

HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY
SCHOOL OF ECONOMIC AND MANAGEMENT

-----o0o-----



BUSINESS FORECASTING
PROJECT 4 – GROUP 3

Full Name	Student ID
Luu Gia Hân	20213056
Nguyễn Thị Ninh	20213088
Trần Hà Yến Nhi	20213086
Phoee Myat Thazin	2023T071

Instructor: Dr. Ha Thi Thu Trang

Ha Noi, April – 2024

TABLE OF CONTENT

1. Define the problem.....	1
Data set Overview.....	1
Objectives	1
2. Data analysis	2
Data cleaning	2
How do firms finance their operations in Vietnam?	3
Describe the characteristics of their financial transactions in Vietnam	4
3. Modeling Regression Model	10
3.1. Pre-processing for regression model.....	10
3.2. Modeling Regression Model.....	10
4. Conclusion.....	16

1. Define the problem

Data set Overview

We use data from World Bank Enterprise Surveys (WBES) during the period April 2023 to September 2023. The research data is firms' day-to-day experiences, and perceptions of enterprises regarding the business environment in which they operate. The organizations providing information for the data set belong to small, medium, and large-sized companies operating in two fields of manufacturing and providing services. These firms are located in 5 stratification regions in Vietnam: Northern Central region, the Central Coast, Central Highlands; Red River Delta; Northern Midlands and Mountains; Southeast; Mekong River Delta. The data set includes 1,028 companies (observations) that completed the survey. The questionnaire covers the topics about: general firm characteristics, infrastructure, sales and supplies, management practices, competition, innovation, capacity, land and permits, finance, business-government relations, exposure to bribery, labor, and performance.

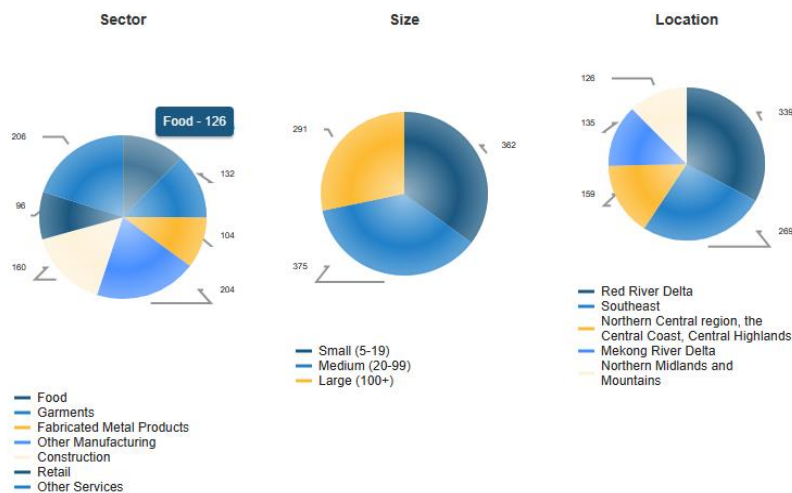


Figure 1: Pie chart for Sector, Size, Location

The data set generally covers the entire Vietnamese economy with the different industries outlined in Figure 1. Industry data is guaranteed to be consistent and fair in terms of industry classification and company size. The two key economic regions are the Red River Delta, which is home to many manufacturing plants and industrial zones, and the Southeast, which is a major center for tourism and agriculture. Figure 3 also accurately reflects the concentration of businesses.

Group 3 of the Business Forecasting class including 4 members: Luu Gia Han, Tran Ha Yen Nhi, Nguyen Thi Ninh, and Phooe Myat Thazin, is responsible for explaining how firms finance their operations and the characteristics of their financial transactions. Therefore, we determined to work mainly on the columns related to the Finance index in the provided data set, including 50 columns encoded with the letter K at the beginning, and some indicators explaining the characteristics of companies that we believe will be important in the regression model.

Objectives

In this report, we mainly focus on investigating the two crucial parts of the question

Q1: How do firms finance their operations?

Q2: What are the characteristics of their financial transactions?

Firstly, is the manner in which firms finance their operations, the operation is related to how the firms can raise their capital, and the mobilized capital comes from two sources: fixed assets and working capital. In terms of finance, fixed assets have internal sources that come from internal funds and retained earnings, and external sources include banks, non-bank financial institutions, suppliers or customers, and others. These two sources are also similar for working capital. Secondly, the next main part we need to mention regarding the finance transaction is selling to customers and purchasing from suppliers.

Next, we will build the model to give detailed information about how credit constraints affect firm performance. Detailed information will include crucial variables and how they affect the objective function.

2. Data analysis

Data cleaning

In this section, we outline the preprocessing steps applied to the Vietnam 2023 data to ready it for subsequent analysis. This original data which contains 355 columns, and 1028 rows is preprocessed by R and loaded using the R package haven.

Regarding the defined problem, we purely focus on the Finance sector starting with the letter 'k' in the original dataset. Initially, we filter the dataset to include only columns relevant to our analysis using the table library. Subsequently, we convert all negative values to NA and assess the presence of missing values.

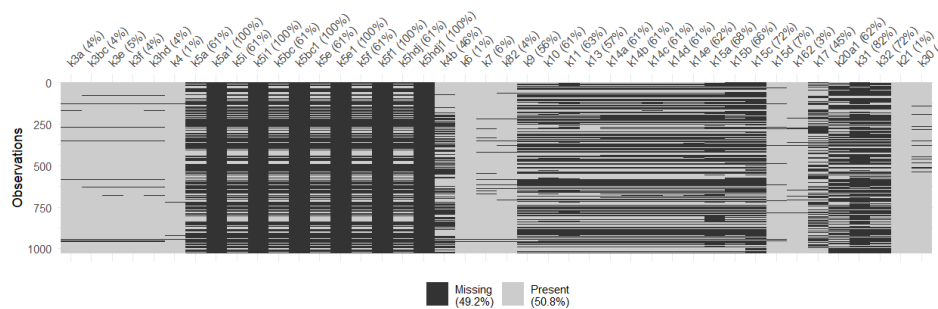


Figure 2: Missing value plot

From the plot, it is evident that there is a significant number of missing values in our data frame, accounting for 49.2% of the total values. Before proceeding to the analysis, it is crucial to handle these missing values. Hence, we implement to exclude all the columns having more than and equal to 1000 missing values. Finally, we have a new data frame that contains 36 columns and 1028 rows with a rate missing value rate of 40.8%. This data frame is used for the subsequent sections of this report.

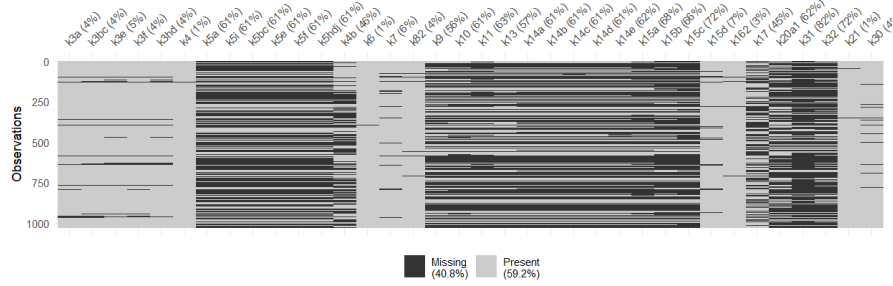


Figure 3: Missing value plot after data pre-processing

How do firms finance their operations in Vietnam?

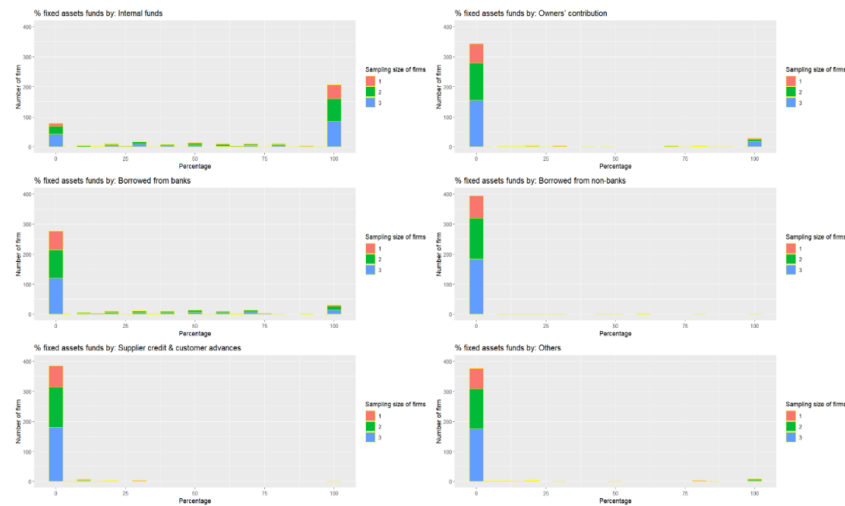


Figure 4: Fixed Assets - Proportion of investments financed by different sources

Firms tend to completely use Internal funds for purchasing fixed assets. Around 90 firms relying on other sources compared to 200 companies depending fully on Internal funds or Retain earnings. For these 90 firms, they mainly choose Equity financing and Bank borrowing for the Investments. In addition, the minority of medium-size firms mobilize 50-70% of their funds for Investment from Bank Borrowed. The options for External institutions and Supplier credit & customer advances are not favorable among Vietnamese firms.

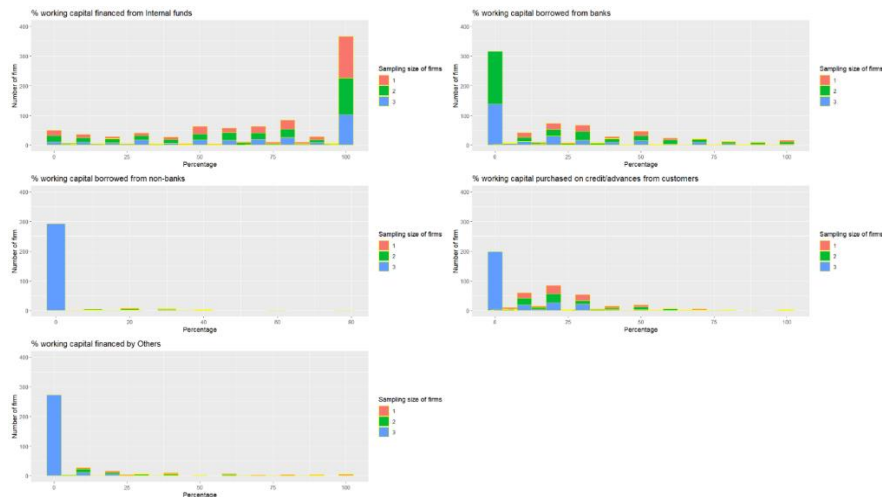


Figure 5: Working Capital - Proportion of investments financed by different sources

Most of the financial resources spent on working capital comes from internal funds or retained earnings, reaching a level greater than 50% for all firms. Larger than 200 medium-sized firms and smaller than 130 large firms are not inclined to mobilize money from banks for working capital investment. It is notable that 200 to more than 300 large firms tend not to use investment money from outside sources. This proves that these companies do not have to raise capital from outside institutions to invest in working capital; in other words, large firms may be more capable of using their own financial resources.

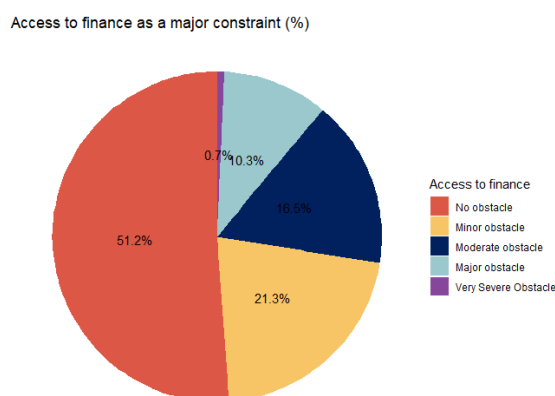


Figure 6: Access to finance as a major constraint

There are 48.8% in the total of 1016 samples reporting that finance appeared as an obstacle for them. 1/4 of firms take finance as moderate to severe problem, therefore in the following parts we will find the insight into the characteristics of their financial transaction.

Describe the characteristics of their financial transactions in Vietnam

Fully credit constrained:

Given that fully credit constrained make up the following conditions:

(1) The firm did not apply for a loan for any reason other than the lack of need for it.

(2) The firm applied for a loan, but the application was rejected, even when it has access to equity financing.

The conditions are checks thoroughly from the questionnaire table:

(1) In k162 we choose the answer decoded as 4, in k17 we choose the answer decoded from 2 to 7. The total number of companies satisfying these conditions equals to 199.

(2) In k162 we choose the answer decoded as 1, 2, and 3 and k20=3, and k5i>0. And the result for these conditions is 9 firms.

Partially credit constrained:

In the data set descriptions, we also know that the conditions 4 partially credit constrained are any of three statements:

(1) The firm applied for a loan and the application was partially approved.

(2) The firm applied for a loan and the application was rejected, but the firm has access to external sources of finance excluding any equity finance (k5i).

(3) The firm has external finance but did not apply for a loan due to any reason other than no need for it.

According to the detailed description, we have the encoded answers from the table K 16.2, K20 and K17, K5 as follow:

(1) $k162=1 \rightarrow 3$ and $k20=2 \Rightarrow$ the total number of firms satisfied these are 57

(2) $k162=1 \rightarrow 3$ and $k20=3$, $\text{sum}(k5bc, k5e, k5f, k5hdj) > 0\% \Rightarrow$ The result is 5 firms.

(3) $k162=4$, $k17=2 \rightarrow 7$, $\text{sum}(k5bc, k5e, k5f, k5hdj) > 0\% \Rightarrow$ The result for this is 50 firms.

Credit unconstrained:

(1) The firm applied for a loan and the application was approved in full.

(2) The firm did not apply for a loan because it did not need it.

$k162=1 \rightarrow 3$, $k20=1 \Rightarrow$ For the first case in credit constraint we have 328 firms.

(3) $k162=4$, $k17=1 \Rightarrow$ And finally there are 368 firms satisfied this condition.

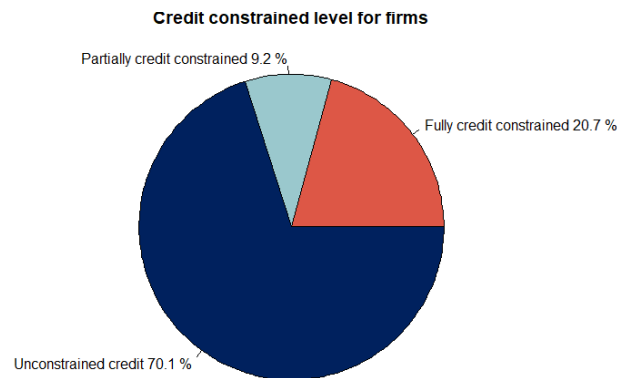


Figure 7: Credit constraints level for firms

In overall, the large occupation is unconstrained credit which made-up to 70.1%. This number of firms whose unconstrained credit is almost thrice the number of fully credit constrained firms. We can conclude that most Vietnamese companies may have not face with too many challenges. However, we should consider 1/4 of firms facing partial or full credit constrained to understand the reason behind this significant number. Therefore, we have the indicator fin21 reflecting the reason for rejected applications.

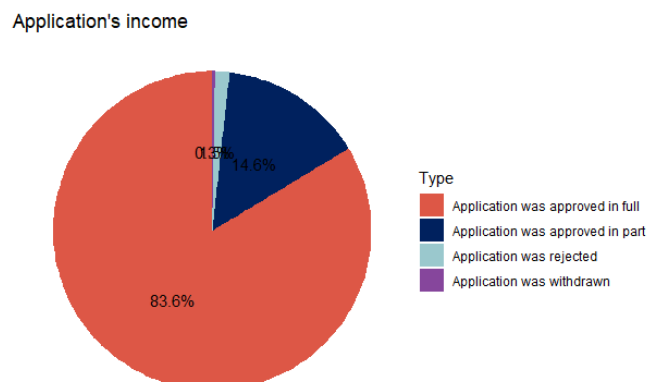


Figure 8: Application's outcome

When analyzing application outcomes, the fully approved applications represent the majority of 84%, while rejected applications constitute a smaller proportion. Following this, the table below details the reasons identified for firm rejections.

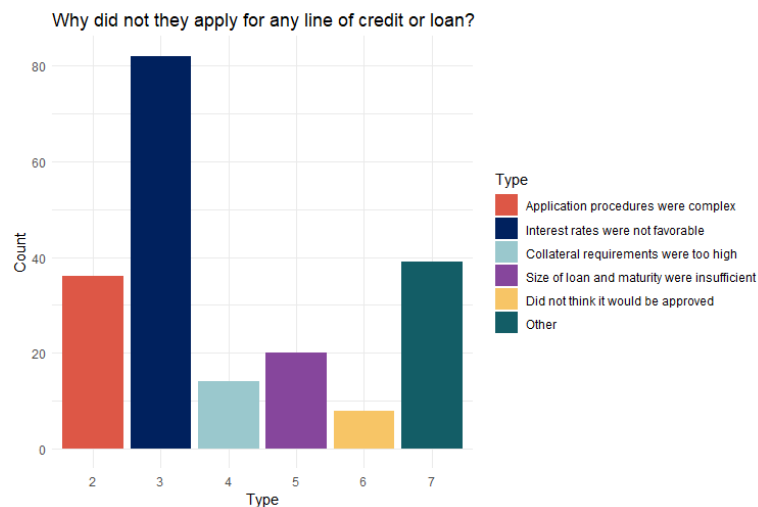


Figure 9: Reasons for rejecting applications

Most firms are inclined to reject applications for loans because of unfulfilled interest rates. Interest rates at 15% and decrease to 10%-13% throughout 2023 compared to previous years. This decrease was driven by the State Bank of Vietnam's (SBV) policy changes aimed at controlling inflation and stimulating economic growth. Eventually, the decrease in interest rate in 2023 still did not meet the need for applying for loans; the interest rate seems to be high. Therefore, the interest rate was considered the most concerning reason for many firms to reject applications. The second concern about applying for loans was related to the complex procedures, which were time-consuming and involved a lot of paperwork and documentation requirements. Moreover, the size of the loan and maturity can be barriers to many firms due to a lack of time to repay the loans, and the loan amount does not satisfy the borrower's demand (it may be too little or too much compared to the loan amount that the borrower ascertains). The next reason why many firms did not apply for a loan is because the amount of collateral needed to borrow is quite high compared to the amount of loan the borrower can receive, or the borrower does not have enough collateral to receive the loan. Lastly, the borrowers don't have enough confidence to take the loan, so they decide to reject the application. This matter can stem from their ability to repay the loan with high interest and their fear of risk as bankrupt banks.

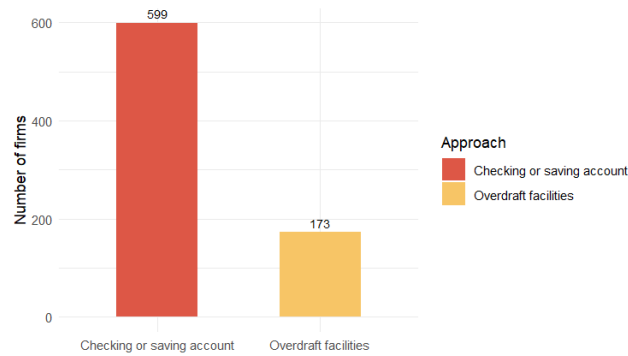


Figure 10: Approaches to financial resource

Savings accounts, a type of deposit account offered by banks and financial institutions, serve as a means for customers to accumulate funds while accruing interest. These accounts are a popular choice for businesses seeking to establish a cash flow buffer within their budgets. Conversely, overdraft facilities, a credit feature offered on a limited basis, allow businesses to access funds even when their account balance falls short. However, there are only 173 firms that use overdraft facilities since overdraft facilities may come with high fees.

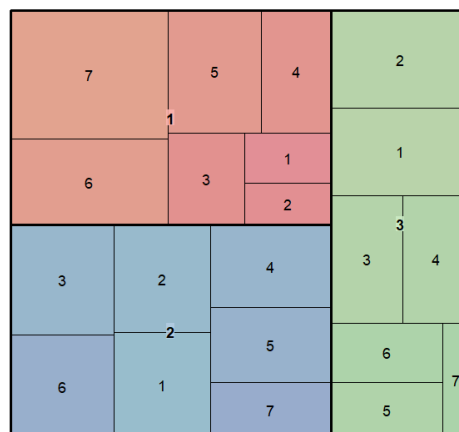


Figure 11: Firms having checking and/ or saving account

Noted that for **Figure 11** and **12**:

Color

Red: Small

Blue: Medium

Green: Large

1. Food

2. Garments

3. Fabricated Metal Products

4. Other Manufacturing

5. Construction

6. Retail

7. Other Services

Service companies are the majority in owning checking and/or saving account making up 1/5 of Small-sized firms, followed by Retail and Construction sector. In contrast, service sector in Medium sized company accounted for having the least number of checking and/or saving account. Moreover, Medium sized companies operating in food, garments, fabricated metal products and retail fields take up the largest part in saving account. Likewise, Large- and

Medium-sized companies also observed service sector as a minor group in having the saving account.

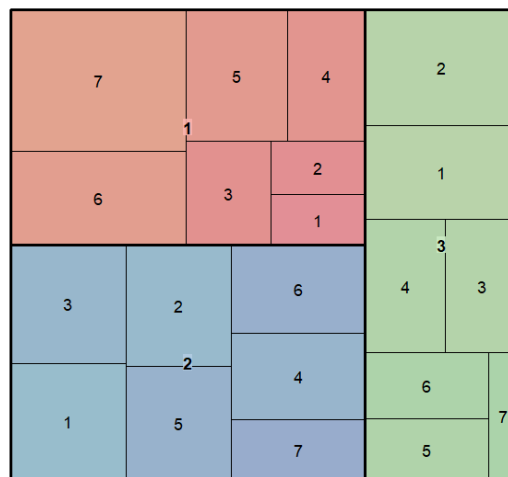


Figure 12: Firms having Overdraft Facilities

Regarding firms choosing the Overdraft Facilities option, the quantity is smaller as shown in **Figure 10**, but the distribution of **Figure 12** mainly shares the same pattern as Firms having checking and/or saving account.

Firms having a line of credit or a loan from a financial institution

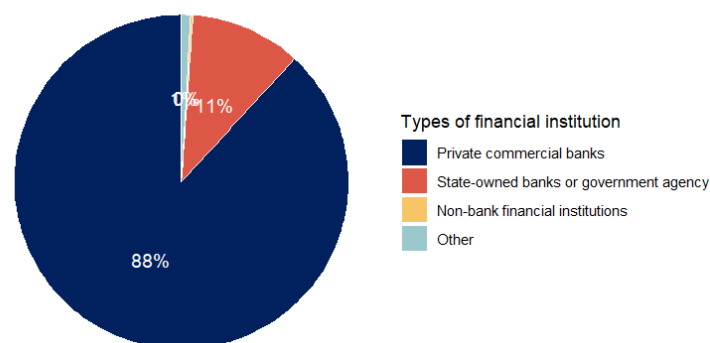


Figure 13: Firms having a line of credit or a loan from a financial institution

In Figure 12, the majority choose Banks as their financial provider than from non-banks and others. On the other hand, there has been no sign of firm having line of credit or loan from purchasing on credit from suppliers and advances from customers.

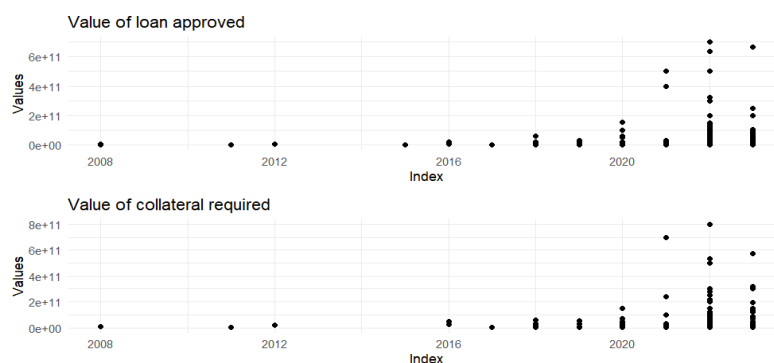


Figure 14: Value of collateral needed for a loan (% of the loan amount)

Due to the severe consequences of post Covid19, company loans tend to increase over the years peaking in the duration of 2020 to 2023. This points out that businesses need loans to maintain their company's business activities. Because of the need for loans, they needed to spend a large amount of collateral to receive the loan. In terms of loans approved, the loan peaked at 27 trillion VND in the year of 2022 compared to the period of financial crisis in 2008 which was 5 billion. This noted that the company has more access to loans during Covid 19 compared to the financial crisis in 2008.

Financial institution requiring collateral

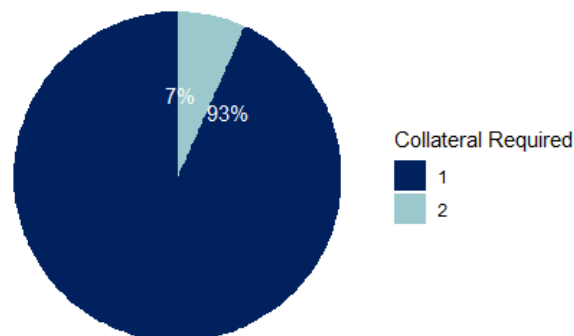


Figure 15: Proportion of loans requiring collateral (1 - Yes, 2 – No)

In this pie chart, it shown that Vietnamese financial institutions generally require the collateral for the loan while it is only 7% of loans not requiring collateral. The reason for most financial institutions selecting collateral trading with loans is that the financial institutions can minus the risks from unsecured loans. Meanwhile, in some cases for not requiring collateral, it can happen when borrowers prove that they are having good credit and a steady income (borrower's creditworthiness) so financial institutions allow the borrowers receive a loan with great deal.

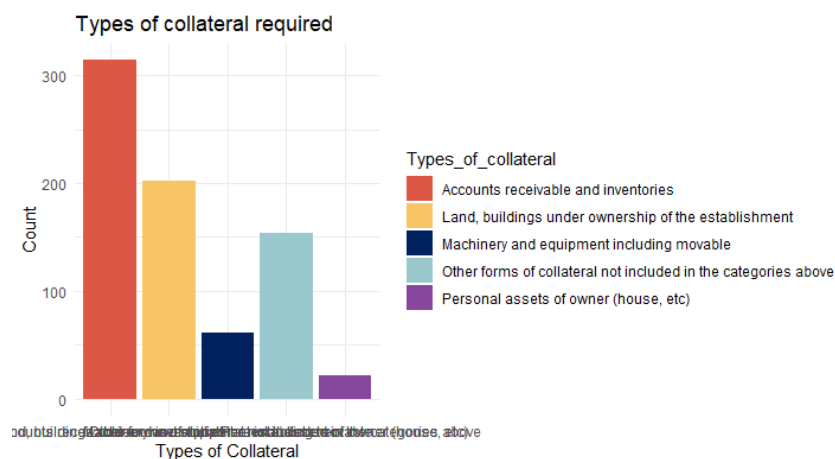


Figure 16: Distribution of collateral required

The bar chart indicates that the main type of collateral required is real estate assets, followed by machinery and equipment. Furthermore, nearly half of the real estate units were equal to personal assets; The amount of collateral used from accounts receivable

and other sources is the least preferred in comparison to the other kinds of collateral needed.

3. Modeling Regression Model

3.1. Pre-processing for regression model

Firstly, to dealing with columns fin23, fin24, fin25 in our dataset, we categorize the data into 3 types: Fully credit constrained, partially credit constrained and Credit unconstrained (the way to classify them was showed in the previous part). Then, we convert them to dummy variables by generating two new columns.

(1) fully: We assign the value “1” to the entries which are Fully credit constrained, and all other values are set to “0”.

(2) partially: Similarly, we assign the value “1” to the entries which are Partially credit constrained, and all other values are set to “0”.

→ Therefore, all the entries with fully=0 and partially=0 are considered to Credit unconstrained.

Apply similar processing to columns sector, size, region, fin15, fin16 and t2 which are the Categorical columns, we have 18 new columns (sector_1-6, size_1-2, region_1-4, fin15_1, fin16_1-4 and t2_1)

We also create a new column called $\text{fin10} = (\text{k15a}/\text{k11}) * 100$ ignoring the missing value to convert the value to percentage value. Next step, we change the name of 12 original columns (k5a, k5bc, k5f, k5i, k5e, k5hdj, k3a, k3bc, k3f, k3e, k11, k8) to (fin1, fin2, fin3, fin4, fin5_1, fin5_2, fin6, fin7, fin8, fin9, loan_value, fin14), respectively.

It is also noted that we use the log-transforming for the column loan_value because loan amounts are often positively skewed, the log-transforming data can help address issues related to skewness, heteroscedasticity, and outliers, and facilitate the interpretation of results in terms of elasticities.

Finally, we get a new data frame with 32 columns and 1016 rows that is used to run the regression model in the following part.

3.2. Modeling Regression Model

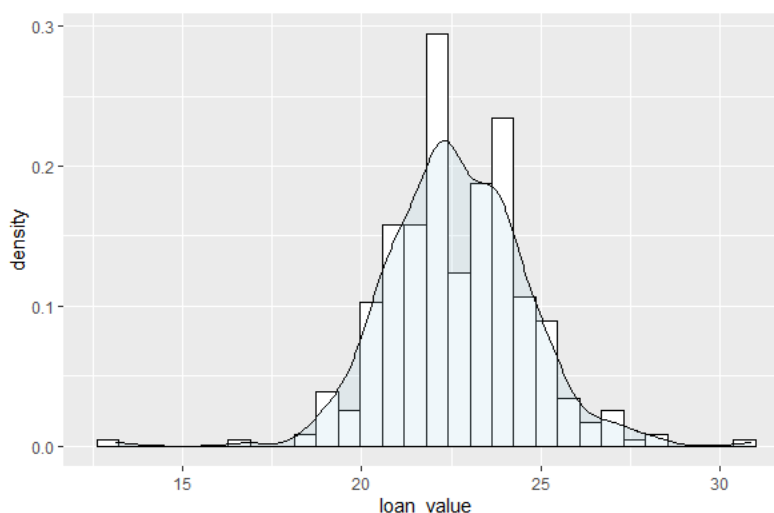


Figure 17: Density plot for Loan value

The plot suggests that logarithm of loan value following a normal distribution. The probability density then tapers off as the loan amount increases and decreases on both sides.

Model 1:

```

Residuals:
    Min       1Q   Median       3Q      Max
-4.4199 -0.9913  0.0000  0.7668  3.6640

Coefficients: (3 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.213e+01  2.034e+00  10.882 < 2e-16 ***
size_2       -4.453e-01  7.236e-01  -0.615  0.539542
size_3       -8.623e-01  7.515e-01  -1.147  0.253596
sector_1      1.012e+00  5.176e-01   1.955  0.053025 .
sector_2      6.405e-01  5.257e-01   1.218  0.225531
sector_3      1.857e-01  5.399e-01   0.344  0.731489
sector_4      4.321e-01  5.547e-01   0.779  0.437548
sector_5      3.517e-01  5.929e-01   0.593  0.554227
sector_6     -5.101e-02  5.357e-01  -0.095  0.924310
region_1     -6.380e-01  7.880e-01  -0.810  0.419824
region_2              NA          NA      NA      NA
region_3              NA          NA      NA      NA
region_4     -4.780e-01  8.055e-01  -0.593  0.554005
fin5_1       -5.065e-02  3.741e-02  -1.354  0.178401
fin5_2       -2.033e-02  1.930e-02  -1.053  0.294359
fin1         -8.523e-03  4.614e-03  -1.847  0.067265 .
fin2         -9.224e-03  5.398e-03  -1.709  0.090171 .
fin3          8.587e-04  3.676e-02   0.023  0.981405
fin4              NA          NA      NA      NA
fully         4.693e-01  5.724e-01   0.820  0.413947
partially    -7.809e-02  4.418e-01  -0.177  0.860027
fin6          9.349e-03  1.081e-02   0.865  0.389031
fin7          1.790e-02  1.198e-02   1.495  0.137719
fin8          6.825e-03  1.484e-02   0.460  0.646388
fin9          4.143e-02  2.113e-02   1.961  0.052322 .
fin10         -1.931e-04  5.069e-05  -3.810  0.000224 ***
fin15_1       8.579e-01  3.485e-01   2.462  0.015294 *
fin16_1       1.059e-01  1.742e+00   0.061  0.951619
fin16_2      -3.395e-01  1.751e+00  -0.194  0.846548
fin16_3      -1.296e-02  1.778e+00  -0.007  0.994196
fin16_4      -4.884e-02  1.803e+00  -0.027  0.978437
t2_1          7.369e-01  2.938e-01   2.508  0.013506 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.525 on 116 degrees of freedom
(871 observations deleted due to missingness)
Multiple R-squared:  0.3729,    Adjusted R-squared:  0.2216
F-statistic: 2.464 on 28 and 116 DF,  p-value: 0.0004258

```

Figure 18: Results for Regression model1

Multiple R-squared (0.3729) indicates that approximately 37.29% of the variance in the dependent variable is explained by the independent variables. Adjusted R-squared = 0.2216 is lower than the multiple R-squared, reflecting the penalty for adding more predictors. F-statistic (2.464) shows the overall significance of the model. A higher absolute value of the t-statistic (10.882) indicates that the coefficient is significantly different from zero. The p-value (0.0004) indicates that the model is statistically significant, meaning at least one of the predictors is significantly related to the dependent variable. Several variables (fin_9, fin_10, fin 15_1, t2_1) are statistically significant and sector_1, which is marginally significant, indicating that they have a meaningful impact on the dependent variable.

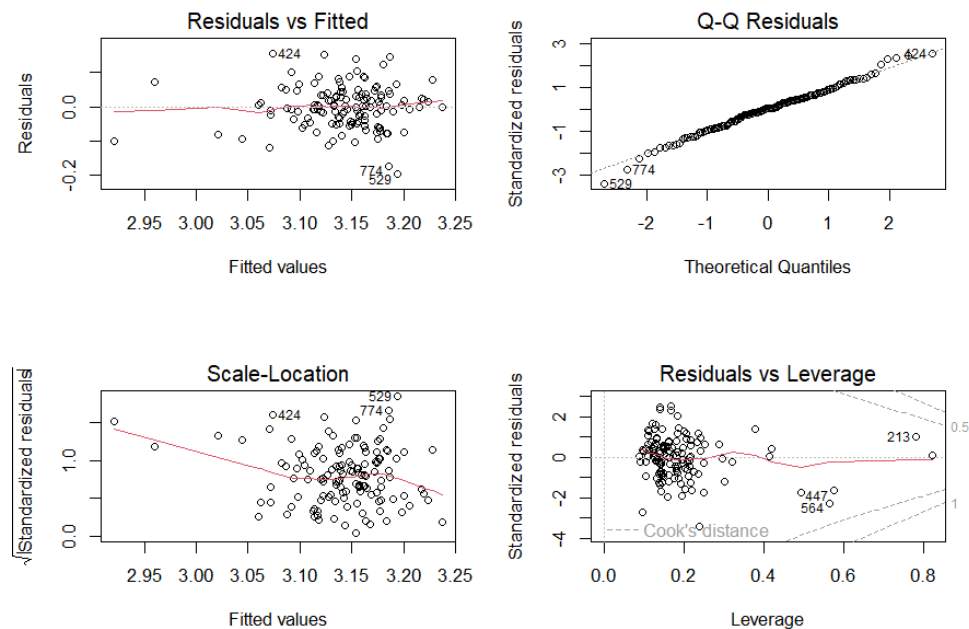


Figure 19: Charts for the results of model1

Residuals vs Fitted (the difference between the observed values and the fitted values): it is good when it is non constant. A random scatter around a horizontal line at zero would indicate that the model fits the data well.

Q-Q residual (comparing the quantiles of the residuals to the quantiles of a normal distribution) – The residual follows a normal distribution if it saturated on the diagnostic lines.

Scale-Location plot shows that there is a random scattering around a horizontal line: This is the desired outcome since it indicates homoscedasticity.

Residual vs leverage (the residuals on the y-axis versus the leverage on the x-axis)
 - There are few potential outliers in the model.

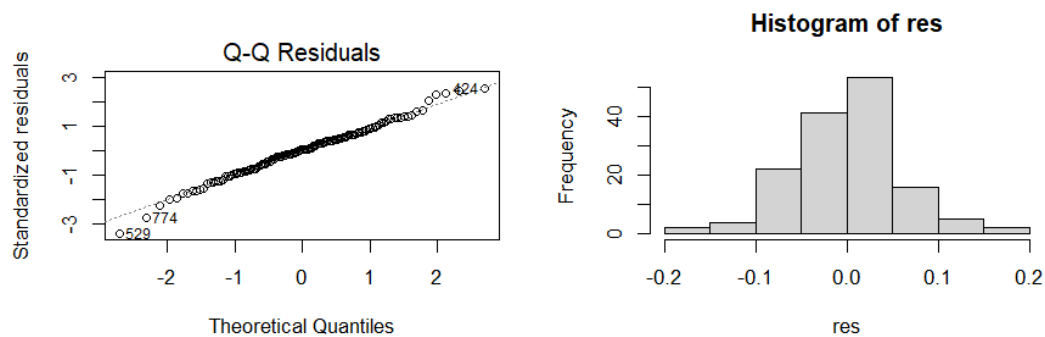


Figure 20: Residuals distribution

```

one sample t-test

data: res
t = -6.9335e-17, df = 144, p-value = 1
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.2247212  0.2247212
sample estimates:
mean of x
-7.882898e-18

```

Figure 21: t-test for mean value of residual

The t-test is nearly zero. The p-value is 1. This means the residuals are centered around zero which is one of the assumptions of a well-fitted linear regression model.

```

studentized Breusch-Pagan test

data: modell
BP = 32.269, df = 28, p-value = 0.2637

```

Figure 22: Breusch-Pagan test

H0: The model has heteroscedasticity.

In this case, the p-value 0.2637 is greater than 0.05. So, we fail to reject the null hypothesis. ➔ There isn't sufficient evidence to conclude that heteroscedasticity is present in the model "modell".

```

> durbinwatsonTest(modell)
lag Autocorrelation D-W Statistic p-value
1 0.01516254 1.966095 0.534
Alternative hypothesis: rho != 0

```

Figure 23: Durbin Waston test for autocorrelation

H0: The model has no autocorrelation.

The Durbin-Watson statistic value is close to 2, suggesting there is no significant autocorrelation in the residuals. P-value 0.534 is higher than 0.05. Therefore, the model has no autocorrelation.

```

> vif(modell_alias)
sector_1 sector_2 sector_3 sector_4 sector_5 sector_6 region_1 region_2 region_3 region_4 fin5_1
fin5_2 fin1 fin2
2.671092 2.457452 2.159775 2.183192 1.788303 2.126588 6.030359 8.070489 4.185302 5.584643 1.620277
1.388274 2.313513 2.528614
fin3 fully partially fin6 fin7 fin8 fin9 fin10 fin15_1 fin16_1 fin16_2
fin16_3 fin16_4 t2_1
1.306683 2.627940 2.227929 7.855608 6.730076 2.997679 1.945700 1.927406 1.578912 45.859795 36.296017
31.538605 22.028940 1.237084

```

Figure 24: VIF results

The set of fin16_1, fin16_2, fin16_3, fin_16_4, fin7, fin6, fin7, region_4, region_2, region_1, has VIF value larger than 5. This means there are some problems with multicollinearity. The remaining VIF value of other variables are between 1 & 5, so there is moderate correlation which is usually not problematic.

Model 2:

```

Residuals:
    Min       1Q   Median       3Q      Max
-4.7625 -0.9102 -0.0154  0.7905  4.0581

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.279e+01  4.347e-01  52.419 < 2e-16 ***
sector_1     7.925e-01  3.271e-01   2.423  0.01668 *
fin1        -7.923e-03  4.055e-03  -1.954  0.05269 .
fin2        -8.466e-03  4.580e-03  -1.848  0.06664 .
fin7         8.445e-03  4.916e-03   1.718  0.08801 .
fin9         2.099e-02  1.672e-02   1.255  0.21139
fin10       -1.919e-04  4.152e-05  -4.621  8.57e-06 ***
fin15_1      7.596e-01  2.801e-01   2.712  0.00752 **
t2_1        5.155e-01  2.678e-01   1.924  0.05632 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.514 on 140 degrees of freedom
(867 observations deleted due to missingness)
Multiple R-squared:  0.2786,    Adjusted R-squared:  0.2374
F-statistic: 6.758 on 8 and 140 DF,  p-value: 1.776e-07

```

Figure 25: Results for regression model 2

Model 2:

$$\log(y) = 22.79 + 0.7925 * sector_1 - 0.008 * fin1 - 0.0085 * fin2 + 0.0084 * fin7 + 0.021 * fin9 - 0.0002 * fin10 + 0.52 * t2_1$$

Where:

y is the loan value

sector1: Is the firm located in Red River Delta location? (yes=1, no=0)

fin1: Proportion of investments financed by internal funds (%)

fin2: Proportion of investments financed by banks (%)

fin7: Proportion of working capital financed by banks (%)

fin9: Proportion of working capital financed by other financing sources (%)

fin10: Value of collateral needed for a loan (% of the loan amount)

fin15_1: Is firm with a checking or savings account (yes=1, no=0)

t2_1: Does the firm have their annual financial statement reviewed by external auditor (yes=1, no=0)

Interpreting the Intercept and Coefficients:

- For a loan of any firms located in other places than Red River, fin1, fin2, fin7, fin9, fin10 are equal to 0, and it have no saving or checking account, the average expected logarit of loan value is 22.79.
- All fin1, fin2, fin7, fin9, fin10 range from 0 to 100 (unit=percent%)
- On average, a company which is located in Red River Delta area has 0.7925 unit higher on the logarit of loan value compared to other firms from other areas, assuming the predictor variables are held constant. (This explanation goes the same with fin15_1 and t2_1)
- On average, each additional proportion in fin1 is associated with an increase of 0.008 units on the logarit of loan value, assuming other predictor variables are held constant. (This explanation goes the same with fin15_1 and t2_1)

Multiple R-squared (0.2786) indicates that approximately 27.86% of the variance in the dependent variable is explained by the independent variables. Adjusted R-squared = 0.2374 is higher than Adjusted R-squared in model1 (0.2216) noting for better explanation ability of the model2. F-statistic (6.758) shows the overall significance of the model. A higher absolute value of the t-statistic (52.419) indicates that the coefficient is significantly different from zero. The p-value indicates that the model is statistically significant, meaning at least one of the predictors is significantly related to the dependent variable. Several variables (sector_1, fin_1, fin_10, fin_15_1) are statistically significant indicating that they have a meaningful impact on the dependent variable.

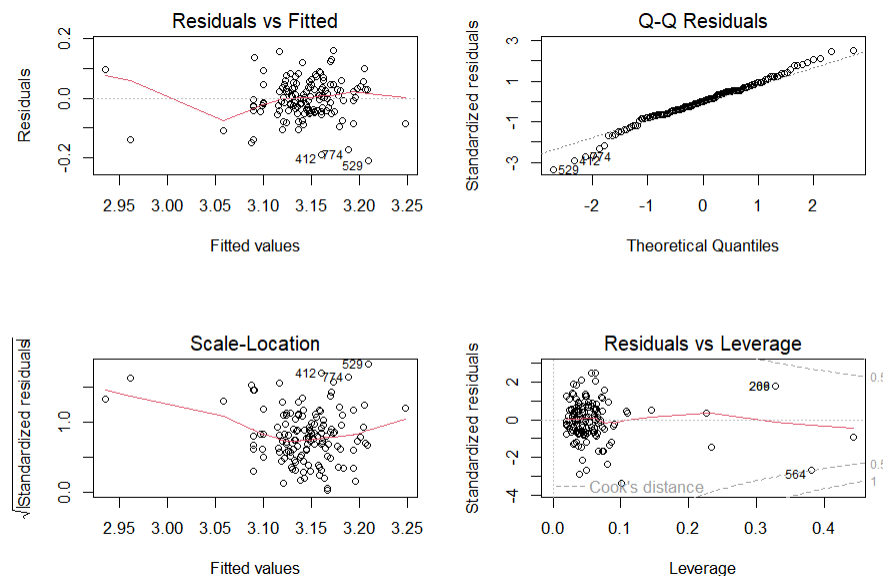


Figure 26: Charts for the results of model2

A random scatter in Residuals vs Fitted plot shows residual value around a horizontal line at zero would indicate that the model fits the data well.

Q-Q residual indicates that residual follows a normal distribution since it saturates on the diagnostic lines.

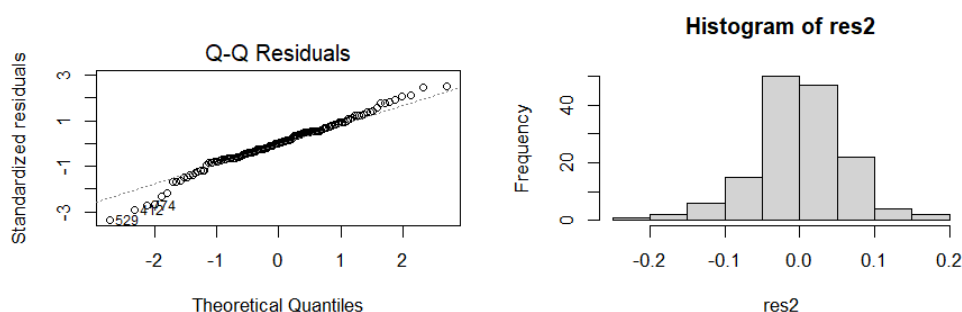


Figure 27: Residual distribution

Scale-Location plot shows that there is a random scattering around a horizontal line: This is the desired outcome since it indicates homoscedasticity.

Residual vs leverage (the residuals on the y-axis versus the leverage on the x-axis)
 - There are few outliers in the model.

```

One Sample t-test

data: res2
t = 7.8746e-17, df = 24, p-value = 1
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -171.0122 171.0122
sample estimates:
mean of x
6.524781e-15

```

Figure 28: t-test for mean value of residual

The t-test is nearly zero. The p-value is 1. This means the residuals are centered around zero which is one of the assumptions of a well-fitted linear regression model.

```

studentized Breusch-Pagan test

data: model2
BP = 11.832, df = 8, p-value = 0.1588

```

Figure 29: Breusch-Pagan test

H0: The model has heteroscedasticity.

In this case, the p-value 0.1588 is greater than 0.05. So, we fail to reject the null hypothesis. ➔ There isn't sufficient evidence to conclude that heteroscedasticity is present in the model "model2".

```

> durbinwatsonTest(model2)
lag Autocorrelation D-W Statistic p-value
1 -0.01543997 2.019386 0.914
Alternative hypothesis: rho != 0

```

Figure 30: Durbin Waston test for autocorrelation

H0: The model has no autocorrelation.

The Durbin-Watson statistic value is around 2, suggesting there is no significant autocorrelation in the residuals. P-value 0.914 is higher than 0.05. Therefore, the model has no autocorrelation.

```

> vif(model2)
sector_1 fin1 fin2 fin7 fin9 fin10 fin15_1 t2_1
1.090206 1.877330 1.917285 1.210854 1.238848 1.314123 1.061098 1.068890

```

Figure 31: VIF results

The VIF value of other variables are between 1 & 5, so there is moderate correlation which is usually not problematic issue with multicollinearity.

4. Conclusion

Our study analyzes data from World Bank Enterprise Surveys (April-September 2023) on firm operations and perceptions in Vietnam (1,028 companies across manufacturing and services, 5 regions). We focus on how firms finance operations (sources of capital) and financial transaction characteristics (credit constraints, etc.) using data on finance and firm characteristics.

In conclusion, we see that Vietnamese firms tend to use Internal funds on fixed assets, investing, and working capitals. Otherwise, Banks would be their second options

for financing investments. Especially, working capitals are purchased on credit/advances from customers with some advantages. By receiving credit or advances from customers, a firm can access immediate funds without needing to tap into other sources like loans or selling assets. This improves their cash flow situation and allows them to meet short-term obligations like payroll and inventory purchases. Compared to traditional loans, credit purchases or customer advances can be less expensive in terms of interest rates and fees. This can improve the firm's financial health and reduce its overall debt burden. The main reasons for Bank not to be the initial option for firms comes from the complicated procedures and unfavorable interest rates during the given period. Therefore, firms who are relying on their own financial strength view the obstacle to finance as minor compared to the ones who choose banks or other financial institutions.

The prediction variable we choose is Loan value. When the loan amount to invest in investments or working capital is larger, the company's financial capacity and performance must be certified in terms of repayment ability and favorable business operations. Therefore, a large loan can also reflect the credit constrained level that the company faces (2).

The results show that the optimal model is:

$$\log(y) = 22.79 + 0.7925 * sector_1 - 0.008 * fin1 - 0.0085 * fin2 + 0.0084 * fin7 + 0.021 * fin9 - 0.0002 * fin10 + 0.52 * t2_1$$

Where fin10 and fin15_1 have the most significant impact, respectively are the Value of collateral needed for a loan (% of the loan amount) and the result of if the firm with a checking or savings account (yes=1, no=0).

The model still has many shortcomings as the current Adjusted R square is only 0.2374. In addition, group 3 thinks the problem lies in the fact that there is a lot of missing data as well as the limited ability to handle missing data in running the model.

Finally, our group is extremely grateful to Instructor Ha Thi Thu Trang for accompanying the group during the process of making this report!

REFERENCES:

- [1] Dataset and figure from <https://www.enterprisesurveys.org/en/data/exploreeconomies/2023/viet-nam#biggest-obstacle>
- [2] <https://documents1.worldbank.org/curated/en/524781502187576101/pdf/117973-BRI-PUBLIC-4-8-2017-12-28-44-DBRN.pdf>
- [3] Reference documents from Instructor: Dr. Ha Thi Thu Trang, slide chapters from 1 to 6
- [4] Hanke, John E. and Dean W. Wichern. (2014) Business forecasting, 9th Edition, Pearson
- [5] <https://r4ds.had.co.nz/index.html>