VIETNAM NATIONAL UNIVERSITY HCMC
UNIVERSITY OF ECONOMICS AND LAW

FINAL PROJECT
PACKAGE FOR FINANCIAL APPLICATION 2

Course:	Package for financial application 2
Lecturer:	PhD. Nguyễn Thanh Liêm
Student ID:	K214140961
Name:	Phạm Duy Tùng

TP. Hồ Chí Minh, ngày 03 tháng 06 năm 2023

CONTENT

ACKNOWLEDGEMENTS

First, I am sincerely grateful for my lecture dedication, time, and effort in teaching and guiding me through the completion of the Financial Application Package 2 course and this final project.

This has truly been an amazing journey for me and definitely one of my most favorite courses in the Fintech program as I get to learn code and explore the new language R and its ability to analyze data in a more efficient and organized manner compared to Python.

And as a student of yours, I truly appreciate your commitment and dedication to every teaching session.

From your student.

Thank you and warmest regards./.

Determine assigned topic:

If your student ID ends with an even number => investment (capital expenditure/total assets)

If your student ID ends with an odd number => debt maturity structure (long debt/total debt)

The assigned student ID number ends with 961, which is part of K214140961. Therefore, the chosen topic for the project is: Debt maturity structure (long debt/total debt).

CHAPTER 1: PERFORM LITERATURE REVIEW

1.1. Literature review

a) Leverage:

Méndez, V. (2013). Firm and country determinants of debt maturity: International evidence: Results for 39 countries indicate that firm-level variables such as leverage, size, firm quality, and asset maturity affect debt maturity structure.

Majumdar, R. (2010). The Determinants of Corporate Debt Maturity: A Study of Indian Firms. *The IUP Journal of Applied Finance*: The study results suggest that collateralizable assets and leverage are the important determinants of debt maturity choice.

Awartani, B., Belkhir, M., Boubaker, S., & Maghyereh, A. (2016). Corporate Debt Maturity in the MENA Region: Does Institutional Quality Matter? *Development Economics: Regional & Country Studies eJournal*: Consistent with the predictions of debt maturity theories and prior empirical findings, we find that leverage, firm size, and asset tangibility are positively associated with the use of more long-term debt while firms facing a higher risk of default tend to use more short-term debt.

Conclusion: These studies suggest that leverage has a positive effect on debt maturity structure, with higher leverage generally leading to more long-term debt and higher risk of default resulting in more short-term debt.

b) Cash holding:

Su, K., & Li, P. (2013). The Effects Of Ultimate Controlling Shareholders On Debt Maturity Structure. *Journal of Applied Business Research*: Ultimate controlling shareholders' cash flow rights are positively related to debt maturity structure.

Memon, Z., Chen, Y., Tauni, M., & Ali, H. (2018). The impact of cash flow volatility on firm leverage and debt maturity structure: evidence from China. *China Finance Review International*: Research implications this study advocates that cash flow volatility is an essential factor for determining both the debt levels and firm's term-to-maturity structure.

Brick, I., & Liao, R. (2017). The joint determinants of cash holdings and debt maturity: the case for financial constraints. *Review of Quantitative Finance and Accounting*: We find that there is a positive relation between debt maturity and cash holdings.

Conclusion: These studies suggest that cash holdings positively affect debt maturity structure and are influenced by factors such as controlling shareholders' cash flow rights and cash flow volatility.

c) Asset tangibility:

Awartani, B., Belkhir, M., Boubaker, S., & Maghyereh, A. (2016). Corporate Debt Maturity in the MENA Region: Does Institutional Quality Matter? *Development Economics: Regional & Country Studies eJournal*: Consistent with the predictions of debt maturity theories and prior empirical findings, we find that leverage, firm size, and asset tangibility are positively associated with the use of more long-term debt while firms facing a higher risk of default tend to use more short-term debt.

Mendoza, J., & Yelpo, S. (2016). Does managerial discretion affect debt maturity in Chilean firms? An agency cost and asymmetric information approach. *Ecos de Economía*: Other conclusions suggest that debt maturity is positively related to firm size, capital structure, and asset tangibility and negatively related to agency costs and membership in a holding company.

Conclusion: These studies suggest that debt maturity structure is positively influenced by factors such as asset tangibility, firm size, and leverage, while negatively affected by agency costs and default risk.

d) Liquidity risk

Chen, H., Xu, Y., & Yang, J. (2012). Systematic Risk, Debt Maturity, and the Term Structure of Credit Spreads. *S&P Global Market Intelligence Research Paper Series*: With both default risk and liquidity costs changing over the business cycle, our calibrated model implies that debt maturity is pro-cyclical, firms with high systematic risk favor longer debt maturity, and that these firms will have more stable maturity structures over the cycle.

Stephan, A., Talavera, O., & Tsapin, A. (2011). Corporate debt maturity choice in emerging financial markets. *The Quarterly Review of Economics and Finance*: This study provides strong evidence that constrained and unconstrained companies react differently to liquidity risk and, hence, pursue different debt maturity strategies.

Dang, V. (2007). Leverage, Debt Maturity and Firm Investment: An Empirical Analysis. *EFA 2008 Athens Meetings (Archive)*: There is a positive relation between leverage and debt maturity as predicted by the liquidity risk hypothesis.

Conclusion: These studies suggest that liquidity risk affects debt maturity structure, with firms having different strategies based on their systematic risk and constraints, and shorter debt maturity potentially increasing exposure to credit and liquidity shocks.

1.2. Dependent variable

Debt maturity structure

Debt maturity structure refers to the composition or proportion of a company's total debt that is classified as long-term debt. It is calculated by dividing the amount of long-term debt by the total debt of the company. The debt maturity structure provides insight into the proportion of debt that has a longer repayment period compared to short-term debt.

Formula: Debt Maturity Structure = Long-term Debt / Total Debt

The debt maturity structure indicates the extent to which a company relies on long-term debt financing for its operations. A higher debt maturity structure suggests that a significant portion of the company's debt is spread over a longer period, potentially reducing short-term liquidity risks. Conversely, a lower debt maturity structure indicates a larger proportion of short-term debt, which may lead to higher refinancing risks and greater sensitivity to interest rate changes.

1.3. Independent variables

Cash holding:

Cash holding is a financial metric that represents the amount of cash and cash equivalents, including short-term investments, held by a company. It is calculated by summing up the values of cash, cash equivalents, and short-term investments.

Formula: Cash Holding = Cash + Cash Equivalents + Short-term Investments

Cash holding is used to evaluate a company's liquidity and its ability to meet short-term obligations. It provides insight into the company's cash management practices, financial stability, and potential for investment or expansion. High cash holding may indicate a conservative financial approach, while low cash holding may suggest a higher degree of financial risk or investment activities. It is often analyzed in comparison to the company's overall financial position and industry benchmarks.

Leverage:

Leverage is a financial ratio that measures the extent to which a company uses debt to finance its assets. It is calculated by dividing the total debt of a company by its total assets.

Formula: Leverage = Total Debt / Total Assets

Leverage provides insight into the financial risk and stability of a company. A higher leverage ratio indicates a greater reliance on debt financing, which can amplify returns but also increase financial vulnerability. It is commonly used by investors and creditors to assess a company's ability to meet its debt obligations and evaluate its financial health. A lower leverage ratio implies a lower level of debt relative to assets, which generally indicates a more conservative financial position.

Asset tangibility:

Asset tangibility is a financial metric that represents the portion of a company's total assets that is tangible and can be physically identified or valued. It is calculated by subtracting the total debt, intangible assets, long-term financial investments, and other long-term investments from the total assets.

Formula: Asset Tangibility = Total Assets - Total Debt - Intangible Assets - Long-term Financial Investment - Other Long-term Investment

Asset tangibility is used to assess the composition of a company's assets and determine the proportion of tangible assets it possesses. A higher asset tangibility ratio indicates a larger portion of tangible assets, which can signify a more stable and less risky financial position. It is often considered in investment analysis and valuation to evaluate the asset base and risk profile of a company.

Liquidity risk:

Liquidity risk is a financial metric that measures a company's ability to meet its short-term financial obligations. It is calculated by subtracting inventory and other short-term assets from short-term assets, and then dividing the result by short-term debt.

Formula: Liquidity Risk = (Short-term Assets - Inventory - Other Short-term Assets) / Short-term Debt

Liquidity risk helps assess the adequacy of a company's short-term liquidity position. A higher liquidity risk ratio indicates a higher proportion of short-term debt relative to available short-term assets, which may suggest a potential challenge in meeting short-term obligations. It is often used by investors, creditors, and analysts to evaluate a company's liquidity profile and assess its ability to manage short-term cash flow needs.

1.4. Dataset

Company description:

Full Name: CMC Technology Group Joint Stock Company - CMC Corporation

Short name: CMC Corp

Address: CMC Tower - 11 Duy Tan Street, Dich Vong Hau Ward, Cau Giay District, Hanoi.

Website: https://www.cmc.com.vn

About the company: CMG is a prominent company operating in the information technology and telecommunications industry in Vietnam. It offers a wide range of services including software services, software outsourcing, system integration, IT services, computer manufacturing, telecommunications infrastructure services, data services, customer care services, e-commerce, digital content services, and more.

Dataset description:

Source: Cafef.vn (dataset was imported directly to excel file)

Company: CMC Technology Group Joint Stock Company (CMG)

Document type: Quarterly Balance Sheet

Format: Excel format (xlsx)

Time period: First quarter of 2010 to fourth quarter of 2022 (Q1-2010 to Q4-2022)

Balance Sheet components: The dataset includes information on various financial components such as assets, liabilities, equity, cash holdings, debt maturity structure, leverage, asset tangibility, liquidity risk, and other relevant financial metrics.

Provided link: https://s.cafef.vn/hose/CMG-cong-ty-co-phan-tap-doan-cong-nghe-cmc.chn

CHAPTER 2: PERFORM CODING TASKS

2. Data collection and input

Import libraries:

```
library(dplyr) # for manipulate data
library(readxl) # library use to read excel file
library(ggplot2) # library for plotting and data visualization
library(tidyverse) # for all the other library
library(car) # for VIF use
library(forecast)
# this package is included in the program because it contains auto.arima which
# is mentioned in the document Phan 7 Time series regression
```

Note: The "forecast" package is introduced in course as it is mentioned in the document "Phần 7 Time series regression" as it contains the auto.arima function needed for training and predicts the p, d, q values later.

Read the data from excel file:

```
df = read_excel("C:/Users/Admin/Desktop/courses/Gói phần mềm ứng dụng trong tài chính 2/final project/K214
View(df)

# According to the chosen variables, we just need to select a few rows in the dataset that we have
# So now Im gonna check for NA values, type of data and fill NA in the needed rows
```

Filter Cols: « ‹ 1 - 50 › »

	Thời gian	Q1 2010	Q2 2010	Q3 2010	Q4 2010	Q1 2011	Q2 2011
1	A- TÀI SẢN NGẮN HẠN	1483175070037	1483175070037	1483175070037	1483175070037	1062596397719	1062596397719
2	I. Tiền và các khoản tương đương tiền	62261127067	62261127067	62261127067	62261127067	62869375345	62869375345
3	1. Tiền	62261127067	62261127067	62261127067	62261127067	42369375345	42369375345
4	2. Các khoản tương đương tiền	NA	NA	NA	NA	20500000000	20500000000
5	II. Các khoản đầu tư tài chính ngắn hạn	1405225000	1405225000	1405225000	1405225000	21083568126	21083568126
6	1. Chứng khoán kinh doanh	NA	NA	NA	NA	NA	NA
7	2. Dự phòng giảm giá chứng khoán kinh doanh	NA	NA	NA	NA	NA	NA
8	3. Đầu tư nắm giữ đến ngày đáo hạn	NA	NA	NA	NA	NA	NA
9	III. Các khoản phải thu ngắn hạn	932355295287	932355295287	932355295287	932355295287	544584247026	544584247026
10	1. Phải thu ngắn hạn của khách hàng	822491548934	822491548934	822491548934	822491548934	480100130854	480100130854

Result of import dataset

Check and handle data:

According to the chosen variables, we just need to select a few variables (rows) in the dataset in order to calculate the needed variables. So now Im gonna check for NA values, type of data and fill NA in the needed rows.

```
# I. Check and handle data

## 1. Debt maturity structure
### 1.1. Long debt
#### check for NA values in long debt
sum(is.na(df[82,]))
#### check for the type of data of long debt
sum(!is.na(as.numeric(df[82,-1])))

### 1.2. Total debt
#### check for NA values in total debt
sum(is.na(df[66,]))
#### check for the type of data of total debt
sum(!is.na(as.numeric(df[66,-1])))
```

Check for NA values, numeric data type and fill values in Long debt and Total debt

This procedure is carried out similarly for the remaining variables: Debt maturity structure, Cash holding, Leverage, Assets tangibility and Liquidity risk. We verify if there are any NA values, check for numeric data type in that row, and fill those cells with appropriate data. Most of the cells are filled with 0 because the corresponding data in the company's balance sheet either does not exist or the company has not utilized that account.

Preprocess and create a new dataframe:

```
## 6. Create a new dataframe with just usable data
df <- df[c(82,67,3,4,5,66,64,43,52,58,1,18,21), ]
#### transpose the dataframe for manipulation
df = t(df)
#### rename columns
column_names <- c("longdebt",
                  "shortdebt",
                  "cash",
                  "equivalentcash",
                  "shortinvest",
                  "totaldebt",
                  "totalassets",
                  "intangibleassets",
                  "longfinancialinvest",
                  "otherlongassets",
                  "shortassets",
                  "inventory",
                  "othershortassets")
colnames(df) = column_names
#### remove the first row
df = df[-1,]
#### apply the df property
df <- as.data.frame(df)
#### turn character to numeric values
df <- as.data.frame(lapply(df, as.numeric))
#### create a list of time periods
listtimenames = list()
for (i in 2010:2022) {
  for (j in 1:4) {
    listtimenames = append(listtimenames, paste0("Q",j,i))
  }
}

df$Time <- listtimenames
df <- df[, c("Time", names(df)[-ncol(df)])]
#### view the new dataframe
View(df)
```

The following steps are the data preprocessing process to make the data usable in R. ***Note:*** The steps below are specific to the input dataset based on its unique characteristics and are not a general processing procedure for all different input datasets. The processing procedure for the current dataset goes through the following steps:

+ Extract the rows with necessary data from the dataframe of the main dataset using index indices.

+ Transpose the dataset to facilitate analysis based on distinct characteristics, where each column represents a specific characteristic, and each row represents observations with quarterly information.

+ Add column names for easy identification and referencing.

+ Delete the first row as it contains labels, not numerical data for analysis.

+ Convert the dataset format to an R dataframe.

+ Convert the data in the dataset to numeric format for computation in R.

+ Create a column with time data using a for loop.

+ Display the new data frame containing only the variables to be used.

	Time	longdebt	shortdebt	cash	equivalentcash	shortinvest	totaldebt	totalassets	intangibleassets	lon
1	Q12010	137088000000	1.104039e+12	62261130000	0	1.405225e+09	1.241127e+12	1.934682e+12	1607660000	
2	Q22010	137088000000	1.104039e+12	62261130000	0	1.405225e+09	1.241127e+12	1.934682e+12	1607660000	
3	Q32010	137088000000	1.104039e+12	62261130000	0	1.405225e+09	1.241127e+12	1.934682e+12	1607660000	
4	Q42010	137088000000	1.104039e+12	62261130000	0	1.405225e+09	1.241127e+12	1.934682e+12	1607660000	
5	Q12011	117409300000	8.245237e+11	42369380000	20500000000	2.108357e+10	9.419330e+11	1.530007e+12	1572936000	
6	Q22011	117409300000	8.245237e+11	42369380000	20500000000	2.108357e+10	9.419330e+11	1.530007e+12	1572936000	
7	Q32011	117409300000	8.245237e+11	42369380000	20500000000	2.108357e+10	9.419330e+11	1.530007e+12	1572936000	
8	Q42011	117409300000	8.245237e+11	42369380000	20500000000	2.108357e+10	9.419330e+11	1.530007e+12	1572936000	
9	Q12012	177946200000	8.737954e+11	41077240000	7075217000	2.156014e+10	1.051742e+12	1.672329e+12	18187410000	
10	Q22012	177946200000	8.737954e+11	41077240000	7075217000	2.156014e+10	1.051742e+12	1.672329e+12	18187410000	

Result of the preprocessed frame

Create a variable dataframe:

```r
# II. Create a variable dataframe
## 1. Debt maturity structure
df$debtmaturitystructure <- df$longdebt / df$totaldebt
variable_df <- data.frame(debtmaturitystructure = df$debtmaturitystructure)

## 2. Cash holding
df$cashholding <- df$cash + df$equivalentcash + df$shortinvest
variable_df <- cbind(variable_df, cashholding = df$cashholding)

## 3. Leverage
df$leverage <- df$totaldebt / df$totalassets
variable_df <- cbind(variable_df, leverage = df$leverage)

## 4. Assets tangibility
df$assetstangibility <- df$totalassets - df$totaldebt - df$intangibleassets - df$longfinancialinvest - df
variable_df <- cbind(variable_df, assetstangibility = df$assetstangibility)

## 5. Liquidity risk
df$liquidityrisk <- (df$shortassets - df$inventory - df$othershortassets) / df$shortdebt
variable_df <- cbind(variable_df, liquidityrisk = df$liquidityrisk)

## 6. View variable dataframe
View(variable_df)
```

Based on the formulas provided in the variable definition section, I will calculate the values and store them in a new dataframe called "variable_df." This dataframe will be the main dataframe used for the subsequent steps of the task.

The variable called names are corresponding with their actual names, are as follows: **debtmaturitystructure** (Debt Maturity Structure), **cashholding** (Cash Holding), **leverage** (Leverage), **assetstangibility** (Assets Tangibility) and **liquidityrisk** (Liquidity Risk).

	debtmaturitystructure	cashholding	leverage	assetstangibility	liquidityrisk
1	0.11045445	6.366636e+10	0.6415147	6.461722e+11	0.9021616
2	0.11045445	6.366636e+10	0.6415147	6.461722e+11	0.9021616
3	0.11045445	6.366636e+10	0.6415147	6.461722e+11	0.9021616
4	0.11045445	6.366636e+10	0.6415147	6.461722e+11	0.9021616

Result of the variable_df

3. Provide descriptive statistics

For the ENTIRE period, BEFORE and AFTER periods, do the following tasks:

3.1. Min, Max, Mean, Median, Standard deviation of all variables.

3.1.1. Entire period

```
# III. Provide descriptive statistics
## 1. Entire period
summary(variable_df)
stats_variable_df <- sapply(variable_df, function(x) c(min = min(x), max = max(x), mean = mean(x), median
print(stats_variable_df)
```

```
> # III. Provide descriptive statistics
> ## 1. Entire period
> summary(variable_df)
 debtmaturitystructure  cashholding          leverage        assetstangibility   liquidityrisk
 Min.   :0.07145        Min.   :6.367e+10   Min.   :0.4849   Min.   :4.691e+11   Min.   :0.6963
 1st Qu.:0.12920        1st Qu.:8.395e+10   1st Qu.:0.5373   1st Qu.:5.487e+11   1st Qu.:0.8771
 Median :0.19299        Median :3.672e+11   Median :0.5822   Median :8.892e+11   Median :0.9550
 Mean   :0.19202        Mean   :5.796e+11   Mean   :0.5771   Mean   :1.071e+12   Mean   :1.0228
 3rd Qu.:0.24437        3rd Qu.:1.356e+12   3rd Qu.:0.6227   3rd Qu.:1.827e+12   3rd Qu.:1.1466
 Max.   :0.30651        Max.   :1.658e+12   Max.   :0.6575   Max.   :2.148e+12   Max.   :1.5956
> stats_variable_df <- sapply(variable_df, function(x) c(min = min(x), max = max(x), mean = mean(x), m
edian = median(x), sd = sd(x)))
> print(stats_variable_df)
       debtmaturitystructure  cashholding   leverage assetstangibility liquidityrisk
min               0.07144617 6.366636e+10 0.48493134      4.690909e+11     0.6962541
max               0.30651396 1.657563e+12 0.65749945      2.147946e+12     1.5956157
mean              0.19201711 5.795517e+11 0.57707981      1.070810e+12     1.0227622
median            0.19298932 3.672469e+11 0.58224125      8.891836e+11     0.9550462
sd                0.06646804 5.714203e+11 0.04835681      5.911159e+11     0.2385642
```

3.1.2. Before period (all the quarters before 2020)

```
## 2. Before period (all the periods before 2020)
summary(variable_df[1:40,])
stats_variable_df_before <- sapply(variable_df[1:40,], function(x) c(min = min(x), max = max(x), mean = me
print(stats_variable_df_before)
```

```
> ## 2. Before period (all the periods before 2020)
> summary(variable_df[1:40,])
 debtmaturitystructure  cashholding           leverage       assetstangibility    liquidityrisk
 Min.    :0.07145       Min.    :6.367e+10   Min.    :0.4849   Min.    :4.691e+11   Min.    :0.6963
 1st Qu.:0.12465        1st Qu.:7.961e+10   1st Qu.:0.5387   1st Qu.:5.293e+11   1st Qu.:0.7968
 Median :0.16633        Median :2.026e+11   Median :0.6091   Median :7.305e+11   Median :0.9022
 Mean   :0.17320        Mean   :3.180e+11   Mean   :0.5861   Mean   :8.006e+11   Mean   :0.9551
 3rd Qu.:0.21075        3rd Qu.:3.939e+11   3rd Qu.:0.6242   3rd Qu.:9.402e+11   3rd Qu.:1.0477
 Max.   :0.30651        Max.   :1.356e+12   Max.   :0.6575   Max.   :1.838e+12   Max.   :1.5956
> stats_variable_df_before <- sapply(variable_df[1:40,], function(x) c(min = min(x), max = max(x), mea
n = mean(x), median = median(x), sd = sd(x)))
> print(stats_variable_df_before)
       debtmaturitystructure  cashholding  leverage assetstangibility liquidityrisk
min               0.07144617 6.366636e+10 0.4849313      4.690909e+11     0.6962541
max               0.30651396 1.355556e+12 0.6574995      1.838154e+12     1.5956157
mean              0.17319692 3.180376e+11 0.5861089      8.005945e+11     0.9550532
median            0.16632929 2.025573e+11 0.6091305      7.305353e+11     0.9021616
sd                0.06358009 3.459522e+11 0.0512694      3.584059e+11     0.2230511
>
```

3.1.3. After period (8 quarters after 2020)

```
## 3. After period (8 periods after 2020)
summary(variable_df[41:48,])
stats_variable_df_after <- sapply(variable_df[41:48,], function(x) c(min = min(x), max = max(x), mean = me
print(stats_variable_df_after)
```

```
> ## 3. After period (8 periods after 2020)
> summary(variable_df[41:48,])
 debtmaturitystructure  cashholding           leverage       assetstangibility    liquidityrisk
 Min.    :0.2428        Min.    :1.258e+12   Min.    :0.5230   Min.    :1.824e+12   Min.    :1.206
 1st Qu.:0.2512         1st Qu.:1.413e+12   1st Qu.:0.5363   1st Qu.:1.842e+12   1st Qu.:1.259
 Median :0.2601         Median :1.442e+12   Median :0.5424   Median :1.894e+12   Median :1.276
 Mean   :0.2658         Mean   :1.446e+12   Mean   :0.5451   Mean   :1.936e+12   Mean   :1.313
 3rd Qu.:0.2767         3rd Qu.:1.498e+12   3rd Qu.:0.5513   3rd Qu.:1.991e+12   3rd Qu.:1.362
 Max.   :0.2985         Max.   :1.613e+12   Max.   :0.5687   Max.   :2.148e+12   Max.   :1.467
> stats_variable_df_after <- sapply(variable_df[41:48,], function(x) c(min = min(x), max = max(x), mea
n = mean(x), median = median(x), sd = sd(x)))
> print(stats_variable_df_after)
       debtmaturitystructure  cashholding   leverage assetstangibility liquidityrisk
min               0.24281165 1.258043e+12 0.52297321      1.823962e+12    1.20600698
max               0.29846092 1.613184e+12 0.56866502      2.147946e+12    1.46668432
mean              0.26578760 1.445725e+12 0.54507135      1.935542e+12    1.31335641
median            0.26005142 1.442448e+12 0.54235273      1.893549e+12    1.27588074
sd                0.02031611 1.024686e+11 0.01602253      1.208367e+11    0.09088358
```

3.2. Give comments on the possible impact of Covid-19 pandemic.

From the given statistics of the financial variables before and after the COVID-19 period, we can observe the following impacts of the pandemic:

Debt Maturity Structure:

Before COVID: Minimum 0.0714, maximum 0.3065, mean 0.1732, median 0.1663, standard deviation 0.0636.

After COVID: Minimum increased to 0.2428, maximum decreased to 0.2985, mean 0.2658, median 0.2601, a slightly lower standard deviation of 0.0203.

Impact of COVID: The debt maturity structure seems to have increased after the COVID-19 period. This suggests that companies may have adjusted their debt obligations to have longer-term maturity, possibly to manage their financial risks during uncertain times.

Cash Holding:

Before COVID: Minimum 6.37e+10, maximum 1.36e+12, mean 3.18e+11, median 2.03e+11, with a relatively high standard deviation of 3.46e+11.

After COVID: Minimum increased to 1.26e+12, maximum increased to 1.61e+12, mean and median increased significantly to 1.45e+12 and 1.44e+12, respectively, standard deviation decreased to 1.02e+11.

Impact of COVID: The cash holding of companies significantly increased during the COVID-19 period. This suggests that companies were likely accumulating more cash reserves to strengthen their liquidity position and navigate the uncertainties and challenges posed by the pandemic.

Leverage:

Before COVID: Ratio ranged from 0.4849 to 0.6575, mean 0.5861, median 0.6091, standard deviation 0.0513.

After COVID: Ratio ranged from 0.5230 to 0.5687, mean 0.5451, median 0.5424, standard deviation decreased to 0.0160.

Impact of COVID: The leverage ratio remained relatively stable during and after the COVID-19 period, with only a slight decrease in the standard deviation. This suggests that the overall leverage of companies did not experience significant changes due to the pandemic.

Assets Tangibility:

Before COVID: Minimum 4.69e+11, maximum 1.83e+12, mean 8.01e+11, median 7.31e+11, standard deviation 3.58e+11.

After COVID: Minimum increased to 1.82e+12, maximum increased to 2.15e+12, mean and median increased significantly to 1.94e+12 and 1.89e+12, standard deviation increased slightly to 1.21e+11.

Impact of COVID: The assets tangibility of companies increased during the COVID-19 period, indicating a higher proportion of tangible assets in their overall asset structure. This could suggest that companies may have shifted their investments towards more physical assets to increase stability and reduce reliance on volatile markets.

Liquidity Risk:

Before COVID: Ranged from 0.6963 to 1.5956, mean 0.9551, median 0.9022, standard deviation 0.2231.

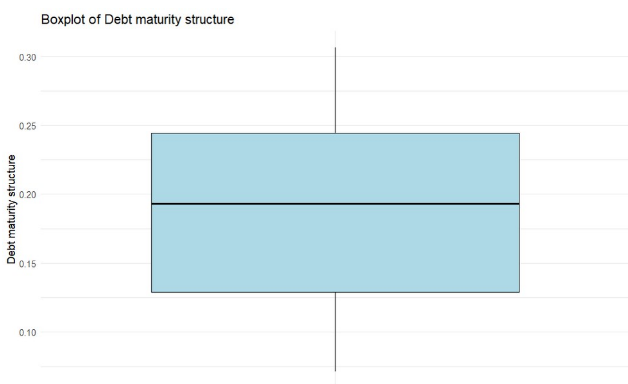
After COVID: Ranged from 1.2060 to 1.4667, mean 1.3134, median 1.2759, standard deviation decreased to 0.0909.

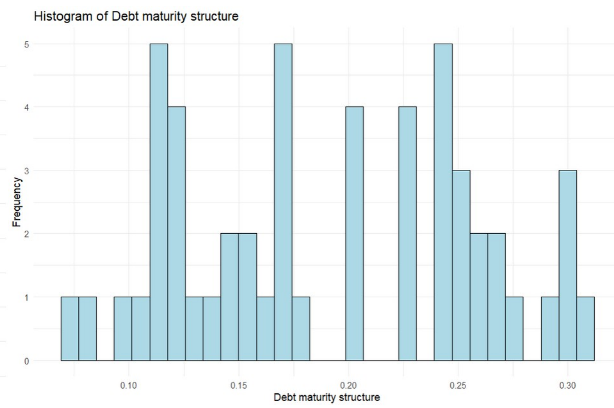
Impact of COVID: The liquidity risk of companies increased during the COVID-19 period, indicating higher vulnerability to liquidity challenges. This suggests that companies may have faced difficulties in accessing sufficient funds and maintaining their financial stability during the pandemic.

4. Provide box & whisker plot and histogram

4.1. Entire period

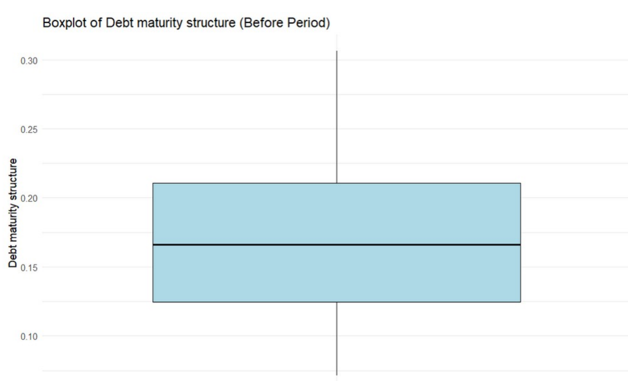
```r
# IV. Provide box & whisker plot and histogram of the debt maturity structure
## 1. Entire period
### Boxplot
ggplot(data = variable_df, aes(x = "", y = debtmaturitystructure)) +
  geom_boxplot(fill = "lightblue", color = "black") +
  labs(title = "Boxplot of Debt maturity structure", x = "", y = "Debt maturity structure") +
  theme_minimal()
### Histogram
ggplot(data = variable_df, aes(x = debtmaturitystructure)) +
  geom_histogram(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Debt maturity structure", x = "Debt maturity structure", y = "Frequency") +
  theme_minimal()
```

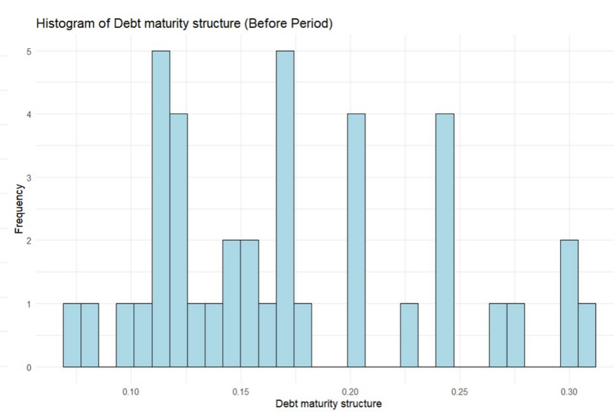




4.2. Before period

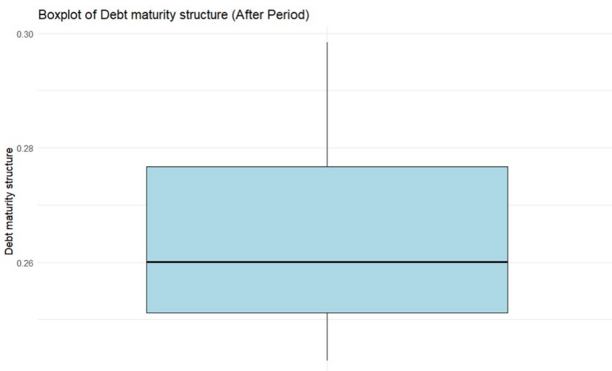
```r
## 2. Before period (all the periods before 2020)
### Boxplot
ggplot(data = variable_df[1:40,], aes(x = "", y = debtmaturitystructure)) +
  geom_boxplot(fill = "lightblue", color = "black") +
  labs(title = "Boxplot of Debt maturity structure (Before Period)", x = "", y = "Debt maturity structure'
  theme_minimal()
### Histogram
ggplot(data = variable_df[1:40,], aes(x = debtmaturitystructure)) +
  geom_histogram(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Debt maturity structure (Before Period)", x = "Debt maturity structure", y =
  theme_minimal()
```

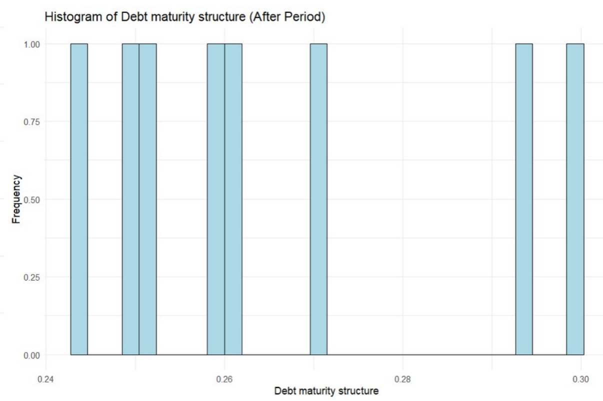




4.3. After period

```r
## 3. After period (8 periods after 2020)
### Boxplot
ggplot(data = variable_df[41:48,], aes(x = "", y = debtmaturitystructure)) +
  geom_boxplot(fill = "lightblue", color = "black") +
  labs(title = "Boxplot of Debt maturity structure (After Period)", x = "", y = "Debt maturity structure"
  theme_minimal()
### Histogram
ggplot(data = variable_df[41:48,], aes(x = debtmaturitystructure)) +
  geom_histogram(fill = "lightblue", color = "black") +
  labs(title = "Histogram of Debt maturity structure (After Period)", x = "Debt maturity structure", y = '
  theme_minimal()
```





4.4. Give comments on the shape of the distribution and any implications drawn.

Based on the graphs, we can draw the following conclusions:

+ Throughout the entire period, the debt maturity structure is distributed in the range of 0.1 to 0.3, with a concentration around 0.125 to 0.25 and a median of 0.2. The frequency distribution is evenly spread over that range and not heavily concentrated on a specific value, indicated by three modes around 0.11, 0.2, and 0.3. This reflects that the company has a flexible and diverse mechanism in using debt maturity structure to effectively manage long-term operations.

+ However, in the boxplot of the debt maturity structure before the onset of covid, there is a higher level of concentration from 0.125 to 0.21, with a lower median below 0.175. In the histogram chart of the pre-covid period, some data values above 0.2 are missing. This suggests that during the covid period, the company had a higher usage of debt maturity structure compared to before. This observation is confirmed by the histogram and boxplot of the covid period, showing a range of values from 0.24 to 0.3, evenly distributed with consistent frequencies.

These observations align with the above assumption that the debt maturity structure seems to have increased after the COVID-19 period. This suggests that companies may have adjusted their debt obligations to have longer-term maturity, possibly to manage their financial risks during uncertain times.

5. Perform multiple regression

5.1. Model 1: With all the individual variables

```r
# V. Perform multiple regression
## 1. Model 1: With all the individual variables
### 1.1. Running the model
model1 <- lm(debtmaturitystructure ~
               cashholding +
               leverage +
               assetstangibility +
               liquidityrisk,data = variable_df)
summary(model1, alpha = 0.1)
```

```
Call:
lm(formula = debtmaturitystructure ~ cashholding + leverage +
    assetstangibility + liquidityrisk, data = variable_df)

Residuals:
      Min        1Q    Median        3Q       Max
-0.090296 -0.026182  0.005582  0.019155  0.082484

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)       -1.445e-01  1.296e-01  -1.115 0.270510
cashholding        2.764e-13  6.273e-14   4.406 6.06e-05 ***
leverage           4.726e-01  1.576e-01   3.000 0.004315 **
assetstangibility -2.203e-13  5.744e-14  -3.835 0.000373 ***
liquidityrisk      1.364e-01  5.051e-02   2.701 0.009579 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03951 on 47 degrees of freedom
Multiple R-squared:  0.6744,    Adjusted R-squared:  0.6466
F-statistic: 24.33 on 4 and 47 DF,  p-value: 5.969e-11
```

Interpret:

b0 = -1.445e-1 with the standard error of 1.296e-1, and the effect of:

+ Cash holding on Debt maturity structure: (b1) = 2.764e-13, with the standard error of 6.273e-14 and p-value of 6.06e-5 (<0.1 requirement value)

+ Leverage on Debt maturity structure: (b2) = 4.726e-1, with the standard error of 1.576e-1 and p-value of 0.004315 (<0.1 requirement value)

+ Assets tangibility on Debt maturity structure: (b3) = -2.203e-13, with the standard error of 5.744e-14 and p-value of 0.000373 (<0.1 requirement value)

+ Liquidity risk on Debt maturity structure: (b4) = 1.364e-1, with the standard error of 5.051e-2 and p-value of 0.009579 (<0.1 requirement value)

+ R-squard = 0.6744 and Adjusted R-squared = 0.6466, with the p-value of 5.969e-11 (<0.1 requirement value)

Conclusion: Cash holding, Leverage and Liquidity risk have a statistically significant positive effect on debt maturity structure, while assets tangibility has a statistically significant negative influence.

The adjusted R-squared value of 0.6466 indicates that the independent variables explain approximately 64.66% of the variation in the debt maturity structure. The F-statistic is significant (p < 0.001), indicating that the overall model is statistically significant in predicting the debt maturity structure.

5.2. Model 2: With the usual individual variables and the interaction

```
## 2. Model 2: With the usual individual variables and the interaction
### 2.1. Add the Covid 19 dummy variable (0 = not covid year, 1 = covid year)
covid_dummy=list()
for (i in 1:52) {
  if (i <= 40) {
    covid_dummy <- append(covid_dummy, 0)
  } else if (i<=48) {
    covid_dummy <- append(covid_dummy, 1)
  } else {
    covid_dummy <- append(covid_dummy, 0)
  }
}
variable_df$covid <- covid_dummy
variable_df$covid <- as.numeric(variable_df$covid)
```

```
### 2.2. Running the model
model2 <- lm(debtmaturitystructure ~
               cashholding +
               leverage +
               assetstangibility +
               liquidityrisk +
               covid:cashholding +
               covid:leverage +
               covid:assetstangibility +
               covid:liquidityrisk, data = variable_df)
summary(model2, alpha = 0.1)
```

```
Call:
lm(formula = debtmaturitystructure ~ cashholding + leverage +
    assetstangibility + liquidityrisk + covid:cashholding + covid:leverage +
    covid:assetstangibility + covid:liquidityrisk, data = variable_df)

Residuals:
      Min        1Q    Median        3Q       Max
-0.092915 -0.019364  0.003015  0.015016  0.081769

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             -1.431e-01  1.320e-01  -1.084 0.284510
cashholding              3.135e-13  7.005e-14   4.475 5.53e-05 ***
leverage                 4.885e-01  1.612e-01   3.031 0.004115 **
assetstangibility       -2.569e-13  6.396e-14  -4.017 0.000233 ***
liquidityrisk            1.436e-01  5.264e-02   2.728 0.009183 **
cashholding:covid       -3.037e-13  2.225e-13  -1.365 0.179313
leverage:covid          -5.109e-01  1.673e+00  -0.305 0.761612
assetstangibility:covid  3.072e-13  3.427e-13   0.897 0.374956
liquidityrisk:covid      9.209e-02  2.294e-01   0.401 0.690127
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04013 on 43 degrees of freedom
Multiple R-squared:  0.6927,    Adjusted R-squared:  0.6355
F-statistic: 12.12 on 8 and 43 DF,  p-value: 7.318e-09
```

Interpret:

b0 = -1.431e-1 with the standard error of 1.32e-1, and the effect of:

+ Cash holding on Debt maturity structure: (b1) = 3.135e-13, with the standard error of 7.005e-14 and p-value of 5.53e-05 (<0.1 requirement value)

+ Leverage on Debt maturity structure: (b2) = 4.885e-01, with the standard error of 1.612e-01 and p-value of 0.004115 (<0.1 requirement value)

+ Assets tangibility on Debt maturity structure: (b3) = -2.569e-13, with the standard error 6.396e-14 and p-value of 0.000233 (<0.1 requirement value)

+ Liquidity risk on Debt maturity structure: (b4) = 1.436e-01, with the standard error of 5.264e-02 and p-value of 0.009183 (<0.1 requirement value)

+ Cash holding and Covid on Debt maturity structure: (b1) = -3.037e-13, with the standard error of 2.225e-13 and p-value of 0.179313 (>0.1 requirement value)

+ Leverage and Covid on Debt maturity structure: (b2) = -5.109e-01, with the standard error of 1.673e+00 and p-value of 0.761612 (>>0.1 requirement value)

+ Assets tangibility and Covid on Debt maturity structure: (b3) = 3.072e-13, with the standard error 3.427e-13 and p-value of 0.374956 (>>0.1 requirement value)

+ Liquidity risk and Covid on Debt maturity structure: (b4) = 9.209e-02, with the standard error of 2.294e-01 and p-value of 0.690127 (>>0.1 requirement value)

+ R-squard = 0.6927 and Adjusted R-squared = 0.6355, with the p-value of 7.318e-09 (<0.1 requirement value)

Conclusion: In the pre-COVID period, the variables cash holding, leverage, assets tangibility, and liquidity risk have similar effects on the debt maturity structure as mentioned in the previous analysis.

During the COVID-19 period, the coefficients for the interaction variables are not statistically significant, indicating that the relationships between these variables and the debt maturity structure did not change significantly during that period.

The adjusted R-squared value of 0.6355 indicates that the independent variables explain approximately 63.55% of the variation in the debt maturity structure, considering both the pre-COVID and COVID-19 periods. The F-statistic is significant (p < 0.001), suggesting that the overall model is statistically significant in predicting the debt maturity structure. But 63.55% is less than 64.66% as model 1 above, so it is not necessary to add the interaction variables in the model, therefore model 1 is the optimal model.

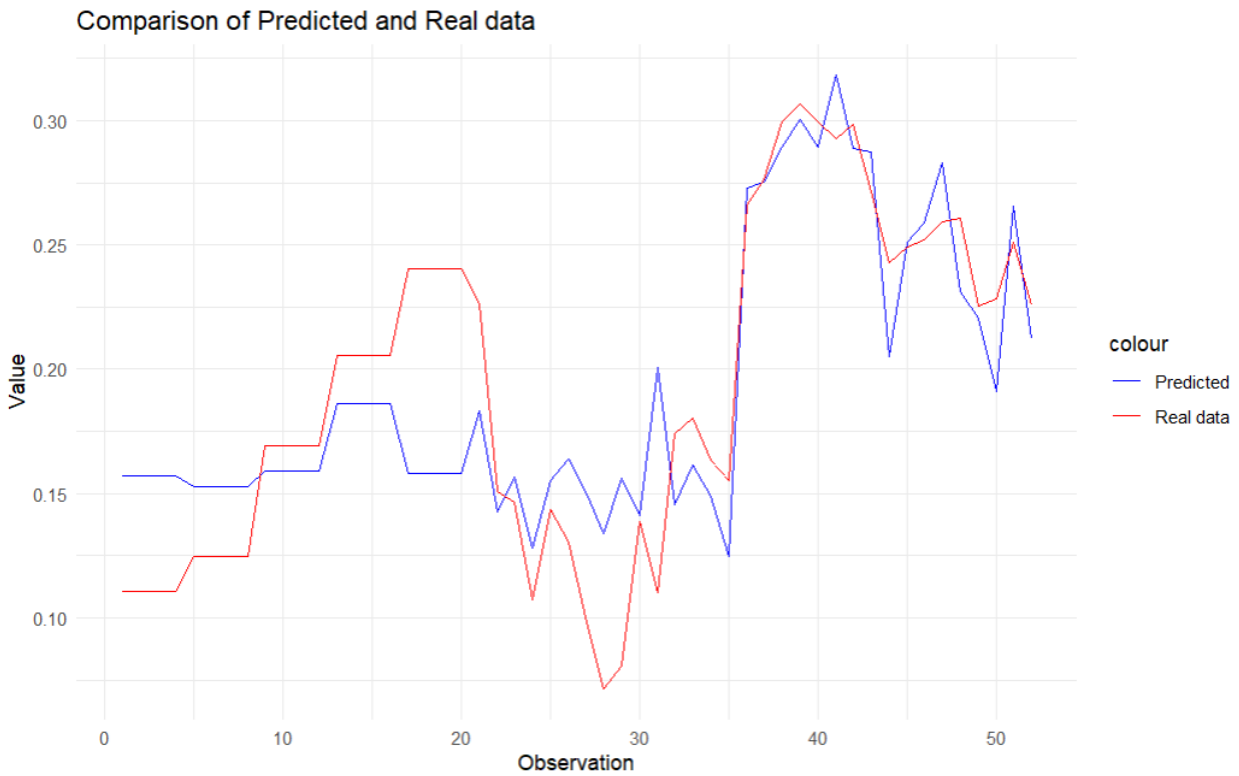
5.3. Predict the debt maturity structure using Model 1

```
## 3. Predict the debt maturity structure using Model 1
### 3.1. Predict the variable
predictions <- predict(model1, data = variable_df)
print(predictions)
variable_df_with_predictions <- cbind(variable_df, predictions)
View(variable_df_with_predictions)
```

Filter

	debtmaturitystructure	cashholding	leverage	assetstangibility	liquidityrisk	covid	predictions
1	0.11045445	6.366636e+10	0.6415147	6.461722e+11	0.9021616	0	0.1569746
2	0.11045445	6.366636e+10	0.6415147	6.461722e+11	0.9021616	0	0.1569746
3	0.11045445	6.366636e+10	0.6415147	6.461722e+11	0.9021616	0	0.1569746
4	0.11045445	6.366636e+10	0.6415147	6.461722e+11	0.9021616	0	0.1569746
5	0.12464719	8.395295e+10	0.6156397	5.486660e+11	0.7623029	0	0.1527528
6	0.12464719	8.395295e+10	0.6156397	5.486660e+11	0.7623029	0	0.1527528
7	0.12464719	8.395295e+10	0.6156397	5.486660e+11	0.7623029	0	0.1527528

```r
### 3.2 Plot the comparison graph
ggplot(variable_df_with_predictions, aes(x = 1:length(predictions))) +
  geom_line(aes(y = predictions, color = "Predicted")) +
  geom_line(aes(y = debtmaturitystructure, color = "Real data")) +
  labs(x = "Observation", y = "Value") +
  scale_color_manual(values = c("Predicted" = "blue", "Real data" = "red")) +
  theme_minimal()
```



6. Perform ARIMA model to predict the variable

Before starting the model training, we divided the available dataset into two parts: the training set and the test set. The training set consists of all observations from 2010 to the end of 2021, while the test set consists of four data samples from the year 2022, with the debtmaturitystructure variable omitted.

```r
# VI. Train and use ARIMA model to predict
## 1. Create the train and test dataset
ts_data <- ts(variable_df[1:48,], frequency = 4)
train_data <- variable_df[1:48,]
test_data <- variable_df[49:52, -1]
View(train_data)
View(test_data)
```

```r
## 2. Fit the ARIMA model using auto.arima
arima_model <- auto.arima(train_data$debtmaturitystructure)

## 3. Forecast variable for 4 quarters of 2022
forecast <- forecast(arima_model, h = 4)

## 4. Extract the predicted values
predicted_values <- forecast$mean

## 5. Compare the predicted variable with realistic data
comparison_data <- variable_df[49:52,]$debtmaturitystructure
comparison_df <- data.frame(Actual = comparison_data, Predicted = predicted_values)

# 6. Print the comparison data
print(comparison_df)
```

The steps to perform the ARIMA model are as follows:

+ Fit the training dataset into the ARIMA model using auto.arima to automatically determine the values of the model's coefficients (p, d, q). Specify the dependent variable of the model as debtmaturitystructure.

+ Perform predictions with the ARIMA model using the forecast function, where the coefficient h = 4 represents the time horizon of the data in quarters.

+ Create a dataframe that contains the predicted values and the actual values. Print the dataframe to compare the predictions with the actual values.

```
> # 6. Print the comparison data
> print(comparison_df)
     Actual  Predicted
1 0.2254500   0.260839
2 0.2281671   0.260839
3 0.2510913   0.260839
4 0.2260039   0.260839
>
```

7. Explain Decision Tree algorithm

Explain in fewer than 150 words how Decision Tree algorithm can be used to make prediction whether the firm will increase/decrease the variable of the assigned topic. No coding is required for this task:

The Decision Tree algorithm could be used to predict whether a firm's debt maturity structure will increase or decrease using cash holding, leverage, assets tangibility, and liquidity risk variables. It constructs a tree-like model, making decisions based on these variables. At each node, the algorithm selects the best variable to split the data, minimizing impurity or maximizing information gain. This recursive process continues until a stopping criterion is met. To make predictions, new instances traverse the decision tree, following branches based on variable values. The majority class in the reached leaf node determines the prediction. By training the algorithm on historical data, it learns patterns and relationships between variables and Debt Maturity Structure. Then, it predicts whether the firm will increase or decrease its Debt Maturity Structure based on the provided variables.

REFERENCES

Méndez, V. (2013). Firm and country determinants of debt maturity: International evidence

Majumdar, R. (2010). The Determinants of Corporate Debt Maturity: A Study of Indian Firms. *The IUP Journal of Applied Finance*

Awartani, B., Belkhir, M., Boubaker, S., & Maghyereh, A. (2016). Corporate Debt Maturity in the MENA Region: Does Institutional Quality Matter? *Development Economics: Regional & Country Studies eJournal*

Su, K., & Li, P. (2013). The Effects Of Ultimate Controlling Shareholders On Debt Maturity Structure. *Journal of Applied Business Research*

Memon, Z., Chen, Y., Tauni, M., & Ali, H. (2018). The impact of cash flow volatility on firm leverage and debt maturity structure: evidence from China. *China Finance Review International*

Brick, I., & Liao, R. (2017). The joint determinants of cash holdings and debt maturity: the case for financial constraints. *Review of Quantitative Finance and Accounting*

Mendoza, J., & Yelpo, S. (2016). Does managerial discretion affect debt maturity in Chilean firms? An agency cost and asymmetric information approach. *Ecos de Economía*

Chen, H., Xu, Y., & Yang, J. (2012). Systematic Risk, Debt Maturity, and the Term Structure of Credit Spreads. *S&P Global Market Intelligence Research Paper Series*

Stephan, A., Talavera, O., & Tsapin, A. (2011). Corporate debt maturity choice in emerging financial markets. *The Quarterly Review of Economics and Finance*

Dang, V. (2007). Leverage, Debt Maturity and Firm Investment: An Empirical Analysis. *EFA 2008 Athens Meetings (Archive)*