether it can pay off its debts as they come due. According to Myers (1977), in the operation, enterprises use debt to compensate for capital shortage, increase the rate of return on equity, and at the same time, the increase in debt increases financial risk for businesses. Businesses use financial leverage in the hope of increasing profits for common shareholders. From there, businesses can use fixed-cost capital sources to generate profits that are greater than the costs of raising fixed-income capital.

Too much debt can be dangerous for a company and its investors. However, if a company's operations can generate a higher rate of return than the interest rate on its loans, then the debt may help to fuel growth. Uncontrolled debt levels can lead to credit downgrades or worse. On the other hand, too few debts can also imply this company is not such a good diverse investment firm. A reluctance or inability to borrow may be a sign that operating margins are tight. A high debt/equity ratio generally indicates that a company has been aggressive in financing its growth with debt. This can result in volatile earnings as a result of the additional interest expense. If the company's interest expense grows too high, it may increase the company's chances of a default or bankruptcy. Typically, a D/E ratio greater than 2.0 indicates a risky scenario for an investor, however, this figure can vary by industry.

Formula of leverage:

$$leverage\ ratio\ =\ \frac{total\ liability}{total\ equity}$$

1.2. Reason for choosing CTD (Coteccons Construction Joint Stock Company) as a company be analyzed factors affecting leverage on.

The construction industry is an important infrastructure sector to promote the country's socio-economic development. Nowadays, with the rapid economic growth, the construction industry not only contributes a vital role in daily life and public services, but also a high important fraction in national GDP. However, in the first half of 2021, most construction projects located in cities that are isolated due to the pandemic have to stop, and projects outside the isolated area are also delayed due to material and human resource disruptions. This causes construction enterprises to face many heavy damages such as maintenance costs, pandemic prevention costs, resource mobilization costs after the lockdowns,... Especially, fluctuation of raw material prices, due to macro factors, is the biggest nightmare for contractors. The main reason is that the purchase price of raw materials accounts for about 65 - 70% of the projected price of the entire project. So, studying the factors affecting the leverage of enterprises operating in the construction industry is necessary to help managers, analysts, and investors look over an objective view of the financial situation of an enterprise in general and forecasts of change in capital structure in particular. Should investors continue to invest in this market or will this situation be worse in the future?

Coteccons Construction Joint Stock Company was officially listed on the Ho Chi Minh City Stock Exchange. Ho Chi Minh City (Hose) on December 9, 2009. Number of shares listed initially 12,000,000 shares with stock code "CTD". COTECCONS is a large-size company by years (based on the total assets and revenue conditions), so it has a capable of carrying out modern and large projects in the country today. The company has a strong and stable management and personnel system and is on the way to strong development. COTECCONS is developing strongly its core business, which is strong in construction, installation, repair civil, industrial areas, infrastructure technical systems, traffic systems, irrigation. In addition, COTECCONS also continues to invest in the field of real estate investment and business. Therefore, the current macro-economy currently is suffering from many fluctuations, and the construction sector is incurring from many unexpected changes in raw material prices, the analysis of factors affecting the leverage (also known as capital structure) will make recommendations and support investors, managers, and creditors in the process of evaluating financial leverage,the financial position and business performance of the enterprise.

2. Literature review

2.1. Tangible assets

A tangible asset is an asset that has a finite monetary value and usually a physical form. Tangible assets can typically always be transacted for some monetary value though the liquidity of different markets will vary. Tangible assets are the opposite of intangible assets which have a theorized value rather than a transactional exchange value, however, they are also usually the easiest to understand and value. Tangible assets are the most basic type of assets on the balance sheet. They are usually the main form of assets in most industries.

Companies with a lot of tangible assets often use leverage to finance production and business activities. Those companies may use tangible assets as collateral, either providing more access to creditors or as a guarantee in case of bankruptcy. Especially, long-term leverage is positively correlated with tangible assets which must be financed in a long period for lower costs and long-term interest rate. Kraus and Litzenberger (1973) proposed the trade-off theory. When making capital structure decisions, managers must trade off between the tax benefits and the projected costs of bankruptcy in the capital structure process. This theory predicts that leverage is positively associated with firm size and tangible assets.

2.2. Liquidity (Cash ratio)

Liquidity is one of the best measures to assess financial strength and repaid ability of a company, especially, cash ratio. The cash ratio is more conservative than other liquidity ratios because it only considers a company's most liquid resources. It specifically calculates the ratio of a company's total cash and cash equivalents to its current liabilities. The metric evaluates a company's ability to repay its short-term debt with cash or near-cash resources, such as easily marketable securities. This information is useful to creditors when they decide how much money they would be willing to loan a company.

According to Tran Hung Son (2013), one of the most factors affecting the capital structure of industrial production enterprises is tangible fixed assets which is positively correlated with capital structure of industrial production enterprises in Ho Chi Minh City and listed industrial production enterprises. On the other hand, liquidity is negatively correlated with the capital structure of both groups of industrial enterprises. Because enterprises with sufficient cash and equivalents, besides, growing steadily in many years indicate an explicit growth and profitability. So these businesses have enough liquidity to finance their investment and business purposes. It is important to notice that a high cash ratio indicates that the firm is not using its cash to its best advantage. Cash should be put to work in operations of the company (Gomes, 2012). In contrast, new or unevenly growing businesses have frequently fluctuating and narrow cash ratios, so debt financing is a popular solution for them to cover operating costs, use tax shields and reduce issuance costs.

Formula of liquidity:

$$Cash\ Ratio = \frac{Cash + Cash\ Equivalents}{Current\ Liabilities}$$

2.3. Revenue growth rate

Growth rates refer to the percentage change of a specific variable within a specific time period, such as growth at revenues, earnings, dividends, or even macro concepts. In this case, we discuss revenue growth rate. Revenue growth rate is computed by dividing the difference between the ending and starting revenue in the period being analyzed and dividing that by the starting revenue, its unit is percentage. Revenue growth rate can be beneficial in assessing a company's operating performance and to predict future sales.

Jermias and Yigit (2019) find that growth opportunity is negatively and significantly associated with leverage. In addition, the trade-off theory of Kraus and Litzenberger (1973) refers to capital structure decisions, in which this theory predicts that leverage is negatively related to growth but positively related to business size, tangible assets.

Formula of revenue growth rate:

$$Revenue\ Growth\ Rate = \frac{Revenue\ this\ period}{Revenue\ previous\ period}$$

2.4. Future free cash flow:

Free cash flow (FCF) represents the cash a company generates after accounting for cash outflows to support operations and maintain its capital assets. Also, it is used for repaying creditors and paying out dividends and interest to investors. Unlike earnings or net income, free cash flow is a measure of profitability that excludes the non-cash expenses of the income statement and includes spending on equipment and assets as well as changes in working capital from the balance sheet.

FCF is also helpful for potential shareholders or lenders to evaluate how likely the company will be able to pay their expected dividends or interest. If the company's debt payments are deducted from FCF, a lender would have a better idea of the quality of cash flows available for additional borrowings. Similarly, shareholders can use FCF minus interest payments to consider the expected stability of future dividend payments. Following the research paper of Shenoy and Koch (1996), the signaling theory implies a positive relationship between current leverage and future free cash flow in the same time period. It means when the companies borrow for future projects, its future cash flow will increase due to those projects' earnings.

2.5. Return on total assets

Return on total assets (ROA) refers to a financial ratio that indicates how profitable a company is in relation to its total assets. Corporate management, analysts, and investors can use ROA to determine how efficiently a company uses its assets to generate a profit. This metric is commonly expressed as a percentage by using a company's net income and its assets. A higher ROA means a company is more efficient and productive at managing its balance sheet to generate profits while a lower ROA indicates there is room for improvement.

Many researches proved that profitability is negatively correlated with financial leverage. Because companies with high business efficiency tend to use internal financing over external financing production and business to achieve higher profitability. This implies that for companies with good business results, shareholders will not want to share this advantage with creditors, because the investment of the business will transfer benefits from shareholders to creditors, bondholders. So making capital structure decisions is very important to gain the highest profit which not only optimizes useful tax-shields but also maximizes the wealth of shareholders.

Formula of return on total assets:

$$Return\ on\ total\ assets\ (ROA) = \frac{Net\ income}{Total\ assets}$$

In this report, the hypothesis proposed in this study is as follows:

- Tangible assets, and future free cash flow are positively correlated with current leverage.
- Cash ratio, revenue growth rate, and return on total assets are negatively correlated with current leverage.

3. Data collection

3.1. Research subjects:

Financial data (balance sheets and income statements) of Coteccons Construction Joint Stock Company (CTD)

3.2. Research scope:

This report is considered from the beginning of first quarter in 2010, when this company started to be listed on HOSE, to the finishing of fourth quarter in 2021 (48 quarters)

3.3. Research variables:

Dependent variable: leverage, rate of return on leverage

Independent variables: tangible asset, cash ratio, revenue growth rate, return on total asset, future free cash flow.

3.4. Data source:

Saigon Securities Incorporation's fundamental analysis platform.

4. Descriptive statistics

4.1. Descriptive statistics for full period from 2010 to 2021

```
descr(data[c('cash_ratio', 'future_fcf', 'growth_rate', 'roa', 'tangible_assets',
'leverage')], transpose = TRUE, stats=c('mean', 'min', 'med', 'max', 'sd'))
```

```
Descriptive Statistics

                              Mean              Min           Median                Max
-------------------- ---------------- ---------------- ---------------- ----------------
          cash_ratio             0.28             0.03             0.27             0.68
          future_fcf   881474406425.36   225702665430.00   681485846811.00   3311114590570.00
         growth_rate             0.12            -0.58             0.15             1.73
            leverage             0.85             0.46             0.84             1.41
                 roa             0.02             0.00             0.02             0.04
     tangible_assets   363762054805.53   117425226052.00   349829182892.00    673342463523.00

Table: Table continues below



                              Std.Dev
-------------------- ----------------
          cash_ratio             0.16
          future_fcf   618282365531.06
         growth_rate             0.44
            leverage             0.23
                 roa             0.01
     tangible_assets   208530252341.24
```

4.2. Descriptive statistics for the before period from 2010 to 2019

```
descr(data_before[c('cash_ratio', 'future_fcf', 'growth_rate', 'roa', 'tangible_assets',
'leverage')], transpose = TRUE, stats=c('mean', 'min', 'med', 'max', 'sd'))
```

```
Descriptive Statistics

                              Mean              Min           Median                Max
-------------------- ---------------- ---------------- ---------------- ----------------
          cash_ratio             0.30             0.06             0.29             0.68
          future_fcf   910652921730.97   251978411295.00   740482607517.00   3311114590570.00
         growth_rate             0.13            -0.52             0.17             0.93
            leverage             0.89             0.46             0.88             1.41
                 roa             0.02             0.01             0.02             0.04
     tangible_assets   335575075021.03   117425226052.00   181895039018.00    673342463523.00

Table: Table continues below



                              Std.Dev
-------------------- ----------------
          cash_ratio             0.15
          future_fcf   653582951074.09
         growth_rate             0.36
            leverage             0.23
                 roa             0.01
     tangible_assets   217419984801.78
```

4.3. Descriptive statistics for the after period from 2020 to 2021

```
descr(data_after[c('cash_ratio', 'future_fcf', 'growth_rate', 'roa', 'tangible_assets',
'leverage')], transpose = TRUE, stats=c('mean', 'min', 'med', 'max', 'sd'))
```

Descriptive Statistics

	Mean	Min	Median	Max
cash_ratio	0.13	0.03	0.13	0.26
future_fcf	739229144310.50	225702665430.00	645023921375.00	1396764826755.00
growth_rate	0.07	-0.58	-0.15	1.73
leverage	0.67	0.56	0.66	0.78
roa	0.00	0.00	0.01	0.01
tangible_assets	501173581255.00	420946301186.00	501308229533.00	581412205517.00

Table: Table continues below

	Std.Dev
cash_ratio	0.08
future_fcf	406553892314.59
growth_rate	0.76
leverage	0.09
roa	0.00
tangible_assets	55994390393.04

According to above descriptive statistic results, we find that the Covid-19 pandemic caused many negative changes in those financial factors of CTD such as:

Cash ratios, future free cash flow significantly decreased all figures in last 8 quarters (included mean, min, median, max) and also narrowed the range of its values during Covid-19 period (standard deviation was dramatically lower). Because in this bad situation, firm not only could not expand investment and business, but also incurred many costs such as labor costs, maintain costs, interests, and so on.

Besides, ***revenue growth rate*** also had a lot of fluctuation during Covid-19 period. In more than half of the Covid-19 period, the revenue growth rate was negative below 15% (this period's median was equal to -15%), and the mean and min values also decreased a lot. However, in the fourth quarter of 2021, CTD started to recover and had a remarkable growth compared to the previous quarter, so the maximum value of the Covid-19 period is 173%, and the standard deviation is 76%. (much larger than the pre-pandemic stage of 36%).

Rate of return on total asset also went down quickly. During Covid-19 stage, ROA is under 1%. At some times, it declined to 0%. So that's why its standard deviation is nearly 0%. However, at the previous pandemic stage, ROA is up to 4% and standard deviation is 1%. This proves in some pandemic times, firm did not earn any income.

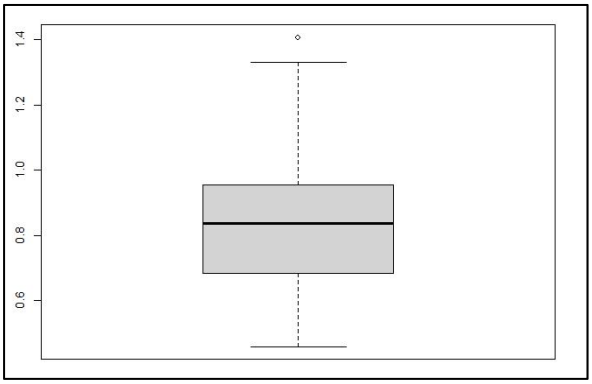
About ***tangible assets,*** objectively, the pre-pandemic stage accounts for 10 years of the company's development, while the Covid-19 pandemic stage accounts for only 2 years, so the tangible asset's standard deviation of the pre-pandemic period will be much larger than the period during the Covid-19 (this cannot be compared by the ratio method like cash ratio, revenue growth rate or ROA). However, when we compare the maximum of tangible assets of the pre-pandemic period with the mean, median, or maximum of tangible assets in the Covid-19 period, we find that tangible assets decreased in the outbreak time.

Finally, about the target variable of this paper - ***leverage***, we find out that leverage significantly decreased in 2-year Covid-19 period. The reasons are the difference in the numbers of observation between 2 periods, and the reduction in cash ratio, revenue growth rate, ROA, and tangible assets, which made the business needed to more focus on repaid ability than concentrate in doing business and investment. Moreover, in the complicated situation of the pandemic, firm was under a lot of pressure from old debts, business performance was ugly. So this made creditors and banks unwilling to lend business more.

5. Visualization

5.1. Box and whisker plot of leverage

```
boxplot(data$leverage)
```



Leverage is almost range from 70% to nearly 100%. It shows that this firm is likely to use more debt than equity to finance its projects. This strategy also bring advantages and disadvantages.

Disadvantages:

- Firm will have more risk of default than other ones which does not use much debt, as the business is required to have sufficient cash balance at all times to make periodic interest payments, regardless of conditions. But declining in sales can create serious problems in meeting loan payment dates.

- Debt financing will reduce corporate cash flow each period, which reduces the amount of money needed for growth.

- A company with too high a debt level also makes creditors and banks unwilling to approve new debt; In addition, investors will also see the financial risk of the company and do not want to invest more equity.

Advantages:

- Cost of equity is usually higher than cost of debt, so using reasonally debt rather than equity will decrease cost of capital, increase net income.

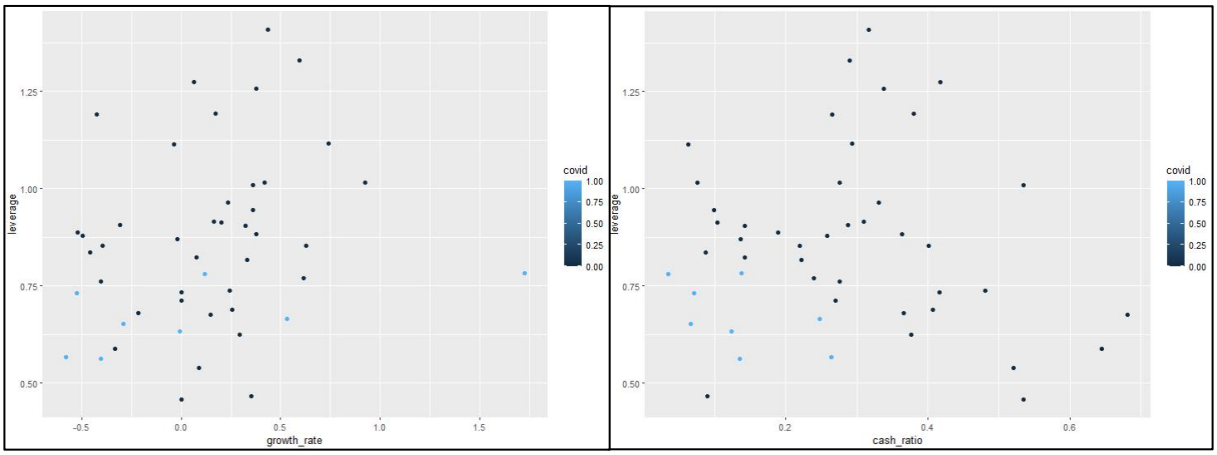
- Less of conflicts between shareholders when firm does not raise much fund by issuing.

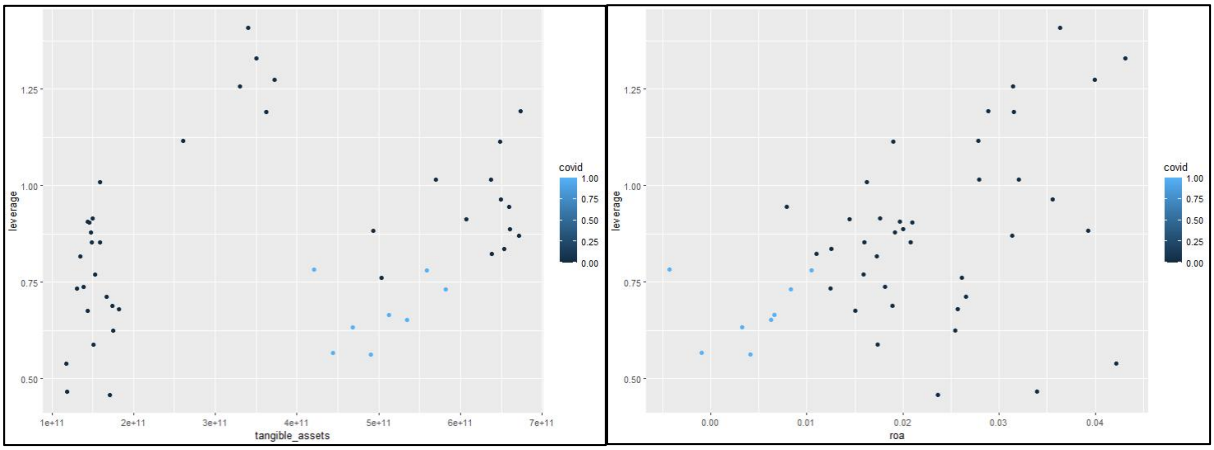
- Firm can enjoy tax shields because loan interest is tax deductible, whereas dividends paid to shareholders are not.

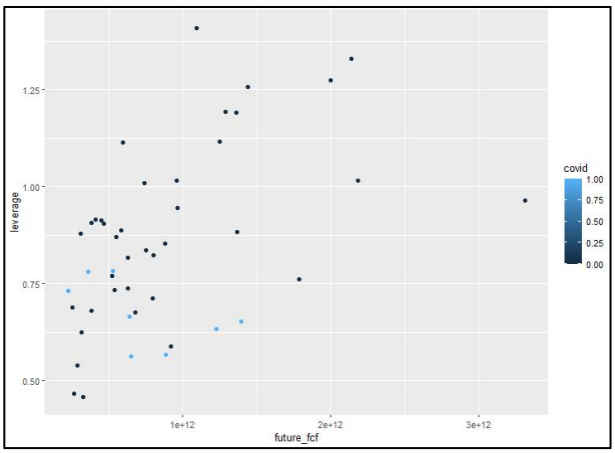
- When financing by debt, managers can previously predict long, medium, or short-term cash flows.

5.2. Scatter plots of interactions between leverage and independent variables

```
# Leverage with cash ratio, covid_19
ggplot(data, aes(x= cash_ratio, y= leverage, color = covid)) +geom_point()
# Leverage with growth rate, covid_19
ggplot(data, aes(x= growth_rate, y= leverage, color = covid)) +geom_point()
# Leverage with tangible assets, covid_19
ggplot(data, aes(x= tangible_assets, y= leverage, color = covid)) +geom_point()
# Leverage with ROA, covid_19
ggplot(data, aes(x= roa, y= leverage, color = covid)) +geom_point()
# Leverage with future FCF, covid_19
ggplot(data, aes(x= future_fcf, y= leverage, color = covid)) +geom_point()
```



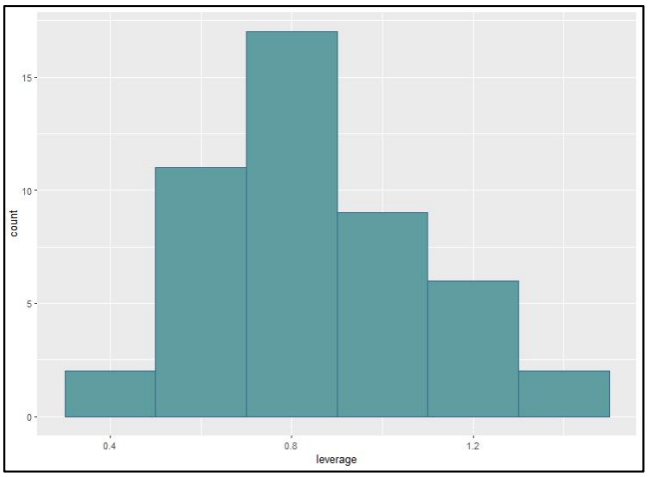




After seeing above scatter charts, we can find out that revenue growth rate, ROA, and future free cash flow are positively correlated with leverage. This is not correctly same as literature reviews but it is reasonal to analyse in practical situations. When revenue growth rate, ROA, and future free cash flow increase, leverage also increase. Meanwhile, cash ratio and tangible asset are not clearly to demonstrate for the relationship between their values and leverage. Especially, about observations which were affected by Covid-19, their leverage values are very small and cause other variables' values decrease too, excepting of tangible assets.

5.3. Histogram of leverage

```
ggplot(data, aes(x=leverage)) +
  geom_histogram(binwidth=.2, fill='cadetblue', color='steelblue4')
```



According to this histogram, CTD's leverage is almost range from 50% to 90%. This says that CTD usually use debt for its financial plans and enjoy the benefit of tax shield. But this firm must consider about the repaid ability in any economic situations.

6. Regression

6.1. Multi regression model with the usual individual variables

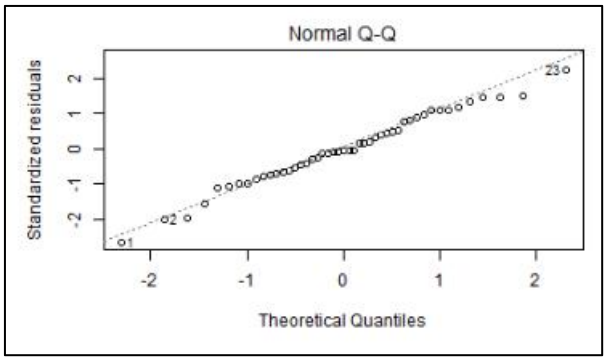
```
leverage.model1<-lm(leverage~tangible_assets+cash_ratio+growth_rate+future_fcf+roa,
data=data)
summary(leverage.model1)
summary(leverage.model1)$coefficient
par(mfrow=c(2,2))
plot(leverage.model1)
```

```
Call:
lm(formula = leverage ~ tangible_assets + cash_ratio + growth_rate +
    future_fcf + roa, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.43058 -0.11675 -0.01064  0.13979  0.39953

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      5.741e-01  1.162e-01   4.941 1.35e-05 ***
tangible_assets  1.722e-13  1.898e-13   0.907  0.36948
cash_ratio      -1.559e-01  2.333e-01  -0.668  0.50778
growth_rate      9.035e-02  6.283e-02   1.438  0.15802
future_fcf       9.956e-14  5.694e-14   1.748  0.08787 .
roa              7.615e+00  2.806e+00   2.714  0.00967 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1846 on 41 degrees of freedom
Multiple R-squared:  0.405,     Adjusted R-squared:  0.3324
F-statistic: 5.581 on 5 and 41 DF,  p-value: 0.0005194
```



Call section:

The call section shows us the formula that R used to fit the regression model. Leverage is dependent variable and I am using both of tangible asset, revenue growth rate, ROA, cash ratio, and future free cash flow as predictors (independent variables) from the full dataset.

Residuals:

Looking at the output of residuals above, it looks like our distribution is not quite symmetrical and is slightly left-skewed. This tells us that our model is not predicting as well at the low leverage ranges as it does for the

high ranges. We can visualize this with a quantile-quantile (Q-Q) plot. Looking at the "Normal Q-Q" chart above, you can see there are outliers on both ends of the chart, but those on the lower end look more severe than those on the top. Overall the residuals look to have a fairly normal distribution.

Coefficients - Estimate:

Following the coefficients of estimate above, we have the equation of "***leverage = 0.57 + 0* tangible assets - 0.16* cash ratio + 0.09* revenue growth rate + 0* future free cash flow + 7.62* ROA***". As a baseline, if this company increase a unit of tangible, that the leverage would not be likely to make any change on average. But if it increase a unit of cash ratio, leverage would decrease 0.16 units. Moreover, in case of all independent variables are zero, leverage is still equal to 0.57 (it means 57%).

Coefficients - standard error:

The standard error of the coefficient is an estimate of the standard deviation of the coefficient. In effect, it is telling us how much uncertainty there is with our coefficient. The standard error is often used to create confidence intervals. For example, as the task requires a 90% significant level, we have the range of ROA is "***7.615 ± 1.68(2.806) = (2.900 ; 12.329)"***. Looking at the confidence interval, we can say we are 90% confident that the actual slope is between 2.900 and 12.329.

Coefficients — t value

The t-statistic is simply the coefficient divided by the standard error. In general, we want our coefficients to have large t-statistics, because it indicates that our standard error is small in comparison to our coefficient. In our example ROA coefficient is 2.714 standard errors away from zero, which statistically, is quite far. The larger our t-statistic is, the more certain we can be that the coefficient is not zero. The t-statistic is then used to find the p-value.

Coefficients — Pr(>|t|)

The p-value is calculated using the t-statistic from the T distribution. The p-value, in association with the t-statistic, help us to understand how significant our coefficient is to the model. In this case, any p-value below 0.10 is deemed as significant. By this report, we only have two per five variables which are significant to the model. There is ROA (p value = 0.0097) and future free cash flow (p-value = 0.0879). Other ones have the p-value higher than 10%. It means only ROA and future free cash flow in fact add value to the model by helping to explain the variance within our dependent variable.

Residual Standard Error:

For our current model, we can see that on average, the actual values are 0.1846 away from the predicted values. Then understanding our full data set, and knowing that the largest leverage is 1.41, having all our predictions be off on average by 0.1846 won't produce a very accurate model.

Multiple R-squared and Adjusted R-squared:

In the result above, those independent variables explain about 40.5% (multiple R-squared) of the variation within leverage, the dependent variable. This means that those independent variables help to explains some of the variation within leverage, but not as much as we would like. Ultimately, our model isn't fitting the data very well (we saw this when looking at the residual standard error). Besides, the adjusted R-squared value is equal to 33.24%, not a far difference from multiple R-squared, so it is not a overfitting model. Finally, a very low R-squared indicates underfitting and adding additional relevant features or using a complex model might help.

F-statistic and p-value:

We can see from this model, the F-statistic is quite large and p-value is so small it is basically zero. This would lead us to reject the null hypothesis and conclude that there is strong evidence that a relationship does exist between leverage and at least a independent variable.

6.2. Multi regression model with the usual individual variables and the interaction between Covid-19 dummy variable and the independent variables

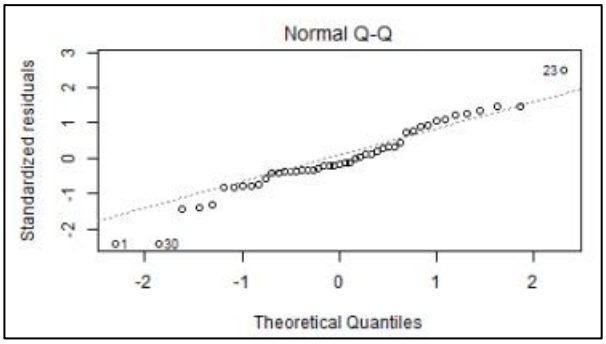
```
leverage.model2 <-
lm(leverage~covid*tangible_assets+covid*cash_ratio+covid*growth_rate+covid*future_fcf+covi
d*roa, data=data)
summary(leverage.model2)
summary(leverage.model2)$coefficient
par(mfrow=c(2,2))
plot(leverage.model2)
```

```
Call:
lm(formula = leverage ~ covid * tangible_assets + covid * cash_ratio +
    covid * growth_rate + covid * future_fcf + covid * roa, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.38992 -0.06921 -0.01138  0.10498  0.44572

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)           7.226e-01  1.385e-01   5.217 8.32e-06 ***
covid                -5.219e-01  2.315e+00  -0.225   0.8229
tangible_assets       1.392e-13  2.009e-13   0.693   0.4929
cash_ratio           -2.671e-01  2.524e-01  -1.058   0.2972
growth_rate           7.811e-02  8.931e-02   0.875   0.3877
future_fcf            1.490e-13  6.446e-14   2.311   0.0268 *
roa                   2.194e+00  3.932e+00   0.558   0.5803
covid:tangible_assets 9.844e-13  4.615e-12   0.213   0.8324
covid:cash_ratio     -5.732e-02  1.184e+00  -0.048   0.9617
covid:growth_rate     8.775e-03  1.518e-01   0.058   0.9542
covid:future_fcf     -1.890e-13  2.247e-13  -0.841   0.4060
covid:roa            -8.148e+00  4.542e+01  -0.179   0.8586
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1868 on 35 degrees of freedom
Multiple R-squared:  0.4796,    Adjusted R-squared:  0.316
F-statistic: 2.932 on 11 and 35 DF,  p-value: 0.00755
```



Call section:

The call section shows us the formula that R used to fit the regression model. Leverage is dependent variable and I am using the interaction between Covid-19 dummy variable and the independent variables from the full dataset.

Residuals:

Looking at the output of residuals above, it looks like our distribution is not quite symmetrical and is slightly right-skewed. This tells us that our model is not predicting as well at the higher leverage ranges as it does for the low ranges. We can visualize this with a quantile-quantile plot. Looking at the chart below, you can see there are outliers on both ends of the chart, but those on the upper end look more severe than those on the bottom. Overall the residuals look to have a fairly normal distribution.

Coefficients - Estimate:

Following the coefficients of estimate above, we have the equation of "***leverage = 0.723 - 0.522* covid + 0* tangible assets - 0.267* cash ratio + 0.0781* revenue growth rate + 0* future free cash flow + 2.194* ROA + 0* tangible assets* covid - 0.0573* cash ratio* covid + 0.00878* revenue growth rate* covid + 0* future free cash flow* covid - 8.148* ROA* covid***". As a baseline, if this company increase a unit of tangible, that the leverage would not be likely to make any change on average, regardless of covid-19 situation. But the addition

in per unit of each independent variable will be expected to affect on leverage differs for two cases of the appearence of covid-19 or the absence of covid-19. Moreover, in case of all independent variables are zero, leverage is still equal to 0.723 (it means 72.3%).

Coefficients - standard error:

As the task requires a 90% significant level, we have the range of ROA*covid is "-8.148 ± 1.68(0.454) = (-8.911 ; -7.385)". Looking at the confidence interval, we can say we are 90% confident that the actual slope of the interaction between ROA and covid-19 is between -8.911 and -7.385.

Coefficients — t value:

According to this result, only intercept t-value and future free cash flow t-value are more than zero, it means those variables are meaning.

Coefficients — Pr(>|t|):

In this case, any p-value below 0.10 is still deemed as significant. By this report, we only have one per eleven variables which are significant to the model. There is future free cash flow (p value = 0.0268). Other ones have the p-value higher than 10%. It means only future free cash flow in fact add value to the model by helping to explain the variance within our dependent variable.

Residual Standard Error:

For our current model, we can see that on average, the actual values are 0.1868 away from the predicted values. Then understanding our full data set, and knowing that the largest leverage is 1.41, having all our predictions be off on average by 0.1868 won't produce a very accurate model.

Multiple R-squared and Adjusted R-squared:

In the result above, those independent variables explain about 47.96% (multiple R-squared) of the variation within leverage, the dependent variable. This means that those independent variables help to explains some of the variation within leverage, but not as much as we would like. Ultimately, our model isn't fitting the data very well (we saw this when looking at the residual standard error). Besides, the adjusted R-squared value is equal to 31.6%, not a far difference from multiple R-squared, so it is not a overfitting model. Finally, a very low R-squared indicates underfitting and adding additional relevant features or using a complex model might help.

F-statistic and p-value:

We can see from this model, the F-statistic is quite large and p-value is so small it is basically zero. This would lead us to reject the null hypothesis and conclude that there is strong evidence that a relationship does exist between leverage and at least a independent variable.

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

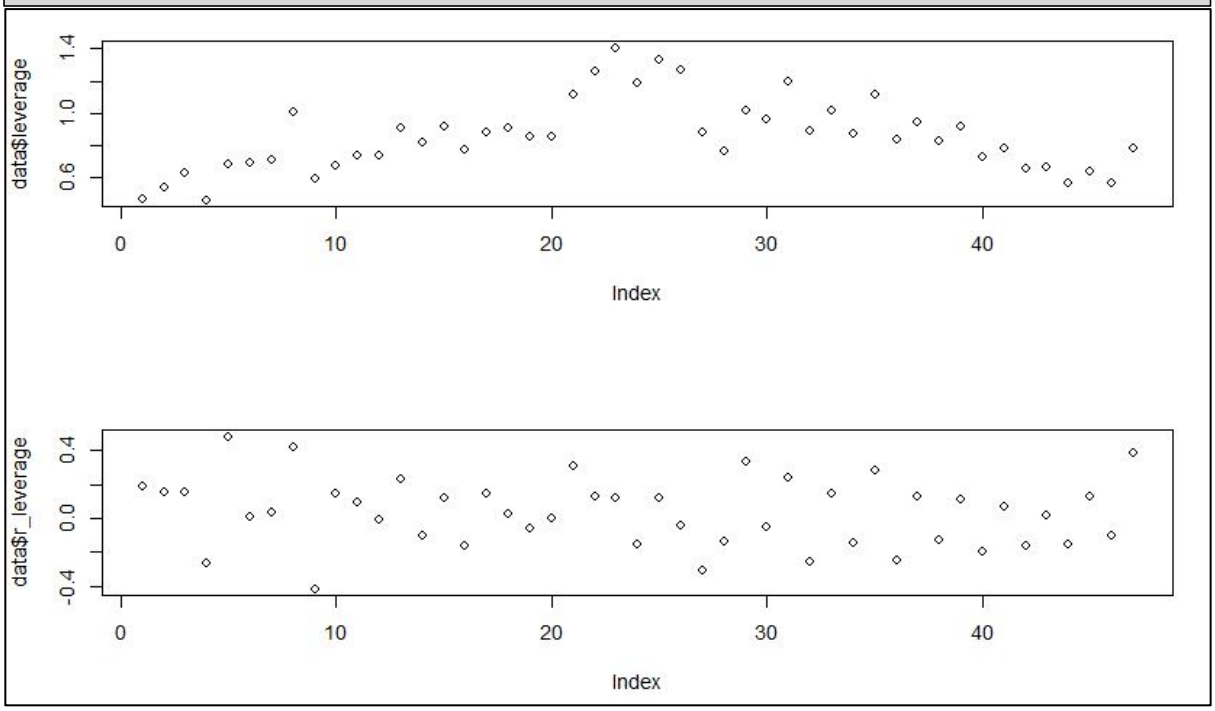
```
predict(leverage.model1, newdata = data)
```

```
> predict(leverage.model1, newdata = data)
        1         2         3         4         5         6         7         8         9        10        11
0.8970761 0.8715652 0.7974362 0.7328129 0.7625548 0.7328893 0.8420346 0.7480584 0.6930948 0.6881701 0.7460713
       12        13        14        15        16        17        18        19        20        21        22
0.6799621 0.7144511 0.7866974 0.7416193 0.7921523 0.6914643 0.8122295 0.8682857 0.7128321 0.9761441 0.9955341
       23        24        25        26        27        28        29        30        31        32        33
1.0082131 0.9322985 1.1840082 1.0819270 1.0714533 0.9579801 1.1281260 1.2557492 0.9943080 0.8224655 1.0638447
       34        35        36        37        38        39        40        41        42        43        44
0.9601739 0.8763923 0.8018472 0.8612933 0.8321911 0.8347449 0.7014321 0.7915139 0.8165017 0.7855980 0.6977546
       45        46        47
0.7817756 0.6381279 0.8009582
```

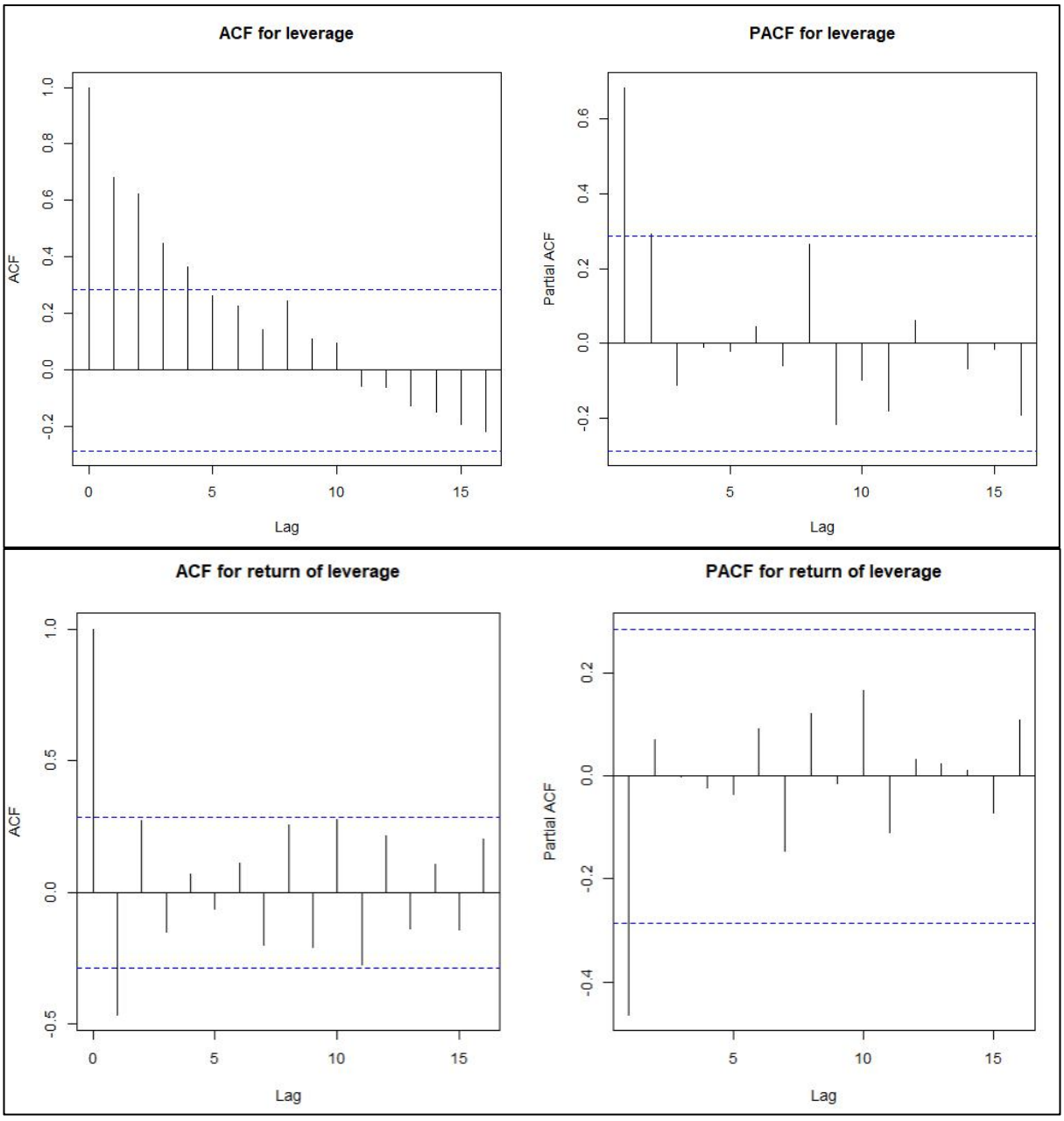
7. ARIMA model

We proceed to run the ARIMA model for leverage and return rate of leverage.

```
# Plot variables
plot(data$leverage)
plot(data$r_leverage)
# Check stationary of leverage
par(mfrow=c(1,1))
acf(data$leverage,main='ACF for leverage')
pacf(data$leverage,main='PACF for leverage')
adf.test(data$leverage)     # p-value = 0.6269 --> not stationary
# Now check whether return of leverage is stationary
acf(data$r_leverage)
pacf(data$r_leverage)
adf.test(data$r_leverage)  # p-value = 0.01705  --> stationary
# Use auto.arima function to determine best P, D, Q
auto=auto.arima(data$r_leverage,seasonal=F,trace = T,max.order=4,
                ic='aic')
coeftest(auto.arima(data$r_leverage,seasonal=F))
acf(auto$residuals)
pacf(auto$residuals)
Box.test(auto$residuals,lag=20,type='Ljung-Box')
# Prediction
term=4
fcastauto=forecast(auto,h=term)
fcastauto    # fcastauto is the predicted values for 4 quarters
plot(fcastauto)
# Test accuracy
length(data$r_leverage)
accuracy(fcastauto)    #train set
```



The chart above demonstrates the distribution of leverage values over time, with leverage ranging from 40% to 140%, not seasonal. For a chart of the return rate of leverage distribution over time, it ranges from -40% to 40%, non-seasonal.



Check stationary of data

```
> adf.test(data$leverage)    # p-value = 0.6269 --> not stationary

        Augmented Dickey-Fuller Test

data:  data$leverage
Dickey-Fuller = -1.8674, Lag order = 3, p-value = 0.6269
alternative hypothesis: stationary
```

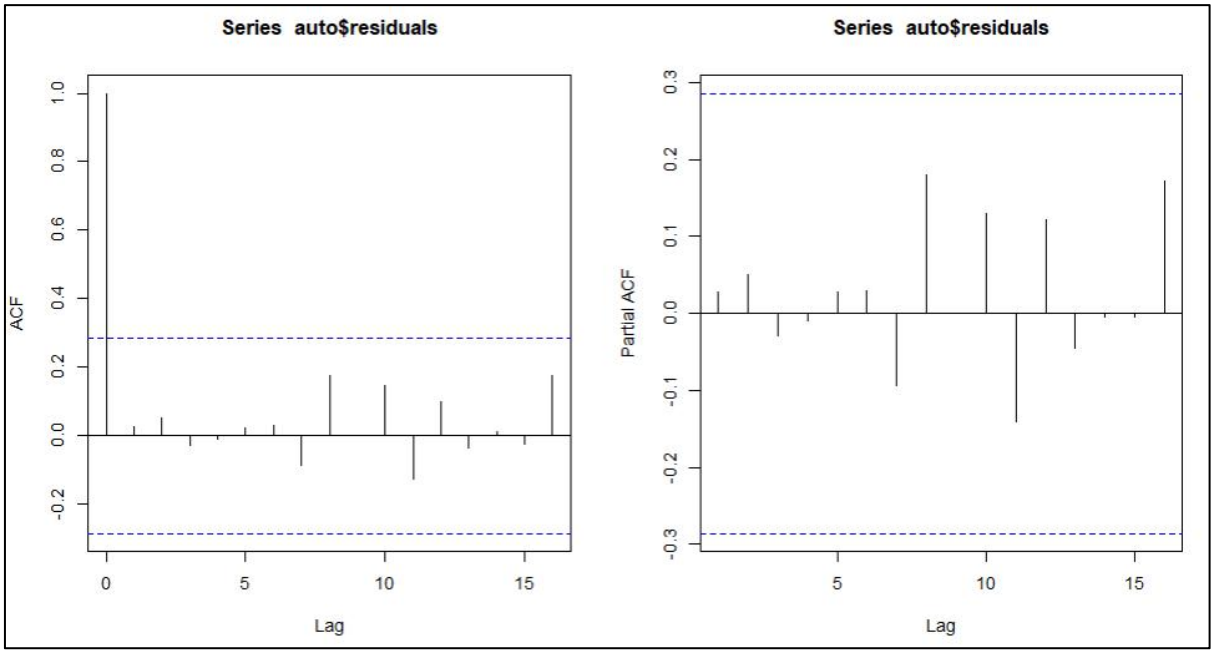
Leverage is not stationary.

```
> adf.test(data$r_leverage)    # p-value = 0.01705  --> stationary

        Augmented Dickey-Fuller Test

data:  data$r_leverage
Dickey-Fuller = -4.0164, Lag order = 3, p-value = 0.01705
alternative hypothesis: stationary
```

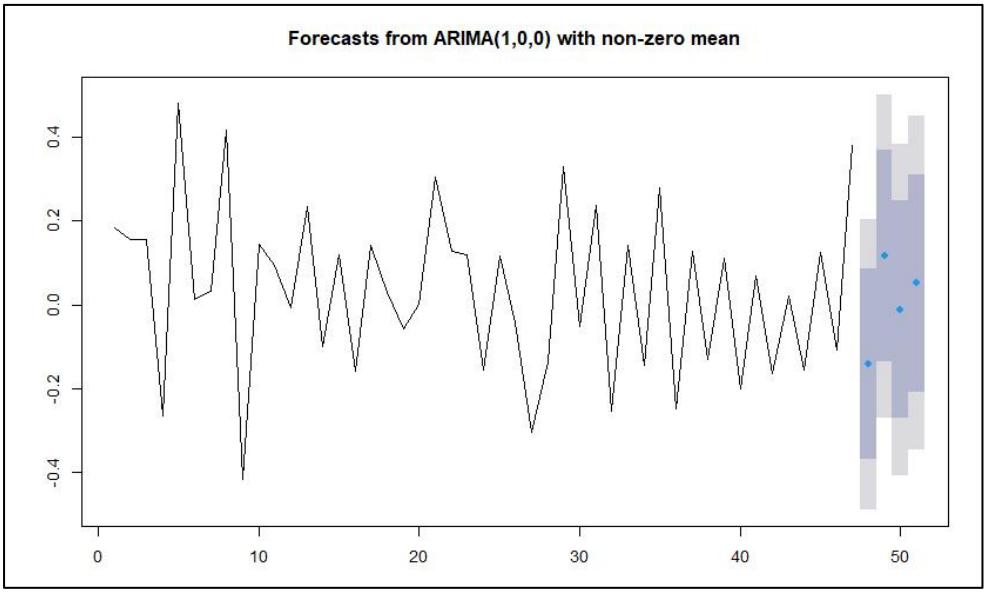
Return rate of leverage is stationary.



The best ARIMA model is ARIMA(1,0,0) based on auto-arima function in R studio.

Predict return rate of leverage for 4 quarters in 2022 by ARIMA model

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
48	-0.14212133	-0.3694436	0.08520092	-0.4897807	0.2055381
49	0.11609122	-0.1373446	0.36952703	-0.2715054	0.5036878
50	-0.01118045	-0.2705638	0.24820290	-0.4078731	0.3855122
51	0.05155112	-0.2092567	0.31235892	-0.3473200	0.4504222



Check accuracy of ARIMA model

```
> accuracy(fcastauto) #train set
                      ME      RMSE       MAE      MPE     MAPE      MASE       ACF1
Training set 0.001191793 0.1735654 0.1416214 9.484734 204.6088 0.4926666 0.02692448
```

8. Explain how Random forest can be used for prediction

There is two ways to use Random Forest algorithm for prediction of leverage of 4 quarters in 2022, includes:

8.1. Random Forest Classification

Classification algorithm requires “leverage” - target variable must be categorical and the data set has more than 1 independent variables. Therefore, we can separate “leverage” variable by two sides such as “increase” and “decrease”. Then we can conduct build Random Forest model for prediction of the next 4 quarters.

The selection of the final output follows the majority-voting system. In this case, the output chosen by the majority of the decision trees becomes the final output of the Random Forest system. Increasing the number of trees increases the precision of the outcome.

8.2. Random Forest Regression

Random Forest Regression allows “leverage” - target variable to be numerical, however, the data set must be only include a target variable and a independent variable. Therefore, we need to choose the most important independent variable for target variable and use this independent variable to predict by Random Forest Regression.

In a random forest regression, each tree produces a specific prediction. The mean prediction of the individual trees is the output of the regression. This is contrary to random forest classification, whose output is determined by the mode of the decision trees’ class.

Reference

1. Vithessonthi, C., & Tongurai, J. 2015. *The effect of firm size on the leverage–performance relationship during the financial crisis of 2007–2009*. Journal of Multinational Financial Management, 29,1-29.
2. Chang, K., Ding, J., Lou, Q., Li, Z., & Yang, J. 2021. *The impact of capital leverage on green firms' investment: New evidence regarding the size and age effects of Chinese green industries*. Finance Research Letters, p. 38.
3. D'Mello. R., Gruskin. M., & Kulchania.M. 2018. *Shareholders valuation of long-term debt and decline in firms' leverage ratio*. Journal of Corporate Finance, 48, 352-374.
4. Jermias, J., & Yigit, F. 2019. *Factors affecting leverage during a financial crisis: Evidence from Turkey*. Borsa Istanbul Review, 19, 171-185.
5. Kraus, A., & Litzenberger, R.H. 1973. *A State-Reference Model of Optimal Financial Leverage*. The Journal of Finance, 28, 911-922.
6. Myers, S. C., & Majluf, N. S. 1984. *Corporate Financing and Investment Decisions when Firms Have Information that Investors Do Not Have.* Journal of Financial Economics, 13, 187-221.
7. Harris, M. and Raviv, A. 1991. *The Theory of Capital Structure*. Journal of Finance, 46, 297-355.
8. Nghia, L., T., T., & Hung, P., M. 2016. *Cac nhan to anh huong den don bay tai chinh cua ngan hang thuong mai Viet Nam va mot so khuyen nghi*. Tap chi khoa hoc va dao tao ngan hang, 173.
9. Son, T., H. 2013. *Xay dung cau truc von muc tieu cho cac doanh nghiep san xuat cong nghiep tai TP Ho Chi Minh*.
10. Tai, T. (2021). *Tieu chi xac dinh doanh nghiep nho va vua moi nhat*. Tin tuc phap luat. Available at https://thuvienphapluat.vn/tintuc/vn/thoi-su-phap-luat/chinh-sach-moi/37296/tieu-chi-xac-dinh-doanh-nghiep-nho-va-vua-moi-nhat
11. Harc, M. (2015). *The relationship between tangible assets and capital structure of small and medium-sized companies in Croatia*.
12. Andersson, A., & Minnema, J. (2018). *The relationship between leverage and profitability.*