1 Acknowledgments

I would like to express my gratitude to my family which has always been supportive of me, as well as Prof. Gabriele Deana, whose guidance and support has been invaluable throughout this thesis.

Contents

2 Introduction

A continuing concern in the world of today for financial investors is to gauge whether the financial risk of the firm they are considering to invest in is considerable enough for them not to invest. In the last two decades, one of the most popular business sectors that has managed to gather billions of dollars in funding has been the tech sector due to a lot of startups being profitable and safe enough. However, the concern harbored by investors still remains as this phenomenon is not true for all the cases. Why is this the case? Why do some companies in the tech sector manage to do well while others do not? The possible factors are countless, and some have been proven to have an impact on a firm's financial health by other academic papers. However, the main factor of interest that this paper seeks to analyze is Research and Development Intensity. The goal of this paper is to show that there is a positive relation supported by data between Research and Development Intensity and Altman Z-score, a commonly used measure of financial risk for firms large enough.

The structure of this paper is in the proceeding order. Firstly, there is the theory section, where a review of the literature is made. Several previously published academic papers are summarized to establish what consists of current knowledge, where does the gap in literature exist, and how does the hypothesis discussed in this study address it.

The following section will be the methods section, which provides an explanation of what data were collected, through what means were they gathered, and in what way were they analyzed. The regression models used and the intuition behind them will also be discussed in this section of the study.

The following section will be the data section. Here, the important details of the sample of data gathered will be discussed as well as the summary statistic table for all the variables used. Then graphical representations of the main findings of the explanatory data analysis will be presented and, if necessary, adjustments will be made. In the end regression results for each model will be shown.

The next section will be the discussion section where all the information presented in the fourth part will be interpreted verbally.

The last section will be the conclusion section where the main findings of this paper will be summarized. Then, there will be a discussion on whether these findings are consistent prove or disprove the initial hypothesis.

3 Theory

3.1 The use of accounting ratios to predict financial distress

Predicting corporate financial distress is a very important topic in the world of finance. This information is extremely useful to investors, who seek to fund profitable and high performing businesses, and to lending institutions like local banks that would prefer to not give out loans to companies that have a high chance of experiencing financial distress in the future.
One of the most early academic papers that established that accounting ratios could be used to predict financial distress was Beaver(1966), which demonstrated that measures such as cashflow to debt and net income to assets could distinguish effectively between firms that had experienced bankruptcy and those that had not some years earlier.
The findings of this paper were later referenced in the Altman(1968) paper, which introduced for the first time the measure of financial distress, the Z-score, which was constructed as follows:

$$Z = 0.12X_1 + 0.14_2X_2 + 0.33X_3 + 0.006X_4 + 0.999X_5$$

where:

- Z represents the Z-score
- X_1 represents Working capital/Total assets, a liquidity ratio
- X_2 represents Retained Earnings/Total assets, a profitability ratio
- X_3 represents Earnings before interest and taxes/Total assets, a productivity ratio
- X_4 represents Market value equity/Book value of total debt, a solvency ratio
- X_5 represents Sales/Total assets, an efficiency ratio

Because of the simplicity and high out-of-sample accuracy of the Z-score, its use became popular at the time.
In the future , more research papers introduced better models for predicting the likelihood of financial distress. A notable paper worth mentioning is that of Merton(1974),

which used in the model information related to equity prices and market volatility to estimate the probability of default. The most recent developments in the field are the use of machine learning techniques to increase the predictive power of the model. Despite these advancements, the Z-score still retains its usefulness due to how easily it can be interpreted and how easy it is to compute it.

3.2 The relationship between R&D and firm performance.

The exploration of the relationship between financial performance and research and development (R&D) has also been the topic of previous academic papers. An example is the study by Griliches (1981) that concluded that investment in R&D is a driver in the creation of intangible assets that enhance firm value.
However, the impact of R&D is not always positive. As R&D investments do not guarantee positive returns and are usually expensive, they can possibly lower short-term liquidity and increase leverage. This claim is supported by the academic paper of Chan et al. (2001) which found that firms with considerable expenses in R&D usually experience higher volatility in their earnings. This finding implies that the result will be a higher likelihood of short-term financial distress.
As the empirical evidence for the studies mentioned previously shows how investing in R&D can have both a positive and a negative effect on the likelihood of short-term financial risk, it suggests that the relationship between the two depends on other factors.

3.3 Identifying the gap in literature

The Altman Z-score, first introduced by Altman (1968) uses different accounting ratios like liquidity, profitability, leverage, and solvency to predict the financial distress of firms. The ratios used in the model that composes the Z-score are directly or indirectly affected by R&D investments. An example is that of an R&D investment big enough that it reduces profitability on the short-term and increases solvency, thus lowering the Z-score.
Despite these nuances, few studies have explicitly linked R&D intensity to the Altman Z-score, representing a clear gap in the literature that this paper wants to address.

3.4 Other factors affecting financial health

Besides R&D, there are other firm-specific factors or macroeconomic factors that can influence the financial health of a company.
One of these factors is firm size because larger firms typically have greater cash flows and more access to credit loans, thus having a positive impact on the Z-score.
Higher levels of profitability and liquidity are also associated with reduced financial risk.
Leverage is a ratio that is used to measure a firm's capability to meet its debt obligations. A famous leverage ratio is debt-to-assets, which observes what proportion of the assets is financed through debt. The lower the leverage, the more capable a company is to pay its obligations as it has less debts to pay. This is likely to result in a higher Z-score, indicating a lower risk of distress.
Another factor that affects financial distress is a macroeconomic shock. Generally, GDP growth affects the financial environment. In the academic paper (Lev & Sougiannis, 1996), it is stated that companies do better during times of expansion. In a recession, they do worse because the likelihood of risk is higher.

3.5 Hypothesis

It is straightforward to utilize the Altman Z-score for evaluating a company's financial stability, as it incorporates liquidity, leverage, productivity, and profitability ratios that are affected by investments in R&D. This leads to the primary hypothesis of this research:
H1: The intensity of R&D (measured as R&D spending relative to total assets) has a positive relationship with the Altman Z-score depending on a large enough firm size.
This premise suggests that although there may be immediate expenses associated with R&D, such investments ultimately foster the long-term financial well-being of organizations large enough to handle the risk. Additionally, control variables are anticipated to affect the Z-score in accordance with previous studies: firms that are larger, more profitable, and possess greater liquidity should exhibit higher Z-scores, whereas increased leverage is expected to diminish them. Furthermore, macroeconomic factors like the impact of a recession will also be considered to reflect broader economic contexts.

4 Methods

4.1 Data collection

The dataset used for this study contains financial information regarding firms in the tech sector in the US, Western Europe and Canada. For the same firms, information was collected from 2014 to 2024, ensuring consistent information between different companies and also accounting for changing effects over time. The method used to obtain this information was through the Bloomberg terminal.

Besides the financial variables directly acquired from the Bloomberg terminal, other variables had to be constructed either by using the raw variables in a formula, or by applying transformations to them.

Initially, through the Bloomberg terminal, the raw information collected was about the following variables:

- The Altman Z-score, a commonly used indicator of firms' financial distress and the dependent variable of interest.
- R&D spending as it is a component of the the main independent variable, R&D intensity which is suspected to affect the Z-score.
- Capital expenditures which was included to control for general investment activity and create a distinction from the investments involving R&D.
- Total assets as it a component of other essential variables which will be discussed a bit later.
- Current ratio as it measures short term liquidity, accounting for the company's capacity to meet short term debts,affecting risk of financial distress.
- Debt to Assets, which measures the proportion of the assets financed through debt. The higher it is, the higher the likelihood of financial distress due to the growing number of financial obligations.
- Operating margin which measures the profitability from core operations. Including it accounts for the firm's operational efficiency which could have an impact on the Z-score or R&D investment levels.

- Revenue growth, which may improve the financial health of the firm the higher it is. Including it in the regression models will separate its effect from the effect of R&D intensity.
- Debt to Equity that measures a firm's leverage by comparing the how much of the financing was through debt and much through equity. Companies with more debt relative to equity are riskier as the obligations to the creditors are greater than to the shareholders and have a higher chance of distress. Including this variable helps to account for the differences in capital structure.

4.2 Data preparation

The next step of the process was to obtain the other variables needed for this paper either by using the current information with the appropriate formulas or by applying transformations.
The variables were as follows:

- R&D intensity - obtained by dividing R&D spending by Total Assets. This is the main independent variable of interest and suspected to affect financial distress.
- Firm size - proxied by the logarithmic transformation of Total Assets. It helps to account that larger firms are financially stable and its effect on the Altman Z-score is independent of R&D intensity or capital expenditures.
- Region - which represents the region of the firm(the U.S., Canada or Western Europe) with the help of a self-created function in Python. The inclusion of this variable in the regression model can help account for macroeconomic factors in different countries, which may have a hidden effect on the financial distress of companies from that country.

4.3 Empirical Strategy and Model Specification

For this particular study, different models were implemented to account for a variety of angles related to the main topic of research.

4.3.1 Fixed effects regression

The first model used in this paper was a panel fixed-effects OLS regression which is structured as follows:

$$AZS_{it} = \beta_0 + \beta_1 X_{it1} + \beta_2 X_{it2} + \beta_3 X_{it3} + \beta_4 X_{it4} + \beta_5 X_{it5} + \beta_6 X_{it6} + \beta_7 X_{it7} + \beta_8 X_{it8} + \beta_9 X_{it9}$$

- AZS_{it} stands for the Z score.
- X_{it1} for firm size.
- X_{it2} for current ratio.
- X_{it3} for operating margin.
- X_{it4} for revenue growth.
- X_{it5} for debt to assets.
- X_{it6} for capital expenditures disinvestment.
- X_{it7} for R&D intensity.
- X_{it8} for R&D intensity squared.
- X_{it9} for the interaction term between R&D intensity and firm size.

The motivation behind choosing such a regression model is because of the data collected which has information for the same entities for multiple years. This is appropriate with the fixed-effects model. The second reason was that the fixed-effects model also accounts for unobserved firm-specific factors that may change over time and have some correlation with both R&D intensity and financial distress and thus influence them. If the pooled OLS regression were to be used instead, these changing factors would be ignored, possibly leading to omitted bias. This does not happen in the fixed effects model as changes within the same firm are compared over time and within a period all firm-specific factors are held constant.

4.3.2 Conditional fixed effects logit regression

The next model used was a conditional fixed-effects logit regression model. To use this model it is necessary to use a binary variable as the dependent variable. It was needed to create a binary variable from the Z-score. In Altman (1968) it is stated that all firms with a Z-score of greater than 2.99 are considered as non-bankrupt while those with a score of lower than 1.81 are considered in financial distress. The firms which have Z-scores between 1.81 and 2.99 are said to be in a "gray" area due to error classification. From this information a binary variable called Distress Binary was created where for each observation with a Z-score of lower than 1.81 the value is 1 and if not, then 0.
The model is as follows:

$$D_{it} = \beta_0 + \beta_1 X_{it1} + \beta_2 X_{it2} + \beta_3 X_{it3} + \beta_4 X_{it4} + \beta_5 X_{it5} + \beta_6 X_{it6} + \beta_7 X_{it7} + \beta_8 X_{it8} + \beta_9 X_{it9}$$

- D_{it} stands for the Distress Binary variable.
- X_{it1} for firm size.
- X_{it2} for current ratio.
- X_{it3} for operating margin.
- X_{it4} for revenue growth.
- X_{it5} for debt to assets.
- X_{it6} for capital expenditures disinvestment.
- X_{it7} for R&D intensity.
- X_{it8} for R&D intensity squared.
- X_{it9} for the interaction term between R&D intensity and firm size.

The motivation for using this classification model was to provide an alternative besides the regression model with a continuous output. A model such as this also holds practical relevance to stakeholders such as they don't just want to know how shifting financial ratios will affect the likelihood of distress, but also are in need of a clear answer whether a business is high risk or low risk.

4.3.3 Fixed effects regression results with interaction terms using Region dummies

The following model is very similar to the first fixed-effects regression model but with one difference. This time additional interaction variables are included to account for two of the regions out of the three. The two regions that were included were Canada and Western Europe. The reason why the US was left out is because it was chosen as the baseline category due to the dummy variable trap. The motivation for choosing the US specifically as the baseline category is because usually it's good practice to choose the category with most observations respective to the other categories.
The model is as follows:

$$\begin{aligned} AZS_{it} = \beta_0 &+ \beta_1 X_{it1} + \beta_2 X_{it2} + \beta_3 X_{it3} \\ &+ \beta_4 X_{it4} + \beta_5 X_{it5} + \beta_6 X_{it6} \\ &+ \beta_7 X_{it7} + \beta_8 X_{it8} + \beta_9 X_{it9} \\ &+ \beta_{10} X_{it10} + \beta_{11} X_{it11} \end{aligned}$$

- AZS_{it} stands for the Z score.
- X_{it1} for firm size.
- X_{it2} for current ratio.
- X_{it3} for operating margin.
- X_{it4} for revenue growth.
- X_{it5} for debt to assets.
- X_{it6} for capital expenditures disinvestment.
- X_{it7} for R&D intensity.
- X_{it8} for R&D intensity squared.
- X_{it9} for the interaction term between R&D intensity and firm size.
- X_{it10} for the interaction term between R&D intensity,firm size and the Canada dummy.

- X_{it11} for the interaction term between R&D intensity,firm size and the Western Europe dummy.

The motivation for using this model was to observe the differences between this regression, which includes interaction terms created from dummy variables corresponding to different regions, and the first model with just the interaction term between R&D intensity and firm size. The reason why the original interaction between R&D intensity and firm size was kept is because if omitted, the triple interactions would represent absolute effects instead of differences relative to the baseline. Therefore, the coefficients of these terms would be misinterpreted. On the other hand, by keeping the original interaction term, the effect in the model is divided into the baseline effect for the US and any differences for the other regions.

4.3.4 Fixed effects regression with lagged R&D intensity

The next model will be a fixed-effects regression model, however, there is a difference from the other models discussed before. In this regression, instead of the normal R&D intensity, the lagged R&D intensity will be used. The model is the following:

$$\begin{aligned} AZS_{it} = \beta_0 &+ \beta_1 X_{it1} + \beta_2 X_{it2} + \beta_3 X_{it3} \\ &+ \beta_4 X_{it4} + \beta_5 X_{it5} + \beta_6 X_{it6} \\ &+ \beta_7 X_{it7} + \beta_8 X_{it8} + \beta_9 X_{it9} \end{aligned}$$

- AZS_{it} stands for the Z score.
- X_{it1} for firm size.
- X_{it2} for current ratio.
- X_{it3} for operating margin.
- X_{it4} for revenue growth.
- X_{it5} for debt to assets.
- X_{it6} for capital expenditures disinvestment.
- X_{it7} for lagged R&D intensity.

- X_{it8} for R&D intensity squared.

- X_{it9} for the interaction term between R&D intensity and firm size.

The motivation for including this model in this paper was to observe whether there were any time-delayed effects of R&D intensity. It is possible that R&D investments may not show their financial returns immediately and affect firm performance negatively in the short-term since they are expenditures, but may have a positive effect on a firm's financial health on the long run, in particular for large enough companies.

4.3.5 Fixed effects regression with Recession dummy

Next is the last regression model that will be used in this paper. This regression will observe how R&D intensity affects a firm's financial health when a macroeconomic shock is introduced like a recession. In the time period of 2014 - 2024, the global economy has experienced only one economic recession, which was in 2020 due to the exogenous shock from the Covid 19 pandemic. To introduce this recession variable in the model, a dummy variable will be added, with a value of 1 if the observation occurs in 2020 and 0 otherwise.

$$\begin{aligned} AZS_{it} = \beta_0 &+ \beta_1 X_{it1} + \beta_2 X_{it2} + \beta_3 X_{it3} \\ &+ \beta_4 X_{it4} + \beta_5 X_{it5} + \beta_6 X_{it6} \\ &+ \beta_7 X_{it7} + \beta_8 X_{it8} + \beta_9 X_{it9} \\ &+ \beta_{10} X_{it10} \end{aligned}$$

- AZS_{it} stands for the Z score.

- X_{it1} for firm size.

- X_{it2} for current ratio.

- X_{it3} for operating margin.

- X_{it4} for revenue growth.

- X_{it5} for debt to assets.

- X_{it6} for capital expenditures disinvestment.

- X_{it7} for lagged R&D intensity.
- X_{it8} for R&D intensity squared.
- X_{it9} for the interaction term between R&D intensity and firm size.
- X_{it10} for the Recession dummy variable.

5 Data

5.1 Sample description

5.1.1 Missing values

First of all, before discussing the results from the regression analysis of the data, it would be better to talk in greater detail about the sample data. In the original dataset, there was a lot of missing information for all the variables as follows.

	Missing Values	Percentage
RD_Expense	4067	42.671283
Debt/Equity	3800	39.869898
Rev growth	3506	36.785227
Operating Margin	3430	35.987829
CAPEX	3384	35.505194
Debt_Assets	3243	34.025811
Current_Ratio	3236	33.952366
Total_Assets	3221	33.794985
AZS	2972	31.182457
Ticker	0	0.000000
Name	0	0.000000
Year	0	0.000000

As it can be seen above, the amount of missing information exceeds 30% for all the variables which is quite a large amount, and for this reason the most rational approach was to completely eliminate all observations of any firm that had any missing values no matter the year. The biggest motivation for eliminating all observations was that this big of an amount of missing data would definitely affect the regression results by introducing bias and reducing statistical. The reason for removing all observations of any firm that had any missing values ensure the consistency of the information after the elimination.

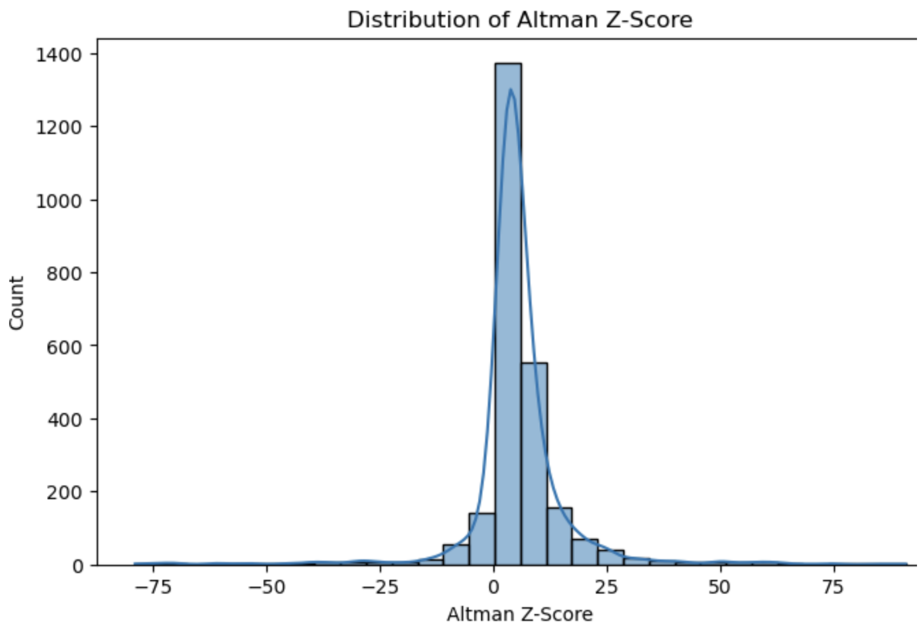
5.1.2 Data adjustment

However, in the data, there were still problems that needed to be addressed. Capital expenditures had only negative values and these values were generally pretty big.

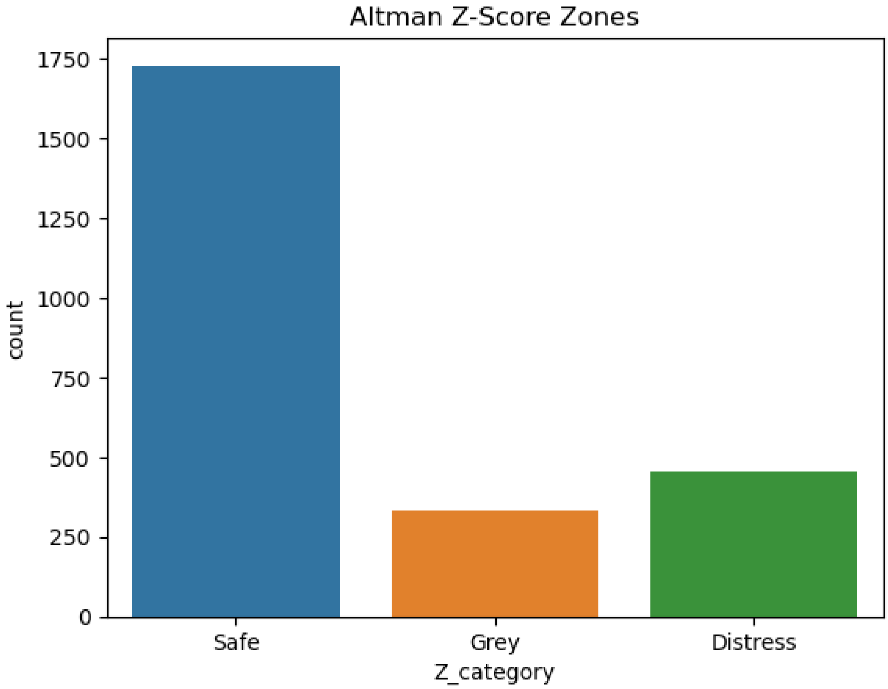
From general accounting knowledge, a positive value of capital expenditures means that the firm has spend cash in investing in assets like property, plant or equipment while a negative value implies that instead there have been more sales of assets than investing. From this definition one may think that the negative values from the dataset also mean the same but this is not the case. From information provided from the Bloomberg terminal for this variable it is known that values are recorded as negative due to accounting convention because all cash outflows to acquire new assets are recorded in the balance sheet with a negative sign. Due to this, it logically made sense to first of all apply scaling by dividing Capital Expenditures by Total Assets. Then, to have all values obtained to be multiplied by -1.

5.1.3 Explanatory data analysis

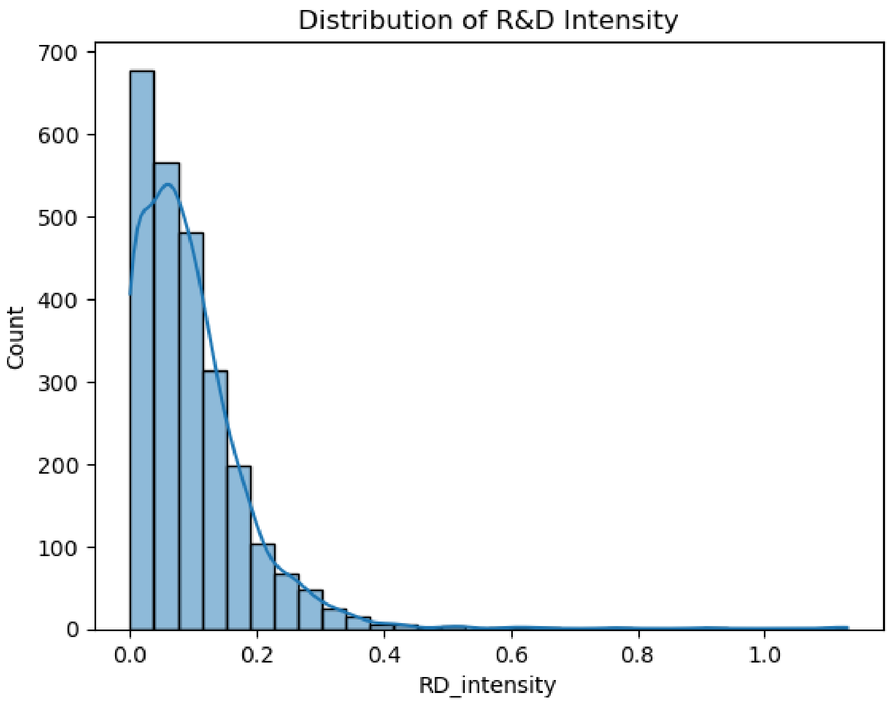
In order to obtain greater information from the sample data, the explanatory data analysis was performed. The motivation for doing such an analysis is to learn about how the values for each variable are distributed and if there are any potential outliers that could be problematic if not handled properly when running the regression analysis later on because of the possibility of introducing a bias.



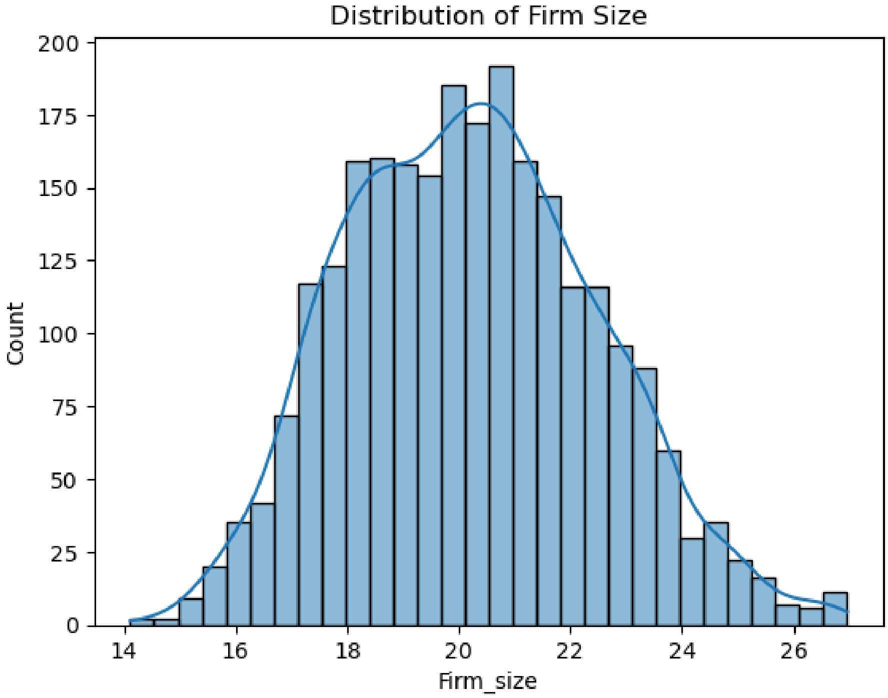
In the above graphical representation of the distribution of the Altman Z-score variable most values are concentrated around the 0-10 range. There are also some outliers outside of this range. In the original study by Altman(1968), any value below 1.81 translates to high financial risk and any value above 2.99 implies low risk. But from the graph above some extreme values can also be observed like -50 or 75. To deal with these outliers the method of winsorizing was used. This method replaced all values above the 99th quantile with the 99th quantile value and all values below the 1st quantile with the 1st quantile value. Since outliers like -50 are 75 are very extreme, this method is appropriate.



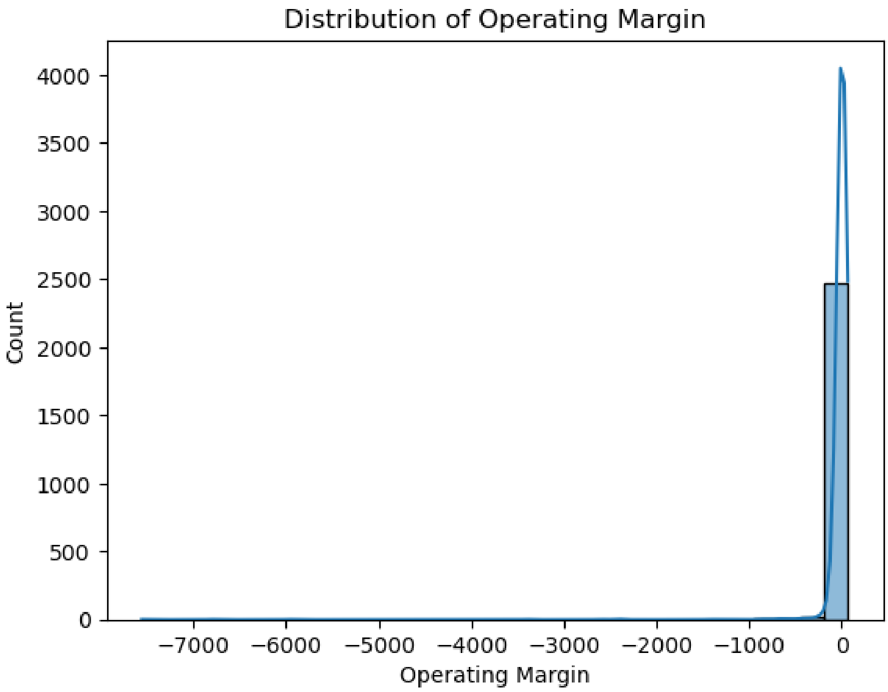
Above is also a bar chart of a self-created variable called Z-category obtained by dividing the values of the Z-score in three regions as discussed also in Altman(1968) which was mentioned before. By the results of the chart it looks like there are much more safe observations than those in the Grey and Distress regions.



This figure is a visual representation of the distribution of the main independent variable of interest, R&D intensity. From the results, it looks like the distribution is right-skewed. Most values are centered around the 0.0 to 0.2 range while those out of this range seem to be at most just a little bigger than 1.0. From observing the figure, it looks like there also aren't any outliers so no winsorizing is needed. However, in the previous section was mentioned that in one of the regression models used in this paper, there will be an interaction variable between R&D intensity and firm size. As both R&D intensity and the interaction term will both be included in the model, the problem of multicollinearity might arise. Therefore in this case it is a good choice to centralize R&D intensity. To centralize a variable means to subtract all of its observation values from the mean.

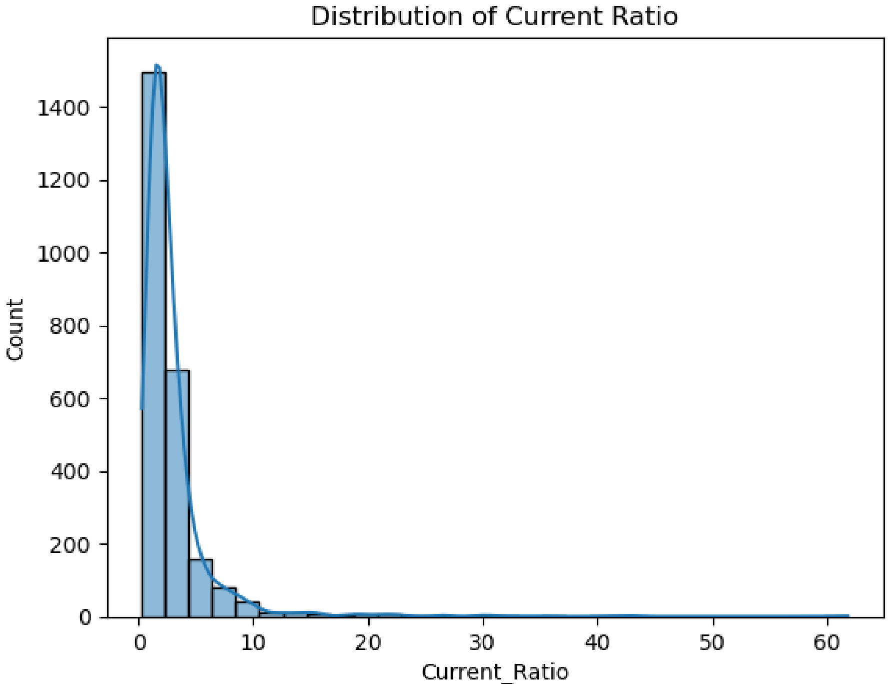


This figure shows the distribution of the variable Firm Size. From the results there don't seem to be any outliers. Most values seem to be around the 14-26 range and no outliers can be observed. Therefore, no winsorizing is needed. However, since both this variable and the interaction term created from R&D intensity and firm size, then firm size also must be centralized.

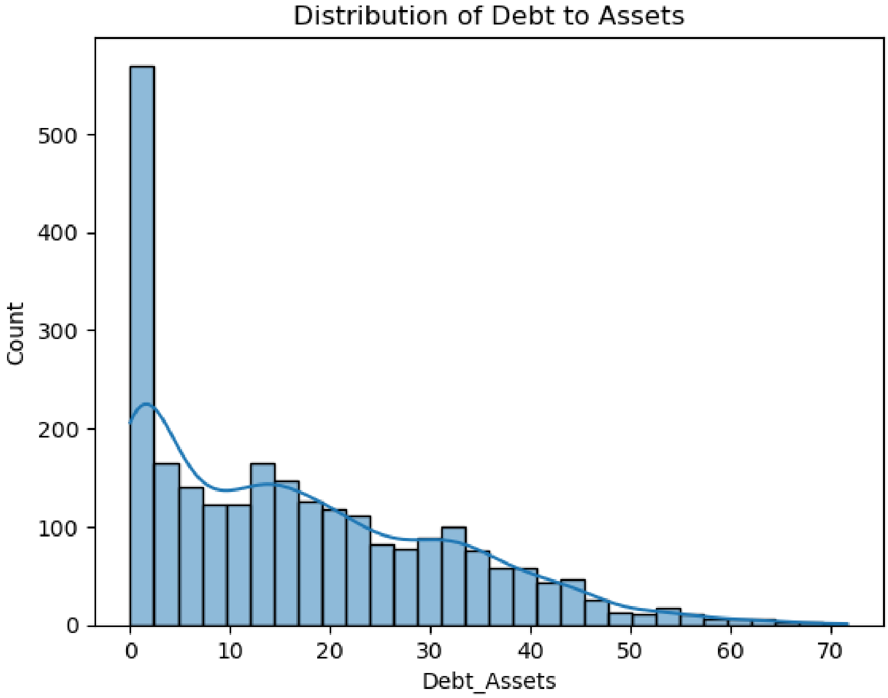


The figure above is the representation of the distribution of the operating margin.

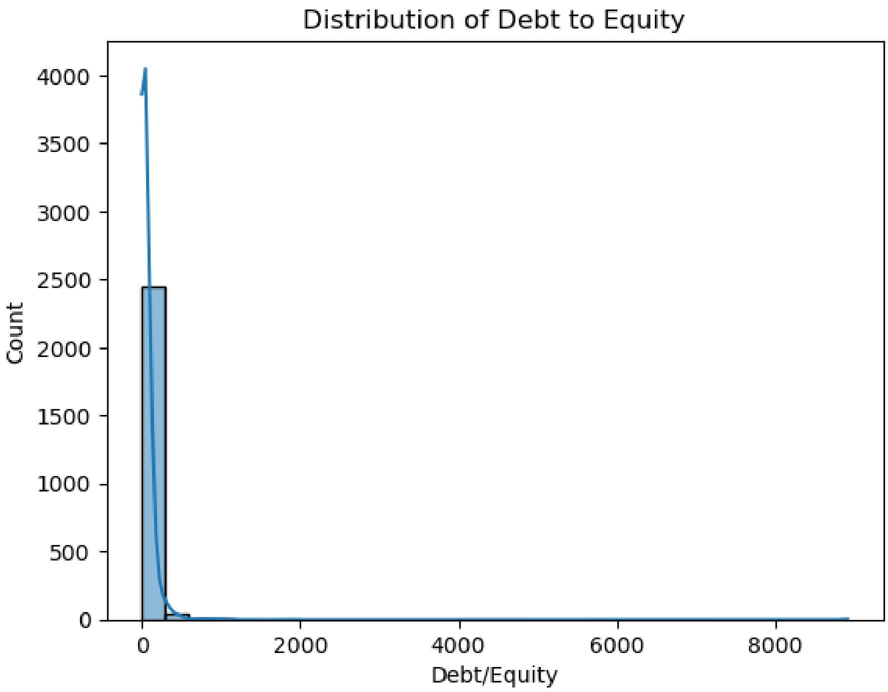
This distribution is a left-skewed one. Most observations in the distribution seem more concentrated around 0 while values like -4000 or -7000 are extreme outliers. Winsorizing in this case is necessary.



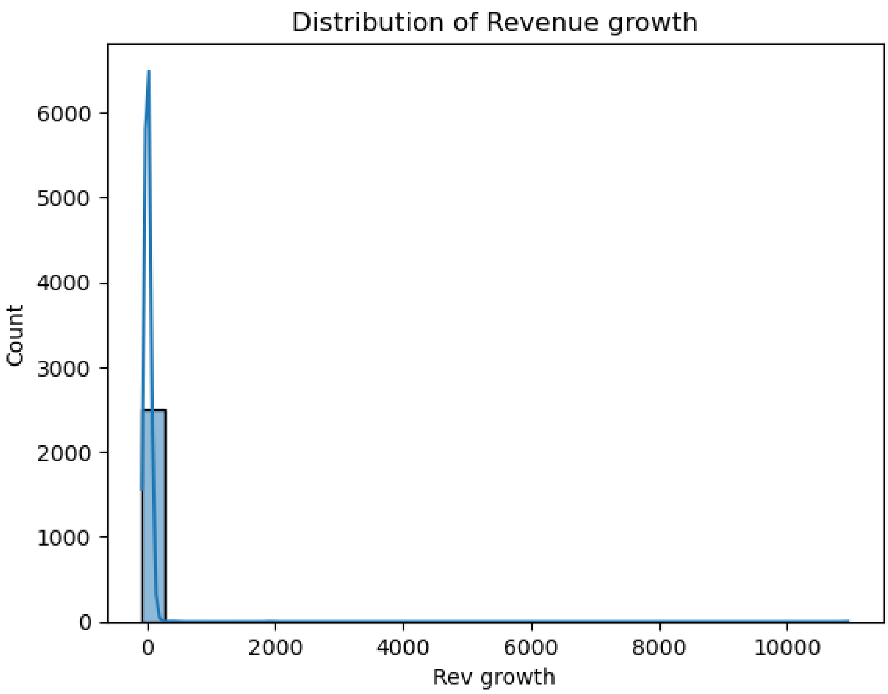
The figure above is the visual representation of the current ratio. Most observations are concentrated in the range of 0-10. From observing the plot, it is a right-skewed distribution. Outliers can also be observed in the plot, for example values like 30 or 50. Winsorizing needs to be applied in this case before using current ratio in the regression analysis.



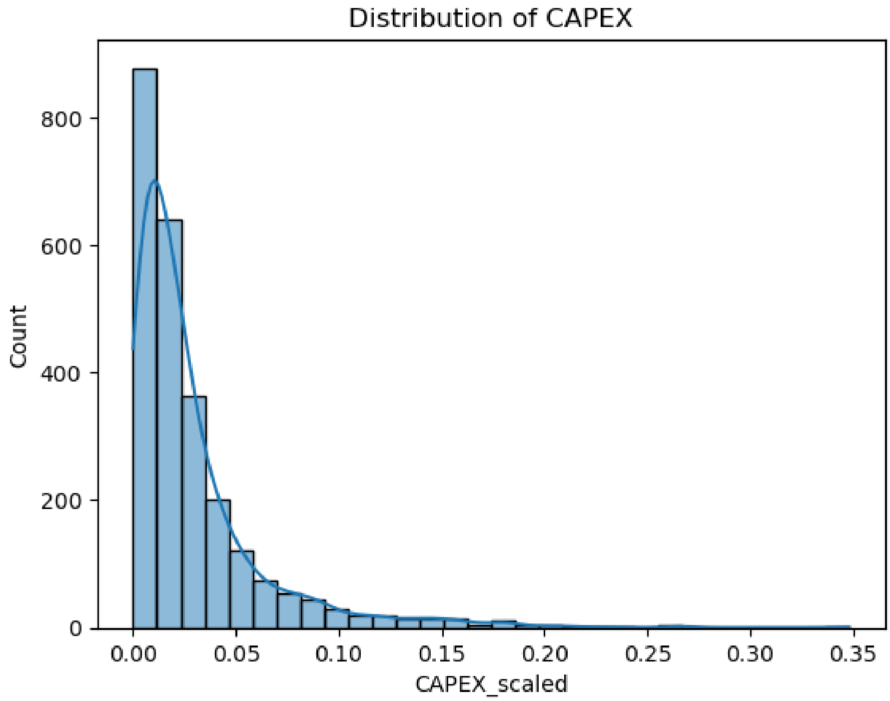
Next is the visual representation of the debt-to-assets variable. From just looking at the plot it can be determined immediately that this is a right-skewed distribution. As debt to assets is a ratio, some values like 20 or 40 are usually considered outliers as they are too big, however not in this case due to the considerable number of observations in that range of values. On the other hand, for values above 50, the number of observations is too low to be acceptable, and those observations will be considered as outliers. From this train of thought, using winsorizing is appropriate.



The next plot shows the distribution of debt to equity ratio. By first glance, most observations in the distribution are centered around 0 and the presence of extremely big outliers like 6000 is concerning. In order for no problems to come up during the regression analysis later on, the debt-to-equity variable has to be winsorized first.

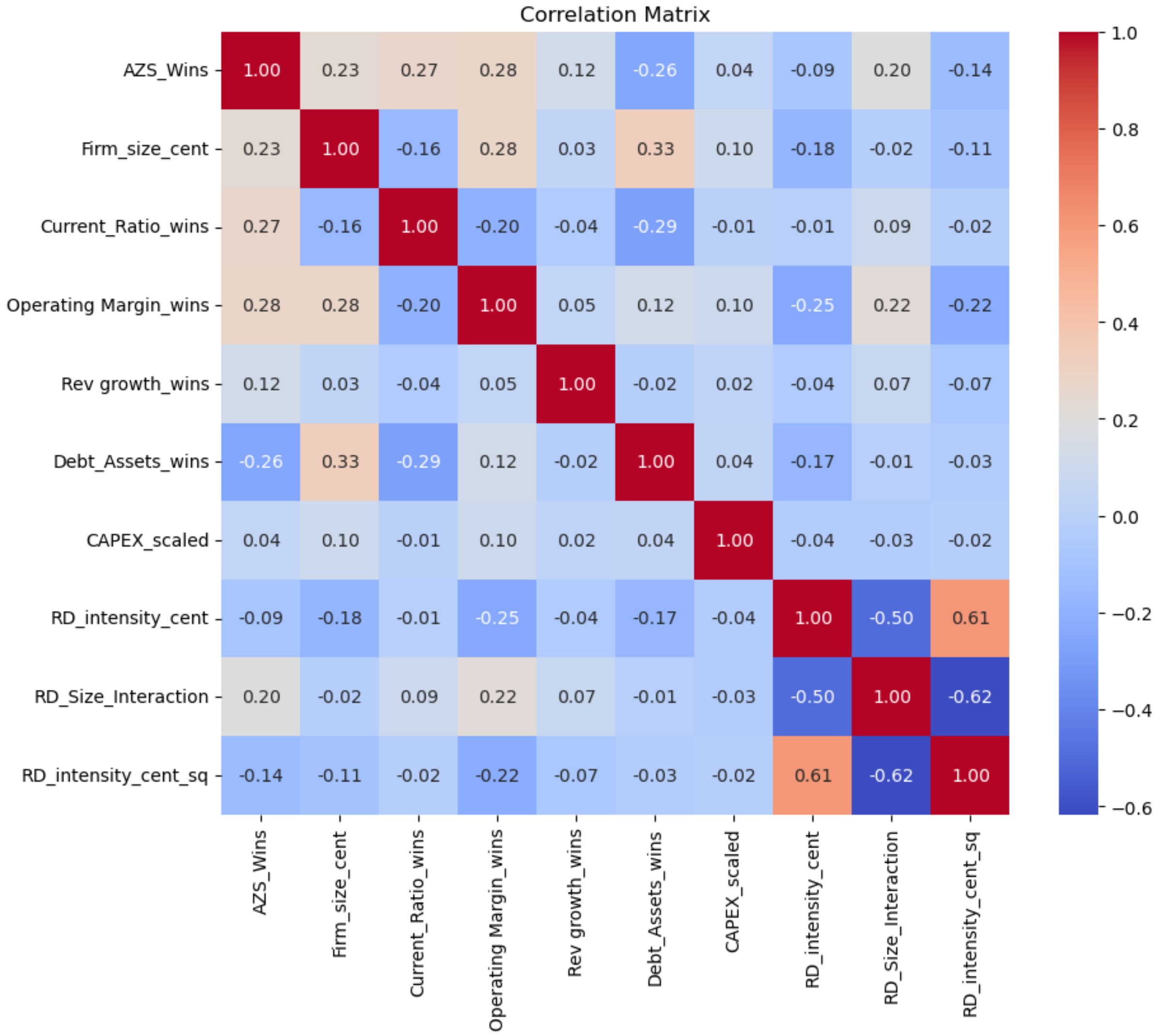


This plot shows the distribution of the revenue growth variable. For most observations, the point they are centered around is near 0. All observations further to the right can be considered outliers, so winsorizing will be applied.



Above is the last plot that will be shown. It shows the distribution of the scaled capital expenditures variable that was created by multiplying the previous values by -1 and dividing them by total assets. Fortunately, in the plot there don't seem to be any outliers so now further intervention is needed.

5.2 Correlation matrix



5.3 Summary statistics

	AZS_Wins	Firm_size_cent	RD_intensity_cent	Current_Ratio_wins
count	2511.000000	2.511000e+03	2.511000e+03	2511.000000
mean	4.433321	3.369489e-14	1.127467e-17	2.834181
std	4.388267	2.239611e+00	8.808085e-02	2.728951
min	-10.000000	-6.109379e+00	-9.402280e-02	0.546770
25%	2.446550	-1.704104e+00	-6.044936e-02	1.352400
50%	4.382200	-7.954774e-02	-1.770393e-02	2.018000
75%	7.619800	1.523762e+00	3.637411e-02	3.195900
max	10.000000	6.742775e+00	1.036375e+00	18.956810

Table 1: Summary Statistics (Columns 1–4)

	Operating Margin_wins	Rev growth_wins	Debt_Assets_wins	CAPEX_scaled
count	2511.000000	2511.000000	2511.000000	2511.000000
mean	-1.733047	10.192435	16.531657	0.028802
std	44.164922	22.999051	14.492424	0.034317
min	-327.422820	-45.064340	0.000000	-0.000000
25%	-2.369300	-2.080250	3.021750	0.008041
50%	6.062300	8.204200	14.039000	0.017959
75%	14.496450	19.540950	26.714850	0.035087
max	39.777810	106.640520	55.395940	0.348251

Table 2: Summary Statistics (Columns 5–8)

5.4 Regression results

5.4.1 Fixed effects regression

Dep. Variable:	AZS_Wins	**R-squared:**	0.2841
Estimator:	PanelOLS	**R-squared (Between):**	-0.0242
No. Observations:	2511	**R-squared (Within):**	0.2830
Date:	Fri, Sep 12 2025	**R-squared (Overall):**	0.0086
Time:	19:40:33	**Log-likelihood**	-4885.1
Cov. Estimator:	Clustered		
		F-statistic:	97.658
Entities:	279	**P-value**	0.0000
Avg Obs:	9.0000	**Distribution:**	F(9,2215)
Min Obs:	9.0000		
Max Obs:	9.0000	**F-statistic (robust):**	21.102
		P-value	0.0000
Time periods:	9	**Distribution:**	F(9,2215)
Avg Obs:	279.00		
Min Obs:	279.00		
Max Obs:	279.00		

	Parameter	**Std. Err.**	**T-stat**	**P-value**	**Lower CI**	**Upper CI**
Firm_size_cent	0.8383	0.2747	3.0522	0.0023	0.2997	1.3770
Current_Ratio_wins	0.3126	0.0819	3.8158	0.0001	0.1519	0.4732
Operating Margin_wins	0.0234	0.0043	5.4123	0.0000	0.0149	0.0318
Rev growth_wins	0.0049	0.0028	1.7280	0.0841	-0.0007	0.0104
Debt_Assets_wins	-0.0983	0.0092	-10.732	0.0000	-0.1163	-0.0803
CAPEX_scaled	7.8043	2.4983	3.1239	0.0018	2.9051	12.703
RD_intensity_cent	1.4202	2.4952	0.5692	0.5693	-3.4730	6.3135
RD_Size_Interaction	2.4080	0.9616	2.5042	0.0123	0.5223	4.2937
RD_intensity_cent_sq	1.5887	4.7677	0.3332	0.7390	-7.7610	10.938

F-test for Poolability: 27.021

P-value: 0.0000

Distribution: F(286,2215)

Included effects: Entity, Time

5.4.2 Conditional fixed effects logit regression

Dep. Variable:	Distress_Binary	**R-squared:**	0.1224
Estimator:	PanelOLS	**R-squared (Between):**	0.2788
No. Observations:	2511	**R-squared (Within):**	0.1236
Date:	Sun, Sep 14 2025	**R-squared (Overall):**	0.2295
Time:	01:55:55	**Log-likelihood**	211.34
Cov. Estimator:	Clustered		
		F-statistic:	34.339
Entities:	279	**P-value**	0.0000
Avg Obs:	9.0000	**Distribution:**	F(9,2215)
Min Obs:	9.0000		
Max Obs:	9.0000	**F-statistic (robust):**	9.2912
		P-value	0.0000
Time periods:	9	**Distribution:**	F(9,2215)
Avg Obs:	279.00		
Min Obs:	279.00		
Max Obs:	279.00		

	Parameter	**Std. Err.**	**T-stat**	**P-value**	**Lower CI**	**Upper CI**
Firm_size_cent	-0.0623	0.0305	-2.0402	0.0415	-0.1222	-0.0024
Current_Ratio_wins	-0.0208	0.0070	-2.9628	0.0031	-0.0346	-0.0070
Operating Margin_wins	-0.0022	0.0005	-4.2433	0.0000	-0.0033	-0.0012
Rev growth_wins	-0.0006	0.0003	-2.0482	0.0407	-0.0012	-2.525e-05
Debt_Assets_wins	0.0079	0.0012	6.5331	0.0000	0.0055	0.0102
CAPEX_scaled	-0.3613	0.4727	-0.7645	0.4447	-1.2883	0.5656
RD_intensity_cent	0.1020	0.2842	0.3589	0.7197	-0.4554	0.6594
RD_Size_Interaction	-0.0558	0.0932	-0.5982	0.5498	-0.2386	0.1270
RD_intensity_cent_sq	-0.1673	0.3806	-0.4397	0.6602	-0.9137	0.5790

F-test for Poolability: 11.101

P-value: 0.0000

Distribution: F(286,2215)

Included effects: Entity, Time

5.4.3 Fixed effects regression results with interaction terms using Region dummies

Dep. Variable:	AZS_Wins	**R-squared:**	0.2844
Estimator:	PanelOLS	**R-squared (Between):**	-0.0254
No. Observations:	2511	**R-squared (Within):**	0.2832
Date:	Fri, Sep 12 2025	**R-squared (Overall):**	0.0076
Time:	19:40:33	**Log-likelihood**	-4884.6
Cov. Estimator:	Clustered		
		F-statistic:	87.982
Entities:	279	**P-value**	0.0000
Avg Obs:	9.0000	**Distribution:**	F(10,2214)
Min Obs:	9.0000		
Max Obs:	9.0000	**F-statistic (robust):**	19.053
		P-value	0.0000
Time periods:	9	**Distribution:**	F(10,2214)
Avg Obs:	279.00		
Min Obs:	279.00		
Max Obs:	279.00		

	Parameter	**Std. Err.**	**T-stat**	**P-value**	**Lower CI**	**Upper CI**
Firm_size_cent	0.8489	0.2814	3.0169	0.0026	0.2971	1.4007
Current_Ratio_wins	0.3121	0.0819	3.8115	0.0001	0.1515	0.4727
Operating Margin_wins	0.0234	0.0044	5.3340	0.0000	0.0148	0.0321
Rev growth_wins	0.0048	0.0029	1.6878	0.0916	-0.0008	0.0104
Debt_Assets_wins	-0.0981	0.0092	-10.615	0.0000	-0.1162	-0.0800
CAPEX_scaled	7.9111	2.5684	3.0802	0.0021	2.8744	12.948
RD_intensity_cent	1.7354	2.1716	0.7992	0.4243	-2.5231	5.9939
RD_Size_Interaction	2.5257	1.0514	2.4022	0.0164	0.4638	4.5877
RD_WesternEurope_Firm	-0.6862	1.4751	-0.4652	0.6419	-3.5790	2.2066
RD_Canada_Firm	-0.7492	1.7169	-0.4364	0.6626	-4.1162	2.6178

F-test for Poolability: 26.499

P-value: 0.0000

Distribution: F(286,2214)

Included effects: Entity, Time

5.4.4 Fixed effects regression with lagged R&D intensity

Dep. Variable:	AZS_Wins	**R-squared:**	0.2891
Estimator:	PanelOLS	**R-squared (Between):**	-0.0231
No. Observations:	2510	**R-squared (Within):**	0.2890
Date:	Fri, Sep 12 2025	**R-squared (Overall):**	0.0096
Time:	19:40:33	**Log-likelihood**	-4874.9
Cov. Estimator:	Clustered		
		F-statistic:	112.58
Entities:	279	**P-value**	0.0000
Avg Obs:	8.9964	**Distribution:**	F(8,2215)
Min Obs:	8.0000		
Max Obs:	9.0000	**F-statistic (robust):**	25.209
		P-value	0.0000
Time periods:	9	**Distribution:**	F(8,2215)
Avg Obs:	278.89		
Min Obs:	278.00		
Max Obs:	279.00		

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
Firm_size_cent	0.7000	0.2438	2.8712	0.0041	0.2219	1.1782
Current_Ratio_wins	0.3058	0.0800	3.8230	0.0001	0.1489	0.4627
Operating Margin_wins	0.0237	0.0043	5.4926	0.0000	0.0152	0.0322
Rev growth_wins	0.0045	0.0029	1.5242	0.1276	-0.0013	0.0102
Debt_Assets_wins	-0.0974	0.0088	-11.022	0.0000	-0.1148	-0.0801
CAPEX_scaled	7.6051	2.4712	3.0775	0.0021	2.7590	12.451
RD_intensity_lag1	-2.8534	1.1462	-2.4895	0.0129	-5.1010	-0.6057
RD_Size_Interaction	1.6832	0.5163	3.2600	0.0011	0.6707	2.6957

F-test for Poolability: 27.401

P-value: 0.0000

Distribution: F(286,2215)

Included effects: Entity, Time

5.4.5 Fixed effects regression with Recession dummy

Dep. Variable:	AZS_Wins	**R-squared:**	0.2851
Estimator:	PanelOLS	**R-squared (Between):**	0.0004
No. Observations:	2511	**R-squared (Within):**	0.2851
Date:	Fri, Sep 12 2025	**R-squared (Overall):**	0.0308
Time:	20:10:12	**Log-likelihood**	-4929.3
Cov. Estimator:	Clustered		
		F-statistic:	88.615
Entities:	279	**P-value**	0.0000
Avg Obs:	9.0000	**Distribution:**	F(10,2222)
Min Obs:	9.0000		
Max Obs:	9.0000	**F-statistic (robust):**	24.780
		P-value	0.0000
Time periods:	9	**Distribution:**	F(10,2222)
Avg Obs:	279.00		
Min Obs:	279.00		
Max Obs:	279.00		

	Parameter	**Std. Err.**	**T-stat**	**P-value**	**Lower CI**	**Upper CI**
Firm_size_cent	0.6364	0.2019	3.1516	0.0016	0.2404	1.0324
Current_Ratio_wins	0.3275	0.0806	4.0627	0.0001	0.1694	0.4855
Operating Margin_wins	0.0238	0.0040	5.9146	0.0000	0.0159	0.0316
Rev growth_wins	0.0066	0.0025	2.6269	0.0087	0.0017	0.0116
Debt_Assets_wins	-0.0991	0.0085	-11.617	0.0000	-0.1158	-0.0824
CAPEX_scaled	7.5499	2.3691	3.1869	0.0015	2.9041	12.196
RD_intensity_cent	0.4825	2.1829	0.2210	0.8251	-3.7983	4.7633
RD_Size_Interaction	2.3632	0.8944	2.6423	0.0083	0.6093	4.1171
RD_intensity_cent_sq	2.1402	4.4678	0.4790	0.6320	-6.6213	10.902
Recession	0.0782	0.1229	0.6364	0.5246	-0.1628	0.3192

F-test for Poolability: 26.643

P-value: 0.0000

Distribution: F(278,2222)

Included effects: Entity

5.5 VIF test results

	Variable	VIF
0	AZS_Wins	2.328435
1	Firm_size_cent	1.267987
2	Current_Ratio_wins	2.275663
3	Operating Margin_wins	1.309262
4	Rev growth_wins	1.200125
5	Debt_Assets_wins	1.616666
6	CAPEX_scaled	1.590870
7	RD_intensity_cent	1.761154
8	RD_Size_Interaction	1.833024
9	RD_intensity_cent_sq	2.099051

6 Discussion

6.1 Correlation matrix

A correlation matrix is a table that shows all pairwise correlations between the variables in the study. Each entry in the table shows how strong the linear relationship is between any two variables and how that relationship is directed. The direction is determined through the positive sign (+) which indicates a positive correlation or the negative sign (-) that indicates a negative correlation. The strength is observed from what value the entry has in the 0-1 range in the case of a positive correlation and -1 to 0 in the case of a negative correlation.

The motivation for including the correlation matrix in this paper is that it can allow any reader to see whether the relationships between the variables align with the theoretical assumptions made earlier. The secondary reason is that the matrix also makes it easy to check for multicollinearity. When the results are not quite from what is expected, changes can be made before running any regression.

The most important thing to look for when analyzing relationships in a correlation matrix is to see the correlations between the dependent variable, the Z-score, and any of the independent variables. In the matrix, the Z-score is represented by AZS_Wins, which means winsorized Altman Z-Score.

The correlation between AZS_Wins and Firm_size_cent (centralized firm size) is 0.23. This is a weak positive correlation, which implies that there exists a positive linear relationship between the two variables. However, the effect of firm size on the Z-score is weak. This matches with the assumption that higher firm size is associated with less financial risk, meaning higher Z-score.

The correlation between AZS_Wins and Current_Ratio_wins (winsorized current ratio) is 0.27. This is again a weak positive correlation, similarly to before. This result aligns with the assumption made before that a higher ratio implies better short-term financial health due to the firm having higher liquidity.

The following correlation between AZS_Wins and Operating Margin_wins (winsorized operating margin) is 0.28. Like in the previous cases, this is a weak positive correlation. This correlation also matches with the previous assumption that a higher operating margin is associated with a higher Z-score since higher profitability implies better financial health.

The correlation between AZS_Wins and Rev growth_wins (winsorized revenue growth) is 0.12. This value shows a weak positive correlation like before. However, the effect that operating margin has on the Z-score is more weaker than that of the previously discussed variables due to the correlation having a lower value positive value. Higher revenue growth is expected to be correlated to a higher Z-score and the results align with this expectation.

The next correlation is that of the Z-score and Debt_Assets_wins (winsorized debt to assets ratio). The value is -0.26, implying a weak negative linear relationship between the two variables. This value aligns with the assumption that higher debt-to-assets is associated with higher financial risk, since a higher ratio means a growing number of debt obligations which will most likely negatively affect financial health.

The next correlation to be discussed is that of the Z score and CAPEX_scaled with a value of 0.04. A positive correlation is expected according to theory since for a firm to have a lot of capital expenditures, it must be in good financial health.

The following correlation is between that of R&D_intensity_cent (R&D_intensity centralized) and the Z-score, which has a value of -0.09. This value implies a weak negative correlation that matches with the theory that higher investments in R&D might put a strain on a firm's short-term financial health.

On the other hand the interaction term between R&D_intensity and Firm_size has a correlation of 0.2 with the Z-score. This value suggests that both terms have a positive linear relationship. This result aligns with the main hypothesis in this study that R&D intensity has a positive effect on the financial health of firms with bigger size.

The last relationship that needs to be discussed is that of R&D_intensity_cent_sq (the squared R&D_intensity centralized). The correlation is -0.14 which is expected from the relationship between R&D intensity and the Z-score.

Besides these relationships, high correlations between independent variables must also be discussed as multicollinearity might happen. There is a positive correlation of 0.61 between R&D_intensity_cent and R&D_intensity_cent_sq, however this is justified since R&D_intensity_cent_sq is the square of R&D_intensity.

Similarly, the correlation of -0.5 between R&D_intensity_cent and the interaction term is justifiable since R&D_intensity_cent is one of the variables used to make the interaction term.

The same holds true also for the correlation of -0.62 between the interaction term and R&D_intensity_cent_sq since R&D_intensity_cent is one of the variables used to make the interaction term and R&D_intensity_cent_sq.

6.2 Regression results

6.2.1 Fixed effects results

In this fixed-effects regression model, from looking at the p-value column it is easy to determine which variables are not statistically significant. The only variables with a p-value higher than 0.05 were RD_intensity_cent, RD_intensity_cent_sq and lastly Rev growth_wins with values of 0.5693, 0.7390 and 0.0841 respectively.

On the other hand, the interaction term ,RD_Size_Interaction, between RD_intensity and firm size, is statistically significant with a p-value of 0.0123 and a coefficient of 2.4080.

It can be interpreted that neither the main effect of RD_intensity, nor its marginal effect, have any clear impact on the Z-score, but the effect changes depending on the firm size. As the coefficient of the interaction term is positive, the strategy of investing in R&D to improve the firm's financial health is more beneficial to larger firms. The reason why it serves more bigger firms is because firm size is significant in the regression with a p-value of 0.0023 and a coefficient of 0.8383. The main effect of firm size on the z-score in this case is clear and positive, which indicates that bigger firms tend to have a higher z-score. A higher z-score for a firm implies that the firm is more likely to be financially safe. These results are consistent with the original hypothesis that R&D intensity positively affects the financial health of firms large enough.

As for revenue growth, due to not being significant, it is indicated that revenue growth does not have any impact on the financial health of a firm. This creates a contradiction with the original hypothesis, where revenue growth was considered a factor that could influence the Z-score.

Current ratio is significant for the regression with a p-value of 0.0001 and a coefficient of 0.3126. From the results it is indicated that there is a positive linear relationship between the current ratio and the Z-score, which aligns with the theory discussed in the earlier sections.

Operating margin is also significant for the regression, having a p-value of approximately 0 and a coefficient of 0.0234. As this coefficient is positive but extremely small, it can be interpreted that the main effect of operating margin on the Z-score is positive, but nearly negligible.

The next significant variable that will be discussed is the debt-to-assets ratio. It's p-value is almost 0 and its coefficient is -0.0983. A negative value is expected and consistent with the previously established theory. The results indicate that even though

the debt-to-assets ratio negatively affects the Z-score, the effect is very weak due to the coefficient being very close to 0.
The next variable to talk about is capital expenditures. Due to being significant from its p-value of 0.0018 and its coefficient of 7.8043, it can be determined that the main effect of capital expenditures has a positive impact on the z-score.

6.2.2 Conditional fixed effects logit regression

In this model, the main effect of R&D intensity, RD_intensity_cent, its interaction with firm size, RD_Size_Interaction and the non-linear effect, RD_intensity_cent_sq, are not statistically significant. This suggests that in this model the effect of in-firm changes of R&D investment in crossing the distress threshold is uncertain.
By comparing both the findings of this model and those of the previous one, it can be stated that R&D intensity is better suited for the continuous Z-score model than its binary counterpart.
Differently from the first regression, the variable that represents capital expenditures, CAPEX scaled, is statistically insignificant, but on the other hand, Rev growth_wins, which was insignificant, now has a p-value of 0.0407 making it statistically significant. However, the effect that revenue growth has on the probability of financial distress is extremely weak and almost negligible due to the coefficient being -0.0006. It can be noticed that the coefficient has a negative sign which implies that an increase of revenue growth is associated with a lower probability of financial distress. This result is consistent with the theoretical assumptions made in the previous sections.
Similarly, Operating_Margin_wins, Current_Ratio_wins and Firm_size_cent are also significant and with negative coefficients which indicates that higher units for any of the variables are associated with a lower probability of financial distress. These results align with the theoretical expectation as well as the results of the first regression.
As for the debt-to-assets ratio, represented by Debt_Assets_wins, while it is significant, its coefficient is 0.0079. It can be interpreted that higher values of the ratio are related to a higher probability of financial distress. This evidence further supports the statement that the debt-to-assets ratio has a negative impact on financial health. However, its effect is very weak as the coefficient is extremely small.

6.2.3 Fixed effects regression results with interaction term using Region dummies

From the results, the variables RD_WesternEurope_Firm and RD_Canada_Firm, the triple interaction terms created using RD_Size_Interaction and the dummies related to Western Europe and Canada don't seem to have any additional difference on the Z-score relative to the US(the baseline) as their p-values are both greater than 0.05, making them statistically insignificant. This does not mean that the terms have no effect on a firm's financial health, but rather that their effects are not statistically different from the United States. Therefore, the geographical location of a firm does not matter when talking about the Z-score.

The effect of the baseline is captured by the main effect, RD_Size_Interaction, and it seems to positively affect the Z-score due to the coefficient being 1.7354.

As RD_intensity_cent is not significant, this indicates the main effect itself does not have an impact on the the probability of financial distress, but depends on the size of the firm, which is consistent with the main hypothesis. This means that investing in R&D is more beneficial to bigger firms.

Rev_growth_wins still remains non-significant and it does not have any impact on the dependent variable. This contradicts the argument that revenue growth has a positive influence in a firm's financial health.

For the rest of the independent variables of this regression, they exhibit similar effects on the Z-score like those in the first model, with slight numerical changes in only the absolute values of the coefficients, not their signs.

6.2.4 Fixed effects regression with lagged R&D intensity

In this regression, the lagged R&D intensity is statistically significant with a p-value of 0.0129 and a coefficient of -2.8534, which implies that investments in R&D previous year, have a negative impact on the Z-score of the current year. This is consistent with the assumption that investing in R&D could potentially increase the likelihood of financial distress for firms in the short-term.

However, the interaction term, RD_Size_Interaction, has a positive impact on the Z-score with a coefficient of 1.6832. It can be interpreted that while investing in R&D could put a financial strain on firms in the short-term, that effect is negative effect is weaker for bigger firms. This is still consistent with the original hypothesis.

Like before, Rev_growth_wins still is non significant and its main effect does not have

any affect on the dependent variable. This reinforces that revenue growth does not influence in any way the Z-score.
The other variables of the regression retain the same effects on the dependent variable. The only observable differences are slight changes in the numerical values of the coefficients, which are negligible.

6.2.5 Fixed effects regression with Recession dummy

At first glance, it be observed that this regression is nearly identical to the first fixed-effects regression model, but with one difference. A dummy variable for the recession in the year 2020 is also included. However, due to Recession having a p-value of 0.5246, it means that is statistically insignificant. From this it can be interpreted that a macroeconomic shock such as a recession has no observable effect on the Altman Z-score.
As for the main independent variable of focus, R&D intensity, while its main effect does not have a clear effect on the Z-score due to being insignificant, its interaction term with firm size is and has a positive impact due to the coefficient also being positive. It can be interpreted that the strategy to invest more in R&D to improve financial health is only beneficial to big enough firms.
Unlike the first regression, in this model the revenue growth is significant and has a coefficient of 0.0066, indicating that it has a positive, yet weak effect of the Z-score. This results serve as evidence for the theoretical claim made in the Methods section that a higher value of revenue growth is positively associated with a lower probability of financial distress.
As for firm size, current ratio, operating margin, debt-to-assets and capital expenditures, they exhibit the same effects as they did in the first model, strengthening the support of the theoretical expectation originally stated earlier in the paper.

6.3 VIF test

From the VIF test results table it can be observed that all VIF values are smaller than 5, which indicates that there is no multicollinearity. This means that the regression models are reliable and stable and the results of their analysis are credible.

7 Conclusions

In the end, after the analysis of all five regressions the following conclusions were made.

Firstly, investing in R&D does not have any immediate returns in the current period for companies not large enough, and its effects for all companies can only be observed starting from the next period of time. In the short-term future, an investment in R&D can potentially put a company in financial distress, but this is more likely only for the smaller firms as they don't have enough capital. For the larger companies, a R&D investment is less likely to experience any financial difficulty as their larger size indicates that they had enough resources to afford the investments. These results prove the original hypothesis to be true.

Secondly, a higher revenue growth positively affects the Z-score since more revenues lead to the firm having more resources, but its contribution is so small that is negligible.

As for the current ratio, it's also a factor that decreases the likelihood of financial distress as a company with a higher ratio is more capable to repay its debts. However, its impact is weak.

Operating margin is also a factor that positively contributes to a firm's financial health. A firm with a higher operating margin leads to higher operational efficiency, which results in a lower likelihood of distress. However, this effect is so weak that it isn't significant at all.

The debt to assets ratio on the other hand has a negative effect as the higher it is, the lower the value of the Z-score. Even though across the models this variable is significant, its impact is very small and doesn't play a major role in affecting the likelihood of financial distress.

Capital expenditures have a considerate positive effect on the Z-score, which means a lower chance of distress. This is consistent with the theoretical assumptions made earlier in this paper.

As for the region where the firm is located, there doesn't seem to be a difference in Western Europe or Canada relative to the US. This means that no matter in which of the three regions a company is located, there is not any difference in the level of the Z-score holding other factors constant.

Lastly, the presence of a macroeconomic shock such as a recession which is expected to have a negative effect on a firm's financial health, doesn't have an impact of the Z-score.

8 References

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. Journal of Finance, 23(4), 589–609. URL: https://doi.org/10.1111/j.1540-6261.1968.tb00843.x

Beaver, W. H. (1966). Financial ratios as predictors of failure. Journal of Accounting Research, 4, 71–111. URL: https://doi.org/10.2307/2490171

Chan, L. K. C., Lakonishok, J., & Sougiannis, T. (2001). The stock market valuation of research and development expenditures. Journal of Finance, 56(6), 2431–2456. URL: https://doi.org/10.1111/0022-1082.00392

Griliches, Z. (1981). Market value, R&D, and patents. Economics Letters, 7(2), 183–187. URL: https://doi.org/10.1016/0165-1765(81)90049-3

Lev, B., & Sougiannis, T. (1996). The capitalization, amortization, and value-relevance of R&D. Journal of Accounting and Economics, 21(1), 107–138. URL : https://doi.org/10.1016/S0165-4101(96)00434-8

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. Journal of Finance, 29(2), 449–470. URL: https://doi.org/10.1111/j.1540-6261.1974.tb03058.x