UNIVERSITY OF LJUBLJANA
SCHOOL OF ECONOMICS AND BUSINESS

MASTER'S THESIS

THE USE OF MERTON'S CREDIT RISK MODEL FOR CREDIT SCORING OF PRIVATE FIRMS

Ljubljana, October 2022 KEVIN KNAVS

AUTHORSHIP STATEMENT

The undersigned Kevin Knavs, a student at the University of Ljubljana, School of Economics and Business, (hereafter: SEB LU), author of this written final work of studies with the title *The use of Merton's credit risk model for credit scoring of private firms*, prepared under supervision of dr. Matej Marinč.

DECLARE

1. this written final work of studies to be based on the results of my own research;
2. the printed form of this written final work of studies to be identical to its electronic form;
3. the text of this written final work of studies to be language-edited and technically in adherence with the SEB LU's Technical Guidelines for Written Works, which means that I cited and / or quoted works and opinions of other authors in this written final work of studies in accordance with the SEB LU's Technical Guidelines for Written Works;
4. to be aware of the fact that plagiarism (in written or graphical form) is a criminal offence and can be prosecuted in accordance with the Criminal Code of the Republic of Slovenia;
5. to be aware of the consequences a proven plagiarism charge based on the this written final work could have for my status at the SEB LU in accordance with the relevant SEB LU Rules;
6. to have obtained all the necessary permits to use the data and works of other authors which are (in written or graphical form) referred to in this written final work of studies and to have clearly marked them;
7. to have acted in accordance with ethical principles during the preparation of this written final work of studies and to have, where necessary, obtained permission of the Ethics Committee;
8. my consent to use the electronic form of this written final work of studies for the detection of content similarity with other written works, using similarity detection software that is connected with the SEB LU Study Information System;
9. to transfer to the University of Ljubljana free of charge, non-exclusively, geographically and time-wise unlimited the right of saving this written final work of studies in the electronic form, the right of its reproduction, as well as the right of making this written final work of studies available to the public on the World Wide Web via the Repository of the University of Ljubljana;
10. my consent to publication of my personal data that are included in this written final work of studies and in this declaration, when this written final work of studies is published.

Ljubljana, ________________________ Author's signature: _________________________

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICIES

INTRODUCTION

Proper and efficient monitoring of corporate portfolio is challenging. In the last few decades, various scope of credit scoring models have been developed to distinguish between good and bad companies (West, 2000). Scoring models usually use accounting data for their prediction, which, in general, are based on historical information as criticized by Yeh, Lin & Hsu (2012). They suggest that, for proper corporate future performance prediction, forward looking market data should be used instead of historical-based accounting ones.

One of the best-known scoring models that use market data in its future prediction is Merton's Distance to default model (Sundaresan, 2013). As described by Zvika, Ohad & Koresh (2016), Merton's Distance to default model, in general, is a structural model used for default prediction. In its prediction, it tries to estimate market value of firm's assets with using equity value, face value of debt and volatility of equity returns. They explain that, as long as estimated asset's value is above value of debt, firm would pay its debt in full. Equity value is in that case higher than zero or exactly zero when assets are lower than debt.

Merton's model is therefore very useful in future assets' value estimation, but since it needs stock market data in its prediction, it is applicable to publicly traded firms only. According to Duan, Kim, Kim & Shin (2018), the model can however be extended with using estimated distance to default parameters of public firms and applicate them to private firms. Application can be done through finding similar characteristics of private firms with a universe of public firms. Such an approach is also called Private firm extension, which utilize timely market information in default prediction of any private or public firm.

In order to check whether a market-based model can really be used for private firms, the above mentioned Private firm extension model will be developed for a set of Slovenian firms in this thesis. However, since the Slovenian stock market is relatively small and has a quite low trading volume, the sample of German public companies will be used for parameter estimation of the original Merton's Distance to default model. The reason to choose German companies is in connection and dependence of Slovenian economy with the German one.

The model developed in this thesis will therefore be built on a sample of German public firms and later used to predict default probabilities for a set of Slovenian private firms. The model should eventually represent some kind of corporate monitoring tool, since obtained records of default probabilities would allow monitoring of corporate portfolio on a monthly basis or even more frequently if needed.

The objective of the thesis is therefore to develop universal corporate monitoring tool, which could be used to properly and timely assess changes in a credit profile of a specific company. The tool will be based on a model that considers all relevant market data that implicitly contain expectations of future performance. Using forward looking market data makes the most important difference compared to other classical default predicting models that mainly use historical accounting data. Considering that, the model using market data should therefore be much faster in predicting distress or defaults than any other classical default predicting model, which will also be the first hypothesis in the thesis.

Another advantage of market-based models over classical ones is also frequency of available data. Market values are changing almost continuously while accounting data are mostly released annually. Models with using market data to predict the probability of default can therefore be run at any time and should provide the most recent market expectations of defaults. Default probabilities, based on accounting data, are on the other hand updated annually and may already not be fully relevant shortly after the announcement. These facts lead to the second hypothesis, namely that the model using market data can be used as an appropriate and efficient corporate monitoring tool by various types of actors in financial industry.

The thesis will therefore try to confirm the following two hypotheses mentioned above:

- The market-based model can predict a company's default earlier than the accounting-based models.
- The market-based model can be used in the financial industry as an efficient corporate monitoring tool.

To confirm or deny these hypotheses, the model based on theoretical construction of original Merton's Distance to default model will be built and extended in accordance with Private firm extension model. Theoretical construction contains several assumptions, some simplification will be made in some parts. Theoretical description of the original Merton's Distance to default model and its Private firm extension is presented in the literature review in Chapter 1.

The methodology of the model presented in Chapter 2 shows full model development from data collection process to estimation of private firms' distance to default values eventually. The data used in the model will be gathered from several domestic and global databases. Gathered data of German public firms will then be used to estimate several parameters, which will result in monthly distance to default values. These values will then also be used to estimate distance to default values of private firms, which would be done through finding some common characteristics of both types of firms. The chapter will therefore conclude with the record of monthly distance to default values of any Slovenian firm.

Obtained distance to default values will lastly be tested using some standard statistical tests as well as case study analysis. Accuracy of the model will be tested using some external probability of default values as a benchmark while the case study analysis will present model implication in practice. Comparison to benchmark probability of default values should therefore give an answer to the first stated hypothesis while its implication with description of the monitoring process based on a case studies would provide an answer to the second one. Model testing and its implication procedure will be presented as Results in Chapter 3.

1 LITERATURE REVIEW

The Literature review chapter will present theoretical background of the original Merton's Distance to default model and its Private firm extension. As already mentioned in the Introduction chapter, Merton's model is one of the best-known scoring models that use market data in its prediction. The assumptions and estimation process is quite straightforward and used by several different authors in their works. However, some minor simplifications or adjustments have been made by some, which will also be described in the first subchapter below.

Private firm extension, on the other hand, has been very rarely used so far. In fact, there is only a paper of Duan, Kim, Kim & Shin (2018) that tried to extend the famous Merton's model to private firms too. Theoretical background of the attempt is therefore very scarce, but the structure of the model could still be set up considering the idea of attempt. The Literature review in that part will thus include some topics from other areas, which would however be very important in later model construction process. Private firm extension attempt will be presented in the second subchapter below.

1.1 Merton's model

According to Zvika, Ohad & Koresh (2016), Merton's Distance to default model is in general a structural model used for default prediction. In its prediction, it tries to estimate market value of a firm's assets (A) with using equity value (E), face value of debt (D) and volatility of equity returns (σ_E). In its prediction, the model makes several assumptions, which should be considered in the estimation process of each parameter. The assumptions of the model are presented in Chapter 1.1.1. while the estimation process of each parameter can be found in Chapter 1.1.2.

1.1.1 Assumptions

As explained by Bharath & Shumway (2004), the model makes two critical assumptions. The first is that each respective firm has issued only one zero-coupon bond (D) with maturity in time T. Therefore, the firm can only default if the value of assets (A) is lower than the value of debt (D) at maturity T. The second important assumption of the model is that assets follow geometric Brownian motion (GBM)

$$dA = \mu_A A dt + \sigma_A A dW \tag{1}$$

where A is the total value of the firm's assets, dA is assets' derivative, μ_A is the expected continuously compounded return of assets, dt is time derivative, σ_A is the volatility of assets returns and dW is the standard Wiener process.

Geometric Brownian motion (GBM) is a stochastic, non-linear process that is usually used in asset price modelling. It can derive various sample paths that the underlying variable may follow. Notation of Brownian motion generally includes a drift term capturing growth over time (μdt) and random shocks to that growth (σdW). Since some degree of randomness is present in the GBM structure, every new simulation will always generate a new path (Paolucci, 2020).

Considering the first assumption again, the firm defaults only when the value of assets is lower than the value of debt at maturity T. In that case, creditors take over the firm and receive the entire value of assets while equity holders are left with nothing. In the opposite case when assets value is higher than the value of debt in time T, creditors receive full repayment of their debt and the equity holders keep the difference (Tanthanongsakkun & Treepongkaruna, 2008).

The payoff structure explained above therefore suggests that creditors with issuing bond actually write short put option on assets of the borrowing firm. Equity holders, on the other hand, actually hold long call option of the firm's assets. Strike price of both options is thus face value of debt (D) while maturity of them is in time T (Sundaresan, 2013).

As explained by Yeh, Lin & Hsu (2012), the value of firm equity (E) is therefore considered as a call option on the underlying assets with maturity in time T and exercise price equal to the value of debt (D). This definition is known as the Black & Scholes formula

$$E = N(d_1)A - De^{-rT}N(d_2) \tag{2}$$

where E is the market value of firm's equity, A is the market value of firm's assets, D is the face value of firm's debt, r is the instantaneous risk free rate, T is the time to maturity and N() is the cumulative normal distribution function with d_1 and d_2 given by

$$d_1 = \frac{\ln\left(\frac{A}{D}\right) + \left[r + \frac{1}{2}\sigma_A^2\right]T}{\sigma_A\sqrt{T}} \tag{3}$$

$$d_2 = d_1 - \sigma_A\sqrt{T} \tag{4}$$

According to Shumway & Bharath (2008), the Black & Scholes formula therefore explains the value of firm's equity as a function of the value of its assets. But in order to estimate value of assets, their volatility is a crucial variable in distribution equations as presented above. Black & Scholes hence further proves that asset volatility can be derived from equity volatility with equation

$$\sigma_E = \left(\frac{A}{E}\right) N(d_1)\sigma_A \tag{5}$$

where all variables have already been defined in equations above.

1.1.2 Estimation process

As further explained by Shumway & Bharath (2008), equations defined in the above chapter regarding market value of equity (E) and equity volatility (σ_E) are crucial in distance to default calculation of Merton's model. All variables defined in the two equations can be observed from the market, except market value of firm's assets (A) and its volatility (σ_A). These two variables must therefore be inferred with solving the two non-linear equations with two unknown variables.

The first step in implementing the model is to estimate equity volatility (σ_E) from either historical stock returns data or option implied volatility data. Since the data of actively traded options is very scarce, historical stock returns are more commonly used in practice. Volatility is usually calculated based on the previous 12 months, using daily equity returns (Yeh, Lin & Hsu, 2012).

The second step is to determine the face value of debt (D). Debt can easily be retrieved from financial statements of the firm, but the main issue is which value should be used. Merton's model assumes only zero-coupon bond with maturity in time T (usually T=1 year), while the firms sign several types of debt contracts with different maturities in reality. In most cases of Merton's model construction, debt is arbitrarily constructed with short term-debt (matures within 1 year) and half of long-term debt. This is the case since several covenants are usually attached to debt contracts determining immediate repayment in case of financial deterioration. In addition to short-term debt, half of long-term debt showed the most statistically significant choice in the majority of past works (Zvika, Ohad & Koresh, 2016).

The third step is to collect values of risk-free rate (r) and the market value of firm's equity (E). As a risk-free rate, annualized rate of 1-year Treasury bond is usually considered in model construction. Market value of equity on the other side is simply the product of the number of shares outstanding with its current stock price (Shumway & Bharath, 2008).

When all mentioned variables are inserted into above equations, values of firm's assets (A) and its volatility (σ_A) can finally be obtained. This is commonly done with iterative procedure as presented by Vassalou & Xing (2004). The iterative procedure begins with using equity volatility (σ_E) estimates as an initial value for assets volatility (σ_A) estimates. Then the Black & Scholes formula is used to estimate value of assets (A), of which standard deviation gives assets volatility (σ_A) estimates for the next iteration. This procedure is repeated until the value of assets volatility (σ_A) from two consecutive iterations converge.

As further explained by Vassalou & Xing (2004), assets volatility estimates (σ_A) usually converge already in a few iterations when tolerance level of 0,001 is used. When converged value of assets volatility (σ_A) is obtained, it can be used to estimate value of assets (A) through already above presented Black & Scholes formula. All obtained values can eventually be used in the equation, that determines measure of distance to default (DD). The distance to default equation is

$$DD = \frac{\ln\left(\frac{A}{D}\right)+\left[\mu_A-\frac{1}{2}\sigma_A^2\right]T}{\sigma_A\sqrt{T}} \tag{6}$$

where all variables have already been presented except μ_A, which is the expected annual return of the firm's assets (Jessen & Lando, 2015).

As described by Zvika, Ohad & Koresh (2016), the expected assets return (μ_A) has to be estimated separately. In their work, they present variety of approaches how to tackle this problem. One approach is to assume that assets have the same expected return as the risk-free rate (r) while the other is that some risk premia should be added as well. Most common estimate is however using equity return of the preceding year ($r_{E,-1}$). Since historical equity returns might sometimes be negative, risk-free rate is often considered as floor in addition. The expected assets return (μ_A) equation would therefore look like

$$\mu_A = \max(r, r_{E,-1}) \tag{7}$$

where *r* is therefore risk-free rate and $r_{E,-1}$is equity return of preceding year.

With estimating expected asset return (μ_A), distance to default (DD) values as presented in the above equation can finally be calculated. According to Benos & Papanastasopoulos (2007), distance to default (DD) value measures the number of standard deviations that the firm's assets value (A) is away from the default point (D). In order to transform distance to default (DD) to probability of default, the following has to be derived

$$PD = N(-DD) \tag{8}$$

where N() is standard normal distribution function.

As slightly criticized by Jessen & Lando (2015), the estimated default probabilities using Merton's Distance to default model would often be far too small, especially for relatively safe firms. However, since the default probability (PD) is a monotone function of distance to default (DD), this measure can still be well used for ranking firms regarding their default risk.

1.2 Private firm extension

Merton's Distance to default model is therefore very useful in future asset's value estimation, but since it needs stock market data in its prediction, it is applicable to publicly traded firms only. According to Duan, Kim, Kim & Shin (2018), the model can however be extended with application of estimated distance to default values of public firms to private firms. Application can be done through finding similar characteristics and matching both types of firms. Matching process is presented in Chapter 1.2.1 while model application can be found in Chapter 1.2.2.

1.2.1 Matching process

Matching can be done through finding similar characteristics of each private firm with a universe of public firms. These characteristics should also represent reasonable determinants of default, which are usually found in financial statements (Duan, Kim, Kim & Shin, 2018).

But since financial statements can significantly differ among companies from different industries, general comparison of public firms with private firms may indicate on somewhat biased relations. To find more appropriate and relevant relations, industry sampling should be done first, which means that companies with similar economic activities are grouped and analysed together (Sun, Li, Huang & He, 2014).

As explained by Perani & Valeria (2015), there are several different industry classification systems used in industry sampling. One of the most widely used classifications is the Statistical Classification of Economic Activities in the European Community or NACE as commonly referred. NACE is a four-digit classification system, ordered in different hierarchical levels.

The most commonly used level in credit scoring industry sampling is Level 1 classification. The Level 1 classification consists of 21 broad industries, that are based on the type of activities companies are involved in. Each company is obliged to register one industry, where the majority of their income is generated (Perani & Valeria, 2015). Codes of Level 1 industries with description of their economic areas are presented in Table 1 below.

Presented industries can therefore be used in industry sampling procedure of DD estimation. More specifically, with allocation of companies used in DD estimation process into industry groups, up to 21 different groups with DD estimations can be obtained. Since some of the presented industries are often limited to public entities or any other specific institutions, the total number of groups would probably be smaller (Perani & Valeria, 2015).

When companies are allocated into industries, searching of default determinants from financial statements can finally start. According to Chouhan, Chandra & Goswami (2014), several different ratios have been derived from financial statements in the past which have largely been using in default prediction. One of the first and most important attempts to obtain default predictive financial ratios was the Z-Score model, which was developed by Altman in 1968.

Altman used Multiple Discriminant Analysis (MDA) to select the five most significant variables for measuring financial distress of firms. MDA assumes that the covariance matrices of two populations are identical and both need to be described by multivariate normal distribution (Chouhan, Chandra & Goswami, 2014).

Table 1: NACE Level 1 industry codes with their description

Code	Economic Area
A	Agriculture, Forestry and Fishing
B	Mining and Quarrying
C	Manufacturing
D	Electricity, Gas, Steam and Air Conditioning Supply
E	Water Supply; Sewerage, Waste Management and Remediation Activities
F	Construction
G	Wholesale and Retail Trade; Repair of Motor Vehicles and Motorcycles
H	Transportation and Storage
I	Accommodation and Food Service Activities
J	Information and Communication
K	Financial and Insurance Activities
L	Real Estate Activities
M	Professional, Scientific and Technical Activities
N	Administrative and Support Service Activities
O	Public Administration and Defence; Compulsory Social Security
P	Education
Q	Human Health and Social Work Activities
R	Arts, Entertainment and Recreation
S	Other Service Activities
T	Activities of Households as Employers; Undifferentiated Goods and Services Producing Activities of Households for Own Use
U	Activities of Extraterritorial Organisations and Bodies

Source: Own work.

The original Z-Score model has initially been intended for publicly traded companies only since market value was incorporated within the model. In year 1983, Altman completely re-estimated the model and adjusted it to be based solely on accounting data (Altman, Iwanicz-Drozdowska, Laitinen & Suvas, 2014). The discriminant function of adjusted Z-score model, using the five most significant variables is

$$Z = 0.7X_1 + 0.8X_2 + 3.1X_3 + 0.4X_4 + 1.0X_5 \qquad (9)$$

where

$$X_1 = \frac{Working\ Capital}{Total\ Assets} \tag{10}$$

$$X_2 = \frac{Retained\ Earnings}{Total\ Assets} \tag{11}$$

$$X_3 = \frac{Earnings\ before\ Interest\ and\ Taxes}{Total\ Assets} \tag{12}$$

$$X_4 = \frac{Book\ Value\ of\ Equity}{Book\ Value\ of\ Total\ Liabilities} \tag{13}$$

$$X_5 = \frac{Total\ Sales}{Total\ Assets} \tag{14}$$

The outcome of the function is a statistical Z-score that classifies companies based on their solvency. The higher the score, the lower the risk of bankruptcy. Each of the above presented variables or financial ratios represents different field of measurement, which are liquidity, leverage, profitability, solvency and activity (Altman, Iwanicz-Drozdowska, Laitinen & Suvas, 2014). The description of each of them in accordance with Altman, Iwanicz-Drozdowska, Laitinen & Suvas (2014) is presented below.

The ratio Working Capital / Total Assets (X_1) is a measure of the liquidity of a firm relative to its entire assets. Working capital represents the difference between current assets and current liabilities, which, in general, is decreasing if the company records losses.

The ratio Retained Earnings / Total Assets (X_2) represents earned surplus of a firm over its entire life. It actually considers the firm's possibility to finance assets growth with own sources or the use of leverage to finance it.

The ratio Earnings before Interest and Taxes / Total Assets (X_3) is a measure of the true profitability of the assets of a firm. Earning before Interest and Taxes (EBIT) represents the possibility to generate profits solely from its operations and are not affected by tax or structure of financing.

The ratio Book Value of Equity / Book Value of Total Liabilities (X_4) shows the solvency ratio of the firm. It represents the degree to which the assets of the firm can decline before the company practically declares bankruptcy.

The ratio Sales / Total Assets (X_5) represents the activity measurement of a company. It shows the efficiency of management in using assets to generate sales in comparison with competition.

1.2.2 Application

The five selected and described financial ratios in the chapter above should therefore represent determinants of default that can be found at any private of public firm. In order to estimate distance to default (DD) values of private firms, indirect projection from public firms' DD values within each industry should be done. As further explained by Duan, Kim, Kim & Shin (2018), the first step of projection should be done with a liner regression of estimated public firms' DD values with their respective financial ratios within each industry. The linear regression of public firm's DD values with their financial ratios for each industry is as follows

$$DD_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \beta_3 X_{3,i} + \beta_4 X_{4,i} + \beta_5 X_{5,i} + \varepsilon_i, \ i = 1, \dots, n \tag{15}$$

where DD_i is distance to default of public firm, variables $X_{1-5,i}$ are financial ratios as already described in the previous chapter, ε_i is the disturbance term and *n* represents the number of public firms.

As further explained by Duan, Kim, Kim & Shin (2018), at least a 3-month lag should be assigned to DD values after the date of financial ratios to allow time for information spread. For example, in a regression where financial ratios are based on date *d*, DD estimations of *d+3 months* should be used at the earliest. Such an approach is presented in regression below

$$DD_{d+m} = \beta_{0,d+m} + \beta_{1,d+m} X_{1d} + \beta_{2,d+m} X_{2d} + \beta_{3,d+m} X_{3d} + \beta_{4,d+m} X_{4d} + \beta_{5,d+m} X_{5d} + \varepsilon_{d+m}, \ m = 3, \dots, 14 \tag{16}$$

where *d* is a certain date and *m* represents the number of months that are limited to one year (m=12). Note that indicator *i* has been neglected for simplicity.

Date *d* in the regression usually represents the last date of the year, which gives values of *m* ranging from March in the next year until February in another year. With the same values of financial ratios from date *d*, twelve separate regressions should therefore be conducted, which results in twelve different sets of β_{0-5} coefficients within *J* industries (Duan, Kim, Kim & Shin, 2018). To span the time frame to *Y* years, the number of separately conducted regressions as well as the number of obtained coefficients is twelve multiplied by the number of industries *J* and the number of years *Y*.

When coefficients of the abovementioned regressions are obtained, mapping with private firms can be done next. As seen in Duan, Kim, Kim & Shin (2018), mapping could simply be done with multiplication of obtained coefficients with their respect private firm's financial ratios within each industry. Multiplied coefficients and financial ratios are then summed up, which eventually gives public-firm DD equivalent. The mapping procedure with multiplication and addition process is presented below

$$\widetilde{DD} = \widehat{\beta_0} + \widehat{\beta_1} X_1 + \widehat{\beta_2} X_2 + \widehat{\beta_3} X_3 + \widehat{\beta_4} X_4 + \widehat{\beta_5} X_5 \tag{17}$$

where $\widetilde{DD}$ is public-firm DD equivalent, variables $X_{1\text{-}5}$ are financial ratios of private firms as already presented and $\widehat{\beta_{0-5}}$ are estimated coefficients obtained from above performed regressions.

According to Duan, Kim, Kim & Shin (2018), the same 3-month lag should be incorporated into the calculation of public-firm equivalent DD values as has been done in the coefficient estimation process. Therefore, financial ratios of private firms based on date *d* could only be multiplied with coefficients estimated in $d+3$ *months* until $d+14$ *months* to obtain their respective public-firm equivalent DD values. The equation showing such an approach is

$$\widetilde{DD}_{d+m} = \hat{\beta}_{0,d+m} + \hat{\beta}_{1,d+m}X_{1d} + \hat{\beta}_{2,d+m}X_{2d} + \hat{\beta}_{3,d+m}X_{3d} + \hat{\beta}_{4,d+m}X_{4d} + \hat{\beta}_{5,d+m}X_{5d}, \quad m = 3, \dots, 14 \tag{18}$$

where *d* is a certain date and *m* represents the number of months that are limited to one year (m=12). Note that indicator *i* has been neglected for simplicity.

As already explained, date *d* within the equation usually represents the last date of the year, which gives values of *m* ranging from March next year until February in another year. With the same values of financial ratios, twelve different public-firm equivalent DD values can be calculated (Duan, Kim, Kim & Shin, 2018).

Although several attempts in estimating of private firms' distance to default have already been made, Duan, Kim, Kim & Shin (2018) point out at least four important merits of their Private firm extension model.

The first merit is that, with a two-step private firm's DD estimation approach, over-fitting problem can be avoided. The two-step approaches penalize overly complex models while alternative one-step approaches result in the excessive degree of freedom as proven by authors. Comparable one-step approaches lead to quite poor predictive performances in general.

The second merit is that presented private firm DD estimation approach provides universal tool for analysing all types of firms in a comparable manner. This is achieved with obtaining coefficients for each industry separately. Furthermore, diversified industry specific characteristics can also be found in addition and can be easily incorporated into the model to optimize its predictive performance.

The third merit with private firm DD estimation is frequent updates of information. Financial statements of private firms usually update annually, which cause the so-called age-of-information issue. This means that financial statements updated 10 months ago cannot provide the same quality of information as the ones updated one month ago, even if their values remain the same. With private firms' DD estimation, information can be updated monthly or even more frequent.

The last but not least merit is utilizing timely stock market data in the monitoring process of the private firm. The estimated DD value of a private firm represents direct projection from the universe of public firms, which implies up to date market information to private firms at any point in time.

2 DATA COLLECTION AND ESTIMATION PROCESS

To confirm or deny hypotheses stated in the Introduction chapter, the model based on a sample of German public firms will be built and used to predict default probabilities for set of Slovenian private firms afterwards. The data used in the analysis will be gathered from several domestic and global databases. More specifically, the data of public firms will be mostly extracted from some global financial data vendors, while private firms' financial data will be fully obtained from domestic data providers.

Gathered data of German public firms will then be used to estimate several parameters as described in Merton's Distance to default model. These parameters will result in distance to default values that would eventually be used in default prediction of any of selected Slovenian company. Translation of public firms' distance to default values to private firms will be done through the selection of some financial variables and through estimation of appropriate coefficients. Coefficients would be obtained on a monthly basis, which means that distance to default should change from month to month and would eventually provide the record of monthly default probabilities.

The data collection process with presentation and dimensions of extracted datasets will be more in detail presented in Chapter 2.1. Chapter 2.2. will then guide through the process of parameter estimation while model construction with translation of public firm's default probabilities to private ones will be described in Chapter 2.3.

2.1 Data collection

The data used in the analysis will be gathered from several databases. Sample of German public firms, together with some accounting variables will be obtained from the global financial vendor Bloomberg Terminal. Stock prices of selected German firms will then be gathered from another financial data vendor Yahoo Finance while 1-year German bond yields would be extracted from MarketWatch database.

Considering Slovenian private firms, the sample will be selected through domestic company register agency AJPES which will also serve as the source for gathering accounting data of selected private firms. To test the model eventually, benchmark credit scores as estimated by some selected Slovenian bank will be obtained for that manner. Gathering process and description of obtained datasets will be presented in detail in the following four subchapters.

2.1.1 Bloomberg Terminal

Bloomberg Terminal is one of the largest and most famous platforms with financial market data (Kelly, 2020). Among various possible actions and analyses that can be performed, Equity screening (EQS) command has been used. Since calculating of distance to default is focused on German firms only, screening with correct country of domicile needs to be done first. Stock's country of domicile is referred to a legal home of an underlying corporation or the country where the corporation has been incorporated (Corporate Finance Institute, 2020), Germany has therefore been chosen in that case.

In order to obtain more relevant data, three additional filters have been applied in stock screening procedures. The first additional filter is that only primary stocks are used. Primary stocks are stocks listed on a stock exchange where they made their Initial Public Offering (IPO) or sale of a stock to public for the first time (Chen, 2022a). The second filter is that the IPO has had to be done in Germany, while the third is that only actively traded stocks are used.

Applying all stated filters with the state as at 31.12.2020, the outcome of screening procedure is 989 German stocks that are active on German market. Since some of screened stocks represent also exchange-traded funds (ETFs), these will also be excluded from the analysis as they do not report financial statements, which will be crucial in later model development. With excluding ETFs from screened stocks, 841 companies' stocks are left for further analysis. Selected stocks should next be used to extract several variables, which will later be needed in model development process. Needed variables will be presented in three sets regarding their dimension or usage in the model construction.

The first set of needed variables are the variables Ticker Symbol, Stock Exchange Code and NACE Sector Code of each individual stock. Ticker Symbol is an abbreviation used to uniquely identify publicly traded stocks on a particular stock exchange, whereas Stock Exchange Code is a unique abbreviation of each stock exchange (Hayes, 2022). NACE Sector Code represents character that uniquely determines each sector based on NACE industry classification (Connects, 2022). Ticker Symbol and Stock Exchange Code will later be used as identifiers to extract financial data from Yahoo Finance platform while NACE Sector Code will be used in industry sampling procedure.

The second set of needed variables are the Number of Shares Outstanding (NSO), the value of Short-Term Debt (STD) and the value of Long-Term Debt (LTD) of each company. Knowing that the analysis of this thesis will be done based on the period 2017-2020, all three variables have been extracted as reported in financial statements at the beginning of each year in this period. The Number of Shares Outstanding represents all issued shares of the firm less own shares (Chen, 2022b). The value of Short-Term Debt represents the entire interest-bearing debt reported in company's balance sheet, which matures within the next reporting year (Ganti, 2020a). Long-Term Debt holds the same definition as the short-term one, except that maturity is scheduled beyond upcoming reporting year (Tuovila, 2021).

The third set of extracted variables are Working Capital (WC), Total Assets (TA), Retained Earnings (RE), Earnings Before Interest and Tax (EBIT), Book Value of Equity (BVE), Total Liabilities (TL) and Total Sales (TS) of each company. All seven variables have also been extracted as reported in financial statements at the beginning of each year in the period

2017-2020. Working Capital (WC) represents all current assets reported in company's balance sheet less current liabilities (Fernando, 2022a). Total Assets (TA) represents the entire balance sheet value, while Book Value of Equity (BVE) contains all equity records including net profit of the year (Seth, 2021). Total Liabilities (TL) are constructed of short-term and long-term ones (Liberto, 2020). Retained Earnings (RE) represents all historical profits retained within the company (Fernando, 2022b). Total Sales (TS) refers to net income from sales, while Earnings Before Interest and Tax (EBIT) are calculated as operating profit (Murphy, 2022).

2.1.2 Yahoo Finance & MarketWatch

Yahoo Finance is a free web side platform that provides financial data of the majority of publicly traded stocks around the world (Somaiya, 2015). Among all possible financial information about a particular stock, adjusted closing prices will be used to calculate returns. Adjusted closing price is adjusted for dividends and stock splits (Ganti, 2020b). Similar to variables extracted from Bloomberg Terminal, adjusted closing prices have also been extracted for the period 2017-2020. Since daily stock prices are needed in the model constriction process as presented in the Literature review, four-year long dataset now becomes roughly thousand days long for each company assuming that each year consists of 250 working days.

Needed stock prices of a particular stock will be extracted by its Ticker and Stock Exchange Code, which has previously been obtained by equity screening. Note that tickers and stock exchange codes are slightly different at Yahoo Finance comparing to Bloomberg Terminal, which therefore have to be adjusted.

MarketWatch is similar to Yahoo Finance also a free web side platform, which provides a variety of financial information (Kramer, 2017). To apply Merton's model in this thesis, annualized yield of 1-year German bond is need to be extracted. The data has been retrieved for period 2017-2020 on a monthly basis, meaning that yield at every end month is considered.

2.1.3 AJPES

When all relevant data of public firms are obtained as presented in the above two chapters, some data of domestic private firms should be obtained next in order that the model can be developed. Data about private firms will be obtained from Slovenian company register agency AJPES which collects, process and publish all kinds of domestic companies' data (AJPES, 2022). Since massive and appropriately structured data export procedure is not readily available on AJPES website, IT tool developed by the selected Slovenian bank will be used for that manner.

Among all available data that can be obtained from AJPES website, only companies' ID numbers, registered industries and some financial statement values will be needed in model construction of this thesis. Before any extraction of data can be performed, some sampling procedure should be done first.

Since larger companies tend to report more reliable financial statements and are more properly industry classified due to general revision requirements, threshold of 10m € of sales will be applied in the sampling process. According to Merritt (2019), the threshold of 10m € is also used in enterprise size classification, which means that only large and medium sized domestic companies will be considered in this thesis. The threshold will be applied based on sales in year 2020, which is also the last year of currently available data.

Technically, sales used in the sampling process are considered as net income from sales, which is classified as AOP 110 according to national accounting classification system. With applying the abovementioned threshold of 10m € of sales in year 2020, sampling procedure results in 946 companies fulfilling that condition.

When companies are selected in accordance with stated threshold, appropriate variables regarding companies' ID numbers, registered industries and some financial statement values have to be obtained as already mentioned before. Needed variables will be put into two different sets considering their dimensions. Both sets are more specifically presented below.

The first set of needed variables are Company Registration Number (CRN) and NACE Sector Code of each individual company. CRN is a unique combination of numbers and additional letters in some cases (Korchak, 2019). It is used to uniquely identify specific company within the country and to verify that it is truly registered in Company Register. NACE Sector Code represents character, that uniquely determines each sector based on NACE industry classification (Connects, 2022). CRN will later be used as companies' identifier through time series of financial data while the NACE Sector Code will be used in the industry matching procedure. The dimension of both extracted variables is therefore the number of companies that fulfil aforementioned sampling procedure, which is 946 companies.

The second set of needed variables represents values from financial statements, which should be comparable to ones obtained at public firms as already described in the Literature review. The extracted financial variables should therefore be Working Capital (WC), Total Assets (TA), Retained Earnings (RE), Earnings Before Interest and Tax (EBIT), Book Value of Equity (BVE), Total Liabilities (TL) and Total Sales (TS). All seven variables should be extracted as reported in financial statements at the beginning of each year during the period 2017-2020. The dimension of dataset will therefore be all companies that fulfil the threshold sampling procedure (946) for the period of four years.

However, since direct extraction of some of the abovementioned financial variables is not possible, derivation from other variables should be performed. The variable construction process will therefore use different values form financial statements, that will be adjusted in accordance with the definition. The process with reference to national accounting classification within brackets (AOP) is described below.

Financial variable Working Capital was constructed as all current assets reported in company's balance sheet (AOP 032) less current liabilities (AOP 085). Total Assets represent the entire balance sheet value (AOP 001), while Book Value of Equity contains all equity records including net profit of the year (AOP 056). Total Liabilities are constructed of Long-Term (AOP 075) and Short-Term (AOP 085) ones. Retained Earnings represent all historical profits retained within the company (AOP 068) plus all retained losses (AOP 069). Total Sales refers to net income from sales (AOP 110) while Earnings Before Interest and Tax are calculated as operating profit (AOP 151) plus operating loss (AOP 152).

2.1.4 Selected Slovenian bank

When the model presented in this thesis is fully developed, testing procedure will have to be done to check for statistical significance and accuracy of the model. To do that, some benchmark credit scores need to be obtained and compared to newly estimated DD values of private firms.

Although there exists lots of different credit scoring models using wide range of input data, credit scores estimated by selected Slovenian bank will be used as a benchmark within this thesis. The reason to choose internally developed credit scoring model of the selected Slovenian bank as a benchmark is its availability and appropriateness to assess financial distress of domestic medium and large enterprises.

The selected Slovenian bank's credit scoring model uses several different parameters in its prediction, but only credit scores based on annual financial statements will be used as a benchmark in this thesis. In order to obtain relevant dataset of credit scores, the same sampling procedure of companies with more that 10m € of sales in year 2020 needs to be performed first.

When companies are selected in accordance with stated threshold, appropriate variables regarding companies' ID numbers and credit scores have to be obtained. To match credit scores with estimated DD values in model testing procedure later, Company Register Number (CRN) should again be used as companies' identifiers. The dimension of extracted values of CRN is therefore the number of companies that fulfil aforementioned sampling procedure, which is 946 companies.

Values of credit scores or probability of default (PD) to be more specific updates on an annual basis, which means that their dimension will be somewhat larger than the one of CRN. The PD values will represent benchmark for estimated monthly DD values, some modification of data dimensions will however need to be done later during the testing procedure. The PD values should be extracted as at end of each year in the period 2017-2020. The dimension of the dataset will therefore be all companies that fulfil threshold sampling procedure (946) for the period of four years.

2.2 Estimation of public firms' parameters

Parameter estimation process can begin when the majority of data about German public firms as presented in the previous chapter are gathered. Estimation process starts with estimating equity volatility, where daily stock prices represents the initial extracted variable. Debt value will be constructed by short-term and long-term debt values while 1-year German bond yields would be used as a risk-free rate. The last estimated parameter using gathered data is market value of equity, where the data about shares outstanding and stock prices will again be used.

When all parameters using gathered data are estimated, asset values and their volatilities can then be obtained. Using all obtained estimates, distance to default values can eventually be modelled. The more detailed estimation process, together with descriptive and visual dataset presentations can be found in the following subchapters.

2.2.1 Equity volatility

As presented in the Literature review, equity volatility is the first variable needed in Merton's Distance to default prediction. Since log returns are usually used to estimate equity volatility, extracted daily stock prices will be used as initial observable values.

Adjusted daily closing stock prices have been extracted for all selected companies in period 2017-2020 as presented in the Data collection chapter. Since the number of working days in each year is about 250, the total of 1,000 adjusted stock prices should therefore be obtained for each company. Returns of obtained adjusted closing prices will then be calculated as natural logs which assumes that returns are compounded continuously. The calculation formula for each company is

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{19}$$

where r_t is daily log return and P_t is its respective stock price in time *t*.

The new dataset with adjusted log returns should then be similar to adjusted stock prices, with only difference that for 1,000 daily prices, 999 returns are obtained. Considering also that the number of selected companies is 841, the number of obtained log returns should be 840,159 in total. This amount therefore represents the number of selected companies (841), multiplied by the number of obtained daily returns for each company (999).

With obtaining log returns, it has however been noticed that many observations have had stable stock prices for longer period of time. To avoid zero equity volatilities and consequently infinite distance to default values later, exclusion of observations with zero average monthly returns from the analysis should be done. Descriptive statistics of adjusted database of daily log returns are presented in Table 2 while their distribution can be seen in Figure 1. In addition, time series of average daily log returns are also presented as time-varying statistics in Figure 2.

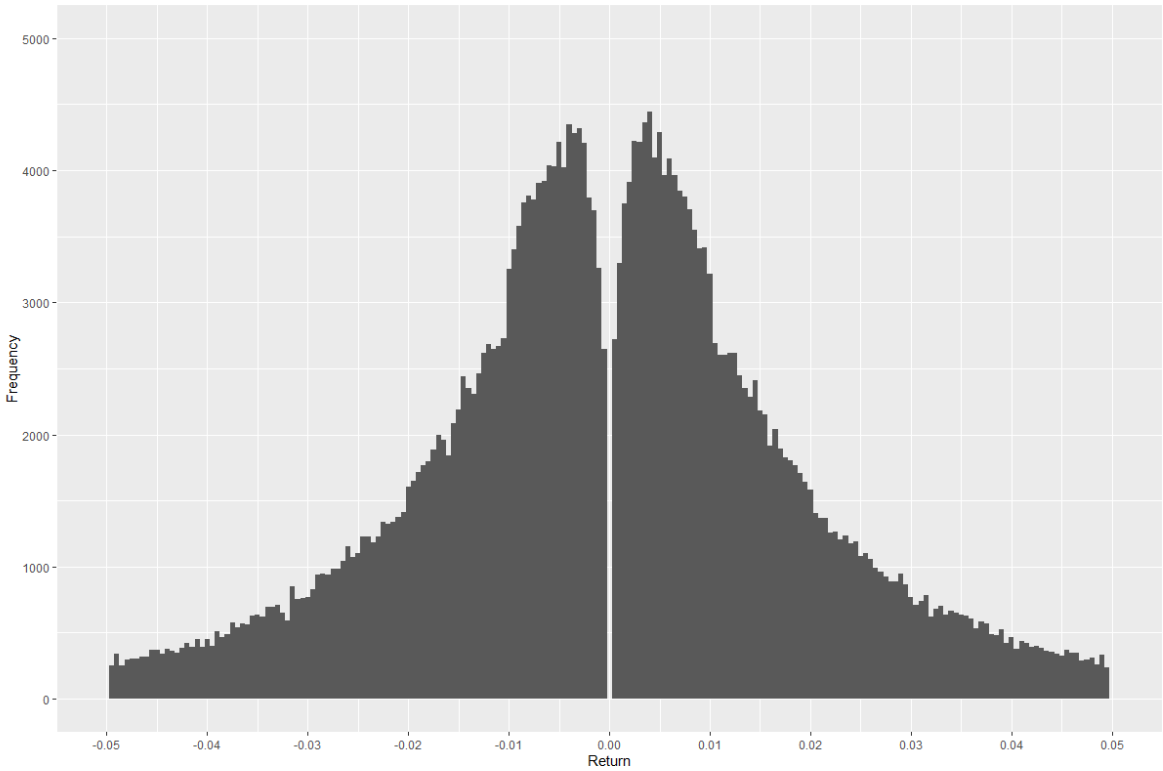
Table 2: Descriptive statistics of daily log returns

N	*N/A*	*N/A %*	*Mean*	*Std*	*Min*	*Q1*	*Median*	*Q3*	*Max*
435,084	*405,075*	*48%*	*0.0000*	*0.0347*	*-2.3639*	*-0.0087*	*0.0000*	*0.0085*	*5.5444*

Source: Own work.

As seen in Table 2, the number of obtained daily return observations is 453,084. The total number of observations should however be much higher as mentioned before (840,159), but around 48% of daily return values have been omitted. Omitted values therefore represented zero returns through longer period of time (monthly average) or missing values in stock price extraction process.

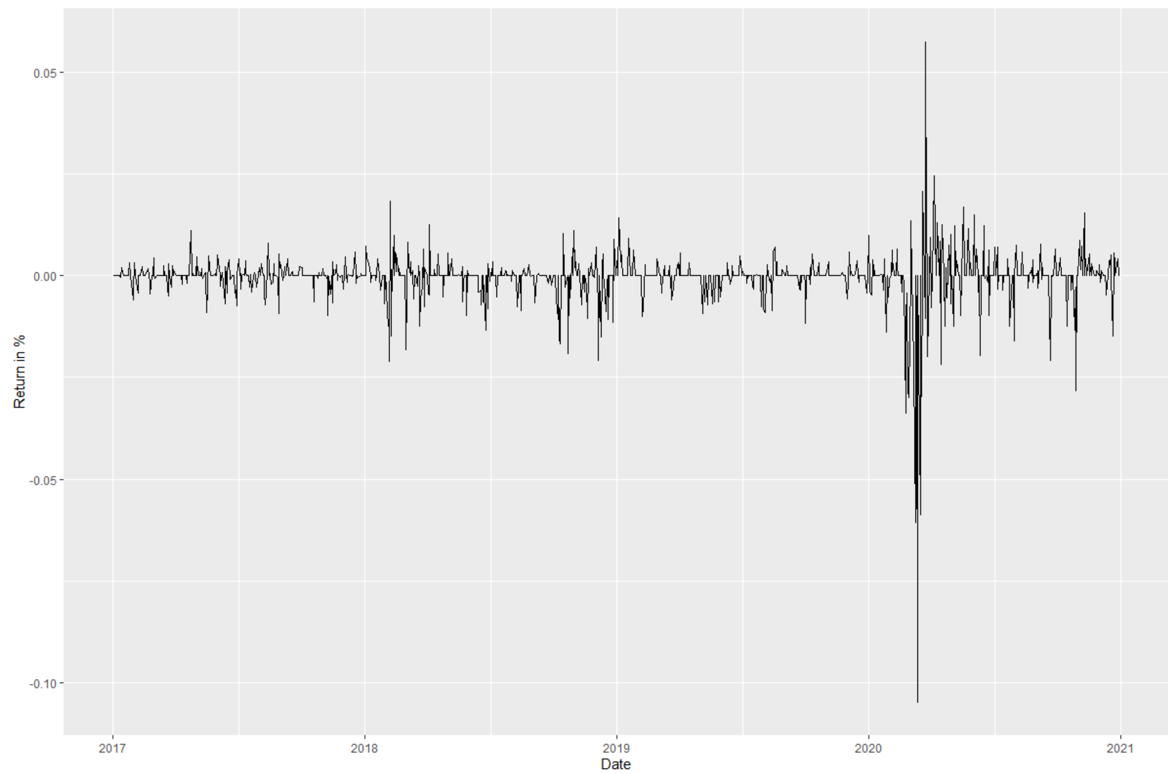
Figure 1: Distribution of daily log returns



Source: Own work.

As further seen in Table 2, descriptive statistics indicate on well-shaped normal distribution, which is also noticed in the graphical presentation (Figure 1) to certain extent. Median and mean values are basically the same at zero while the standard deviation is at 3.4%. Half of the observations centred around median falls between -0.9% and 0.9%, both tails are therefore quite similar. Note only that quite large data gap just around the mean value as seen in Figure 1 is a consequence of omitted average monthly zero returns as already mentioned before.

Figure 2: Time-series of average daily log returns



Source: Own work.

Considering the time series of average daily stock returns, they can be seen that values have been quite stable at around zero in the majority of analysed timeframe as presented in Figure 2. Some occasional jumps to up to 2.5% or -2.5% occurred until year 2020, when jumps of stock returns to both directions extended significantly. More specifically, stocks decreased in value for even more that 10% in a single day in the beginning of year 2020 on average, which was then followed with fast partial recovery of more than 5% of average daily increase few days after. Average daily jumps somewhat stabilized afterwards, but still remain on higher level compared to pre-2020 average daily jumps.

When daily equity returns are obtained, the estimation procedure of equity volatility (σ_E) can start. Equity volatility will be estimated for the period 2017-2020 at the end of each month, using daily log returns of each respective month. Volatility will be expressed with standard deviations, equation used for that purpose is

$$\sigma_E = \sqrt{\frac{T}{(T-1)} \sum_{t=1}^{T} (r_t - r)^2} \tag{20}$$

where σ_E is monthly standard deviation, T is the number of working days within each month, r_t is daily log return and r is monthly mean of daily returns, calculated as

$$r = \frac{1}{T} \sum_{t=1}^{T} r_t \tag{21}$$

where all variables have already been defined above. The dimensions of estimated volatility dataset now shorten from 999 of daily log returns to 48 monthly standard deviations.

Considering that the number of companies used in analysis is 841, the number of estimated equity volatilities should thus be 40,368. This amount therefore represents the number of selected companies (841), multiplied by the number of estimated monthly volatilities for each company (48). Since stocks with zero average returns have been excluded from the analysis as already mentioned before, a somewhat lower number of observations is expected to be obtained, however. Moreover, any additional zero equity volatility values would also be excluded from the analysis. Descriptive statistics of estimated equity volatilities are presented in Table 3 while their distribution can be seen in Figure 3. In addition, time series of average equity volatility is also presented as time-varying statistics in Figure 4.

Table 3: Descriptive statistics of monthly equity volatilities

N	*N/A*	*N/A %*	*Mean*	*Std*	*Min*	*Q1*	*Median*	*Q3*	*Max*
19,767	*20,601*	*51%*	*0.0245*	*0.0259*	*0.0008*	*0.0131*	*0.0193*	*0.0284*	*11.793*

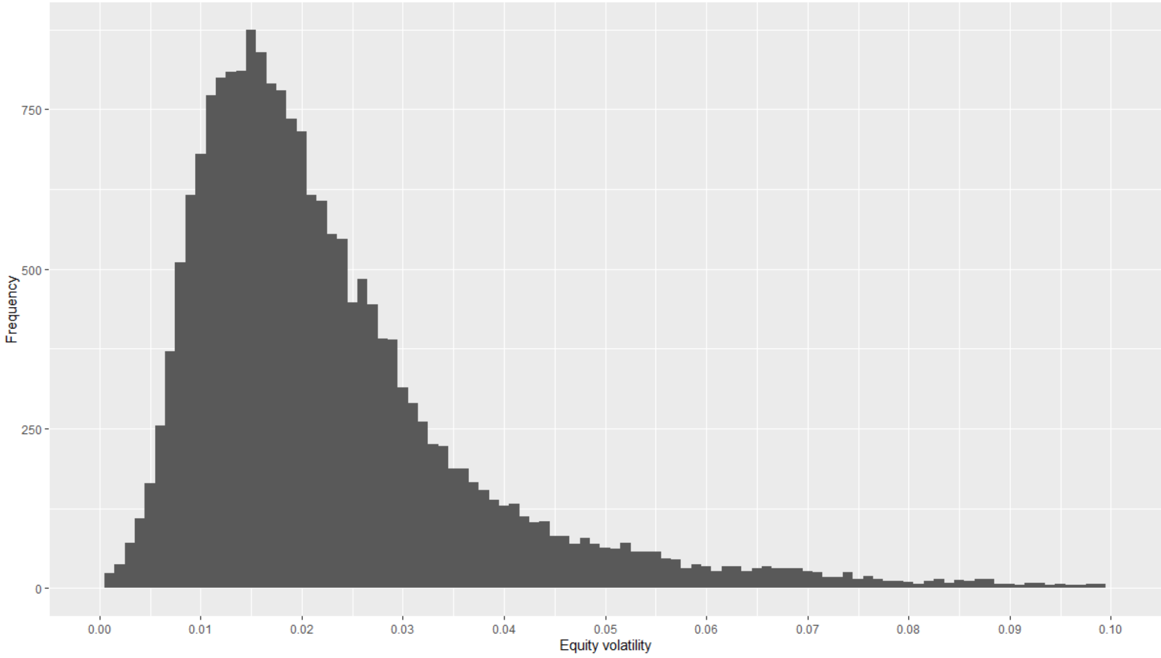
Source: Own work.

As seen in Table 3, the number of estimated monthly volatility observations is 19,767. The total number of observations should however be much higher as presented before (40,368), but around 51% of equity volatility estimates have been omitted. Omitted values are majorly related to exclusion of daily returns with average monthly zero returns as already mentioned before, some additional exclusions also occurred due to zero equity volatility values estimated afterwards.

Visual distribution (Figure 3) of equity volatilities indicates on normal distribution, but heavily skewed to the right. Mean value of 2.5% average volatility is therefore somewhat higher than the median of 1.9% while standard deviation is relatively high at 2.6%. The range of the middle half of observations fits between 1.3% and 2.8% of equity volatility.

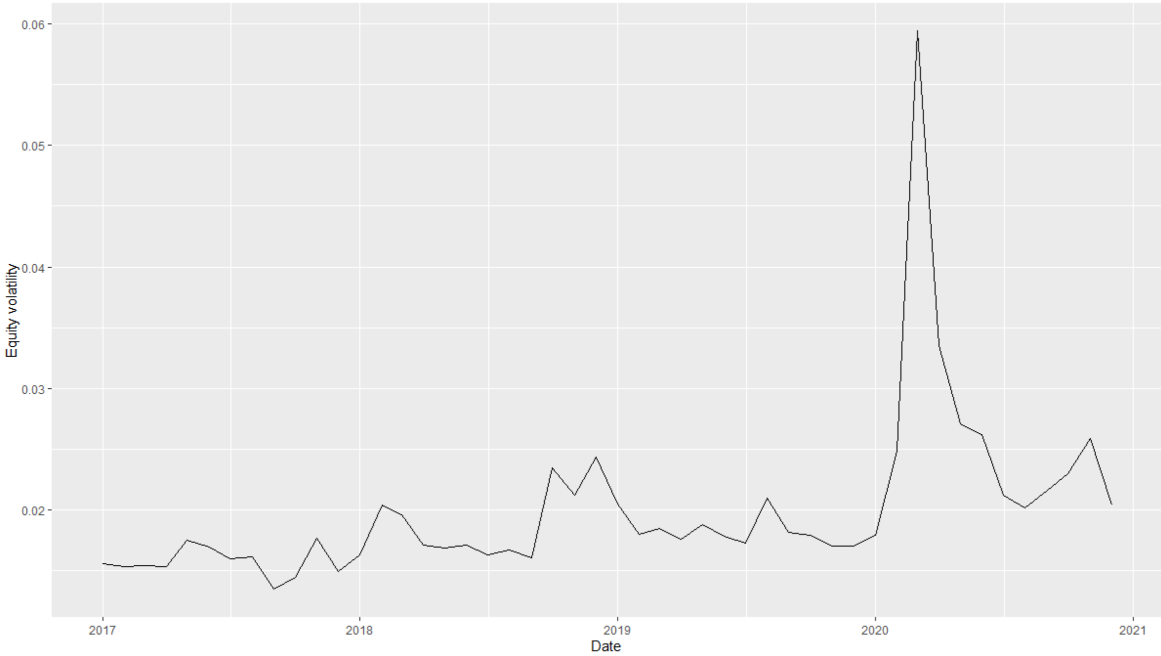
As seen in time series of presented data (Figure 4), average equity volatilities had been quite stable between 1% and 2% until the beginning of year 2020, when increase to high 6% was recorded on average. Average volatility, however, significantly lowered until end of year 2020, but remained at somewhat higher level of around 2.5% compared to pre-year 2020 period. Movements of volatilities through time will be very important in later model testing procedures.

Figure 3: Distribution of monthly equity volatilities



Source: Own work.

Figure 4: Time-series of average monthly equity volatilities



Source: Own work.

2.2.2 Debt value

After obtaining equity volatilities, face value of debt (D) is next crucial variable that should be retrieved. As described in the Literature review, short-term debt with half of long-term debt would be the most appropriate variable selection for determining face value of debt. Equation that determines face value of debt of each company is therefore

$$D_t = STD_t + \frac{1}{2} LTD_t \tag{22}$$

where STD_t is current interest-bearing debt and LTD_t is non-current interest-bearing debt in time *t*.

To calculate monthly distance to default values, face value of debt on a monthly basis would need to be obtained for each company. Since debt values update on an annual level only, constant face value of debt for each company would be assumed throughout the year. More specifically, debt values from the beginning of the year will be used as a constant throughout the same year.

Considering that the number of selected companies is 841, the number of calculated monthly debt value observations should be 40,368. This amount therefore represents the number of selected companies (841), multiplied by the number of months of the analysis (48). However, in order to avoid infinite distance to default values later in the model, debt value observations equal to zero would be omitted from the database. Descriptive statistics of calculated debt values (D) are presented in Table 4 while their distribution can be seen in Figure 5. In addition, time series of average debt values is also presented as time-varying statistics in Figure 6.

Table 4: Descriptive statistics of monthly debt values (in million €)

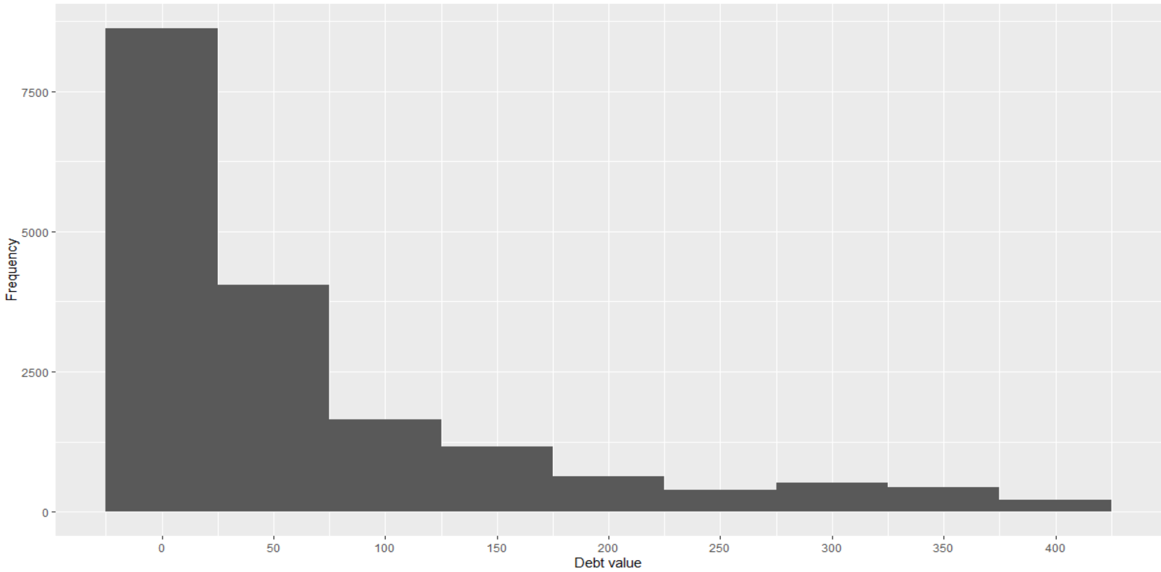
N	*N/A*	*N/A %*	*Mean*	*Std*	*Min*	*Q1*	*Median*	*Q3*	*Max*
24,036	*16,332*	*40%*	1,231	6,272	*1*	*11*	*51*	*352*	*90,721*

Source: Own work.

As seen in Table 4, the number of calculated debt value observations is 24,036. The total number of observations should however be much higher as mentioned before (40,368), but around 40% of debt values have been omitted. Omitted values majorly represented zero debt or missing data in minor extent.

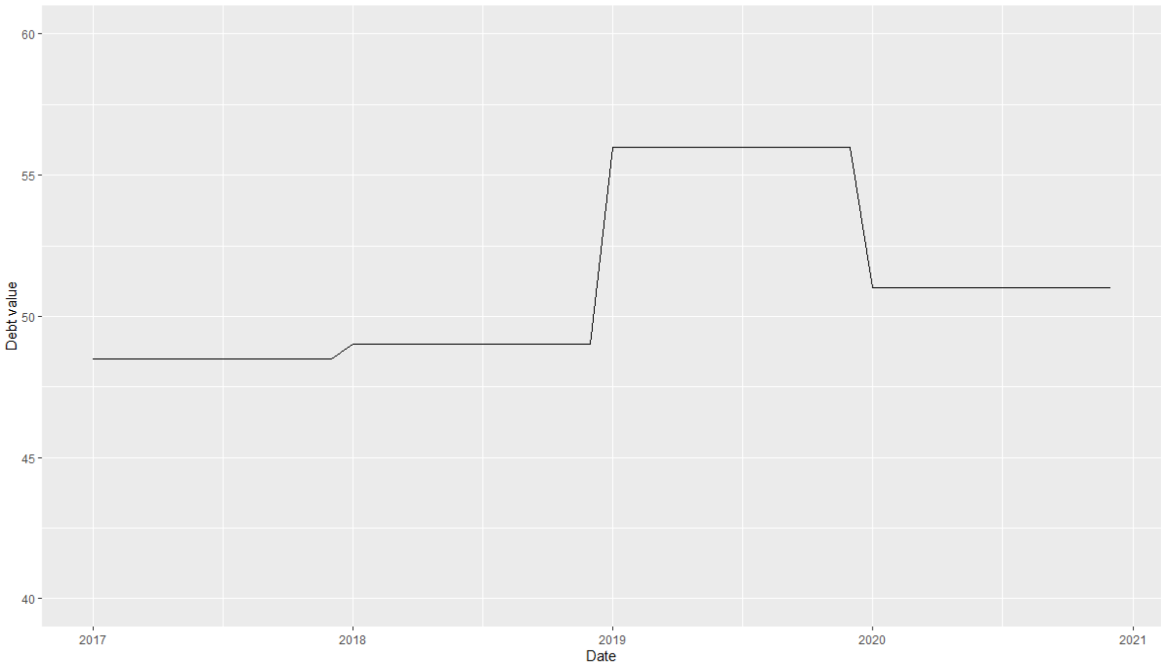
As further presented in Table 4, the face value of debt from three quarters of the observed companies is below 350 million €, while half of them reports debt value of less than 50 million € as indicated with the median. Mean value and standard deviation have no relevance in this case since some outliers with very large nominal debt values are present in the database. Distribution Figure 5 therefore presents only roughly the first three quarters of observations, while a very long tail to the right was excluded from the graph.

Figure 5: Distribution of monthly debt values (in million €)



Source: Own work.

Figure 6: Time-series of average monthly debt values (in million €)



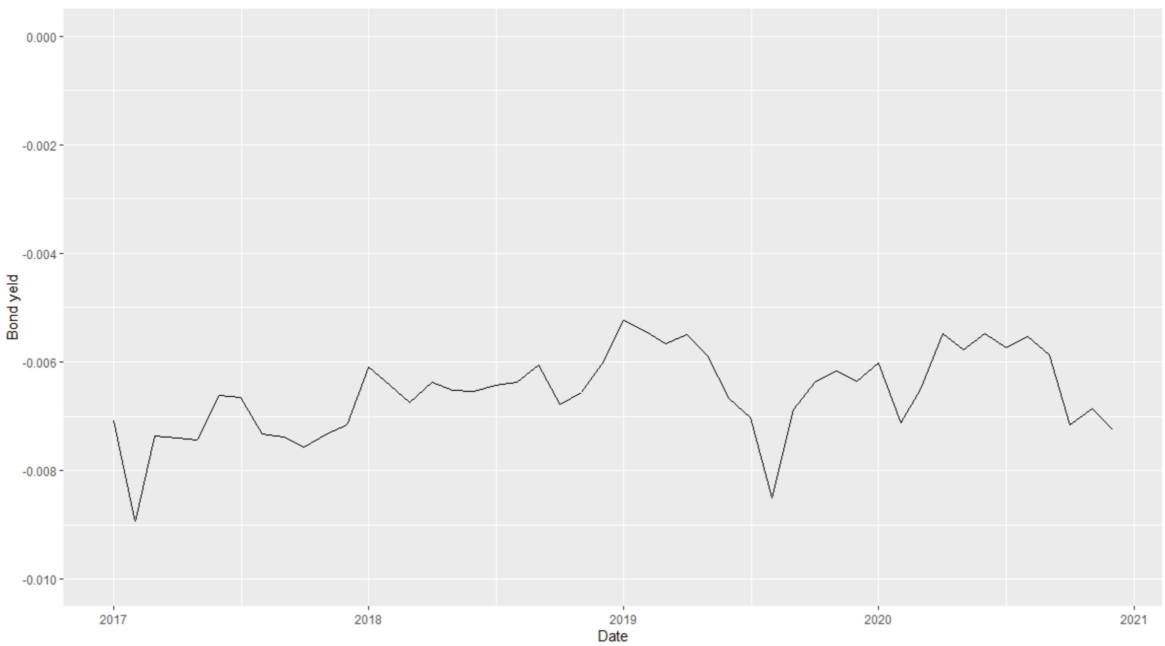
Source: Own work.

Time series of average debt values (Figure 6) indicate a slightly upward trend throughout the years. The only exception is year 2019, when the average debt value increased from around 48 million € to 56 million €, but this growth roughly halved later in year 2020.

In order to be able to discount face value of debt within the model, risk-free rate (r) would have to be determined. As mentioned in the Literature review, annualized rate of 1-year Treasury bond is usually considered for that purpose. Since distance to default values will be calculated for German firms only, yields of 1-year German bond are the most reasonable to consider.

As presented in the Data collection chapter, yields of 1-year German bond have been retrieved for period 2017-2020 on a monthly basis, which represents 48 months of data. Time-series of bond yields are presented in Figure 7, while their descriptive statistics and distribution can be seen in Table 5 and Figure 8, respectively.

Figure 7: Time-series of monthly 1-year German bond yields



Source: Own work.

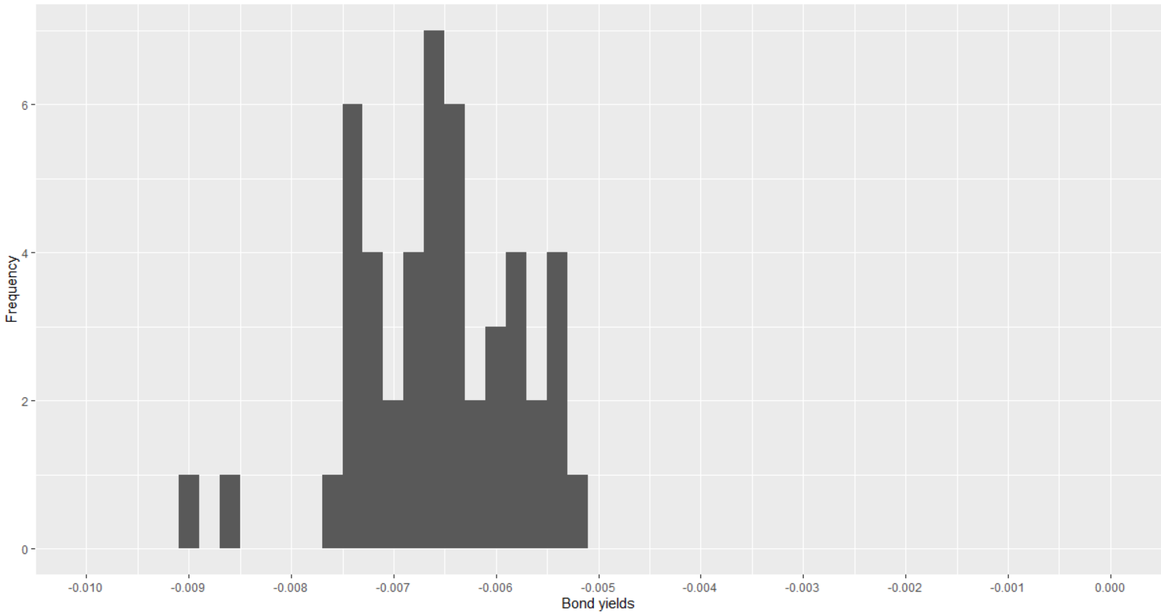
Table 5: Descriptive statistics of monthly 1-year German bond yields

N	*N/A*	*N/A %*	*Mean*	*Std*	*Min*	*Q1*	*Median*	*Q3*	*Max*
48	*0*	*0%*	*-0.0066*	*0.0008*	*-0.0090*	*-0.0071*	*-0.0065*	*-0.0060*	*-0.0052*

Source: Own work.

As seen in Figure 7, monthly yields of 1-year German bond have been constantly negative in the observed period. Descriptive statistics (Table 5) show a quite normal distribution of yields with mean and median roughly the same at -0.7%. Standard deviation has been even below 0.1% in the observed period. Figure 8 shows that also graphically.

Figure 8: Distribution of monthly 1-year German bond yields



Source: Own work.

2.2.3 Market value of equity

The last market variable needed in DD calculation is market value of equity. Market value of equity is according to the Literature review considered as a product of the number of shares outstanding with the stock price at each point in time. The equation representing market value of equity calculation for each company is

$$E_t = S_t P_t \tag{23}$$

where S_t represents the number of shares outstanding and P_t is the stock price in time *t*.

To calculate monthly distance to default values, market cap on a monthly basis would need to be obtained for each company. Monthly stock prices are easily observable on the market while the number of shares outstanding is often available only on annual financial reports. To estimate monthly values, the number of shares outstanding from the beginning of the year will be used as a constant throughout the same year as was the case with debt value. Regarding prices of stocks, end-of-month prices will be used in aforementioned product.

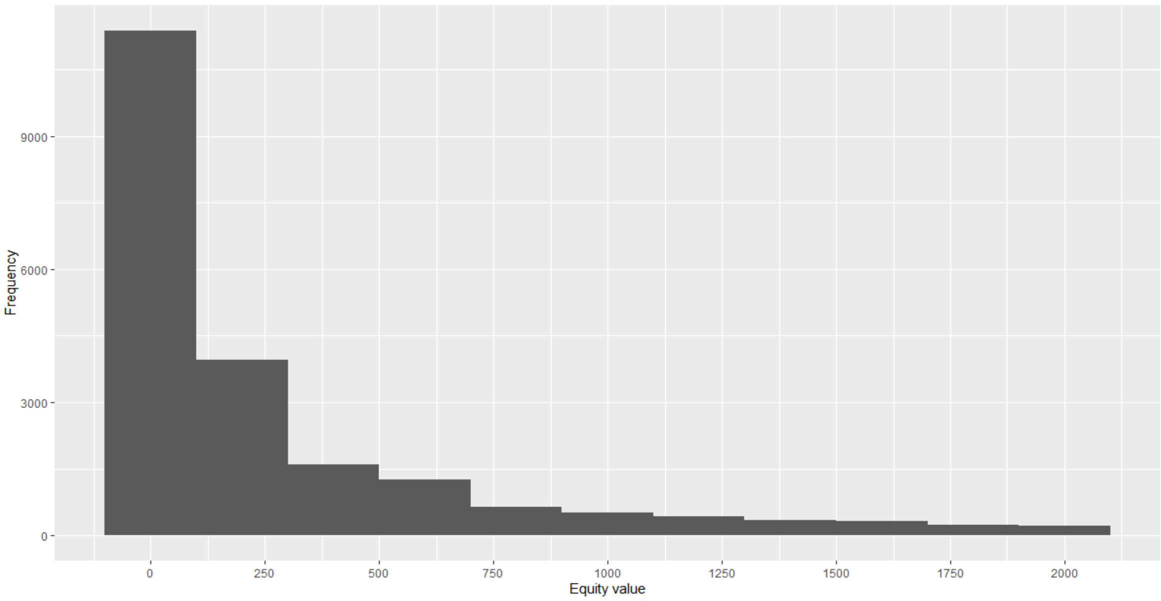
Similar as presented in the debt value calculation, the number of calculated monthly equity value observations should be 40,368. This is amount again represents the number of selected companies (841), multiplied by the number of months of the analysis (48). To avoid infinite estimates of distance to default values later, zero values will be omitted also from equity value database. Descriptive statistics of calculated equity values (E) are presented in Table 6 while their distribution can be seen in Figure 9. Time series of average equity values can be seen in Figure 10.

Table 6: Descriptive statistics of monthly equity values (in million €)

N	*N/A*	*N/A %*	*Mean*	*Std*	*Min*	*Q1*	*Median*	*Q3*	*Max*
28,905	*11,463*	*28%*	6,751	27,531	*1*	*29*	*198*	*1,959*	*834,569*

Source: Own work.

Figure 9: Distribution of monthly equity values (in million €)



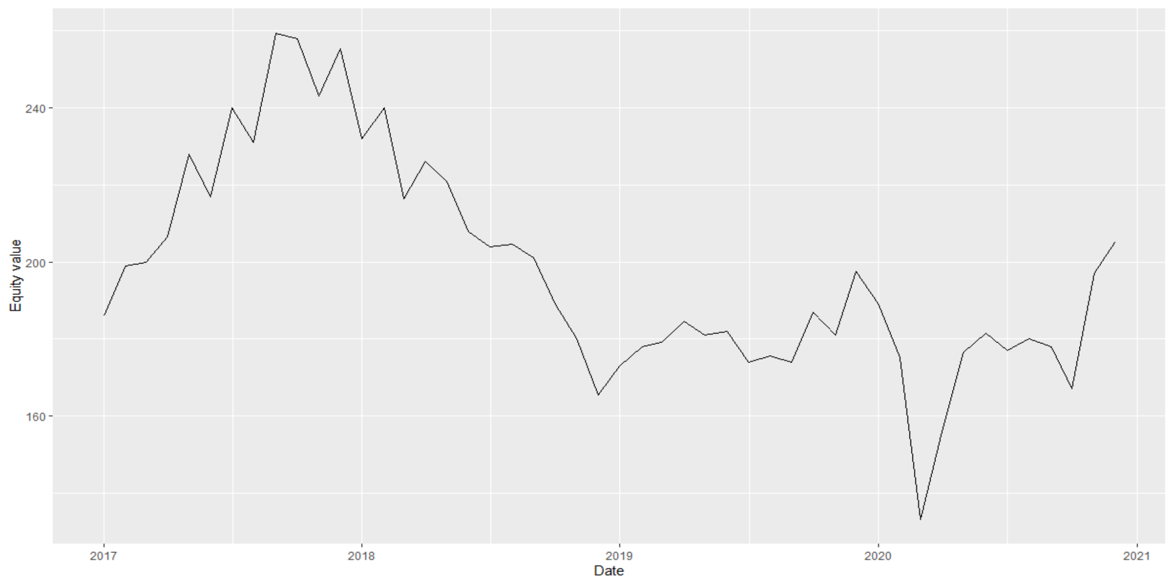
Source: Own work.

As seen in Table 6, the number of calculated equity value observations is 28,905. The total number of observations should be somewhat higher as mentioned before (40,368), but 28% of equity values have been omitted due to zero or missing values.

As further presented in Table 6, equity value of the half of observed companies is below 195 million € while only the top quarter reaches 1,950 million €. Since few companies have quite extremely high equity values, mean value and standard deviation do not provide any relevant information in that case. Distribution in Figure 9 presents only roughly the first three quarters of observations, similar to visual presentation of debt values.

Time series of average equity values (Figure 10) shows positive trend until second half of year 2017, when average equity values reached peak of roughly 260 million €. Afterwards, a somewhat negative trend with bottom line in the first half of year 2020 can be noticed.

Figure 10: Time-series of average monthly equity values (in million €)



Source: Own work.

2.2.4 Assets value and volatility

When all variables defined in previous chapters are estimated, values of firm's assets (A) and their volatility (σ_A) can finally be obtained. This will be done with iterative procedure of VX algorithm as presented in the Literature review. The two equations with two unknowns are

$$N(d_1)A - De^{-rT}N(d_2) - E = 0 \tag{24}$$

$$\left(\frac{A}{E}\right)N(d_1)\sigma_A - \sigma_E = 0 \tag{25}$$

where

$$d_1 = \frac{\ln\left(\frac{A}{D}\right) + \left[r + \frac{1}{2}\sigma_A^2\right]T}{\sigma_A\sqrt{T}} \tag{26}$$

$$d_2 = d_1 - \sigma_A\sqrt{T} \tag{27}$$

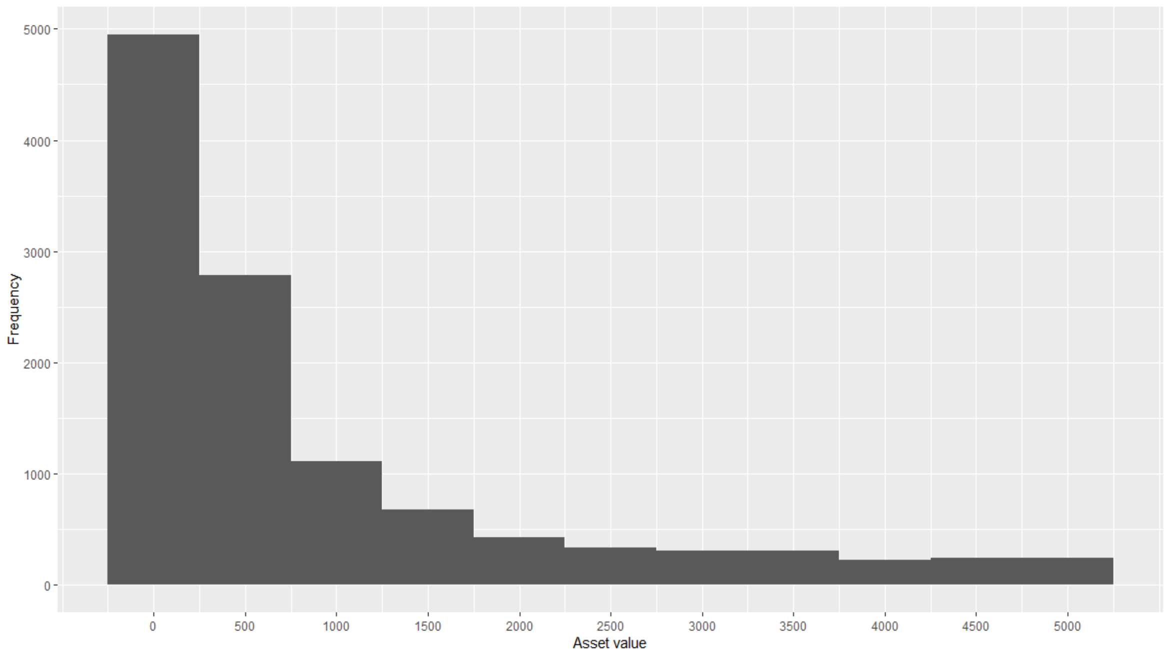
When all observable variables are inserted into above two equations, monthly asset values and volatilities for every company are obtained. The number of observations in each of the two datasets should be 40,368, which represents 841 of selected companies multiplied by 48 months. Descriptive statistics of asset values are presented in Table 7 while their distribution can be seen in Figure 11. Descriptive statistics of asset volatilities are presented in Table 8 while their distribution can be seen in Figure 13. Time series of average asset values and average assets volatilities are shown in Figure 12 and Figure 14, respectively.

Table 7: Descriptive statistics of monthly asset values (in million €)

N	N/A	N/A %	Mean	Std	Min	Q1	Median	Q3	Max
15,703	24,665	61%	9,675	32,392	5	180	763	5,163	839,625

Source: Own work.

Figure 11: Distribution of monthly asset values (in million €)



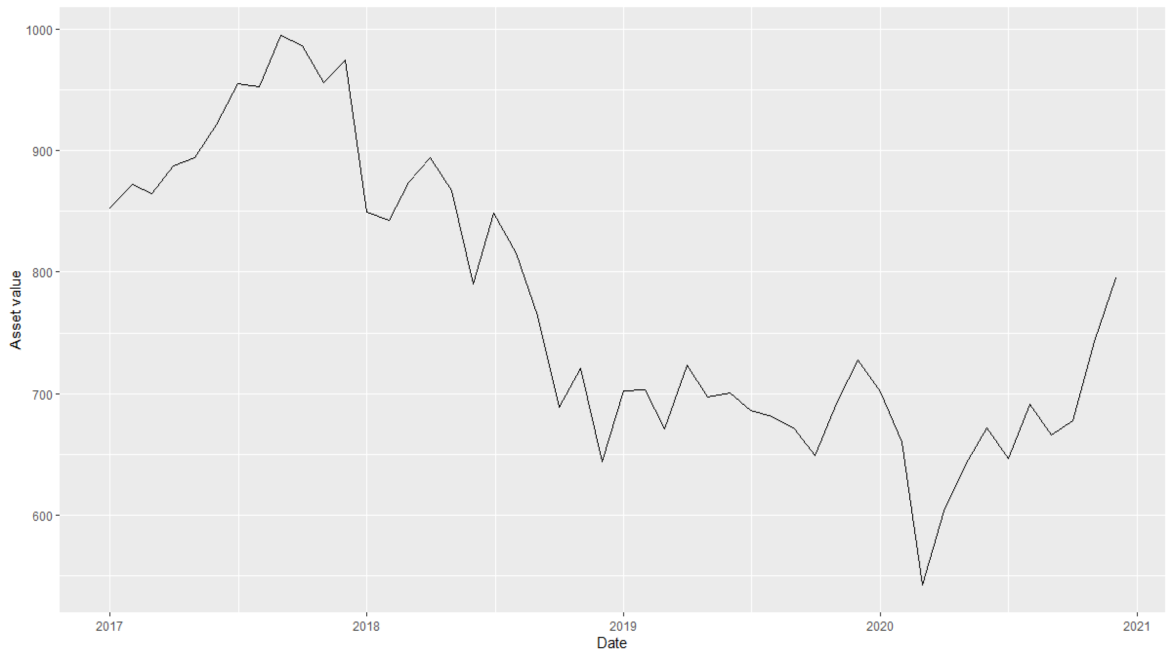
Source: Own work.

As seen in Table 7, the number of estimated assets value observations is 15,703. The total number of observations should be somewhat higher as mentioned before (40,368), but 61% of estimated assets values have been omitted due to missing values of input data.

As further presented in Table 7, estimated values of assets have similar distribution structure as equity and debt values, which was naturally expected. Estimated assets value of roughly half of companies is below 760 million €, which is indicated with median while additional quarter already exceeds 5,160 million €. Mean and standard deviation values similar as at equity and debt value analysis do not provide any relevance. Distribution in Figure 11 again presents only left three quarter of the dataset due extremely long tail present to the right.

Time series of average asset values (Figure 12) show a similar path as equity values, but on a much larger scale. More specifically, the peak in the second half of the year 2017 almost reaches 1,000 million € while the lowest drop in the beginning of year 2020 stops at around 550 million €. Similar path to equity values was however expected considering quite stable average debt values throughout the years noticed in face value of debt analysis.

Figure 12: Time-series of average monthly asset values (in million €)



Source: Own work.

Table 8: Descriptive statistics of monthly asset volatilities

N	*N/A*	*N/A %*	*Mean*	*Std*	*Min*	*Q1*	*Median*	*Q3*	*Max*
15,703	*24,665*	*61%*	*0.0180*	*0.0173*	*0.0000*	*0.0096*	*0.0152*	*0.0226*	*11.1990*

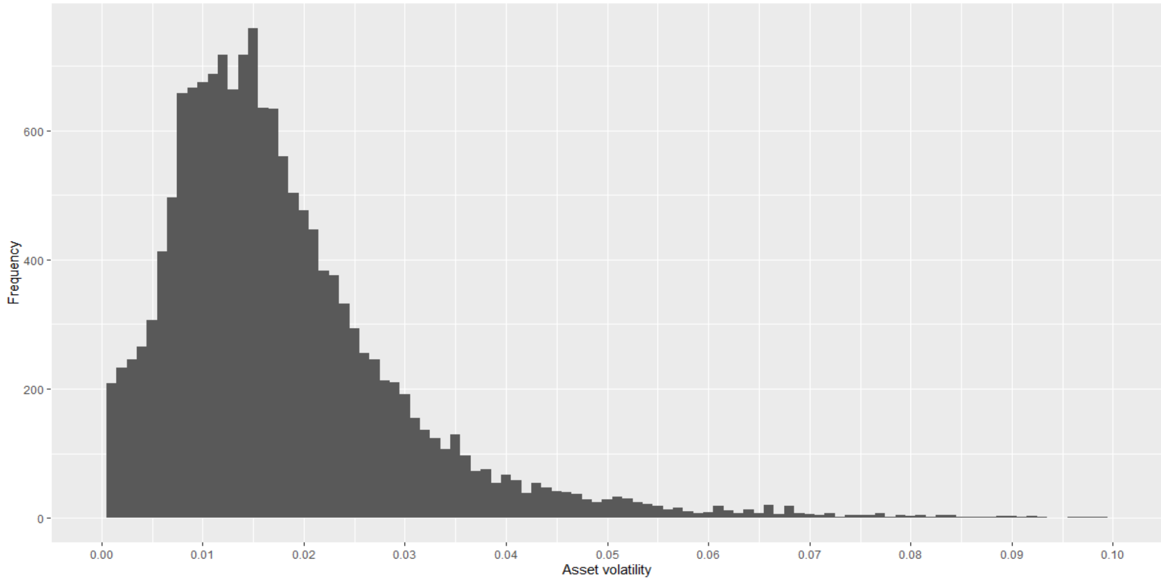
Source: Own work.

As seen in Table 8, the number of estimated asset volatility observations is 15,703. The total number of observations should be somewhat higher as mentioned before (40,368), but 61% of estimated asset volatilities have been omitted due to missing values.

Similar to equity volatility, asset volatility distribution indicates on heavily right skewed normal distribution with the mean being slightly higher than the median as seen in Table 8 and Figure 13. The distribution is naturally limited with zero on the left side while no limits have been implied on the right side. Asset volatility of three quarters of observations is lower than 2.3% while half of them falls below 1.5% as indicated with median. Mean value is at 1.8% while standard deviation is relatively high at 1.7%.

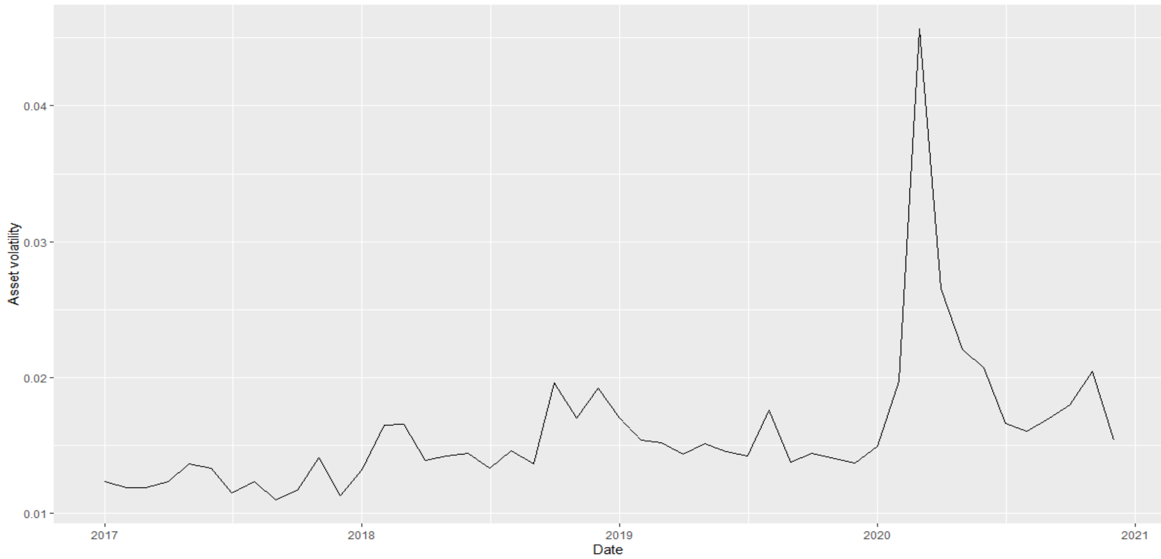
As seen in time series presentation in Figure 14, the average asset volatilities also follow a similar path as equity volatilities on average. Average estimated asset volatilities had been slightly increasing until the beginning of year 2020 within range of 1% and 2%. In year 2020, volatility extremely increased almost to 5% and then shortly dropped back to below 2% level.

Figure 13: Distribution of monthly asset volatilities



Source: Own work.

Figure 14: Time-series of average monthly asset volatilities



Source: Own work.

2.2.5 Distance to default

After estimation of asset values and volatilities is completed, all obtained values can finally be used in distance to default (DD) equation

$$DD = \frac{\ln\left(\frac{A}{D}\right) + \left[\mu_A - \frac{1}{2}\sigma_A^2\right]T}{\sigma_A\sqrt{T}} \tag{28}$$

The only missing value within the equation however remains the expected assets return (μ_A). As presented in the Literature review, various approaches have been used to estimate that value appropriately. Some authors assumed that assets have the same expected return as the risk-free rate (r) while the others that some risk premia should be added as well. In this thesis, the usage of risk-free rate (r) as the expected assets return (μ_A) will be made to somewhat simplify the process.

After inserting the last remaining values of expected assets returns, the distance to default (DD) equation can now be solved. Considering that the initial number of selected companies was 841, the total number of DD observations should therefore be 40,368. This amount represents the number of selected companies (841), multiplied by the number of months of the analysis (48). Since quite a large part of observations have been omitted from the analysis due to several specific values as presented in the previous chapters, a significantly smaller dataset of estimated DD values is however anticipated. Descriptive statistics of distance to default (DD) values are presented in Table 9, while their distribution can be seen in Figure 15. In addition, time series of average distance to default (DD) values is also presented as time-varying statistics in Figure 16.

Table 9: Descriptive statistics of monthly distance to default (DD) values

N	*N/A*	*N/A %*	*Mean*	*Std*	*Min*	*Q1*	*Median*	*Q3*	*Max*
15,444	*24,924*	*62%*	*190*	*188*	*1*	*89*	*145*	*230*	*4,212*

Source: Own work.

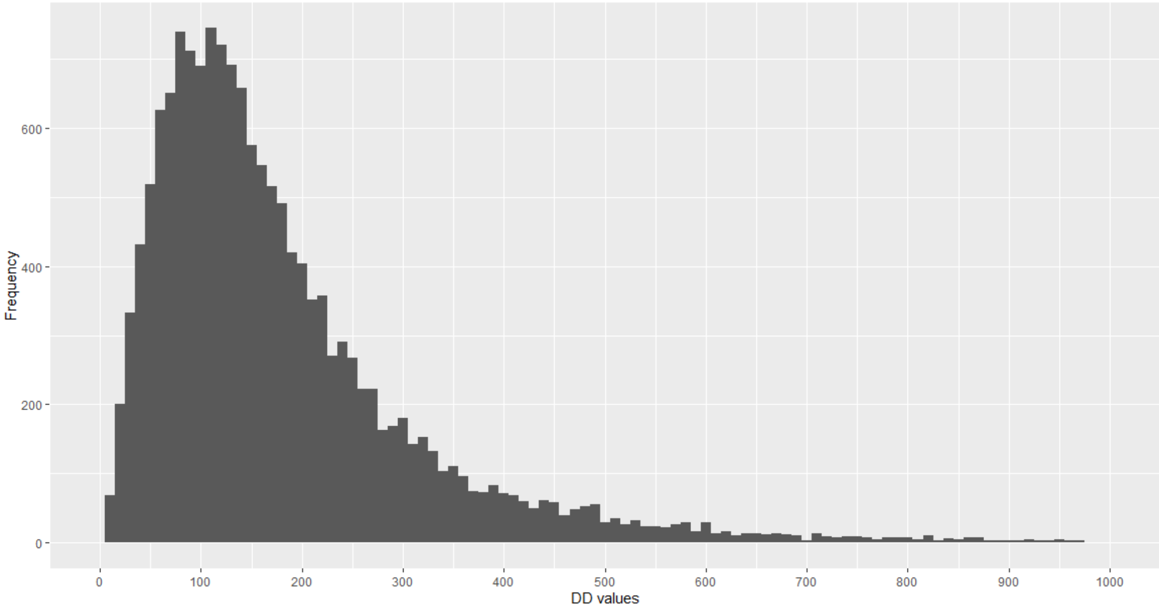
As seen in Table 9, the number of estimated monthly DD value observations is 15,444. The total number of observations should be significantly higher as mentioned before (40,368), but 62% of values that were needed for DD value estimation have been omitted in previous chapters.

Distribution in Figure 15 shows that the estimated distance to default values have similar characteristics as equity or asset volatility values. All three distributions follow a quite well shaped normal distribution with strong skewness to the right. Contrary to a long tail of observations noticed on the right, the left side is naturally bounded at zero. However, note that few estimates of distance to default were still below zero, but have been excluded from the analysis as they do not provide any explanatory power. These few observations had relatively high debt values compared to its market equity values and extremely high equity volatility, which brought the outcome of the DD equation even below zero.

As seen from descriptive statistics (Table 9), the median of estimated distance to default values is 145 while roughly half of the observations centred around median falls between 90 and 230. The mean value of DD estimates is 190 while standard deviation at relatively high 188. As presented in the Literature review, estimated DD values should measure the number of standard deviations firm is away from default. Considering the estimated mean value from

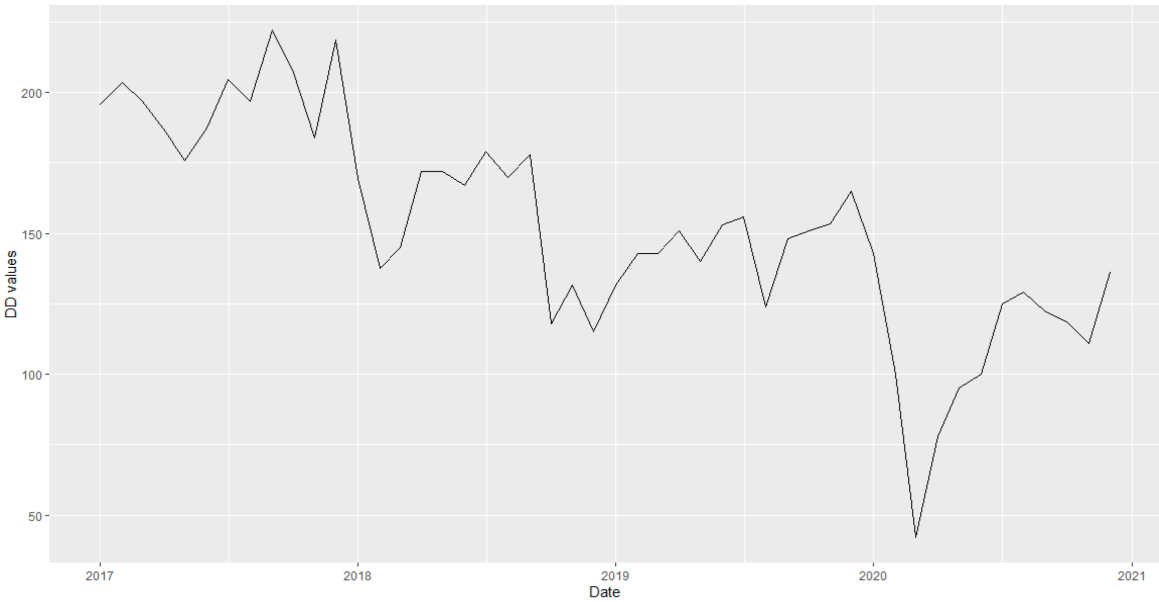
Table 9, analysed firms are on average 190 standard deviations away from default. Since this number suggest on extremely low probability of default, estimated DD values should still be very useful for ranking of firms regarding their default risk.

Figure 15: Distribution of monthly distance to default (DD) values



Source: Own work.

Figure 16: Time-series of average monthly distance to default (DD) values



Source: Own work.

Time series presentation (Figure 16) shows that average distance to default values follow a very similar path as average equity and asset values. Distance to default values had been decreasing during the observed period where the highest drop occurred in the beginning of year 2020. Average distance to default values were at around 200 in year 2017 while the year 2020 was ended with values of around 130. The highest drop in the beginning of year 2020 caused average distance to default value to fall even below 50 at one point.

2.3 Inclusion of private firms into the model

Estimated distance to default (DD) values as obtained in the previous chapter can now be used to construct credit scoring model for private firms. Developed credit scoring model should eventually provide monthly DD values of any Slovenian firm that has been initially selected. To obtain monthly DD values, some coefficients should be estimated first, which would be able to translate DD values of German public firms to Slovenian private firms.

Before aforementioned coefficients can be estimated, public and private firms should be matched through some common characteristics. Since these characteristics should also represents some reasonable determinants of default, financial variables as extracted from financial statements can be used for that purpose. Additionally, industry sampling procedure would also need to be done first as financial variables might significantly differ among companies from different industries. A more detailed process of model construction is presented in the following four subchapters.

2.3.1 Industry sampling

As presented in the Literature review, financial statements can significantly differ among companies from different industries, which may indicate on biased relations if general comparisons are made. In order to find more appropriate and relevant relations, industry sampling would therefore represent the first step in private firm extension DD model construction.

The industry sampling procedure will be based on NACE code classification, which is one of the most widely used classification systems around the world as explained in the Literature review. The NACE code contains four different hierarchical levels and only the top level (Level 1) will be used in sampling procedure of the model.

The Level 1 classification consists of 21 broad industries, which are based on the type of activities companies are involved in. Since some industries are often limited to public entities or any other specific institutions, the total number of groups in corporate sampling procedures is usually somewhat smaller.

To match private firms with public ones in extended DD model construction, industry sampling for both groups should be done. The distribution of all German public companies that have been initially collected for DD parameter estimation process within industry groups is presented in Table 10, while the distribution of retrieved Slovenian private firms can be seen in Table 11.

Table 10: Industry groups with the number and shares of public firms

NACE code	*n*	*n %*
A	*1*	*0%*
B	*8*	*1%*
C	*254*	*30%*
D	*17*	*2%*
E	*4*	*0%*
F	*14*	*2%*
G	*46*	*5%*
H	*16*	*2%*
I	*3*	*0%*
J	*115*	*14%*
K	*228*	*27%*
L	*62*	*7%*
M	*36*	*4%*
N	*7*	*1%*
O	*0*	*0%*
P	*2*	*0%*
Q	*12*	*1%*
R	*16*	*2%*
S	*0*	*0%*
T	*0*	*0%*
U	*0*	*0%*
Total	***841***	***100%***

Source: Own work.

According to Table 10, the total number of distributed German public companies is 841. The far largest part operates in industry C, which represents 30% of all sampled companies. Other industries have significantly lower shares, some of them are even without a single representative.

As seen in Table 11, the total number of sampled Slovenian private companies is 946. The largest represented industry in this case is industry G, which is extremely higher compared to the same industry in the public firm allocation table. Industry G represents high 36% share of total sampled private companies while only 5% share is seen considering public firms. The second highest share of 33% in private firm distribution represents industry C, which is quite similar to 30% share in public firm distribution. Other industries have almost negligible shares, which is also similar to public firms.

Table 11: Industry groups with the number and shares of private firms

NACE code	*n*	*n %*
A	*7*	*1%*
B	*2*	*0%*
C	*315*	*33%*
D	*29*	*3%*
E	*23*	*2%*
F	*41*	*4%*
G	*337*	*36%*
H	*61*	*6%*
I	*9*	*1%*
J	*39*	*4%*
K	*8*	*1%*
L	*7*	*1%*
M	*45*	*5%*
N	*18*	*2%*
O	*0*	*0%*
P	*0*	*0%*
Q	*2*	*0%*
R	*3*	*0%*
S	*0*	*0%*
T	*0*	*0%*
U	*0*	*0%*
Total	***946***	***100%***

Source: Own work.

2.3.2 Variable selection

When companies are allocated into industries, variable selection processes can be initiated. As presented in the Literature review, financial ratios from Altman's Z-Score model would be the right choice for that manner.

Altman's Z-Score model suggests five financial ratios, which have been proven to be the most significant determinants of default in his work. The model is based on Multiple Discriminant Analysis using only accounting data in its equation. Each of the variables represents different filed of measurement, their definition is again presented below.

$$X_1 = \frac{Working\ Capital}{Total\ Assets} \quad (29)$$

$$X_2 = \frac{Retained\ Earnings}{Total\ Assets} \quad (30)$$

$$X_3 = \frac{Earnings\ before\ Interest\ and\ Taxes}{Total\ Assets} \quad (31)$$

$$X_4 = \frac{Book\,Value\,of\,Equity}{Book\,Value\,of\,Total\,Liabilities} \tag{32}$$

$$X_5 = \frac{Total\,Sales}{Total\,Assets} \tag{33}$$

In private firm extension DD model, presented variables will be extracted from financial statements for public and private firms. Since accounting data usually updates annually, four different values of the same financial ratio should be obtained for each firm. With using five different financial ratios, 20 records would therefore be obtained for each company.

In the case of public firms, the number of financial ratio sets should therefore be 3,364. This amount represents the number of public companies used in the model (841), multiplied by the number of years of the analysis (4). To get some insights into financial ratios, average ratios of each industry as well as overall average ratios are presented in Table 12.

Table 12: Average of public firms' financial ratios, grouped by industries

NACE code	*X1*	*X2*	*X3*	*X4*	*X5*	*n*	*n %*
A	*1.26*	*5.10*	*65.72*	*0.32*	*0.39*	*4*	*0%*
B	*0.18*	*0.21*	*-0.02*	*0.60*	*0.01*	*28*	*1%*
C	***0.25***	***0.31***	***0.05***	***0.79***	***0.95***	***945***	***40%***
D	*0.08*	*0.16*	*0.03*	*0.38*	*0.47*	*62*	*3%*
E	*0.56*	*0.15*	*0.03*	*1.83*	*0.74*	*16*	*1%*
F	*0.15*	*0.04*	*0.02*	*0.09*	*0.65*	*45*	*2%*
G	*0.19*	*0.23*	*0.03*	*0.64*	*1.56*	*177*	*7%*
H	*0.05*	*0.22*	*0.03*	*0.36*	*0.59*	*54*	*2%*
I	*0.18*	*0.20*	*-0.03*	*0.39*	*0.53*	*12*	*1%*
J	*0.12*	*0.30*	*0.05*	*0.84*	*0.87*	*427*	*18%*
K	*0.11*	*0.19*	*0.00*	*0.63*	*0.12*	*161*	*7%*
L	*0.06*	*0.30*	*0.05*	*0.71*	*0.08*	*180*	*8%*
M	*0.34*	*0.16*	*0.00*	*0.74*	*0.43*	*121*	*5%*
N	*-0.07*	*0.00*	*0.06*	*0.19*	*0.80*	*27*	*1%*
O	*/*	*/*	*/*	*/*	*/*	*0*	*0%*
P	*-9.44*	*0.56*	*-126.19*	*0.02*	*7.98*	*5*	*0%*
Q	*0.07*	*0.25*	*0.04*	*0.66*	*0.63*	*44*	*2%*
R	*0.19*	*0.31*	*0.11*	*0.86*	*0.72*	*59*	*2%*
S	*/*	*/*	*/*	*/*	*/*	*0*	*0%*
T	*/*	*/*	*/*	*/*	*/*	*0*	*0%*
U	*/*	*/*	*/*	*/*	*/*	*0*	*0%*
Total	***0.19***	***0.27***	***0.04***	***0.73***	***0.80***	***2,367***	***100%***
A-U	*NA*	*NA*	*NA*	*NA*	*NA*	*997*	*30%*

Source: Own work.

As seen in Table 12, the total number of analysed financial ratio sets is 2,367. The total number of sets should however be much higher (3,364) as mentioned before, but around 30% of total sampled observations has been omitted due to missing values. More specifically, the data of at least one of the five financial ratios has been insufficient for 997 observations. The largest part of omitted observations refers to industry K as seen from share decrease compared to the initial industry sampling table presented before. The reason for the large proportion of insufficient financial ratios of industry K is probably in its somewhat specific financial statements reporting.

Table 13: Average of private firms' financial ratios, grouped by industries

NACE code	X1	X2	X3	X4	X5	n	n %
A	0.20	0.14	0.04	1.49	1.87	25	1%
B	-0.10	0.05	0.02	0.69	0.87	8	0%
C	**0.18**	**0.20**	**0.07**	**1.18**	**1.23**	**1,237**	**34%**
D	0.08	0.00	0.03	2.46	0.55	112	3%
E	0.19	0.03	0.03	1.17	1.57	92	2%
F	0.12	0.19	0.05	0.77	1.54	160	4%
G	**0.22**	**0.16**	**0.06**	**0.66**	**2.38**	**1,309**	**35%**
H	0.11	0.18	0.06	0.76	1.42	239	6%
I	-0.03	0.05	0.04	1.26	0.46	36	1%
J	0.17	0.15	0.06	0.96	1.44	151	4%
K	0.14	0.14	0.09	7.63	0.56	32	1%
L	0.20	0.00	0.03	0.25	0.59	25	1%
M	0.15	0.13	0.05	0.73	1.56	174	5%
N	0.14	0.11	0.05	0.73	1.84	72	2%
O	/	/	/	/	/	0	0%
P	/	/	/	/	/	0	0%
Q	-0.05	0.10	0.05	0.36	0.62	8	0%
R	0.09	0.00	0.12	1.32	2.62	12	0%
S	/	/	/	/	/	0	0%
T	/	/	/	/	/	0	0%
U	/	/	/	/	/	0	0%
Total	**0.18**	**0.15**	**0.06**	**0.90**	**1.60**	**3,692**	**100%**
A-U	NA	NA	NA	NA	NA	92	2%

Source: Own work.

Another important outcome noticed in Table 12 is a quite high diversification of average financial ratio values among different industries. The number of observations within the majority of industries is however probably too small to make some relevant conclusions, but the comparison between quite well represented industry C and overall industry averages may give some clues. As seen in the table, the average values of all five financial ratios are somewhat higher in case of industry C compared to overall averages. According to Altman model presented in the Literature review, this suggests that industry C may be on average less risky compared to overall general economy.

With moving the analysis towards private firms, 3,784 sets of financial ratios would be expected knowing that 946 of private companies are considered in the model. The expected number similar to public firms represents the number of private companies (946) multiplied by the number of years of the analysis (4). Average ratios of each industry as well as overall average ratios of all industries are presented in Table 13.

As seen in Table 13, the total number of analysed financial ratio sets of private firms is 3,692. This number is contrary to public firm dataset only slightly lower than the expected number of sets (3,784). More specifically, only 2% of total expected observations has been omitted due to missing values or insufficient data of at least one of the five financial ratios. Shares of all industries remain quite the same compared to initial structure of companies, and missing values are therefore fully random in case of private firms.

By comparing overall average values of financial ratios with other industries, high diversification of values can be seen also in case of private firms. Focusing on the most representative industries C and G, all financial ratios are at least the same to overall averages. The only exception is variable X4, which is somewhat lower than the overall average in case of industry G. These results suggest that even in private firms' sample, the largest industries tend to be less risky than the overall economy on average.

2.3.3 Coefficient estimation

The five selected financial ratios should therefore represent determinants of default that can be found at any private or public firm. To estimate distance to default (DD) values of private firms, indirect projection from public firms' DD values within each industry should be done as presented in the Literature review.

Since public firm DD values have been estimated on a monthly basis in Parameter estimation chapter, 12 distance to default estimates should be matched with each of the 2,367 annual financial ratio sets presented in previous section. Therefore, multiplying the number of obtained financial ratio sets (2,367) by 12 months, 28,404 monthly DD estimates of public firms are expected in total. Average values of DD estimates for public firms within each industry are presented in Table 14.

As seen in Table 14, the total number of monthly DD estimates of public firms is 14,469. This number represents only around half (51%) of the initially expected observations (28,404) if the number of financial ratio sets is considered. The reasons for quite a large proportion of omitted observations are predominately missing or extreme values in distance to default estimation process.

Table 14: Average DD values of public firms, grouped by industries

NACE code	*DD values*	*n*	*n %*
A	*138*	*24*	*0%*
B	*81*	*48*	*0%*
C	***145***	***6,437***	***44%***
D	*163*	*384*	*3%*
E	*/*	*0*	*0%*
F	*141*	*249*	*2%*
G	*146*	*994*	*7%*
H	*159*	*336*	*2%*
I	*/*	*0*	*0%*
J	*147*	*2,889*	*20%*
K	*123*	*626*	*4%*
L	*184*	*1,027*	*7%*
M	*173*	*704*	*5%*
N	*116*	*144*	*1%*
O	*/*	*0*	*0%*
P	*/*	*0*	*0%*
Q	*157*	*383*	*3%*
R	*232*	*224*	*2%*
S	*/*	*0*	*0%*
T	*/*	*0*	*0%*
U	*/*	*0*	*0%*
Total	***150***	***14,469***	***100%***
A-U	*NA*	*13,935*	*49%*

Source: Own work.

The largest share of the remaining observations refers to estimated DD values in industry C. The share of industry C is 44%, which is even slightly higher compared to the structure of financial ratio sets as presented in the previous chapter. Other industries have significantly lower shares, some of them are even without single representative.

Considering presented distance to default estimates, it can be seen that average values quite differ among industries. The safest industry with the largest average DD value seems to be industry R while average DD estimates suggest that industry B might be the riskiest. However, since the majority of industries are rather weakly represented, fully unambiguous conclusions cannot be made.

The highest represented industry C records average DD value of 145, which is somewhat less than the average of all industries. This suggests that industry C might be slightly riskier than the overall economy, which is not fully in line with Altman's Z score model as presented in the previous chapter. The difference between the two outcomes might however be significantly lower dataset in case of estimated DD values compared to sets of financial ratios.

To start with projection of public firms DD values to private firms, the first step should be done with a liner regression of estimated public firms' DD values with their respective financial ratios within each industry as presented in the Literature review. The linear regression, which would be done for each industry and for each month separately is as follows

$$DD_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \beta_3 X_{3,i} + \beta_4 X_{4,i} + \beta_5 X_{5,i} + \varepsilon_i,\ i = 1, \ldots, n \qquad (34)$$

where DD_i is distance to default of public firms, variables $X_{1-5,i}$ are financial ratios, ε_i is the disturbance term and *n* represents the number of firms.

As further presented in the Literature review, at least 3-month lag should be assigned to DD values after the date of financial ratios to allow time for information spread. In a regression where financial ratios are based on date *d*, DD estimations of *d+3 months* would therefore be used. Such an approach is presented in regression equation below

$$DD_{d+m} = \beta_{0,d+m} + \beta_{1,d+m} X_{1d} + \beta_{2,d+m} X_{2d} + \beta_{3,d+m} X_{3d} + \beta_{4,d+m} X_{4d} + \beta_{5,d+m} X_{5d} + \varepsilon_{d+m},\ m = 3, \ldots, 14 \qquad (35)$$

where *d* is a certain date and *m* represents the number of months that are limited to one year (m=12). Note that indicator *i* has been neglected for simplicity.

With conduction of above presented regressions, 48 monthly sets of β_{0-5} coefficients should be obtained within each industry. Since some of industries have no representatives, coefficient estimation of all industries can unfortunately not be done. Furthermore, each of the aforementioned 48 monthly regressions within each industry needs to have sufficient observations in each point in time that coefficients can be appropriately estimated. For insight, the average number of observations (companies) within each month and industry are presented in Table 15.

As can be seen from Table 15, the average number of companies within each month is 301 considering all industries together. Among all industries, only industry C shows an appropriate average number of companies in each point in time (134), while others contain less than a half of that value. Since the only industry with sufficiently high average number of companies within each month is industry C, coefficient estimation for only this industry will be done. Performing therefore regression analysis as presented before, 48 monthly sets of β_{0-5} coefficients are obtained for industry C. Descriptive statistics for each of the six coefficients for industry C are presented in Table 16.

Table 15: Average number of public firms in each month, grouped by industries

NACE code	*n*	*n / m*
A	*24*	*1*
B	*48*	*1*
C	***6,437***	***134***
D	*384*	*8*
E	*0*	*0*
F	*249*	*5*
G	*994*	*21*
H	*336*	*7*
I	*0*	*0*
J	*2,889*	*60*
K	*626*	*13*
L	*1,027*	*21*
M	*704*	*15*
N	*144*	*3*
O	*0*	*0*
P	*0*	*0*
Q	*383*	*8*
R	*224*	*5*
S	*0*	*0*
T	*0*	*0*
U	*0*	*0*
Total	***14,469***	***301***

Source: Own work.

Table 16: Descriptive statistics of all six estimated coefficient for industry C

	Min	*Q1*	*Median*	*Mean*	*Q3*	*Max*	*Std*	*n*
β_0	*49.11*	*128.69*	*169.63*	*170.81*	*216.65*	*270.30*	*50.77*	*48*
β_1	*-34.64*	*36.48*	*146.00*	*146.05*	*225.13*	*352.59*	*107.50*	*48*
β_2	*-79.58*	*-18.07*	*-2.84*	*-11.06*	*-1.15*	*41.93*	*19.57*	*48*
β_3	*-4.33*	*-1.59*	*0.67*	*12.57*	*12.40*	*173.65*	*30.50*	*48*
β_4	*0.41*	*1.35*	*2.06*	*2.06*	*2.80*	*5.72*	*1.03*	*48*
β_5	*-48.99*	*-12.24*	*-7.56*	*-9.62*	*-3.35*	*19.11*	*12.04*	*48*

Source: Own work.

Table 16 therefore presents some features of estimated β_{0-5} coefficients. As can be seen, a quite normal distribution with similar median and mean values is noticed at β_0, β_1, β_4 and

β_5 coefficients while significantly different centred values suggest heavily skewed tails at β_2 and β_3.

2.3.4 DD calculation

When the above presented beta coefficients are obtained, private firms mapping represents the next step. As presented in the Literature review, mapping can be done with simple multiplication of obtained coefficients with their respect financial ratios of private firms. Multiplied coefficients and financial ratios are then summed up which eventually gives public firm equivalent DD value. The mapping procedure with multiplication and addition process is presented below

$$\widetilde{DD} = \widehat{\beta_0} + \widehat{\beta_1}X_1 + \widehat{\beta_2}X_2 + \widehat{\beta_3}X_3 + \widehat{\beta_4}X_4 + \widehat{\beta_5}X_5 \tag{36}$$

where $\widetilde{DD}$ is public firm equivalent DD value, variables $X_{1\text{-}5}$ are financial ratios of private firms and $\widehat{\beta_{0-5}}$ are estimated coefficients obtained from regressions presented in the previous chapter.

Similar to coefficient estimation process, the same 3-month lag should also be incorporated into calculation of public-firm equivalent DD values as presented in the Literature review. Therefore, financial ratios of private firms based on date *d* should only be multiplied with coefficients estimated in *d+3 months* until *d+14 months* to obtain their respective public-firm equivalent DD values. The equation showing such an approach is

$$\widetilde{DD}_{d+m} = \hat{\beta}_{0,d+m} + \hat{\beta}_{1,d+m}X_{1d} + \hat{\beta}_{2,d+m}X_{2d} + \hat{\beta}_{3,d+m}X_{3d} + \hat{\beta}_{4,d+m}X_{4d} + \hat{\beta}_{5,d+m}X_{5d},$$
$$m = 3, \dots, 14 \tag{37}$$

where *d* is a certain date and *m* represents the number of months that are limited to one year (m=12). Note that indicator *i* has been neglected for simplicity.

With multiplication and addition of previously obtained 48 monthly sets of β_{0-5} coefficients and five calculated financial ratios of each private firm from industry C, 48 monthly public firm equivalent DD values should be obtained for each of firm. The number of initially sampled private firms in industry C was 315, therefore maximum of 15,120 monthly public firm equivalent DD values could be obtained. Since some missing financial ratios have been omitted as presented in the Variable selection chapter, the total expected number of monthly public firm equivalent DD estimates decreases to 14,844. This is calculated as the number of annual financial ratio sets of industry C (1,237) multiplied with 12 months.

Obtained public firm equivalent DD values will be presented in following tables and figures where comparison with initial DD values of public firms from industry C will also be made. Descriptive statistics of public and private firms' DD values are therefore presented in Table

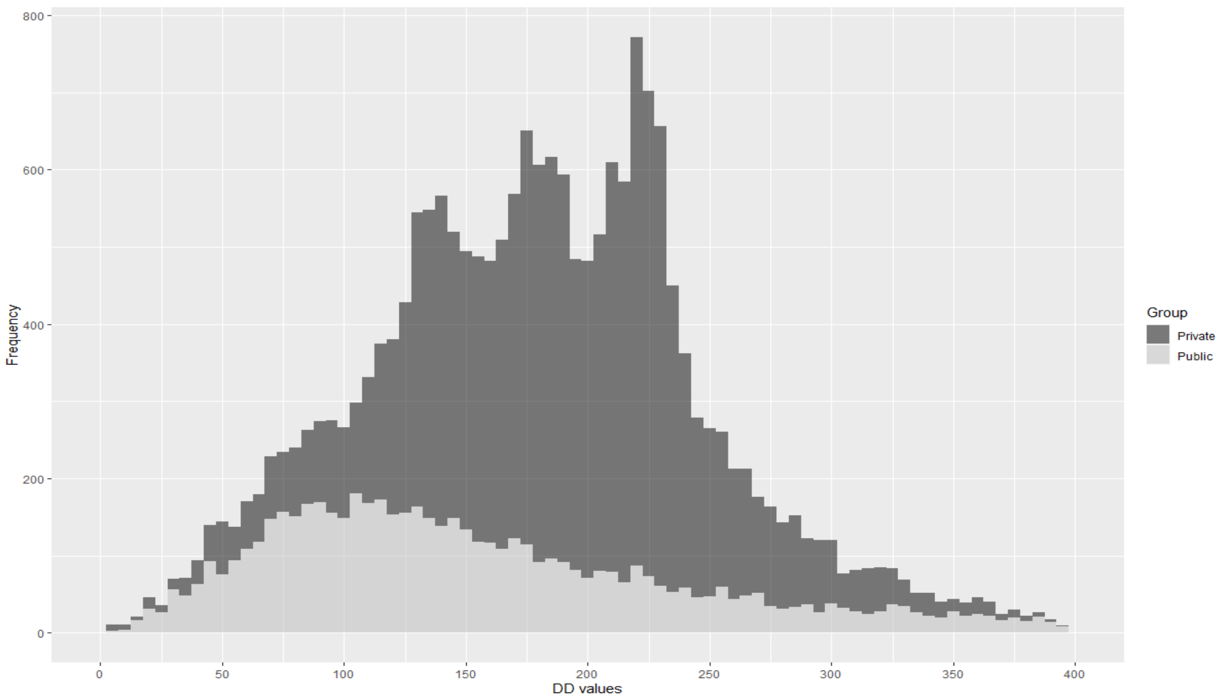
17, while their distribution can also be seen in Figure 17. In addition, time series of both average values is also presented as time-varying statistics in Figure 18.

Table 17: Descriptive statistics of monthly DD values for industry C

Private firms (industry C)									
N	*N/A*	*N/A %*	*Mean*	*Std*	*Min*	*Q1*	*Median*	*Q3*	*Max*
14,804	*40*	*0%*	*189*	*60*	*1*	*148*	*190*	*226*	*504*
Public firms (industry C)									
N	*N/A*	*N/A %*	*Mean*	*Std*	*Min*	*Q1*	*Median*	*Q3*	*Max*
6,437	*4,903*	*43%*	*195*	*203*	*1*	*95*	*145*	*230*	*4,212*

Source: Own work.

Figure 17: Distribution of monthly DD values for industry C



Source: Own work.

As seen in Table 17, the number of public firm equivalent DD estimates in industry C is 14,804. This number is only slightly lower than expected considering financial ratio sets, since some public firm equivalent DD values have been estimated below zero and thus omitted. However, omitted values represented only negligible 40 observations, which represents less than 1% of all values.

Considering the distribution of public firm equivalent DD estimates, it can be further seen in Table 17 that mean and median values are practically the same at 190. The centred half of the estimates falls between 148 and 226 while the standard deviation is 60. All these values suggest on normal distribution of obtained public firm equivalent DD estimates, which can also be seen in distribution Figure 17.

Comparing both private and public firm sample groups, it seems that private firms might, on average, be somewhat less risky than public firms in industry C. More specifically, average estimated DD value of private firms is 190 while distance to default was 145 in case of public firms in industry C. Somewhat different is consequently also the distribution figure of private firms compared to public firms in industry C, as can further be seen in Figure 17.

Figure 18: Time-series of average monthly DD values for industry C

Source: Own work.

Time series Figure 18 shows that the average distance to default values of private firms follow a very similar path as public firms. Both distance to default values had been decreasing in the observed period where the highest drop occurred in the beginning of year 2020. Average distance to default values of private firms were at around 250 in year 2017 while the year 2020 ended with values of around 150. The highest drop in the beginning of year 2020 caused that average distance to default value fell even close to 50 at one point.

3 RESULTS

Once the public firm equivalent DD values are obtained, the testing procedure can start. The testing procedure will be based on some standard statistical tests considered with regression analysis. Regression analysis should provide forecasting accuracy strength of the model, external probability of default values should serve as a benchmark within this analysis. To choose the most appropriate regression type, the relation between estimated DD values and benchmark PD values should be determined first. The relation will be checked with ranking power analysis of the newly developed model.

Conclusions made with testing procedure will then also be checked with case study analysis of some selected companies. More specifically, the ranking power and forecasting accuracy will be checked with comparative analysis of selected companies with overall industry. Case study analysis should eventually present model implication in practice, where movements of estimated DD values of selected companies will also be analysed in detail.

Model testing procedure will be more in detail presented in Chapter 3.1, which would also try to answer the first hypothesis of this thesis that market-based model can predict defaults earlier than the accounting-based one. Description of model implication on the other side will be presented in Chapter 3.2, where the second hypothesis that the market-based model can be used in the financial industry as an efficient monitoring tool will be considered.

3.1 Model testing

The testing procedure will start with assessment of ranking power of the model. Ranking power should test whether obtained DD estimates can indeed correctly rank companies regarding their riskiness. To do that, the probability of default (PD) values as presented in the Data collection chapter will be used as a benchmark. Since benchmark PD values have been extracted on an annual basis, some adjustments of monthly DD estimates will be done first.

Ranking power analysis should give some insight of relation between DD estimates and benchmark PD values. The relation assumption will be crucial in model accuracy testing performed next. Model forecasting accuracy will be checked with performing regression analysis, the outcome should answer the first hypothesis of this thesis. Ranking power and forecasting accuracy analysis will be more in detail presented in two subchapters below.

3.1.1 Ranking power

Ranking power of newly developed model will be checked with comparison of obtained public firm equivalent DD values and a benchmark. As already revealed in Data collection chapter, probability of default (PD) values as estimated by the selected Slovenian bank will be used as a benchmark in this thesis.

The selected Slovenian bank estimates PD values based on several parameters, but only PD values based on annual financial statements will be used as a comparison with estimated DD values. The reason to use PD values based only on annual financial statements is to compare market-based model of DD estimates with pure accounting model of PD estimates.

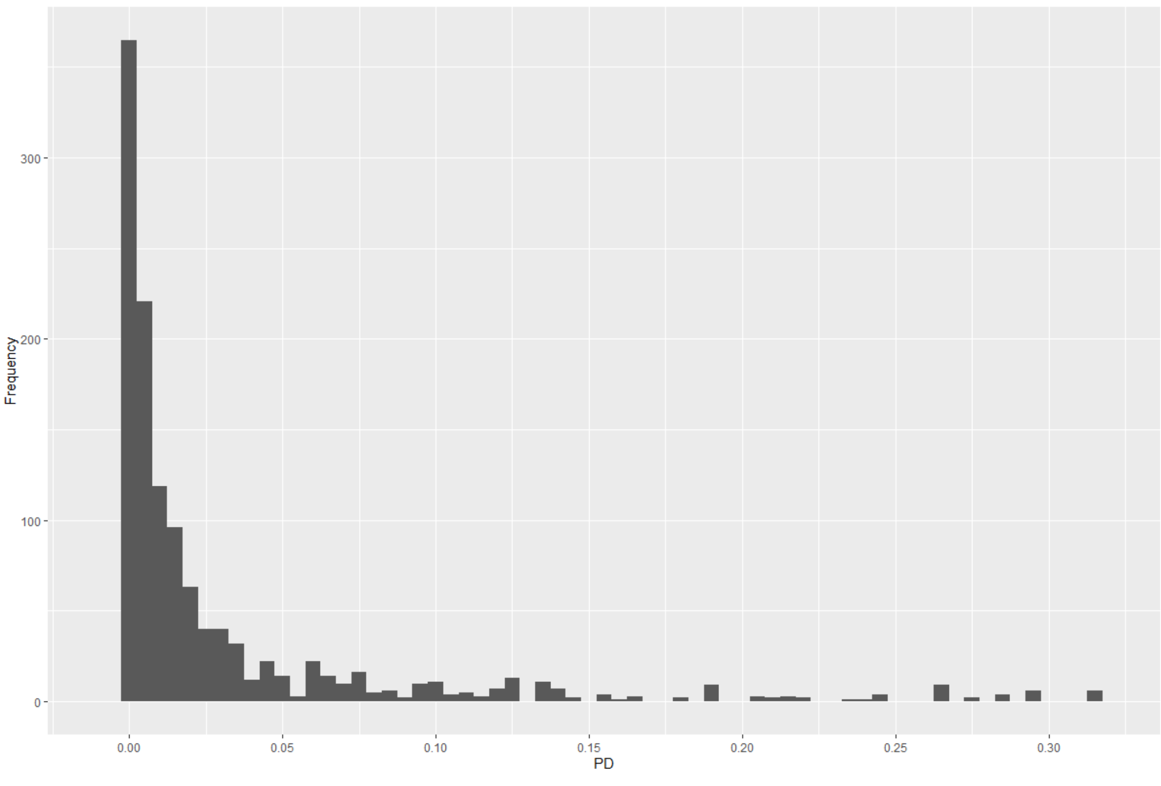
Annual PD values have been extracted for all initially selected Slovenian companies, therefore four PD values for each of the 946 companies from period 2017-2020 were obtained. Since public firm equivalent DD values have been estimated only for 315 companies from industry C, the database of PD values will adjust in accordance with that. The total number of observations should therefore be 1,260, which is calculated as the number of companies (315) multiplied by four years of the analysis. However, on account of some missing values in DD estimation process as presented in the Model construction chapter, the total number of observations would be slightly smaller. Descriptive statistics of accordingly adjusted PD value dataset are presented in Table 18 while their distribution can be seen in Figure 19.

Table 18: Descriptive statistics of annual extracted PD values

N	*N/A*	*N/A %*	*Mean*	*Std*	*Min*	*Q1*	*Median*	*Q3*	*Max*
1,237	*0*	*0%*	*0.0320*	*0.0571*	*0.0013*	*0.0023*	*0.0088*	*0.0298*	*0.3150*

Source: Own work.

Figure 19: Distribution of annual extracted PD values



Source: Own work.

As seen in Table 18, the number of extracted PD value observations therefore is 1,237 which is in line with the number of financial ratios of selected companies presented in Variable selection chapter. As indicated with median, roughly half of PD observations fall below 1%

while three quarters almost reach 3%. Mean and standard deviation are almost irrelevant in that case due to a quite long tail present on the right, which is also seen in distribution Figure 19.

Before any comparison of annual PD values with estimated monthly public firm equivalent DD values can be made, some adjustments of time frequency component should be done first. Since less frequent PD values cannot be extended, shortening of estimated DD values would be the only possible solution. Among all possible data shortening approaches, simple average of estimated DD values within each year would be the most appropriate. The formula for calculation of average DD value for one year is

$$\overline{DD} = \frac{1}{12}\sum_{i=1}^{12} DD_i \tag{38}$$

where $\overline{DD}$ is average DD value for one year and DD_i is monthly DD estimation.

When the above presented calculation is performed, four annual average DD values for each of the 315 companies from industry C in period 2017-2020 should be obtained. The total expected observations would therefore be 1,260, but since some missing data have been noticed in initial DD estimation process the total number would be slightly lower. Descriptive statistics of annual average DD values are presented in Table 19 while their distribution can be seen in Figure 20.

Table 19: Descriptive statistics of annual average DD values

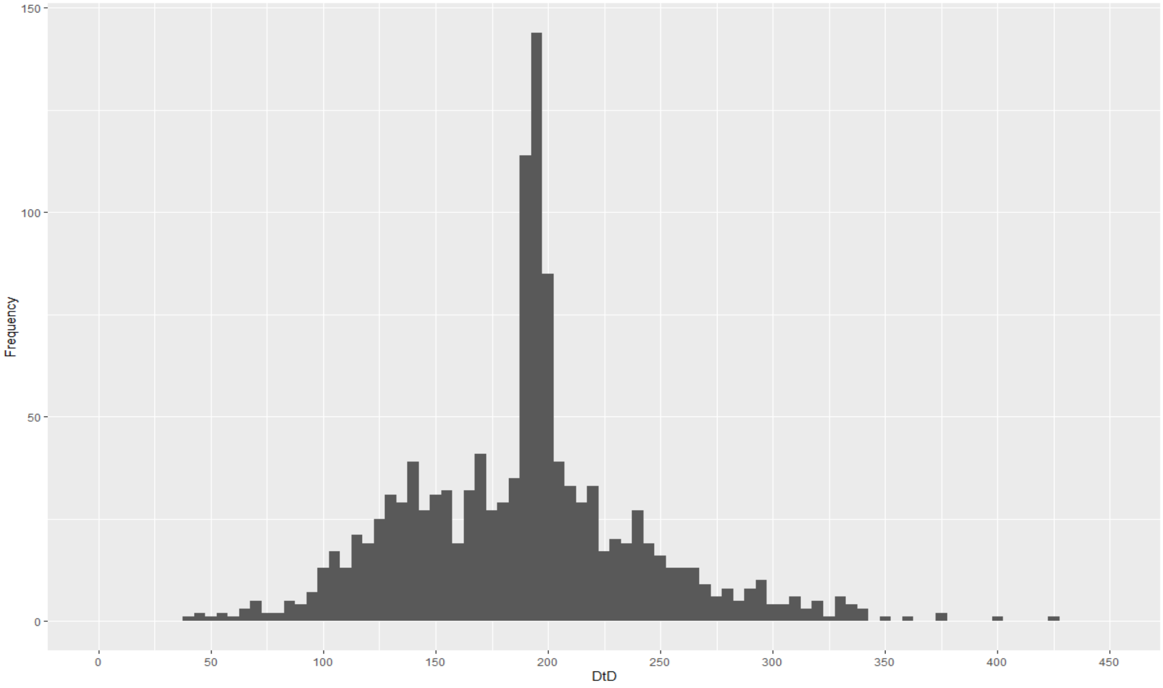
N	*N/A*	*N/A %*	*Mean*	*Std*	*Min*	*Q1*	*Median*	*Q3*	*Max*
1,237	*0*	*0%*	*189*	*52*	*40*	*154*	*192*	*212*	*426*

Source: Own work.

As seen in Table 19, the number of annual average DD values in industry C is 1,237, which is slightly below than expected as already mentioned above. Considering distribution of presented DD estimates, it can be seen that mean and median values are almost the same at around 190. The centred half of the estimates falls between 154 and 212 while standard deviation is 52. All these values suggest on slightly skewed normal distribution with values of around mean and median being extremely well represented in the dataset as seen in distribution Figure 20.

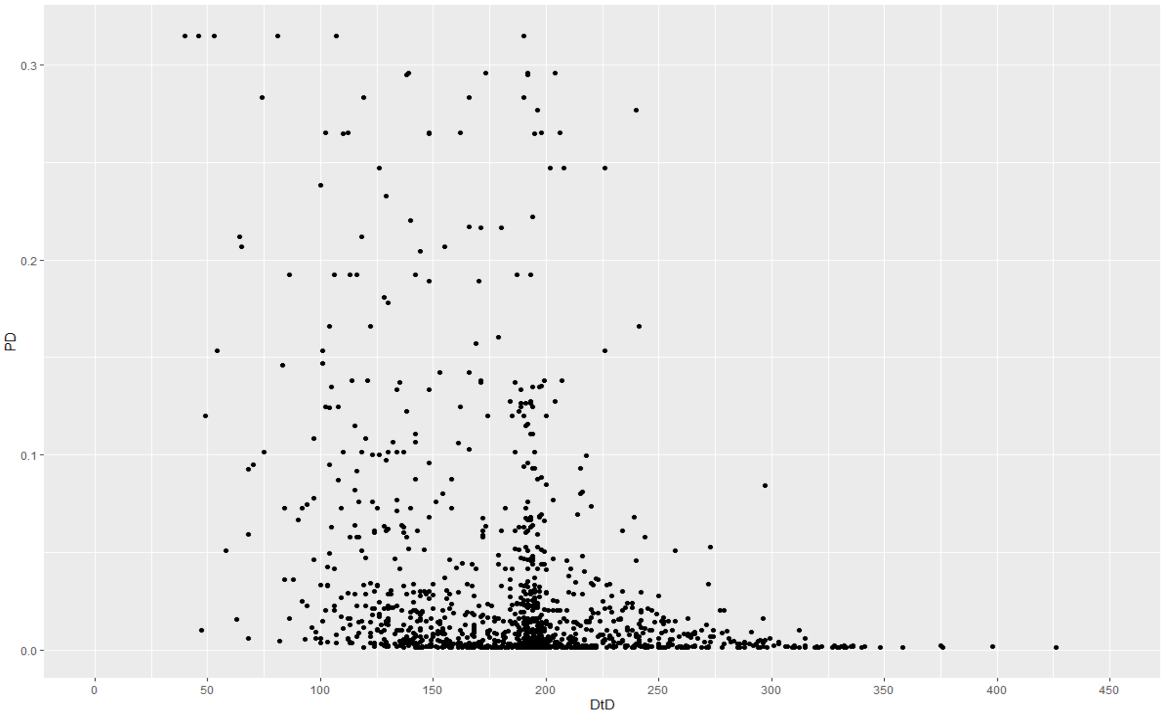
The dimension of annual average DD dataset has now become the same as the one of extracted PD values, which has also been a condition that comparison analysis can be performed. The easiest, but probably also the most appropriate way to compare both datasets in a ranking power point of view is to plot them. Simple scatter plot would give clear visual comparison of both sets and would also suggest the type of their relation. The scatter plot with combination of annual average DD values and PD values is presented in Figure 21.

Figure 20: Distribution of annual average DD values



Source: Own work.

Figure 21: Scatter plot with combination of annual extracted PD values and average DD values



Source: Own work.

As seen in Figure 21, some negative relation between the two sets of variables is noticed which was also expected based on theoretical construction of DD values. Besides negative relation, Figure 21 also indicates on non-linear relation between annual average DD values and extracted PD values. This outcome is also in line with the theory, which only rarely assumes that PD values are linearly dependent on a set of explanatory variables (Lopez, 2004). Non-linear negative relation however cannot be seen from the figure in fully unbiased manner, but the assumption of such relation will be very important in the following regression analysis chapter.

3.1.2 Forecasting accuracy

When we see that relation between average annual DD estimates and extracted PD values might be non-linear and negative, focus on forecasting power can be made. Since the aim of this thesis is to develop some kind of monitoring tool rather than classical credit scoring model, forecasting of changes in PD values instead of exact PD estimation should be tested. To test that, some further adjustments of afore presented dataset should be done first.

Since we are interested in forecasting changes in PD values, the existing dataset of PD values should be modified in a way to obtain changes through time. Regarding changes of PD values, only the trend of change would be important within the scope of this thesis rather than the exact percentage change. Therefore, the PD value increase from one year to another will be labelled as 1 in the transformed dataset, while a decrease or no change should be labelled as 0. The notation of presented data transformation is

$$\frac{PD_t}{PD_{t-1}} > 1 \quad \rightarrow \quad PD_{ch} = 1 \tag{39}$$

$$\frac{PD_t}{PD_{t-1}} \leq 1 \quad \rightarrow \quad PD_{ch} = 0 \tag{40}$$

where *t* represents years 2018-2020 in the dataset for each company.

When the above presented transformation of PD values is performed, three PD value changes for each of the 315 companies would be expected which results in 945 observations. On account of some missing values as already presented in the previous chapter, the number of observations will be slightly lower than expected. The descriptive statistics of PD value changes are presented in Table 20.

Table 20: Descriptive statistics of annual PD value changes

N	*N/A*	*N/A %*	*Mean*	*Std*	*Min*	*Q1*	*Median*	*Q3*	*Max*
936	*0*	*0%*	*0.38*	*0.48*	*0.00*	*0.00*	*0.00*	*1.00*	*1.00*

Source: Own work.

As seen in Table 20, the number of observations is 936, which is slightly lower than 945 expected. Mean is at 0.38, which means that there is 38% of observations that represents PD value increases while the rest 62% refers to stable values or their decreases. Since the number of both types of PD value changes is quite similar, the dataset can be considered as well balanced. The structure of this dataset is also similar to standard dataset of defaults, but the number of defaults (labelled as 1) is generally much smaller than the number of PD increases in above presented dataset. Since lack of defaults within datasets usually represents an issue in credit scoring model, this should not be the case during the testing procedure of this thesis.

As presented in the Ranking power chapter, scatter plot of extracted PD values and average annual DD values indicated on negative non-linear relation between the two variables. This outcome is fully in line with the theory, which only rarely assumes that PD values are linearly dependent on a set of explanatory variables. Moreover, non-linear Logistic or Cumulative Normal distribution is often proven when PD values are considered (Yeh & Lien, 2009).

According to Alkasasbeh & Raqab (2009), Logistic and Cumulative Normal distributions are continuous probability distributions which are used in several different areas. The only difference between them is that the logistic distribution has wider tails (higher kurtosis) as Cumulative Normal one. Since wider tails are usually more consistent with the underlying data and provide more insight into the likelihood of extreme values, logistic distribution will be tested also in this thesis. The logistic distribution function can be written as

$$f(x) = \frac{1}{1+e^{-z}}, \quad z = \frac{x+\mu}{\sigma} \tag{41}$$

where x is independent variable, μ is estimated mean of independent variable while σ is its standard deviation.

To prove that relation between DD values and PD value changes follow logistic distribution, logistic regression should therefore be performed. Logistic regression or Logit model is a binary statistical model which estimates probability that certain event occurred (Alkasasbeh & Raqab, 2009). The notation of the model is similar to logistic distribution function with variable z to be estimated through linear function. The logistic regression function therefore is

$$p(x) = \frac{1}{1+e^{-z}}, \quad z = \beta_0 + \beta_1 x \tag{42}$$

where β_0 is intercept and β_1 is the slope of the function.

As further explained by Alkasasbeh & Raqab (2009), the β_0 and β_1 coefficients are usually estimated through maximum likelihood method, which is calculated by the product of the likelihood function for each individual observation. The function which should be maximised in order that coefficients are estimated is

$$L = \prod_{i=1}^{N} p(x_i)^{y_i} \, (1 - p(x_i))^{1-y_i} \tag{43}$$

where all parameters have already been presented before.

With applying the above presented Logit model on DD values as independent variable (x) and PD value changes as dependent variable (y), coefficients, as presented in Table 21, are obtained.

Table 21: Summary of Logit model, applied on average DD values (x) and PD value changes (y)

	Estimate	*Standard error*	*z-value*	*p-value*
β_0	*0.2226*	*0.2230*	*0.9980*	*0.3181*
β_1	*-0.0039*	*0.0012*	*-3.3520*	*0.0008****
**** significance at 1% level, ** significance at 5% level, * significance at 10% level*				

Source: Own work.

Table 21 therefore shows coefficient estimates with their standard errors and z-tests as well as statistical p-value tests that evaluates whether DD values are helpful in explaining PD value changes. According to Beers (2022), the p-value is therefore a statistical measurement used to validate a hypothesis against observed data. It measures the probability of obtaining the observed results, assuming that null hypothesis is true. The lower the p-value, the greater the statistical significance of the observed difference.

As seen in Table 21, the estimated β_1 coefficient is therefore significantly different from zero on a 99% confidence level. The estimated coefficient is negative, which suggests that the lower the DD value, the higher the probability of PD value increase. This outcome also indicates on acceptance of the first hypothesis of the thesis that market-based model can predict defaults earlier than the accounting-based models. More specifically, lower DD values during the year would probably result in increased PD values at the end of year.

3.2 Model implication

When statistical significance of the model is proven, model implication based on a case study will be presented. Two companies will be selected for that manner and compared to overall industry average. Comparison will start with analysing financial ratios, extracted PD values and average DD estimates for each of the selected company. It should basically test the ranking power and model accuracy as developed in previous chapter.

Next, the analysis will shift from comparative analysis on an annual level towards description of a corporate monitoring process on a monthly basis. Within the monitoring process, the distribution specifics of a monthly DD values will be analysed for each of selected company and then presented as a timeseries graphically. The time series analysis should also at least partly answer to the second hypothesis of this thesis. The comparative analysis and monitoring process will be more in detail presented in subchapters below.

3.2.1 Comparative analysis

Comparative analysis will focus on comparing of some selected companies with the overall industry. Comparison will be based on all information and data obtained through the model construction process of this thesis. To get the most possible diverse outcomes, one extremely good company and one extremely bad company will be selected.

Good company will therefore represent a private firm, which can today be considered as very safe firm. Since extracted PD values as presented in the Data collection chapter show risk of each company through time, the company with the lowest risk or PD value in year 2020 will be selected for that manner. Selection of the bad company will be just the opposite. The company with the highest risk or PD value in year 2020 will be chosen as a bad company in the analysis.

With selection of both types of companies, comparison analysis with the industry can start. The first analysis will represent financial statements' review. As already retrieved from Altman in the Literature review, the analysis of only five financial ratios can be done in order that financial health of the company is determined. Financial ratios as presented by Altman are

$$X_1 = \frac{Working\ Capital}{Total\ Assets} \tag{44}$$

$$X_2 = \frac{Retained\ Earnings}{Total\ Assets} \tag{45}$$

$$X_3 = \frac{Earnings\ before\ Interest\ and\ Taxes}{Total\ Assets} \tag{46}$$

$$X_4 = \frac{Book\ Value\ of\ Equity}{Book\ Value\ of\ Total\ Liabilities} \tag{47}$$

$$X_5 = \frac{Total\ Sales}{Total\ Assets} \tag{48}$$

where higher values in all of the five ratios should result in lower risk or PD value of the firm.

Comparing obtained financial ratios with DD estimates, some positive relation would therefore be expected among them. If this relation is fulfilled, negative relation between DD estimates and extracted PD values would consequently be expected as already suggested in the Ranking power chapter. Moreover, regression analysis also proved that companies with overall low DD estimate tend to increase in their PD value. All three aspects will be presented and analysed for both types of companies in Table 22 and Table 23, respectively. Table 22 therefore shows records of financial ratios, average DD estimates and PD values of a good company while Table 23 presents all these values of a bad company. Both tables also include average industry values for a comparison.

Table 22: Financial ratios, average DD estimates and PD values of a good company

Year	*X1*	*X2*	*X3*	*X4*	*X5*	*DtD*	*PD*
2017	*0.52*	*0.61*	*0.13*	*2.91*	*0.77*	*320*	*0.20%*
2018	*0.67*	*0.66*	*0.20*	*6.54*	*1.54*	*201*	*0.25%*
2019	*0.56*	*0.76*	*0.12*	*7.83*	*0.96*	*266*	*0.25%*
2020	*0.57*	*0.75*	*0.14*	*5.76*	*1.01*	*218*	*0.13%*
Average	***0.58***	***0.70***	***0.15***	***5.76***	***1.07***	***251***	***0.21%***
Industry average	0.18	0.20	0.07	1.18	1.23	*189*	*3.20%*

Source: Own work.

Table 23: Financial ratios, average DD estimates and PD values of a bad company

Year	*X1*	*X2*	*X3*	*X4*	*X5*	*DtD*	*PD*
2017	*0.00*	*0.11*	*0.13*	*0.27*	*1.38*	*186*	*1.83%*
2018	*0.00*	*0.13*	*0.03*	*0.21*	*1.53*	*192*	*11.61%*
2019	*-0.50*	*0.16*	*0.01*	*0.00*	*1.73*	*53*	*31.50%*
2020	*-0.40*	*0.26*	*0.01*	*-0.10*	*1.46*	*46*	*31.50%*
Average	***-0.23***	***0.17***	***0.05***	***0.10***	***1.53***	***119***	***19.11%***
Industry average	*0.18*	*0.20*	*0.07*	*1.18*	*1.23*	*189*	*3.20%*

Source: Own work.

As seen in Table 22, all financial ratios had always been significantly higher than the overall industry in almost all years. The only exception was Total Sales / Total Assets ratio (X5), which is above industry average only in one year. This ratio suggests that good company generates lower sales with the same value of assets compared to total industry on average. On the other side, DD estimates of a good company had also been always above industry average. Considering these findings, some positive relation is indeed present between financial ratios and DD estimates in a case of a good company.

Focusing now on bad company presented in Table 23, we can see that the majority of financial ratios had been somewhat lower than the overall industry in all years as expected. The largest exception has also been in that case Total Sales / Total Assets ratio (X5), which was above industry average in all years. This suggests that selected bad company generates higher revenue with the same value of assets compared to total industry average. On the other side, DD estimates of a bad company were quite similar to industry average in the first two years of the analysis and highly below average in the other two. Considering these findings, some positive relation might also be present between financial ratios and DD estimates in a case of a bad company as well.

When we see that financial ratios and DD estimates might be positively related, ranking power of the latter can be checked. As presented in the good company case in Table 22, some deterioration of estimated DD value occurred in year 2018 which is fully in line with increased PD in that year. In the next two years, PD value did not increase anymore while DD estimates had been increasing as expected. The only distortion is year 2020, when DD value somewhat declined while the opposite would be expected knowing that PD value declined as well. However, considering that DD values ranked the firm correctly in three out of four years, ranking suggests working quite well in case of a good company.

In a case of a bad company presented in Table 23, slight increase of estimated DD value occurred in year 2018 which is not in line with increased PD in that year. In the next three years, DD values had been constantly decreasing which is fully in line with positive trend of PD values. Considering therefore that DD values ranked the firm correctly in three out of four years, ranking power might work well also in case of a bad company.

Lastly, with comparison of DD estimates and PD movements, another important conclusion can be made. As proven in Regression analysis chapter, companies with relatively low DD estimates tend to increase in PD value while companies with high DD estimates tend to at least remain stable. As seen in case of a good company in Table 22, PD value increased only once out of four. Just the opposite is seen in Table 23 of a bad company, when PD value only once remained stable. These findings therefore confirm that companies with relatively low DD estimates tend to increase in their PD values and vice versa.

3.2.2 Monitoring process

In this chapter, the analysis from financial ratios, average annual DD estimates and PD values towards monthly DD estimates will be shifted. Since the aim of this thesis was to develop some kind of monitoring tool of companies on a monthly basis, it would be interesting to see how monthly DD values of both types of companies moved through time.

First, the distribution of monthly DD values of both companies will be analysed and then presented in time graphically. Descriptive statistics of estimated monthly DD values of a good and a bad company, as well as of industry average for a comparison are presented in Table 24. Time series of monthly estimates of a good and a bad company are then shown in Figure 22 and Figure 23, respectively.

As seen in Table 24, 48 monthly DD values have been included into analysis in case of a good company which is fully in line with the dataset of four years of monthly observations. In a case of a bad company, only 46 monthly DD values have been included which is slightly below 48 observations expected. Two values have therefore been excluded from the analysis;

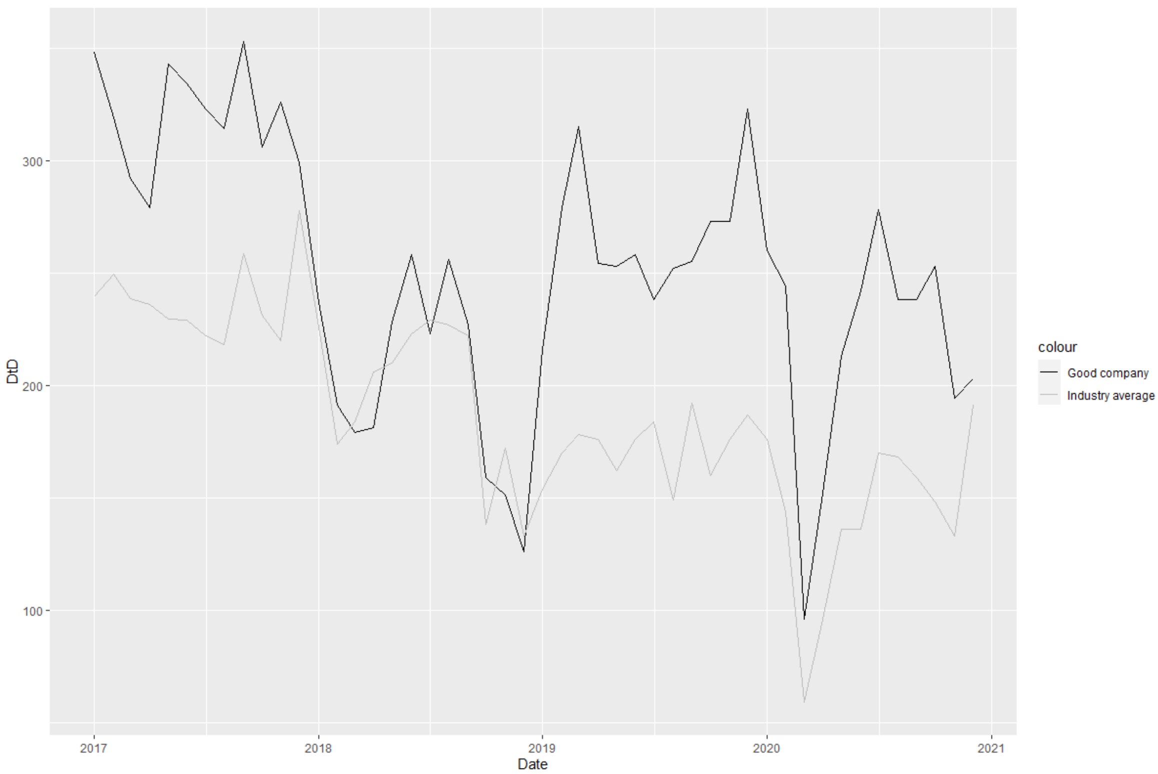
the reason is in estimation of negative values. In contrast, 14,804 observations were included within industry average, while additional 40 observations have been omitted due to estimation of negative values.

Table 24: Descriptive statistics of monthly DD values of a good and a bad company

Good company									
N	*N/A*	*N/A %*	*Mean*	*Std*	*Min*	*Q1*	*Median*	*Q3*	*Max*
48	*0*	*0%*	*251*	*60*	*96*	*221*	*254*	*294*	*353*
Industry average									
N	*N/A*	*N/A %*	*Mean*	*Std*	*Min*	*Q1*	*Median*	*Q3*	*Max*
14,804	*40*	*0%*	*189*	*60*	*1*	*148*	*190*	*226*	*504*
Bad company									
N	*N/A*	*N/A %*	*Mean*	*Std*	*Min*	*Q1*	*Median*	*Q3*	*Max*
46	*2*	*4%*	*122*	*81*	*1*	*54*	*140*	*197*	*254*

Source: Own work.

Figure 22: Time series of DD estimates of a good company



Source: Own work.

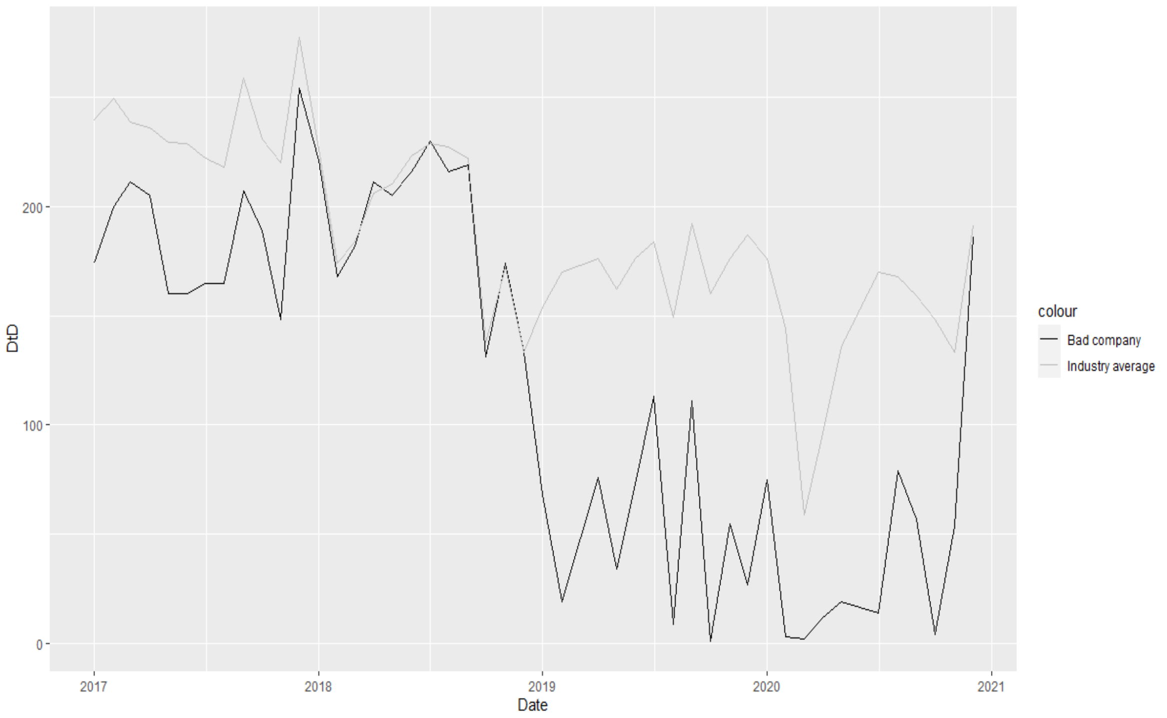
Comparing next mean and standard deviation of good and bad company with industry average, some difference among them is noticed in Table 24 as expected. In a case of a good company, standard deviation of estimated DD values is exactly the same as of industry average while DD values of bad company had been more volatile. Comparing mean values, good company has somewhat higher mean that the industry while mean of a bad company is lower than industry as expected considering selection process in the previous chapter.

Considering the distribution of DD values, we can see that the distribution of good company as well as overall industry are quite normally distributed with median values being very close to mean values. The only difference is again that the distribution of good company is somewhat shifted to the right and that tails of industry average are wider, which is clearly expected. On the other hand, distribution of a bad company might be slightly skewed to the left since median is somewhat higher than the mean.

When looking into time series of DD estimates of a good company in Figure 22, we can see that three quite severe drops occurred during the analysed period. At the first larger drop in the beginning of year 2018, estimated DD values came near to average industry DD values and followed very similar path until the second larger drop at the end of the same year. Afterwards, good company's DD values again separated from industry average and became closer only in third larger drop in the beginning of year 2020.

The timeseries analysis therefore suggests that good company was always safer compared to overall industry average. The only exception was year 2018, when the company came very close to the average risk of the industry. Knowing that extracted PD value increased solely in this year as seen in the previous chapter, developed monitoring model seems to be quite accurate.

Figure 23: Time series of DD estimates of a bad company



Source: Own work.

Analysing time series of a bad company in Figure 23, we can see that DD values had almost always been below industry average. The exception was similar to good company only year 2018, when DD values of a bad company and industry average followed very similar path. After that year, a quite severe drop of bad company's DD values occurred with some strong volatility recorded until the end of analysing period. Right at the end, DD value however again significantly improved and again reached industry average.

The timeseries analysis therefore shows that bad company was only slightly riskier than the industry in the first two years and significantly riskier in the rest of the analysed period. As seen in the previous chapter, the PD values of a bad company have been constantly increasing. The highest incline is recorded right after the first two years, which also suggest that developed monitoring tool might be quite accurate.

Since timeseries analysis of estimated DD values based on good and bad company provides quite accurate monthly movements of their riskiness, the model developed in this thesis could be used as an efficient corporate monitoring tool. This therefore also suggests acceptance of the second hypothesis of this thesis. However, the hypothesis can only be partly accepted since only industry C has been eventually considered in model development. With inclusion of other industries into the model, monitoring tool should yet significantly improve with adding additional dimension of overall average values into the analysis.

CONCLUSION

Proper and efficient monitoring of corporate portfolio is challenging. The monitoring tool developed in this thesis shows innovative and rarely used approach to assess credit profile of a specific company. It is based on well-known Merton's Distance to default model, which mainly uses timely market data in its prediction. Since the original Merton's model is applicable to public firms only, private firm extension had to be done in model development process. Private extension has been done through finding similar characteristics of private and public firms, monthly probability of default values for any private or public company can therefore be obtained.

Since the developed monitoring tool is based on a market-based model, it was expected that it would be much faster in predicting distress compared to any other accounting-based model. This also represented the first hypothesis of the thesis which was accepted after statistical tests were performed. More specifically, logistic regression proved that estimated DD values and extracted benchmark PD values are negatively related at a quite high significance level. This suggested that the lower the DD value of a specific company during some year is, the higher the probability that PD value will increase at the end of that year. Market-based model can therefore indeed predict default earlier than the selected accounting-based one.

The second hypothesis of the thesis suggested that the developed model could be used in a financial industry as an efficient monitoring tool. Since the accounting data are in majority released annually and may already not be fully relevant shortly after announcement, the market-based model as developed in this thesis should therefore be more appropriate for frequent monitoring needs. This was tested with case study analysis, that considered two different types of companies. The conclusion of the analysis was that selected companies can indeed be efficiently monitored with developed model, since their estimated monthly DD values show a quite accurate movements of their true riskiness. Stated hypothesis can however be only partly accepted, since very important industry-specific risk could not have been tested due to lack of data. More specifically, monitoring process of companies from industry C could have been performed only while inclusion of some additional industry would also introduce industry-specific factors.

Suggestion for further work would therefore be to significantly enlarge initial dataset of public firms to obtain DD estimates for companies from several industries eventually. In this thesis, the dataset of German public firms was only used, but extension to other EU countries might also be considered to obtain sufficiently large dataset. The other suggestion would also be to test estimated DD values with comparison to real default indicators. In this thesis, the comparison focused on an already built credit scoring model, while the construction of own database of defaulted companies could result in direct probability of default estimate. With this approach, forecasting power of the model can be even more accurate and could potentially be also calibrated with several calibration technics.

Although the initial dataset was too small to obtain proper and robust results, some basic overall ideas can still be extracted from this thesis. The model was developed to serve as some kind of corporate monitoring tool, which would help to detect deterioration of any firm. The model is not meant to be a substitute for currently widely used accounting-based credit scoring models, but it should be complementary. Accounting-based models usually provide annual credit scores while the model developed in this thesis can improve frequency to monthly or even daily level. The outcome of the model is distance to default value, which nominally does not hold any explanatory power, but it can be used to rank companies quite well regarding their riskiness.

Putting the model in context of current market situation, companies that are very sensitive to energy prices should have probably recorded severe drops in estimated distance to default values in last months. Since accounting-based credit scores are currently still modelled based on year 2021 financial statements, the newly developed model would suggest adjusting them accordingly. Similar was also during recent COVID pandemic period, when some industries were much more negatively affected than the others. During that period, the model could have also exposed affected companies in accordance with market movements and suggest credit scores' adjustments. The newly developed model can in practice therefore be very useful in monitoring of companies, especially in case of larger economic shocks.

REFERENCE LIST

1. AJPES. (2022). *O AJPES - Predstavitev*. Retrieved 18 February 2022 from AJPES: https://www.ajpes.si/O_AJPES/Predstavitev/Vizija,_poslanstvo_in_strateski_cilji
2. Alkasasbeh, M. R. & Raqab, M. Z. (2009). Estimation of the generalized logistic distribution parameters: Comparative study. *Statistical Methodology, 6(3)*, 262–279.
3. Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K. & Suvas, A. (2014). Distressed firm and bankruptcy prediction in an international context: A review and empirical analysis of Altman's Z-score model. *Available at SSRN 2536340*.
4. Beers, B. (2022, May 18). *P-Value*. Retrieved on 20 June 2022 from Investopedia: https://www.investopedia.com/terms/p/p-value.asp#toc-what-is-p-value.
5. Benos, A. & Papanastasopoulos, G. (2007). Extending the Merton Model: A hybrid approach to assessing credit quality. *Mathematical and Computer Modelling, 46(1–2)*, 47–68.
6. Bharath, S. T. & Shumway, T. (2004). Forecasting default with the KMV-Merton model. *AFA 2006 Boston Meetings Paper*.
7. Chen, J. (2022a, April 2). *Primary Market*. Retrieved on 13 April 2022 from Investopedia: https://www.investopedia.com/terms/p/primarymarket.asp
8. Chen, J. (2022b, March 3). *Shares Outstanding*. Retrieved on 15 April 2022 from Investopedia: https://www.investopedia.com/terms/o/outstandingshares.asp
9. Chouhan, V., Chandra, B. & Goswami, S. (2014). Predicting financial stability of select BSE companies revisiting Altman Z score. *International Letters of Social and Humanistic Sciences, 15(2)*, 91–105.
10. Connects. (2022, April 7). *NACE Codes: What Are They and Why Do They Always Matter?* Retrieved on 15 May 2022 from Connects: https://connects.world/nace-codes/
11. Corporate Finance Institute. (2020, September 25). *Domicile*. Retrieved on 15 April 2022 from Corporate Finance Institute: https://corporatefinanceinstitute.com/resources/knowledge/other/domicile/
12. Duan, J. C., Kim, B., Kim, W. & Shin, D. (2018). Default probabilities of privately held firms. *Journal of Banking & Finance, 94,* 235–250.
13. Fernando, J. (2022b, August 1). *Retained Earnings*. Retrieved on 5 August 2022 from Investopedia: https://www.investopedia.com/terms/r/retainedearnings.asp
14. Fernando, J. (2022a, June 29). *What Is Working Capital?* Retrieved on 23 July 2022 from Investopedia: https://www.investopedia.com/terms/w/workingcapital.asp#toc-what-is-working-capital.
15. Ganti, A. (2020b, December 28). *Adjusted Closing Price*. Retrieved on 18 April 2022 from Investopedia: https://www.investopedia.com/terms/a/adjusted_closing_price.asp.
16. Ganti, A. (2020a, October 31). *Short-Term Debt*. Retrieved on 15 April 2022 from Investopedia: https://www.investopedia.com/terms/s/shorttermdebt.asp
17. Hayes, A. (2022, February 20). *Stock Symbol (Ticker)*. Retrieved on 6 March 2022 from Investopedia: https://www.investopedia.com/terms/s/stocksymbol.asp

18. Jessen, C. & Lando, D. (2015). Robustness of distance-to-default. *Journal of Banking & Finance*, *50*, 493–505.
19. Kelly, R. (2020, November 27). *Bloomberg Terminal*. Retrieved on 3 March 2022 from Investopedia: https://www.investopedia.com/articles/professionaleducation/11/bloomberg-terminal.asp
20. Korchak, J. (2019, February 18). *What is a company registration number?* Retrieved on 26 March 2022 from Inform Direct: https://www.informdirect.co.uk/company-records/company-registration-number-crn-what-is-it/
21. Kramer, L. (2017, October 30). *The history of MarketWatch: How a sports data startup became a half-billion-dollar financial news site*. Retrieved on 22 April 2022 from MarketWatch: https://www.marketwatch.com/story/the-history-of-marketwatch-how-a-sports-data-startup-became-a-half-billion-dollar-financial-news-site-2017-10-30
22. Liberto, D. (2020, November 23). *Total Liabilities*. Retrieved on 17 April 2022 from Investopedia: https://www.investopedia.com/terms/t/total-liabilities.asp.
23. Lopez, J. A. (2004). The empirical relationship between average asset correlation, firm probability of default, and asset size. *Journal of Financial Intermediation, 13(2)*, 265–283.
24. Merritt, C. (2019, March 8). *What Size Company Is Considered a Mid-Size Company?* Retrieved on 16 March 2022 from CHRON: https://smallbusiness.chron.com/size-company-considered-midsize-company-71776.html
25. Murphy, C. (2022, July 5). *Earnings Before Interest and Taxes (EBIT)*. Retrieved on 27 July 2022 from Investopedia: https://www.investopedia.com/terms/e/ebit.asp#toc-what-is-earnings-before-interest-and-taxes-ebit.
26. Paolucci, R. (2020, July 2). *Geometric Brownian Motion*. Retrieved on 5 February 2022 from Towards Data Science: https://towardsdatascience.com/geometric-brownian-motion-559e25382a55
27. Perani, G. & Valeria, C. (2015). Matching industry classifications. A method for converting Nace Rev. 2 to Nace Rev. 1. *(No. 1502).*
28. Seth, S. (2021, January 17). *Book Value vs. Market Value: What's the Difference?* Retrieved on 21 June 2022 from Investopedia: https://www.investopedia.com/articles/investing/110613/market-value-versus-book-value.asp
29. Shumway, T. & Bharath, T. S. (2008). Forecasting Default with the Merton Distance to Default Model. *The Review of Financial Studies*, *21(3)*, 1339–1369.
30. Somaiya, R. (2015, Februar 18). *Ex-Fortune Editor Joins Yahoo Finance*. Retrieved on 26 March 2022 from The New York Times: https://www.nytimes.com/2015/02/19/business/ex-fortune-editor-joins-yahoo-finance.html
31. Sun, J., Li, H., Huang, Q. H. & He, K. Y. (2014). Redicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches. *Knowledge-Based Systems*, *57*, 41–56.

32. Sundaresan, S. (2013). A Review of Merton's Model of the Firm's Capital Structure with Its Wide Applications. *Annual Review of Financial Economics, 5*, 21–41.
33. Tanthanongsakkun, S. & Treepongkaruna, S. (2008). Explaining Credit Ratings of Australian Companies—An Application of the Merton Model. *Australian Journal of Management, 33(2)*, 261–275.
34. Tuovila, A. (2021, March 4). *Long-Term Debt*. Retrieved on 18 April 2022 from Investopedia: https://www.investopedia.com/terms/l/longtermdebt.asp
35. Vassalou, M. & Xing, Y. (2004). Default risk in equity returns. *The journal of finance, 59(2)*, 831–868.
36. West, D. (2000). Neural network credit scoring models. *Computers & Operations Research, 27(11–12),* 1131–1152.
37. Yeh, C. C., Lin, F. & Hsu, C. Y. (2012). A hybrid KMV model, random forests and rough set theory approach for credit rating. *Knowledge-Based Systems, 33*, 166–172.
38. Yeh, I. C. & Lien, C. H. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert systems with applications, 36(2)*, 2473–2480.
39. Zvika, A., Ohad, A. & Koresh, G. (2016). Using Merton model for default prediction: An empirical assessment of selected alternatives. *Journal of Empirical Finance, 35*, 43–67.

APPENDICES

Appendix 1: Slovenian summary of Master's thesis.

Pravilna in učinkovita spremljava portfelja podjetij je izziv. V zadnjih nekaj desetletjih je bilo razvitih veliko različnih modelov ocenjevanja kreditnega tveganja, ki temeljijo predvsem na računovodskih podatkih podjetij. Računovodski podatki praviloma prikazujejo preteklo dogajanje v podjetju, kar pa ni povsem skladno z idejo o napovedovanju prihodnosti. Nekateri avtorji tako predlagajo, da bi za napovedovanje prihodnje uspešnosti podjetij morali v modelu biti uporabljeni predvsem tekoči tržni podatki.

Eden najbolj znanih in uporabljenih modelov, ki pri svoji napovedi uporablja tekoče tržne podatke je Mertonov model napovedovanja verjetnosti neplačila. Mertonov model v svoji napovedi poskuša oceniti tržno vrednost sredstev podjetij, in sicer z uporabo vrednosti kapitala, dolga in gibanja delnic. A ker model pri napovedovanju potrebuje podatke o delnicah, je uporaben le za podjetja, ki kotirajo na borzi. Model je po napovedi nekaterih avtorjev mogoče razširiti tudi na zasebna podjetja, in sicer z iskanjem podobnih značilnosti le teh z naborom izbranih kotirajočih podjetij.

V želji, da bi se preverila uporabnost modela za napovedovanje verjetnosti neplačila slovenskih zasebnih podjetij, je bil Mertonov model z zgoraj omenjeno razširitvijo razvit v tem magistrskem delu. Ker je slovenski delniški trg relativno majhen in ima razmeroma nizek obseg trgovanja, so bila za oceno parametrov originalnega Mertonovega modela uporabljena kotirajoča nemška podjetja. Razlog za izbiro nemških podjetij je v povezanosti in odvisnosti slovenskega gospodarstva od nemškega.

Model je bil torej zgrajen na vzorcu nemških kotirajočih podjetij in nato uporabljen za napovedovanje verjetnosti neplačila slovenskih zasebnih podjetij. Prvotna ideja je bila, da bi model predstavljal nekakšno orodje za spremljavo podjetij, s katerim bi lahko pravilno in pravočasno ocenili spremembe v kreditnem profilu posameznega podjetja. Pojma pravočasnosti zaznave in splošne uporabnosti orodja sta predstavljali tudi hipotezi magistrske naloge, ki sta bili na koncu testirani in ovrednoteni.

Pravočasnost zaznave je bilo testirano s primerjavo ocenjenih vrednosti verjetnosti neplačila razvitega modela tekom leta z nekaterim drugim že razvitim in uveljavljenim modelom na koncu istega leta. Test je bil opravljen s pomočjo logistične regresije, ki je dokazala statistično značilno negativno razmerje med omenjenima spremenljivkama. Prva hipoteza je tako bila sprejeta, razvit model pa je dokazano sposoben pravočasno zaznati poslabšanje verjetnosti neplačila nekega podjetja.

Druga hipoteza, ki se nanaša na splošno uporabnost orodja pa je bila testirana s pomočjo študije primera dveh popolnoma različnih tipov podjetij. Zaključek analize nakazuje, da je izbrana podjetja z razvitim modelom res mogoče učinkovito spremljati, saj njihove ocenjene mesečne vrednosti verjetnosti neplačila kažejo precej natančno gibanje njihove resnične tveganosti. Navedeno hipotezo pa je bilo mogoče le delno sprejeti, saj zaradi pomanjkanja podatkov ni bilo mogoče testirati zelo pomembnega panožnega tveganja. Natančneje, tekom razvoja modela je bilo možno pridobiti le ocene verjetnosti neplačil za podjetja iz panoge C, medtem ko bi z vključitvijo vsaj še ene dodatne panoge bilo možno primerjati tudi panožno specifične dejavnike.