VŠB TECHNICKÁ UNIVERZITA OSTRAVA | EKONOMICKÁ FAKULTA | SERIES ON ADVANCED ECONOMIC ISSUES

MICRO-MODELLING APPROACHES FOR CREDIT RATING AND CORPORATE SURVIVAL

Martina Novotná

Series on Advanced Economic Issues
Faculty of Economics, VŠB-TUO

Martina Novotná

MICRO-MODELLING APPROACHES FOR CREDIT RATING AND CORPORATE SURVIVAL

Ostrava, 2024

Martina Novotná
Department of Finance
Faculty of Economics
VŠB-Technical University Ostrava
17. Listopadu2172/15
708 00 Ostrava-Poruba, CZ
martina.novotna@vsb.cz

Reviews

Jiří Witzany, Prague University of Economics and Business

Lumír Kulhánek, VSB – Technical University of Ostrava

This publication is the output of research activity by the research team of the project No. SP 2019/132.

The text should be cited as follows: Novotná, M. (2024). *Micro-Modelling Approaches for Credit Rating and Corporate Survival*, SAEI, vol. 69. Ostrava: VSB-TUO.

ISBN 978-80-248-4728-3 (print)
ISBN 978-80-248-4729-0 (on-line)
DOI 10.31490/9788024847290

Preface

The essential issue of this book is the term credit, either in the context of credit markets, credit risk or credit rating. Credit markets' existence is associated with credit risk, which refers to the risk of an economic loss from the failure of a counterparty to meet its contractual obligations. Due to credit risk, suppliers of credit need to assess the creditworthiness of prospective borrowers. Although modern approaches to credit risk analysis have been developed in recent decades, examining borrowers' ability to repay their funds is one of the oldest lending activities.

The main goal of this monograph is to apply and verify certain methods of credit risk modelling to real data from selected CEE countries. For the main purpose of this work, a micro approach is used to measure credit risk based on monitoring basic indicators and allowing creditors to take the necessary actions in time. The book aims at two partial financial and methodological objectives related to credit risk modelling in this context. Both of them are interconnected, and they complement each other throughout the book.

In terms of financial application, this book's principal objective is to analyse credit risk based on real data, assess its main factors, explore mutual relations, and draw conclusions related to risk assessment and market behaviour. The application is focused on two approaches of individual credit risk assessment within the areas of credit rating and corporate survival.

The methodological purpose is the application and verification of rating and bankruptcy models. Rating models estimated using conventional approaches such as discriminant analysis or logistic regression are supplemented by an alternative survival analysis approach to determine the probability of rating downgrade over time. Survival analysis is subsequently used in the following empirical studies on corporate bankruptcy. We will investigate the relationship between the rating and corporate bankruptcy rates and estimate the rating assessment depending on the used model, input variables and the company's age.

This book is intended for everyone interested in credit risk, particularly rating and corporate survival modelling, mainly for academia and students at all levels of study. This monograph aims to provide complex information on credit risk fundamentals, current trends and rating systems' principles. However, the primary purpose is the practical application and estimating models using real corporate data. Thus, we can determine the main factors of rating assessment and corporate survival and demonstrate how these models can be developed through different statistical methods.

The text is structured into three central parts: The theoretical background on credit risk and the credit rating industry, a description of econometric approaches used in the applications, and empirical studies on credit rating and corporate bankruptcy modelling. If the reader is particularly interested in estimating models and their comparison and interpretation, then it is suggested that they go directly to the practical application. However, reading the book step by step is recommended to understand the essence and main principles and use them in the application.

Martina Novotná, Ostrava, April 2024

Brief Contents

Contents

List of Abbreviations

ABS	Asset Backed Security
BIS	Bank for International Settlements
CEE	Central and Eastern Europe
CDO	Collaterized Debt Obligation
CDS	Credit Default Swap
CHAID	Chi-squared Automatic Interaction Detection
CLN	Credit Linked Note
CBR	Cumulative Bankruptcy Rate
CDR	Cumulative Default Rate
CRA	Credit Rating Agency
DDA	Descriptive Discriminant Analysis
EAD	Exposure to Default
EBA	European Banking Authority
EC	European Commission
EDF	Expected Default Frequency
ESMA	European Securities and Markets Authority
EU	European Union
FDIC	Federal Deposit Insurance Corporation
FICO	Fair Isaac Corporation
GDP	Gross Domestic Product
HHI	Herfindahl-Herfindahl Index
HR	Hazard Ratio
IRB	Internal Ratings-Based Approach
ISDA	International Swaps and Derivatives Association
KMV	Kealhofer, McQuown, Vasicek
LDA	Linear Discriminant Analysis
LGD	Loss Given Default
LTV	Loan-to-Value Ratio
M	Maturity of Exposures

MBS	Mortgage Backed Security
MLR	Multinomial Logistic Regression
OLR	Ordinal Logistic Regression
PAYG	Pay-As-You-Go
PCC	Percentage Correctly Classified
PD	Probability of Default
PDA	Predictive Discriminant Analysis
PH	Proportional Hazards
PTI	Payment-to-Income Ratio
ROC	Receiver Operating Characteristic
RR	Recovery Rate
S&P	Standard&Poor's
SEC	Securities and Exchange Commission
SWOT	Strengths, Weaknesses, Opportunities, Threats

Chapter 1

Introduction

Credit and credit markets are associated with credit risk, which refers to the risk of an economic loss from the failure of a counterparty to meet its contractual obligations. Due to credit risk, credit suppliers need to assess prospective borrowers' creditworthiness. Credit rating agencies developed rating systems extensively used for risk monitoring and management by financial institutions, governments, investors and other market participants. As their ratings are often incorporated in financial institutions’ regulation in many countries, they have an essential role in the financial markets. Thus, it is vital to understand the credit rating principles, the process and factors of rating assessment and current trends in this industry.

In the current research, many studies use various techniques to predict ratings or corporate bankruptcy. However, there is still less attention to modelling the individual credit risk using time-to-event methods. This fact is one of the motivations of our research, which is to use and compare conventional approaches with less frequently applied survival analysis methods. The main contribution of this monograph is the expansion of existing research in this area and the application of selected models to specific data from CEE countries, respectively, from the Czech Republic. The primary purpose is to find a link between rating and survival models, identify the main predictive variables and propose a procedure to convert bankruptcy rates into rating assessment. As a result, the association between the probability of survival and the rating assessment over time can be better understood. Furthermore, as the rating is widespread and used globally, we believe the interpretation of credit risk using the rating is more suitable for users, especially individual investors.

The main objective of this monograph is to apply and verify certain methods of credit risk modelling to real data of selected CEE countries. For the main purpose of this work, a micro approach is used to measure credit risk based on monitoring basic indicators and allowing creditors to take the necessary actions in time. The book aims at two partial financial and methodological objectives related to credit risk modelling in this context. Both of them are interconnected, and they complement each other throughout the book.

From the perspective of applied finance and financial markets, this monograph aims to analyse credit risk based on real data, assess main factors, explore mutual relations, and draw conclusions related to credit rating assessment and market behaviour. Furthermore, attention is paid to regional markets of CEE countries and the narrower market within the Czech Republic. This way, the current credit market's overall characteristics, the main factors of credit rating, and corporate survival can be more generally assessed.

The methodological purpose of this work is the application and verification of rating and bankruptcy models. Rating models estimated using conventional approaches such as discriminant analysis or logistic regression are supplemented by an alternative survival analysis approach to determine the probability of rating downgrade over time. Survival analysis is subsequently used in the following empirical studies on corporate bankruptcy. We will investigate the relationship between the rating and corporate bankruptcy rates and estimate the rating assessment depending on the used model, input variables and the company's age.

All participants in credit contracts can use the main findings of this work. Nevertheless, we see the main use on the part of retail investors, whether individuals or companies, who can use the partial results of the work in several directions, mainly for a better understanding of the factors that significantly influence the survival probability and, thus, the overall rating evaluation. Furthermore, the results of this work can be further used to apply selected models to their data and subsequent use to measure credit risk. Finally, this work can also be used in academic research as an example of connecting two micro approaches to estimating credit risk models.

Consistent with the primary objective, this monograph is structured into three central parts: (i) the theoretical background on credit risk and the credit rating industry, (ii) the literature review and a description of econometric approaches used in the applications, (iii) four empirical studies on credit rating and corporate bankruptcy modelling.

The first chapter is devoted to the theoretical introduction to credit risk and the primary purpose of this monograph. The essentials of credit risk and rating assessment are described in *Chapter 2*, where attention is paid to credit risk factors, the procedure of credit rating assessment and the role of credit rating agencies. Agencies such as Moody's, Standard & Poor's or Fitch issue ratings and play an important role in global financial markets. However, due to the problems of misleading some ratings of asset-backed securities during the subprime mortgage crisis of 2007-2008, they came under intense criticism. For their practices leading to loss of credibility, many countries' efforts have been made to strengthen their regulation. Hence, current issues in the credit rating industry and central areas of regulation in the European Union are also mentioned in this section. Nevertheless, despite the rating industry's recent problems, rating assessment remains a widely accepted evaluation of credit quality. Thus, it will be used as a credit risk measurement in the application studies. Furthermore, since a particular part of this application is focused on estimating rating models, their use and interpretation, it is essential to understand how rating agencies provide rating assessments. Thus, the main principles of rating systems and the credit rating process are described in the second chapter.

Various regulations stimulate the formal quantification of credit risk and the use of credit portfolio models in financial institutions. Thus, many issues must be considered when selecting the appropriate approach for credit risk modelling. For example, suppose the credit risk analyst evaluates credit risk as a discrete event and concentrates merely on a potential default event. In that case, the fundamental-based models provide a suitable way of assessing credit risk. On the other hand, structural and different quantitative approaches should be applied if the modeller analyses the dynamics of the debt value and the associated credit spread over the whole time interval to maturity. Throughout this book, and especially the application part, credit risk is considered a discrete event, such as a potential default or bankruptcy event represented by rating grade or the probability of survival. In such cases, the main task is to assess the credit risk of a particular issue or issuer, typically through credit risk models developed to discriminate between lower and higher credit risk. Such models are usually based on the statistical analysis of past characteristics of debtors or issuers, mainly quantitative variables such as corporate financial ratios. Since these models are focused on evaluating individual subjects, mostly based on data from financial statements, they are also referred to as micro models or fundamental-based models. In the application part, attention is paid to the procedure and development of such models and their use and interpretation.

The econometric approaches selected based on the recent studies and used in the application part are described in *Chapter 3*. Firstly, we provide some research review, a summary of approaches used in the chronological context, and the main findings of selected studies in this section. Following, we will find the possible extension of the current research and emphasize the contribution of this work in the context of the application part, which is focused on CEE countries and the use of survival analysis. Next, the selected methods used in the application part are described. First, discriminant and logistic regression analyses are described and used in Chapter 4 to estimate rating models. Then, the principles of survival analysis are explained, and selected survival models are described, focusing on the Kaplan-Meier estimates, the Cox proportional hazards and the Weibull model. These approaches are then used in Chapters 5, 6 and 7.

The first application study in *Chapter 4* is focused on credit rating modelling. Firstly, selected methods are used to estimate rating models based on data from CEE countries. Then, based on the main findings, we determine the main accounting-based variables of the credit rating of non-financial companies from selected CEE countries representing economies with a shorter history and tradition of capital markets. The analysis is based on a corporate rating evaluation known as MORE Rating. This study's methodological objective is to compare models developed by different approaches, such as discriminant and logistic regression analysis and suggest a more suitable method for rating modelling. Finally, this part is followed by the modelling of rating downgrade employing survival analysis methods. Therefore, we can compare the main findings and identify the key influential variables on rating assessment and the hazard of rating deterioration based on two different approaches.

The aim of *Chapter 5* is to assess the relationship between the rating and corporate bankruptcy rates. This study is based exclusively on data from Czech companies and thus complements the main findings from CEE countries. In this section, we compare published default rates with estimated bankruptcy rates. Based on the comparison, we propose the procedure for rating estimation using bankruptcy rates and average spreads. Next, the Cox proportional hazard and the Weibull model are used in *Chapter 6* to assess the impact of industry, legal form and company size on the survival probability, followed by evaluating the influence of financial variables on the survival probability in *Chapter 7*. Both studies identify the effect of selected variables on corporate survival, whether categorical or quantitative. In addition, cumulative bankruptcy rates are estimated using the models and converted to a rating assessment. A significant advantage of this approach is using the time variable in survival models, which allows us to determine the rating not only depending on financial performance or other characteristics but also on the company's age. Thus, this procedure represents a dynamic approach to modelling and predicting individual ratings.

Finally, the main findings and overall suggestions are summarized in the conclusion in *Chapter 8*.

This book used two statistical software packages for data analysis: IBM SPSS Software and Stata statistics. All models are estimated based on unique corporate data from the MORE Rating and corporate data on Czech companies from the Magnusweb database.

2024 Martina Novotná

Chapter 2

The Essentials of Credit Rating Assessment

The crucial issue in this monograph is the term credit, which refers to credit markets, credit risk, or credit rating. This chapter provides an introduction to credit and credit markets critical role; however, the primary attention will be paid to the explanation of credit risk, its measurement and analysis.

This monograph is focused on two areas of application: credit rating and corporate survival modelling. Both topics are associated with credit risk; however, the assessment uses different methods and data and is conducted from different perspectives. Nevertheless, both approaches are used to obtain more information about corporate credit risk and are not mutually exclusive. On the contrary, it is appropriate to use both ways to analyse the credit risk and its dynamic in certain cases.

This section aims to provide an introduction and a literature review of credit rating and corporate survival modelling. First, the emphasis will be on the purpose and goals of modelling, some recent research, and a summary of the approaches used. Then, we gradually focus on the theoretical background and the current state of credit rating modelling. Finally, attention will be paid to the overview of the corporate survival problem.

The structure of the monograph's remaining part corresponds with the particular objectives of this work, as they are mentioned in the introduction. First, a literature review will provide the principles and purpose of rating and survival models (Chapter 2). Then, the statistical methods used in the application will be described (Chapter 3). The main part is the application, in which the estimated models, procedures and main results will be presented (Chapters 4–7). Finally, the main findings of the rating and bankruptcy analysis will be compared and used to draw this monograph's main conclusions and recommendations.

2.1 Role of Credit Markets

Credit markets are markets for credit that can be described as transactions between the creditor (the lender) and the debtor (the borrower). The creditor supplies money or non-monetary assets such as goods, services or securities to the debtor in return for a promise of future payment, which typically includes the amount of interest (Joseph, 2013). The creditors generally have no right of ownership, and the interest represents compensation for undertaken risk.

The economic role of credit lies in the fact that borrowers with insufficient resources can get funds from lenders, usually through financial intermediaries. Thus, when used effectively, credit enables the economic growth of borrowers, increasing household consumption and business investment. Credit is used by businesses, individuals, and governments, and typically, it leads to economic growth (Joseph, 2013). In financial markets, credit is supplied by financial institutions such as commercial banks in loans. It can be provided by investors who purchase bonds issued by deficit units. Debt holders are known as creditors or lenders, and they typically grant loans or hold bonds. Conversely, the use of credit by deficit units or borrowers can be considered a primary source of debt financing. While bonds are typically traded, loans are not assumed to be tradable in debt markets (De Servigny and Renault, 2004).

The proportion of loans and bonds on the total amount of debt financing can differ in various countries, depending on tradition, legal environment (e.g., property rights system), macroeconomic conditions or the development of capital markets. The proportion of these two ways of financing in the Czech Republic can be seen in Figure 2-1. In this graph, the total loans include short-term (less than one year), medium-term (1 – 5 years) and long-term loans (more than five years) to clients provided in CZK, and total bonds consist of all bonds issued in CZK, including government and corporate bonds. As can be seen, the amount of loans exceeds the number of bonds during the whole period. From 2012 to 2019, we can see relatively stable development, with the share of loans being around 56% on total debt financing and the proportion of bonds moving about 44%. However, since 2019, we have seen that the percentage share of bonds has risen, mainly due to increased government bond issuance during the COVID-19 pandemic.

While the ratio of both types of financing remains relatively stable, except in the post-pandemic period, annual changes are somewhat volatile. Bond issuance rose by 21.2% from 2011 to 2012, mainly due to an increase in corporate bonds by 40.4%. The decrease in the bonds issued in 2017 is primarily due to the decline in government bond issuance, with the opposite trend since 2019 (see Figure 2-2).

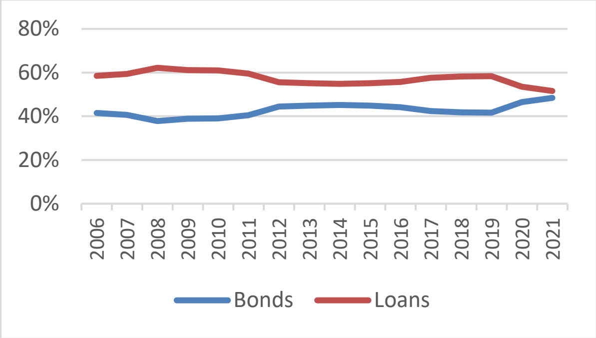


Figure 2–1 Proportion of loans and bonds in the Czech Republic (end of the year)

Source: Czech National Bank (ARAD, 28. 9. 2022), author

In most advanced economies, banking intermediation has reduced over the past years due to the increasing breadth of **credit markets,** and the trend toward market-based finance seems to be very strong (De Servigny and Renault, 2004). However, banks as intermediaries still fulfil a crucial role in three primary functions: liquidity, risk and information intermediation.

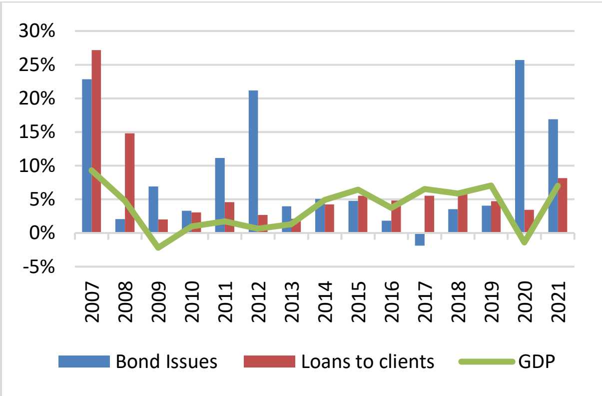


Figure 2–2 Annual change of loans and bond issues in the Czech Republic (in %)

Source: Czech National Bank (28. 9. 2022), Czech Statistical Office (28. 9. 2022), author

2.2 Classification of Credit Risk

Risk can generally be defined as the volatility of returns leading to unexpected losses, as defined by Crouhy et al. (2014). Several risk factors can influence this volatility of returns, including:

- **Market risk** that changes in market prices and rates will negatively affect a security or portfolio value.
- **Credit risk** of an economic loss from a counterparty's failure to fulfil their contractual obligations.

- **Liquidity risk** includes the risk that a firm cannot raise the necessary cash (funding risk) or a transaction will not be executed (trading risk).
- **Operational risk** refers to potential losses from operational failures (management, controls, fraud, human factors). It is closely related to legal and regulatory risk or reputation risk.
- **Business risk** refers to uncertainty about the demand for products, prices, and production costs.
- **Strategic risk** is the risk of significant investments with high uncertainty about success and profitability.

The existence of credit and credit markets is associated with **credit risk**. In the text, credit risk refers to the risk of economic loss from a counterparty's failure to meet its contractual obligations, such as interest payments or principal repayment. However, credit risk involves the possibility of non-payment on a future commitment and during a transaction. This type of risk is called settlement risk. It arises from exchanging principals in different currencies or payments in different time zones during a short window, typically a day. Traditionally, credit risk is considered a pre-settlement risk, which arises during the obligation's life (Jorion, 2011). Overall, credit risk can be decomposed into the following four categories:

- **Default risk** refers to the debtor's capacity or refusal to meet debt obligations such as interest or principal payments by more than a reasonable relief period from the due date (usually 60 days in the banking industry).
- **Bankruptcy risk** can be considered the risk of taking over a defaulting borrower's assets or counterparty. In this case, debt holders are taking over the control of the company from the shareholders.
- **Downgrade risk** is the risk that the creditworthiness of the borrower or counterparty might deteriorate in the future when a significant deterioration can be seen as the default.
- **Settlement risk** refers to the risk due to the exchange of cash flows when a transaction is settled. It can be caused by counterparty default, liquidity constraints, or operational issues (Crouhy et al., 2014).

Due to all types of credit risk, suppliers must assess their creditworthiness before granting credit to prospective borrowers (Joseph, 2013). In addition, since traditional banks typically hold the loan until maturity, they analyze the riskiness of the borrowers' activities both before and after the loan is made because they face the risk that the borrower's credit quality could deteriorate during the life of the loan.

Some lenders, such as banks, use financial innovations in various strategies to reduce credit risk and increase returns. These innovations primarily involve securitization, syndication of loans, proprietary trading and investment in non-traditional assets, or increased use of financial derivatives (Saunders and Allen, 2010; Stowell, 2010):

- **Securitization** represents an innovative way for lenders to raise funds in the capital market by selling their assets' future receivable cash flows, such as mortgage, student or credit card loans. The loans and other assets are packaged and sold as asset-backed securities. This process transfers credit risk to investors, while banks can free up capital for other lending and investment activities.
- **Loan syndication** is another way how banks can reduce risk exposures. Firstly, a bank originates a loan and then sells parts of the loan to outside investors. The outside investors include other banks, hedge funds, mutual funds, insurance companies and other investors. **Banks' proprietary** investment activities involve non-client-related investments in securities or other assets for their accounts; for example, banks establish hedge funds, private equity**,** or venture capital funds. These subsidiaries are then involved in investment activities that are considered too risky for banks.
- The use of **financial derivatives** covers the use of credit default swaps designed to transfer the credit risk on a portfolio of banks to nonbanks, typically insurance and reinsurance companies.

As we can see in the previous text, lenders such as banks can reduce credit risk in different ways. These innovative activities also slightly change the traditional view as an institution that issues short-term deposits and offers long-term loans. Even though these innovative strategies have been increasing in recent years, banks still face a substantial credit risk resulting from their traditional activities. For this reason, they pay considerable attention to credit risk measurement and management. Not only do banks face credit risk from their operations, but also persons placing deposits with banks or investors purchasing corporate bonds.

We already defined credit risk as the probability of loss due to the failure or counterparty's unwillingness to meet contractual obligations. According to Joseph (2013), credit risk generally exists whenever a product or service is obtained without paying for it. A single borrower (obligor) exposure is known as firm-credit risk, while credit exposure to a group of borrowers is called a portfolio-credit risk. Credit risk is the product of various events and factors, such as domestic, international or company-specific issues. As we can see from the scheme in Figure 2-3, some of the causes are more controllable than others.

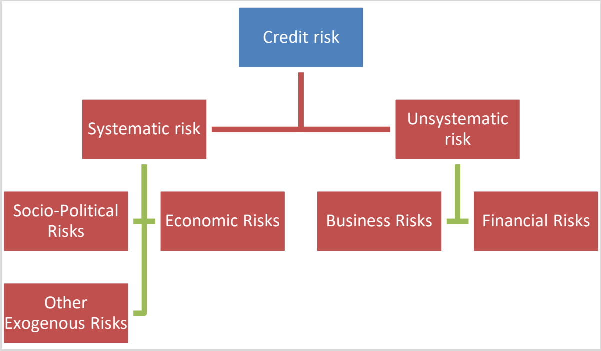


Figure 2–3 Major sources of credit risk

Source: Joseph (2013), p. 16

Uncontrollable risks are called systematic risks, and they are associated with external forces that affect all businesses and households in the country. For instance, the frequency of defaults or bankruptcies typically increases during the economic recession, causing credit losses for the lenders (Joseph, 2013). Figure 2-4 shows the number of corporate defaults of companies rated by Standard & Poor's from 1981 to 2015. The bankruptcy number increased during each of three periods of economic downturn: The recession of the early 1990s that came after the Black Monday of October 1987, the first 2000s recession, and the great recession of 2008.

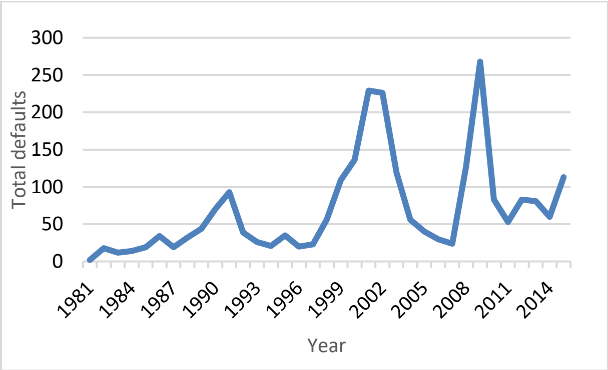


Figure 2–4 Total number of corporate defaults (1981–2015)

Source: S&P Global Ratings (2015), author

Unsystematic risk can be considered controllable because these risks do not affect the entire economy or all businesses or households. On the other hand, these risks are mainly industry or company-specific. Lenders might reduce unsystematic risk through diversification or extending credit to various customers.

2.3 Factors of Credit Risk

The principal problem in credit risk measurement is to quantify the risk of losses due to counterparty default. As Jorion (2011) suggests, the distribution of credit risk can be considered a compound process driven by the following three variables:

- Default,
- loss given default,
- credit exposure.

Default is the principal issue in credit risk measurement, so it is essential to pay some attention to its explanation. The definition of default of an obligor typically includes the following characteristics:

- Days past due criterion for default identification,
- indications of unlikeness to pay,
- conditions for a return to non-defaulted status.

Due to the absence of specific rules and other aspects of the application, various approaches have been adopted across institutions and jurisdictions. Based on the European Banking Authority (EBA, 2016), institutions use differing practices regarding default. As stated in the report[1], specific rules adopted in most jurisdictions usually focus on counting days past due and applying the material threshold. On the other hand, particular rules on different aspects of the definition of default are much less common. To harmonize a consistent use of default meaning, the EBA suggests guidelines to increase comparability of risk estimates and own funds requirements, especially when using internal rating-based or IRB models. For example, in the Czech Republic, the default subject is regulated by the Act on Bankruptcy and Settlement, known as the Insolvency Act[2]. This Act aims to control the resolution of the debtor's insolvency and imminent bankruptcy and the debtor's discharge of debts. According to this Act, a debtor is insolvent if they have several creditors, outstanding financial liabilities overdue for more than 30 days, and cannot fulfil such liabilities. While insolvency is a specific legal term meaning that a debtor cannot pay their debts, default generally means that a debtor has not yet paid a debt as required.

We can consider default as a discrete state for the counterparty with some **probability of default** (PD). The determination of the likelihood of default is the crucial issue in the credit risk management approach and can be achieved through various methods (De Laurentis, 2010):

- The observation of historical default frequencies and allocation to different credit classes (ex-post),
- the use of mathematical and statistical tools to expect the probability (ex-post),

---

[1] The guidelines will apply from 1 January 2021

[2] Act No. 182/2006 on Bankruptcy and Settlement

- the combination of judgmental and mechanical approaches or the approach based on market prices.

The probability of default can be considered the default risk measure within a specified time horizon, usually one year. Alternatively, when exposures are more than one year, the assessment is typically based on cumulative probabilities (De Laurentis et al., 2010).

The typical technique to assess the creditworthiness of retail and commercial loans' counterparty is scoring models (De Servigny and Renault, 2004). Although the credit scoring method was explored and introduced by Altman (1968) several decades ago, it is still a topical theme for researchers and practitioners. Today, different and more sophisticated methods, such as nonparametric techniques or machine learning methods, can be applied in credit risk management. Another approach to assessing default risk is based on firm-value-based or structural models that describe the default process as the explicit outcome of the firm value's deterioration. Based on this approach, corporate securities are considered contingent claims or options on the issuing firm's value. This method was introduced by Merton (1974) as the first example of an application of option pricing methodology to price corporate securities. Credit scoring models can be applied to any borrower, whereas structural models can be primarily used for the largest companies listed on stock exchanges (De Servigny and Renault, 2004).

Loss given default (LGD) is the second key variable in a credit risk analysis, and it can be considered the fractional loss due to default, provided that default is given in this case. The complement to one is called recovery rate; for example, if a fractional recovery rate is 30%, 70% of the exposure is LGD. The recovery rate is expressed as a percentage of the par amount recovered on defaulted debts and refers to the amount of money recovered. LGD can be defined as

$$LGD = 1 - f_i \tag{2.1}$$

where f_i is the recovery rate (Jorion, 2011).

The main difference between the probability of default (PD) and loss given default (LGD) is that a distribution better represents LGD than a single figure. As De Servigny and Renault (2004) suggest, uncertainty about recovery depends on quantifiable factors and more fuzzy factors such as debtors or creditors' bargaining power.

For example, there is a clear link between seniority and the recovery level, as shown in Table 2-1. The table shows the debt recoveries of companies rated by Moody's during 1985 – 2016. We can see that recoveries correlate with their priority of claim in the capital structure in most cases, where claims with higher priority have higher average recovery rates. There are small differences in recovery rates between Europe and the rest of the world; however, it must be noted that these results, particularly European recoveries, are based on a relatively small sample of loans and bonds.

Table 2–1 Recovery rates (1985 – 2016)

	Europe		Global	
	Recoveries	No of Issuers	Recoveries	No of Issuers
First Lien Loan	65.69%	11	66.84%	460
Senior Unsecured Loan	50.71%	7	46.42%	66
Senior Secured Bond	46.14%	37	50.85%	358
Senior Unsecured Bond	38.39%	107	36.99%	972
Senior Subordinated Bond	33.98%	11	30.97%	507
Subordinated Bond	36.87%	39	31.14%	386
Junior Subordinated Bond	14.00%	1	23.23%	24

Source: Moody's Investors Service (2016)

Altman (2008) points out that while significant attention by the credit risk literature has been devoted to estimating PD, much less attention has been paid to the recovery rate in the event of default (RR) and the relationship between PD and RR. It can be a result of two factors:

- Credit pricing models and risk management applications usually focus on the systematic risk components of credit risk (they attract risk premia).
- Credit risk models traditionally assume that RR depends on individual features such as collateral or seniority, which do not respond to systematic factors. For this reason, RR is supposed to be independent of PD.

As Altman (2008) argues, recent studies on RR estimation and RR and PD's relationship have reversed this traditional focus on default analysis. Based on recent empirical evidence, it can be said that core factors of recovery rate involve recovery procedures in different countries, such as legal system and jurisdiction, general economic conditions, the industry of the issuer and the availability of collateral or guarantees (De Laurentis et al., 2010; De Servigny and Renault, 2004). De Servigny and Renault (2004) claim that the same macroeconomic indicators influence default probabilities and recovery rates; thus, there might be a relation between these two variables. The link was empirically examined in the study by Altman et al. (2001). The main results suggest that high default rates are historically associated with low recovery rates. The statement can be explained by the fact that recession periods are connected with an asset increase being liquidated and decreased investment and demand. As lower demand drives prices down, recovery rates deteriorate and eventually co-move with default rates (De Servigny and Renault, 2004). For example, Finger (1999) and Gordy (2000) proposed models based on the assumption that the same economic conditions that cause defaults to rise might cause RRs to decline. In other research, the distribution of recovery is different in high-default periods from low-default ones. Thus, the correlation between these two components is driven by their mutual dependence on the systematic factor.

Credit exposure (CE) is the economic or market value of the claim on the counterparty, and it can also be called exposure to default (EAD) at the time of default. Credit exposure is the measure of exposure risk, which is the amount of risk in the event of default. Jorion (2011) states that credit risk is traditionally measured in the context of loans or bonds for which the asset's exposure, or economic value, is close to its notional or face value.

Then, the credit exposure can be defined as

$$CE_t = Max(V_t, 0) \tag{2.2}$$

where CE_t is the credit exposure and V_t is the value of the assets. If the counterparty defaults with money owed, the full amount must be paid; however, only a fraction must be recovered (Jorion, 2011).

Measures of credit risk can be based on:

- Notional amounts (simple exposures),
- risk-weighted amounts (risk-adjusted exposures),
- notional amounts combined with credit rating (exposures adjusted for default probabilities),
- internal portfolio credit models (a complex measurement of credit risk).

The simplest way of credit risk measurement tools is based on the total notional amounts when a multiplier is applied to get the quantity of capital required to hold as a reserve against credit risk. Because this approach ignores the probability of default, the Basel Committee in 1998 implemented a rough categorization of credit risk by risk classes. However, these risk weights proved too simplistic, not sufficiently preventing banks from taking higher risks. Thus, Basel II rules were introduced, allowing banks to use their internal or external credit rating.

2.4 Credit Risk Analysis

Credit risk analysis is the procedure a credit supplier uses to assess a borrower's creditworthiness. Although modern approaches to credit risk analysis have been developed in recent decades, we can say that examining the borrower's ability to repay the funds is one of the oldest activities accompanying money lending (Joseph, 2013). Modern credit analysis techniques involve approaches based on accounting and financial information (fundamental-based models) and methods connected with security prices (market-based models).

Credit risk is directly connected with possible bad debts or credit losses for financial and non-financial businesses. In the context of recent events and developments in the financial markets, primarily following the 2008 financial crisis, there has been an increasing focus on credit risk management. The primary attention is paid, for example, to prudence, increase in bankruptcies, increase in competition, the volatility of asset values, low asset quality, high impact of the credit losses, existence of limited liability entities, or off-balance sheet activities (Joseph, 2013).

2.4.1 Principles of Risk Grading

The principal aim of the credit risk analysis is to evaluate the counterparty's creditworthiness and the exposure or financial impact in the event of default. The focus is on determining the ability and willingness to meet payment obligations when due. However, credit analysis refers to a comprehensive study that involves the following four areas (Levine and Sarchese, 2010):

- Business risk: Business risk is assessed through the company (size, life cycle stage, market position, cost position, or diversification) and industry analysis (cyclicality, life cycle stage, capital intensity, or competition).
- Financial risk: Financial risk is based on analysing cash flow, stress testing, financial ratios, liquidity, quality of assets, access to capital markets, or capital structure.
- Management risk: Management risk is focused on the quality of management, experience, reputation, track record, strategy and vision, and financing philosophy.
- Covenants: Covenants refer to the analysis of bond covenants such as debt insurance, restricted payments, change of control or assets sale, and bank loan covenants, such as mandatory prepayments or maintenance tests.

The analysis of the four key areas should focus on studying historical data, performance and future prospects. All the sites should be thoroughly examined, and the specific features of individual companies or industries should be considered. For this reason, credit risk analysis can be seen as a very complex assessment of all potential risks associated with the borrower's current and future situation. The final assessment evaluates the probability of default and credit loss linked to borrowers' risk grades or scores. Thus, the rates should reflect the degree of credit risk associated with the borrower. An essential part of this process is updating the grades based on consistent risk monitoring and management (Joseph, 2013, p. 161). In the credit risk assessment process, the credit risk evaluation is converted into credit risk grades. Credit risk grades can be assigned based on judgmental (subjective) evaluation or objective assessment based on credit risk modelling. The credit risk classifying system can comprise different grade levels; for example, according to the Basel Accords, at least seven risk grades are required (Joseph, 2013).

In credit risk grading, the probability of default is frequently used to distinguish between good and bad credits. PD grading scales are usually broken down into more categories to evaluate the borrower better. From a statistical point of view, PD represents the percentage probability of a borrower defaulting, usually within a one-year time horizon (Joseph, 2013). Unlike the method of credit risk grading, PD assigns a statistical value for default. We can consider the estimation of PD as the core part of the credit risk analysis. It is also incorporated in the IRB[3] approach suggested by the Basel Capital Accord. For example, it is defined as the greater of the one-year PD associated with the internal borrower grade to which that exposure is assigned, or 0.03 per cent for IRB purposes (Bank for International Settlements, 2001). To quantify PD, banks can use various methodologies and data sources, where the approaches are typically based on a bank's own default experience. Commonly, better credit risk grades are associated with lower probabilities of default. Figure 2-5 shows the hypothetical relationship between credit risk scores ranging from 1 – 20 (1 – very low risk, 20 – loss) and PD behaviour.

Usually, PD values are derived from historical data of borrowers' defaults over time. Lenders can use the PD values for capital allocation, credit risk pricing, or economic capital (Joseph, 2013).

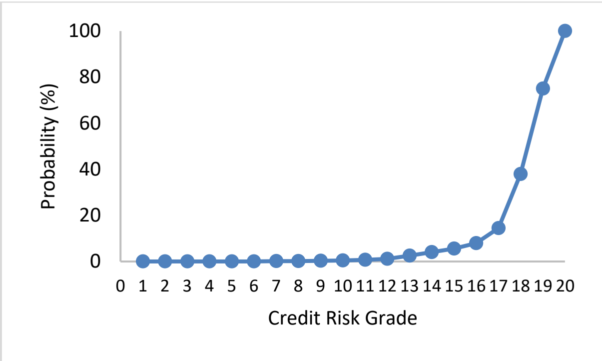


Figure 2–5 Credit risk grades and PD

Source: Joseph (2013, p. 165), author

2.4.2 Credit Rating Assessment

Credit rating can be seen as an ordinal measure of the probability of default on a given time horizon, providing a relatively easy way to assess credit risk. Credit rating systems are very similar to credit scores; however, we can find a distinction. A credit rating is usually expressed as a letter grade to evaluate a business or government's creditworthiness. On the other hand, the credit score is expressed in numerical form and used mainly for individuals.

There are two types of credit rating:

- External ratings,
- internal ratings.

[3] The Internal Ratings-Based Approach (Basel Committee on Banking Supervision)

While external ratings are published by external agencies such as Standard &Poor's (S&P), Moody's or Fitch, internal ratings are assigned by lenders that use their rating systems. Although these internal ratings may differ, there are requirements for banks and financial institutions covered by the Basel Accords. As De Laurentis et al. (2013) suggest, ratings only indicate that some choices are riskier than others. To ensure the objectivity of the assessment and responsibility of lenders, rating systems have the following features:

- Measurability: Ratings give correct expectations in terms of PDs, and they are adequately and continuously backtested.
- Objectivity and homogeneity: Rating systems provide judgements only based on credit risk considerations, and ratings are comparable among portfolios, market segments, or customer types.
- Specificity: The rating system measures the distance from the default event without regard to other corporate financial features not directly related to it.

Rating systems can be developed using different approaches and methodologies; for example, De Laurentis et al. (2010) distinguish between:

- Expert-based approaches,
- statistical-based models,
- heuristic and numerical methods.

Expert-based approaches include agencies' ratings and expert-based internal ratings used by banks and financial institutions. Agencies' ratings are provided and published by specific companies, typically called credit rating agencies (CRA). The rating assessment by these agencies is usually a combination of judgmental-based and model-based approaches. Many national and international rating agencies operate in various countries, from importance agencies to agencies, the "Big Three", three rating agencies with the largest market share. These rating agencies include Standard &Poor's, Moody's and Fitch, controlling most of the rating business industry. Another type of expert-based method refers to internal ratings developed by banks or financial institutions for credit risk assessment purposes. An example of the internal expert-based approach is the IRB approach to capital requirements for credit risk.

According to the relevant document (BIS, 2001), the primary purpose of the IRB approach is to provide a single framework by which a given set of risk components are translated into minimum capital requirements. As proposed by this approach, banks are suggested to take into account not only the probability of default (PD) of a borrower or a group of borrowers when assessing credit risk but also loss given default (LGD) and exposure at default (EAD). Besides, the approach also considers the maturity of exposures (M). Thus, these four components (PD, LGD, EAD, and M) form the IRB approach's necessary inputs and capital requirements.

While expert-based approaches are usually based on judgmental and model-based analyses, **statistical-based models** refer to quantitative financial models.

Although the models are based on simplifying assumptions about the variable to be predicted, they should also incorporate unquantifiable factors such as the vision of organisations' behaviour or possible economic events. For this reason, these models usually represent a mixture of statistics, behavioural psychology, and numerical methods. Different assumptions and varying proposed uses lead to other models; however, these models are typically developed to classify a borrower's creditworthiness or predict the probability of default.

The application of artificial intelligence methods mostly drives **heuristic and numerical approaches**. These methods have been applied to default prediction and provide an alternative credit risk management approach in recent years. Heuristic methods mimic human decision-making procedures, generating new knowledge on trial by error rather than statistical modelling. On the other hand, numerical methods are used to reach optimal solutions by trained algorithms to make decisions in highly complex environments characterised by inefficient and fuzzy information. An example of this approach is the suitable neural networks method, especially in massive quantitative data analysis. Artificial neural networks originate from biological sciences when artificial neurons refer to hierarchical nodes, or steps, connected in a system by mathematical models. Since some nodes can be based on statistical methods, neural networks can also be considered statistical models in some cases (De Laurentis et al., 2010).

All statistical methods described in the previous text are well suited for using quantitative data, for example, income payments, financial ratios, and macroeconomic or industry variables. However, as many other non-quantifiable factors potentially impact credit risk, qualitative information can also be incorporated into the models to increase their accuracy. Thus, credit risk models are typically based on both quantitative and qualitative risk factors.

2.5 Obligor-Level Credit Risk

The primary objective of credit risk analysis is to assess a borrower's creditworthiness or a group of borrowers, as mentioned in the previous section. Lenders such as banks, financial institutions, or bond investors make decisions based on potential borrowers' credit risk analysis. This chapter is devoted to the main subjects of credit risk assessment, and attention is paid to the obligor-specific credit risk assessment. In general, the obligor (the debtor) is an individual or company that owes a debt to another individual or company (the creditor), usually due to borrowing or issuing bonds. This chapter focuses on obligor-level credit risk's primary fundamentals, both at a company and an individual level. Since these two categories have some specifics regarding credit risk, they are described in more detail in the following subchapters.

2.5.1 Individual Credit Risk

Individual credit risk is associated with personal lending when a borrower, typically a bank customer, applies for a personal loan. Banks have already developed their internal credit scoring systems to identify critical factors that determine the probability of default, usually based on their experience. Generally, the main factors affecting an individual credit risk are related to the borrower's personal and economic position, which banks have used for many years. For example, Chapman (1940) specifies two types of aspects related to credit risk in personal lending:

- Personal characteristics such as age, sex, family status, and
- occupational and economic position, for example, income and borrower's net worth.

The ability to pay is primarily determined by the applicant's employment and the industry in which they are engaged. These factors are related to borrowers' income, assets such as real estate, automobiles, securities, and debts, such as mortgages, credit cards, and other personal loans that can be used to identify the financial capacity. In many countries, the credit history of a borrower's responsible repayment of debts is recorded and used by lenders as an essential aspect to determine individual creditworthiness or an individual's ability to repay a debt. The importance of each of the former factors can differ for different banks, and it is usually subject to their assessment. In assessing the credit quality of a loan applicant, lenders look at various measures. The starting point is the applicant's credit score, a numerical grade of the borrower's credit history (Fabozzi, 2013).

The well-known and widely used credit score system by lenders in the United States is the FICO Score, developed by Fair Isaac Corporation and first introduced in 1989. This system is used to assess the credit risk of individual borrowers and determine whether to extend credit. This assessment is based on account payment history, the current level of indebtedness, types of credit used, lengths of credit history or new credit accounts (Fair Isaac Corporation, 2017). FICO scores range from 300-850, with industry-specific scores from 250-900, where the higher the score, the lower the credit risk. The basic scheme of FICO scores and their definitions are in the table below (Table 2-2). According to recent data, American consumers' average FICO score reached 699 in late 2016 (Karimzad, 2015).

Table 2–2 FICO Credit Scores

FICO Scores	**Definition**
800 +	Excellent, an exceptional borrower
749-799	Good, a very dependable borrower
670-739	Average, a good score borrower
580-669	Fair, below the average borrower
579 and lower	Poor, a very risky borrower

Source: Karimzad (2015), author

The system of FICO Scores is based on the following five categories:

- **Payment history** refers to a borrower's historical ability to pay their payment on time; this category represents the most critical factor in the credit assessment. Credit history usually includes credit cards, retail accounts, instalment loans, or mortgage loans.

- **Amounts owed** show the amounts owed on specific accounts, including credit card balances, instalment loans, and other revolving credit accounts.
- **Length of credit history** positively affects the credit score; the longer the record, the higher the score.
- **Credit mix** is another crucial determinant of the score, especially the total number and types of borrowers' accounts.
- **The new credit** category suggests that opening several credit accounts in a short period represents a greater risk, especially for people with a brief credit history.

The contribution of each category to the total FICO score is shown in Figure 2-6. As we can see, the significant factors in credit assessment are payment history and amounts owed.

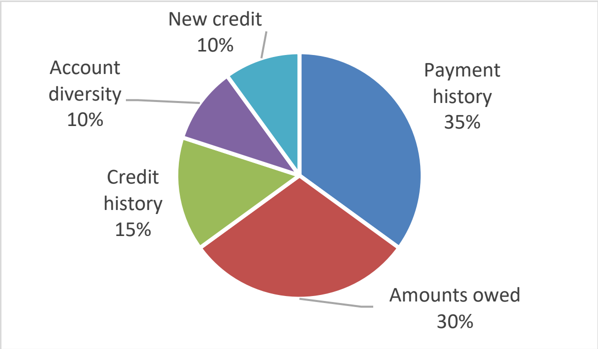


Figure 2–6 FICO Score categories

Source: Fair Isaac Corporation (2017), author

The process by which the lender decides whether an applicant is creditworthy and should receive a loan is called underwriting. The requirements specified by the lender to grant the loan are called underwriting standards. The approval process can be judgmental, fully automated, or a combination of the abovementioned types; however, it should consider all necessary information to support loan granting decisions. In the case of secured loans, collateral identification should also be considered (FDIC, 2017).

For example, the two primary quantitative underwriting standards for granting residential mortgage loans are:

- Payment-to-income ratio (PTI) that refers to the rate of monthly payments to monthly income. PTI is used to measure an applicant's ability to make monthly payments. The higher the ratio, the lower the risk.
- Loan-to-value ratio (LTV) is the ratio of the loan amount to the market or appraised property's value. The lower the rate, the lower the risk for a lender (Fabozzi, 2013).

2.5.2 Corporate Credit Risk

Corporate credit risk assessment examines firm-level credit risk that can be affected by various factors. Firm credit risk analysis typically involves two parts of the evaluation: business and financial risks.

Firstly, business or operating risks are associated with risks that originate from other than the company's financial aspects. These risks include outside and inside events with a potential impact on the business credit risk, for example, changes in economic, regulatory, climatic, industry, demographic, geo-political, product innovations, quality of management, or other factors. For example, Joseph (2013) suggests the following three categories of risks from the operating environment:

- External,
- industry,
- internal.

External risks can be seen as systematic risks that involve the impact of the business cycle, economic conditions (private consumption, government spending, investment, imports and exports), inflation, the balance of payments, exchange rates, political factors, fiscal policy, monetary policy, demographic factors, regulatory framework, technology, environmental issues, international developments and other types of systematic risks. It should also be considered that these external variables are usually interrelated.

Industry analysis focused on industry life stage, composition, nature, or structure is another crucial part of credit risk analysis. In this part of the study, the stage of the industry life cycle, government support, factors of production, the sensitivity of industry to the business cycle and industry profitability should be examined, followed by competitor group analysis. Industry profitability assessment is usually based on the analysis of forces that determine the potential of an industry, known as Porter's model, which provides a basis for analysing the level of competition Figure 2-7.

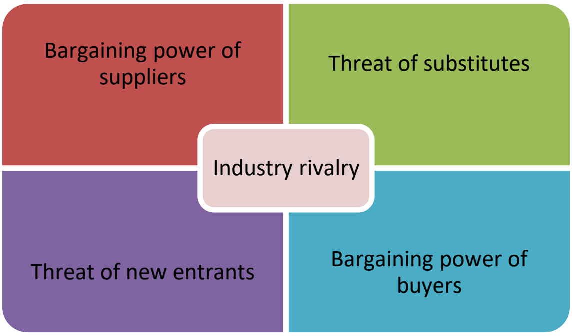


Figure 2–7 Porter's model

Source: Joseph (2013, p. 67), author

Finally, internal or company credit risk analysis is focused on the capabilities, resources strategies, competencies, strengths and weaknesses of the borrowers (Joseph, 2013). In addition, attention is paid to business activities and identifying internal risks, including peer comparison and SWOT analysis. Other internal risks include, for example, production, human resource, product, customer/supplier concentration, legal, reputation or financial risks.

The second type of firm credit risk is originated solely from the financial aspects of a business. Because even a successful business may go bankrupt due to inappropriate financial decisions, substantial attention is paid to analysing **financial risks**. Financial risks are linked to a company's financing policies, strategies, and decision-making that can substantially affect the credit risk level. Financial risk analysis is primarily based on the analysis of financial statements, the balance sheet, income statement, and cash flow statement; however, other financial statement information, such as a statement of stockholders' equity, can also be useful. For credit risk assessment, a business's comprehensive economic analysis is conducted to identify the financial strengths and weaknesses and warning signals of financial risks. The study involves common size analysis, indexed trend analysis and financial ratio analysis. Typically, the focus is paid to all categories of financial ratios (liquidity, solvency, activity and profitability ratios), including studying their relationships and eventually predicting financial default. The procedure of the analysis of financial statements and their interpretation have already been discussed in many publications, see for example, Fridson and Alvarez (2011), Berk and DeMarzo (2017), Brealey et al. (2014), Megginson et al. (2008), Joseph (2013) or Dluhošová et al. (2014). The analysis of relations among financial ratios is usually examined through the DuPont Model; however, scoring models are typically applied for default prediction. To summarise, business and financial risks should be studied together, and the final credit risk assessment should be based on the company's overall situation.

2.5.3 Corporate Credit Scoring Models

We can understand credit scores as statistically derived indicators of risk that indicate the relative risk that a borrower will experience an adverse credit event, for example, delinquency or default. When the credit scoring model is built, the statistical model's output is usually transferred to generate a set number of score points or the probability of a credit event occurrence (Mays and Lynas, 2011). Lenders develop models to assess borrowers' credit risk, both at an individual and corporate level. While an example of a well-known individual credit scoring model used by financial institutions is the FICO model, specific models are developed to assess corporate credit risk. Corporate credit scoring models include both the models developed by banks based on their borrowers' behaviour and publically available scoring models. Different entities may use the later models to get an overall picture of a counterparty's creditworthiness, for example, business partners. In contrast to individual borrower credit scoring models, corporate credit scoring models use different input variables, typically financial ratios and other corporate financial performance indicators.

Several methods can be used to derive scoring models, as explained further in the text. The proposed models are then used to calculate the score values of entities and can be used to classify them into pre-defined categories. One of the best-known models in this area was developed by E. I. Altman (1968), whose default model is often known as the Altman's model or Z- Score model as a tool in the financial analysis of a company. This model can identify companies with possible financial problems, namely default risk, and it can be proposed based on multivariate discriminant analysis, whose product is a so-called Z-score, classifying companies.

The original version of the Z-Score model can be used to predict the likelihood of a firm going bankrupt, and the score can be calculated using the following formula (Joseph, 2013),

$$Z = 1.2X_1 + 1.4X_2 + 3.3X_3 + 0.6X_4 + 0.999X_5, \tag{2.3}$$

where the variables in the formula refer to the following ratios: X_1 working capital/total assets; X_2 retained earnings since inception/total assets; X_3 profit before interest and tax/total assets; X_4 market value of equity/book value of total debt; X_5 sales/total assets.

The first Altman's prediction model is based on a weighting system of five financial ratios. It was developed based on statistical data from sizeable public manufacturing companies with more than $1 million in assets. Its primary purpose is to measure a company's financial health and predict the probability of bankruptcy within two years. Although some empirical studies show that the model has a 72% – 80% reliability of predicting bankruptcy, we should realise that it can only be used to forecast if a company being analysed can be compared to the database. The resulting scores of the original Z-Score model for public manufacturing companies and their implications can be seen in Table 2-3.

Table 2–3 Original Z-Score model

Z-Score	**Forecast**
Above 3.0	Bankruptcy is not likely
1.8 to 3.0	Bankruptcy cannot be predicted – GREY AREA
Below 1.8	Bankruptcy is likely

Source: Wilkinson (2013)

Although the model is relatively simple, it is still used and is mainly relevant for manufacturing companies. According to Cao (2016), Altman decided on two potentially very powerful variables among all possible financial variables that had not been used yet. One of the variables is the retained earnings because, as Altman explains, "*a firm that has grown its assets mainly by reinvesting earnings is healthier than a firm that has grown the assets by using other people's money. Retained earnings is also a measure of the company's age and leverage" The other* *variable is the market value of the equity relative to the book value of the debt, an indication of the company's abilit*y to raise money from capital markets. Today, equity's market value is a fundamental part of structural models provided, for example, by Merton (1974) or the KMV model by Moody's Analytics (2017).

Since the first version of Altman's model, several modifications have been suggested, or some comments have been published by Altman, Altman et al. (i.e. 1970, 1977, 2005, 2007, 2010) to other authors. It is necessary to realise that such models' development is highly demanding due to data intensity and modelling specifics. These techniques are difficult to employ without information technologies and specific mathematical-statistical applications.

2.6 Issue-Specific Credit Risk

Issue-specific credit risk typically refers to a bond issuer's credit risk, specific bond issues, or other issues of debt securities, which represent a contractual agreement between a lender (investor or bondholder) and a borrower (issuer). However, credit risk can also be associated with innovative contracts such as asset-backed securities or credit derivatives. In this chapter, these three categories of securities will be discussed in more detail, particularly in the context of credit risk.

2.6.1 Credit Analysis of Bonds

A bond can be defined as a debt instrument requiring the issuer to repay the investor the amount borrowed plus interest over a specified period (Fabozzi, 2013). Mostly, the principal must be repaid on the maturity date. Thus, we can see an analogy between financial institutions or other entities lending money and the issuance of securities from a credit risk perspective. Similarly, the credit risk assessment will be conducted based on the borrower's ability to repay all contractual payments when the amounts are due. Therefore, the analysis is focused on the study of issuer business and financial risks, including the analysis of financial statements. On the other hand, bond credit risk analysis should consider some features specific to bonds, such as the study of indenture and covenants.

As Fabozzi (2013) suggests, the issuer's nature is a vital feature of a bond. There are three types of issuers of bonds: governments, municipalities, and corporations. Some bonds are issued with an amortisation feature, meaning that the principal repayment can be repaid over the bond's life; these securities are called amortising securities. In addition to simple or 'plain vanilla' bonds, there are also bonds with embedded options, for example:

- Bonds with a call provision: The issuer has the right to retire the debt before the scheduled maturity date.
- Bonds with a put provision: The bondholder has the right to sell the issue back to the issuer at par value on pre-specified dates.
- Convertible bonds: The bondholder has the right to exchange the bond for a specified number of shares of common stock.

- Exchangeable bonds: The bondholder can exchange the issue for a specified number of common stock shares of a corporation different from the bond issuer.

Investing in bonds is associated with some risks, such as interest rate, reinvestment, call, credit, inflation, exchange, liquidity, or volatility risks. While attention in this chapter will be paid to bond credit risk, the description of other types of risks can be found in a vast literature on this subject, for example, Fabozzi (2013), Bodie et al. (2011), Reilly and Brown (2015), Petitt et al. (2015).

Corporate bond credit analysis consists of three areas (Fabozzi, 2013):

- Analysis of covenants,
- analysis of collateral, and
- assessing an issuer's ability to pay.

Analysis of covenants is linked to the study of the indenture provisions that form rules for essential areas of operation for corporate management. These provisions, including bond covenants, can be found in a company's prospectus for its bond offering. There are generally two covenants: affirmative (promises by the corporation) and negative or restricted (limitations on the borrower). Restrictive covenants may limit the absolute amount of outstanding debt or a fixed charge coverage ratio test. For example, the maintenance test requires the borrower's earnings ratio to be available for interest or fixed charges at a minimum for a certain period. On the other hand, the debt incurrence test is used to adjust interest or fixed charge coverage when the company takes on additional debt. In some indentures, we can also find limitations on subsidiaries' borrowing from all other companies except the parent.

Analysis of collateral refers to the careful understanding of a corporate debt obligation security when the debt can be secured or unsecured. Generally, if the company is liquidated, proceeds from bankruptcy are preferably distributed to creditors. Secured bonds are collateralized by an asset (i.e. property, equipment). In the event of default, investors claim the issuer's assets to recover their loss to some extent. However, most corporate bonds are unsecured, and investors have no claim on specific collateral. We can see this fact in Table 2-4, which shows the proportion of secured and unsecured bonds issued by European industrial companies as of the end of 2010[4]. While the balance of unsecured bonds is more than 90% of total bonds, secured bonds represent a minority in both groups of countries.

[4] There are 23 countries included and divided into two groups EU-15 (Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden, United Kingdom) and EU-8 (Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovakia, Slovenia).

Table 2–4 Proportion of secured and unsecured bonds

	EU-15	EU-8	Total proportion (%)
Secured/Senior Secured	80	2	7.9
Unsecured/Senior/Subordinated Unsecured	875	85	92.1
Total	955	87	100

Source: Reuters database (accessed 1st December 2010), author's calculations

The third area of the bond credit analysis is focused on assessing an issuer's ability to make timely payments of interest and principal. Although a substantial part is based on the analysis of financial statements, we should also analyse other factors that may impact the ability to generate cash flow, thus service the debt. This part of the credit analysis is analogical to the research described in Chapter 2.1.2. It involves studying business risk and financial risk, including assessing corporate governance risk with an emphasis on the ownership structure of the corporation, the practices followed by management and policies for financial disclosure.

Government bond credit risk is associated with the country's overall situation and institutional strength, especially the banking system's stability and policy credibility. Traditionally, government bonds have been considered relatively risk-free securities; however, we can find significant differences among different countries' risks. Government bond credit risk, so-called sovereign credit risk, is usually assessed by rating agencies in terms of rating. However, financial institutions, other entities or investors may also use their credit scoring models, similar to corporate loans. According to Moody's rating agency[5], there are four factors of government credit risk analysis:

- Economic strength: GDP (per-capita), economy size and degree of diversification, medium-term trends (productivity, infrastructure).
- Institutional strength: Policy predictability (continuity), institutional quality, regulatory framework.
- Government financial strength: Fiscal balance, debt indicators (ratios to GDP and revenue), debt affordability (interest/revenue), debt structure, and market access.
- Susceptibility to event risk: The impact of economic, financial and political events.

The two critical factors in the creditworthiness analysis are government financial strength, monetary policy, and economic power, indicating the economic trend. Susceptibility to event risk is the factor that shows the shock resistance of a country, for example, the impact of the financial crisis, Brexit or a US presidential election on the economy and fiscal outlook.

[5] Moody's Investors Service, Moody's 7th Annual CEE Credit Risk Conference, Czech National Prague, 16 April 2013.

In addition to government bonds, there are debt securities issued by local governments, districts, or cities called **municipal bonds.** The credit risk of municipal bonds is also assessed and published by rating agencies to help investors make investment decisions.

The main factors of credit risk analysis cover (SEC, 2017; Peterson, 1998):

- Sources of funds to pay principal and interest,
- purpose of the financing, and
- the financial condition of the issuer.

In this case, financial condition analysis is focused on the magnitude and structure of local debt, including a proposed borrowing. Economic analysis can be used as an adequate indicator of municipal debt burden and municipal borrower's ability to service the debt. For example, the most important financial ratios include debt service related to recurring revenues, operating surplus or total income and total debt to the tax base.

Finally, credit risk is associated with **short-term securities**, such as commercial papers and other short-term debts with maturities of up to one year. The credit analysis is slightly different from long-term bonds as it usually does not consider the likely recovery of the debt instruments. The principal factors of the credit analysis include assessing the fundamental long-term credit quality. However, the short-term credit risk is driven predominantly by the issuer's liquidity position, which indicates the ability to repay the debt from internal or external sources.

2.6.2 Credit Risk of Asset-Backed Securities

Asset-backed securities are considered an innovative and alternative way corporations or lenders can raise funds. Through securitization, a corporation pools loans or receivables and uses the pool of assets as collateral for security issuance (Fabozzi, 2013). Since the cash flows are sold in the form of securities backed by the cash flows of the very assets sold, the securities are called asset-backed securities. Compared to traditional ways of debt financing, such as borrowing in the form of a loan or issuing bonds, securitization is associated with specific features and represents a different way of financing. Issuers of asset-backed securities typically raise funds to finance the origination of loans, and from an accounting point of view, this issuance is considered an asset sale. There are various backing assets, such as residential or commercial mortgages, consumer loans, commercial leases, or any financial instruments with predictable and stable receivable cash flows (i.e. credit card receivables, auto loans, student loans). As lenders issue asset-backed securities by structuring future receivable cash flows of underlying assets, they are also called structured finance securities.

There are the following specific features of the securitization process (Hu, 2011):

- The asset-backed securities are issued through a special purpose entity,
- to accounting aspects, the issuing of asset-backed securities is an asset sale (not a debt financing),

- servicing of the underlying assets for the investor is required,
- the credit of the asset-backed security depends on the credit of the underlying asset,
- credit enhancement is usually needed.

Asset-backed securities are issued in the form of certificates entitling the investors to receive a pre-determined share in a specific pool of assets' cash flows. From this perspective, they are similar to bonds because investors accept regular payments, usually based on a coupon rate. However, in asset-backed securities, the payments depend on the cash flows generated by underlying assets. Thus, principally, we can distinguish between two types of asset-backed securities:

- Pass-through securities that are issued as single-class mortgage-backed securities (i.e. agency MBS) and
- securities structured in several bond classes called tranches (i.e. non-agency MBS, ABS).

A mortgage pass-through security is issued as one bond class, which means that investors are entitled to receive a pro-rata share of the cash flows of the specific mortgage loan pool. When pass-through security is first issued, the principal is known; however, over time, due to regularly scheduled principal payments and prepayments, the amount of the pool's outstanding loan balance declines. Payments of pass-through security are made each month, and the monthly cash flow is less than the monthly cash flow of the loan pool by an amount equal to servicing and other fees. Agency mortgage pass-through securities, so-called mortgage-backed securities (agency MBS), are issued by US government agencies known as Freddie Mac or Fannie Mae, and Ginnie Mae, which is not the issuer; however, it provides guarantees. Since these pass-through securities carry their warranty and fulfil underwriting standards, they can be considered asset-backed securities with the lowest level of credit risk.

Securities structured in bond classes are created to redistribute credit risk using a senior-subordinate structure. While bond classes with the lowest credit risk and the highest rating are referred to as the senior bond classes, the subordinated classes have a lower rating or are not rated. Losses are distributed based on the bond class's position in the structure when losses start from the bottom and move to the senior level. The rules for the cash flow distribution (interest and principal) and losses are explained in the prospectus. They are usually referred to as cash flow waterfall (Fabozzi, 2013; Hu, 2011; Choudry, 2010).

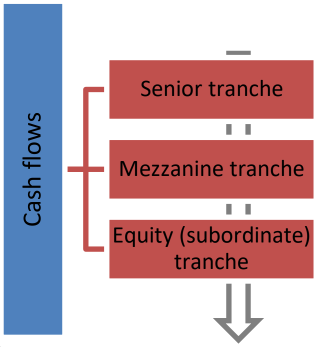


Figure 2–8 Cash flow waterfall

Source: Fabozzi (2013), Author

The simple scheme of the cash flow waterfall is shown in Figure 2-8. Since potential losses are distributed from the bottom to the top, senior tranches have the lowest credit risk and, thus, the lowest expected returns. Typically, these securities are structured with additional credit support to receive an investment-grade rating. This extra credit support is needed to absorb expected losses from the underlying loan pool due to defaults, and it is referred to as a credit enhancement. Credit enhancement is usually provided through the excess spread, over-collateralization or reserve funds.

The **credit risk** of non-agency MBS and other asset-backed securities (ABS) depends on various factors (e.g. quality of the underlying loan pool, structure and bond classes, credit enhancement). Thus, the credit risk assessment is a complex evaluation provided by specialized rating agencies. Although they can use different approaches, they typically focus on the same areas of analysis. For example, Moody’s agency investigates (Fabozzi, 2013):

- Asset risks,
- structural risks,
- third-party providers’ risks.

Asset risk evaluation is associated with the collateral's credit quality, particularly the underlying borrower’s ability to pay and the borrower’s equity in the asset. The key determinants are the experience of the originators of the underlying loans and the concentration of loans. The concentration risk refers to the risk of a not sufficiently diversified pool of loans relative to the entire pool balance. As a result, rating agencies established concentration limits on the amount or percentage of receivables from any borrower, region, country, or industry to reduce the concentration risk.

Structural risks are linked to the structural scheme and the extent to which the cash flow from the backing assets, or collateral, can satisfy all the bond classes' obligations in securitization. While the cash flow of the underlying collateral includes interest and principal repayments, the cash flow payments should cover interest and principal to investors, service fees, and other expenses. According to Fabozzi (2013), rating agencies consider some factors to be the potential for early amortization and credit enhancement changes over time. For example, they include loss allocation, cash flow allocation, and the interest rate spread between the interest earned on the collateral and the interest paid to the bond class, including the servicing fee.

Since several third parties are involved in the securitization, the **third-party providers' risks** are also examined when assessing the credit risk. These third parties include, for example, credit guarantors whose credit risk is associated with their ability to pay. Another third party involved in securitization is a servicer. Since servicers are responsible for collecting payments, delinquencies, recovering and disposing of collateral if necessary, they perform a vital role in securitization. Therefore, rating agencies usually evaluate their abilities to perform all these activities based on their servicing history, experience, underwriting standard for loan origination, servicing capacities or financial condition (Fabozzi, 2013).

2.6.3 Derivatives Credit Risk

Credit derivatives refer to financial instruments that allow a lender or a borrower to transfer the default risk of a loan to a third party. In these contracts, one party is a credit protection buyer, and the other is a seller. For a fee, the credit protection seller provides credit protection to the buyer against some credit events. Hence, the most straightforward credit derivative works like an insurance policy (Choudhry, 2010; Fabozzi, 2013).

Credit derivatives are relatively innovative products in the capital markets, first introduced in 1994. There is a wide range of credit derivatives products, for example, the most commonly used credit default swaps (CDS), or credit debt obligations (CDO), credit-linked notes (CLN), asset swaps, total return swaps, basket CDS and synthetic CDO. All these products are designed to reduce or eliminate credit risk exposure due to credit events that must be specified in the trade documentation, including other terms such as:

- Reference entity – the issuer of the debt instrument on which credit protection is bought or sold and
- reference obligation – the particular debt issue.

Credit derivatives are used mostly by banks as both protection sellers and buyers, where banks are net buyers of protection, while insurance companies are net sellers. Since credit derivatives are over-the-counter products, they are very flexible and can be designed to meet both a buyer's and a seller's specific requirements. Credit derivatives can generally be used to hedge credit risk, reduce credit risk with a particular client and diversify investment options. In addition, these instruments enable trade-in credit as an asset because they isolate and transfer credit risk. Due to this fact, their values reflect only the credit quality of the reference entity. The most common credit derivative is the credit default swap. Generally, there are two types of CDS: single-name CDS and index CDS. A single-name CDS is the simplest CDS associated with one reference entity or a specific asset (i.e., a corporate debt issuer, sovereign issuer, municipal bond issuer, or a tranche of asset-backed security). On the other hand, the index, also called basket CDS, is linked to a group of reference entities. The transfer of the default protection payment from the protection seller to the protection buyer depends on the occurrence of a specified credit event (Choudhry, 2010):

- Downgrade in credit rating below a specified minimum level,
- financial or debt restructuring,
- bankruptcy or insolvency of the reference asset obligor,
- default on payment obligations,
- a change in credit spread payable by the obligor above a specified maximum level.

Credit events are defined in more detail by the International Swaps and Derivatives Association (ISDA), including bankruptcy, obligation acceleration, obligation default, failure to pay, repudiation/moratorium, and restructuring. For more information, see Fabozzi (2013) and ISDA Definitions (ISDA, 2003).

CDS can also be written on asset-backed securities, typically ABS (so-called ABS CDS). However, because of an ABS's unique aspects, modifying the ISDA documentation to credit event definitions was required when the reference entity is an ABS tranche. Thus, in 2005, the ISDA published its pay-as-you-go (PAYG) template for ABS[6] (Fabozzi, 2013). In this template, the focus is paid to the cash flow adequacy of the ABS structure.

There are three credit events defined:

- Failure to pay (the underlying reference obligation fails to make scheduled interest or principal payment),
- writedown (the central component of the underlying reference obligation is written down and deemed irrecoverable),
- distressed rating downgrade (the underlying reference obligation is downgraded to a Caa2/CCC rating or lower).

As Karagozoglu and Jacobs (2010) emphasize, because the CDS prices reflect the credit risk of the reference entity or the obligation, they theoretically represent the credit quality of a firm. Various authors have studied the relationship between CDS spreads and credit ratings and suggested an alternative measure of credit risk. For example, Kiesel and Spohnholtz (2017) show a linear relationship between logarithmized CDS spreads and the issuer credit rating of European and US non-financial corporates. Flannery et al. (2010) focused on CDS spreads of large financial institutions, and their results suggest that CDS spreads may serve as a possible substitute for credit rating.

[6] Available at: http://www.isda.org/press/press012306.html [Accessed 17 July, 2017]

2.7 Fundamentals of Rating Assessment

Credit rating agencies (CRAs) produce ratings and play an essential role in local and international markets. However, due to the problem of misleading some ratings of asset-backed securities in the context of the subprime mortgage crisis of 2007 – 2008, following the global financial crisis of 2008 – 2009, they came under intense criticism. Since their practices led to a loss of credibility, efforts to strengthen their regulation were adopted in many countries. This subchapter provides an overview and the main objectives of the ordinance adopted in the European Union.

Although the main purpose of rating agencies is to evaluate issuers of securities, they are supposed to serve other roles. According to Schroeder (2015), the main functions of rating agencies involve:

- Evaluation of creditworthiness: Rating agencies specialize in evaluating creditworthiness.
- Discrimination: Ratings represent a relative measurement of riskiness.
- Rebalancing asymmetric information: Investors in debt securities have less data than issuers. Rating agencies can facilitate the flow of information from potential borrowers to investors.
- Dissemination of information: Rating agencies provide ratings and opinions of creditworthiness, outlooks, and rating reviews to the public.
- Use for regulatory purposes: The discrimination function of ratings is incorporated in financial market regulation (i.e. Basel II and Basel III).
- Enhancement of liquidity: Due to reducing information asymmetry, ratings support the financial market's liquidity.
- Market efficiency: Rating agencies help ensure the accuracy of prices with the public dissemination of information.
- Allocation of investment: Ratings help investors allocate their funds to achieve better return rates; borrowers can allocate sources more effectively.
- Cost of capital: Ratings provide information on relative risk; thus, the higher the risk, the lower the cost of capital for borrowers.
- Benchmarking internal systems: Ratings are used to benchmark internal rating systems and validate their performance.
- Evaluation of the overall economy: Sovereign rating can assess the economy's overall health.

On the other hand, rating agencies' activities may also be linked with problems and malfunctions, such as potential conflict of interest, lack of transparency, rating methods, pro-cyclicality, market structure and competition, and overreliance on ratings or lack of accountability. Regulatory authorities in many countries are aware of these potential failures and have adopted some rules to strengthen oversight, especially after the global financial crisis of 2008 and 2009.

Credit rating agencies assess the creditworthiness of various borrowers such as corporations, governments, and municipalities, and they assign a rating to debt securities, including bonds, commercial papers or structured products. Although some rating agencies are available in different countries, there are three agencies with global importance: Moody's Investors Service, Standard & Poor's Ratings Group, and Fitch Ratings, hereafter referred to as Moody's, S&P, and Fitch. These three rating agencies developed rating systems extensively used for risk monitoring and management by financial institutions, governments, investors and other market participants. Since their ratings are also incorporated in financial institutions' regulation in many countries, they have a very responsible and essential role in the financial markets. Thus, knowledge and understanding of the credit rating industry, the principles of credit rating, and the rating assessment system can be advantageous in making investments and financing decisions. This chapter's primary purpose is to define rating as used by credit rating agencies, summarize the main benefits and costs of rating, and discuss current issues in the rating industry, emphasising the regulation of CRAs in the European Union.

2.7.1 Definition of Rating

This section will focus on the basics and definition of external ratings published by the two most critical global agencies, Moody's and Standard & Poor's. Since these CRAs issue ratings internally recognized by investors, financial institutions and other participants in the financial market, this section also covers the basics of rating, the meaning of rating symbols, and comparison. Both agencies use similar rating systems; however, we can find differences in their rating definitions. For this reason, the essentials of their rating grading system will also be described in this chapter.

The system of rating securities originated by John Moody in 1909 when a simple method for assessing creditworthiness was developed. Today, both global agencies with the most significant market share, Moody's and Standard & Poor's, issue ratings assigned on long-term and short-term rating scales as forward-looking opinions of the relative credit risks of financial obligations issued by various entities. Generally, long-term ratings are assigned to issuers or obligations with an original maturity of one year or more. Short-term ratings are assigned to commitments with an original maturity of thirteen months or less (Moody's) or those obligations considered short-term in the relevant market (S&P). In the case of Moody's, ratings reflect both the likelihood of a default on contractually promised payments and the expected financial loss in the event of default. A forward-looking approach is also emphasised in the Standard &Poor's (S&P) rating definition. The issue rating is considered an opinion about an obligor's creditworthiness to a specific obligation, a particular class of financial obligations, or a specific financial program (Moody's Corporation, 2017; S&P Global Market Intelligence, 2016; S&P Global Ratings, 2017).

The agencies assign ratings using letters to distinguish between more or less risky obligations, both on the long-term and short-term rating scales. Table 2-5 shows the symbols and basic meaning of long-term rating scales for **issue rating**. Both agencies distinguish between investment and speculative grades, and definitions for issue rating are relatively similar. Furthermore, both agencies use more detailed classification within the global ranks, for example, Aa1 or AA+, to show their relative standing in the category.

Table 2–5 Long-term issue rating scales

Moody's		S&P	
Aaa	The highest quality, the lowest level of credit risk	**AAA**	The highest rating, extremely strong capacity to meet the obligation
Aa	High quality, very low credit risk	**AA**	Very strong capacity to meet the obligation
A	Upper-medium grade, low credit risk	**A**	Susceptible to adverse effects of changes in circumstances and economic conditions, still strong capacity to meet the obligation
Baa	Medium-grade, moderate credit risk, certain speculative characteristics	**BBB**	Adverse effects of changes in circumstances and economic conditions are likely to lead to a weakened capacity of the obligor to meet the obligation
Ba	Speculative, substantial credit risk	**BB**	Speculative, less vulnerable to non-payment than other speculative issues
B	Speculative, high credit risk	**B**	Speculative, more vulnerable to non-payment
Caa	Speculative of poor standing, very high credit risk	**CCC**	Speculative, currently vulnerable to non-payment, the obligor is not likely to have the capacity to meet the obligation
Ca	Speculative, likely in, or very near, default, with some prospect of recovery of principal and interest	**CC**	Speculative, currently highly vulnerable to non-payment, anticipated default
C	The lowest rated, typically in default, with little prospect for recovery of principal or interest	**C**	Speculative, currently highly vulnerable to non-payment, lower ultimate recovery

Source: S&P Global Market Intelligence (2016), Moody's Corporation (2017), author

When compared to issue rating, the **issuer ratings** provide a forward-looking opinion about the obligor's overall creditworthiness (S&P) or entities' ability to honour senior unsecured debt and debt-like obligations (Moody's). According to Moody's, issuer ratings do not incorporate support agreements, such as guarantees, that apply only to specific senior unsecured financial obligations and contracts.

Furthermore, ratings are used to measure the credit risk associated with a variety of obligations or entities, for example (Moody's, S&P):

- Bank deposits,
- clearing counterparties,
- corporate families,

- countries,
- credit default swaps,
- insurance companies,
- municipalities,
- structured finance counterparty instruments,
- structured finance counterparties,
- structured finance interest-only securities.

Both agencies also use **national scale ratings** for the opinions of issuers' relative creditworthiness and financial obligations within a particular country. They are not designed to be compared among states; conversely, they address relative credit risk within a given country. Thus, they are opinions of an obligor's creditworthiness or overall capacity to meet specific financial obligations relative to other issuers and issues in a given country or region. The notation of national scale rating is based on the characters that indicate the state in which the issuer is located, for example, 'Aaa.br' (Moody's), or 'brAAA' (S&P) ratings demonstrate the strongest creditworthiness relative to other domestic issuers in Brazil.

National scale systems are maintained only for some countries. We can use the case of the Czech Republic to demonstrate national scale ratings and their use (Table 2-6). There are municipal ratings in the table, both long-term and national scale ratings. Although most long-term ratings are A2, there is a relative difference among issuers according to national scale ratings within the country.

Table 2–6 Long-term and national scale municipal ratings in the Czech Republic

Issuer	**Long-term rating**	**National scale rating**
Ceska Lipa, City of	A2	Aa2.cz
Klatovy, City of	A2	Aa2.cz
Liberec, City of	Baa1	A3.cz
Liberec, Region of	A2	Aa3.cz
Prostejov, City of	A1	Aa1.cz
South-Moravian Region	A2	Aa3.cz
Trebic, City of	A2	Aa2.cz
Uherske Hradiste, City of	A2	Aa3.cz
Usti, Region of	A2	Aa3.cz
Zdar nad Sazavou, City of	A2	Aa2.cz

Source: Moodys' Corporation (2017), author

Without considering some current issues in the credit rating industry and the problem of misleading some ratings, they generally provide a handy tool for all participants in the financial market. There are many rating users in the financial market; however, the most important rating beneficiaries are investors and issuers. For **investors**, ratings provide an easily understandable and reliable guide about the likelihood of issuer default on a particular fixed-income instrument. Thus, rating offers the basis for investment decisions. In addition, since rating increases knowledge and transparency, it can reduce uncertainty and provide a benchmark for comparisons.

2.7.2 Benefits and Costs of Rating

The main roles of rating for investors are as follows (Nye, 2014):

- Ratings can help long-term investors (e.g. pension funds, insurance companies) to evaluate various long-term investment options.
- Ratings provide the basis for investors to make a more informed investment decision and to match the relative credit risk or debt with their risk tolerance.
- The rating system allows the issuer’s credit fundamentals to be compared against the industry peer group.
- Ratings reduce investors’ costs of gathering, analysing and monitoring borrowers' financial positions.

Ratings also have benefits for other market participants, such as debt **issuers**. Issuers such as corporations, financial institutions, governments, cities and municipalities use credit ratings to provide independent views of their creditworthiness and the credit quality of their debt issues. Thus, ratings play a useful role in enabling issuers to raise money in the capital markets and facilitate the process of issuing and purchasing bonds by providing a measure of relative credit risk. In addition, debt issuers that acquire ratings may benefit from:

- A broader base of investors, lenders, customers or business partners,
- alternative financing options,
- a standard measure of creditworthiness relative to other issuers in an industry or a country,
- lower financing costs and more dependable access to liquidity,
- a more accurate estimate of borrowing costs,
- a useful discipline on senior management (Nye, 2014).

There are additional benefits of ratings for an issuer that is a corporate enterprise. These include lower financing costs, better negotiation power with banks, the basis for offering debt issues in the market, higher visibility and credibility or a peer benchmark. For commercial banks, rating gives confidence to depositors, regulators and may help attract equity investors. If the issuer is a sovereign government, a rating helps attract international capital, support the development of local capital markets, meet international standards of transparency and cooperation or achieve more excellent funding stability (Nye, 2014).

Even though ratings play a crucial role in the financial market, we should also consider some rating scales' weaknesses. The disadvantages of rating are related to the issues discussed in Chapter 2.9, primarily the negative influence of the high concentration of the CRA market on financial stability, conflicts of interest, and transparency. The lack of accountability in the rating industry may lead to the limited reliability of rating scales and consequently to increased inefficiency in the financial markets. As mentioned in the previous chapter, rating scales are easy to read and use; however, there might also be a potential problem when unqualified users use and interpret these scales. It is essential to realize that ratings are opinions about credit risk and do not provide an absolute measure of default probability.

The overall credit rating industry is based on the scheme that issuers pay the credit rating agency. From the point of view of new issuers, agency fees should be considered, and the issuer should conduct a cost-benefit analysis of acquiring a rating. Typically, agency fees are negotiable, and issuers can discuss fee options with the agencies. According to Nye (2014), competition among CRAs keeps prices relatively modest, depending on the rated entity. For example, according to the S&P[7], the minimum fee for industrial corporations and financial service companies is $150,000.

2.8 Description of Credit Rating Process

The credit rating process typically involves several steps, requiring close cooperation between the rating agency and the issuer. First, new issuers must choose and contact a rating agency that fits their needs. Second, the rating process begins with an initial meeting to introduce the issuer's approach, methodology, and products. Finally, the steps of the rating process (Figure 2-9) include the following activities:

- The issuer signs an application,
- rating agency analysts are assigned to the customer to review the relevant information,
- analysts meet with the management team to review and discuss information,
- analysts evaluate information and propose the rating to a rating committee,
- the committee reviews the lead analyst’s rating recommendation and then votes on the credit rating,
- the issuer is generally provided with a pre-publication rationale for its credit rating for fact-checking and accuracy purposes,
- a press release announcing the public rating is typically published,
- the rating is kept current by identifying issues that may result in either an upgrade or a downgrade (S&P Global Ratings, 2017).

[7] Global Ratings U.S. Ratings Fee Disclosure [online]. Available at: http://www.standardandpoors.com/, [Accessed 19 September, 2017]

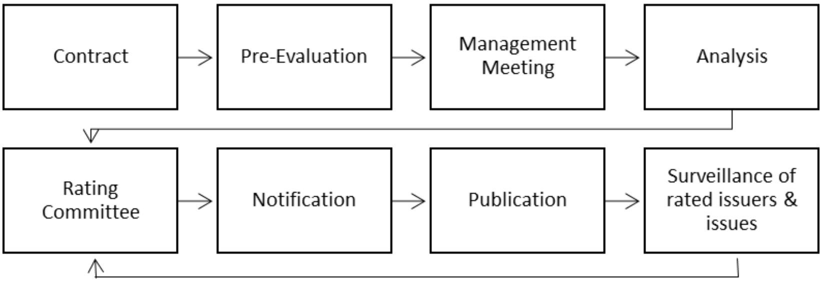


Figure 2–9 Steps of rating process

Source: S&P Global Ratings (2017), author

During the rating process, the issuer, an industrial company, provides relevant financial and non-financial information. Then, the discussion at the management meeting generally focuses on the following subjects[8]:

- Background and history of the entity,
- industry and sector trends,
- the national political and regulatory environment,
- management policies, experience, track record, attitude toward risk-taking,
- management structure,
- primary operating and competitive position, corporate strategy and competitive philosophy,
- debt structure, including structural subordination and priority of claim,
- financial situation and liquidity sources (cash flow, operating margin, a balance sheet analysis of debt profile and maturity).

As we can see, assessing companies' creditworthiness is a complex task based on both **quantitative** and **qualitative** analysis. CRAs emphasize the qualitative side, arguing that no ratios can capture the full complexity of a company's financial position, cash available to meet its future obligations, and management's willingness to pay principal to bondholders. Above all, CRAs evaluate long-term fundamentals related to the company's credit strengths and weaknesses related to its ability to generate cash, plausible stress scenarios, and elements of future cash flow (Nye, 2014). Rating agencies use different assignment methodologies according to the counterparty's nature (i.e. corporations, countries, public entities) and the nature of products (i.e. bonds, structured finance). Since the empirical analysis in the application part focuses on corporate borrowers, the primary attention is concentrated on risk factors for corporate ratings or individual debt issues specifically.

[8] Moody's Investors Service (2017)

Analysts of CRAs typically start with evaluating the issuer's creditworthiness before assessing the credit quality of a specific debt issue. The main factors of rating assessment described in the previous text consider financial and non-financial factors, including key performance indicators, economic, regulatory and geopolitical influences, management and corporate governance attributes, and competitive position. In a rating of an individual debt issue, analysts primarily focus on (S&P Financial Services, 2014):

- The terms and conditions of the debt security, including its legal structure,
- the relative seniority of the issue concerning the issuer's other debt issues and priority of repayment in the event of default,
- the existence of external support or credit enhancements (i.e. letters of credit, guarantees, insurance or collateral).

The critical factors of assessing the corporate rating can be summarized using the **rating analytical pyramid** (Figure 2-10).

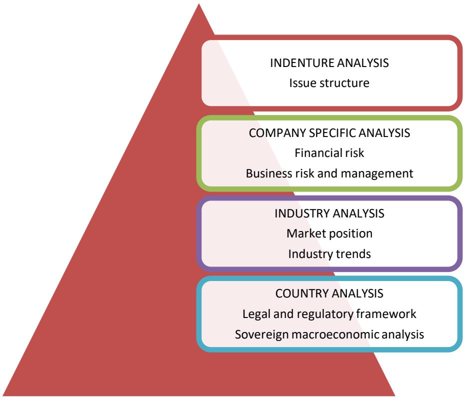


Figure 2–10 Rating analysis pyramid

Source: Moodys' Corporation (2017), author

Rating agencies can use both top-down and bottom-up approaches to assess the rating, with the main parts of the assessment focusing on the country, industry, company, and indenture analysis.

Credit ratings can be upgraded, downgraded, or unchanged, and the percentage of unchanged ratings can be used to measure rating stability. According to the rating changes, we can calculate the transition frequency from one rating class to another. Thus, we can assess the migration risk. The entire possible states a rating can take over a given time horizon is usually called rating transition or migration matrix. Transition matrices are based on time series of rating changes, and they can be estimated using various approaches. For example, Gunnvald (2014) use the Markov chain method and the duration method, Hu et al. (2002) use the combination of a Bayesian approach and ordered probit estimate of sovereign ratings, and Koopman et al. (2008) use an intensity-based duration model.

Frydman and Schuermann (2008) applied Markov mixture models or a mixture of two Markov chains. Their results show that a firm’s rating depends on its current credit ratings and its past rating history.

Transition matrices are published by CRAs and can be used to analyse transition rates and rating behaviour. For example, according to S&P (2015), investment-grade rated issuers exhibit greater rating stability (measured by the frequency of rating transition) than speculative-grade issuers. It is confirmed by the one-year transition rates (Table 2-7): The lower the rating, the lower the rating stability. For instance, while 93.26% of AA issuers are still rated as AA one year later, 79.97% of BB issuers stay within the BB rating category in one year. The rating category AAA is relatively stable; however, it should be considered that because the number of issuers with AAA ratings is typically deficient, the downgrade of even one issuer may have a large effect on this category’s rating stability.

Table 2–7 One-year corporate global transition rates (2015, in %)

From/to	**AAA**	**AA**	**A**	**BBB**	**BB**	**B**	**CCC/C**	**D**	**NR**
AAA	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
AA	0.29	93.26	4.40	0.00	0.00	0.00	0.00	0.00	2.05
A	0.00	1.43	89.87	5.48	0.00	0.00	0.00	0.00	3.23
BBB	0.00	0.06	3.12	85.52	4.90	0.00	0.00	0.00	6.40
BB	0.00	0.00	0.00	3.63	79.97	6.87	0.24	0.16	9.13
B	0.00	0.00	0.00	0.15	3.58	76.04	4.57	2.39	13.27
C/CCC	0.00	0.00	0.00	0.00	0.00	5.85	49.71	25.73	18.71

Source: S&P (2015), author

The level of rating stability depends not only on the rating category but also on the time horizon. Over the long term, rating stability is also consistent with higher ratings; however, the average long-term transition rates are usually higher than one year. For example, from 1981 to 2015, AAA-rated issuers were still rated AAA one year later 87.1% of the time, while CCC/C ratings remained CCC/C 44.2% of the time (Table 2-8).

The category denoted as D means payment default on one or more of an obligor’s financial obligations (according to rating agency’ definition of default), or issuer’s filing for bankruptcy, and NR stands for not rated.

The analysis of credit migration is the basis from the CreditMetrics approach developed by JP Morgan, the U.S. bank in 1997, subsequently revised as RiskMetrics, Inc. and acquired by MSCI in 2010. In practice, banks widely use this approach to estimate the full one-year forward distribution of any bond or loan portfolio values, where the changes in values are only related to credit migration. The critical assumption is that rated bonds' past migration history accurately describes the probability of migration in the next period (Crouhy et al., 2014).

Table 2–8 Long-term corporate global transition rates (1981 – 2015, in %)

From/to	AAA	AA	A	BBB	BB	B	CCC/C	D	NR
AAA	87.08	9.00	0.53	0.05	0.08	0.03	0.05	0.00	3.18
AA	0.53	86.69	8.06	0.53	0.06	0.02	0.02	0.02	4.02
A	0.03	1.81	87.65	5.39	0.33	0.02	0.02	0.06	4.58
BBB	0.01	0.11	3.55	85.43	3.82	0.12	0.12	0.19	6.24
BB	0.00	0.03	0.13	5.08	76.78	0.64	0.64	0.73	9.63
B	0.00	0.03	0.09	0.21	5.25	4.39	4.39	3.77	11.99
C/CCC	0.00	0.00	0.14	0.20	0.61	44.19	44.19	26.36	15.66

Source: Source: S&P (2015), author

2.9 Current Issues in Rating Industry

CRAs produce ratings in both local and international markets. The credit rating industry is regulated by Regulation (EC) No 462/2013[9] and Directive 2013/14/EU[10] in the European Union. This regulation (hereafter referred to as EC Regulation) aims to regulate the activity of credit rating agencies to protect investors and European financial markets against the risk of malpractice. The first incentives to strengthen the regulatory and supervisory framework of CRAs in the EU and the first set of rules came into effect at the end of 2009, connected with the global financial crisis 2008 – 2009. To be registered in the European Union (EU), credit rating agencies must (EC, 2017):

- Avoid conflicts of interest: For example, credit rating analysts must not rate an entity in which they have a holding;
- ensure the quality of their ratings and rating methods;
- provide a high degree of transparency, for example, by publishing an annual transparency report.

In addition to the regulatory framework, the European Securities and Markets Authority (ESMA) was created in 2011 to supervise CRAs registered in the EU. Since July 2011, ESMA has been responsible for registering credit rating agencies and has exclusive supervisory powers concerning such agencies. Under the CRA regulation, it is possible for a rating agency established outside the EU to have its rating recognised and used for regulatory purposes in the EU. It can happen in two ways: certification through the equivalence regime and endorsement. The equivalence certification applies to the CRAs established and supervised outside the EU that have no affiliation in the EU. These CRAs that meet the requirements of the regulation may apply to the ESMA for certification. The endorsement regimes apply to CRAs affiliated with or working closely with EU-registered agencies, and they must also comply with specific legal requirements. The list of registered or certified CRAs in the EU is available in Appendix 1. For example, we can see that as of 24 March 2022, there are 33 CRAs regulated in the EU.

[9] Regulation (EC) No 462/2013 of the European Parliament and the Council of 21 May 2013 on credit rating agencies

[10] Directive 2013/14/EU of the European Parliament and the Council of 21 May 2013

Regarding the critical provisions of the last EC Regulation reforms, a recent study on the state of the credit rating market was published in 2016 (EC, 2016). According to this document, hereafter referred to as Report, the current credit rating market has an oligopolistic structure, and it is dominated by the three global CRAs: Moody's, S&P and Fitch. The EC Regulation on credit rating agencies is aware that CRAs play a crucial role in the global securities and banking markets. A significant number of financial institutions use their credit ratings to estimate their capital requirements for solvency purposes or evaluate risks. It is suggested that these entities make their credit risk assessment and thus less reliance on credit ratings.

Concerning credit rating agencies, the main issues that concern the European Commission include (EC, 2016):

- Conflicts of interest due to the issuer-pays model,
- conflicts of interest due to the remuneration model of credit rating agencies,
- disclosure for structured finance instruments,
- transparency,
- procedural requirements and the timing of publication specifically for a specific period.

Thus, specific measures have been adopted in the European Union to improve the situation. For example, the Credit Rating Agency (CRA3) Regulation's last reforms focused on enhancing competition in the credit rating market, further addressing conflicts of interest, enhancing disclosure on structured finance instruments, and the rotation mechanism for CRAs (EC, 2017).

2.9.1 Market Concentration

As said in the Report, market shares based on total revenues and the Herfindahl-Herfindahl Index (HHI)[11] suggest that the market of CRAs is highly concentrated, both overall and at the individual product category level, with a small increase of concentration between 2012 and 2014. The summary of market shares for credit rating activity and ancillary services provided in the European Union can be seen in Table 2-9. The three largest CRAs cover most of the market, reaching around 96% during the reference period, while the remaining share, about 4%, is filled by the other, much less critical CRAs.

[11] According to the report, HHI provides an indication of concentration within markets, with an HHI over 1,000 generally considered to be concentrated.

Table 2–9 Market shares for credit rating activity and ancillary services

CRA	**2012**	**2013**	**2014**
Moody's	36.69%	36.38%	36.99%
S&P	32.88%	36.00%	38.43%
Fitch	17.67%	16.47%	18.40%
Euler Hermes Rating GmbH	0.24%	0.26%	0.25%
Feri EuroRating Services AG	0.84%	0.78%	0.76%
BCRA-Credit Rating Agency AD	0.02%	0.03%	0.00%
Creditreform Rating AG	0.51%	0.54%	0.51%
Scope Ratings AG	0.10%	0.20%	0.17%
GBB-Rating Gesellschaft für Bonitätsbeurteilung GmbH	0.34%	0.34%	0.33%
ASSEKURATA (Assekuranz Rating-Agentur GmbH)	0.30%	0.31%	0.27%
ARC Ratings, S.A. (previously Companhia	0.04%	0.03%	0.02%
AM Best Europe-Rating Services Ltd. (AMBERS)	0.75%	0.73%	0.47%
DBRS Ratings Limited	0.82%	1.23%	1.39%
CRIF S.p.A.	0.35%	0.77%	0.07%
Capital Intelligence (Cyprus) Ltd	0.00%	0.00%	0.03%
European Rating Agency, a.s.	0.00%	0.00%	0.00%
Axesor SA	1.85%	1.41%	0.73%
The Economist Intelligence Unit Ltd	6.48%	4.41%	1.05%
Dagong Europe Credit Rating Srl (Dagong Europe)	0.01%	0.01%	0.01%
Spread Research	0.09%	0.09%	0.13%
EuroRating Sp. z o.o.	0.01%	0.01%	0.00%
HHI-index	*2,787*	*2,916*	*3,189*

Source: EC (2016)

2.9.2 Conflicts of Interest

CRAs, similarly to other financial intermediaries, play a crucial role in the financial system because their expertise in interpreting signals and collecting information from their customers gives them a cost advantage in producing information. Thus, CRAs positively affect the problem of asymmetric information in the financial market because borrowers have some information they do not disclose to lenders. Hence, one party in the contract often does not have enough information to make accurate decisions. However, there is a potential problem of conflicts of interest: One party in a financial agreement may have incentives to act in their interest rather than in the interest of the other party. Since CRAs provide very specialized, usually multiple financial services, conflicts of interest may arise in the rating industry; see, for example, Mishkin and Eakins (2009) or Cecchetti and Schoenholtz (2015).

Conflicts of interest can substantially reduce the quality of information and even increase asymmetric information problems. For this reason, the next issue of the Report is the problem of conflicts of interest arising from the fact that while investors and regulators demand a well-researched credit quality assessment, issuers need a favourable rating. Since the issuers of securities pay a rating firm to have their securities rated, investors and regulators may question their rating quality (Mishkin and Eakins, 2009). Other conflicts of interest might be between CRAs and shareholders and CRAs on a firm level and its employees (e.g. rating analysts). They are linked to ancillary consulting services provided by CRAs in terms of auditing and consultancy services. CRAs may deliver favourable ratings to attract more clients and thus increase asymmetric information in financial markets (EC, 2016). Another problem of conflicts of interest might be connected to structured debt instruments (Efing and Hau, 2015).

The Report (EC, 2016) highlights that conflicts of interest are related to competition in the industry. There are two phenomena linked to this problem and discussed in the literature (EC, 2016, p. 72):

- Rating shopping: In this situation, issuers solicit ratings from multiple agencies and then choose the most favourable one. This phenomenon can lead to rating inflation.
- Rating catering: A situation strictly related to rating shopping. CRAs may be incentivised to loosen their standards to compete with more favourable ratings from other CRAs, usually in booming markets, when CRAs have fewer concerns about their reputation and market shares (Sangiorgi and Spatt, 2013).

2.9.3 Transparency

If the operating activities and actions of CRAs are easy for other participants to see and understand, then the CRAs are considered transparent. As Nye (2014) suggests, the rating agencies have become a potent force in demanding transparency in the accounts of issuers they rate since the Asian financial crisis of 1997 – 1998. Thus, CRAs emphasised more precise data, more open accounting, and conforming to international standards to increase their reputation. However, during the financial crisis of 2008 – 2009, rating agencies failed to publish accurate and adequate ratings in many cases, especially in assessing structured products. Due to the Basle rules, there was a strong demand for high ratings by financial institutions that are not permitted to hold assets with lower rating grades. For example, as Crotty (2009) says, in 2005, more than 40% of Moody’s revenue came from a rating of securitized debt. Due to financial boom, inexperience or ignorance, CRAs issued absurdly high ratings to illiquid, non-transparent structured financial products. While the explosion of these innovative securities created large profits at financial institutions, it also destroyed the transparency necessary for any aspect of market efficiency.

Although CRAs aim to improve their credibility and transparency, there is still a lack of accountability in the credit rating industry. Some arguments provided (e.g. Pagano and Volpin, 2010) that issuers should disclose all the information relevant to assessing the products' risk instead of requiring CRAs to disclose the information they used. On the other hand, there are also arguments regarding transparency (i.e. the disclosure requirements for CRAs) rather than the issuers' disclosure. Avgouleas (2009) argues that such disclosures are adequate if useful and easily understood. Pagano and Volpin (2010) claim that increased transparency may lead to better market decisions. Still, there is a problem that the amount of time and resources needed to make these decisions is limited to the strongest players in the market. As a result, sophisticated investors have an advantage over a naïve majority.

In the EU, there are incentives to increase transparency in the rating industry. For example, according to the EC Regulation (Article 8B), the issuer, the originator and the sponsor of all structured finance instruments established in the EU to jointly publish information on these instruments and the performance of their underlying assets (EC, 2016). It is assumed that the information available to investors could reduce their reliance on CRAs' ratings; prices of rated products could be more informative and facilitate a better capital allocation.

2.9.4 CRAs Rotation

The rotation mechanism refers to the mandatory rotation between a credit rating agency and an issuer. This relatively new issue in the credit rating industry is discussed and proposed in the Report; however, there is still insufficient empirical evidence on potential effects. The rotation system has been used in the audit sector as the practice of mandatory changes in auditors to prevent overfamiliarity that could lead to misstatements and misrepresentation in financial accounts (Hodgson, 2016).

However, there are essential differences between the audit and credit rating sectors (EC, 2016):

- Auditors have perceived themselves as protectors of the public trust. They are subject to state licensing and regulation; CRAs have perceived themselves as publishers and have defended their opinions in courts as free speech.
- Auditors provide a statement on backwards-looking information; CRAs provide forward-looking information and analysis.
- CRAs have more substantial incentives to generate short-term profits.

For example, the EC Regulation introduced a common approach to regulation and supervision within the EU. CRAs shall not enter into a new contract for ratings on the same originator's assets for a period equal to the previous contract's duration but not more than four years (ESMA, 2017). The EC provisions should change the time horizon of the CRAs, influence their cost structure and affect CRAs' sectoral knowledge. The main objective of rotation is to reduce the efforts of CRAs to establish a long-lasting relationship, enhance their independence and reduce the incentives to inflate ratings.

2.10 Chapter Summary

As shown in this chapter, the measurement and evaluation of credit risk is a central topic of financial management. Two approaches can be used to measure credit risk: macroeconomic and microeconomic. The macroeconomic approach is based on common national and sometimes international regulations and is determined mainly by the quality of the portfolio and the size of the exposures. The parameters probability of default (PD), loss given default (LGD) or exposure at default (EAD) are used as basic indicators.

On the other hand, the microeconomic approach to credit risk measurement is based on the internal regulations of financial institutions and the monitoring of basic parameters so that it is possible to take the necessary measures in time and thus prevent non-payment of debts, particularly loans. A typical example of this approach is personal credit scoring or rating agency services.

Under the main focus of this monograph, the next chapter of this work will first be devoted to the purpose and use of micro models of credit risk, including an overview of current research. Subsequently, selected econometric approaches used in this monograph's application parts will be described.

Chapter 3

Approaches for Credit Rating and Corporate Bankruptcy Modelling

This chapter and the rest of the work focus on applying selected models within the micro approach to credit risk measurement. For this reason, in this section, we first focus on the main motivation of our research and the possibilities of using rating or survival models. In the following section, the econometric models used in the application part of the work are mentioned and briefly described.

The structure of this section is as follows. Firstly, the main role of credit rating and survival models is clarified, including an overview of current research. Finally, the main principles of discriminant, logistic regression and survival analyses are provided. In this chapter, the methods are described with appropriate references to additional literature, as this book mainly focuses on applying models and their use.

3.1 Introduction and Research Background

Based on the previous chapter, it is clear that the rating assesses the degree of credit risk. Therefore, it is necessary to have data containing ratings for selected companies or instruments to estimate rating models that will replace this role in a certain way.

The following sections provide the primary arguments for why rating and corporate survival models are needed. Firstly, we focus on the practical use of rating models and, based on the current research, summarize the main approaches for modelling. Then, we provide some research reviews on rating and scoring models. Finally, we present some background for corporate survival modelling.

3.1.1 Why Micro Credit Risk Models

The concept of rating assessment is not new; however, this area has attracted massive attention in the last decades, especially during and after the 2008 – 2009 financial crisis. An increasing need to actively and effectively manage credit risk across various sectors of the economy has resulted in sophisticated and readily available tools and techniques.

Rating agencies play an essential role in the financial markets, and their product – rating – is generally accepted as the default risk assessment criterion. Rating is publically available for market participants, and its usage is simple. Anyone with basic economic education can “read” and interpret ratings and changes in rating grades. The global CRAs such as Moody’s, Standard & Poor’s and Fitch are considered important international players. Although the general concept and rating description are analogical, minor differences depend on each rating agency. Thus, rating agencies' rating assessments are not directly comparable; for example, Cantor and Packer (1996). Resti (2002) states that, unlike the U.S., agency ratings are not widely adopted in continental Europe, and the demand for external ratings comes mostly from financial companies. Since there is a lack of rating information in the financial markets, particularly in emerging markets, a need for parsimonious models arises. The contribution of own credit models is, for example, evaluated by Rerolle and Rimaud (2009). As they suggest, research in a credit risk area and credit models have important value-added compared with the certified rating because it enables one to react to changes and new information sooner than in the case of complete dependency.

Companies with agency ratings are traded on capital markets, quoted and listed on stock exchanges. However, there is a lack of rating information in some countries, such as Central and Eastern Europe (CEE). Although the capital markets of CEE countries have made remarkable progress during the last decades, the tradition of the bank-based economy and a relatively low number of issues still result in a relative absence of rating information. Thus, the lack of ratings accentuates a need for inherent economic models. Such models' main principle is to give the necessary information about the company’s operating characteristics, identify alert factors, and assign a corresponding rating. With this assessment, comparisons among companies can be made, and problematic companies' detection can be done. Thus, the models enable researchers, analytics, or creditors to signal potential companies with higher default risk or bankruptcy probability. In addition, since the rating models are quantitative models that typically use publicly available information, mostly financial statements, we can assess unquoted companies, compare them according to the credit risk, or identify those with the highest probability of default.

Some credit agencies focus almost exclusively on quantitative data, which they incorporate into a mathematical model. Thus, a well-estimated model partially fulfils the role of rating agencies. We should point out the term ‘partially’ because rating models cannot fully substitute rating agencies. For example, the agencies incorporate other factors into the rating; they consider an economic development perspective and country forecasts or world-leading economic indicators. In addition, rating agencies consider other analytic areas such as management’s reputation, reliability, experience, and past performance. While rating agencies emphasise that financial and non-financial factors matter in predicting default and bond ratings, the academic literature has focused on economic indicators' ability to predict ratings. However, recent studies have shown that other factors, such as the state and perspective of the overall economy or the industry characteristics, may improve rating models' predictive ability.

The credit rating models are typically based on historical performance and the credit risk event, for example, default, bankruptcy or the change of credit rating assessment. Therefore, a crucial step in any credit risk modelling is determining the probability of default. For example, de Laurentis et al. (2010, p. 6) point out the following approaches for the estimation of credit risk models:

- The observation of historical default frequencies of borrowers' homogenous classes (for example, assigned ratings and default rates observed ex-post per rating class),
- the use of mathematical and statistical tools, models are based on large databases and enable an ex-ante measure of expected probability,
- hybrid methods that combine both judgmental and mechanical approaches (qualitative aspects correct quantitative results),
- the use of an entirely different system extracting the implicit probability of default embedded in market prices (securities and stock).

The purpose of models based on traditional statistical approaches is to distinguish between 'good' and 'bad' borrowers in terms of their creditworthiness in an automated way. Financial institutions widely use these techniques for retail or SME loans, and scoring models typically represent them. The main principles of scoring models can also be applied to rating to distinguish between 'more' and 'less' credit risk in rating categories. Although some authors argue that credit scoring is not a highly sophisticated approach (i.e. De Servigny and Renault, 2004), it is undoubtedly widely used and suitable for assessing the credit risk applied to borrowers. With new technologies and financial innovations such as Fintech, the importance of rating models is rising. However, risks are still involved besides advantages such as higher interest rates or lower fees than bank loans. Whether crowdfunding, peer-to-peer lending or borrowing, assessing credit risk becomes essential to all finance providers and investors.

Multivariate statistical models, sometimes referred to as the traditional way of credit risk modelling, are widely applied by rating agencies and financial institutions to assess the credit risk of bonds and corporate loans. Since they are based on a critical approach supported by specific vital financial numbers, they can be considered fundamental-based models (i.e. scoring and rating models). In addition to these traditional methods, alternative approaches, such as the Merton model, were developed to model and measure credit risk. As Crouhy et al. (2014) emphasize, these models complement and, to a degree, compete with traditional approaches to measuring credit risk and offer an independent check on judgment ratings.

Micro credit risk measuring approaches use fundamental-based or firm-specific credit risk models primarily utilising the company's financial reports. This approach's main principle is to find critical factors that predict the likelihood that cash flow generated from the firm’s assets will be sufficient to cover its debt obligations (Moody’s Analytics, 2012). This approach relies on historical values recorded and carried over from one period to the next, so historical values may differ from current or expected values. As a result, the company’s prospects may not be accurate in some cases. On the other hand, the advantage of these models lies in the fact that since financial reporting is subject to accounting rules, it increases the objectivity and comparability of financial ratios. Thus, fundamentally-oriented models are statistical-based models that are useful for assessing a borrower's creditworthiness or predicting default probability. In other words, the primary purpose of these models is to provide an automated rating or scoring system to assess credit risk.

3.1.2 Research Review

The first rating and scoring models were developed using the linear discriminant analysis suitable for separating groups. Different approaches can be used for rating modelling, and the choice depends on many factors, including who will use the models and for what purpose. Typically, the models are derived using logistic regression analysis methods, classification trees, random forests, or neural networks. Big data brings about new options for scoring and rating, which are typically based on the development of scoring algorithms. Financial institutions widely use these techniques for retail or SME loans, and scoring models represent them. However, the scoring model's main principles can also be applied to rating to distinguish between ‘more’ and ‘less’ credit risk. Though some authors may see credit scoring methods as not highly sophisticated, they are widely, easily used, and functional approaches for assessing the credit risk applied to any borrower type.

Naturally, many studies deal with model estimation; however, there is significantly more attention to scoring models than rating models. Scoring models will find application mainly in predicting credit default or, for example, the probability that a company will go bankrupt. Various modifications of these models have a wide application in corporate finance when assessing the company's financial health, especially in the banking sector, where the probability of default in loan repayment is analysed. Rating models work similarly but have slightly different uses. Their application is important, especially where a certified rating is not available. The analyst has a reason or needs to examine or estimate the rating of the instrument or entity. These models can play a significant role in valuing bonds not traded in the public market and can be an attractive asset for investment. In these cases, it is especially necessary to consider the risk of default carefully, and it is when the models can be used. Although their application is possible, this issue is not given significant attention in the literature, for example, compared to scoring models.

The rating modelling is a crucial financial topic, as it is related to assessing the credibility of the debtor or issuer of debt securities. Some research on bond rating dates back to the fifties of the twentieth century, for example, Hickman (1958) or

Fisher (1959). Regression analysis became one of the most used methods to estimate ratings in this period. An alternative approach to predicting bond ratings is the multiple discriminant analysis introduced, for example, by Pinches and Mingo (1973), Ang and Patel (1975), Altman and Katz (1976) and Belkaoui (1980). Other research compared particular statistical methods; e.g. Kaplan and Urwitz (1979) compare ordered probit analysis with ordinary least square regression, and Wingler and Watts (1980) compare ordered probit analysis with multiple discriminant analysis. Other studies replicate the process of bond rating model estimation and modify the approaches by considering new variables, such as the study of Chan and Jegadeesh (2004), or examine the impact of financial variables on credit rating for a given country or region, e.g. Gray et al. (2006). Recent studies come from the theoretical framework mentioned above and extend statistical methods for new non-conservative approaches such as neural networks (Dutta and Shekhar, 1988; Surkan and Singleton, 1990). For example, Waagepetersen (2010) assesses the relationship between quantitative models and expert rating evaluation. More recently, Altman et al. (2010) focused on the importance of non-financial information within risk management.

Research in rating prediction has shifted mainly to applying machine learning methods in recent years. Hsu et al. (2018) propose a model based on the artificial bee colony approach and support vector machine technique. The authors find that their bio-inspired computing mechanism improves prediction accuracy compared to other statistical methods. Golbayani et al. (2020) compare neural networks, support vector machines and decision trees. The authors find that the decision tree-based model achieves the best performance. In their research, they apply conventional accuracy measures and introduce the notch distance approach, which is suitable for comparing the performance of various machine learning methods. As it turns out, all the above techniques are appropriate for rating prediction. The individual models differ mainly in the methodology, used variables, and ability to predict the rating. Therefore, research in this area is focused primarily on improving the predictive accuracy of models. For example, Wang and Ku (2021) developed the parallel artificial neural networks model that creates several independent artificial neural networks. As the authors suggest, their approach achieved competitive results compared to conventional artificial intelligence techniques.

Current research shows that conventional approaches and newer methods based on artificial intelligence are widely used to model credit ratings. The huge advantage of these models is their practical applicability and the possibility of using them for potential rating revisions. In addition, research studies show that models' predictive power is sufficient and comparable to other commonly used methods in valuing historical data based on averages or growth rates. For example, Jones et al. (2015) examine the predictive performance of binary classifiers using a large sample of international credit ratings. They apply conventional techniques (logit and probit regression, linear discriminant analysis) and fully nonlinear classifiers (neural networks, support vector machines, general boosting, AdaBoost and random forests). The authors conclude that although the newer classifiers outperform others, simpler ones can be viable alternatives to more sophisticated approaches, particularly if interpretability is an important objective of predictive models.

The literature review shows that various studies are focused on modelling ratings using fundamental-based approaches. However, from the application point of view, these studies do not pay sufficient attention to Czech entities or companies from CEE countries. At the same time, knowledge of the rating and its influencing factors has utilization in many areas. It is primarily an assessment of the investment quality of unrated bonds. Another application is, for example, corporate finance and the issue of determining the cost of capital. In any case, the absence and unavailability of rating models are the main motivation for this study. The main benefit is applying CEE countries' data, including the application of models corresponding to Czech conditions and needs. The partial goal is to present the models' estimation process, interpretation, and mutual comparison.

An alternative way provided in this work to model ratings is based on time-to-event or survival analysis. For example, Glennon and Nigro (2005) use a discrete-time hazard framework to measure the default risk of small business loans. Roa et al. (2009) propose a survival analysis methodology to analyze falling rating duration. They test macroeconomic variables to predict this event in selected countries and find differences between developed and emerging economies. In further research, Zhang and Thomas (2012) compare linear regression and survival analysis for modelling recovery rates. The authors find that linear regression is better for recovery rate modelling; however, they suggest some adjustments and additional validation. Overall, survival analysis methods are typically used for rating or credit transitions and time series rating patterns (e.g. Parnes, 2007; Figlewski et al., 2012; Louis et al., 2013; Leow and Crook, 2014). Thus, we can model the rating behaviour over time and measure, for example, the probability of a certain change in the rating depending on time and other relevant variables. Therefore, survival analysis allows us better to understand the data and its dynamics over time. In addition, it is a method used by rating agencies to estimate default rates, which are regularly published and used by analysts and researchers in the financial market.

As part of credit risk measurement, attention is also heavily paid to bankruptcy prediction. The bankruptcy of companies is typically analysed based on credit score models, which are statistically derived models for predicting credit risk. Among all the studies on scoring models, the study by Altman (1968) and the model known as Altman´s or Z-score model are worth mentioning. Since the first publication of this model, extensive research has been conducted in bankruptcy prediction and the application of discriminant analysis, logistic regression, classification trees, and neural networks. In addition, the survival analysis approach can be seen as an alternative way to examine corporate bankruptcy. However, this area has not yet attracted adequate attention compared to the traditional methods mentioned above.

Nevertheless, some studies apply survival analysis to predict corporate failure in different countries. For example, the earlier research includes Lane et al. (1986), who employed the Cox model to predict bank failure using a sample of 130 banks. As the authors suggest, the overall accuracy of their model is similar to the discriminant analysis results. Among other studies, Laitinen and Kankaanpää (1999) discuss the six most popular alternative methods to financial failure prediction, including survival analysis. However, their research proposes no only one way, even though the accuracy of failure prediction varies depending on the technique applied. Other empirical results include the study by Agarwal and Audretsch (2001), who focus on the effect of companies' size on their survival. Their research finds that smaller companies are less likely to survive than larger companies. However, they suggest that general pronouncements are hazardous because the size changes over the industry cycle and with the technological demands of that industry.

Similarly, Glennon and Nigro (2005) examined the effect of time on the probability of default on medium–maturity loans under a loan guarantee program for small firms. The authors find that the default behaviour of these loans is time-sensitive. As loan seasons, the probability of default initially increases, and it declines after the second year. They also suggest that the likelihood of default is conditional on the borrower, lender, loan characteristics and changes in economic conditions. Finally, De Leonardis and Rocci (2008) used a discrete-time survival analysis approach to assess the default risk of small and medium-sized Italian companies from 1995-1998. The authors suggest that the prediction accuracy of the duration model is better than that provided by a single-period logistic model.

In addition to examining the effect of financial and economic factors on corporate failure, some studies assess the impact of other factors. For example, Mokarami and Motefares (2013) examine the effect of the internal mechanisms of corporate governance on the bankruptcy of firms enlisted in the Tehran Stock Exchange. Using the Cox model of survival analysis, the authors claim a significant relationship between CEO replacement and bankruptcy. Other research includes, for example, Pereira (2014), who applied the Cox proportion hazard model in predicting business failure of companies in the textile industry, and Kelly et al. (2015), who focused on corporate liquidations in Ireland. Louzada et al. (2014) modelled the time to default on a personal loan portfolio. They state that survival models are being proposed in financial risk management as alternative tools due to the continuous monitoring of risk over time. Their empirical study is illustrated by credit data from a Brazilian commercial bank. Their results show that attention should be paid to continuously checking the validity of requirements for using the available models. Besides the problems of loan default and bankruptcy, Kristanti and Isynuwardhana (2018) examined the effect of certain predictors on the probability of financial distress of companies enlisted on the Indonesia Stock Exchange. Applying the Cox hazard model of survival analysis, they found evidence that there is an inverse relationship between the control of corruption and the probability of financial distress, except for the predictors such as leverage, operational risk, and size.

Overall, there is a vast literature on predicting corporate bankruptcy using various techniques. However, there is still little attention to modelling corporate bankruptcy using time-to-event methods. This fact is one of the motivations of our research, which is to compare commonly used approaches with less frequently applied survival analysis methods. The main contribution of this monograph is the expansion of existing research in this area and the application of selected models to specific data from CEE countries, respectively, from the Czech Republic. The main goal is to find a link between rating and survival models, identify the main predictive variables and propose a procedure to convert bankruptcy rates into rating evaluations. As a result, the association between the probability of survival and the rating, or the predicted development of the rating over time, can be better understood. Furthermore, as the rating is widespread and used globally, we believe the interpretation of credit risk using the rating is more suitable for users, especially for individual investors.

All participants in credit contracts can use the main findings of this work. In addition, it is useful for analytical departments that create mathematical-statistical models for monitoring and measuring credit risk in banks and other financial institutions. Nevertheless, we see the main use on the part of retail investors, whether individuals or companies, who can use the partial results of the work in several directions, mainly for a better understanding of the factors that significantly influence the survival probability and, thus, the overall rating evaluation. Furthermore, the results of this work can be further used to apply selected models to their own data and their subsequent use to measure credit risk. Finally, this work can also be used in academic research as an example of connecting two different approaches to creating micro credit risk models and possibly expanding further.

3.2 Discriminant Analysis

Discriminant analysis is a standard statistical method used to separate groups and, thus, a suitable method for credit scoring or bond rating modelling. The analysis can be used for two primary objectives: first, the description of group separation; second, predicting or allocating observations to groups. Huberty and Olejnik (2006) distinguish between descriptive discriminant analysis (DDA) and predictive discriminant analysis (PDA). The purpose of DDA is usually the study of comparison among a certain number of groups, for each of which we have several outcome variable scores. However, suppose a single set of response variables is used as predictors, and there is a single grouping variable. In that case, the primary purpose is to analyse how well group membership of analysis units may be predicted using PDA. Correspondingly, Rencher (2002) differentiates between discriminant and classification functions. Discriminant functions separate groups, while classification functions assign individual units to one or more groups. In group separation, linear functions of variables describe the differences between two or more groups. The main objective is to identify the relative contribution of p variables to split. The latter problem is focused on the prediction or allocation of observations to groups, which is a common goal of discriminant analysis. A prediction rule then consists of a set of linear combinations of predictors, where the number of combinations reflects the number of groups.

Discriminant functions are linear combinations of variables that best separate groups, for example, the *k* groups of multivariate observations. The description of discriminant analysis and methods can be found, for instance, in Rencher (2002), Manly (2005), Huberty and Olejnik (2006), Tabachnik and Fidell (2007), Harrell (2010) or Hair et al. (2014). As Rencher (2002) suggests, linear functions of variables (discriminant functions) describe the difference between two or more groups for group separation. The goal is to identify the relative contribution of the *p* variables to separation and derive discriminant functions as linear combinations of variables that best separate groups.

Firstly, we assume the **discriminant function for two groups** (the following definitions and equations are taken from Rencher, 2002):

- The two populations to be compared have the same covariance matrix but distinct mean vectors μ_1 and μ_2,
- we assume samples $\mathbf{y}_{11}, \mathbf{y}_{12}, \ldots, \mathbf{y}_{1n_1}$ and $\mathbf{y}_{21}, \mathbf{y}_{22}, \ldots, \mathbf{y}_{2n_2}$ from the two populations,
- each vector $\mathbf{y}_{ij}$ consists of measurement on *p* variables.

The discriminant function is the linear combination of these *p* variables that maximizes the distance between the two (transformed) group vectors. Thus, a linear combination $z = \mathbf{a}'\mathbf{y}$ transforms each observation vector to a scalar:

$$z_{1i} = \mathbf{a}'\mathbf{y}_{1i} = a_1 y_{1i1} + a_2 y_{1i2} + ... + a_p y_{1ip}, \quad i = 1, 2, ..., n_{1,},$$

$$z_{1i} = \mathbf{a}'\mathbf{y}_{2i} = a_1 y_{2i1} + a_2 y_{2i2} + ... + a_p y_{2ip}, \quad i = 1, 2, ..., n_{2,}.$$

Then the $n_1 + n_2$ observations in two samples,

$$\begin{matrix} \mathbf{y}_{11} & \mathbf{y}_{21} \\ \mathbf{y}_{12} & \mathbf{y}_{11} \\ \vdots & \vdots \\ \mathbf{y}_{1n_1} & \mathbf{y}_{2n_2}, \end{matrix}$$

are transformed into scalars,

$$\begin{matrix} z_{11} & z_{21} \\ z_{12} & z_{11} \\ \vdots & \vdots \\ z_{1n_1} & z_{2n_2}. \end{matrix}$$

We find the means $\bar{z}_1 = \sum_{i=1}^{n_1} z_{i1} / n_1 = \mathbf{a}'\bar{\mathbf{y}}_1$ and $\bar{z}_2 = \mathbf{a}'\bar{\mathbf{y}}_2$, where $\bar{\mathbf{y}}_1 = \sum_{i=1}^{n_1} \mathbf{y}_{1i} / n_1$ and $\bar{\mathbf{y}}_2 = \sum_{i=1}^{n2} \mathbf{y}_{2i} / n_2$. Then, we find the vector **a** that maximises the standard difference $(\bar{z}_1 - \bar{z}_2) / s_z$. Finally, we use the squared distance $(\bar{z}_1 - \bar{z}_2)^2 / {s_z}^2$ so that the result is positive:

$$\frac{(\bar{z}_1 - \bar{z}_2)^2}{{s_z}^2} = \frac{\left[\mathbf{a}'\left(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2\right)\right]^2}{\mathbf{a}'\mathbf{S}_{\mathrm{pl}}\mathbf{a}}, \tag{3.1}$$

where $\mathbf{S}_{\mathrm{pl}}$ is the pooled covariance matrix and $n_1 + n_2 - 2 > p$. The maximum of (3.1) occurs when

$$\mathbf{a} = \mathbf{S}_{\mathrm{pl}}^{-1}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_1), \tag{3.2}$$

or when **a** is any multiple of $\mathbf{a} = \mathbf{S}_{\mathrm{pl}}^{-1}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_1)$. The maximizing vector is not unique; however, its direction, or the relative values or ratios of $a_1, a_2, \ldots, a_p$ are unique, and $z = \mathbf{a}'\mathbf{y}$ projects points **y** onto the line on which $(\bar{z}_1 - \bar{z}_2)^2 / {s_z}^2$ is maximized.

The linear discriminant analysis (LDA) can be used for more than two groups. Thus, we can extend the previous case for the study of **several groups**. The objective is to find linear combinations of variables that best separate the k groups of multivariate observations. We assume that for k groups with n_i observations in the ith group, we transform each observation vector $\mathbf{y}_{ij}$ to obtain $z_{ij} = \mathbf{a}'\mathbf{y}_{ij}$, where i = 1, 2,..., k and j = 1, 2,..., n_i. Then, we find the means $\bar{z}_{ij} = \mathbf{a}'\bar{\mathbf{y}}_i$, where $\bar{\mathbf{y}}_{\mathbf{i}} = \sum_{j=1}^{n_i} \mathbf{y}_{ij} / n_i$. Similarly, to the two-group analysis, we seek the vector **a** that maximally separates $z_1, z_2, \ldots, z_k$. In this case, the formula (3.2) will be extended for k-groups. Assuming that $\mathbf{a}'\left(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2\right) = \left(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2\right)' \mathbf{a}$, then

$$\frac{\left(\bar{z}_1 - \bar{z}_2\right)^2}{s_z^2} = \frac{\left[\mathbf{a}'\left(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2\right)\right]^2}{\mathbf{a}'\mathbf{S}_{\mathrm{pl}}\mathbf{a}} = \frac{\mathbf{a}'\left(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2\right)\left(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2\right)'\mathbf{a}}{\mathbf{a}'\mathbf{S}_{\mathrm{pl}}\mathbf{a}} \tag{3.3}$$

In the case of k-groups, the separation criterion among $z_1, z_2, \ldots, z_k$ can be expressed in terms of matrices, where the **H** matrix denotes $\left(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2\right)\left(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2\right)'$ and the matrix **E** replaces $\mathbf{S}_{\mathrm{pl}}$,

$$\lambda = \frac{\mathbf{a}'\mathbf{H}\mathbf{a}}{\mathbf{a}'\mathbf{E}\mathbf{a}}, \tag{3.4}$$

Matrix **H** has a between sum of squares on the diagonal for each of the *p* variables, and matrix **E** has a within sum of squares for each variable on the diagonal.

Alternatively, the separation criterion can be expressed as

$$\lambda = \frac{\mathrm{SSZ}(z)}{\mathrm{SSE}(z)}, \tag{3.5}$$

where SSH (*z*) and SSE (*z*) are the between and within sums of squares for *z*. The formula (3.4) can be rewritten as:

$$\begin{gathered} \mathbf{a}'\mathbf{H}\mathbf{a} = \lambda\mathbf{a}'\mathbf{E}\mathbf{a}, \\ \mathbf{a}'\left(\mathbf{H}\mathbf{a} - \lambda\mathbf{E}\mathbf{a}\right) = 0. \end{gathered} \tag{3.6}$$

Next, we examine values of λ and **a** that are solutions of (3.6):

$$\begin{gathered} \mathbf{H}\mathbf{a} - \lambda\mathbf{E}\mathbf{a} = 0, \\ (\mathbf{E}^{-1}\mathbf{H} - \lambda\mathbf{I})\mathbf{a} = 0, \end{gathered} \tag{3.7}$$

where **I** refers to the inversion matrix. The solutions are the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_s$ and corresponding eigenvectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_s$ of $\mathbf{E}^{-1}\mathbf{H}$. From the *s* eigenvectors, we obtain *s* discriminant functions $z_1 = \mathbf{a}_1'\mathbf{y}, z_2 = \mathbf{a}_2'\mathbf{y}, \ldots, z_s = \mathbf{a}_s'\mathbf{y}$ which show the dimensions or directions of differences among $\bar{\mathbf{y}}_1, \bar{\mathbf{y}}_2, \ldots, \bar{\mathbf{y}}_k$. The ratio of its eigenvalue can calculate the relative importance of each discriminant function as a proportion to the total:

$$\frac{\lambda_i}{\sum_{j=1}^{s} \lambda_j} \tag{3.8}$$

The coefficients in discriminant functions can be used to assess the contribution of the *y*'s to the separation of groups. As Rencher (2002, p. 283) suggests, this comparison is informative if the *y*'s are measured on the same scale and with comparable differences. For this reason, we use standardized discriminant functions (see, for example, Rencher (2000) for a more detailed description).

3.2.1 Tests of Significance

To test hypotheses, we assume multivariate normality. The discriminant criterion (3.8) is maximized by λ_1, the largest eigenvalue. The remaining eigenvalues $\lambda_2, \ldots, \lambda_s$ are associated with other discriminant dimensions.

The significance test is usually based on the Wilks' lambda (Wilks' Λ), and the eigenvalues are used in the test for significant differences among mean vectors. If the hypothesis H_0 is rejected, we conclude that at least one λ's is significantly different from zero. Therefore, there is at least one dimension of separation of mean vectors. Each λ_i is gradually tested until a test fails to reject H_0. The test statistic at the *m*th, step ($m = 2, 3, \ldots, s$) is:

$$\Lambda_m = \prod_{i=m}^{s} \frac{1}{1+\lambda_i}\,, \tag{3.9}$$

which is distributed as $\Lambda_{p-m+1,k-m,N-k-m+1}$. The statistic

$$V_m = -\left[N-1-\frac{1}{2}(p+k)\right]\ln \Lambda_m = \left[N-1-\frac{1}{2}(p+k)\right]\sum_{i=m}^{s}\ln(1+\lambda_i) \tag{3.10}$$

has an approximate χ^2 - distribution with $(p-m+1)(k-m)$ degrees of freedom. If more λ's are statistically significant, we may not consider the associated discriminant function if $\lambda_i / \sum_j \lambda_j$ is small, even if it is significant (Rencher, 2002).

3.2.2 Interpretation of Discriminant Functions

The main purpose of interpretation is to assess the contribution of each variable. However, the signs of the coefficients are considered. According to Rencher (2002), there are three general approaches to assessing the contribution of each variable to separating the groups:

- Standardized discriminant function coefficients,
- partial F-test for each variable,
- correlation between each variable and the discriminant function.

Standardized coefficients are useful when the variables are measured on differing scales. In that case, coefficients are adjusted so that they apply to standardized variables. For example, for the observations in the first of two groups,

$$z_{1i} = a_1^* \frac{y_{1i1}-\bar{y}_{11}}{s_1} + a_2^* \frac{y_{1i2}-\bar{y}_{12}}{s_2} + \ldots + a_p^* \frac{y_{1ip}-\bar{y}_{1p}}{s_p}\,, \tag{3.11}$$

where $i = 1, 2, \ldots, n_1$. The standardized variables $y_{1ir} - \bar{y}_{1r} / s_r$ are scale-free, and the standardized coefficients $a_r^* = s_r a_r$, where $r = 1, 2, \ldots, p$. This standardization is applied to each of the s discriminant functions. The contribution of the variables to separating the groups is based on the absolute values of the coefficients. The stability of coefficients may vary from sample to sample; for example, if N/p is too small, one sample's important variables may emerge as less important in another sample.

The second approach to assessing the contribution of each variable is based on a partial F-value. We can calculate a partial F-test for any variable y_r and rank the variables. For example, in the case of two groups, the partial F-value is:

$$F = (\nu - p + 1)\frac{T_p^2 - T_{p-1}^2}{\nu + T_{p-1}^2}, \tag{3.12}$$

where T_p^2 is the two-sample Hotelling T^2 with all p variables, T_{p-1}^2 is the T^2-statistic with all variables except y_r, and $\nu = n_1 + n_2 - 2$. The F-statistic is distributed as $F_{1,\nu-p+1}$.

Conversely to standardized coefficients, the partial F-values are not associated with a single dimension of group separation. For example, y_2 will have a different contribution in each of the s discriminant functions; however, the partial F for y_2 creates an overall index of the contribution of y_2 to group separation, considering all dimensions.

Finally, the correlation between variables and discriminant functions can be used to assess each variable's contribution. These correlations are usually referred to as loadings or structure coefficients. Rencher (2002) points out that these correlations show each variable's contribution in a univariate context rather than in a multivariate one.

In addition to LDA, which is one of the well-known methods, we use quadratic and logistic methods of discriminant analysis in the application part.

- Quadratic discriminant analysis (QDA) is a variant of LDA that allows for the non-linear separation of data.
- Logistic discriminant analysis (LogDA) is when the posterior probabilities are estimated by multi-nominal logistic regression (MLR).

3.2.3 Selection of Variables

There are usually a large number of dependent variables available in discriminant analysis applications. For this reason, it is useful to select only some of the variables that will be finally considered for separating groups. For the selection of variables, we can use the following methods of discriminant analysis (Rencher, 2002):

- Forward selection is when we begin with a single variable (the one that maximally separates groups). Then, the variable entered at each step is the one that maximises the partial F-statistic based on Wilks's lambda.
- Backward selection, when we begin with all the variables, and then at each step, the variable that contributes least is deleted according to the partial F-statistic.
- Stepwise selection is when we combine the forward and backward approaches. Variables are added one at a time, and at each step, they are re-examined. The procedure ends when the largest partial F among variables available for entry fails to exceed a pre-set threshold value.

3.2.4 Classification Analysis

The attention in the previous text was paid primarily to a descriptive aspect of discriminant analysis. On the other hand, the discriminant analysis can also solve allocation and group membership prediction. In classification, a sampling unit with unknown group membership is assigned based on the vector of *p* measured values, **y**. As Rencher (2002) suggests, the one approach is to compare **y** with the mean vectors $\overline{\mathbf{y}}_1, \overline{\mathbf{y}}_2, \ldots, \overline{\mathbf{y}}_k$. of the *k* samples. Then, the unit is assigned to the group whose $\overline{\mathbf{y}}_i$ is closest to **y**.

We can use a classification procedure suggested by Fisher (1936, cited in Rencher, 2002) in two populations. Using Fisher's approach, we assume that the two populations have the same covariance matrix, $\Sigma_1 = \Sigma_2$, normality is not required. Supposing that we get two samples from two populations, we can compute $\overline{\mathbf{y}}_1, \overline{\mathbf{y}}_2$ and $\mathbf{S}_{\text{pl}}$. The classification is based on the discriminant function,

$$z = \mathbf{a}'\mathbf{y} = \left(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_1\right)' \mathbf{S}_{\text{pl}}^{-1}\mathbf{y}\ , \tag{3.13}$$

where **y** is the vector of measurements on a new sampling unit that we classify into one of the two groups.

To determine the group membership, we compare *z* with the transformed mean $\overline{z}_1$ or $\overline{z}_2$. For each observation $\mathbf{y}_{1i}$ from the first sample, we evaluate (3.13), obtain $z_{11}, z_{12}, \ldots, z_{1n_1}$ and get $\overline{z}_1 = \sum_{i=1}^{n_1} z_{1i} / n_1 = \mathbf{a}'\overline{\mathbf{y}}_1 = \left(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2\right)' \mathbf{S}_{\text{pl}}^{-1}\overline{\mathbf{y}}_1$, similarly $\overline{z}_2 = \mathbf{a}'\overline{\mathbf{y}}_2$. Assuming **two groups** are referred to as G_1 and G_2, **y** is assigned to G_1 if $\overline{z} = \mathbf{a}'\mathbf{y}$ is closer to $\overline{z}_1$ than to $\overline{z}_2$ according to the Fisher's linear classification procedure, or to G_2 if $\overline{z}$ is closer to $\overline{z}_2$.

In the two-group case, the linear classification function is expressed as the discriminant function for group separating. However, the classification functions are different in the several-group case. Supposing classification of **several groups**, *k*, we find the sample mean vectors $\overline{\mathbf{y}}_1, \overline{\mathbf{y}}_2, \ldots, \overline{\mathbf{y}}_k$. We can use a distance function to assign a group membership for a vector y to find the mean vector that **y** is closest to and set **y** to the corresponding group.

If we assume equal population covariance matrices, $\Sigma_1 = \Sigma_2 = \ldots = \Sigma_k$, then we obtain linear classification functions. A linear function $L_i(\mathbf{y})$ can be expressed as:

$$L_i(\mathbf{y}) = \mathbf{c}_i'\mathbf{y} + c_{i0} = c_{i1}y_1 + c_{i2}y_2 + \ldots + c_{ip}y_p + c_{i0}, \tag{3.14}$$

where $\mathbf{c}_i' = \overline{\mathbf{y}}_i'\mathbf{S}_{\text{pl}}^{-1}$ and $\mathbf{c}_{i0}' = -\frac{1}{2}\overline{\mathbf{y}}_i'\mathbf{S}_{\text{pl}}^{-1}\overline{\mathbf{y}}_i$. Using this procedure, $\mathbf{y}$ is allocated to the group for which $L_i(\mathbf{y})$ is the largest.

If the population covariance matrices are not equal, observations tend to be classified more frequently into groups whose covariance matrices have larger variances on the diagonal. In this case, the rules are based on quadratic classification functions (Rencher, 2002).

The classification procedure for group membership prediction is usually based on the probability of misclassification or the error rate. The complement to the error rate is called the correct classification rate (Rencher, 2002).

The classification procedure can be carried out using the same data used to compute the classification functions. Then the method is called resubstitution. Each observation vector y_{ij} is assigned to a group according to the rules. The correct classifications and misclassifications are counted, and the error rate is calculated as the misclassification proportion. The results are typically shown in a classification table (Table 3-1).

Table 3–1 Classification table for two groups

Actual group	**Number of observations**	**Predicted group**	
		1	**2**
1	n_1	n_{11}	n_{12}
2	n_2	n_{21}	n_{22}

Source: Rencher (2002, p. 307)

We can find the apparent correct classification rate as:

$$\frac{n_{11}+n_{22}}{n_1+n_2}. \tag{3.15}$$

The apparent error rate tends to be biased for small samples, and for this reason, there are techniques for reducing the bias:

- The method of partitioning the sample, when the sample is split into two parts. A training (experimental) sample used to construct the classification rule, and a validation sample used to evaluate it. However, this method is not suitable for small samples. Rencher (2002) recommends using all the data to construct the functions to minimise the variance of the error rate estimate.
- The holdout method, or cross-validation, when all but one observation is used to compute the classification rule, is then used to classify the omitted observations.
- We can also use nonparametric classification procedures and resubstitution and holdout methods (Rencher, 2002).

3.3 Logistic Regression Analysis

This chapter aims to provide the main principles and techniques of logistic regression, especially for our empirical study. Logistic regression is a rather different approach to discriminant analysis. In finance, logistic regression is mostly used in its bivariate context. However, it can be easily modified for the outcome variable with more than two possible values. The common problem where logistic regression can be applied is the prediction of default. Most bankruptcy models are based on scoring methodology, where there are two possible values of the outcome variable, for example, default and non-default. Multinomial logistic regression must be applied when exploring relationships among rating and firms' indicators since there are more than two dependent variable categories. In this case, the number of categories comes from the number of rating groups. The simplest case is when there are just two rating categories: investment and speculative grades. Then, the outcome rating is dichotomous or binary. Finally, univariate or multiple logistic regressions can be used to estimate the prediction model. There is a vast literature on logistic regression methods, for example, Hosmer et al. (2013), Menard (2010), Harrel (2010), or Tabachnik and Fidell (2007).

Generally, if we intend to describe the relationship between an outcome (dependent) variable and a set of independent (predictor or explanatory) variables, logistic regression is a suitable method. Hosmer et al. (2013) distinguish among several types of logistic regression models according to the number of variables used in the model, for example:

- Binary (dichotomous) models:
    - the model with a single variable,
    - the multiple logistic regression model.
- Polychotomous models:
    - the multinomial logistic regression model,
    - the ordinal logistic regression model.

The simplest binary model contains only one independent variable and a dependent variable with two possible outcome values. If we consider more than one independent variable in the model with two possible outcomes, the model is called the multiple or multivariable logistic regression model. The model can be further modified for the outcome variable with more than two levels or responses. Then, it is called a multinomial, polychotomous, or polytomous logistic regression model. Moreover, if the outcome is an ordinal scale, we can use ordinal logistic regression. The definitions and derivations used in the following text are based on Hosmer et al. (2013).

3.3.1 Binary Logistic Model

The binary logistic model's purpose is to predict cases into one of two dependent variable categories by one or more independent variables. In contrast with the discriminant analysis, we do not predict the arbitrary value associated with a category. Still, we try to predict the probability that a case will be classified into one group.

Provided the model of a single independent variable, the relationship between the outcome and the variable can be described using the conditional mean $E(Y|x)$. Y denotes the outcome variable and x represents a specific value of the independent variable. The relationship can be expressed as (3.16) assuming the case of linear regression,

$$E(Y|x) = \beta_0 + \beta_{1x}. \tag{3.16}$$

The value of mean $E(Y|x)$ may take any value, because x ranges between $-\infty$ and $+\infty$. Compared to the linear regression, the conditional mean lies in the interval $\langle 0,1 \rangle$ in the dichotomous outcome variable. We denote the probability of being classified into the first category $\pi(Y=0)$ and the likelihood of being classified into the second category $\pi(Y=1)$. Then it follows that $\pi(Y=0) = 1 - \pi(Y=1)$.

Since the observed values $\pi(Y=1)$ must lie between 0 and 1, the model $\pi(Y=1) = \beta_0 + \beta_1 x$ is not suitable for this problem. For this reason, the model for dichotomous outcome variable is based on logistic distribution, and we use the quantity $\pi(x) = E(Y|x)$ to represent the conditional mean of Y (the outcome variable) given x (a specific value of the independent variable) when the logistic distribution is used,

$$\pi(x) = \frac{e^{\beta_0 + \beta_{1x}}}{1 + e^{\beta_0 + \beta_{1x}}}. \tag{3.17}$$

We can convert the equation (3.17) using the natural logarithm and express *g(x)* called a logit. The logit is linear in its parameters, may be continuous and may range from $-\infty$ and $+\infty$ depending on the variable x:

$$g(x) = \ln\left[\frac{e^{\beta_0 + \beta_{1x}}}{1 + e^{\beta_0 + \beta_{1x}}}\right] = \beta_0 + \beta_{1x}. \tag{3.18}$$

The conditional distribution of the outcome variable follows a binomial distribution. The conditional mean gives the probability $\pi(x)$.

To fit the logistic regression model, we estimate the values of parameters β_0 and β_1 maximize the probability of obtaining the observed data set. Thus, the method is based on the maximum likelihood. We must first construct the function that describes the observed data's probability as a function of the unknown parameters. This function is called the likelihood function, and the values of parameters that maximize this function are called maximum likelihood estimators.

According to Hosmer et al. (2013), the contribution to the likelihood function for the pair (x_i, y_i) can be expressed as:

$$\pi(x_i)^{y_i}\left[1-\pi(x_i)\right]^{1-y_i}. \tag{3.19}$$

Then, the likelihood function is the product of the terms given in (3.19) as:

$$l(\boldsymbol{\beta})=\prod_{i=1}^{n}\pi(x_i)^{y_i}\left[1-\pi(x_i)\right]^{1-y_i}. \tag{3.20}$$

Based on the maximum likelihood principle, the value that maximises the expression (3.20) is used as the estimate of $\boldsymbol{\beta}$. Alternatively, we can use the log of equation (3.20) that is called log-likelihood:

$$L(\boldsymbol{\beta})=\ln\left[l(\boldsymbol{\beta})\right]=\sum_{i=1}^{n}\left\{y_i\ln\left[\pi(x_i)\right]+(1-y_i)\ln\left[1-\pi(x_i)\right]\right\}. \tag{3.21}$$

Once the model is fitted, we assess the significance of the estimated coefficients. The basic principle is to compare observed values of the outcome variable and predicted values obtained from the model, with and without the variable in question. The comparison is based on the likelihood ratio test,

$$D=-2\sum_{i=1}^{n}\left[y_i\ln\left(\frac{\hat{\pi}_i}{y_i}\right)+(1-y_i)\ln\left(\frac{1-\hat{\pi}_i}{1-y_i}\right)\right], \tag{3.22}$$

where $\hat{\pi}_i=\hat{\pi}(x_i)$ and the statistic D is called the deviance. The independent variable significance is assessed by the comparison of the value D with and without the independent variable,

$$G=D(\text{model without the variable})-D(\text{model with the variable}), \tag{3.23}$$

$$G=-2D\left[\frac{\text{likelihood without the variable}}{\text{likelihood with the variable}}\right]. \tag{3.24}$$

This dichotomous model of a single independent variable can be further generalized for more than one independent variable. We assume a collection of p independent variables denoted by the vector $\mathbf{x}'=(x_1, x_2, \ldots, x_p)$, where each of these variables is at least interval scaled. If we denote the conditional probability that the outcome is present by the expression $\Pr\left(Y=1|\mathbf{x}\right)=\pi\left(x\right)$, then the logit of the multiple logistic regression model has the following form,

$$g\left(x\right)=\ln\left[\frac{\pi\left(\mathbf{x}\right)}{1-\pi\left(\mathbf{x}\right)}\right]=\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_p x_p, \qquad (3.25)$$

and the multivariable logistic model is given by:

$$\pi\left(x\right)=\frac{e^{g(\mathbf{x})}}{1+e^{g(\mathbf{x})}}. \qquad (3.26)$$

Similar to the univariable case, the multivariable model is estimated by the maximum likelihood method. The likelihood ratio test is used for the overall significance of the *p* coefficient for the independent variables in the model. Assuming the null hypothesis that the *p* slope coefficients for the covariates in the model are equal to zero, the test is based on the statistic G (3.24) of χ^2 distribution with *p* degrees of freedom.

The possible values of the logistic probabilities from a single dichotomous model are displayed in Table 3-2.

Table 3–2 Probability values of a single dichotomous model

Outcome variable (*y*)	Independent variable (*x*)	
	$x = 1$	$x = 0$
$y = 1$	$\pi(1)=\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$	$\pi(0)=\frac{e^{\beta_0}}{1+e^{\beta_0}}$
$y = 1$	$1-\pi(1)=\frac{1}{1+e^{\beta_0+\beta_1}}$	$1-\pi(0)=\frac{1}{1+e^{\beta_0}}$
Total	1	1

Source: Hosmer, Lemeshow and Strudivant (2013, p. 52)

The logistic model is usually interpreted using the **odds ratio** (OR). The OR is the ratio of the odds for $x=1$ to the odds for $x=0$:

$$OR=\pi\left(x\right)=\frac{\dfrac{\pi\left(1\right)}{1-\pi\left(1\right)}}{\dfrac{\pi\left(0\right)}{1-\pi\left(0\right)}}, \qquad (3.27)$$

where the numerator refers to the odds of the outcome being present among individuals with $x=1$ and the denominator is the odds of the outcome being present among individuals with $x=0$. We can replace the expressions in (3.27) with those in the table (Table 3-2), modify them and finally obtain the following equation,

$$OR = e^{\beta_1}, \tag{3.28}$$

indicating the relationship between the odds ratio and the regression coefficient. The odds ratio approximates how much more likely or unlikely it is for the outcome to be present among those subjects compared to those with $x = 0$.

3.3.2 Multinomial Logistic Regression

The multinomial logistic regression is the modification of the binary alternative. We assume that the outcome variable has more than two levels or categories to extend the previous case. For example, the outcome variable can be the bond rating because there are more than two credit rating categories. As Hosmer et al. (2013) suggest, the multinomial logistic regression model can be sufficiently described for the case of three categories.

We assume the outcome variable, Y, is a nominal scale coded 0, 1 and 2. Thus, we need two logit functions. Usually, the category $Y = 1$ is used as a baseline, and logit functions compare each category to this reference value. Assuming we have p covariates and a constant term, denoted by the vector **x**, of length $p+1$, and $x_0 = 1$, then the two logit functions have the following forms:

$$g_1(x) = \ln\left[\frac{\Pr(Y=1|\mathbf{x})}{\Pr(Y=0|\mathbf{x})}\right] = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \ldots + \beta_{1p}x_p = \mathbf{x}'\boldsymbol{\beta}_1, \tag{3.29}$$

$$g_2(x) = \ln\left[\frac{\Pr(Y=2|\mathbf{x})}{\Pr(Y=0|\mathbf{x})}\right] = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \ldots + \beta_{2p}x_p = \mathbf{x}'\boldsymbol{\beta}_2. \tag{3.30}$$

Then, the conditional probabilities of each outcome variable given the covariate vector **x** are:

$$\Pr(Y=0|\mathbf{x}) = \frac{1}{1+e^{g_1(\mathbf{x})}+e^{g_2(\mathbf{x})}}, \tag{3.31}$$

$$\Pr(Y=1|\mathbf{x}) = \frac{e^{g_1(\mathbf{x})}}{1+e^{g_1(\mathbf{x})}+e^{g_2(\mathbf{x})}}, \tag{3.32}$$

$$\Pr(Y=2|\mathbf{x}) = \frac{e^{g_2(\mathbf{x})}}{1+e^{g_1(\mathbf{x})}+e^{g_2(\mathbf{x})}}. \tag{3.33}$$

Similarly to the binary model, we let $\pi_j(\mathbf{x}) = \Pr(Y = j|\mathbf{x})$ for j = 0, 1, 2. Then each probability is a function of the vector of $2(p+1)$ parameters $\boldsymbol{\beta}' = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')$.

In the case of the three category model, the conditional probability can be expressed as:

$$\pi_j\left(\mathbf{x}\right)=\Pr\left(Y=j\middle|\mathbf{x}\right)=\frac{e^{g_j(\mathbf{x})}}{\sum_{k=0}^{2}e^{g_k(\mathbf{x})}}, \tag{3.34}$$

where the vector $\boldsymbol{\beta}_0=0$ and $g_0\left(\mathbf{x}\right)=0$.

The variables are coded as follows:

- If $Y = 0$, then $Y_0 = 1$, $Y_1 = 0$, $Y_2 = 0$,
- if $Y = 1$, then $Y_0 = 0$, $Y_1 = 1$, $Y_2 = 0$,
- if $Y = 2$, then $Y_0 = 0$, $Y_1 = 0$, $Y_2 = 1$,

and the sum of these variables is $\sum_{j=0}^{2}Y_j=1$. Then the conditional likelihood function for a sample of n independent observations is:

$$l\left(\boldsymbol{\beta}\right)=\prod_{i=1}^{n}\left[\pi_0\left(\mathbf{x}_i\right)^{y_{0i}}\ \pi_1\left(\mathbf{x}_i\right)^{y_{1i}}\ \pi_2\left(\mathbf{x}_i\right)^{y_{2i}}\right], \tag{3.35}$$

and the log-likelihood function can be expressed as:

$$L\left(\boldsymbol{\beta}\right)=\sum_{i=1}^{n}y_{1i}g_1\left(\mathbf{x}_i\right)+y_{2i}g_2\left(\mathbf{x}_i\right)-\ln\left(1+e^{g_1(\mathbf{x}_i)}+e^{g_2(\mathbf{x}_i)}\right). \tag{3.36}$$

The likelihood equations can be found by taking the first partial derivatives of $L\left(\boldsymbol{\beta}\right)$ with respect to each of the $2(p + 1)$ unknown parameters (see for example Hosmer et al. (2013) for more details).

Analogically to the binary model, the multivariable model is interpreted by odds ratios. For example, if we assume that the outcome variable $Y=0$ is the reference outcome, then the odds ratio of the outcome $Y = j$ versus outcome $Y=0$ for covariate values of $x = a$ versus $x = b$ is:

$$OR_j\left(a,b\right)=\frac{\Pr\left(Y=j\middle|x=a\right)/\Pr\left(Y=0\middle|x=a\right)}{\Pr\left(Y=j\middle|x=b\right)/\Pr\left(Y=0\middle|x=b\right)}. \tag{3.37}$$

The importance of the variable in the model is based on the likelihood ratio test. To test the significance of coefficients, we compare the log-likelihood from the fitted model containing the coefficients $\left(L_1\right)$ to the log-likelihood for the model containing only constant terms $\left(L_0\right)$, one for each logit function. The test statistic can be expressed as:

$$G=-2\times\left[L_0-L_1\right]. \tag{3.38}$$

The significance of the coefficients for a variable has degrees of freedom equal to the number of outcome categories minus one times the degrees of freedom for the variable in each logit.

Concerning applying multinomial logistic regression to the bond rating prediction, our credit rating analysis to five categories will require four logit functions and determination of the baseline rating category, which is then compared with other logits. A general expression for the conditional probability in the five-category model is:

$$\pi_j(\mathbf{x}) = \Pr(Y = j|\mathbf{x}) = \frac{e^{g_j(\mathbf{x})}}{\sum_{k=1}^{5} e^{g_k(\mathbf{x})}}, \tag{3.39}$$

and we form four logits comparing $Y = 1$, $Y = 2$, $Y = 3$ and $Y = 4$ to it. The four logit functions are then denoted as:

$$g_1(x) = \ln\left[\frac{\Pr(Y=1|\mathbf{x})}{\Pr(Y=5|\mathbf{x})}\right] = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \ldots + \beta_{1p}x_p = \mathbf{x}'\boldsymbol{\beta}_1, \tag{3.40}$$

$$g_2(x) = \ln\left[\frac{\Pr(Y=2|\mathbf{x})}{\Pr(Y=5|\mathbf{x})}\right] = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \ldots + \beta_{2p}x_p = \mathbf{x}'\boldsymbol{\beta}_2, \tag{3.41}$$

$$g_3(x) = \ln\left[\frac{\Pr(Y=3|\mathbf{x})}{\Pr(Y=5|\mathbf{x})}\right] = \beta_{30} + \beta_{31}x_1 + \beta_{32}x_2 + \ldots + \beta_{3p}x_p = \mathbf{x}'\boldsymbol{\beta}_3, \tag{3.42}$$

$$g_4(x) = \ln\left[\frac{\Pr(Y=4|\mathbf{x})}{\Pr(Y=5|\mathbf{x})}\right] = \beta_{40} + \beta_{41}x_1 + \beta_{42}x_2 + \ldots + \beta_{4p}x_p = \mathbf{x}'\boldsymbol{\beta}_4. \tag{3.43}$$

In addition to multinomial logistic regression analysis, we can also use an ordinal logistic regression approach.

3.3.3 Ordinal Logistic Regression

Generally, multinomial logistic regression can be used if the outcome variables are not nominal but ordinal scale. However, it is suggested to use the ordinal logistic regression, which respects the categorical outcome's natural ranking in some cases. Menard (2010) describes ordinal variables as either crude measurement of a variable that could be measured on an interval or ratio scale or measurement of an abstract characteristic for which there is no natural metric or unit of measurement. Hosmer et al. (2013) argue that multinomial logistic regression could be used even in these cases. Still, it must be realised that not considering the natural ordering of outcome variables may lead to the problem that estimated models might not address the analysis's questions. Thus, the ordinal logistic regression seems to be an appropriate method for assessing bond rating that considers the rank ordering of rating categories. As Menard (2010) says, different models make different assumptions about whether the dependent variable is intrinsically ordinal or reflects an underlying continuous interval or ratio variable. Various models can be proposed to analyse ordinal dependent variables, such as the cumulative logit model, the continuation ratio logit model, the adjacent categories logit model or the stereotype model. The cumulative logit model is the most widely used logistic regression model, and it is described in more detail in this chapter.

According to Hosmer et al. (2013), the ordinal logistic model can be described using the following definitions. We assume that the ordinal outcome variable, Y, can take on $K+1$ values coded $0,1,\ldots,K$. The probability that the outcome is equal to k conditional on a vector $\mathbf{x}$ of p covariates is denoted $\Pr\left(Y=k\middle|\mathbf{x}\right)=\phi_k\left(\mathbf{x}\right)$. If the model is assumed to be multinomial, then $\phi_k\left(\mathbf{x}\right)=\pi_k\left(\mathbf{x}\right)$, where the model is given in equations (3.31 – 3.33) for K = 2.

The multinomial model is called the baseline logit model in the case of ordinal logistic regression. When we assume an ordinal model, we must decide what outcomes to compare and the most reasonable model for the logit. There are three following main alternatives:

- Compare each response to the next larger response (adjacent-category logistic model),
- compare each response to all lower responses (continuation-ratio logistic model),
- compare the probability of an equal or smaller response to the likelihood of a larger response (proportional odds model).

As the proportional odds model will be used to model bond rating in the application part, we describe this approach's principles in the following text. Using the proportional odds model, we compare the probability of an equal or smaller response, $Y \le k$, to the likelihood of a larger response, $Y > k$,

$$c_k\left(\mathbf{x}\right)=\ln\left[\frac{\Pr\left(Y\le k\middle|\mathbf{x}\right)}{\Pr\left(Y>k\middle|\mathbf{x}\right)}\right]=\ln\left[\frac{\phi_0\left(\mathbf{x}\right)+\phi_1\left(\mathbf{x}\right)+\ldots+\phi_k\left(\mathbf{x}\right)}{\phi_{k+1}\left(\mathbf{x}\right)+\phi_{k+2}\left(\mathbf{x}\right)+\ldots+\phi_K\left(\mathbf{x}\right)}\right]=\tau_k-\mathbf{x}'\boldsymbol{\beta}, \quad (3.44)$$

for k = 0, 1, …, K – 1. Supposing $K=1$, the model is simplified to the usual logistic regression model in that it yields odds ratios of $Y=0$ versus $Y=1$.

The method used to fit the ordinal model is based on adapting the multinomial likelihood and log (3.36) for K = 2. The basic procedure involves the following steps (Hosmer et al., 2013, p. 292):

a) The expressions defining the model-specific logits are used to create an equation defining $\phi_k(\mathbf{x})$ as a function of the unknown parameters.

b) The values of a K + 1-dimensional multinomial outcome, $\mathbf{z}' = (z_0, z_1, \ldots, z_k)$ are created from the ordinal outcome as $z_k = 1$ if $y = k$ and $z_k = 0$ otherwise, where only one value of z equals 1.

The general form of the likelihood for a sample of n independent observations $(y_i, \mathbf{x}_i), i = 1, 2, \ldots, n$, is:

$$l(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left[\phi_0(\mathbf{x}_i)^{z_{0i}} \, \phi_1(\mathbf{x}_i)^{z_{1i}} \times \ldots \times \phi_K(\mathbf{x}_i)^{z_{Ki}} \right], \tag{3.45}$$

where $\boldsymbol{\beta}$ denotes both the p slope coefficients and the K model-specific intercept coefficients. Then, the log-likelihood function can be expressed as:

$$L(\boldsymbol{\beta}) = \sum_{i=1}^{n} z_{0i} \ln\left[\phi_o(\mathbf{x}_i)\right] + z_{1i} \ln\left[\phi_1(\mathbf{x}_i)\right] + \ldots + z_{Ki} \ln\left[\phi_K(\mathbf{x}_i)\right]. \tag{3.46}$$

When applying the method, it should be checked whether the data support the assumption of proportional odds. Tests for assessing the goodness of fit are based on the comparison of the model to an augmented model in which the coefficients for the model covariates are allowed to be different:

$$c_k(\mathbf{x}) = \ln\left[\frac{\Pr(Y \le k \mid \mathbf{x})}{\Pr(Y > k \mid \mathbf{x})}\right] = \tau_k - \mathbf{x}'\boldsymbol{\beta}_k, \tag{3.47}$$

where $\tau_k < \tau_{k+1}$ for $k = 1, \ldots, K$.

Norušis (2012, p. 70) explains a minus sign before the coefficients for the predictor variables, which is done so that larger coefficients indicate an association with larger scores. For a continuous variable, positive coefficients suggest that the likelihood of a larger score increases as the variable's values increase. Each logit has its τ_k term but the same coefficient, which means that the independent variable's effect is the same for different logit functions. The τ_k terms are called threshold values, and they are used in the calculations of predicted values.

In the application section, the assigned rating category is considered as the outcome variable in the model. For instance, in our analysis (Chapter 4), the ratings are classified into five ordinal categories, ranging from the lowest (BB) to the highest (A) in terms of bond investment quality. Therefore, we account for five distinct rating grades, with the event of interest being the observation of a specific rating grade or lower. In this context, the ordinal model will evaluate a series of dichotomies between successive rating categories:

- grade (1) versus grades (2, 3, 4 or 5),

- grades (1 or 2) versus grades (3, 4 or 5),
- grades (1 or 2 or 3) versus grades (4 or 5),
- grades (1 or 2 or 3 or 4) versus grade (5).

Thus, in terms of bond rating groups, we will estimate the following odds based on the proportional odds model:

$$\begin{aligned}
\theta_1 &= \frac{\Pr(\text{rating } 1)}{\Pr(\text{rating greater than } 1)}, \\
\theta_2 &= \frac{\Pr(\text{rating } 1 \text{ or } 2)}{\Pr(\text{rating greater than } 2)}, \\
\theta_3 &= \frac{\Pr(\text{rating } 1 \text{ or } 2 \text{ or } 3)}{\Pr(\text{rating greater than } 3)}, \\
\theta_4 &= \frac{\Pr(\text{rating } 1 \text{ or } 2 \text{ or } 3 \text{ or } 4)}{\Pr(\text{rating greater than } 4)}.
\end{aligned} \tag{3.48}$$

The highest category, 5, does not have associated odds because the probability of being in that category or any lower category is 1. Unlike the multinomial approach, ordinal logistic regression requires estimating only one set of regression coefficients.

3.3.4 ROC Analysis

The estimated models represent a classification rule used to assign objects into classes. Thus, we need to know how effectively this classification rule works, preferably using the validation sample. It means that the available data are split into two datasets: an experimental sample (training) used for constructing the rule and the validation sample used for assessing the performance. There are other ways to split the data, for example, the leave-one-out method, when only one data point is put in the validation sample and others in the experimental sample, or the bootstrap methods.

To understand how the performance is measured, we need to specify some terms required for further analysis. According to Krzanowski and Hand (2009), a classification rule yields a score $s(\boldsymbol{X})$ for each object. It will result in distribution $p(s|\mathrm{P})$ for objects in the positive group, P, and distribution $p(s|\mathrm{N})$ for objects in the negative group, N. Then, the classifications are given by comparing the scores with a threshold, *T*. The authors claim that if we can find a threshold $T = t$ such that all members of class P have scores that are all greater than t, and all members of class N have scores all less or equal to t, we attain the perfect classification. However, the two sets of scores typically overlap to some extent, and perfect classification is impossible. In this case, performance is measured by the extent to which scores for objects in class P tend to take large values, and scores for objects in class N tend to take small values. The methods used for these measurements are based on the two-by-two classification table resulting from cross-classifying the true class of each object by its predicted class. Krzanowski and Hand (2009) state that the proportions of the validation set that fall in this table's cells are empirical realisations of joint probabilities $p(s>t,P), p(s>t,N), p(s \le t,P), p(s \le t,N)$. Then, different ways of summarising these four joint probabilities yield various measures of classification performance. Generally, we can use the misclassification or error rate measure, which is the probability of a class N object having a score greater than t or a class P object having a score less than t. The misclassification rate is a widely used criterion; however, it weights two kinds of classification (class N misclassified as P, and vice versa) equally important.

We use the following two conditional probabilities and one marginal probability in the evaluation of classification ability (Krzanowski and Hand, 2009):

- The false positive rate (fp) – the probability that an object from class N yields a score greater than $t: p(s>t|\mathrm{N})$,
- the true positive rate (tp) – the probability that an object from class P yields a score greater than $t: p(s>t|\mathrm{P})$,
- the marginal probability that an object belongs to class $\mathrm{P}: p(\mathrm{P})$.

Next, we use two complementary conditional rates and one complementary marginal probability (Krzanowski and Hand, 2009):

- The true negative rate (tn), $p(s \le t|\mathrm{N})$ - the proportion of class N objects which are correctly classified as class N, equal to $1-fp$,
- the false negative rate (fn), $p(s \le t|\mathrm{P})$ - the proportion of class N objects which are correctly classified as class N, equal to $1-fp$,
- the marginal probability that an object belongs to class $N: p(\mathrm{N}) = 1-p(\mathrm{P})$.

The true positive rate is typically called the Sensitivity (Se), and the true negative rate is the Specificity (Sp). The rates described above are all conditional probabilities of having a particular predicted class given the true class. As Krzanowski and Hand (2009) point out, there are obvious relationships between the various conditional, marginal, and joint probabilities. For example, the misclassification rate e of a classification rule can be expressed as a weighted sum of the true positive and false positive rate:

$$e = (1-tp)\cdot p(\mathrm{P}) + fp \cdot p(\mathrm{N}). \tag{3.49}$$

Since a classification rule's true positive and negative rates are complementary, they are typically used together as joint performance measures. Generally, the true positive rate increases as t decreases, while the true negative rate decreases with a lower t. Thus, we can find the misclassification rate as the value of t which leads to the overall minimum of the weighted sum e in the formula (3.49). Krzanowski and Hand (2009) suggest another way to determine the threshold by choosing the maximum $tp - fp$, or $tp + tn - 1$ (Sensitivity + Specifity – 1). The maximum value is called the Youden index (YI).

Generally, the performance measures are based on comparing the distributions of the scores for the positive and negative populations. A good classification rule tends to produce high scores for the positive population and low scores for the negative population. The larger the extent to which these distributions differ, the better the classifier. The graphical depiction of both two distributions is presented by the ROC (Receiver Operating Characteristic) curve. The method based on the ROC curve is a commonly used way for assessing the performance of classification rules. First, the graph shows the true positive rate (tp) on the vertical axis and the false positive rate (fp) on the horizontal axis, as the classification threshold t varies. Then, the misclassification rate is the minimum distance between the curve and the upper left corner of the square containing the ROC plot. If we develop a classification rule for more than two classes, we face a more complex problem. For example, the rating models assign objects to several rating categories. In this case, we combine multiple ROC curves and use different approaches to assess performance. Krzanowski and Hand (2009) suggest treating the situation using two-class analyses.

There are two main approaches to how the ROC analyses can be achieved:

- Assuming k classes, we produce k different ROC curves by considering each class in turn as population P and the union of all other classes as population N,
- we have all $k(k-1)$ distinct pairwise-class ROC curves.

Both approaches are suitable for summary statistics, such as the AUC (Area Under the Curve). In the case of perfect separation of P and N, AUC is the area under the upper borders of the ROC (the area of a square of side one, so the upper bound is 1). In random allocation, AUC is the area under the chance diagonal (the area of a triangle whose base and height are equal to 1, so the lower band is 0.5). Based on Krzanowski and Hand (2009), the AUC can be generally expressed as

$$AUC = \int_0^1 y(x)dx. \tag{3.50}$$

The AUC can be defined as the average positive rate, taken uniformly over all possible false-positive rates in the range (0,1). A frequently used interpretation of AUC is that it is a probability that the classifier will allocate a higher score to a randomly chosen individual from population P than it will to a randomly and independently chosen individual from population N.

3.4 Survival Analysis

The latter application study focuses on applying the most popular survival analysis techniques. It is suggested to use the regression models appropriate for survivor data to analyse time to event. As Hosmer et al. (2008, p. 3) claim, the most important differences between the outcome variables modelled via linear and logistic regression analyses and the time variable are that we may only partially observe the survival time. If the event's occurrence is unimportant, the event can be analysed as a binary outcome using the logistic regression model. As Harrell (2010) points out, survival analysis is used to analyse the data in which the time until the event is of interest. The input variable is the time until the event or duration time. The survival analysis allows the response to be incompletely determined for some subjects; perhaps we cannot follow all observations in the dataset. For example, some companies are still alive after the observation time or lost to follow-up. As we face incomplete information, we need to analyse the data using specialised survival techniques. The analysis involves a censoring mechanism when we define the censored and uncensored observations. For example, Hosmer et al. (2008, p. 18) explain a censored observation as one whose value is incomplete due to random factors for each subject. If we analyse data using the survival procedure, the dates of start and finish are not dealt with because they are different. Rather, we consider the length of time before the initial event, $t = 0$, and the terminal event or date of the last information about the object, $t = 1$. Usually, when an observation begins at the defined time $t = 0$ and terminates before the outcome of interest, it is assumed to be a censored observation. If no responses are censored, standard regression models for continuous responses could analyse the failure times (Harrell, 2010).

Based on the distribution of failure times, we use parametric, semiparametric, and nonparametric methods. Survival analysis is the approach that allows working with incomplete data and modelling the time to an event, such as a corporate failure or default. Two time points must be clearly defined for time to event modelling: the beginning point and an endpoint when the event of interest occurs. In this context, survival time refers to the distance on the time scale between these two points (Hosmer et al., 2008).

3.4.1 Censoring

When applying the survival analysis, we deal with censoring the data that comes from the fact that we can face the problem of incomplete observation of time. Two mechanisms can lead to incomplete observation in time: censoring and truncation. These two terms can be defined as follows:

- A censored observation: the value of an observation is incomplete due to random factors for each subject.
- A truncated observation: the value of an observation is incomplete due to a selection process inherent in the study design.

According to Hosmer et al. (2008), there are several types of incomplete observations:

- Right censoring,
- left censoring,
- interval censoring,
- left truncation,
- right truncation.

For survival analysis, we have to specify a point when observation ends on all subjects. Thus, subjects may enter the study at different times; however, they will have variable lengths of maximum follow-up time.

For example, in Figure 3-1, we can see a hypothetical study of four subjects, the end of the study is September 2016. The bond issuer 1 entered the study in January 2015 and defaulted in February 2016. The bond issuer 2 joined the study in February 2015 and was lost to follow up in December 2015. The bond issuer 3 entered the study in May 2015, and there was no default until September 2016, the end of the study. Finally, bond issuer 4 joined the study in August 2015 and defaulted in April 2016.

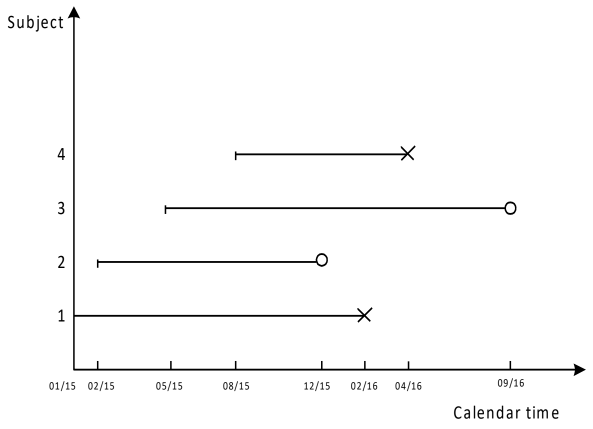


Figure 3–1 Line plot in calendar time (follow-up study)

Source: Hosmer et al. (2008, p. 7), author

For the practical reasons of survival analysis, we can assume that all subjects entered the study at the same calendar time and were followed until their respective endpoint. Thus, we must convert the collecting data from calendar time to analysis time (Figure 3-2).

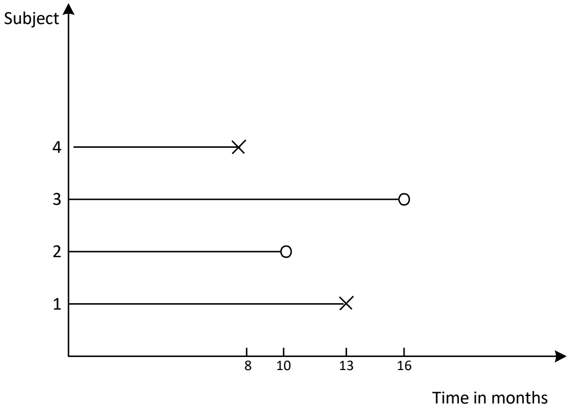


Figure 3–2 Line plot in the time scale (follow-up study)

Source: Hosmer et al. (2008, p. 7), author

The previous example is the case of the most common type of censoring, right censoring. The incomplete observations occur in the right tail of the time axis, usually when the observation begins at the defined time and terminates before the outcome of interest is observed. In some cases, left censoring can be used if the event of interest has already occurred when observation begins. If the time is not observable continuously, we can use interval censoring. It is a special type of failure data that include the right-censored failure time data but have a much more complex structure; for more details, you can see, for example, Sun and Li (2014). Since the left and right truncation are less common forms of incomplete data, they will not be considered in this text, and for practical reasons, the focus will be especially on right censoring.

3.4.2 Survival and Hazard Functions

In the case of right censoring, two random variables need to be defined (Houwelingen and Stijnen, 2014):

- The survival time (T_{surv}),
- the censoring time (T_{cens}),

where the termination of the study usually determines the censoring time. The necessary condition for statistical analysis is that survival time and censoring time are independent. This condition can be weakened to the independence of survival time and censoring time conditional on the explanatory variables in the presence of explanatory variables.

Then, we can define cumulative distribution functions for both random variables:

$$F_{surv}(t) = \Pr(T_{surv} \le t), \tag{3.51}$$

$$F_{cens}(t) = \Pr(T_{cens} \le t). \tag{3.52}$$

According to Houwelingen and Stijnen (2014), the distribution function of the survival time is commonly called the failure function. However, for practical reasons, it is often more suitable to use complimentary functions in survival analysis, the survival (or survivor) function *S(t)* and the censoring function *G(t)*:

$$S(t) = 1 - F_{surv}(t) = \Pr(T_{surv} > t), \tag{3.53}$$

$$G(t) = 1 - F_{cens}(t) = \Pr(T_{cens} > t). \tag{3.54}$$

We assume that T_{surv} has a continuous distribution, and thus the survival function *S(t)* is continuous and differentiable. Although the survival function gives the most important information of survival dataset, the censoring function can also be relevant to describe the distribution of the follow-up times if no subject would have died, for example, if no issuer would have defaulted. In practice, it is mostly impossible to observe both, T_{surv} and T_{cens} , thus the observed survival time *T* is the smallest of the two,

$$T = \min(T_{surv}, T_{cens}). \tag{3.55}$$

In our survival dataset, we can identify the time that is observed according to the event indicator *D*. The usual definition is:

$$D = \begin{cases} 0, & \text{if } T = T_{cens} \\ 1, & \text{if } T = T_{surv} \end{cases}, \tag{3.56}$$

and the information on the survival status is given by the pair (*T*, *D*).

Following the previous description of the observed time, the survival time can be referred to as *T*. Then, the survivor function and cumulative distribution can be expressed as:

$$S(t) = 1 - F(t) = \Pr(T > t), \tag{3.57}$$

$$F(t) = \Pr(T \le t), \tag{3.58}$$

where *T* is a nonnegative random variable denoting the time to a failure event.

The survival function evaluated at time *t* can be considered the probability that a subject will live for at least time *t*, taking values between 0 and 1. The survival function is equal to one at *t* = 0 and decreases toward zero as *t* goes to infinity (Gourieroux and Jasiak, 2007), (Cleves et al., 2010).

Using the survival function, we can estimate the probability of surviving beyond time *t*. In other words, we can estimate the likelihood that there is no failure event before *t*.

The density function *f(t)* can be obtained both from *S(t)* or F*(t)*:

$$f(t) = \frac{dF(t)}{dt} = \frac{d}{dt}\{1 - S(t)\} = -S'(t). \tag{3.59}$$

The hazard function or rate *h(t)* at time *t* can be explained as the probability that the subject will die. More specifically, the company will bankrupt or default very shortly after reaching time *t*, provided it gets time *t* (Gourieroux and Jasiak, 2007). Cleves et al. (2010) explain the hazard rate as the conditional failure rate or the intensity function. As they emphasize, the hazard rate represents the instantaneous rate of failure with 1/*t* units:

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t + \Delta t > T > t \mid T > t)}{\Delta t} = \frac{f(t)}{S(t)}. \tag{3.60}$$

The hazard function can range from zero (no risk) to infinity (the certainty of failure at that instant) and can decrease, increase, or constant, or even take on other shapes. The relationship between the hazard and the survival function can be described as

$$h(t) = \frac{f(t)}{S(t)}. \tag{3.61}$$

Gourieroux and Jasiak (2007) use duration dependence to describe the relationship between the exit rate and the time spent in a given state by a subject. The form of the hazard function determines it. For example, the positive duration dependence in a sequence of failure events occurring randomly in time means that the more time elapsed since the last failure event, the greater the probability of an instantaneous occurrence of another failure. There are three types of duration dependence:

- Negative, associated with decreasing hazard functions,
- positive, associated with increasing hazard functions,
- the absence of duration dependence, and no relationship between the exit rate and the duration.

Analysis of survival data can be based on parametric, semiparametric and nonparametric modelling. While parametric models require assumptions about the distribution of failure times, semiparametric models are parametric in the sense that the effect of the covariates is assumed to take a certain form. In this case, no parametric form of the survival function is specified, yet the effects of covariates are parametrized to modify the baseline survivor function. Compared to the previous approaches, nonparametric models do not require any assumptions about the distribution of failure times. Survival models are estimated based on different methods. Some of the widely used approaches are summarized in Table 3-3.

Table 3–3 Survival analysis models

Nonparametric	Kaplan-Meier Nelson-Aalen
Semiparametric	Cox-proportional hazard
Parametric	Exponential Weibull Lognormal Loglogistic Gamma Gompertz

Source: Cleves et al. (2010)

3.4.3 Nonparametric Models

According to Cleves et al. (2010), when no covariates exist or are qualitative in nature, we can use nonparametric methods such as Kaplan-Meier and Nelson-Aalen to estimate survival probability past a certain time or to compare survival experiences for different groups. The common characteristic of nonparametric models is that they do not make any assumptions about the distribution of failure times or how covariates change the survival experience.

The **Kaplan-Meier** method will be described based on Hosmer et al. (2008) and Houwelingen and Stijnen (2014) as follows. We assume a sample of independent observations denoted $(t_i, c_i), i = 1, 2, \ldots, n$ of the underlying survival time variable *T* and the censoring indicator *C* (we take right-censoring). If $c_i = 1$, the observed times are considered survival times, if $c_i = 0$, they are called the censoring times.

Assuming that there are $m \leq n$ recorded times of failure and $n - m$ censored values among *n* observations, we denote the rank-ordered survival times as $t_{(1)} < t_{(2)} < \ldots < t_{(m)}$. The Kaplan-Meier estimator of the survivorship function at time *t* can be obtained from the equation

$$\hat{S}(t) = \prod_{t_{(i)} \leq t} \frac{n_i - d_i}{n_i}, \tag{3.62}$$

where n_i is the number at risk of dying (company failure) at $t_{(i)}$, d_i refers to the observed number of deaths (company failures), and $\hat{S}(t) = 1 \; if \; t < t_{(i)}$. The calculated estimators of the survival function are used for interpretation and used to derive point estimates of quantiles of the distribution. As Hosmer et al. (2008) state, there is no typical shape of the survivorship function because many factors can influence it. Various approaches can be used to derive the variance of the Kaplan-Meier estimator (Hosmer et al., 2008).

The Kaplan-Meier estimator can be considered the most frequently used estimator. On the other side, if we assume that the time variable is continuous, then the survival function may be expressed as:

$$S(t) = e^{-H(t)}, \tag{3.63}$$

where *H(t),* the cumulative hazard function, can be written as

$$H(t) = -\ln(S(t)). \tag{3.64}$$

The expression (3.64) shows that the survival function can be estimated using an estimator of *H(t)* instead of *S(t)* (e.g. Kaplan-Meier estimator). For example, Aalen, Nelson and Altshuler proposed the indicator *H(t)* that is referred to as **Nelson-Aalen** estimator (Hosmer et al., 2008). The Nelson-Aalen estimator of *H(t)* can be expressed as:

$$\tilde{H}(t) = \sum_{t_{(i)} \le t} \frac{d_i}{n_i}. \tag{3.65}$$

Then, the Nelson-Aalen estimator of the survival function is

$$\tilde{S}(t) = e^{-\tilde{H}(t)}. \tag{3.66}$$

As Hosmer et al. (2008) emphasize, the expression (3.633.63) is valid for continuous-time, but the estimator in (3.66) is discrete. As the authors show, using the Taylor series expansion, $d_i/n_i \le -\ln(1-d_i/n_i)$ for each survival time. For that reason, the Nelson-Aalen estimator will always be greater than or equal to the Kaplan-Meier estimator. If the size of the risk sets relative to the number of events is large, then $d_i/n_i \cong -\ln(1-d_i/n_i)$ and there will be little difference between Nelson-Aalen and the Kaplan-Meier estimators of the survival function.

The function *H(t)* is also called the cumulative or integrated intensity process. There is a reverse direction between *H(t)* and *S(t),* and an increase in the cumulative hazard is associated with a decrease in *S(t).* The term hazard can be used to describe the concept of the risk of failure in an interval after time *t*, conditional on the subject having survived to time *t*. The term cumulative refers to the fact that its value is the total hazard up to time *t*. Therefore, the cumulative hazard function is defined as integral of the hazard from time 0 to time *t*,

$$H(t) = \int_0^t h(u)du, \tag{3.67}$$

where $h()$ is the hazard function (Cleves et al., 2010; Hosmer et al., 2008). For example, assuming a company that survived ten years had to be alive during the ninth year of follow-up. Thus, we can say that the hazard at ten years is the failure rate per year, conditional on the fact that the subject has lived nine years. The conditional failure rate applies only to that subset of the sample that has survived to a particular time, and the hazard function can be defined as:

$$h(t) = \frac{f(t)}{S(t)}, \tag{3.68}$$

where *S(t)* is the survival function and *f(t)* is the probability density function of the time variable. In more practical terms, the numerator in the hazard function (3.61)) is the unconditional probability of experiencing the event at time *t*, which is then scaled by the fraction of alive at time *t*.

The Nelson-Aalen estimator of the hazard function at observed time *t* can be written as:

$$\tilde{h}(t_i) = \tilde{H}(t_{i+1}) - \tilde{H}(t_i) = \frac{d_i}{n_i}. \tag{3.69}$$

As Hosmer et al. (2008) argue, the estimator in (3.69) requires many failures at each point to be useful. Otherwise, there is so much variability in the values that we cannot draw any conclusions about its base shape. However, the variability in the values can be smoothed out by averaging, which is called kernel smoothing. As the authors emphasize, the kernel smoothed estimator estimates a smoothed hazard function, not the hazard function itself. Nevertheless, using the smoothed hazard function, we can obtain a visual impression instead of providing precise point-wise estimates.

3.4.4 Cox-Proportional Hazard Model

The Cox-proportional hazard model is a semiparametric model of survival analysis. The effect of the covariates is assumed to take a certain form compared to the nonparametric approach. In this case, no parametric form of the survival function is specified, yet the effects of covariates are parametrized to modify the baseline survivor function. In general, the baseline survival function is the function for which all covariates are equal to zero in a certain way. According to Hosmer et al. (2008), one form of a regression model for the hazard function can be expressed as:

$$h(t,x,\beta) = h_0(t) r(x,\beta). \tag{3.70}$$

As the authors emphasize, the hazard function in (3.70) is the product of two parts, which characterizes how the hazard function changes as a function of survival time and the function $r = (x,\beta)$ that describes how the hazard function changes as a function of subject covariates.

It follows from the model that:

- The functions must be chosen such that $h(t,x,\beta) > 0$,
- $h_0(t)$ is the hazard function when $r(x,\beta) = 1$,

- $h_0(t)$ is referred to as the baseline hazard function when the function $r(x,\beta)$ is parametrized such that $r(x=0,\beta)=1$.

Thus, the baseline hazard function can be seen as a generalization of the intercept or constant term found in parametric regression models. We do not make any assumptions about $h_0(t)$, however, at the cost of a loss in efficiency. Although the model makes no assumptions about the shape of the hazard over time, the general shape is assumed to be the same for everyone.

The ratio of the hazard functions for two subjects with covariate values denoted x_0 and x_1 is:

$$\begin{aligned} HR(t,x_1,x_0) &= \frac{h(t,x_1,\beta)}{h(t,x_0,\beta)}, \text{or} \\ HR(t,x_1,x_0) &= \frac{h_0(t)r(x_1,\beta)}{h_0(t)r(x_0,\beta)} = \frac{r(x_1,\beta)}{r(x_0,\beta)}. \end{aligned} \tag{3.71}$$

As shown in (3.71), the hazard ratio (HR) depends only on the function $r(x,\beta)$. This model was originally proposed by Cox in 1972, who suggested using $r(x,\beta)=\exp(x\beta)$ for practical reasons. Then, the hazard function can be expressed as:

$$h(t,x,\beta)=h_0(t)e^{x\beta}, \tag{3.72}$$

and the hazard ratio is

$$HR(t,x_1,x_0)=e^{x\beta(x_1-x_0)}. \tag{3.73}$$

The Cox model is the most used semiparametric model called Cox model, Cox proportional hazards model, or the proportional hazards model. The term proportional hazards (PH) refers to the fact that the hazard functions are multiplicatively related; thus, their HR is constant over time (Hosmer et al., 2008, p. 70). In other words, we assume that the covariates multiplicatively shift the baseline hazard function. Then, one subject's hazard is a multiplicative replica of another's (Cleves et al., 2010). For example, when a covariate is dichotomous, with a value $x_1=1$ for small companies and $x_0=0$ large companies, the hazard ratio can be written $HR(t,x_1,x_0)=e^{\beta}$. For instance, if the value of the coefficient is $\beta=\ln(2)$, then small companies are failing at twice $(e^{\beta}=2)$ the rate of large companies.

Besides the assumption of proportional hazards, other parametrizations can be used, for example, additive models. These parametrisation approaches are described in the relevant literature (Hosmer et al., 2008; Klein et al., 2014).

The regression coefficients can be estimated by the maximum likelihood method (see Gourieroux and Jasiak, 2007, p. 99; Hosmer et al., 2008). After fitting the model, the significance of the model and the formation of a confidence interval for key estimated parameters should follow. The latter mentioned authors suggest three related tests to assess the significance of the coefficient:

- The partial likelihood ratio test,
- the Wald test,
- the score test.

The **partial likelihood test** is based on the following statistic:

$$G = 2\left\{L_p(\hat{\beta}) - L_p(0)\right\}, \tag{3.74}$$

where L_p refers to the log partial likelihood of the model containing the covariate and L_0 is the log partial likelihood for the model not containing the covariate. The log partial likelihood evaluated at $\beta = 0$ is:

$$L_p(0) = -\sum_{i=1}^{m} \ln(n_i), \tag{3.75}$$

where n_i denotes the number of subjects in the risk set at observed survival time t_i. Under the null hypothesis that the coefficient is equal to zero, this statistic will follow a χ^2-distribution with 1 degree of freedom and thus can be used to obtain *p*-values to test the significance of a coefficient (Hosmer et al., 2008).

The **Wald statistic** test is based on the ratio of the estimated coefficient to its estimated standard error, assuming that the statistic follows a standard normal distribution. Unlike the linear regression, the Wald and log partial likelihood ratio test are not numerically related. The Wald statistic is given by

$$z = \frac{\hat{\beta}}{SE(\hat{\beta})}. \tag{3.76}$$

The third approach, the **score test**, is based on the ratio of the derivative of the log partial likelihood to the square root of the observed information all evaluated at $\beta = 0$ (see Hosmer, et al., 2008).

Cleves et al. (2010) use the term relative hazard for $e^{x\beta}$, and the log relative hazard, or risk score, for $x\beta$. To verify the model's specification $x\beta$ and adequate parameterisation, we can use tests called tests of the proportional-hazard assumptions (P-H assumptions). In the application, the tests will be based on the analysis of residuals. As to the fact that the proportional hazards model to censored survival data is fit using the partial likelihood, the calculation of residuals differs from the usual regression models. For this reason, various approaches have been developed for Cox proportional model. The residuals used in the empirical study will be based on Schoenfeld residuals. For more details, see, for example, Hosmer et al. (2008), Cleves et al. (2010), Harrel (2010). Since the survival models estimate the time to event, the explained variation should also be assessed after fitting the model. The measures of explained variation for use with censored survival data differ from the traditional concept of variation using the index of determination. In our case, we apply the measure proposed by Royston (2006) with the character of explained variation in proportional hazards models, which can be used as an adjusted index of determination in PH models.

3.4.5 Parametric Models

While nonparametric analysis is a useful tool for describing our survivor data, semiparametric models are used especially for the estimation of hazard ratios and their further explanations. In many cases, semiparametric analysis based on the Cox model can sufficiently analyse our data. However, if we aim to predict the time to failure, some parametric assumption is necessary. Parametric models generally provide smooth estimates of the hazard and survival functions and enable us to model also a nonproportional effect on the hazard scale (Royston and Lambert, 2011). Compared to semiparametric models, parametric models are used when the distribution of survival time has a known parametric form. In this case, the fully parametric model enables us to better analyze survival data. Cleves et al. (2010) describe six standard parametric survival models: exponential, Weibull, Gompertz, lognormal, and generalized gamma.

According to Hosmer et al. (2008), using these models may have the following advantages:

- Full maximum likelihood may be used to estimate the parameters,
- the estimated coefficients or their transformations can provide clinically meaningful estimates of effect,
- fitted values from the model can provide estimates of survival time,
- residuals can be computed as differences between observed and predicted values of the time.

In parametric models, we assume that the distribution of time to event (T) can be described as a function of a single covariate:

$$T = e^{\beta_0 + \beta_1 x} \times \varepsilon. \tag{3.77}$$

Since the time must always be positive, the equation (3.77) can be expressed as the product of a positive systematic component, $\exp(\beta_0 + \beta_1 x)$, and an error component, ε, that also takes only positive values.

a) Exponential Regression Model

The exponential survival model is the simplest parametric model with a constant hazard function. We denote the exponential distribution with survival function

$S(t) = \exp(-t)$ as $E(1)$. The model can be expressed by taking the natural log of each side of the equation (3.77):

$$\ln(T) = \beta_0 + \beta_1 x + \varepsilon^*, \tag{3.78}$$

where $\varepsilon^* = \ln(\varepsilon)$. If the error component ε follows the exponential distribution, then the error component ε^* follows the extreme minimum value distribution denoted as $G(0,1)$. The model in (3.76) is called the exponential regression model. If this model is generalized by allowing the shape parameter to be different from 1 by using $G(0,\sigma)$ distribution:

$$\ln(T) = \beta_0 + \beta_1 x + \sigma \times \varepsilon^*, \tag{3.79}$$

we get the Weibull regression model (3.79).

Survival time models that are linearized by taking logs are called accelerated failure time models. The covariate effect in these models is multiplicative on the time scale, as we can see in (3.77). In other words, the impact of the covariate is said to accelerate survival time.

To describe the method of the exponential regression model, firstly, we assume the single covariate model (3.77), where the error distribution is log-exponential. Then, the survival function for the model can be expressed as:

$$S(t, x, \boldsymbol{\beta}) = \exp(-1/e^{\beta_0 + \beta_1 x}). \tag{3.80}$$

If we set the right-hand side of this equation equal to 0.5 and solve the resulting equation, we get an equation for the covariate specific median survival time of

$$t_{50}(x, \boldsymbol{\beta}) = -e^{\beta_0 + \beta_1 x} \times \ln(0.5). \tag{3.81}$$

Assuming the dichotomous covariate in (3.79) coded 0 or 1, then the ratio of the median survival time for the group with $x=1$ to the group with $x=0$ is:

$$\mathrm{TR}(x=1, x=0) = \frac{t_{50}(x=1, \boldsymbol{\beta})}{t_{50}(x=0, \boldsymbol{\beta})} = \frac{-e^{\beta_0 + \beta_1} \times \ln(0.5)}{-e^{\beta_0} \times \ln(0.5)} = e^{\beta_1}, \tag{3.82}$$

where TR denotes time ratio. The relationship between the two median times can be written as:

$$t_{50}(x=1, \boldsymbol{\beta}) = e^{\beta_1} t_{50}(x=0, \boldsymbol{\beta}). \tag{3.83}$$

For example, if $\exp(\beta_1) = 2$, then the median survival time in the group with $x=1$ is twice the median survival time in the group with $x=0$. The quantity $\exp(\beta_1)$ is usually called the acceleration factor, although it can accelerate or decelerate survival time.

The multiplicative covariate effect can be presented using the following form of survival function,

$$S(t, x=1, \boldsymbol{\beta}) = S(te^{-\beta_1}, x=0, \boldsymbol{\beta}). \tag{3.84}$$

The equation shows that the value of the survival function at time t for the group with $x=1$ can be obtained by evaluating the survival function for the group with $x=1$ at time $t\exp(-\beta_1)$.

In addition to the survival function, the model in (3.80) can be expressed in terms of the hazard function as follows,

$$h(t, x, \boldsymbol{\beta}) = e^{\beta_0 + \beta_1 x}, \tag{3.85}$$

that is constant over time because it depends only on model coefficients and covariate values. Thus, on the one hand, the hazard function is relatively simple. However, it may be too simple to provide a realistic description of the survival data. The hazard ratio for a dichotomous covariate is

$$\mathrm{HR}(x=1, x=0) = e^{-\beta_1}. \tag{3.86}$$

The model can be estimated using the maximum likelihood method. The fitted values are predictions of values from a censored exponential distribution. The estimator of variances and covariances of the estimator of the coefficients are obtained using the second partial derivative of the log-likelihood function. The influence of individual subjects on the values of the estimated parameters is based on the score residuals (see Hosmer et al., 2008).

In parametric models, the assumption of proportional hazards is replaced by the procedure that determines whether the data support the particular parametric form of the hazard function. Hosmer et al. (2008) suggest using the model-based estimate of the cumulative hazard function to form the Cox-Snell residuals. The estimated cumulative hazard function values can be considered observations from a censored sample from an exponential distribution with a parameter equal to one. Then, the model diagnosis is based on the plot that compares the model-based cumulative hazard to the hazard obtained from a nonparametric estimator (Kaplan-Meier, Nelson-Aalen). As the authors say, the nonparametric estimator uses the model-based estimates of the cumulative hazard at each observed time as the time variable and the censoring indicator from the survival time variable as the censoring variable. Thus, this plot should follow a line through the origin with a slope equal to one if the parametric model is correct. The estimator of Cox-Snell residuals can be obtained by exponentiating the additive residuals on the log time scale (see Hosmer et al., 2008, p. 257 for more details). Similar to Cox proportional hazards model, the significance of variables can be assessed using the score test, likelihood ratio or Wald test.

b) Weibull Regression Model

We consider the Weibull distribution a natural generalization of the exponential (Roysten and Lambert, 2011). The Weibull regression model can be expressed using the natural log as in equation (3.80). Compared to the exponential model, the Weibull model allows the shape parameter σ to be different from 1 by using $G(0,\sigma)$ distribution.

The hazard function for the single covariate model is

$$h(t,x,\boldsymbol{\beta},\lambda)=\frac{\lambda t^{\lambda-1}}{e^{(\beta_0+\beta_1 x)^{\lambda}}}\lambda, \tag{3.87}$$

and we assume that $\lambda = 1/\sigma$.

The **proportional hazards form** of the function can be expressed as

$$h(t,x,\boldsymbol{\beta},\lambda)=\lambda t^{\lambda-1}e^{-\lambda(\beta_0+\beta_1 x)}=\lambda t^{\lambda-1}e^{-\lambda\beta_0}e^{-\lambda\beta_1 x}\text{, or} \tag{3.88}$$

$$h(t,x,\boldsymbol{\beta},\lambda)=\lambda\gamma t^{\lambda-1}e^{-\lambda\beta_0}e^{-\lambda\beta_1 x}=h_0(t)e^{\theta_1 x}, \tag{3.89}$$

where $\gamma=\exp(-\beta_0/\sigma)=\exp(\theta_0),\theta_1=-\beta_1/\sigma$ and the baseline hazard function is

$$h_0(t)=\lambda\gamma t^{\lambda-1}, \tag{3.90}$$

and $\lambda = 1/\sigma$ is usually called the shape parameter. Although the parameter σ is a variance-like parameter on the log-scale, we can refer to σ as the shape parameter in this text, as suggested by Hosmer et al. (2008, p. 261). The parameter γ is called the scale parameter. For example, the expression in (3.80) leads to a hazard ratio interpretation of the parameter θ_1. The Weibull distribution can provide variety of shapes of the hazard function determined by the estimated parameter λ. When $\lambda = 1$, the hazard is constant, and the Weibull model reduces to the exponential model. When $\lambda < 1$, the hazard in monotone decreasing, and when $\lambda > 1$, it is monotone increasing. Thus, the Weibull model is suitable for modelling data that exhibit monotone hazard rates (Cleves et al., 2010).

The **accelerated failure-time form** of the hazard function can be expressed as

$$h(t,x,\boldsymbol{\beta},\lambda)=\lambda t^{\lambda-1}e^{-\lambda(\beta_0+\beta_1 x)}=\lambda\gamma(te^{-\beta_1 x})^{\lambda-1}e^{-\beta_1 x}. \tag{3.91}$$

The relationship between the two sets of estimated coefficients using the proportional hazards form and the accelerated failure-time form of the hazard function is $\boldsymbol{\theta} = \textbf{-}\boldsymbol{\beta}/\boldsymbol{\sigma}$.

The survival function that corresponds to the accelerated failure-time form of the hazard function in (3.92) is

$$S(t,x,\boldsymbol{\beta},\sigma)=\exp\left\{-t^{\lambda}\exp\left[(-1/\sigma)(\beta_0+\beta_{1x})\right]\right\}. \tag{3.92}$$

Then, the median survival time can be obtained by setting the survival function equal to 0.5 and solving for time,

$$S_{50}(x,\boldsymbol{\beta},\sigma)=\left[-\ln(0.5)\right]^{\sigma}e^{\beta_0+\beta_{1x}}. \tag{3.93}$$

If the covariate is dichotomous and coded 0/1, then the time ratio at the median survival time is (Hosmer et al., 2008, p. 262):

$$\mathrm{TR}(x=1,x=0)=\frac{t_{50}(x=1,\boldsymbol{\beta},\sigma)}{t_{50}(x=0,\boldsymbol{\beta},\sigma)}=\frac{\left[-\ln(0.5)\right]^{\sigma}e^{\beta_0+\beta_1}}{\left[-\ln(0.5)\right]^{\sigma}e^{\beta_0}}=e^{\beta_1}. \tag{3.94}$$

The interpretation of the $\boldsymbol{\beta}$ form of the coefficients is the same as in the exponential regression model. The assessment of model fit is based on the score residuals, similarly to the exponential regression model. See, for example, Hosmer et al. (2008) for a detailed description.

c) Flexible Parametric Models

Royston and Lambert (2011) suggest that simpler parametric models may not be flexible enough to represent the hazard function adequately and thus fit our data well. For example, the hazard function of a Weibull model always goes in the same direction with time. Therefore, the authors propose new parametric models that include flexible PH models, flexible proportional odds (PO), and probit-scale models. Thus, we get alternative models which extend the range of survival distributions that can be estimated. Furthermore, these models allow nonlinearity functions and thus increase their practical use and applications.

Royston and Lambert (2011) propose Royston-Parmar (RP) models, which have considerably greater flexibility concerning the shapes of the survival distributions they can model. In this case, the baseline distribution function is a restricted cubic spline function of log time instead of simply as a linear function of log time. The complexity of models with spline functions is determined by the number and the positions of the connection points in log time (knots) of the spline's cubic polynomial segments. The parameters of models are estimated based on maximum likelihood; for more details and description of models, see, for example, Royston and Lambert (2011). The authors consider RP models as an extension of the Weibull, loglogistic, and lognormal models. While we assume that the effect of covariates is proportional on the appropriate scale (hazard, odds of failure, or probit of failure probability) in these models, the assumption of linearity is relaxed in RP models. The generalisation of the Weibull model using spline functions is described by the authors as follows. Firstly, we express the Weibull cumulative hazard function in logarithmic form:

$$\ln H(t)=\ln\lambda+\gamma_1\ln t=\gamma_0+\gamma_1\ln t\ , \tag{3.95}$$

where $\ln H(t)$ is a sum of two components: a constant (γ_0) and a linear function of log time ($\gamma_1 \ln t$). In the case when the latter component does not correctly capture the shape of the (log) cumulative hazard function, we might need a more flexible model:

$$\ln H(t) = f(t;\gamma)\,, \tag{3.96}$$

where $f(t;\gamma)$ represents some general family of a nonlinear function of time t, having some parameter vector γ. Royston and Lambert (2011) suggest fractional polynomials and splines as suitable functions. For example, they describe a restricted cubic spline function as $s(\ln t;\gamma)$ with *s* standing for splines and ln*t* to show that we use the scale of log time:

$$\ln H(t) = s(\ln t;\gamma) = \gamma_0 + \gamma_1 \ln t + \gamma_2 z_1(\ln t) + \gamma_3 z_2(\ln t) + ..., \tag{3.97}$$

where $\ln t$, $z_1(\ln t)$, $z_2(\ln t)$ and so on are the basis functions of the restricted cubic spline. Thus, if there are no knots, then $s(\ln t;\gamma_0,\gamma_1) = \gamma_0 + \gamma_1 \ln t$, which is the Weibull model. The parameters γ are estimated by maximum likelihood, as proposed by Lambert and Royston (2009). In practical application and estimation of models, the chosen number of interior knots specifies the degrees of freedom (one plus the number of knots). So then, the PH(*d*) model is a PH model whose spline function has *d* degrees of freedom (*d* – 1 interior knots and when *d* > 1, two boundary knots).

The fit of estimated parametric models can be compared based on the Akaike information criterion (AIC), defined as the deviance plus $2k$, where k is the dimension of the model (the number of fitted parameters).

Alternatively, we can use the Bayes information criterion (BIC), which is the deviance penalized by adding $k \log n$, where n is the sample size. Because parametric models are estimated by the maximum likelihood method, both criteria, AIC and BIC, can be used to compare fitted models. However, since the Cox model is estimated by the maximum partial likelihood method, this model cannot be compared with parametric models based on AIC and BIC criteria (Royston and Lambert, 2011).

3.4.6 Multiple Failure-Time Data

Cleves (2000) describes multiple failure-time data as data when two or more events (failures) occur for the same subject or from identical events occurring to related subjects. The typical feature is that failure times are correlated within a cluster (subject or group), violating the independence of failure times assumption required in traditional survival analysis. As the author points out, failure events should be classified according to whether they have a natural order and recurrences of the same type of events. The events are supposed to be ordered when the second event cannot occur before the first event. On the contrary, unordered events can happen in any sequence.

There are more approaches to examining multiple failure-time data. Firstly, we can examine the time to the first event, ignoring additional failures. However, it means we do not use all available data. The second method is based on the available data analysis while accounting for the lack of independence of the failure times. Cleves (2000) suggests corresponding procedures for estimating these models using the Cox proportional hazard model. Under the proportional hazard assumption, the hazard function (3.72) of the *i*th cluster for the kth failure type is as follows:

$$h_k(t, Z_{ki}) = h_0(t)e^{Z_i{}'\beta}, \tag{3.98}$$

where Z_{ki} is a *p*-vector of possibly time-dependent covariates for *i*th cluster to the *k*th failure type. While we presume in equation (3.98) that the baseline hazard function is equal for every failure type, the baseline hazard function is allowed to differ by failure type in the following formula:

$$h_k(t, Z_{ki}) = h_{0k}(t)e^{Z_i{}'\beta}. \tag{3.99}$$

As Cleves (2000) suggests, the maximum likelihood estimates for the models (3.98) and (3.99) are obtained from Cox's partial likelihood function $L(\beta)$, assuming independence of failure times.

Concerning the analysis of multiple failure-time data, Cleves (2000) emphasizes the need to determine whether it is ordered or unordered data and select a suitable method for estimating models accordingly. In the case of unordered times, which is the case for rating analysis, it is first necessary to determine whether the events are of the same or different types. Similarly, deciding whether the baseline hazard is the same or different for all event failures is necessary. In any case, it is essential to implement the methods for correctly structuring the data, including identifying individual failure events.

3.5 Chapter Summary

The purpose of this chapter was to clarify the main motives for conducting an application study in this work. Therefore, the basic characteristics and meaning of micro approaches for measuring credit risk were provided at the beginning of this section. Subsequently, attention was paid to the research review, based on which it was determined which procedures are suitable for modelling individual credit risk and which results the selected authors reached. Hence, the motivation and the main goal were specified, and the main contribution to the current research was outlined.

The next part of the chapter was devoted to describing the econometric models used in the application part of this work, namely discriminant, regression and survival analysis. Since plenty of professional publications deal with these approaches in great detail and professionally, the selected methods were only briefly described in this section. The main attention was paid to understanding the main principles and the possible use of models.

Finally, the methods described in this chapter are applied in the following four sections. The first study aims at modelling the influence of selected factors on ratings and their downgrades. Next, we find out whether there is a relationship between rating and corporate bankruptcy rates. Then, we analyse the effect of selected corporate characteristics on the survival probability. Finally, we formulate parametric survival models based on the previous results and estimate bankruptcy rates and ratings.

Chapter 4

The Effect of Selected Factors on Rating and its Dynamics

The aim of this chapter is to provide an application of rating models based on the sample formed by companies from selected CEE countries. The application consists of two empirical studies, whereby both of which are focused on credit rating modelling. The primary purpose is to find the main factors affecting rating assessment and its dynamics, focusing on rating downgrade. In addition, attention is paid to the practical aspects of rating models and their application for rating prediction. The overall methodological objective of the following applications is to estimate models based on selected statistical approaches. Therefore, we compare the models and their ability to predict ratings, emphasize their differences and recommend the most suitable method for modelling our data.

The first study is focused on the rating analysis of selected European non-financial companies. The main objective is to estimate rating models and identify the key predictive variables for rating assessment. We analyse data from CEE countries that have not received adequate attention in recent research on rating, and in addition, there is a lack of certified ratings in the market. The rating models are developed based on two econometric methods: discriminant and logistic regression analysis.

The selected multivariable methods will be supplemented by applying survival analysis to assess the dynamics and behaviour of the rating over time. Using this approach, we examine what particularly affects the changes in rating. We will focus primarily on the rating downgrade, as this change can be considered the riskiest development from the creditor's point of view. The application of survival analysis will allow us to understand better rating behaviour and the role of selected variables in rating deterioration.

Rating models in both studies will be developed based on the published MORE Rating. Thus, we can compare whether the main elements of credit rating assessment and the time to rating downgrade are consistent and have a wider interpretation capacity.

The following structure corresponds to this chapter's objectives, as mentioned above in the text. Firstly, we focus on modelling corporate credit rating in Chapter 4.1. The study uses data from CEE countries and MORE Rating assessment, discriminant analysis and logistic regression methods. Next, we apply survival analysis to estimate the Cox model for a rating downgrade in Chapter 4.2. Finally, we summarise the main findings and draw overall conclusions.

4.1 Corporate Credit Rating Assessment Models

The main goal of this study is to assess the role of selected financial variables in corporate rating prediction. From the financial perspective, the study's main objective is to analyse the corporate credit rating based on real data and evaluate the key factors of rating assessment, which are essential for the financial decision-making. In addition, attention is paid to the association between corporate financial performance and rating assessment of non-financial companies. This study builds on and expands on the author's previous work (Novotná, 2012a; Novotná, 2012b; Novotná, 2013; Novotná, 2015).

The models' estimation is based on multinomial logistic regression and discriminant analysis. From the methodological perspective, we compare both approaches to find the most suitable model for rating prediction.

Thus, we can summarize the partial objectives of this study as follows:

- Assess the role of selected financial variables on rating assessment (financial objectives) and
- compare and evaluate models estimated by different statistical methods (methodological goal).

4.1.1 Description of Data

This study is focused on analysing corporate credit ratings from eight countries in Central and Eastern Europe (CEE): the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovakia, and Slovenia. Because global rating agencies in these countries do not rate many companies, the models are estimated based on MORE Rating[12] (the Multi-Objective Rating Evaluation). As part of issuing credit ratings, the MORE methodology complies with Regulation (EC) N. 1060/2009 of the European Parliament and the Council of 16 September 2009 (the Credit Rating Agencies Regulation). Therefore, since July 10th, 2015, it has been registered as a credit rating agency per this regulation[13]. Thus, the MORE Rating assessment is completely in line with the objectives of this work.

[12] MORE ratings classify companies similarly as rating agencies (Bureau van Dijk Electronic Publishing, 2008). The MORE rating is calculated using a unique model that references the company’s financial data to create an indication of the company’s financial risk level.
[13] MORE Rating by the first Fintech Rating Agency – S-Peek, online access: https://www.s-peek.com/en/more-rating (17th June, 2019)

The used dataset[14] includes 1249 very large industrial companies from the mining, manufacturing, and construction sectors for 2002 – 2007. We record annual rating assessments and selected financial data for each company and the observed period. In addition, we recognize when the company was established or first rated. Thus, we have several observations for each company according to individual years, which gives a total of 6646 cases with an assigned rating for all years.

The dataset is divided into an experimental and hold-out sample, where the experimental sample is used for estimation and the hold-out sample for validation of models. The samples are prepared in two ways, and they are referred to as random and non-random in the following text.

Table 4-1 summarises the first random way of dataset split. All companies' and years' observations are randomly divided into the experimental sample (83%) and the hold-out sample (17%).

Table 4–1 Experimental and hold-out sample I (random)

Rating	**Code**	**Experimental sample**		**Hold-out sample**	
		Number of observations	**Percentage**	**Number of observations**	**Percentage**
B	1	277	5.51	91	5.62
BB	2	893	17.76	290	17.92
BBB	3	2353	46.80	719	44.44
A	4	1212	24.11	414	25.59
AA	5	292	5.83	104	6.43
Total		*5028*	*100*	*1618*	*100*

Table 4-2 shows the second way of splitting the original sample. In this case, the experimental sample (84.3%) contains observations from the first five years (2002-2006) and the hold-out sample (15.7%) data from the last year, 2007. Although the hold-out sample does not have observations in the outer categories B and AA, it will be used for validation because it reflects the actual rating assessment in this particular year. Compared to the first approach, this separation is given by the year of observation, and for this study, the dataset will be considered non-random.

The partial aim of this study is to compare the models developed through the two ways of data splitting. While the random sample ignores the time continuity of the rating, the non-random sample is estimated on previous ratings and validated

[14] Data on companies were obtained from the Amadeus database based on the following filters: They come from the eight CEE countries mentioned in the text, are very large companies from three industries (mining, manufacturing, construction), have MORE Rating from AAA to B, and have data on selected indicators for the set period.

based on the rating of the following year. The hypothesis is that the non-random approach might provide a more accurate predictive rating model. Thus, we also focus on the sample selection effect from the methodological perspective.

Table 4–2 Experimental and hold-out sample II (non-random)

Rating	Code	Experimental sample		Hold-out sample	
		Number of observations	Percentage	Number of observations	Percentage
B	1	368	6.57	0	0
BB	2	1183	21.12	644	61.63
BBB	3	2428	43.35	312	29.86
A	4	1314	23.46	89	8.52
AA	5	308	5.5	0	0
Total		*5601*	*100*	*1045*	*100*

4.1.2 Description of Input Variables

Regardless of the sample, this study aims to analyse the relationship between corporate rating and financial variables that reflect a company's size, profitability, capitalization, liquidity, and interest coverage. In addition, some variables were transformed using the logarithm to adjust for different measurement scales for further analyses (Table 4-3). Finally, we consider ten financial variables as predictors with a potential effect on rating: logarithm of total assets (*lnta*), return on assets (*roa*), return on equity (*roe*), equity to total assets (*eqta*), logarithm of interest coverage (*lnintcov*), logarithm of liquidity ratio (*lnliqr*), logarithm of cash flow (*lncf*), logarithm of current ratio (*lncurr*), logarithm of long-term debt to total assets (*lnltdta*) and the ratio of ebitda to total debt (*ebitdar*).

Table 4–3 Financial variables

Category	Economic rationale	Financial variables	
Size	Adequate protection	Total assets	*lnta*
Profitability	Ability to earn satisfactory returns	Return on total assets Return on equity The ratio of EBITDA to total debt	*roa* *roe* *ebitdar*
Capitalization	A measure of capital structure and leverage	Long-term debt to total assets Equity to total assets	*lnltdta* *eqta*
Liquidity	The flow of financial resources	Current ratio Liquidity (acid) ratio Cash flow	*lncurr* *lnliqr* *lncf*
Interest coverage	Ability to service the financial charges	Interest cover	*lnintcov*

This study aims to apply logistic and discriminant analysis, estimate credit rating models, assess their classification ability, and evaluate the suitability of the methods used for credit rating modelling. We develop rating models according to the used dataset and input variables, as summarized in Table 4-4.

Table 4–4 Rating models

	Discriminant analysis	**Logistic regression analysis**
Method	Linear	Multinomial Ordinal
Dataset	Random Non-random	Random Non-random
Number of categories	5 3	5 3
Number of predictors	10	10

The models are developed based on the following procedure and criteria:

- **Method**: We apply linear discriminant analysis and two methods of logistic regression analysis (multinomial - MLR, ordinal - OLR).
- **Dataset**: We develop models using both random and non-random samples, as described in the above text.
- **The number of categories**: We use five and three rating categories as dependent variables in the models.
- **The selection of predictors** (variables): The potential association between each independent variable and rating was first examined through the univariable analysis ($p<0.2$). Based on the results, we use ten financial variables as predictors with a potential effect on the rating.

4.1.3 Discriminant Models

The four discriminant models are presented in detail in Appendix 2. The overview of all estimated models is summarized in Table 4-5.

Table 4–5 Discriminant models

Model	**No. of predictors**	**No. of categories**	**Sample**
Model 1	10	5	Non-random
Model 2	10	5	Random
Model 3	10	3	Non-random
Model 4	10	3	Random

In the following sections, we explain and interpret the following characteristics of estimated models, as presented in Appendix 2:

- The general description of the estimation sample (by rating),

- canonical linear discriminant analysis results that show the canonical correlation for each dimension (the number of groups minus one),
- classification functions, which are functions used for classifications of subjects (the subject is assigned to the group with the highest value),
- standardized canonical function coefficients (loadings), which apply to variables that have been standardized using the pooled within-group covariance matrix,
- canonical structure, which represents correlations between each discriminating variable and the discriminant functions.

a) Main Results and Interpretation

The overall results show that the discriminant models are similar and have minor differences. Model 1 (LDA, non-random, five categories) will be described in more detail in the next part of the text, and the other models can be understood accordingly. In addition, some tables from Appendix 2 (Table 2-A) are rewritten for easier commentary in the following text. However, since the interpretation of the other three models is analogous, it is unnecessary to describe them in more detail.

The estimated coefficients of model 1 are summarized in Table 4-6. Since we have five groups, we get standardized coefficients of four discriminant functions. We use the estimated coefficients to assess the relative importance and relationship between the discriminating variables and the discriminant functions. The actual signs of the coefficients are arbitrary, and we compare just coefficients with the same sign to determine how these variables relate to the groups. For example, large *roa, eqta, lnintcov, lncf, lnliqr, lncurr, ebitdar* and low values *roe, lnta, lnltdta* result in large values of the discriminant score in the first function. However, some of the coefficients are close to zero. For example, the variable *ebitdar* has a negative standard coefficient in the third function but is not as distinct as the difference in the first function.

Additionally, the contribution of each variable to the discriminant score can be assessed based on Pearson correlations between function and variable values. The correlations are presented in a structure matrix in which the variables are sorted based on the absolute values of the correlation coefficients (Canonical structure in Appendix 2). According to the structure matrix, we can see that *roa, eqta, lnintcov, lncurr* have the highest association with the discriminant scores; however, there are differences among particular functions. For example, on average, the lowest contribution to the score is associated with *lnta*.

The eigenvalues of discriminant functions are used to assess how strongly the functions are related to groups. For each group, the eigenvalue is the ratio of the between-groups to the within-groups sum of squares for the discriminant functions scores. According to Appendix 2 (Canonical linear discriminant analysis), the first function has the largest canonical correlations (4.9258), indicating that the first function is strongly related to rating groups.

Table 4–6 Standardized canonical discriminant function coefficients (model 1)

Variable	**Function**			
	1	**2**	**3**	**4**
roa	0.8919	1.2621	0.8018	0.2660
roe	-0.3216	-0.4274	-0.0586	0.0642
eqta	0.7286	-0.5229	0.3506	0.2142
lnta	-0.3680	1.6715	1.7264	0.7928
lnintcov	0.4202	-0.0856	-0.6401	-0.2755
lncf	0.3627	-1.6822	-2.1066	-0.8115
lnliqr	0.2756	-0.0771	0.5208	-0.7117
lncurr	0.0955	-0.0001	-0.7138	1.0340
lnltdta	-0.0644	-0.0112	0.0435	0.1917
ebitdar	0.0371	0.1531	-0.5950	0.3354

b) Classification ability

The coefficients of classification functions (Fisher's linear discriminant functions) of the estimated models are shown in Table 4-7. The classification functions are used to classify individual cases: First, the values of five functions are computed and compared. Then, the group corresponding to the function with the highest value is selected as the target rating category.

Table 4–7 Classification function coefficients (model 1)

Variable	**1**	**2**	**3**	**4**	**5**
roa	1.2600	1.4491	1.6511	2.0131	2.7745
roe	0.1019	0.0861	0.0694	0.0346	-0.0262
eqta	9.2790	22.2835	37.4865	51.4774	62.9647
lnta	34.2175	33.1949	31.7998	30.5193	32.5883
lnintcov	0.3915	1.0540	1.9416	3.3279	4.1108
lncf	-26.6527	-25.8531	-24.6153	-23.3636	-25.4225
lnliqr	-4.5765	-2.4185	-1.2134	0.0995	1.8487
lncurr	-0.3275	-1.6194	-0.9198	0.0149	0.0925
lnltdta	0.4275	0.3187	0.2539	0.1220	0.0150
ebitdar	-0.3804	-0.4616	-0.4653	-0.4115	-0.3934
constant	-77.2415	-73.9966	-77.9575	-93.3122	-126.6945

For example, using the mean, minimum and maximum values of input variables representing three hypothetical companies, the classification functions assign them to different rating groups (Table 4-8). The average company is given to the middle group 3 – BBB. This result is not surprising because, as mentioned above, most companies have the middle rating assessment. Since the groups are not equally sized, the classification functions are weighted more heavily to classify group three in our case. The hypothetical company with the minimum (maximum) values is classified as group 1 – B (5 – AA). This result is also not surprising, as the higher value of most variables is generally associated with a better company's financial situation.

Table 4–8 Example of classification

	1	**2**	**3**	**4**	**5**
Mean	65.48	74.35	**78.48**	74.84	60.93
Minimum	**33.05**	28.52	13.21	-10.44	-31.02
Maximum	121.33	122.26	143.74	175.39	**191.59**

The main characteristics of the classification ability of models are summarised in Table 4-9 . The criteria for comparison and ranking the models are classification tables (confusion matrices) and error rates:

- The resubstitution classification table, obtained by classifying the observations used to build the discriminant model, is Class. (ES).
- The classification table, based on the hold-out sample's estimation ability, is called Class. (hold).
- The error rate, which represents the overall error rate and the error rate for each group and corresponds to the classification table, is based on the count-based estimate.

Table 4–9 Percentage correctly classified (PCC) – 5cat models

Method	**Model**	**Experimental sample (ES)**	**Number of obs. (ES)**	**Class. (ES)**	**Class. (hold)**
LDA	1	non-random	3518	0.8511	0.8775
LDA	2	random	3274	0.8525	0.8600

Although the classification ability of models is similar (Table 4-9), it differs across the rating groups (Table 4-10). For example, while only 8.35% of subjects from rating 3 are misclassified, it is 47.01% from rating 1. The least accurate is, therefore, the classification into rating 1, i.e. category B. From the lender's point of view, the model tends to reduce the classification ability of the companies with the worst ratings. The results of the second model are proportionally similar.

Table 4–10 Error rate (5 cat)

Model	**Misclas.**	**Rating**					**Total**
		1	**2**	**3**	**4**	**5**	
1	Error rate	0.4701	0.1978	0.0835	0.1816	0.2123	0.1489
	Priors	0.3326	0.2041	0.4801	0.2317	0.0509	
2	Error rate	0.4205	0.2750	0.0677	0.1745	0.2473	0.1476
	Priors	0.0269	0.1677	0.5100	0.2398	0.0559	

According to the overall error rate criterion, both five-category discriminant models achieve a similar classification ability. For example, the total error rate of model 1 shows that the proportion of misclassified observations is 14.89%.

Overall, the non-random LDA model has a better classification ability on the hold-out sample. However, the general results of both models do not allow a clear choice of a more suitable model. Furthermore, the classification accuracy of these models is strongly influenced by the fact that the boundary categories are not evenly represented in the sample. Therefore, we eliminate this shortcoming by deriving models for only three internal rating categories: 2, 3, and 4 (Table 4-11).

Table 4–11 Percentage correctly classified (PCC) – 3cat models

Method	Model	Experimental sample (ES)	Number of obs. (ES)	Class. (ES)	Class. (hold)
LDA	3	non-random	3222	0.8858	0.8986
LDA	4	random	3274	0.8878	0.8882

Unsurprisingly, both the classification ability and the error rate provide better results. The detailed error rates are summarized in Table 4-12.

Table 4–12 Error rate (3 cat)

Model	Misclas.	Rating			Total
		2	3	4	
3	Error rate	0.1671	0.0787	0.1411	0.1142
	Priors	0.2228	0.5242	0.2529	
4	Error rate	0.2332	0.0641	0.1300	0.1122
	Priors	0.1828	0.5560	0.2613	

In this case, the classification ability of the 3-category models is similar. However, model 3 (non-random) achieves a lower misclassification at the lowest rating and higher classification accuracy on the hold-out sample. For this reason, this model can be considered more suitable for rating prediction.

4.1.4 Multinomial Logistic Models

Since we use two logistic regression analysis methods, we estimate eight models (Appendix 3). The models are summarized in Table 4-13.

Table 4–13 Overview of logistic models

Model	Method	No. of predictors	No. of categories	Sample
Model 5	MLR	10	5	Non-random
Model 6	MLR	10	5	Random
Model 7	OLR	10	5	Non-random
Model 8	OLR	10	5	Random
Model 9	MLR	10	3	Non-random
Model 10	MLR	10	3	Random
Model 11	OLR	10	3	Non-random
Model 12	OLR	10	3	Random

a) Main Results and Interpretation

The overall results based on the logistic regression suggest that the estimated models do not differ significantly. Therefore, similarly to discriminant analysis, only one model will be described in more detail in the following text. The interpretation for other models is analogous.

The following text focuses on model 5 (MLR, non-random, five categories). However, we will pay attention to both approaches because we use two methods. In addition, some tables from Appendix 3 are rewritten in the next sections for better clarification. First, four logit functions are estimated because it is a five-category model. We arbitrarily use the middle rating category 3 (BBB) as a reference value. Thus, we form four logits, $g_1(x), g_2(x), g_4(x), g_5(x)$, comparing each group to the reference category. Equations (3.40) – (3.43) are used to estimate the unknown parameters based on the maximum likelihood method to fit the model. The parameter estimates are summarized in Table 4-14, and the details are provided in Appendix 3. The assessment of parameter estimates and their significance is based on the Wald test used to test the null hypothesis that each of the individual coefficients is zero for each logit.

Table 4–14 MLR parameter estimates (model 5)

Variable	Rating			
	1	2	4	5
roa	-0.8719*	-0.3489*	0.4810*	0.8939*
roe	0.0288*	0.0253*	-0.1126*	-0.2060*
eqta	-50.6921*	-25.1934*	7.2455*	1.7041
lnta	2.3379*	1.9669*	-1.1749*	-1.0128
lnintcov	-1.8130*	-1.3457*	2.2807*	4.3186*
lncf	-.19509*	-1.8375*	1.1257*	1.3755
lnliqr	-4.5343*	-2.4685*	1.4070*	2.7621*
lncurr	-3.6769*	-2.6221*	0.9934*	0.4867
lnltdta	0.3388*	0.1937*	-0.1992*	-0.3323*
ebitdar	-0.0629*	-0.0588*	-0.9185*	-3.4539*
constant	12.2038*	9.1524*	-12.0212*	-26.3962*

*significant at .05 level

The parameter estimates compare pairs of outcome variables (note that Rating 3 is a reference category). Thus, for example, the first part of the table labelled Rating 1 compares this category against Rating 3, Rating 2 compares this category against Rating 3 and so on. Therefore, the interpretation is similar to the binary logistic regression. The coefficients in Table 4-14 are expressed in terms of the log odds. For example, the coefficient -0.8719 (Rating 1, *roa*) implies that a one-unit change in *roa* results in a -0.8719 unit change in the log of odds.

However, we typically prefer using the odds ratios for interpretation, computed by exponentiating the coefficients (StataCorp, 2021; UCLA, 2021a; UCLA, 2021b). The odds ratios of model 5 are summarized in Table 4-15.

Table 4–15 MLR odds ratios (model 5)

Variable	Rating			
	1	**2**	**4**	**5**
roa	0.4182	0.7055	1.6177	2.4446
roe	1.0292	1.0256	0.8935	0.8138
eqta	0.0000	0.0000	1401.7826	5.4964
lnta	10.3595	7.1485	0.3088	0.3632
lnintcov	0.1632	0.2604	9.7835	75.0834
lncf	0.8228	0.1592	3.0824	3.9571
lnliqr	0.0107	0.0847	4.0837	15.8331
lncurr	0.0253	0.0727	2.7004	1.6269
lnltdta	1.4033	1.2137	0.8194	0.7173
ebitdar	0.9390	0.9429	0.3991	0.0316

The multinomial logistic model is a simple extension of the binary model. However, the interpretation is more difficult because of relevant binary comparisons. For example, with five outcomes (rating groups), we have ten binary comparisons for one model: R1 versus R2, R1 versus R3, R1 versus R4, R1 versus R5, R2 versus R3, R2 versus R4, R2 versus R5, R3 versus R4, R3 versus R5, and R4 versusR5.

To demonstrate the interpretation of odds ratios, we compare Rating 1 (B) against the reference category, Rating 3 (BBB):

- The odds ratio of *roa* is 0.4182, meaning that as *roa* increases, the company is likely to get a rating 3, assuming that all other variables in the model are held constant. Similarly, for *eqta, lnintcov, lncf, lnliqr, lncurr* and *ebitdar* (their odds ratios are less than one).
- On the other hand, the odds ratio of *roe* is 1.0292. Thus, as roe increases, the company is likely to get rating 1 relatively to rating 3, similarly, for *lnta* and *lnltdta* (their odds ratios are more than one).

Based on the estimated coefficients, the four logit functions can be written as:

$$g_1 = -0.87roa + 0.03roe - 50.69eqta + 2.34lnta - 1.81lnintcov -$$
$$- 0.2lncf - 4.53lnliqr - 3.68lncurr + 0.34lnltdta - 0.06ebitdar + 12.2,$$

$$g_2 = -0.35roa + 0.03roe - 25.19eqta + 1.97lnta - 1.35lnintcov -$$
$$-1.84lncf - 2.47lnliqr - 2.62lncurr + 0.19lnltdta - 0.06ebitdar + 9.15,$$

$$g_4 = 0.48roa - 0.11roe + 7.25eqta - 1.17lnta + 2.28lnintcov +$$
$$+1.26lncf + 1.41lnliqr + 0.99lncurr - 0.2lnltdta - 0.92ebitdar - 12.02,$$

$$g_5 = 0.89roa - 0.21roe + 1.7eqta - 1.01lnta + 4.32lnintcov +$$
$$+1.38lncf + 2.76lnliqr + 0.49lncurr - 0.33lnltdta - 3.45ebitdar - 26.4.$$

The conditional probability for each rating category can be expressed as:

$$\pi_1\left(Y=1\middle|\mathbf{x}\right)=\frac{e^{g_1(\mathbf{x})}}{1+e^{g_1(\mathbf{x})}+e^{g_2(\mathbf{x})}+e^{g_4(\mathbf{x})}+e^{g_5(\mathbf{x})}},$$

$$\pi_2\left(Y=2\middle|\mathbf{x}\right)=\frac{e^{g_2(\mathbf{x})}}{1+e^{g_1(\mathbf{x})}+e^{g_2(\mathbf{x})}+e^{g_4(\mathbf{x})}+e^{g_5(\mathbf{x})}},$$

$$\pi_3\left(Y=3\middle|\mathbf{x}\right)=\frac{1}{1+e^{g_1(\mathbf{x})}+e^{g_2(\mathbf{x})}+e^{g_4(\mathbf{x})}+e^{g_5(\mathbf{x})}},$$

$$\pi_4\left(Y=4\middle|\mathbf{x}\right)=\frac{e^{g_4(\mathbf{x})}}{1+e^{g_1(\mathbf{x})}+e^{g_2(\mathbf{x})}+e^{g_4(\mathbf{x})}+e^{g_5(\mathbf{x})}},$$

$$\pi_5\left(Y=5\middle|\mathbf{x}\right)=\frac{e^{g_5(\mathbf{x})}}{1+e^{g_1(\mathbf{x})}+e^{g_2(\mathbf{x})}+e^{g_4(\mathbf{x})}+e^{g_5(\mathbf{x})}}.$$

b) Model fitting

The overview and main characteristics of estimated models are summarized in Table 4-16, including:

- Deviance ($D=-2LL_M$): The value of a likelihood-ratio chi-squared for the test of the null hypothesis that all the coefficients associated with independent variables are simultaneously equal to zero (including degrees of freedom as the number of constrained parameters),
- pseudo R^2 measured as McFadden's,
- AIC and BIC.

Likelihood ratio tests for the overall models test the null hypothesis that all coefficients in the model are zero. Firstly, we calculate the value -2 log-likelihood for models with only an intercept term and all variables. Then, we get the difference between these values, Chi-square. If the observed significance is small, we can reject the null hypothesis that all coefficients are zero and conclude that the final model is significantly better than the intercept-only. According to the values in Table 4-16, we conclude that all models are better than the intercept-only models.

Table 4–16 Model-fitting

Model	**Deviance**	**LR Chi2(df)**	**Pseudo R^2**	**AIC**	**BIC**
Model 5	2176.797	6830.33 (40)	0.7583	2264.794	2536.082
Model 6	2004.445	6135.13 (40)	0.7537	2092.445	2360.571
Model 7	2683.156	6323.97 (10)	0.7021	2711.156	2797.475
Model 8	2450.653	5688.92 (10)	0.6989	2478.653	2563.965
Model 9	1509.218	5068.91 (20)	0.7706	1553.218	1686.928
Model 10	1347.236	4586.88 (20)	0.7730	1391.236	1523.406
Model 11	1812.961	4765.17 (10)	0.7244	1836.961	1909.894
Model 12	1602.200	4331.91 (10)	0.7300	1626.200	1698.292

Next, we consider the information criteria AIC and BIC (for the description of the measure, see, for example, Long and Freese, 2014). Based on the criterion BIC, we prefer model 6 among 5-cat models and model 10 among 3-cat models (they have the smallest value of BIC). These results are also suggested by the criterion AIC.

c) Classification ability

Similar to the discriminant analysis, we use the model to determine the rating of a hypothetical firm with average, minimum and maximum values of input variables. However, unlike the discriminant analysis, we calculate the probability of belonging to a group and classify it into the group with the highest probability (Table 4-17).

Table 4–17 Example of classification

	1	**2**	**3**	**4**	**5**
Mean	4.89E-05	9.98E-	**9.94E-01**	4.89E-	7.37E-10
Min	**1.00E+00**	1.81E-	2.02E-24	2.63E-	1.98E-23
Max	5.56E-45	2.42E-	**1.00E+00**	4.94E-	0.00E+00

The average company is assigned to the middle rating 3 – BBB. The hypothetical company with the minimum (maximum) values is rated 1 – B (3 – BBB). Compared with the discriminant model (Table 4-8), the logistic model predicts a different rating group using the minimum values of predictors.

Next, we examine the classification accuracy of estimation and hold-out samples (Table 4-18). All models achieve a relatively high overall classification accuracy, and their differences are small. For example, the highest classification accuracy on a hold-out sample is achieved by Model 9. Overall, it is clear that MLR models have a higher classification accuracy than OLR models. Also, 3-cat models perform better than 5-cat models.

Table 4–18 Percentage correctly classified (PCC)

Model	**Class. (ES)**	**Class. (hold)**
Model 5	0.8801	0.8850
Model 6	0.8861	0.8851
Model 7	0.8649	0.8468
Model 8	0.8494	0.8600
Model 9	0.9096	0.9060
Model 10	0.9105	0.8977
Model 11	0.8818	0.8882
Model 12	0.8986	0.8811

4.1.5 Comparison of Estimated Rating Models

Our application developed classification rules for five and three rating classes, which is more complex than the binary task. Therefore, we will combine multiple ROC curves to assess the performance of estimated models. As mentioned in

Chapter 3.2.4, there are two main approaches to multiple ROC analysis: Each class versus the union of other classes, or distinct pairwise-class ROC curves. Both methods are suitable for summary statistics, such as the AUC (Area Under the Curve).

We use the first approach in our application to compare the ability to predict a category versus a union of other categories. This way is sufficient for our purposes and effective for overall comparison. Thus, for each 5-cat model, we produce ROC curves and calculate the AUC as follows:

- Cat1 versus the union of other categories (cat2 + cat3 + cat4 + cat5),
- cat2 versus the union of other categories (cat1 + cat3 + cat4 + cat5),
- cat3 versus the union of other categories (cat1 + cat2 + cat4 + cat5),
- cat4 versus the union of other categories (cat1 + cat2 + cat3 + cat5),
- cat5 versus the union of other categories (cat1 + cat2 + cat3 + cat4).

The procedure is analogical for 3-cat models when considering only three rating categories. The ROC analysis is based on a parametric model, using the maximum likelihood estimation. We analyse the whole experimental and hold-out sample to determine whether the classification ability varies with the used sample selection. Preferably, we focus on the classification ability of the hold-out sample that is not used to estimate models.

Table 4–19 AUC (5-cat models)

Model/AUC	Cat1	Cat2	Cat3	Cat4	Cat5
LDA (Model 1)	0.9658	0.9876	0.9689	0.9803	0.9836
Est. sample	0.9647	0.9887	0.9793	0.9840	0.9797
Hold-out	++	++	0.9199	0.9696	0.9931
LDA (Model 2)	0.9444	0.9815	0.9689	0.9804	0.9836
Est. sample	0.9556	0.9814	0.9676	0.9813	0.9721
Hold-out	0.9200	0.9816	0.9730	0.9779	0.9998
MLR (Model 5)	0.9950	0.9942	0.9807	0.9902	0.9955
Est. sample	0.9947	0.9956	0.9892	0.9940	0.9957
Hold-out	++	++	0.9253	0.9741	0.9967
MLR (Model 6)	0.9946	0.9927	0.9807	0.9902	0.9955
Est. sample	0.9926	0.9917	0.9794	0.9912	0.9924
Hold-out	**0.9987**	**0.9953**	**0.9849**	**0.9866**	**0.9998**
OLR (Model 7)	0.9890	0.9867	0.9654	0.9798	0.9800
Est. sample	0.9884	0.9884	0.9749	0.9823	0.9746
Hold-out	++	++	0.9130	0.9701	0.9829
OLR (Model 8)	0.9879	0.9832	0.9654	0.9798	0.9800
Est. sample	0.9854	0.9817	0.9635	0.9808	0.9712
Hold-out	0.9944	0.9878	0.9713	0.9772	0.9975

++ not sufficient data to perform ROC analysis; random models (white), nonrandom models (grey)

Firstly, we compare the 5-cat models. Based on the AUC values of 5-cat models (Table 4-19), the classification ability is sufficient, and there are only minor differences among the models. Nevertheless, we conclude that the best classification ability is performed by model 6 (MLR, random). We can also see that the best model is estimated by multivariate regression analysis no matter what sample we use (random, nonrandom).

Table 4–20 AUC (3-cat models)

Model/AUC	**Cat2**	**Cat3**	**Cat4**
LDA (Model 3)	0.9913	0.9604	0.9927
Est. sample	0.9934	0.9732	0.9951
Hold-out	++	0.8992	0.9803
LDA (Model 4)	0.9870	0.9600	0.9923
Est. sample	0.9858	0.9589	0.9935
Hold-out	0.9911	0.9633	0.9877
MLR (Model 9)	0.9900	0.9711	0.9960
Est. sample	0.9918	0.9822	0.9979
Hold-out	++	0.9121	0.9837
MLR (Model 10)	0.9936	0.9753	0.9966
Est. sample	0.9929	0.9738	0.9974
Hold-out	**0.9957**	**0.9798**	**0.9937**
OLR (Model 11)	0.9912	0.9584	0.9930
Est. sample	0.9931	0.9718	0.9954
Hold-out	++	0.8945	0.9803
OLR (Model 12)	0.9875	0.9586	0.9923
Est. sample	0.9862	0.9571	0.9942
Hold-out	0.9911	0.9635	0.9894

++ not sufficient data to perform ROC analysis; random models (white), nonrandom models (grey)

Next, we compare 3-cat models (Table 4-20). Overall, the classification ability is slightly higher compared to the 5-cat models. However, all models perform sufficient classification ability, and there are minor differences among the AUC. The results support the main findings from the 5-cat models because the best classification ability is performed by model 10 (MLR, random).

The ROC curves are presented only for 5-cat models (Appendix 4). They reflect the data on AUC shown in Table 4-19. The classification of models is similar and relatively high because we test the ability to predict one category against the union of other categories. Overall, we prefer model 6, which provides the best results.

4.1.6 Summary of Results

We estimated twelve rating models in this study. We assumed ten financial variables and five or three output categories. In addition, we used two samples to determine whether sample selection affects the models and their prediction ability.

All estimated models suggest that all used financial variables are good predictors of rating. The main findings of this study provide evidence that accounting-based variables significantly impact corporate rating in our sample. However, the direct effect on particular ratings is not easily interpretable and must be explained in the context of estimated models.

The impact of selected variables on rating in discriminant models is measured by estimating the coefficients' contributions. However, since the contribution of the variables for the other variables in the model differs in each discriminant function (standardised coefficients), it is not conceivable to draw clear conclusions. Thus, ignoring other variables in the models, the association of individual variables with each discriminant function is compared through the correlation coefficients (structure matrix). According to the structure matrix, the highest average association with discriminant functions apply for *roa, eqta, lnintcov, lnliqr* and *lncurr*.

The impact of the variables on rating in the logistic regression models is determined through their statistical significance, particularly in logit functions. For example, variables *eqta, lnta, lncurr* and *ebitdar* are not statistically significant in some logistic functions; thus, they are not considered key rating factors. Furthermore, based on the logistic models, the higher the value of *roa, lnintcov, lnliqr* and *lncf,* the greater the probability of a better rating assessment. Contrary, greater values of *roe* and *lnltdta* increase the likelihood of a lower rating.

To summarise all partial results of the role of financial variables, we conclude that *roa, roe, lnintcov*, and *lnliqr* achieve the highest correlations with discriminant functions and are statistically significant in all logit functions. Thus, they are considered as the primary factors of rating prediction, followed by *lncf* and *lnltdta*.

We provide evidence that the following financial variables are the main factors of rating assessment:

- Return on total assets,
- return on equity,
- interest cover,
- liquidity ratio,
- cash flow,
- long-term debt to total assets.

The classification accuracy of all estimated models was determined based on the overall percentage correctly classified (PCC). The PCC, or hit ratio, ranges from 88% to 90.6% for a hold-out sample for our models. To assess the overall classification ability, we must set an acceptable level and compare the hit ratio to the standard. Firstly, we determine the percentage that could be determined correctly by chance. Since we compare the hit ratio for unequal group sizes, we consider just the largest group. For example, the largest random sample group in our study is represented by rating BBB (2353 observations in the experimental sample and 719 observations in the hold-out sample). Thus, we can arbitrarily assign all the subjects to the largest group. Therefore, if we classify each observation into this largest group in the case of random models, we would achieve a classification accuracy of 46.8% (experimental sample) and 44.4% (hold-out sample). Assuming nonrandom models, we get 43.4% (experimental sample) and 61.63% (hold-out sample). This approach is referred to as the maximum chance criterion, and unless a model achieves accuracy more than the computed values, it should be disregarded. Hair et al. suggest (2014) that the standard classification accuracy should be at least one-fourth greater than that achieved by chance. In our case, the standard is 58.5% and 54.3% (experimental sample), and 55.5% and 77% (holdout sample). All estimated models satisfy the maximum chance criterion. However, we should consider that this criterion provides only a rough estimate of the acceptable level of predictive accuracy. On the other hand, the criterion is relatively easy to apply (Hair et al., 2014).

The ROC analysis performed a more accurate and suitable diagnosis of predictive ability. Since we compare models with more than two output categories, we used the adjusted ROC analysis for a binary case. In our application, we tested the ability to predict a particular rating category (first binary output) versus the ability to predict the union of other remaining rating categories (second binary output). All estimated models achieve a high capability to predict a particular rating category against the different types – the values of AUC range from 0.8945 to 0.9998, suggesting sufficient predictive accuracy.

To summarize the predictive ability of estimated models, it is unsurprising that 3-cat models generally outperform the 5-cat models. However, among the 5-cat models, model 6 achieves the highest PCC on the hold-out sample, and we chose this model based on the ROC analysis. Thus, we conclude that the multinomial logistic regression analysis should preferably be used to estimate rating models. In addition, the results suggest that the sample selection method does not substantially affect the overall quality of models and their predictive accuracy.

Overall, we provide evidence that the statistical methods used in this study are suitable for credit rating modelling. The models are relatively simple to use and achieve sufficient predictive ability. Furthermore, the main findings show that logistic models gain a higher classification ability than discriminant models. Finally, when model assumptions, their interpretation and predictive power are considered, the multinomial logistic regression analysis seems to provide the most appropriate method for estimating the credit rating models.

In the next section, we extend the main results from the previous study by applying survival analysis to examine the main factors for a rating downgrade. The purpose of the following application is to analyse rating behaviour over time and find the main aspects of rating deterioration. Finally, we compare, discuss and summarize the main findings from both applications.

4.2 Modelling of Rating Downgrades Based on Multiple Failure-Time Data

The aim of this section is to examine the relationship between time and corporate rating downgrades, including the role of used variables on negative changes in rating. Furthermore, we extend the MORE rating analysis from the previous part by applying survival analysis methods. Thus, we get more information about the rating behaviour over time and the main factors of its deterioration.

In this study, the event, or failure in terms of survival analysis, is defined as a rating downgrade. As the rating can be downgraded more than once during the period under our observation, multiple failure-time analysis approaches should be used, as Cleves (2000) suggested, see Chapter 3.4.6. Therefore, this section will estimate rating models using the Cox proportional hazard model for multiple failure-time data. This study builds on and expands on the author's previous work (Novotná, 2019; Novotná, 2021).

This application aims to model the rating downgrade depending on time and annual changes in financial variables. Therefore, we use yearly changes in rating as the dependent variable and annual changes in financial variables as independent variables in our model. Specifically, the dependent variable will be a rating downgrade. This choice is natural because detecting a potential investment quality deterioration is crucial, increasing the credit risk.

Since we extend the main findings from the previous chapter on MORE Rating (Chapter 4.1), the models are estimated based on the same data sample. However, since we analyse the factors of a rating downgrade and use survival analysis methods, we must first prepare and adjust data for this application. In this study, we observe rating assessments and financial variables annually, which is 7494 observations.

4.2.1 Description of Data

The overall dataset description is presented in Table 4-21. We can see that the estimation sample used for modelling consists of 5694 observations. Next, we use two data structures referred to as “single” and “multiple”, which differ in the definition of the failure event:

- If we assume only one rating change for each company, the sample is called single,
- assuming that a rating can change repeatedly and we account for all these changes, the sample is called multiple.

Finally, our two datasets define 705 failures in single-record and 870 failure events in multiple-record data, referred to as a rating downgrade. Using two datasets with different specifications of failure events will allow us to examine whether the multiple-failure time analysis will lead to a more suitable model than the approach when considering only single events. We consider all rating downgrades as the same event type and do not distinguish the downgrade size. Regardless, the rating was changed by more than one degree in a marginal number of cases, so we will not consider this fact.

Table 4–21 Dataset description

	Event (Rating downgrade)	Estimation sample	Hold-out sample	Total
Single	No (Survival)	5169 (90.78%)	1620 (90%)	6789 (90.59%)
	Yes (Failure)	525 (9.22%)	180 (10%)	705 (9.41%)
	Total	5694 (75.98%)	1800 (24.02%)	7494 (100%)
Multiple	No (Survival)	5041 (88.53%)	1583 (87.94%)	6624 (88.39%)
	Yes (Failure)	653 (11.47%)	217 (12.06%)	870 (11.61%)
	Total	5694 (75.98%)	1800 (24.02%)	7494 (100%)

Similarly to the previous study on MORE Rating, we use the same financial variables. However, this application calculates their annual changes and examines their impact on the yearly rating downgrade. We compute annual growth rates (%) of financial variables and use them as input variables in the model. The data are adjusted for outliers (over the 95th percentile and below the 5th percentile) by replacing them with the closest value of not-outlier. The main characteristics of adjusted financial variables are summarized in Table 4-22. The mean values are generally positive, meaning all the variables increased on average during the observed period. Note that mean and median values vary, suggesting skewed distributions.

Table 4–22 Financial variables

Financial variable (annual change in %)	Symbol	Mean	Median
Total assets	*tag*	20.31	14.22
Return on assets	*roag*	34.68	-3.57
Return on equity	*roeg*	21.88	-8.94
EBITDA to total debt	*ebitdarg*	21.64	4.69
Equity to total assets	*eqtag*	9.26	3.82
Cash flow	*cfg*	39.37	20.84
Interest coverage	*intcovg*	60.18	18.25
Long-term debt to total assets	*ltdtag*	28.61	-12.26
Current ratio	*currg*	6.92	4.29
Liquidity ratio	*liqrg*	1.17	4.24

4.2.2 Application of Survival Models

The models are developed based on the Cox proportional hazards model. Since we consider single and multiple failure-time data, we estimate two models called single and multiple. First, we perform a survival analysis, considering only the first rating downgrade, ignoring additional ones for each company. The time is measured in years after the time of origin, set as the year 2002.

The Kaplan-Meier estimates of survival functions are shown in Figure 4-1. In both cases, the survival curves have a descending, stepped shape. We can see that the probability of survival, i.e. a stable or improved rating, is higher in the case of single data. It is a consequence of the assumptions used in the survival analysis, as in this case, only one event is allowed for each subject. The resulting models are summarised in Table 4-23. We can see estimated coefficients of independent variables (Coeff.), including the standard error in the brackets. Hazard ratios (HR) are also listed in the table for completeness. The final models include six statistically significant financial variables. The remaining insignificant variables were removed from the model (*ebitdarg, eqtag, ltdtag, currg*).

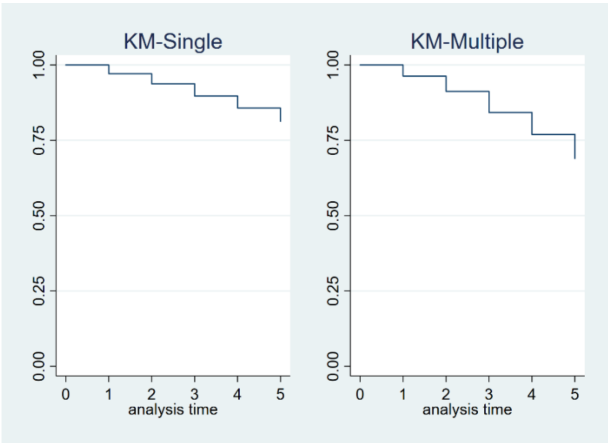


Figure 4–1 Kaplan-Meier survival estimates

If we compare both models, we see that the estimated coefficients are very similar. Thus, the effect of the used variables on the probability of survival, i.e. a stable or upgraded rating, is similar. Furthermore, all coefficients are statistically significant at a 0.05 level in both models. Estimated coefficients can be used to interpret the effect of individual variables, but it is more appropriate to use the hazard rates. For example, an increase in *tag* by one unit (annual change of total assets by one percent) increases the hazard of a rating downgrade by 1.16% in the single model, or 1.38% in the multiple model, respectively. The same effect has the variable *roeg*. Contrary, if the covariate *roag* increases by one unit, the hazard of rating downgrade decreases by 2.71% in single and 2.79% in multiple model. The same effect have variables *cfg*, *intovg* and *liqrg*.

Table 4–23 Estimated coefficients of Cox models

Variable	Coeff. Single	HR Single	Coeff. Multiple	HR Multiple
tag	0.0115* (0.002)	1.0116	0.0137* (0.002)	1.0138
roag	-0.0275* (0.002)	0.9729	-0.0283* (0.002)	0.9721
roeg	0.0195* (0.002)	1.0196	0.0212* (0.002)	1.0214
cfg	-0.0137* (0.002)	0.9864	-0.0173* (0.003)	0.9829
intcovg	-0.0071* (0.002)	0.9929	-0.0061* (0.001)	0.9939
liqrg	-0.0066* (0.002)	0.9934	-0.0063* (0.002)	0.9937

*significant at 0.05, standard error adjusted for 737 clusters

The overall results show that the annual growth rate in the variables *roa*, *cf*, *intcov* and *liqr* reduces the hazard of the rating downgrade. In contrast, the hazard is increased by the growth rate of *ta* and *roe*. In general, we conclude that all used variables can be used as significant indicators of rating downgrade.

4.2.3 Comparison of Baseline and Average Hazard

To assess the influence of variables on hazard, we determine the hazard for the so-called average company (H1), i.e. the values of the variables are equal to their mean values. Subsequently, we compare this hazard with baseline hazard, when the values of all variables are equal to zero (H0).

The graphical illustration of cumulative hazard functions is given in Figure 4-2. H1 lies below the baseline hazard, H0, in both graphs. It follows from the fact that baseline hazard corresponds to a situation where all covariates are zero. However, a zero annual change in the financial indicators means that the risk of a rating downgrade must be higher than the average company's yearly changes. Based on the figure, the hazard of a rating downgrade in time is more elevated in the multiple model. It is because we account for all rating downgrades for each company, not only the first ones, as in the single model.

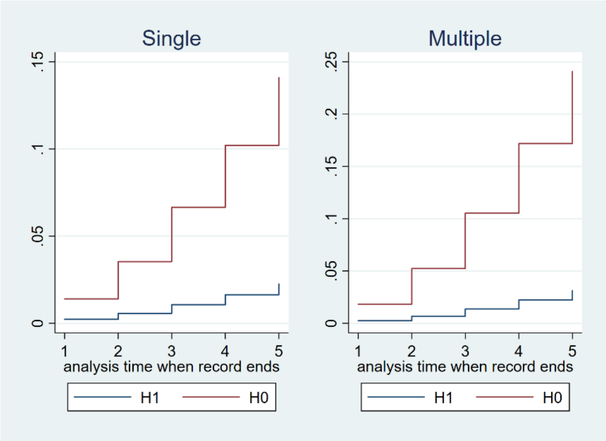


Figure 4–2 Baseline and average hazard

For comparison, we plot the cumulative hazard functions with median values (Figure 4-3). In other words, we construct the hazard function for a hypothetical middle-values company.

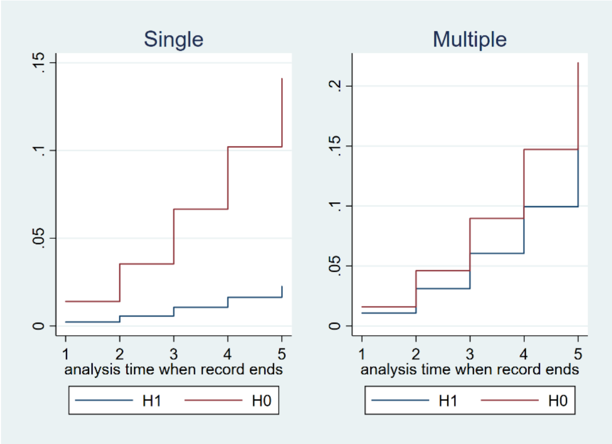


Figure 4–3 Baseline and median hazard

Relating Figure 4-2 with Figure 4-3, we can see that the median model's multiple hazard function (H1) is above the average model. It suggests that the middle company has a greater hazard of rating downgrade than the average company – the difference results from the data distributions.

In addition to the curves of cumulative hazard obtained by transforming the Kaplan-Meier product-limit estimator, we estimate hazard functions, *h(t)*. Cleves et al. (2010) suggest taking the steps of the Nelson-Aalen cumulative hazard and smoothing them with a kernel smoother. Figure 4-4 depicts the smoothed hazard functions for single (hs) and multiple (hm) models. The hazard functions are constructed for models with null values of variables (h0) and for a hypothetical company with i) mean and ii) medium values (h1). Comparing Figures 4-3, 4-4 and 4-5, or specifically the cumulative hazard with smoothed hazard curves, we can see that plotting ranges are narrower. As Cleves et al. (2010) explain, this is because kernel smoothing requires averaging values over a moving window of data.

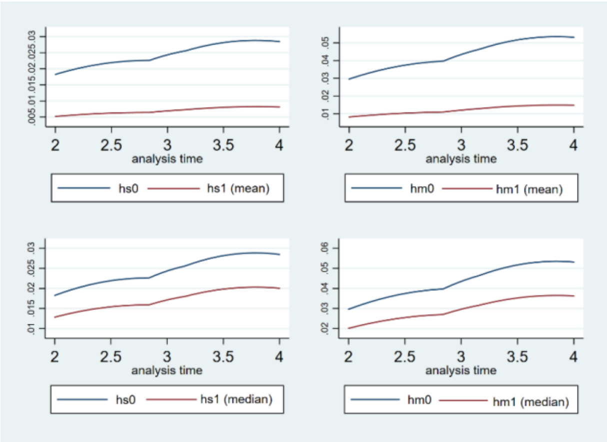


Figure 4–4 Smoothed hazard functions

4.2.4 Model Verification

Verifying whether the hazard functions are multiplicatively related is advisable when using the Cox model. Therefore, we assess the proportional hazards assumption by plotting the estimated hazards on a log scale. The lines in all graphs (Figure 4-5) seem parallel. Thus, we conclude that the proportionality assumption in both models is not violated.

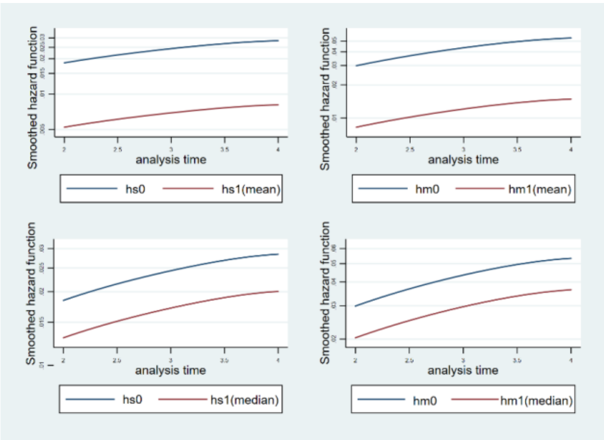


Figure 4–5 Smoothed hazard functions (log scale)

Table 4–24 Test of PH assumptions

Indep. variable	Single rho (Chi2)	Multiple rho (Chi2)
tag	0.0217 (0.17)	-0.0060 (0.02)
roag	0.0387 (0.59)	0.0270 (0.39)
roeg	0.0048 (0.01)	0.0260 (0.40)
cfg	0.0060 (0.03)	0.0131 (0.20)
intcovg	-0.0716* (6.55)	-0.0734* (10.27)
liqrg	0.0598 (2.05)	0.0403 (1.24)
Global	8.86	12.42

*significant at 0.05, standard error adjusted for 737 clusters

The test of the proportional-hazards specification is based on the Schoenfeld residuals after fitting the model. It is used to test the independence between residuals and time. The test results in Table 4-24 suggest that the hazard assumption is not proportional to the variable *intcovg*. Furthermore, we find no evidence that our specification violates the proportional-hazard assumption regarding other variables. Therefore, our specification does not violate the proportional-hazards assumption in both models based on the global test.

The models are evaluated by the overall model fit using Cox-Snell residuals. Figure 4-6 shows the Nelson-Aalen cumulative hazard estimator plots for Cox-Snell residuals for both models. We can see some variability around the 45°line, particularly in the right-hand tail. Cleves et al. (2010) argue that some variability is expected due to the reduced effective sample caused by prior failures and censoring. However, we can see that both graphs fit the data adequately based on the charts.

(a) Single (b) Multiple

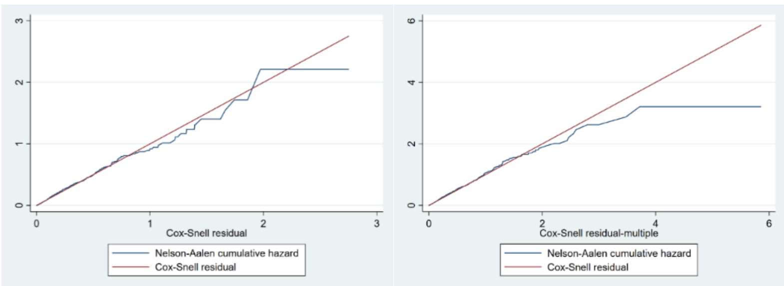


Figure 4–6 Cumulative hazard of Cox-Snell residuals

However, we cannot choose a better model based on the graphical illustration. For this reason, we evaluate the predictive power by computing the Harrell's C concordance statistics, which measures the agreement of predictions with observed failure order. The statistics can be defined as the proportion of all usable subject pairs in which the predictions and outcomes are concordant (Cleves et al., 2010). The values of C range between 0 and 1. Additionally, we can use Somers'D, which reports the rank correlation value, ranging from -1 to 1. Both measures are related as $D = 2(C - 0.5).$ As Cleves et al. (2010) state, a value of 0.5 Harrell's C and 0 of Somer's D indicate no predictive ability of the model.

The values of Harrell's C and Somer's D and their calculation procedure are shown in Table 4-25. Harrell's C values are 0.8586 (single) and 0.8705 (multiple). Thus, we can correctly identify the order of the survival times for pairs of subjects 85.86%, or 87.05% of the time. Although the results are similar, the values are slightly higher for the multiple model. These results are supported by the value of Somer's D, which is 0.7172 (single) and 0.7411 (multiple).

The results provide evidence that both models have sufficient predictive accuracy. We prefer the multiple model over the single one based on the values.

Table 4–25 Harrell's C and Somer's D

	Single model	**Multiple model**
Number of subjects (N)	3554	3554
Number of comparison pairs (P)	924134	984815
Number of orderings as expected (E)	793479	857329
Number of tied predictions (T)	0	0
Harrell's C = (E + T/2) / P	0.8586	0.8705
Somer's D	0.7172	0.7411

To conclude, the multiple failure-time data analysis leads to a more suitable model based on the statistical significance of the estimated coefficients and goodness of fit. On the other hand, it should be noted that both survival models are very similar based on estimated coefficients and used criteria.

4.2.5 Summary of Results

This study aimed to develop rating models using survival analysis methods. Specifically, we applied the Cox proportional hazards model to analyze the survival time until the event. In our case, we focused on using survival analysis to model the time to a rating downgrade. As a part of the analysis, we examined the effect of financial variables on the probability of negative annual rating change.

We provide evidence that annual changes in some variables are related to the rating downgrade, specifically using covariates *tag, roag, roeg, cfg, intcovg* and *liqrg*. On the contrary, variables *eqtag, ebitdarg, ltdtag* and *currg* are not statistically significant. Thus, we can conclude that annual changes in the following financial variables are good predictors of potential rating deterioration measured as rating downgrade in our sample:

- Total assets,
- return on total assets,
- return on equity,
- cash flow,
- interest cover,
- liquidity ratio.

Two different approaches were used to estimate the models, depending on whether we considered only one or more events (rating downgrades) for one company. First, the single model was derived, assuming that the event can occur only once for each subject. On the other hand, the multiple models accept that the event can occur repeatedly. Due to these different assumptions, the input data and structure also had to be adjusted.

The resulting models are presented based on the estimated coefficients for the variables used in the analysis. Both models are statistically significant, as are the estimated coefficients of the individual variables in the multiple model.

We used baseline hazard and the hazard of the so-called average (medium) company to interpret the models based on mean (medium) values of variables. The fit of both models was assessed using Cox residuals. Based on the main findings of this study, we conclude that the multiple model approach is more suitable for events that might occur repeatedly. The simple model should be preferably used when the survival time until the first event is a matter of interest. In other cases, we should use multiple failure-time data analyses, making data use better.

Overall, the findings of this study show that survival analysis is, in addition to typical financial problems, suitable for other types of tasks, such as the analysis of survival to bankruptcy or default. However, it is necessary to consider the specific data structure when applying it, especially whether the event can repeatedly occur for one subject or whether more events can occur for a given subject. In these cases, it is appropriate to use multiple failure-time analysis, which better corresponds to the problem.

4.3 Chapter Summary

The fourth chapter's main goal was to analyse corporate data of selected CEE countries and understand the influence of selected financial variables on the MORE rating. Therefore, we first focused on estimating rating models using discriminant analysis and two logistic regression methods. In total, we obtained twelve rating models, which were compared with each other. Next, we identified financial variables that can be considered predictive factors of the rating evaluation. Even though the individual models differed slightly in terms of their formulation and classification ability, the overall conclusions confirm the main role of used financial variables in rating assessment.

In the next section, we used the same data set and examined the relationship between financial variables and rating downgrades. In this case, we applied the method of survival analysis using the Cox model. Survival analysis allows us to estimate the survival probability of subjects to a predefined event. Since rating deterioration is quite a fundamental problem, especially for lenders, it is certainly important to recognize an impending rating change early. For this reason, the rating downgrade was chosen as the event in the survival analysis. Using the Cox model, we subsequently found six financial variables that have a fundamental connection with the annual deterioration of the rating assessment.

The main influential financial variables based on both studies are summarized in Table 4-26. The results confirm that common financial indicators influence the rating assessment and its annual deterioration, regardless of the method used or the output variable. These are five commonly used financial indicators in financial analysis and evaluation of the financial performance of companies. Even if these indicators seem basic and simple, they still play a major role in assessing the borrower's credit quality and, thus, the rating. Of course, it is necessary to take them in the context of their change and its possible impact on the rating, or rather on its deterioration.

Table 4–26 Main factors of rating and its downgrade

	Rating assessment	**Rating downgrade**
Total assets	x	✓
Return on total assets	✓	✓
Return on equity	✓	✓
Liquidity ratio	✓	✓
Cash flow	✓	✓
Interest cover	✓	✓
Long-term debt to total assets	✓	x

In general, it has been shown in this chapter that the first approach, based on discriminant analysis and logistic regression, and the second method, using survival analysis, lead to similar results. In addition, with the help of survival analysis, we could estimate the hazard and survival functions, which can be used to predict the probability of survival or the time until the rating deteriorates. Furthermore, it means we can look deeper into the development of the monitored variable over time and its dynamics. For this reason, survival analysis will be used in the following section, where the survival time of firms until bankruptcy will be modelled. Moreover, as this is a specific event related to the borrower's credit quality, a sub-goal of the following chapter will be to understand and link the relationship between corporate survival probability and credit rating.

Chapter 5

Relationship Between Rating and Corporate Bankruptcy Rates

This chapter provides an alternative view on measuring and predicting firm-based credit risk. While we estimated rating models in the previous section, the focus is on bankruptcy models in the following application. Bankruptcy, as a terminal state of the company, and rating are related because the worst rating assessment is typically issued to insolvent companies in financial trouble, often close to bankruptcy. Thus, we also examine and model corporate bankruptcy in this research to extend the previous findings of the main predictors of rating assessment.

The knowledge and understanding of rating and bankruptcy factors are essential for credit risk management. The evidence of corporate survival and non-default rates helps investors and lenders assess the credit quality of borrowers and predict potential problems of default, insolvency or even corporate bankruptcy. For example, the analysis of time to default, which CRAs conduct, is based on the cumulative distribution of defaulters by the time of default and survival rates. According to the historical default rates published by CRAs, there is a clear association between rating and default rates. Thus, assuming this relationship, we can estimate the survival rates of a certain sample of companies and compare them with historical non-default rates published by CRAs.

The main goal of this chapter is to determine whether there is a measurable relationship between estimated bankruptcy rates and default rates published by rating agencies using empirical data from Czech companies. In a positive case, this relationship can be described in a certain way. Then, a procedure can be suggested in which the detected bankruptcy rates could be used for a rating assessment corresponding to the rating agencies' assessment. This procedure has meaning and main application, especially in cases where we have data on corporate bankruptcies. We can process them statistically, and our goal is to translate them in a certain way into the “language” of rating agencies.

As mentioned in Chapter 3.1, there is a vast literature on predicting corporate bankruptcy using various techniques. However, little attention is paid to estimating corporate bankruptcy using survival analysis methods compared to discriminant, logistic, neural network methods or classification trees. Therefore, the partial aim of this study is to analyse survivor data of Czech companies and assess the impact of various factors on corporate survival.

During the observed time interval, our analysis considers bankruptcy the failure event. Thus, the time between the start of the business and bankruptcy is used to estimate survival and hazard functions. Eventually, the estimated survival rates can be compared with CRAs' historical default rates and corresponding rating assessments. Hence, we consider three credit risk measures in the following study: bankruptcy rates, default rates, and rating.

The structure is as follows. Firstly, the association between rating and corporate defaults observed and published by rating agencies is studied. Next, the Kaplan-Meier method is used to assess the effect of selected characteristics of Czech companies on the survival probability. Then, the relationship between estimated cumulative bankruptcy rates and published default rates is explored. Finally, based on the main findings, a procedure is proposed for converting the bankruptcy rates into rating assessments.

5.1 Association Between Rating and Corporate Defaults

This chapter provides an overview of the relationship between rating and default rates based on historical data from rating agencies. Next, we describe the approach rating agencies use to calculate cumulative default rates.

5.1.1 CRA Annual Default Rates

The purpose of this section is to explore the relationship between credit rating grades and corporate default rates. According to the historical occurrence of defaults within rating grades observed and published by CRAs, there is an evident correlation between the initial rating of a firm and its time to default. Typically, the historical number of defaults within an investment grade is substantially lower when compared to a speculative grade (Table 5-1).

The highest number of defaults, or the highest default rates, occurred during the financial crisis of 2008 – 2009, even within the investment grade. During the last 30 years, from 1990 to 2020, the investment grade achieved the highest default rate, 0.42%, during the two financial crises in 2002 and 2008 (S&P, 2021).

Table 5–1 Corporate default summary, 2005 – 2020

Year	Total defaults	Inv.-grade defaults	Spec.-grade defaults	Default rate (%)	Inv.-grade default rate (%)	Spec.-grade default rate (%)
2005	40	1	31	0.60	0.03	1.51
2006	30	0	26	0.48	0.00	1.19
2007	24	0	21	0.37	0.00	0.91
2008	127	14	89	1.80	0.42	3.71
2009	268	11	224	4.18	0.33	9.95
2010	83	0	64	1.21	0.00	3.02
2011	53	1	44	0.80	0.03	1.85
2012	83	0	66	1.14	0.00	2.59
2013	81	0	64	1.06	0.00	2.31
2014	60	0	45	0.69	0.00	1.44
2015	113	0	94	1.36	0.00	2.78
2016	163	1	143	2.09	0.03	4.24
2017	95	0	83	1.21	0.00	2.47
2018	82	0	72	1.03	0.00	2.10
2019	118	2	92	1.30	0.06	2.54
2020	226	0	198	2.74	0.00	5.50

Source: S&P (2021)

CRAs also calculate and publish annual default rates of rating grades, showing the historical trend of corporate defaults within rating categories (Table 5-2). Most rated defaulters come from the lowest categories, B and CCC/C, while the default rates of the AAA category are zero.

Table 5–2 Global annual default rates by rating category (%), 2005 – 2020

Year	AAA	AA	A	BBB	BB	B	CCC/C
2005	0.00	0.00	0.00	0.07	0.31	1.74	9.09
2006	0.00	0.00	0.00	0.00	0.30	0.82	13.33
2007	0.00	0.00	0.00	0.00	0.20	0.25	15.24
2008	0.00	0.38	0.39	0.49	0.81	4.08	27.27
2009	0.00	0.00	0.22	0.55	0.75	10.91	49.46
2010	0.00	0.00	0.00	0.00	0.58	0.85	22.73
2011	0.00	0.00	0.00	0.07	0.00	1.66	16.42
2012	0.00	0.00	0.00	0.00	0.30	1.56	27.33
2013	0.00	0.00	0.00	0.00	0.10	1.63	24.34
2014	0.00	0.00	0.00	0.00	0.00	0.77	17.03
2015	0.00	0.00	0.00	0.00	0.16	2.39	25.73
2016	0.00	0.00	0.00	0.06	0.47	3.76	33.17
2017	0.00	0.00	0.00	0.00	0.08	1.00	26.56
2018	0.00	0.00	0.00	0.00	0.00	0.99	27.18
2019	0.00	0.00	0.00	0.11	0.00	1.49	29.76
2020	0.00	0.00	0.00	0.00	0.93	3.52	47.48

Source: S&P (2021)

The historical trend of the default rates is relatively stable and shows a clear association between rating grade and the general default rate or credit risk. There is a negative correlation between the initial rating of a firm and its time to default. For example, the average time to a default of entities initially rated A grade (the time between first rating and date of default) is 14.1 years. In comparison, the average time to default among entities originally B was 5.1 years based on 1981 – 2020 in the study by S&P Global Ratings (S&P, 2021). Next, the average time for AAA rating is 18 years, BBB is 9.2 years, and CCC/C is only 2.2 years.

In addition to calculating time to default, the cumulative distribution of defaulters by the time of default and survival, or non-default rates, can also be used to get more information on the rating dynamics and defaults. For example, the survival rate or the percentage of B corporate issuers still alive for one, three or five years was 97.6%, 93.4% and 89.6% for 2011 – 2015 (Table 5-3).

Table 5–3 Corporate defaults and survival rates (2011 – 2015)

Rating	**One-year pool (2015)**		**Three-year pool (2013 – 2015)**		**Five-year pool (2011 – 2015)**	
	Number of defaults	Non-default rate	Number of defaults	Non-default rate	Number of defaults	Non-default rate
AAA	0	100.0%	0	100.0%	0	100.0%
AA	0	100.0%	0	100.0%	0	100.0%
A	0	100.0%	0	100.0%	1	99.8%
BBB	0	100.0%	0	100.0%	0	100.0%
BB	2	99.8%	7	99.1%	22	97.1%
B	42	97.6%	93	93.4%	122	89.6%
CCC/C	38	73.8%	56	57.9%	47	59.8%

Source: Source: S&P (2015)

For comparison, Table 5-4 summarises default rates for 2016 – 2020, partly already reflecting the consequences of the COVID-19 pandemic.

Table 5–4 Corporate defaults and survival rates (2016 – 2020)

Rating	**One-year pool (2020)**		**Three-year pool (2018 – 2020)**		**Five-year pool (2016 – 2020)**	
	Number of defaults	Non-default rate	Number of defaults	Non-default rate	Number of defaults	Non-default rate
AAA	0	100.0%	0	100.0%	0	100.0%
AA	0	100.0%	0	100.0%	0	100.0%
A	0	100.0%	2	99.9%	0	100.0%
BBB	0	100.0%	2	99.9%	13	99.3%
BB	12	99.1%	17	98.7%	28	97.8%
B	73	96.5%	175	90.9%	283	85.0%
CCC/C	113	52.5%	103	47.2%	103	48.2%

Source: S&P (2021)

CRAs' default and non-default rates are typically based on issuer-level ratings and are calculated using a discrete-time hazard rate method of survival analysis. Given the historical track of defaults, we can estimate issuers' expected cumulative default probabilities based on the average cumulative default rates, which may be a useful benchmark for the expected likelihood of default for obligations (Moody's Investors Service, 2006). Rating agencies widely use survival analysis; however, there is still little attention paid to using survival analysis for corporate bankruptcy in academic research.

5.1.2 Methodology of CRA's Cumulative Default Rates

In our study, the bankruptcy rates of the observed firms are compared with the historical data of ratings and defaults tracked and published by CRAs. For example, S&P (2017a) calculates cumulative default rates based on each static pool's average annual marginal default rates for each time horizon. The steps are as follows:

- Calculation of annual marginal default rates,
- calculation of conditional default rates,
- calculation of cumulative default rates.

The cumulative default rates average the experience of all static pools by first calculating marginal default rates for each possible time horizon and each static pool, weight averaging the marginal default rates conditional on survival, and accumulating the average conditional marginal default rates. Then, the conditional default rates are calculated by dividing the number of issuers in a static pool that default at a specific horizon by the number of issuers that survived to that point. Finally, cumulative default rates are one minus the product of the proportion of survivors.

Table 5-5 summarises the weighted-average default rates for all investment and speculative-grade rated companies for 40 pools by S&P (2021). For example, the first-year default rate of a speculative grade was 3.71%, meaning that an average of 96.29% survived one year. Similarly, the second-year conditional marginal average was 3.61%, the third-year conditional marginal average was 3.23%, and so on. Thus, 96.39 % of those companies that did not default in the first year survived the second year, and 96.77% of those companies that did not default by the second year survived the third year. Multiplying 96.29% by 96.39%, we get a 92.81% survival rate to the end of the second year, which is a two-year average cumulative default rate of 7.19% (S&P, 2021). There is a clear difference between investment and speculative-grade default rates. For example, while a survival rate to the end of the tenth year is 98.12% for investment-grade ratings, it is 79.19% for a speculative grade.

Table 5–5 Cumulative corporate default rates (1981 – 2020)

	Investment-grade ratings		**Speculative ratings**	
Time horizon (years)	Marginal average	Cumulative average	Marginal average	Cumulative average
1	0.09	0.09	3.71	3.71
2	0.15	0.24	3.61	7.19
3	0.18	0.41	3.23	10.18
4	0.22	0.63	2.73	12.63
5	0.23	0.86	2.30	14.64
6	0.23	1.09	1.94	16.30
7	0.21	1.30	1.65	17.68
8	0.20	1.50	1.40	18.83
9	0.19	1.69	1.27	19.86
10	0.19	1.88	1.18	20.81
11	0.18	2.05	1.02	21.61
12	0.15	2.20	0.86	22.29
13	0.15	2.35	0.82	22.93
14	0.15	2.49	0.73	23.49
15	0.16	2.65	0.72	24.04

Source: S&P (2021)

5.2 Modelling of Corporate Survival Based on Kaplan-Meier Estimates

This study uses survival analysis methods to analyse survivor data of Czech companies and assess the impact of industry, business entity type, and size on corporate survival. The main contribution to the current research is the application of survival analysis on real data of Czech companies, examining selected factors on the probability of corporate survival and comparing estimated bankruptcy rates. This study builds on and expands on the author's previous work (Novotná, 2016; Novotná, 2017; Novotná, 2020). The research used a dataset of companies from different economic sectors from 1988 to 2015. In this study, we focus on three hypotheses:

(i) The first hypothesis is that cyclical sectors such as construction or transportation are riskier, and therefore, corporate bankruptcy rates are higher than in other sectors.

(ii) The second hypothesis is that joint-stock companies are less risky than other business entities, such as self-employers, foreign persons, or cooperatives. This statement is based on the fact that joint-stock companies typically allow vast capital mobilisation and their contribution to business expansion.

(iii) The third hypothesis is that smaller companies are riskier and have a higher probability of bankruptcy. This premise assumes that smaller companies are more vulnerable to changes in the economic environment, especially during periods of economic contraction.

The structure of this section is as follows. Firstly, we describe the used dataset, and next, the Kaplan-Meier method is used to assess the effect of selected corporate characteristics on the bankruptcy of companies from the Czech Republic.

5.2.1 Description of Data

The dataset[15] comprises 16,727 subjects, including 1,481 bankruptcies considered failure events. Each record in our data sample documents the time of a particular company; the original duration time is measured in days. The mean time to failure is 8568.4 days, 95% [8536.6, 8600.2]. Transferred to years, the average time to bankruptcy is 23.5 years. In our study, a censored observation is when companies did not go bankrupt during the observed period or were no longer registered in the dataset. Otherwise, the observation is uncensored. Each record in our data sample documents the time span of a particular company, and the duration is measured in days.

The companies are grouped according to the following characteristics (Table 5-6): Industry classification, legal form and business size.

Table 5–6 Description of groups

Group	Sector	Legal form	Size
1	Services	Joint-stock comp.	Micro
2	Industrials	Cooperative	Small
3	Agriculture	Limited-liability comp.	Medium
4	Utilities	Other	Large

a) Industry

Companies from nine industries are considered in our dataset (Table 5-7), and the classification is based on MSCI (2019). Construction and information technology are the sectors that are represented the most, while the lowest number of firms comes from health care and water supply. Accordingly, the largest number of bankruptcies come from the construction sector, followed by transportation and accommodation. These basic descriptive statistics already suggest that construction might be the riskiest industry. On the contrary, health care and water supply seem less risky according to the relative number of failures. As shown in

[15] Data on Czech companies were obtained from the information database Magnusweb, supplemented by the justice.cz server. All companies registered in the commercial register from 1990 to 2005 and operating in nine selected industries are included in the analysis. The start date is considered the company's establishment date according to the commercial register. The end date represents the company's bankruptcy day if bankruptcy has occurred. If the bankruptcy did not happen in the monitored period, the end date is 31 March 2015, considered the observation's end.

the table, some industry categories were merged, and we used four sectors in the following analysis: Services, industrials, agriculture, and utilities.

Table 5–7 Industry groups

Industry	Cat.	Sector	Group	No. of subj.	No. of subj. (%)	No. of events	No. of events (%)
Hotels, restaurants, leisure	**1**	S	**1**	1905	11.4 %	193	13.0 %
Diversified consumer services	**2**	S	**1**	2060	12.3 %	126	8.5 %
Agriculture	**3**	A	**3**	2238	13.4 %	102	6.9 %
Construction	**4**	I	**2**	4550	27.2 %	657	44.4 %
Entertainment	**5**	S	**1**	789	4.7 %	40	2.7 %
Health care equipment and services	**6**	S	**1**	644	3.9 %	12	0.8 %
IT services, software	**7**	S	**1**	2634	15.7 %	95	6.4 %
Transportation	**8**	I	**2**	1228	7.3 %	223	15.1 %
Water supply	**9**	U	**4**	678	4.1 %	33	2.2 %
Total				***16726***	*100 %*	***1481***	*100 %*

A-Agriculture, I-Industrials, S-Services, U-Utilities

b) Legal form

The legal form classification of commercial companies and cooperatives in the Czech Republic depends on the minimum number of founders, minimum registered capital or financial liability of members. The most common form of entrepreneurship in the Czech Republic is limited-liability companies (Businessinfo, 2019), which corresponds with the structure of our data sample, followed by joint-stock companies and cooperatives (Table 5-8). Finally, the least represented categories are unlimited and limited partnerships, European and foreign companies and foundations, referred to as Other in our analysis.

Table 5–8 Legal form groups

Legal form	Group	No. of subj.	No. of subj. (%)	No. of events	No. of events (%)
Joint-stock company	**1**	2265	13.5 %	216	14.6 %
Cooperatives	**2**	493	2.9 %	31	2.1 %
Limited-liability	**3**	13789	82.4 %	1230	83.1 %
Other	**4**	179	1.1 %	4	0.3 %
Total		***16726***	*100 %*	***1481***	*100 %*

c) Business size

Enterprises are classified according to size based on the OECD approach (OECD, 2019). Overall, small and medium-sized enterprises employ less than 250 employees, and in the opposite case, companies are assumed to be large. Most companies in our data sample are micro-companies with less than ten employees, followed by small and medium-sized enterprises. Since the indicator of business size is not known for 2,839 companies, the total number of firms and events is lower when compared to the full sample size (Table 5-9).

Table 5–9 Business size groups

Business class	**Group**	**No. of subj.**	**No. of subj. (%)**	**No. of events**	**No. of events (%)**
Micro (0-9)	**1**	7408	53.3 %	511	45.3 %
Small (10-49)	**2**	4453	32.1 %	430	38.2 %
Medium (50-249)	**3**	1716	12.4 %	153	13.6 %
Large (>250)	**4**	311	2.2 %	33	2.9 %
Total		***13888***	*100 %*	***1127***	*100 %*

5.2.2 K-M Estimation of Bankruptcy Rates

We use Kaplan-Meier (K-M) estimates based on equation (3.62). The overall K-M estimates of our data can be summarized using the table that shows the number of subjects at risk (eligible to fail), the number of fails, the number of lost observations, the estimate of survival function, standard error and confidence interval. The K-M estimates are displayed for ten equally spaced time intervals (Table 5-10). The equal period is 1152 days (3.16 years), and the survival function is calculated using all data in the dataset using the formula (3.62). The estimated survival and cumulative hazard functions based on the whole data sample are shown in Figure 5-1.

Table 5–10 K-M estimated at equally spaced time intervals

Time	**No. at risk**	**No. failed**	**Survival function**	**Standard error**	**[95% Conf. Int.]**	
4	16727	1	0.9999	0.0001	0.9996	1.0000
1156	16125	322	0.9806	0.0011	0.9784	0.9826
2308	15362	280	0.9633	0.0015	0.9603	0.9661
3460	13708	313	0.943	0.0018	0.9393	0.9465
4612	10586	278	0.9214	0.0022	0.917	0.9256
5764	7609	280	0.8946	0.0027	0.8893	0.8997
6916	4810	4	0.894	0.0027	0.8887	0.8992
8068	1892	3	0.8932	0.0027	0.8878	0.8984
9220	2	0	0.8932	0.0027	0.8878	0.8984

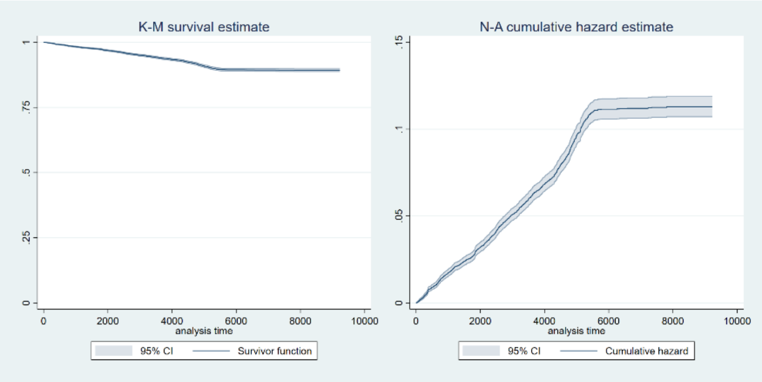


Figure 5–1 Survival and cumulative hazard functions

a) The Effect of Industry

Based on the mean estimated survival time (Table 5-11), industrial sectors such as transportation and construction have the lowest estimated survival time: 7809 and 8118 days. These sectors' highest probability of failure can be associated with their higher sensitivity to the business cycle. In contrast, IT services & software have the highest estimated survival time of 8951 days, followed by agriculture and consumer discretionary sectors such as entertainment and diversified consumer services.

Table 5–11 Estimated mean survival time by industry

Category	Industry	Mean	Standard error	Confidence interval (95%)	
1	Hotels & restaurants	8307.18	56.2484	8196.94	8417.43
2	Consumer services	8752.37	40.4705	8673.05	8831.69
3	Agriculture	8787.60	30.7435	8727.35	8847.86
4	Construction	8118.35	36.6590	8046.49	8190.20
5	Entertainment	8761.25	58.2390	8647.10	8875.40
6	Health care	8693.96	38.7573	8618.00	8769.92
7	IT services & software	8951.35	27.1386	8898.16	9004.54
8	Transportation	7808.62	76.8673	7657.96	7959.28
9	Water supply	8543.17	54.1554	8437.02	8649.31
Total		*8568.35*	*16.2208*	*8536.56*	*8600.15*

For further analysis, some industries are merged, and we analyse four sectors: (1) services, (2) industrials, (3) agriculture and (4) utility. The codes of these four sectors are denoted in Table 5-7. Among all sectors, the estimated survivorship of industrial companies is the lowest among all groups, followed by utility, services and agriculture (see Figure 5-2). The tests of equality of overall survival functions across groups based on the log-rank test (Chi2(3)=421.24), Wilcoxon (Chi2(3)=384.24) and Peto-Peto test (Chi2(3)=416.54) reject the hypothesis that the survival functions are the same.

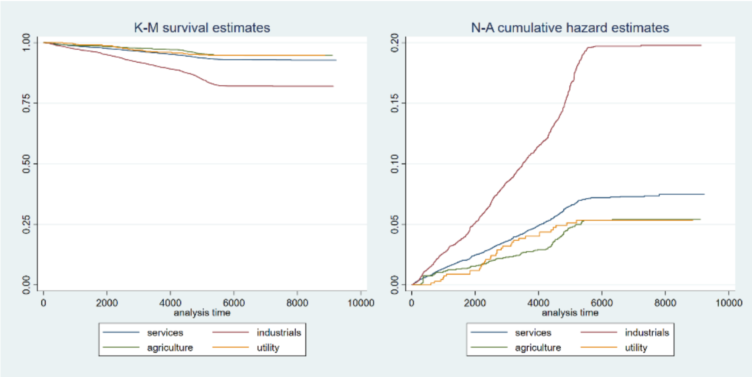


Figure 5–2 Survival and cumulative functions by sectors

b) The Effect of Legal Form

The mean estimated times are summarised in Table 5-12. Joint-stock and limited-liability companies have the lowest estimated survival time, followed by cooperatives and other legal forms.

Table 5–12 Estimated mean survival time by legal status

Group	Legal status	Mean	Standard error	Confidence interval (95%)	
1	Joint-stock	8517.66	40.5545	8438.17	8597.15
2	Cooperatives	8726.66	86.3793	8557.36	8895.96
3	Limited-liab.	8552.97	18.1951	8517.29	8588.62
4	Other	8996.49	65.7681	5567.59	9125.39
Total		*8568.40*	*16.2197*	*8536.61*	*8600.19*

The survival and cumulative functions of each legal form are depicted in Figure 5-3. The tests of equality of overall survival functions across groups based on the log-rank test (Chi2(3)=14.31), Wilcoxon (Chi2(3)=14.54) and Peto-Peto test (Chi2(3)=15.96) reject the hypothesis that the survival functions are the same.

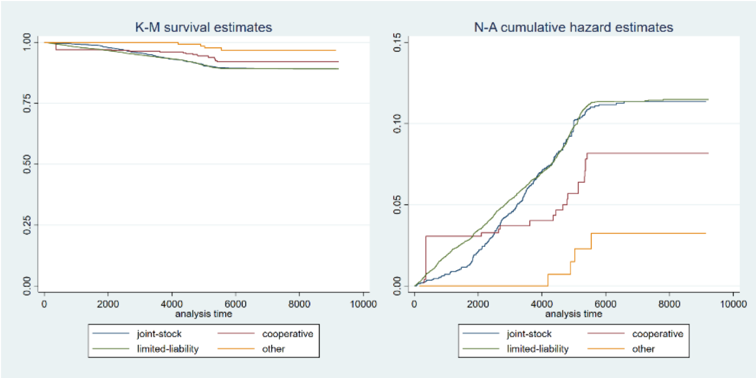


Figure 5–3 Survival and cumulative hazard functions by legal form

c) The Effect of Business Size

While the lowest estimated mean survival time is associated with small companies, the mean survival times of other categories are similar (Table 5-13). Micro companies have the highest estimated mean survival time, followed by medium and large companies. The visual results suggest that small companies have the lowest probability of survival (Figure 5-4).

Table 5–13 Estimated mean survival time by size

Group	Business category	Mean	Standard error	Confidence interval (95%)	
1	Micro	8595.66	21.6411	8553.25	8638.08
2	Small	8488.22	29.7732	8429.86	8546.57
3	Medium	8568.99	44.0947	8482.56	8655.41
4	Large	8556.97	109.99	8341.39	8772.54
Total		*8644.55*	*16.4998*	*8612.21*	*8676.89*

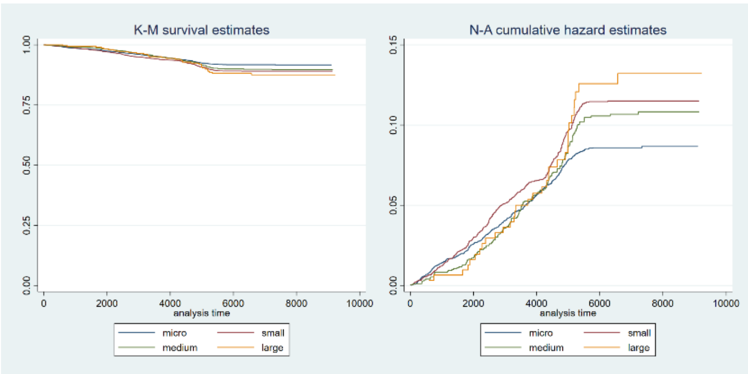


Figure 5–4 Survival and cumulative hazard functions by size

The tests of equality of overall survival functions across groups based on the log-rank test (Chi2(3)=16.58), Wilcoxon (Chi2(3)=12.54) and Peto-Peto test (Chi2(3)=15.86) reject the hypothesis that the survival functions are the same.

5.3 The Relationship Between Bankruptcy Rates and Rating Assessment

Based on the results of our prior analysis, the estimated bankruptcy rates are compared with CRAs' historical default rates to assess the average credit quality of the observed firms. Since bankruptcy can be considered a legal procedure for liquidating a business that cannot fully pay its debts, a strong correlation between bankruptcy and default rates is assumed. Thus, we can compare default rates with bankruptcy rates without much impact on the overall findings and their interpretation.

The aim of this section is to examine whether there is an association between the estimated corporate bankruptcy rates and rating cumulative default rates. First, the estimated survival functions are used to determine cumulative bankruptcy rates at the end of a particular year. Next, they are compared with CRAs' historical cumulative default rates and, eventually, corresponding rating assessments. Finally, through peer comparison, we propose a way that bankruptcy rates can be translated into the rating. Hence, the main purpose of this study is to link the bankruptcy rates and rating assessment.

5.3.1 Association Between Bankruptcy Rates and Rating

To assess the average rating quality of companies in our data sample, we examine the association between corporate bankruptcy rates and S&P rating grades. Firstly, cumulative bankruptcy rates (CBR) of the sample are calculated using the full data set according to the time horizon. Then, the bankruptcy rates are compared with long-term average global cumulative default rates of S&P rating (CDR, 1981-2020); see Table 5-14.

The association between corporate bankruptcy rates and S&P cumulative default rates is graphically presented in Figure 5-5. It can be seen from this figure that there are substantial differences between default rates of rating groups. For example, cumulative default rate CCC/C, B, and BB curves lie above all other curves, reflecting the greatest credit risk. It is also evident that the bankruptcy rates lie between BBB and BB's rating categories, approaching BB with a longer time horizon. Hence, according to the rating definition, our companies' investment quality changes with time.

Table 5–14 Cumulative bankruptcy rates and average cumulative default rates

Time horiz.	CBR	CDR AAA	CDR AA	CDR A	CDR BBB	CDR BB	CDR B	CDR CCC/C
1	0.0072	0.0000	0.0002	0.0005	0.0016	0.0063	0.0334	0.2830
2	0.0129	0.0003	0.0006	0.0013	0.0043	0.0193	0.0780	0.3833
3	0.0184	0.0013	0.0011	0.0022	0.0075	0.0346	0.1175	0.4342
4	0.0234	0.0024	0.0021	0.0033	0.0114	0.0499	0.1489	0.4636
5	0.0281	0.0034	0.0030	0.0046	0.0154	0.0643	0.1735	0.4858
6	0.0350	0.0045	0.0041	0.0060	0.0194	0.0775	0.1936	0.4961
7	0.0428	0.0051	0.0049	0.0076	0.0227	0.0889	0.2099	0.5075
8	0.0494	0.0059	0.0056	0.0900	0.0261	0.0990	0.2231	0.5149
9	0.0556	0.0064	0.0063	0.0105	0.0293	0.1082	0.2350	0.5216
10	0.0629	0.0070	0.0070	0.0120	0.0324	0.1164	0.2462	0.5276
11	0.0685	0.0072	0.0076	0.0134	0.0355	0.1233	0.2558	0.5321
12	0.0753	0.0075	0.0082	0.0146	0.0380	0.1299	0.2631	0.5368
13	0.0856	0.0078	0.0088	0.0159	0.0403	0.1359	0.2699	0.5423
14	0.0985	0.0084	0.0093	0.0171	0.0428	0.1409	0.2763	0.5469
15	0.1095	0.0090	0.0099	0.0184	0.0454	0.1465	0.2824	0.5476

Source: S&P (2021), author

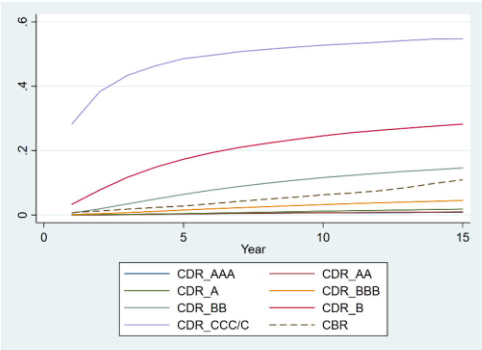


Figure 5–5 Cumulative bankruptcy and default rates

However, it is important to compare cumulative bankruptcy rates with default rates. Unlike a default event that refers to the debtor's incapacity or refusal to meet their debt obligations when due, bankruptcy is the legal status of an entity that cannot repay debts to creditors. Based on their definitions, we assume default rates to be generally higher than bankruptcy rates. Nevertheless, this comparison can give us an interesting look at the association between bankruptcy and rating default rates and the overall credit risk of corporates in the data sample.

In addition to the average values, we summarize the bankruptcy rates by sector, legal form and size (Appendix 5). The rates correspond to the main findings in Chapter 5.2.2, specifically Figure 5-2, Figure 5-3 and Figure 5-4, showing survival and hazard functions.

Next, we examine credit default rates (CDRs) and their relationship with estimated corporate bankruptcy rates. Table 5-15 summarizes the average credit default rates for the time horizon of ten years based on the statistics by S&P (2021). We can see the global CDRs and the rates from Europe and emerging countries. In addition to the average rates by rating groups, note the overall rates for investment and speculative grades.

Table 5–15 Average CDRs (t = 10 years)

	Global CDR	Europe CDR	Emerging CDR
AAA	0.0036	0.0000	0.0000
AA	0.0035	0.0017	0.0000
A	0.0057	0.0025	0.0003
BBB	0.0170	0.0070	0.0146
BB	0.0664	0.0363	0.0452
B	0.1659	0.1281	0.1191
CCC/C	0.4618	0.4599	0.2800
Investment	0.0097	0.0036	0.0109
Speculative	0.1418	0.1065	0.0888

Source: S&P (2021)

Based on Table 5-14, we calculate the average cumulative bankruptcy rate for a time horizon of 1–10 years ($CBR = 0.0336$). Then, we find the differences between the average CDRs (Table 5-15) and the calculated average CBR. Finally, we can see the differences between CDRs and CBR, referred to as average spreads (AS), in Table 5-16.

Table 5–16 Average spreads CDR-CBR (t = 10 years)

	Global AS	Europe AS	Emerging AS
AAA	-0.0299	-0.0336	-0.0336
AA	-0.0301	-0.0319	-0.0336
A	-0.0279	-0.0310	-0.0333
BBB	-0.0166	-0.0265	-0.0189
BB	0.0329	0.0027	0.0117
B	0.1323	0.0945	0.0855
CCC/C	0.4282	0.4263	0.2464
Investment	-0.0239	-0.0300	-0.0227
Speculative	0.1083	0.0729	0.0552

The relationship between the calculated average spreads and ratings is graphically presented in Figure 5-6. The higher the spread, the lower the average rating assessment, coded from 1 (AAA) to 7 (CCC/C). Hence, based on the results, bankruptcy rates seem to be good indicators of rating quality.

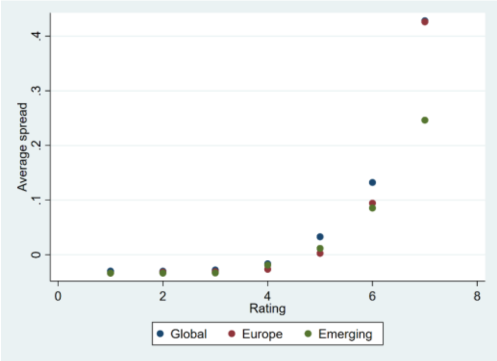


Figure 5–6 Rating and average spreads

In the next section, a method is proposed for how the calculated average spreads can be used to assign the probable corporate rating. As shown above, the CBRs are based on the N-A estimates of cumulative hazard rates in this study. Therefore, for rating estimation, we need average spreads by rating grades. Furthermore, since we model the data of Czech companies, the spreads are based on European CDRs (Table 5-17).

Table 5–17 Spreads by rating and time horizon

Time horiz.	AAA	AA	A	BBB	BB	B	CCC/C	Inv.	Spec.
1	-0.0072	-0.0072	-0.0069	-0.0066	-0.0036	0.0150	0.2792	-0.0069	0.0216
2	-0.0129	-0.0127	-0.0122	-0.0112	-0.0016	0.0442	0.3743	-0.0120	0.0430
3	-0.0184	-0.0179	-0.0174	-0.0154	0.0005	0.0705	0.4091	-0.0169	0.0594
4	-0.0234	-0.0224	-0.0219	-0.0191	0.0034	0.0910	0.4429	-0.0212	0.0730
5	-0.0281	-0.0265	-0.0258	-0.0227	0.0078	0.1075	0.4614	-0.0251	0.0842
6	-0.0350	-0.0329	-0.0321	-0.0272	0.0080	0.1170	0.4635	-0.0309	0.0889
7	-0.0428	-0.0404	-0.0390	-0.0332	0.0070	0.1222	0.4611	-0.0378	0.0906
8	-0.0494	-0.0467	-0.0453	-0.0383	0.0048	0.1242	0.4612	-0.0438	0.0904
9	-0.0556	-0.0526	-0.0513	-0.0429	0.0022	0.1265	0.4550	-0.0494	0.0897
10	-0.0629	-0.0599	-0.0585	-0.0487	-0.0011	0.1270	0.4557	-0.0562	0.0881

Source: S&P (2021); author

5.3.2 Transmission of Bankruptcy Rates to Rating Assessment

In the previous section, we observed the relationship between the estimated bankruptcy rates and the rating using the spreads between the average bankruptcy rates and default rates for different time horizons. In this part, we propose a procedure for converting the bankruptcy rates to ratings using the calculated average spreads.

The procedure steps are as follows:

1. Estimate a cumulative bankruptcy rate (*ECBR*), for example, for the particular combination of variables (based on the table in Appendix 5).
2. Calculate the estimated spread between *CDR* and *ECBR* for each rating category (based on Table 5-14) using the equation (5.1),

$$ES = CDR - ECBR\,. \tag{5.1}$$

3. Compare *ES* with the average spread *AS* summarized in Table 5-16 . Then, choose the rating category with the lowest absolute value of deviance based on equation (5.2).

$$D = \left| ES - AS \right|\,. \tag{5.2}$$

Although this procedure is based on the European credit default rates published by the S&P agency, by analogy, it can be used for different geographical locations, rating agencies and time horizons. An application example of this procedure is shown below in the text.

Assuming that bankruptcy rates depend only on the sector, legal form and corporate size, we can estimate a rating based on the following procedure. For example, the above technique will be used to determine the probable rating for the following combinations of corporate characteristics:

a) Industrials, limited-liability, micro (i = 1),
b) services, joint-stock, large (i = 2),
c) agriculture, cooperative, medium (i = 3).

We assume time horizons $t_1 = 5$ and $t_2 = 10$ years to assess the effect of time on the development of rating quality.

Firstly, we find the estimated cumulative bankruptcy rates (*ECBR*s) as simple averages of 5-year and 10-year *CBR*s. Since our cumulative hazard rate estimates are based on a nonparametric analysis, we use the same weight for each categorical variable. Then, for ith combination of variables (i = 1, 2, 3) and time horizon t (t = 5, 10), the estimated cumulative bankruptcy rate is found as:

$$ECBR_{i,t} = \frac{1}{3}(CBR\_sector + CBR\_legal + CBR\_size). \tag{5.3}$$

For example, using the equation (5.3), we get the following *ECBR*s:

$$ECBR_{1,t=5} = \frac{1}{3}\left(0.045 + 0.0304 + 0.0223\right) = 0.0326,$$

$$ECBR_{2,t=5} = \frac{1}{3}\left(0.0216 + 0.0161 + 0.0097\right) = 0.0158,$$

$$ECBR_{3,t=5} = \frac{1}{3}\left(0.0145 + 0.0306 + 0.159\right) = 0.0203.$$

$$ECBR_{1,t=10} = \frac{1}{3}(0.105 + 0.0647 + 0.0503) = 0.0733,$$

$$ECBR_{2,t=10} = \frac{1}{3}(0.0448 + 0.0618 + 0.05) = 0.0522,$$

$$ECBR_{3,t=10} = \frac{1}{3}(0.0275 + 0.0402 + 0.0526) = 0.0401.$$

The results suggest differences between the estimated cumulative bankruptcy rates of two time periods. For example, assuming the time horizon of 5 years, the greatest cumulative hazard rate is associated with a combination of $i = 1$ (industrials, limited-liability, micro), followed by $i = 3$ (agriculture, cooperative, medium) and $i = 2$ (services, joint-stock, large). In the 10-year time horizon, the greatest cumulative hazard rate is again associated with the scenario $i = 1$. It is followed by $i = 2$, suggesting that the cumulative hazard rates vary with time for our combinations of variables.

Next, we calculate *ES* and *D* using the equations (5.1) and (5.2) and find the corresponding rating categories with the minimum absolute values of *D*. The results are summarized in Table 5-18. Thus, assuming the time horizon of 5 years (since the company's founding), we assign the middle rating BBB to the companies with the combination of variables $i = 2$ and $i = 3$. In the longer time horizon, the assigned rating is the same for $i = 3$; however, it is estimated to be BB for $i = 2$. The first corporate characteristics ($i = 1$) indicate a BB rating category regardless of the time horizon. Nevertheless, the credit rating seems to have worsened for a long time based on the suggested speculative rating grade.

Table 5–18 Rating estimation (K-M model)

	AAA	**AA**	**A**	**BBB**	**BB**	**B**	**CCC/C**
AS	-0.0336	-0.0319	-0.0310	-0.0265	0.0027	0.0945	0.4263
$ES_{1,t=5}$	-0.0326	-0.0310	-0.0303	-0.0272	0.0033	0.1030	0.4569
$ES_{1,t=10}$	-0.0733	-0.0703	-0.0689	-0.0591	-0.0115	0.1166	0.4453
$ES_{2,t=5}$	-0.0158	-0.0142	-0.0135	-0.0104	0.0201	0.1198	0.4737
$ES_{2,t=10}$	-0.0522	-0.0492	-0.0478	-0.0380	0.0096	0.1377	0.4664
$ES_{3,t=5}$	-0.0203	-0.0187	-0.0180	-0.0149	0.0156	0.1153	0.4692
$ES_{3,t=10}$	-0.0401	-0.0371	-0.0357	-0.0259	0.0217	0.1498	0.4785
$D_{1,t=5}$	0.0010	0.0010	0.0008	0.00064	0.00059	0.0085	0.0306
$D_{1,t=10}$	0.0398	0.0384	0.0379	0.0326	0.0143	0.0221	0.0189
$D_{2,t=5}$	0.0178	0.0177	0.0175	0.0161	0.0174	0.0253	0.0474
$D_{2,t=10}$	0.0186	0.0173	0.0168	0.0115	0.0069	0.0432	0.0401
$D_{3,t=5}$	0.0132	0.0132	0.0130	0.0116	0.0128	0.0208	0.0428
$D_{3,t=10}$	0.0065	0.0052	0.0047	0.0006	0.0190	0.0553	0.0522

The overall results suggest that the first combination of variables (industrials, limited-liability, micro) is the riskiest. The second case (services, joint-stock, large) is associated with the average risk, potentially worsening with a longer time horizon. Finally, based on the estimated rating, we found the combination with the lowest and most stable level of credit risk (agriculture, cooperative, medium). The main findings show that based on estimated bankruptcy rates, we can estimate the future rating development according to the time horizon.

5.4 Chapter Summary

This chapter focused on understanding the relationship between bankruptcy rates, default rates published by CRAs, and credit ratings. As was shown in the introductory part of this chapter, methods based on survivorship analysis are used by rating agencies to determine the default rates of rated subjects, both according to rating categories and to the considered time horizon.

This section used the survival method to assess the influence of selected corporate characteristics on the survival probability of Czech companies. For this purpose, the basic procedure used was the Kaplan-Meier method. It is a non-parametric analysis with which we can discover the basic relationships and understand the data survivorship. We focused on assessing the influence of industry, legal form and size on survival time. The results indicate that all used factors are related to the probability of corporate survival. Two hypotheses were confirmed, suggesting that small, industrial firms are riskier compared to the other considered sectors and sizes. On the other hand, our findings suggest that joint-stock companies are riskier compared to other legal forms, unlike the assumption. The overall results are summarized in Table 5-19. For example, a small, industrial joint-stock company can be considered the riskiest compared to other cases.

Table 5–19 Summary of results

Corporate characteristics	Category	Mean survival time
Industry	Services	•••
	Industrials	•
	Agriculture	••••
	Utility	••
Legal form	Joint-stock	•
	Cooperatives	•••
	Limited-liability	••
	Other	••••
Size	Micro	••••
	Small	•
	Medium	•••
	Large	••

• (lowest) •••• (greatest)

Using the procedures described above in Chapter 5.3, we derived survival and cumulative hazard functions, which were subsequently used to estimate the rating. They were determined using the proposed approach based on the average spread between cumulative bankruptcy and default rates. Based on the average spread, it was confirmed that the higher the spread, the lower the rating. Thus, the proposed procedure was built on this finding. The relevant rating category was determined using the absolute deviation between the estimated and average spread. This procedure was subsequently used to determine the rating of three hypothetical companies with different combinations of the considered characteristics. Even though this is a greatly simplified approach to determining the rating based on only three categorical variables, this method clearly shows that the risk associated with the probability of survival can be better understood with the help of the variables used.

The calculation above considered only three categories of corporate characteristics, ignoring the potential effect of other factors. Although we cannot accurately determine the rating based on the sector, legal form and corporate size, converting the cumulative bankruptcy rate into a rating will be further examined in the following chapters. The procedure will be additionally used in the next section based on survival and cumulative hazard functions estimated by other survival analysis methods. The Cox model will be used first, and then the Weibull model. The individual steps are analogous to the last part. First, the survival and cumulative hazard functions will be estimated. Then, the cumulative bankruptcy rates for the selected time horizons will be determined, and the proposed procedure will be used to determine the rating.

Chapter 6

Survival Models with Categorical Variables

Based on Chapter 5, there is evidence that the probability of survival depends on corporate characteristics such as industry, size and legal form. It was also shown that estimated cumulative bankruptcy rates are associated with credit rating default rates and can be used to estimate the rating assessment. Therefore, we use the Cox proportional hazards, the Weibull and flexible survival analysis models in this chapter to better understand these relationships.

6.1 Application of the Cox Proportional Hazards Model

The models in this section are estimated based on the dataset described in Chapter 5.2.1, using randomly selected 12547 observations (including 1125 failures). The Cox proportional hazards model is used to estimate hazard ratios, assuming that it acts multiplicatively. The estimated coefficients and hazard ratios are summarized in Table 6-1. For each category, we get coefficients of the baseline dummy variable (1) with other dummy variables (2, 3, 4). If the coefficient is negative, there is an inverse relationship between variables and vice versa. For example, the hazard ratio of agriculture related to services is $e^{-0.58} = 0.56$, suggesting that the hazard of agriculture is decreased by 44% when associated with the hazard of services. However, some estimated coefficients are statistically insignificant, meaning an inappropriate choice of variables in the model, especially the variable size.

Next, we assess the assumption of proportional-hazards, which is graphically presented in Figure 6-1, on the left-hand side. The lines do not seem parallel, suggesting that the PH assumption may be violated. Moreover, some lines cross in the data region, which indicates a potential problem with the PH assumption.

Table 6–1 Multivariate Cox model (CM1)

Variable	Coeff.	Hazard ratio	Standard error	Confidence interval (95%)	
Industry					
2	1.00*	2.72	0.078	0.848	1.152
3	-0.58*	0.56	0.167	-0.912	-0.256
4	-0.32	0.73	0.228	-0.765	0.128
Legal					
2	-0.91*	0.40	0.428	-1.744	-0.067
3	-0.25*	0.78	0.097	-0.435	-0.056
4	-1.52*	0.22	0.713	-2.920	-0.126
Size					
2	0.09	1.10	0.077	-0.057	0.245
3	0.03	1.03	0.111	-0.193	0.244
4	0.30	1.35	0.200	-0.093	0.692

* significant at 0.05, baseline dummy variable = 1

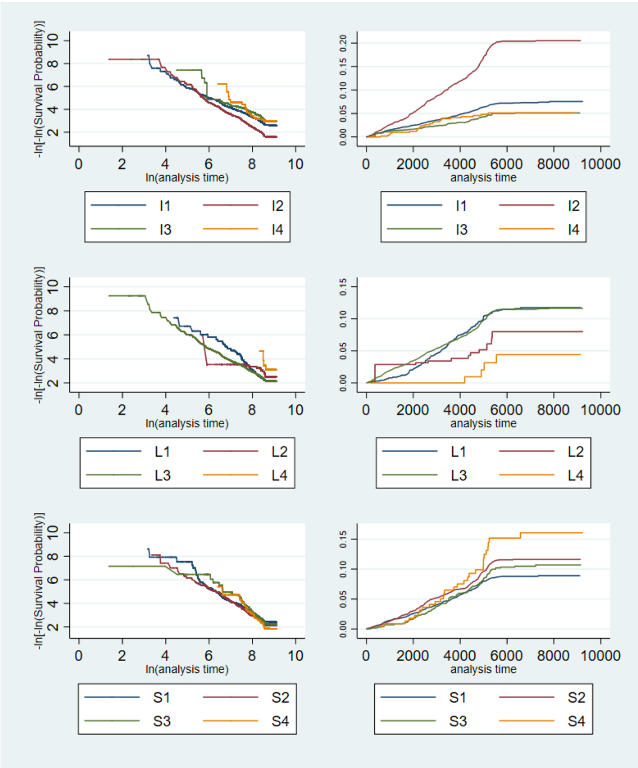


Figure 6–1 PH assessment and N-A cumulative hazard estimates

The results based on the graphical visualisations are supported by applying a Grambsch-Therneau test based on Cleves et al. (p. 209, 2010) of the scaled

Schoenfeld residuals from a Cox model on used variables (Table 6-2). The rejection of the null hypothesis indicates a deviation from the proportional-hazards assumption for industry2, size3, size4 and the global test.

Table 6–2 PH test

Variable	rho	Chi2	df	Prob>Chi2
Industry2	0.0959	7.58	1	0.0059
Industry3	0.0122	0.12	1	0.7248
Industry4	-0.0397	1.37	1	0.2421
Legal2	-0.0301	0.74	1	0.3894
Legal3	-0.0347	1.03	1	0.3091
Legal4	0.0405	1.41	1	0.2354
Size2	0.0247	0.52	1	0.4721
Size3	0.0768	4.94	1	0.0262
Size4	0.0710	4.24	1	0.0017
Global test		*26.45*	*9*	*0.0017*

The test results suggest that the estimated model is not well specified for the above reasons. Thus, we modify the used variables and assess a new model in the following section.

6.1.1 The Cox Models with Binary Categorical Variables

We use the same three categorical variables in this section; however, they are modified and defined as binary. This modification and division of all variables into two groups results from a previous examination and selection of the most suitable alternative. Finally, the new variables used in the analysis are defined as presented in Table 6-3.

Table 6–3 Binary categorical variables

Variable	Group	
	1	2
Industry_b	agriculture, utility	services, industrials
Legal_b	cooperatives	joint-stock, limited-liability, other
Size_b	micro	small, medium, large

Next, we estimate the Cox proportional hazards model with new binary variables (Table 6-4). The results show that compared to the variable baseline levels, the hazard decreases by 48% for services and industrials, increases by 100% for other legal forms than cooperatives, and increases by 37% for small, medium and large companies.

Table 6–4 Multivariate Cox model (CM2_1)

Variable	**Coeff.**	**Hazard ratio**	**Standard error**	**Confidence interval (95%)**	
industry_b	-0.65*	0.52	0.065	-0.782	-0.525
legal_b	0.69*	2.00	0.212	0.277	1.107
size_b	0.31*	1.37	0.063	0.187	0.435

* significant at 0.05, baseline dummy variable = 1

The results of PH tests are summarized in Table 6-5. The global test suggests that the PH assumption is not violated in this model. However, although the global test means no violation, the assumption is not fulfilled for the variable industry.

Table 6–5 PH test (CM2_1)

Variable	**rho**	**Chi2**	**df**	**Prob>Chi2**
industry_b	-0.0717	5.69	1	0.0171
legal_b	0.0292	0.96	1	0.3271
size_b	-0.0002	0.00	1	0.9936
Global test		*6.40*	*3*	*0.0939*

Since the variable indicating the industry violates the PH assumption, we will use it as a stratum in the following model. Thus, we can estimate the models for two subsets of population, agriculture & utility (1) and services & industry (2), see Table 6-6.

Table 6–6 Multivariate Cox model (CM2_2)

Variable	**Coeff.**	**Hazard ratio**	**Standard error**	**Confidence interval (95%)**	
legal_b	0.69*	2.00	0.218	0.276	1.106
size_b	0.31*	1.36	0.063	0.186	0.434

* significant at 0.05, baseline dummy variable = 1, stratified by industry_b

The PH assumption test suggests no violation of proportional hazards in this model (Table 6-7).

Table 6–7 PH test (CM2_2**)**

Variable	**rho**	**Chi2**	**df**	**Prob>Chi2**
Legal	0.02871	0.93	1	0.3361
Size	-0.00008	0.00	1	0.9979
Global test		*0.93*	*2*	*0.6289*

The estimated coefficients are equal across the used strata, however, with a baseline hazard unique to each value of industry indicator (1 = agriculture & utility, 2 = services & industry) in the new model CM2_2.

The baseline cumulative hazard and survival functions for two groups of industry variables used as strata are shown in Figure 6-2. There is a much higher slope for agriculture & utility (H01) than services & industry (H02) until $t = 5700$.

Moreover, while the slope is fairly constant for industry_b = 2, there is a higher constant slope after t = 4200 for industry_b = 1. These findings suggest an increasing hazard rate for companies in agriculture & utility after 4200 days (11.5 years).

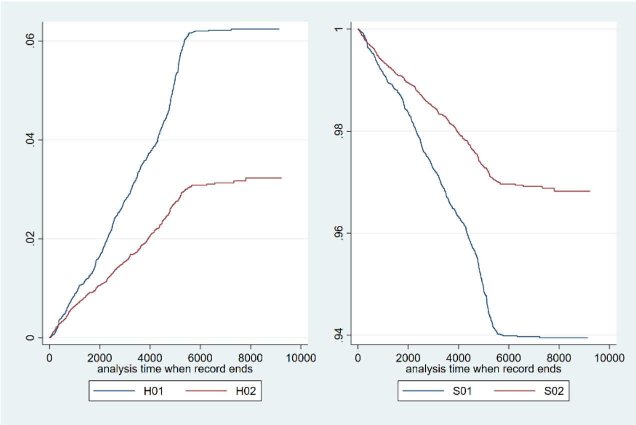


Figure 6–2 Estimated baseline cumulative hazard and survival functions

Based on the estimated coefficients, we can express survival and cumulative hazard functions for i = 1, 2, 3 combinations of variables as in Chapter 5.2. However, we adjust them for binary classification. The functions are formulated in Table 6-8. We denote S_{01} as the baseline survival function for strata1 (industry_b = agriculture & utility) and S_{02} for strata 2 (industry_b = services & industry). The baseline cumulative hazard functions are denoted analogically.

Table 6–8 Stratified Cox models

i	Variable Indicator	Survival function Cum. hazard function
1	Strata = 2 Limited-liab (2) Micro (1)	$S_1(t\|x) = S_{02}(t)^{\exp(0.69+0)} = S_{02}(t)^{1.99}$ $H_1(t\|x) = H_{02}(t) \cdot 1.99$
2	Strata = 2 Joint-stock (2) Large (2)	$S_2(t\|x) = S_{02}(t)^{\exp(0.69+0.31)} = S_{02}(t)^{2.72}$ $H_2(t\|x) = H_{02}(t) \cdot 2.72$
3	Strata = 1 Cooperatives (1) Medium (2)	$S_3(t\|x) = S_{01}(t)^{\exp(0+0.31)} = S_{01}(t)^{1.36}$ $H_3(t\|x) = H_{01}(t) \cdot 1.36$

The graphical visualisation of estimated functions is presented in Figure 6-3. We prove that corporate characteristics affect survival probability or the cumulative hazard rate. Combination 2 is the riskiest, followed by alternative 1 with average risk and 3 with the lowest risk.

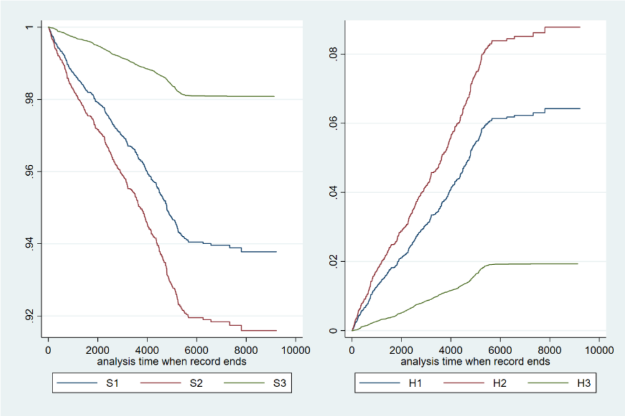


Figure 6–3 Estimated survival and cumulative hazard functions

The results support the main findings from Chapter 5.2. However, the advantage of the Cox proportional hazards model is that we can express survival and cumulative hazard functions as multiples of the baseline functions. So, we can measure the relationship between the functions and estimate the bankruptcy rates more accurately.

6.1.2 Estimation of Cumulative Bankruptcy Rates

Based on the estimation sample, the baseline survival functions $S_{01} = 0.9853, S_{02} = 0.9904$ and the cumulative hazard functions $H_{01} = 0.0148, H_{02} = 0.0097$ for $t = 5$. Assuming the time horizon $t = 10$, then $S_{01} = 0.9658, S_{02} = 0.9815$ and $H_{01} = 0.0348, H_{02} = 0.0187.$ These findings reflect the baseline curves as presented in Figure 6-2.

Estimated survival and cumulative hazard rates of specific combinations of variable indicators are calculated using the formulas in Table 6-8. Based on the baseline values, we get the following *ECBR*s:

$$ECBR_{1,t=5} = 0.0097 \cdot 1.99 = 0.0193, \qquad ECBR_{1,t=10} = 0.0187 \cdot 1.99 = 0.0372,$$
$$ECBR_{2,t=5} = 0.0097 \cdot 2.72 = 0.0264, \qquad ECBR_{2,t=10} = 0.0187 \cdot 2.72 = 0.0509,$$
$$ECBR_{3,t=5} = 0.0148 \cdot 1.36 = 0.0201. \qquad ECBR_{3,t=10} = 0.0348 \cdot 1.36 = 0.0473.$$

We used the same steps as in Chapter 5.3.2. Since the detailed procedure was described there, we provide only the main results and rating estimations in Table 6-9. In the five-year time horizon, all cases are assigned the same initial BBB rating, which changes in the longer time horizon. Unlike the K-M model, we see a rating upgrade for the first combination of variables $(i = 1)$.

Table 6–9 Rating estimation (Cox model)

	AAA	AA	A	BBB	BB	B	CCC/C
AS	-0.0336	-0.0319	-0.0310	-0.0265	0.0027	0.0945	0.4263
$ES_{1,t=5}$	-0.0193	-0.0177	-0.0170	-0.0139	0.0166	0.1163	0.4702
$ES_{1,t=10}$	-0.0373	-0.0357	-0.0350	-0.0319	-0.0014	0.0983	0.4522
$ES_{2,t=5}$	-0.0264	-0.0248	-0.0241	-0.0210	0.0095	0.1092	0.4631
$ES_{2,t=10}$	-0.0508	-0.0492	-0.0485	-0.0454	-0.0149	0.0848	0.4387
$ES_{3,t=5}$	-0.0202	-0.0186	-0.0179	-0.0148	0.0157	0.1154	0.4693
$ES_{3,t=10}$	-0.0474	-0.0458	-0.0451	-0.0420	-0.0115	0.0882	0.4421
$D_{1,t=5}$	0.0142	0.0142	0.0140	0.0126	0.0138	0.0218	0.0438
$D_{1,t=10}$	0.0037	0.0038	0.0039	0.0054	0.0041	0.0038	0.0259
$D_{2,t=5}$	0.0072	0.0072	0.0070	0.0056	0.0068	0.0147	0.0368
$D_{2,t=10}$	0.0173	0.0173	0.0175	0.0189	0.0177	0.0097	0.0123
$D_{3,t=5}$	0.0134	0.0133	0.0132	0.0118	0.0130	0.0209	0.0430
$D_{3,t=10}$	0.0139	0.0139	0.0141	0.0155	0.0143	0.0064	0.0157

The used model is a much-simplified model that contains only two broadly defined binary variables. Nevertheless, we demonstrated the use of the stratified Cox model and showed its application procedure to predict bankruptcy rates and, eventually, the rating. Furthermore, based on the model, we provided evidence that the variables used correspondingly affect the probability of corporate survival and cumulative hazard rates. However, this model has no practical use. For this reason, we will continue modelling our data using parametric survival analysis.

6.2 Parametric Models

In the previous section, we used the Cox proportional hazards model that assumes that the functions are multiplicatively shifted based on the baseline functions. As mentioned above, since the baseline hazard function is unspecified, the model is not fully parametric but semiparametric. Although we know the regression parameters, the outcome distribution is unknown.

To further explore the probability of corporate survival, we use parametric survival models in this section. Thus, we can estimate the parameters of the outcome distribution. This section aims to apply selected parametric models and find the most suitable model for our data. The focus will be on using the Weibull and some extended or flexible parametric models. Firstly, we apply the models without covariates to better understand our data and compare the survival and hazard functions based on different assumptions. Next, we consider the effect of the same categorical variables as in the previous section, 6.1. Finally, we compare the main findings with the Cox model.

6.2.1 Estimation of the Weibull and Extended Models

Firstly, we apply the Weibull and extended (flexible) parametric models without categorical variables to explore our data. Specifically, according to the number of interior knots, we estimate six models: the Weibull model without one interior knot and extended models with a different number of interior knots.

a) Models without variables

The baseline cumulative and survival functions of estimated models are shown in Figure 6-4. For comparison, we also plot N-A and K-M estimates. All models perform well up to approximately t = 700 days (2 years). Then, the Weibull model begins deviating from other models and tends to slightly underestimate the N-A cumulative hazard until t = 7600 days (21 years). Extended models perform similarly up to the time of approximately t = 1400 days (3.8 years). After that, the PH(6) model best fits our data, followed by other similarly performing models.

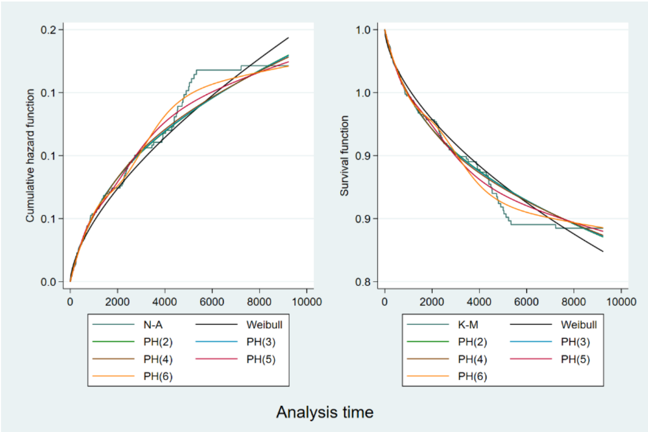


Figure 6–4 Cumulative hazard and survival functions

Overall, we cannot choose the best fitting model based on the graphical presentations of cumulative hazard and survival functions. Thus, we compare the models according to the information criteria. Table 6-10 shows the degree of freedom (d.f.), parameter estimates, standard error (SE) and goodness of fit (AIC and BIC) of six estimated models. All models are summarized in Appendix 6.

Table 6–10 Coefficient estimates (no variables)

Model	d.f.	$\hat{\beta}$ (_rcs1)	Standard error (SE)	AIC	BIC
Weibull	1	0.6304*	0.0546	958.5875	972.1075
PH(2)	2	1.1240*	0.2262	954.1481	974.4280
PH(3)	3	1.1195*	0.3298	956.1206	983.1605
PH(4)	4	1.1598*	0.4101	957.9758	991.7757
PH(5)	5	0.9584*	0.4391	957.2599	997.8198
PH(6)	6	1.0980*	0.4943	955.4142	1002.734

* significant at 0.05

The minimum AIC is associated with PH(2), followed by PH(6) and PH(3), with very similar AIC values. The models that minimise BIC are the Weibull model PH(1), PH(2) and PH(3). Based on the overall results, we suggest four models for further comparison: the Weibull model and the extended PH(2), PH(3) and PH(6) models in which the baseline cumulative hazards are spline functions with one, two and five interior knots.

The cumulative (left-hand side) and hazard (right-hand side) functions of well-performed models PH(1), PH(2), PH(3) and PH(6) are shown in Figure 6-5. While the Weibull model hazard function monotonically decreases with time (the shape is determined by parameter p = 0.59), the hazard functions of extended models fluctuate, especially the PH(6) model with five interior knots. For example, the hazard rates of this model are firstly above the Weibull model, up to t = 1200 days. Then, the hazard rates are lower when t = 1200 – 2100, higher when t = 2100 – 4400 days. Finally, the hazard function is below the Weibull beyond t = 4400 days. It is clear from the graph that extended models assume more flexible hazard functions than the basic model without interior knots. It can be seen as an advantage of these models, which can better fit the real data.

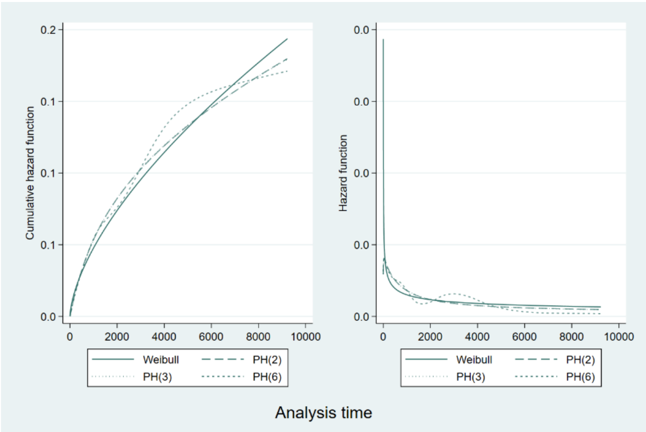


Figure 6–5 Hazard functions

b) Models with categorical variables

To explore the effect of variables on the probability of survival, we apply the Weibull and extended (flexible) parametric models using industry, legal and size indicators. Analogically to the previous section, we estimate six models according to the number of interior knots. The estimates _rcs1, AIC and BIC are presented in Table 6-11. All models are summarized in Appendix 7.

Table 6–11 Coefficients estimates (multiclass variables)

Model	d.f.	$\hat{\beta}$ (_rcs1)	Standard error (SE)	AIC	BIC
Weibull	1	0.6414^{*}	0.0687	675.2899	748.1626
PH(2)	2	1.0216^{*}	0.2728	674.7651	754.2626
PH(3)	3	0.8827^{*}	0.3628	676.4007	762.5230
PH(4)	4	0.9347^{*}	0.4269	677.8724	770.6195
PH(5)	5	0.6763^{**}	0.3993	673.9622	773.3341
PH(6)	6	0.7032^{**}	0.4268	674.0831	780.0798

*significant at 0.05, **significant at 0.10

While the minimum AICs are associated with PH(5), PH(6), and PH(2), the minimum BIC values are achieved by the Weibull model, PH(2) and PH(3). Therefore, we consider the Weibull model and PH(2) potentially appropriate models based on BIC values for our data. Although the estimated coefficients _rcs1 are statistically significant at 0.05, other estimated coefficients of most models are not statistically significant. Thus, we use the binary categorical variables defined in Table 6-3 in the following model (Table 6-12).

Table 6–12 Coefficients estimates (binary variables)

Model	d.f.	$\hat{\beta}$ (_rcs1)	Standard error (SE)	AIC	BIC
Weibull	1	0.9464^{*}	0.0267	9705.694	9752.893
PH(2)	2	1.3893^{*}	0.1332	9694.558	9751.196
PH(3)	3	0.8665^{*}	0.1490	9666.651	9732.728
PH(4)	4	1.5874^{*}	0.2051	9574.518	9650.036
PH(5)	5	1.4857^{*}	0.2328	9519.670	9604.627
PH(6)	6	1.8279^{*}	0.2741	9468.017	9562.413

*significant at 0.05

All six estimated models are summarized in Appendix 8 (the baseline dummy variable = 1.industry_b). The minimum values of information criteria AIC and BIC are associated with PH(6) model with five interior knots, followed by PH(5) and PH(4). Since the estimated coefficients of all parameters are statistically significant at a 0.05 level, we choose model PH(6) as the most suitable model for our data.

6.2.2 Estimation of Bankruptcy Rates and Rating

Finally, we use two selected models with binary variables, the Weibull and PH(6), and calculate the cumulative hazard rates based on the equation (3.95). As can be seen in Appendix 8, the estimated Weibull model has the prediction _rcs1, which is the spline model parameter $\gamma_1 = 0.9464$, and constant _cons, $\gamma_0 = -11.2105$. There are also estimated coefficients xb for each value of a categorical variable in the table. We use the Weibull model to estimate the cumulative hazard rates analogically to Chapter 6.1 when we applied the Cox model. We get the rates for three hypothetical companies and use them to estimate ratings.

We follow the equation (3.95) using the Weibull model and predict the log cumulative hazard rates. Firstly, we need to calculate the log time $\ln(t):\ln(1825)=7.51, \ln(3650)=8.2$. Then, we calculate the log cumulative hazard rates as follows:

$$\ln(H_{1w,t=5}) = -11.2105 + 0.9464 \cdot 7.51 - 0.6378 + 0.7599 + 0 = -3.9809,$$
$$\ln(H_{2w,t=5}) = -11.2105 + 0.9464 \cdot 7.51 - 0.6378 + 0.7599 + 0.2915 = -3.6894,$$
$$\ln(H_{3w,t=5}) = -11.2105 + 0.9464 \cdot 7.51 + 0 + 0 + 0.2915 = -3.8115.$$

$$\ln(H_{1w,t=10}) = -11.2105 + 0.9464 \cdot 8.2 - 0.6378 + 0.7599 + 0 = -3.3279,$$
$$\ln(H_{2w,t=10}) = -11.2105 + 0.9464 \cdot 8.2 - 0.6378 + 0.7599 + 0.2915 = -3.0364,$$
$$\ln(H_{3w,t=10}) = -11.2105 + 0.9464 \cdot 7.51 + 0 + 0 + 0.2915 = 3.1585.$$

Next, we express the exponentiated cumulative hazard rates, which are used as the cumulative bankruptcy rates:

$$ECBR_{1w,t=5} = \exp(-3.9809) = 0.0187, \quad ECBR_{1w,t=10} = \exp(-3.3279) = 0.0359,$$
$$ECBR_{2w,t=5} = \exp(-3.6894) = 0.0250, \quad ECBR_{2w,t=10} = \exp(-3.0364) = 0.0480,$$
$$ECBR_{3w,t=5} = \exp(-3.8115) = 0.0221. \quad ECBR_{3w,t=10} = \exp(-3.1585) = 0.0425.$$

The cumulative bankruptcy rates are calculated analogically for the PH(6) model. Generally, if we apply flexible models with interior knots, we need to use the functions of the restricted cubic splines (basis functions of ln*t* for each interior knot). We use predictions of survival probabilities from Stata software and calculate the cumulative hazards. Then, the estimated cumulative hazard rates are as follows:

$$ECBR_{1ph(6),t=5} = 0.0165, \quad ECBR_{1ph(6),t=10} = 0.0346,$$
$$ECBR_{2ph(6),t=5} = 0.0225, \quad ECBR_{2ph(6),t=10} = 0.0470,$$
$$ECBR_{3ph(6),t=5} = 0.0220. \quad ECBR_{3ph(6),t=10} = 0.0455.$$

Comparing the estimated cumulative hazard rates at $t = 5$ with $t = 10$, they are higher for a longer time horizon. However, the rates calculated by both models are similar, and we can get only small differences.

Next, we use the estimated rates to determine the rating. The estimated ratings are summarized in Table 6-13. Since the estimated cumulative bankruptcy rates of both models are similar, there are no substantial differences in rating categories. The models estimate rating BBB for all cases if $t = 5$ years. In the time horizon $t = 10$ years, the models suggest the same rating BBB for cases 2 and 3, and A or AA for case 1. Thus, according to our results, micro, limited-liability industrial companies are likely to upgrade their rating over time, suggesting a lower credit risk compared to other cases with an unchanged rating.

Table 6–13 Rating estimation (Weibull model, PH6)

	AAA	AA	A	BBB	BB	B	CCC/C
AS	-0.0336	-0.0319	-0.0310	-0.0265	0.0027	0.0945	0.4263
$D_{1w,t=5}$	0.0151	0.0150	0.0148	0.0134	0.0147	0.0226	0.0447
$D_{1w,t=10}$	0.0017	0.0004	0.0001	0.0054	0.0238	0.0601	0.0570
$D_{2w,t=5}$	0.0089	0.0088	0.0086	0.0072	0.0085	0.0164	0.0385
$D_{2w,t=10}$	0.0134	0.0121	0.0116	0.0063	0.0121	0.0484	0.0453
$D_{3w,t=5}$	0.0117	0.0116	0.0114	0.0100	0.0113	0.0192	0.0413
$D_{3w,t=10}$	0.0081	0.0068	0.0063	0.0010	0.0174	0.0537	0.0506
$D_{1ph6,t=5}$	0.0171	0.0170	0.0168	0.0154	0.0167	0.0246	0.0467
$D_{1ph6,t=10}$	0.0010	0.0003	0.0008	0.0061	0.0245	0.0608	0.0577
$D_{2ph6,t=5}$	0.0111	0.0110	0.0108	0.0094	0.0107	0.0186	0.0407
$D_{2ph6,t=10}$	0.0134	0.0121	0.0116	0.0063	0.0121	0.0484	0.0453
$D_{3ph6,t=5}$	0.0116	0.0115	0.0113	0.0099	0.0112	0.0191	0.0412
$D_{3ph6,t=10}$	0.0119	0.0106	0.0101	0.0048	0.0136	0.0499	0.0468

When we compare our main findings with the Cox model (Table 6-9), we can see that the parametric models assign a more stable rating in some cases. Since the parametric models use broadly defined categorical variables, they are useless for a practical rating estimation. However, they provide solid evidence that the probability of corporate survival and, thus, the cumulative bankruptcy rates depend on the chosen corporate characteristics such as the industry, legal form and size.

6.3 Chapter Summary

This chapter followed a previous study in which it was found that attributes such as size, legal form or sector affect corporate survival probability. Therefore, these results were further verified using other selected survival analysis approaches, such as semiparametric and parametric models. The goal was to obtain survival models and to express survival and cumulative hazard functions, which are used to estimate the probability of survival or hazard rates depending on the period of operation of the company.

First, the Cox proportional hazards model was applied. Then, since some categorical variables were not statistically significant and due to some deviation from the proportional hazards assumption, the variables were adjusted and used as binary. Binary variables were created simply by some groups of categorical variables being combined after a thorough examination to have some justification in the models. Due to the proportional hazards assumption violation, the sector variable was finally excluded from the model and used as a loss. Subsequently, a stratified Cox model with two categorical variables, legal form and size, was derived. It has been proven that the survival probability of companies from the service and industrial sectors is higher compared to other sectors.

Subsequently, the stratified Cox model was used for practical application and expression of survival and cumulative hazards function for three chosen combinations of categorical variables, analogously to Chapter 5.2. The results confirmed that corporate characteristics influence survival probabilities, expressed using the formulation of survival and cumulative hazard functions. These were then practically used to formulate baseline functions for selected variable values and to estimate cumulative hazard rates for a time horizon of five and ten years. Subsequently, cumulative hazard rates were used to determine the rating, using the procedure proposed and described in Chapter 5.3.2.

Similarly, parametric models were applied in the next part of this chapter, specifically the Weibull model and the flexible model with multiple interior knots. First, the models were used without variables to assess the data fit. Based on this, categorical variables were then involved in the models. Due to the statistical insignificance of some variables, six models with modified binary categorical variables were finally estimated. These were subsequently compared, after which the two Weibull and PH(6) models were used for practical application and calculating cumulative hazard rates for two periods. Finally, the procedure for converting these measures into ratings was applied, and ratings were determined for the selected combinations of categorical variables.

The results of using the final models to determine the cumulative bankruptcy rates and the subsequent rating assessment for the three selected combinations of variables (i = 1, 2, 3) are summarized in Table 6-14, where the symbol i denotes the following combinations:

a) Industrials, limited-liability, micro (i = 1),
b) services, joint-stock, large (i = 2),
c) agriculture, cooperative, medium (i = 3).

Table 6–14 Comparison of estimated ratings

t	***i***	**Cox**		**Weibull**		**PH(6)**	
		ECBR	Rating	ECBR	Rating	ECBR	Rating
5	1	0.0193	BBB	0.0187	BBB	0.0165	BBB
5	2	0.0264	BBB	0.0250	BBB	0.0225	BBB
5	3	0.0201	BBB	0.0221	BBB	0.0220	BBB
10	1	0.0372	AAA	0.0359	A	0.0346	AA
10	2	0.0509	B	0.0480	BBB	0.0470	BBB
10	3	0.0473	B	0.0425	BBB	0.0455	BBB

The main findings of this application confirm that the chosen corporate characteristics affect the survival probability and rating, which is in line with the contribution of Chapter 5. The influence of individual variables on the likelihood of survival varies and is determined with the help of estimated models. We can see in Table 6-14 that the cumulative bankruptcy rates calculated using each of the three models are very similar, which also leads to similar ratings. Nevertheless, the parametric models might capture some variables' influence more accurately, leading to a less volatile rating in the ten-year horizon compared to the Cox model.

Furthermore, the overall results, following the estimate of cumulative bankruptcy rates, indicate a dynamic development of the rating over time.

Although the results of this study brought us interesting findings, it is clear that the general use of models with only three broadly defined categorical variables has no practical use. On the other hand, though, they provide solid evidence that the probability of corporate survival depends on the industry, legal form and size. To extend our main findings, we will focus on the role of financial ratios in the following chapter.

Chapter 7

The Use of Financial Performance Indicators in Survival Analysis

The results of previous chapters suggest that survival depends on corporate characteristics such as industry, legal form and size. In this section, we focus on the role of financial performance indicators and their potential use in survival analysis. Thus, the aim of this chapter is to use the Cox proportional hazards model and selected parametric models to analyse the effect of financial variables on time to corporate bankruptcy.

All models are estimated based on the data sample described in the previous section. Each record documents the life of a particular company and corresponding 22 time series of variables based on financial statements, considered quantitative variables in this study. Hence, they provide some information on the financial performance, including the size of each company measured as the total assets and financial ratios of activity, profitability, liquidity and solvency observed at the end of each year. In this application, we build on the main findings of credit rating analysis (Chapter 4) and focus on using the same variables. We especially emphasise the practical use of all models and their appropriateness for modelling survivorship data and corresponding rating prediction.

The structure of this chapter is as follows. First, the data sample and corporate survivorship are analysed using the Cox proportional hazards model. Then, we use parametric survival models to estimate survival functions. Finally, the estimated models are compared, and the main findings are summarized.

7.1 Estimation of Survival Models

In this section, the semi-parametric and parametric approaches are applied to assess the effect of financial variables on time to corporate bankruptcy. The Cox proportional hazards model is first used, and the selected parametric models are then applied. Finally, the chapter is focused on identifying the main factors of the corporate survival time, estimation of cumulative bankruptcy rates and their usage for rating prediction.

We employ the same data sample as in Chapter 6 for the survival analysis, where 12547 observations with 1125 failures will be used for model estimations. The original time series of 22 financial variables were checked for potential outliers. The outlying values were transformed and changed using the winsorizing procedure when the tails of the distribution were recoded to less extreme values. The 5% of the lowest values are recoded to the value of the 5th percentile, and the 5% of the highest values are recoded to the value of the 95th percentile (Ludwig-Mayerhofer, 2020). Then, the transformed and adjusted time series were used in the analysis.

7.1.1 Cox Proportional Hazards Models with Financial Variables

Firstly, financial variables with a significant effect on the survival probability were identified using the Cox proportional hazards models. Then, the models were checked for PH assumptions, and variables violating this assumption were removed from further analysis. Finally, after adjustments, the six most influential variables are considered in the Cox models (Table 7-1).

Table 7–1 Financial variables

Variable	**Financial indicator**	**Mean**	**Median**
roa	Return on assets	0.0284	0.023
roc	Return on costs	0.1313	0.0544
cla	Coverage long assets (equity+long term liab)/fixed assets	2.5867	1.0200
er	Equity ratio (equity/total assets)	0.4671	0.4900
liab_t	Liability turnover (days)	191.432	79.5538
td	Debt ratio (debt/equity)	0.5123	0.4833

Next, three Cox models were estimated using the selected variables (Table 7-2). When using six variables (model 3), two are not statistically significant. Therefore, two other models (model 1, and model 2) with five statistically significant variables are estimated and largely described in the following text.

Table 7–2 Estimated Cox models

	Model 1	**Model 2**	**Model 3**
roa	x	-0.6896**	-0.6425
roc	-0.3413*	-0.1967**	-0.1966**
cla	0.0188**	0.0200**	0.0207*
er	-0.9997**	x	-0.9617
liab_t	-0.0006*	-0.0006*	-0.0006*
td	2.3486*	3.3199*	2.3774*
PH test (global)	0.5063	0.3646	0.4777
Violation	x	x	x
Explained variation R_D^2	0.3970	0.3944	0.3952
AIC	12174.72	11917.48	11917.41
BIC	12220.18	11962.71	11971.70

*significant at 0.05, **significant at 0.10

According to the global PH test, models 1 and 2 are suitable, and all used variables satisfy the PH assumption. The graphical representation of the PH assumptions is shown in Appendix 9. As to other goodness-of-fit tests, the values of the explained variation of both models and the information criteria of AIC and BIC are similar.

Estimated coefficients explain the association between each variable and corporate survival time. For example, in model 1, we can see that an increase in three following variables *roc*, *er*, *liab_t* decreases the hazard, while the rise in *cla* and *td* increases the hazard. The same holds for model 2, where increasing additional variable *roa* reduces the hazard.

Next, we use the exponentiated individual coefficients to interpret the results. Finally, the hazard ratios are summarized in Table 7-3, representing the hazard ratio for a 1-unit change in the corresponding variable.

Table 7–3 Hazard ratios

	Model 1	**Model 2**	**Model 3**
roa	x	0.5018	0.5260
roc	0.7108	0.8212	0.8215
cla	1.0190	1.0202	1.0209
er	0.3680	x	0.3822
liab_t	0.9994	0.9994	0.9994
td	10.4726	4.3646	10.7771

For example, in model 1, a 1-unit increase in *roc* decreases the hazard by 28.9% because exp(-0.3413) = 0.7108. From the economic point of view, the results are consistent with theoretical assumptions. The higher the return on costs and equity ratio, the lower the bankruptcy hazard. Next, total indebtedness is another significantly important factor in the model, with an opposite impact on the hazard. A 1-unit increase in *td* increases the hazard substantially, even by 947%. The variable *liab_t* decreases the hazard but with a minimal effect. All estimated coefficients, or hazard ratios, respectively, are similar in all three models, except the variable *td*. The association between the hazard and the variable is the same in model 2; however, the impact is smaller than in the other two models. The hazard by a 1-unit rise in *td* is increased by 336% in this case.

The explained variation measured by the adjusted index of determination R^2 equals 0.3970 in model 1 and 0.3944 in model 2. Comparing models with a changing number of predictors (Table 7-4, Table 7-5), the greatest contribution to the explained variation is carried by covariates *td*. The contribution is 84.33% relative to the explained variation in model 1 and 84.89% in model 2. It is followed by *roc* (5.79%, 5.68%), *cla* (5.47%, 5.43%) and *liab_t* (4.31%, 4.34%). The remaining variables contribute only minimally.

Table 7–4 Explained variation (model 1)

Variables in model	R_D^2	St. error	95% conf. interval	
roc, cla, er, liab_t, td	0.3970	0.0183	0.3603	0.4320
cla, er, liab_t, td	0.3740	0.0182	0.3378	0.4089
er, liab_t, td	0.3523	0.0173	0.3179	0.3857
liab_t, td	0.3519	0.0173	0.3175	0.3853
td	0.3348	0.0181	0.2989	0.3698

Table 7–5 Explained variation (model 2)

Variables in model	R_D^2	St. error	95% conf. interval	
roa, roc, cla, liab_t, td	0.3944	0.0185	0.3574	0.4298
roc, cla, liab_t, td	0.3957	0.0183	0.3590	0.4308
cla, liab_t, td	0.3733	0.0182	0.3370	0.4083
liab_t, td	0.3519	0.0173	0.3175	0.3853
td	0.3348	0.0181	0.2989	0.3698

Since model 1 includes more statistically significant variables at a 0.05 level than model 2, with similar other characteristics, this model is further used for a detailed description (the interpretation of model 2 is analogical).

The cumulative hazard function under this model for a company with covariates $x_1, x_2, ..., x_5$, where k = 5 can be expressed using the general formula (3.72),

$$H\left(t \middle| x_1, x_2, ..., x_5\right) = H_0(t)\exp(-0.3413x_1 + 0.0188x_2 - 0.9997x_3 - 0.0006x_4 + 2.3486x_5), \tag{7.1}$$

where $x_1, x_2, ..., x_5$, refer to variables included in model 1 (Table 7-2). For example, we can express the hazard rate of a hypothetical (average) company with mean values of variables as shown in Table 7-1 as:

$$H_1\left(t \middle| x_1, x_2, ..., x_5\right) = H_0(t)\exp(0.6252) = H_0(t)1.8186. \tag{7.2}$$

The expression in (7.1),

$$\exp\left(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5\right) = \exp(\mathbf{x\beta}), \tag{7.3}$$

is referred to as the log relative hazard or the risk score. Similarly, the "mean" survival function is expressed as follows,

$$S_1(t \mid x_1, x_2, ..., x_5) = S_0(t)^{\exp(0.6252)} = S_0(t)^{1.8186}. \tag{7.4}$$

Figure 7-1 demonstrates that the cumulative baseline hazard and the "mean" hazard functions of model 1 increase. As already mentioned, the baseline function (H0) is a function when the values of all variables are equal to zero, so in our case, all financial indicators are zero. Considering a hazard function with average variables (H1) values, the hazard is higher for every value of time (*t*). The hazard functions are plotted on the left side of the figure, while survival functions on the right side show the decreasing probability of survival as the time is greater. The baseline survival function (S0) is evaluated with all the covariates equal to zero, created as the Kaplan-Meier estimate. The survival and cumulative hazard functions of model 2 (Figure 7-2) have a similar shape; however, the baseline functions are slightly different because they are not based on the same sample of companies due to other use of variables. Thus, the baseline cumulative hazard is lower when *roa* is used instead of *er*. On the other hand, both models' "mean" cumulative hazard and survival functions are almost the same.

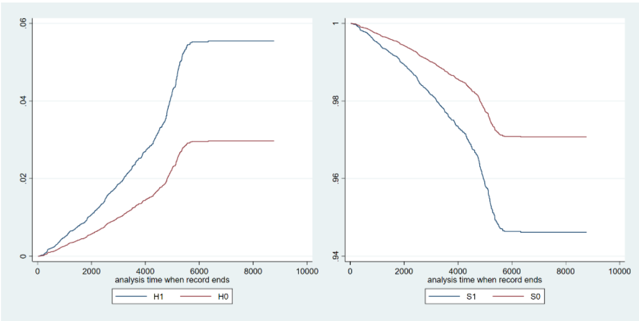


Figure 7–1 Baseline versus average company functions (model 1)

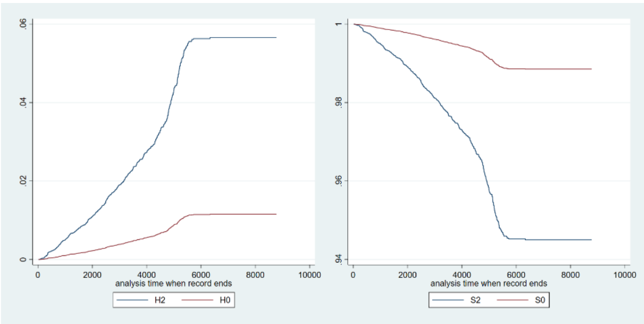


Figure 7–2 Baseline versus average company functions (model 2)

Similarly to the non-parametric approach, we estimate the hazard function using the kernel smoother. Both smoothed hazard functions, the baseline and the "mean" function are shown in Figure 7-3. Although the models consider a different set of variables and, thus, the baseline functions differ, the estimated "mean" hazard functions have almost the same shape with very similar functional values. Based on these results, it can be said that both models determine very similar hazard and survival functions, regardless of which combination of variables we use.

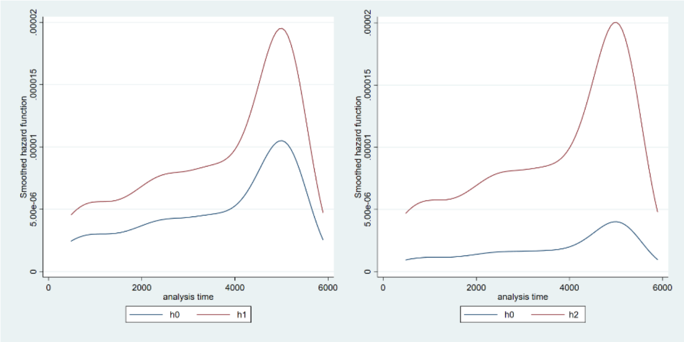


Figure 7–3 Hazard functions (model 1, model 2)

The hazard rates change, meaning the risk of bankruptcy for a "mean" company is not constant over time. The hazard rate increases with time and is highest when $t = 5000$ days. Growth is modest initially, but the curve is steep from $t = 4000$ days. Therefore, it can be said that the most critical period for corporate survival is between 4000 – 5000 days of the company's life. The hazard rate decreases significantly in the longer term, i.e. more than 5000 days. Thus, companies that have survived the previous critical period might be strong companies with a lower bankruptcy risk in the next period.

The estimated models are evaluated through the overall model fit using Cox-Schnell residuals. If the Cox regression model fits our data, then the true cumulative hazard function conditional on the covariate vector has an exponential distribution with a hazard rate of one (Cleves et al., 2010). The figures below (Figure 7-4) show both models' Nelson-Aalen cumulative hazard estimator plots for Cox-Snell residuals. We can see some variability about the 45º, particularly in the right-hand tail. Cleves et al. (2010) argue that some variability is expected due to the reduced effective sample caused by prior failures and censoring. However, there is a better fit for model 2, as seen on the right side.

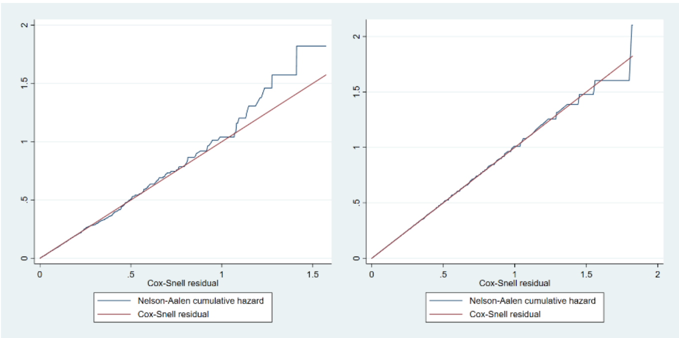


Figure 7–4 Cumulative hazard of Cox-Snell residuals (model 1, model 2)

To sum it up, considering the adjusted explained variability, information criteria and the goodness of fit, model 2 with five covariates *roa, roc, cla, td, liab_t* is preferred to model 1. Based on the overall results from the application of the Cox models, financial variables with a significant impact on survival time were identified, and only variables that satisfy the PH assumption were used in the models. Undesirably, this condition was not met for most variables in the data sample; therefore, the set of possible variables entering the model was quite limited.

7.1.2 Parametric Models with Financial Variables

To compare with the Cox model, the Weibull and extended (flexible) parametric models are applied in this section to explore the effect of financial variables on the probability of survival. Based on the application of the Weibull model, nine variables are identified as potentially influential variables on the probability of survival (Table 7-6).

Table 7–6 Financial variables

Variable	Financial indicator	Mean	Median	Min	Max
roa	Return on assets	0.0284	0.0230	-0.2608	0.338
roc	Return on costs	0.1313	0.0544	-1.231	1.793
cla	Coverage long assets (equity+long term liab)/fixed assets	2.5867	1.0200	-0.0622	17.07
er	Equity ratio (equity/total assets)	0.4671	0.4900	-0.2492	0.9741
liab_t	Liability turnover (days)	191.432	79.5538	13.668	1252.99
td	Debt ratio (debt/equity)	0.5123	0.4833	0.02	1.22
cr	Current ratio	3.9182	2.0301	0.2384	19.248
lnta	Ln(Total assets)	9.0552	9.1135	5.3845	12.34
rec_to_ca	Receivables to current assets	0.4751	0.4653	0.0009	0.9597

However, when all nine variables are used, *roc* is not statistically significant at a 0.05 significance level and can only be accepted at a 0.10 level. For this reason, the other two models with a different mix of eight variables are developed (model 1, model 2). Overall, three Weibull models are estimated, and the main results are presented in Table 7-7.

All models' estimated coefficients, values of explained variation, and information criteria are similar – however, the AIC and BIC values of model 3 are slightly lower than in other models. In addition, since it includes all nine variables, this model is further used for a more detailed description and interpretation of results.

Table 7–7 Weibull models

	Model 1	**Model 2**	**Model 3**
roa	-1.3801*	x	-0.8511*
roc	x	-0.3419*	-0.1907**
cla	0.0294*	0.0278*	0.0300*
er	-1.4211*	-1.5433*	-1.4885*
liab_t	-0.0006*	-0.0006*	-0.0006*
td	1.9628*	1.8981*	1.9334*
cr	-0.0836*	-0.0867*	-0.0834*
lnta	0.1482*	0.1424*	0.1534*
rec_to_ca	0.9681*	0.9595*	0.9821*
p	1.2701*	1.2865*	1.2813*
_const	-16.1520*	-16.1509*	-16.2687*
Explained variation R_D^2	0.4417	0.4433	0.4463
AIC	5330.788	5393.084	5274.162
BIC	5421.211	5483.898	5373.565

*significant at 0.05, **significant at 0.10

The association between each variable and corporate survival time is based on the estimated coefficients analogically to the Cox model. For example, an increase in *roa, roc, er, liab_t* and *cr* decreases the hazard, while the rise in *cla, td, lnta and rec_to_ca* increases the bankruptcy hazard. The same holds for other models with a lower number of variables. These results support the main findings of the Cox models in Chapter 7.1.1 above.

The baseline hazard function of model 3 can be expressed using the formula (3.90) when all variables equal zero,

$$\widehat{H}_0(t) \approx 2.28t^{1.28}\exp(-16.15) = 0.000000221t^{1.28}. \tag{7.5}$$

Then, the overall hazard is:

$$\begin{aligned} H(t|x_1, x_2, ..., x_9) = 0.000000221t^{1.28}\exp(&-0.8511x_1 - 0.1907x_2 + \\ &+ 0.3x_3 - 1.4885x_4 - 0.0006x_5 + 1.9334x_6 - \\ &- 0.0834x_7 + 0.9821x_8 + 1.2813x_9). \end{aligned} \tag{7.6}$$

7.1.3 Estimation of Hazard Functions by Flexible Models

Next, in this section, we estimate flexible models with various numbers of interior knots analogically to section 6.2.1. The flexible models are based on the Weibull model (model 3) with nine financial variables. Thanks to this additional analysis using the interior knots, we will find more detailed evidence about the shape of the hazard functions and, thus, more information about the probability of survival depending on the age of the companies.

The estimates _rcs1 (*p*-parameters) and information criteria are presented in Table 7-8 (all models are summarized in Appendix 10). The characteristics of all models are comparable. Therefore, according to the information criteria, we prefer models with minimum values of BIC – PH(6), PH(5), and the Weibull model, followed by PH(2), PH(4) and PH(3). Since some of the estimated coefficients _rcs in PH(2), PH(3), and PH(5) models are not statistically significant, we finally choose the Weibull model, PH(4) and PH(6) as suitable models for our data.

Table 7–8 Flexible models

Model	**d.f.**	$\hat{\beta}$ **(_rcs1)**	**Standard error (SE)**	**AIC**	**BIC**	**Explained variation**
Weibull	1	1.2813*	0.0442	5274.162	5373.565	0.4463
PH(2)	2	1.3506*	0.1585	5275.950	5384.389	0.4455
PH(3)	3	1.2936*	0.1942	5277.554	5395.030	0.4445
PH(4)	4	1.6828*	0.2440	5261.063	5387.575	0.4416
PH(5)	5	1.6060*	0.2696	5231.471	5367.020	0.4372
PH(6)	6	2.0159*	0.3223	5164.083	5308.669	0.4351

*significant at 0.05, **significant at 0.10

The cumulative and hazard baseline functions of well-performed models PH(1), PH(4), and PH(6) are shown on the left side in Figure 7-5, where the functions are constructed for the models with zero covariates. Based on the right-hand side plot, all models suggest that the intensity of the bankruptcy varies with time, or in other words, with the length of company life. So, for example, the overall hazard rises as companies get older based on the Weibull model. Initially, the curve is steeper up to approximately 2000 days, suggesting a rapidly increasing hazard of bankruptcy at the beginning of company life (up to 5.5 years). Then, the curve rises at a slower pace. Based on the spline functions, it can be assumed that the bankruptcy hazard is the highest for companies aged between 7.5 and 13.5 years. Then the hazard rate decreases and stabilize after approximately 16.5 years.

The Weibull model's hazard function on the right side monotonically increases with time (the shape is determined by parameter $p = 1.28$), and flexible models' hazard functions fluctuate around it. For example, the hazard rates of model PH(4) are above the Weibull rates when $t = 200 – 1000$ and 2700 - 4800 days; otherwise, they are lower. Overall, hazard rates of flexible models range slightly around the Weibull model to approximately 5000 days (13.7 years). However, they are lower over a longer period, with a declining trend compared to the Weibull, slowly increasing hazard function.

The hazard ratios of the Weibull, PH(4), and PH(6) models can be further used for interpretation (Table 7-9). The hazard ratios of all three models are similar and support the main findings from the previous section, 7.1.1, suggesting only slight differences compared to the Cox model. For example, a 1-unit increase in *td* increases the bankruptcy hazard by 591%, while a 1-unit increase in *er* decreases the hazard by 77.4% in the Weibull model without interior knots.

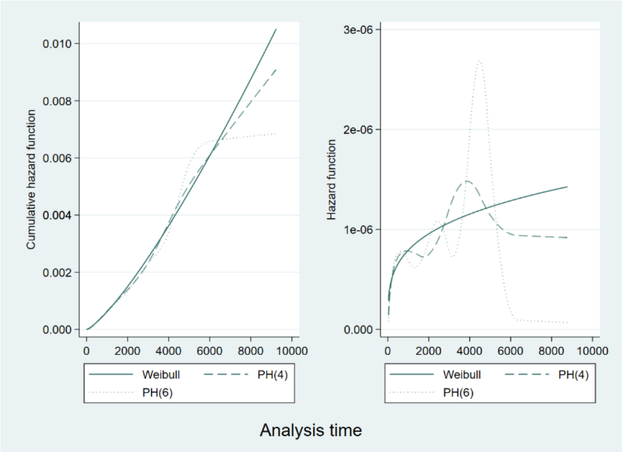


Figure 7–5 Baseline hazard functions

Although we used flexible models to examine our data more thoroughly and derive the relevant hazard functions, all models are similar in terms of the comparative criteria used. For this reason, in the next part of the work, we will consider the basic Weibull model without interior knots, which shows the highest explained variation and minimal standard error. In addition, its use is simpler and more practical.

Table 7–9 Hazard ratios

	Weibull	PH(4)	PH(6)
roa	0.4269	0.3964	0.3370
roc	0.8264	0.8303	0.8411
cla	1.0304	1.0301	1.0298
er	0.2257	0.2354	0.2489
liab_t	0.9994	0.9994	0.9994
td	6.9143	6.8153	6.5041
cr	0.9200	0.9203	0.9204
lnta	1.1658	1.1690	1.1717
rec_to_ca	2.6700	2.6624	2.6737

7.1.4 Combined Weibull Models

Finally, we estimate the Weibull model using both financial and categorical variables. Hence, the variables industry, legal, and size (see Chapter 6) are included in the model to explore the role of these factors on estimated rating in combination with financial indicators. Adding new variables affects the originally obtained coefficients and their statistical significance. Finally, we get two models with acceptable results (Table 7-10). The detailed coefficient estimations are attached in Appendix 11.

Table 7–10 Combined Models

	Model 4		Model 5	
	Coeff.	HR	Coeff.	HR
2.size_b	0.3600*	1.43	0.3531*	1.42
2.industry_b	-0.8332*	0.43	x	x
roa	x	x	-0.8475**	0.43
roc	-0.3681*	0.69	-0.1811**	0.83
cla	0.0297*	1.03	0.0321*	1.03
er	-3.4656*	0.03	-1.4744*	0.23
liab_t	-0.0004*	1.00	-0.0005*	1.00
td	x	x	1.9323*	6.91
cr	-0.0972*	0.91	-0.0819*	0.92
lnta	0.0636*	1.07	0.1137*	1.12
rec_to_ca	1.0017*	2.72	0.9738*	2.65
p	1.2886*	x	1.2845*	x
cons	-13.5523*	1.30e-06	-16.1810*	9.39e-08

*significant at 0.05, **significant at 0.10

Both models include categorical and financial variables statistically significant at 0.05 or 0.10 in some cases. While model 4 is based on seven financial variables and two binary categorical variables of size and industry, model 5 contains nine financial variables and accounts only for the size.

The estimated coefficients of both models are mostly similar, which is also reflected in the hazard rates (HR). Recall that categorical variables are compared with the baseline levels in interpretation. For example, in model 4, the hazard rate increases by 43% for small, medium and large companies compared to micro-sized firms. On the other hand, regarding the effect of industry, the hazard rates decrease by 57% for services and industrials. Model 5 supports that the hazard rate for small, medium and large firms relative to micro companies increases, by 42% in this case.

The effect of financial variables on hazard rates is analogous to the models without considering the categorical variables. The hazard rates are positively affected by variables *td*, *rec_to_ca*, *lnta* and *cla*, ranked by their influence. While the relationship of variable *liab_t* is rather neutral to the hazards rates, other variables decrease them.

However, since it is unclear whether categorical variables improve the models' quality, the combined models will be compared with a model containing only financial variables in the following section.

7.1.5 Model Comparison

In previous sections, estimated parametric models differ according to what and how many input variables were used. In this section, the Weibull model containing nine financial variables (model 3) will be compared with models combining categorical and financial variables (model 4 and model 5). These models are summarized in Table 7-11.

Table 7–11 Summary of Weibull models

	No. of categorical variables	No. of continuous variables	AIC	BIC
Model 3	x	9	5274.162	5373.565
Model 4	2	7	5279.170	5379.066
Model 5	1	9	5259.722	5368.161

To assess the goodness of fit, we plot the Nelson-Aalen cumulative hazard estimator for Cox-Snell residuals analogically to the Cox model in Chapter 7.1.1 (Figure 7-6). We observe that model 3 and model 5 perform a similar fit. However, there is some lack of fitting in all models. For example, although model 4 performs relatively well, it has higher variability about the 45° line on the right-hand tail. Nevertheless, Cleves et al. (2010) suggest that this phenomenon is expected, especially in the right-hand tail, due to reduced effective sample caused by prior failures and censoring. However, based on the graphs presented, it is clear that the remaining models have a variability around the 45-degree line lower and, therefore, can be considered better-constructed models. The AIC and BIC criteria information values correspond to this deduction, preferring models 3 and 5 over model 4. This conclusion suggests a significant influence of firm size and nine selected financial variables on the time to corporate bankruptcy.

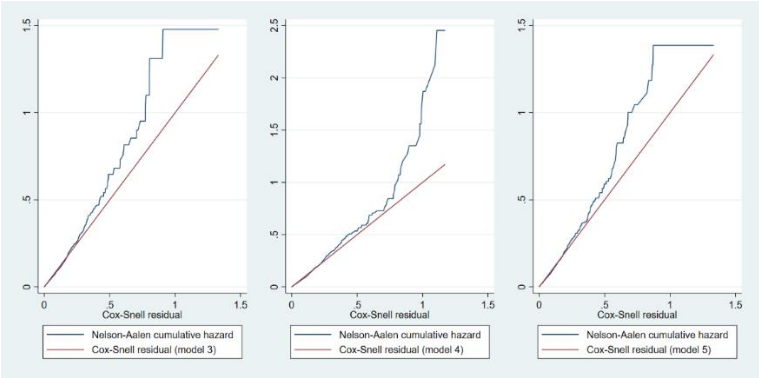


Figure 7–6 Cumulative hazard of Cox-Snell residuals (models 3–5)

Overall, the differences between the values of the information criteria are very small, and we can say that the models are very similar. However, it is necessary to realize that some influences can only be captured in the model to a limited extent, such as categorical variables, which are used very simply as binary. Therefore, using models with only financial variables is more appropriate. It is also easier to apply them in practice and adjust the rating individually regarding the results achieved in this study. So, if we are, for example, assessing small, medium, or large-sized companies compared to micro companies, we will be more cautious and possibly adjust the resulting rating slightly down.

Assuming that we are aware of the influence of selected corporate characteristics and can take them into account in the final assessment, we will focus on applying models containing only financial variables in the following section.

7.2 Cumulative Bankruptcy Rates and Ratings for Specific Parameters

In this chapter, two models containing only financial variables, the Cox proportional hazards and the Weibull model, are used for bankruptcy rate calculations and rating assessment. The models are applied to estimate the survival functions for two periods, 5 and 10 years, analogically to chapters 6.1.2 and 6.2.2. First, the survival functions are estimated using specific parameters such as the variable values and the time range. Next, the cumulative hazard rates representing the cumulative bankruptcy rates are calculated based on equation (3.64). This chapter is divided into two sections. Firstly, the models are used for specific values of all input variables for a given time range. In the second part, only the time range is specified, and rates are calculated based on individual companies' values.

7.2.1 Application for Specific Values of Financial Variables

As mentioned above in the text, we apply two models chosen as the most suitable based on the results from the previous sections. Remind the estimated coefficients of both models in Table 7-12.

Table 7–12 Cox and Weibull model

	Cox (model 2)	**Weibull (model 3)**
roa	-0.6896	-0.8511
roc	-0.1967	-0.1907
cla	0.0200	0.0300
er	x	-1.4885
liab_t	-0.0006	-0.0006
td	3.3199	1.9336
cr	x	-0.0837
lnta	x	0.1534
rec_to_ca	x	0.9821

To use the Cox model, the baseline survival and hazard functions must be firstly determined. Assuming the time horizon of five years, $S_0 = 0.998$ and $H_0 = 0.002$ correspondingly. When the time horizon is ten years, the probability of survival decreases to $S_0 = 0.995$, and the equivalent cumulative hazard is $H_0 = 0.005$. Thus, the Cox model can be expressed as

$$S_1(t|x) = S_0(t)^{\exp(-0.6896roa - 0.1967roc + 0.02cla - 0.0006liab\_t + 3.3199td)}, \tag{7.7}$$

$$H_1(t|x) = H_0(t) \cdot \begin{bmatrix} \exp(-0.6896roa - 0.1967roc + \\ +0.02cla - 0.0006liab\_t + 3.3199td) \end{bmatrix}. \tag{7.8}$$

To apply the Weibull model, we use the estimated coefficients and two parameters, the constant $\gamma_0 = -16.2687$ and the spline model parameter $\gamma_1 = 1.2813.$ Then, the model can be written as

$$\begin{aligned} H(t|x) = \exp(&\gamma_0 + \gamma_1 \ln(t) - 0.8511roa - 0.1907roc + \\ &+ 0.03cla - 1.4885er - 0.0006liab\_t + 1.9336td - \\ &- 0.0837cr + 0.1534lnta + 0.9821rec\_to\_ca), \end{aligned} \tag{7.9}$$

where log times $\ln(t)$ for two time horizons are $\ln(1825) = 7.51$ and $\ln(3650) = 8.2.$

Next, we calculate the cumulative bankruptcy rates for different values of variables included in the models, where the financial situation of a particular company gives the values. Then, we assign the individual company rating using the same method as in Chapter 5.3

To apply and assess the ability of estimated models for rating prediction, we assume six hypothetical companies as scenarios *i* = 1, 2, …, 6 with the following variable values: mean, median, minimum and maximum (Table 7-6). In addition, based on our findings, we consider combinations minimizing and maximizing the bankruptcy hazards: Max (*roa, roc, er, liab_t, cr*) and Min (*cla, td, lnta, rec_to_ca*) to minimize and Min (*roa, roc, er, liab_t, cr*) and Max (*cla, td, lnta, rec_to_ca*) to maximize the bankruptcy hazard. The considered scenarios are summarized in Table 7-13.

Table 7–13 Scenario overview

i	Measure	Predicted survival probability	Predicted cumulative hazard rate
1	Mean	s_mean	ch_mean
2	Median	s_median	ch_median
3	Minimum	s_min	ch_min
4	Maximum	s_max	ch_max
5	Minimize hazard	s_min_h	ch_min_h
6	Maximize hazard	s_max_h	ch_max_h

Thus, we assume six combinations of variable values considered hypothetical companies, for which we estimate the cumulative bankruptcy rates and assign a rating for $t_1 = 5$ and $t_2 = 10$ years. We follow the procedure as described in Chapter 5.3.2. Recall that firstly, we calculate the estimated spread ES between CDR and estimated cumulative bankruptcy rates, $ECBR.$ Next, we compare ES

with the average spread AS and choose the category with the lowest absolute value of the difference D.

Based on the resulting difference values D , the assigned ratings for each case of a hypothetical company *i*, where *i* = 1, 2,…,6, are summarized in Table 7-14. Although the ratings are similar, those derived from the Cox model seem more volatile when considering both time horizons and different cases. On the other hand, the ratings based on the Weibull model are more consistent and make better economic sense from a longer-term perspective. For this reason, we focus on the results of the second approach in more detail.

Table 7–14 Rating estimation

	Cox model			**Weibull model**		
i	5 years	10 years	change	5 years	10 years	change
1	BBB	AAA	↑	BBB	AAA	↑
2	BBB	AAA	↑	BBB	AAA	↑
3	BBB	AAA	↑	BBB	AAA	↑
4	B	B	→	BBB	AAA	↑
5	BBB	AAA	↑	BBB	AAA	↑
6	CCC	B	↑	CCC	B	↑

Over a five-year horizon, we use the Weibull model to estimate the mid-range BBB rating for five cases, including case *i* = 5, where the values are assumed to minimize bankruptcy risk. In contrast, for case *i* = 6, the predicted rating is CCC, which aligns with values that maximize the bankruptcy hazard.

Above all, we can see the rating dynamics over the chosen time horizon when considering the predicted ratings if $t = 5$ and $t = 10$ years. As a result, there is an overall tendency for rating improvement in the longer-term horizon, supported by both models. While the rating upgrade for the majority of cases is from BBB to AAA, hence by three rating grades, the improvement of the last case changed by one rank, from CCC to B, kept in the speculative grade.

Figure 7-7 presents the predicted survival functions for all cases (*i* = 1, 2, ..., 6) over the time range of *t* = 0 to 3650 days. The survival functions for the first four scenarios are displayed on the left in the graph, while the remaining two scenarios are shown on the right for comparative purposes.

Overall, six survival functions are predicted based on the mean, median, minimum and maximum values, assuming all companies start at the same time point. The predicted survival probabilities are very similar based on the plots. Nevertheless, while the lowest survival function is associated with the median variable values, the highest function is expected for the maximum values of used variables. The right side of the graph shows the predicted survival functions for two additional cases, one for minimizing and the second for maximizing the bankruptcy hazard. Now, there is a significant visible difference between the predicted survival functions. The depicted curves support the main findings of the influence of variables used in the model. As can be seen, the minimum values of *roa*, *roc*, *er*, *liab_t*, *cr*, and the maximum values of *cla*, *td*, *lnta*, *rec_to_ca* substantially increase the bankruptcy hazard, and it is the riskiest combination of variables. On the other hand, the opposite values lead to the maximum probability of survival of all considered cases.

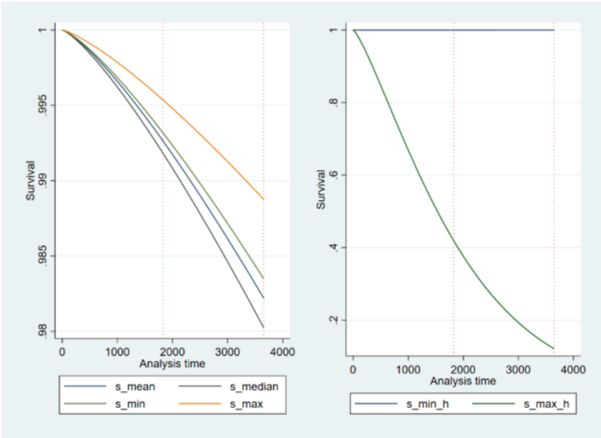


Figure 7–7 Predicted survival functions

The corresponding cumulative hazard rate functions are shown in Figure 7-8. These rates will be used to estimate ratings following the procedure outlined in Chapter 5.3.

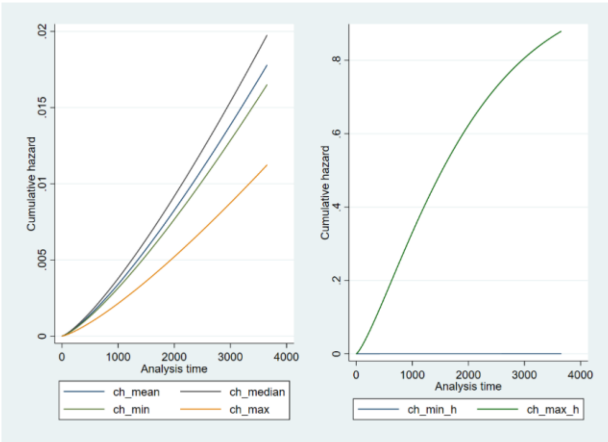


Figure 7–8 Predicted cumulative hazard functions

7.2.2 Application for Specific Values of Categorical and Financial Variables

While the models without categorical variables can be used to estimate the survival probability, hazard rate or, finally, rating based on the results from financial statements, the models extended for corporate characteristics can provide more specific estimates. Therefore, in the following text, these models will be used to determine the cumulative bankruptcy rates and subsequently determine the rating according to the procedure used in the above applications. Based on the results from the previous section, we consider only Weibull models in this part.

We use the estimated coefficients, the constant and the spline parameter to apply the Weibull models. For example, $\gamma_0 = -13.5533$ and the spline model parameter $\gamma_1 = 1.2886$ in model 4 with two categorical and seven financial variables (Table 7-10). The categorical variables are described in Table 6-3. The model can be written as

$$\begin{aligned} H(t|x) = \exp(&-13.553 + 1.289\ln(t) + 0.36 \cdot 2.size\_b - \\ &-0.83 \cdot 2.industry\_b - 0.3681roc + 0.03cla - \\ &-3.466er - 0.0004liab\_t - 0.097cr + \\ &+0.064lnta + 1.002rec\_to\_ca). \end{aligned} \tag{7.10}$$

where log times $\ln(t)$ for two time horizons are $\ln(1825) = 7.51$ and $\ln(3650) = 8.2$.

Next, we estimate the cumulative bankruptcy rates and rating using the same method as in the previous chapters. Finally, we consider a scenario with mean values of financial variables and corresponding combinations of corporate characteristics based on the used model. The overview of used variables and estimated ratings are summarized in Table 7-15. There is also a model without categorical variables for comparison (model 3).

Table 7–15 Rating estimation with categorical variables

Financial var. (mean values)	**Categorical var.**		**Rating**	
	Size	*Industry*	5 years	10 years
Model 3 *roa, roc, cla, er, liab_t, td, cr, lnta, rec_to_ca*	x	x	BBB	AAA
Model 4 *roc, cla, er, liab_t, cr, lnta, rec_to_ca*	1	1	BBB	AAA
	1	2	BBB	AAA
	2	1	BBB	AAA
	2	2	BBB	AAA
Model 5 *roa, roc, cla, er, liab_t, td, cr, lnta, rec_to_ca*	1	x	BBB	AAA
	2	x	BBB	AAA

The overall results suggest that the cumulative bankruptcy rates depend on the used variables and time, as proposed in previous chapters. The estimated ratings using the average spreads approach are the same based on all three models, suggesting that although the size and industry variables affect the cumulative bankruptcy rates, this influence is negligible from the point of view of the impact on the rating estimation. In addition, all models suggest a tendency to rating upgrade during a longer time-horizon. This result has an economic rationality, as the first years of a new company's operation can be considered riskier.

7.2.3 Application for a Specific Time Frame

After fitting the survival regression models for specific values of variables, we compute the survival functions for a particular time frame in this section. Compared to the last subchapter, when certain variables' values were used, they are not specified in this part. On the contrary, the actual observed values of each company are considered.

The steps described by Boswell and Gutierrez (2011) are followed in this part to obtain the predicted survival functions for a particular range of time. First, the predictions of the survival function for companies over the time range [0,5] and [0,10] years are calculated based on unconditional survival. Then, the cumulative hazards are computed based on survival functions, and the ratings are determined using the procedure described in Chapter 5.3. The relative proportions of rating grades based on three selected models are graphically shown in Figure 7-9.

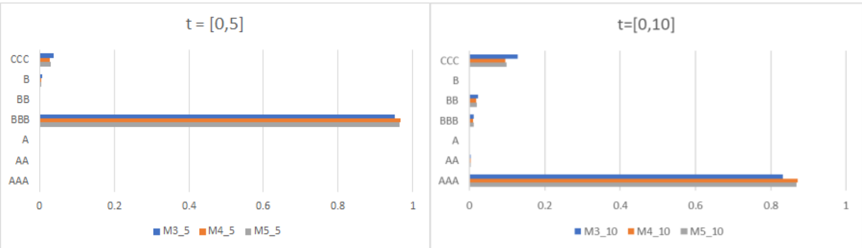


Figure 7–9 Estimated ratings

Note: Percentage share of estimated ratings for the time range [0,5] are on the left side, and for the time range [0,10] on the right side. For example, the notation M3_5 refers to model 3 for the time range [0,5], and other symbols are analogical.

All models show that over a five-year time horizon, most firms are rated BBB. If we consider a longer time horizon of ten years, then the rating of most companies is upgraded to AAA. These results apply to all models, however, model 3 without categorical variables has slightly different proportional results, as compared to the other two models. Overall, the results indicate that an improved rating can be expected for most companies in the longer term of ten years. On the other hand, the percentage share of companies rated with the lowest CCC rating is increasing with time.

For these findings, however, it is necessary to point out that different observations of the same company at different times are considered separate observations. Furthermore, it should also be noted that we assume that all companies start operating at the same time $t_0 = 0$. Therefore, there are certain assumptions and simplifications on which it is impossible to derive general results.

Overall, this section showed how the estimated survival functions could be used to calculate the cumulative bankruptcy rates by which the rating can be assigned for specific values of variables or a particular time frame. For this purpose, the procedure is based on average deviations of cumulative hazard rates from cumulative default rating rates published by credit rating agencies. In the context of data modelling, this is understandably a certain simplification. On the other hand, this approach provides a better understanding of the relationship between the selected variables and ratings, allowing for more accurate predictions of potential rating developments over time.

7.3 Chapter Summary

Chapter 7 was primarily devoted to the use of financial performance indicators in survival analysis. The aim was to derive survival models with financial variables and to determine their role in estimating the probability of corporate survival.

First, the Cox proportional hazards model was used, based on which six variables with a significant influence on survival probability were found: *roa*, *roc*, *cla*, *er*, *liab_t* and *td*. While the variables *cla* and *td* increase the hazards, the other variables reduce them. The greatest contribution to the explained variation is carried by *td*, and then *roc*, *cla*, *liab_t*. Furthermore, hazard functions were formed, and the role of baseline functions was clarified, including their graphical representation. The overall results based on the comparison of the two Cox models confirm the role of the following variables:

- Coverage of long-term assets (increase hazard),
- debt ratio (increase hazard),
- liability turnover (decrease hazard),
- return on assets (decrease hazard),
- return on costs (decrease hazard).

Since the use and application of the Cox model have certain limitations, we used the Weibull model with financial variables for comparison. Based on the main results, the importance of these indicators for estimating the probability of survival was confirmed. Specifically, it was found that while *roa*, *roc*, *er*, *liab_t* and *cr* decrease the hazards, the variables *cla*, *td*, *lnta* and *rec_to_ca* increase them. These results are consistent with those found by the Cox model.

The so-called flexible models with more interior knots were used in the next part to obtain a better and more detailed knowledge of the hazard function and, thus, the probability of survival depending on the companies' age and financial performance. Since the differences between the basic Weibull model and flexible models were small based on used criteria, the Weibull model without interior knots was further used in the next part. In addition, it achieved the lowest standard error value and the highest explained variation.

The Weibull model with nine financial variables was subsequently applied and combined with categorical variables characterizing size, industry and legal form. Hence, two models were developed, one with two and one with one categorical binary variable. Eventually, it turned out that the Weibull model containing merely financial variables is the most suitable for determining the probability of survival and, therefore, the possible rating estimation. Furthermore, the influence of categorical variables is considerably limited in the model. For this reason, it is a more appropriate alternative to consider corporate characteristics only after the subsequent use of the model.

Overall, it can be summarized that after detailed analysis and estimation of models with different types of variables, two resulting models, Cox and Weibull without categorical variables, were selected for further comparison and practical application. Both models were first used to calculate cumulative bankruptcy rates for specific values of variables defined for six scenarios and two periods. Based on Chapter 5, a rating was subsequently determined for all these cases, namely for 5 and 10 years of the company's life. Both models provided very similar rating grade estimations. The results of the Weibull are slightly more consistent and therefore it can be considered more suitable for our purposes. Moreover, both models agreed on the ratings when the values of the variables maximizing the hazard of bankruptcy were used. In this case, they predicted the same CCC rating in a period of five years, respectively B in a longer time horizon of ten years. In any case, both models put this scenario in the speculative grade, which is economically rational.

Finally, for completeness, Weibull models with and without categorical variables were applied to determine the rating for specific time frames based on observed values of variables in the data sample. However, this procedure was a certain simplification, which assumes that different observations of the same company at different times are considered separate observations. On the other hand, this approach can be used better to understand the relationship between the chosen variables and ratings and anticipate possible rating development over time.

The main findings of this application study confirm the role of nine financial variables and their influence on the survival probability:

- Coverage of long-term assets (increase hazard),
- current ratio (decrease hazard),
- debt ratio (increase hazard),
- equity ratio (decrease hazard),
- liability turnover (decrease hazard),
- the logarithm of total assets (increase hazard),
- receivables to current assets (increase hazard),
- return on assets (decrease hazard),
- return on costs (decrease hazard).

Furthermore, it can be said that the basic Weibull model without categorical variables and interior knots is most suitable for modelling our data and can serve as an appropriate tool for estimating the probability of survival. It can also be used to determine the probable rating and estimate its development over time, depending on the company's age and financial performance. This estimate should also be subjectively adjusted for the influence of other corporate characteristics (size, industry, legal form), the effect of which was demonstrated in Chapter 5.

Finally, the overview of all significant variables found by different approaches used in this monograph is summarized in Appendix 12. The tables show a clear comparison of the main results of both the rating and bankruptcy models. These main findings, regardless of the used model, confirm the influence of the key variables as listed above.

Chapter 8

Conclusion

This monograph was focused on modelling credit rating and the related probability of survival. For the main purpose of this work, a micro approach was used to measure credit risk based on monitoring basic indicators and allowing creditors to take the necessary actions in time. This method is used primarily by banks that have enough information about their clients and create personal credit scoring based on this approach. The second way is relying on rating agencies, which assign a certain rating grade according to the probability of repaying the obligation. In addition to banks and financial institutions, which set their regulations and credit risk measurement strategies, other financial market entities do not have specialised analytical unit possibilities or expertise. Still, they also need to assess the level of credit risk they undertake or intend to undertake. Therefore, both institutional and retail investors can most often use the services of rating agencies. In the Czech Republic, the issuance of corporate bonds has expanded in recent years, which has been supported by the economic boom and the environment of low interest rates. At that time, however, many bonds appeared on the market, some of which can be considered highly risky. These are bonds sold outside of organized markets, often through a direct offer by companies to potential investors.

As mentioned, important information about bonds is the rating, which is not used in some countries to the same extent as in highly developed countries with advanced capital markets. There, then, we often lack this information, which is why different approaches are used to measure credit risk using proprietary rating models. Their goal is to create mathematical-statistical multicriteria models, which can be used to assess credit risk and assign a rating based on selected indicators. Of course, this approach is considerably simplified compared to rating agencies' assessment. Still, it provides enough basic information about the rating quality, especially the main parameters influencing the rating. The models can then be used in practice for continuous monitoring and detection of potential risks and possible predictions for the future.

The principal aim of this book was to analyse the credit risk based on real data, assess main factors, explore mutual relations and draw conclusions related to credit rating assessment and survival probability. The attention was focused on regional markets and, subsequently, the narrower market within the Czech Republic. The main results of the application studies suggest that regardless of the data, methodology or even the output variable – credit rating or corporate survival, there are several key accounting-based variables or their modifications that determine the individual corporate credit risk. Overall, we suggest that these well-known and widely used variables that reflect all main areas of financial analysis, such as profitability, solvency, liquidity, activity and size, are crucial for credit risk assessment. Hence, from the point of view of credit risk management, the indicators mentioned in the above text should primarily be monitored long-term by company creditors, such as banks or debt-holders.

In this work, the key principles of rating and scoring evaluation, along with the challenges currently facing the rating industry, were clarified in the second chapter. Furthermore, the primary purpose of rating models and the main motivation of our research in the context of recent studies were emphasized. Following this, the main econometric approaches for rating and survival models employed in this research were outlined in chapter three. The remaining sections of the work focus on applications. Specifically, in the fourth chapter, selected methods were used to estimate rating models based on data from CEE countries. The most appropriate model was then utilised to identify variables with a significant impact on ratings. This was followed by the modelling of rating downgrades using survival analysis. While the first study focused on financial variables at a specific point in time, and the second study analyzed annual changes in these variables, both approaches yielded consistent results. Additionally, survival models allowed for the interpretation of the variables' influence on the hazard of a rating downgrade.

The subsequent chapter five focused on evaluating the relationship between credit ratings and corporate bankruptcy rates, with an exclusive emphasis on data from Czech companies. This analysis provided a more localized perspective, complementing the broader findings on ratings derived from CEE countries. The relationship was assessed using published default rates and calculated bankruptcy rates. As a result, the relationship between these quantities can be used to determine the rating through the proposed procedure with average spreads. In chapters six and seven, survival models were further applied to examine the influence of two different types of variables on survival probability. The key result was the identification of the effects of selected variables on corporate survival, as well as an analysis of how industry, legal form, and company size affect survival outcomes. Additionally, cumulative bankruptcy rates were estimated and converted into a rating assessment. A significant advantage of this approach is the incorporation of the time variable in survival models, which enables ratings to be dynamically assessed based not only on financial performance or other characteristics but also on the company’s age. This dynamic procedure offers a more comprehensive method for modelling and determining individual ratings.

Overall, the results of this work confirm the conclusions of previous studies mentioned in the text, which emphasize the importance of basic financial variables on rating or the probability of bankruptcy. The contribution of this monograph is mainly the verification of their influence using different approaches and two datasets, with a focus on CEE countries. The practical result is finding the relationship between default and bankruptcy rates and proposing a procedure to determine the rating using the bankruptcy rates. This procedure can be applied to specific time frames and covariate values. In general, it can be said that in terms of the methodology used, all applied approaches are suitable for modelling rating or corporate survival. Although some authors claim that these techniques are conventional compared to nonlinear classifiers, on the other hand, they emphasize their relatively good predictive ability and an easy way of applying and interpreting the models (De Servigny and Renault, 2004; Rerolle and Rimand, 2009; Jones et al. 2015). For this reason, they are particularly useful for retail investors. Furthermore, they can serve as a useful tool for obtaining basic information about the credit risk of the entity or debt instrument or signal an impending change in credit quality.

The models in this work are based on specific data from CEE countries, emphasising the Czech Republic. Therefore, they are suitable for companies from this region. However, thanks to certain standards in the financial performance reporting and accounting rules of companies, it can be said that they apply to any company from any country. Therefore, the following research could be devoted to other countries and companies and their mutual comparison. It may also be challenging to apply the models to companies with a certified rating and verify their compliance with agency ratings. To summarize, although the estimated models do not fully replace the certified rating or the professional assessment of the probability of bankruptcy, they help understand their main determinants, serve as primary and basic indicators, and signal any changes in credit quality.

Appendix

Appendix 1	List of authorized CRAs in the EU
Appendix 2	Linear discriminant models
Appendix 3	Logistic models
Appendix 4	ROC Curves
Appendix 5	Estimated CBRs
Appendix 6	Parametric models without covariates
Appendix 7	Parametric models with categorical covariates
Appendix 8	Parametric models with binary categorical variables
Appendix 9	PH assumptions (Cox model)
Appendix 10	Flexible models (Model 3)
Appendix 11	Combined models
Appendix 12	Summary of significant variables

Appendix 1

Name of CRA	Country of residence	Status	Effective date
Scope Hamburg GmbH (previously Euler	Germany	Registered	16 November
Japan Credit Rating Agency Ltd	Japan	Certified	6 January 2011
BCRA-Credit Rating Agency AD	Bulgaria	Registered	6 April 2011
Creditreform Rating AG	Germany	Registered	18 May 2011
Scope Ratings GmbH (previously Scope Ratings AG and PSR Rating GmbH)	Germany	Registered	24 May 2011
ICAP CRIF S.A. (previously ICAP S.A.)	Greece	Registered	7 July 2011
GBB-Rating Gesellschaft für Bonitätsbeurteilung GmbH	Germany	Registered	28 July 2011
ASSEKURATA	Germany	Registered	18 August 2011
ARC Ratings, S.A. (previously Companhia Portuguesa de Rating, S.A)	Portugal	Registered	26 August 2011
Fitch Ratings Ireland Limited	Ireland	Registered	31 October 2011
Moody’s Investors Service Cyprus Ltd	Cyprus	Registered	31 October 2011
Moody’s France S.A.S.	France	Registered	31 October 2011
Moody’s Deutschland GmbH	Germany	Registered	31 October 2011
Moody’s Italia S.r.l.	Italy	Registered	31 October 2011
Moody’s Investors Service España S.A.	Spain	Registered	31 October 2011
S&P Global Ratings Europe Limited	Ireland	Registered	31 October 2011
CRIF Ratings S.r.l. (previously CRIF S.p.a.)	Italy	Registered	22 December
Capital Intelligence Ratings Ltd	Cyprus	Registered	8 May 2012
EthiFinance Ratings, S.L. (previously Axesor Risk Management, S.L.)	Spain	Registered	1 October 2012
Cerved Rating Agency S.p.A. (previously CERVED Group S.p.A.)	Italy	Registered	20 December 2012
QIVALIO SAS (previously Spread Research)	France	Registered	1 July 2013
EuroRating Sp. z o.o.	Poland	Registered	7 May 2014
HR Ratings de México, S.A. de C.V. (HR Ratings)	Mexico	Certified	7 November 2014
Egan-Jones Ratings Co. (EJR)	USA	Certified	12 December
modeFinance S.r.l.	Italy	Registered	10 July 2015
Rating-Agentur Expert RA GmbH	Germany	Registered	1 December 2015
Kroll Bond Rating Agency Europe Limited	Ireland	Registered	13 November
Nordic Credit Rating AS	Norway	Registered	3 August 2018
Moody’s Investors Service (Nordics) AB	Sweden	Registered	13 August 2018
A.M. Best (EU) Rating Services B.V.	The Netherlands	Registered	3 December 2018
DBRS Ratings GmbH	Germany	Registered	14 December
Inbonis S.A.	Spain	Registered	27 May 2019

Source: ESMA (2022)

Appendix 2

Table 2-A Model 1 (LDA, non-random, 5 categories)

Estimation sample discrim lda
Summarized by rating

Mean	rating 1	2	3	4	5	Total
roa	2.193248	5.737577	9.03984	17.57556	32.41313	11.30486
roe	26.96803	25.546	21.72895	29.88464	49.30832	25.97489
eqta	.1710648	.3183587	.5073716	.6388264	.6978374	.4977553
lnta	8.868195	8.80307	8.865966	9.010979	9.012495	8.894253
lnintcov	.9280927	1.52686	2.231812	3.598151	4.724367	2.487936
lncf	6.072837	6.336812	6.683156	7.343005	7.83025	6.803402
lnliqr	-.8052876	-.46903	-.0527532	.3109041	.5774003	-.04643
lncurr	-.1297439	.0407714	.4244067	.7317776	.9046224	.4233209
lnltdta	-2.603659	-3.052501	-3.403263	-3.796831	-4.021998	-3.42774
ebitdar	10.48564	6.256535	3.38856	1.466098	.7597026	3.630798
N	117	718	1689	815	179	3518

Canonical linear discriminant analysis

Fcn	Canon. Corr.	Eigen-value	Variance Prop.	Variance Cumul.	Like-lihood Ratio	F	df1	df2	Prob>F
1	0.9117	4.92577	0.9572	0.9572	0.1376	228.3	40	1.3e+04	0.0000a
2	0.3986	.188913	0.0367	0.9939	0.8154	27.45	27	1.0e+04	0.0000a
3	0.1555	.024775	0.0048	0.9987	0.9694	6.8673	16	7012	0.0000 e
4	0.0812	.006645	0.0013	1.0000	0.9934	3.3292	7	3507	0.0015 e

Ho: this and smaller canon. corr. are zero; e = exact F, a = approximate F

Classification functions

	rating 1	2	3	4	5
roa	1.25951	1.449069	1.651068	2.013122	2.774612
roe	.1018749	.0860851	.0694353	.0345938	-.0261763
eqta	9.279019	22.28352	37.48646	51.47743	62.96465
lnta	34.21748	33.19485	31.79981	30.5193	32.58834
lnintcov	.3914805	1.05402	1.94155	3.327866	4.110825
lncf	-26.65267	-25.85311	-24.61532	-23.36355	-25.42252
lnliqr	-4.576533	-2.418458	-1.213412	.0994657	1.848741
lncurr	-.3274796	-1.61941	-.9197677	.0140922	.0924849
lnltdta	.4275029	.3186982	.253931	.1220489	.0150388
ebitdar	-.3804411	-.4615594	-.4653281	-.4115075	-.3934377
_cons	-77.2415	-73.99657	-77.95752	-93.31223	-126.6945
Priors	.0332575	.2040932	.4801023	.2316657	.0508812

Standardized canonical discriminant function coefficients

	function1	function2	function3	function4
roa	.891933	1.262051	.8017989	.266007
roe	-.3216331	-.4274077	-.0586236	.0641569
eqta	.728584	-.5228711	.3505561	.2142251
lnta	-.3680145	1.671466	1.726428	.7928341
lnintcov	.4201541	-.0856055	-.640141	-.2755164
lncf	.3627294	-1.682173	-2.106613	-.8114754
lnliqr	.2755986	-.0771478	.5207953	-.711704
lncurr	.0955224	-.0001297	-.7137809	1.039632
lnltdta	-.0644268	-.0112287	.0434867	.1917375
ebitdar	.0370822	.1530808	-.5950172	.335438

Canonical structure

	function1	function2	function3	function4
roa	.4377888	.6760415	-.0169238	-.2580644
roe	.0716345	.3998873	-.1506577	-.2455939
eqta	.4225021	-.6332764	.2953211	.3980326
lnta	.0277969	-.0006454	-.1823951	.0707038
lnintcov	.4115247	.1458912	-.5711483	-.4059625
lncf	.1572273	.0406031	-.2555312	-.1345257
lnliqr	.2878036	-.2507739	.1244314	.1502626
lncurr	.2763432	-.2869469	-.0584975	.6337375
lnltdta	-.0889352	.074549	-.0180928	.1657307
ebitdar	-.1730388	.3260823	-.3468995	.4125257

Table 2-B Model 2 (LDA, random , 5 categories)

Estimation sample discrim lda
Summarized by rating

Mean	rating 1	2	3	4	5	Total
roa	2.338636	5.679526	9.105874	17.49361	32.305	11.65017
roe	27.76989	25.59505	21.56647	29.87208	49.06022	25.92852
eqta	.1654249	.3193856	.5055008	.6394739	.6985521	.5080055
lnta	8.957169	8.810367	8.962073	9.105444	9.139335	8.980732
lnintcov	.9865285	1.553033	2.233993	3.68865	4.74235	2.574495
lncf	6.074221	6.322329	6.78123	7.43561	7.953954	6.907366
lnliqr	-.7464654	-.4694504	-.0861725	.3143515	.5734044	-.0354917
lncurr	-.0750167	.0403075	.4053312	.7357511	.8974123	.4377899
lnltdta	-2.681752	-3.120707	-3.459351	-3.812252	-4.095114	-3.501621
ebitdar	11.53334	6.358842	3.411175	1.480872	.7727544	3.514272
N	88	549	1670	785	182	3274

Canonical linear discriminant analysis

Fcn	Canon. Corr.	Eigen-value	Variance Prop.	Variance Cumul.	Like-lihood Ratio	F	df1	df2	Prob>F
1	0.9067	4.6185	0.9525	0.9525	0.1437	206.47	40	1.2e+04	0.0000a
2	0.3992	.189517	0.0391	0.9915	0.8074	26.817	27	9524	0.0000 a
3	0.1773	.032442	0.0067	0.9982	0.9604	8.3283	16	6524	0.0000 e
4	0.0921	.008547	0.0018	1.0000	0.9915	3.9843	7	3263	0.0002 e

Ho: this and smaller canon. corr. are zero; e = exact F, a = approximate F

Classification functions

	rating 1	2	3	4	5
roa	1.278251	1.425057	1.621704	1.95875	2.695607
roe	.1044083	.0905082	.0721828	.0389411	-.0205019
eqta	7.318283	20.92005	35.40541	49.02092	59.7408
lnta	34.84051	33.25814	31.83304	30.46373	32.55227
lnintcov	.2737411	.8805808	1.6952	3.135519	3.800604
lncf	-27.51177	-26.25614	-24.94467	-23.6325	-25.70231
lnliqr	-3.55494	-1.57155	-.4945931	.961885	2.89468
lncurr	-.1521573	-1.687478	-.9334868	.0340394	.0750888
lnltdta	.3653	.2443488	.1531677	.0209476	-.1280564
ebitdar	-.3478393	-.4515869	-.4589353	-.4078347	-.3918927
_cons	-78.61788	-73.0414	-76.52901	-91.25291	-123.3156
Priors	.0268784	.1676848	.5100794	.2397679	.0555895

Standardized canonical discriminant function coefficients

	function1	function2	function3	function4
roa	.8836248	1.262926	.7846844	-.2686077
roe	-.3129231	-.3715761	-.1166245	.0043921
eqta	.6982283	-.5232665	.3872112	-.173332
lnta	-.3870928	1.856544	1.313462	-.8974731
lnintcov	.4204638	-.0988689	-.7081107	.3326431
lncf	.3799422	-1.874078	-1.597985	.7095367
lnliqr	.28535	.0132871	.3439525	.5796318
lncurr	.0996098	-.0036004	-.6412883	-.9670562
lnltdta	-.0750212	-.0083899	-.0278363	-.0884735
ebitdar	.0342697	.1674534	-.5628519	-.3135336

Canonical structure

	function1	function2	function3	function4
roa	.4496794	.6492283	.0513664	.3029508
roe	.0841503	.4159143	-.164123	.3340838
eqta	.4155877	-.617556	.3016841	-.4590644
lnta	.0373212	-.0208471	-.1270363	-.2779526
lnintcov	.4256647	.1233813	-.625108	.464296
lncf	.1713345	-.0027022	-.1451953	-.0671213
lnliqr	.2854331	-.1913603	-.0461536	-.1729966
lncurr	.2731867	-.2433969	-.1657369	-.6398386
lnltdta	-.0855687	.0609827	-.0631585	-.0813899
ebitdar	-.1720863	.3552419	-.397548	-.3775538

Table 2-C Model 2 (LDA, non-random, 3 categories)

Estimation sample **discrim lda**
Summarized by rating

Mean	rating 2	3	4	Total
roa	5.737577	9.03984	17.57556	10.46305
roe	25.546	21.72895	29.88464	24.64252
eqta	.3183587	.5073716	.6388264	.4985027
lnta	8.80307	8.865966	9.010979	8.888631
lnintcov	1.52686	2.231812	3.598151	2.420332
lncf	6.336812	6.683156	7.343005	6.772884
lnliqr	-.46903	-.0527532	.3109041	-.053531
lncurr	.0407714	.4244067	.7317776	.4166653
lnltdta	-3.052501	-3.403263	-3.796831	-3.424651
ebitdar	6.256535	3.38856	1.466098	3.541384
N	718	1689	815	3222

Canonical linear discriminant analysis

Fcn	Canon. Corr.	Eigen-value	Variance Prop.	Cumul.	Like-lihood Ratio	F	df1	df2	Prob>F
1	0.8722	3.17954	0.9840	0.9840	0.2275	351.99	20	6420	0.0000 e
2	0.2216	.051651	0.0160	1.0000	0.9509	18.428	9	3211	0.0000 e

Ho: this and smaller canon. corr. are zero; e = exact F

Classification functions

	rating 2	3	4
roa	1.496691	1.739775	2.183524
roe	.1515258	.1214814	.0635478
eqta	26.83959	41.44539	54.60978
lnta	35.39423	34.01192	32.88522
lnintcov	.8594306	1.774678	3.183076
lncf	-28.09034	-26.8733	-25.80067
lnliqr	-2.192841	-1.006613	.2851679
lncurr	-2.54848	-1.729267	-.6344219
lnltdta	.3459011	.2694612	.1206118
ebitdar	-.4837224	-.4857704	-.4306812
_cons	-77.86736	-81.6769	-97.38644
Priors	.222843	.5242086	.2529485

Standardized canonical discriminant function coefficients

	function1	function2
roa	.8609197	.8750223
roe	-.4229165	-.4867342
eqta	.7124752	-.6473262
lnta	-.5430063	.6505696
lnintcov	.4478823	.259588
lncf	.5461345	-.5314011
lnliqr	.2389107	-.0879394
lncurr	.1613545	.0325695
lnltdta	-.0705606	-.0828867
ebitdar	.0574122	.2872596

Canonical structure

	function1	function2
roa	.3742801	.681001
roe	.0430037	.5174829
eqta	.4544354	-.7514998
lnta	.036613	.057458
lnintcov	.4184258	.4867864
lncf	.1601752	.1796464
lnliqr	.2998421	-.2993759
lncurr	.3053891	-.3809885
lnltdta	-.0912952	.0258934
ebitdar	-.1770874	.311453

Table 2-D Model 2 (LDA, random, 3 categories)

Estimation sample discrim lda
Summarized by rating

Mean	rating 2	3	4	Total
roa	5.679526	9.105874	17.49361	10.67155
roe	25.59505	21.56647	29.87208	24.47313
eqta	.3193856	.5055008	.6394739	.5064967
lnta	8.810367	8.962073	9.105444	8.971814
lnintcov	1.553033	2.233993	3.68865	2.489672
lncf	6.322329	6.78123	7.43561	6.868364
lnliqr	-.4694504	-.0861725	.3143515	-.0515547
lncurr	.0403075	.4053312	.7357511	.4249656
lnltdta	-3.120707	-3.459351	-3.812252	-3.489681
ebitdar	6.358842	3.411175	1.480872	3.445456
N	549	1670	785	3004

Canonical linear discriminant analysis

Fcn	Canon. Corr.	Eigen- value	Variance Prop.	Variance Cumul.	Like- lihood Ratio	F	df1	df2	Prob>F	
1	0.8657	2.99066	0.9848	0.9848	0.2395	312.13	20	5984	0.0000	e
2	0.2100	.04612	0.0152	1.0000	0.9559	15.337	9	2993	0.0000	e

Ho: this and smaller canon. corr. are zero; e = exact F

Classification functions

	rating 2	3	4
roa	1.480171	1.729483	2.156227
roe	.1718863	.1364475	.0776781
eqta	27.08754	40.90851	53.74412
lnta	35.58178	34.18606	32.98927
lnintcov	.5444646	1.403623	2.891621
lncf	-28.68608	-27.40949	-26.29257
lnliqr	-1.263114	-.1717268	1.307274
lncurr	-2.813459	-1.898183	-.7299488
lnltdta	.2933359	.1916541	.0465658
ebitdar	-.4824391	-.4882204	-.4358344
_cons	-77.16217	-80.55266	-95.84519
Priors	.1827563	.5559254	.2613182

Standardized canonical discriminant function coefficients

	function1	function2
roa	.8520154	.6723546
roe	-.4297179	-.3037564
eqta	.6733164	-.6427832
lnta	-.5626518	.6749223
lnintcov	.4620343	.3785943
lncf	.5654349	-.6431988
lnliqr	.2504184	.0397647
lncurr	.1740468	-.0019092
lnltdta	-.075902	-.0229589
ebitdar	.0558685	.3077855

Canonical structure

	function1	function2
roa	.3806158	.6245731
roe	.0526006	.5799675
eqta	.4452476	-.7754893
lnta	.0485494	-.0437633
lnintcov	.43691	.5828203
lncf	.1743345	.0524237
lnliqr	.3008696	-.17976
lncurr	.3061413	-.3162677
lnltdta	-.084406	.0511321
ebitdar	-.1725151	.3511672

Appendix 3

Table 3-A Model 5 (MLR, non-random, 5 categories)

Multinomial logistic regression	Number of obs = 3,518
	LR chi2(40) = 6830.33
	Prob > chi2 = 0.0000
Log likelihood = -1088.3968	Pseudo R2 = 0.7583

rating	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
1						
roa	-.8718716	.101053	-8.63	0.000	-1.069932	-.6738114
roe	.0288418	.0122541	2.35	0.019	.0048242	.0528595
eqta	-50.69209	3.193373	-15.87	0.000	-56.95099	-44.4332
lnta	2.337893	.5146432	4.54	0.000	1.329211	3.346575
lnintcov	-1.812987	.2155753	-8.41	0.000	-2.235507	-1.390467
lncf	-1.950915	.4760596	-4.10	0.000	-2.883974	-1.017855
lnliqr	-4.53426	.4636259	-9.78	0.000	-5.44295	-3.62557
lncurr	-3.676856	.6242642	-5.89	0.000	-4.900392	-2.453321
lnltdta	.3387546	.1127981	3.00	0.003	.1176744	.5598348
ebitdar	-.0629338	.0183886	-3.42	0.001	-.0989748	-.0268928
_cons	12.20383	2.055179	5.94	0.000	8.175753	16.23191
2						
roa	-.3488892	.0486066	-7.18	0.000	-.4441565	-.253622
roe	.0252744	.010932	2.31	0.021	.003848	.0467008
eqta	-25.19343	1.536032	-16.40	0.000	-28.204	-22.18287
lnta	1.966889	.3053352	6.44	0.000	1.368443	2.565335
lnintcov	-1.345666	.1179606	-11.41	0.000	-1.576865	-1.114468
lncf	-1.837547	.2916285	-6.30	0.000	-2.409128	-1.265966
lnliqr	-2.468488	.2977994	-8.29	0.000	-3.052164	-1.884812
lncurr	-2.622097	.3502034	-7.49	0.000	-3.308483	-1.935711
lnltdta	.1936708	.0590373	3.28	0.001	.0779598	.3093819
ebitdar	-.0588433	.0144689	-4.07	0.000	-.0872018	-.0304848
_cons	9.152363	1.190255	7.69	0.000	6.819505	11.48522
3	(base outcome)					
4						
roa	.4810173	.0487996	9.86	0.000	.3853718	.5766629
roe	-.1125707	.0168357	-6.69	0.000	-.145568	-.0795733
eqta	7.245488	1.630551	4.44	0.000	4.049668	10.44131
lnta	-1.174941	.5124099	-2.29	0.022	-2.179246	-.1706365
lnintcov	2.280726	.1549713	14.72	0.000	1.976988	2.584465
lncf	1.125697	.4989072	2.26	0.024	.1478573	2.103538
lnliqr	1.406972	.3169969	4.44	0.000	.7856697	2.028275
lncurr	.993776	.4028839	2.47	0.014	.2041381	1.783414
lnltdta	-.1992168	.0653998	-3.05	0.002	-.327398	-.0710355
ebitdar	-.9185144	.2204053	-4.17	0.000	-1.350501	-.4865279
_cons	-12.0212	1.660966	-7.24	0.000	-15.27664	-8.765771
5						
roa	.893902	.0752345	11.88	0.000	.7464451	1.041359
roe	-.2059876	.0237095	-8.69	0.000	-.2524573	-.1595179
eqta	1.704079	3.755006	0.45	0.650	-5.655597	9.063754
lnta	-1.012847	1.252292	-0.81	0.419	-3.467294	1.4416
lnintcov	4.318553	.284643	15.17	0.000	3.760663	4.876443
lncf	1.375477	1.241551	1.11	0.268	-1.057918	3.808872
lnliqr	2.762139	.6843357	4.04	0.000	1.420866	4.103412
lncurr	.4867153	.8284927	0.59	0.557	-1.137101	2.110531
lnltdta	-.3323168	.1274203	-2.61	0.009	-.5820561	-.0825775
ebitdar	-3.453901	.8734257	-3.95	0.000	-5.165784	-1.742018
_cons	-26.3962	4.02156	-6.56	0.000	-34.27832	-18.51409

Table 3-B Model 6 (MLR, random, 5 categories)

Multinomial logistic regression

Number of obs	=	3,274
LR chi2(40)	=	6135.13
Prob > chi2	=	0.0000
Pseudo R2	=	0.7537

Log likelihood = -1002.2227

rating	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
1						
roa	-.7598721	.1072226	-7.09	0.000	-.9700246	-.5497196
roe	.0257439	.0131376	1.96	0.050	-5.32e-06	.0514931
eqta	-47.96046	3.30881	-14.49	0.000	-54.44561	-41.47531
lnta	2.515261	.5573991	4.51	0.000	1.422779	3.607743
lnintcov	-1.687978	.2262245	-7.46	0.000	-2.131369	-1.244586
lncf	-2.206877	.515337	-4.28	0.000	-3.216919	-1.196835
lnliqr	-4.106548	.4760127	-8.63	0.000	-5.039516	-3.17358
lncurr	-3.636054	.6707748	-5.42	0.000	-4.950748	-2.321359
lnltdta	.2536894	.1192275	2.13	0.033	.0200077	.4873711
ebitdar	-.0607702	.0188982	-3.22	0.001	-.0978099	-.0237304
_cons	10.51605	2.245376	4.68	0.000	6.115199	14.91691
2						
roa	-.3223003	.0512918	-6.28	0.000	-.4228303	-.2217703
roe	.0247848	.0118769	2.09	0.037	.0015065	.048063
eqta	-24.04493	1.627909	-14.77	0.000	-27.23558	-20.85429
lnta	1.876271	.3304874	5.68	0.000	1.228528	2.524015
lnintcov	-1.275826	.1212193	-10.52	0.000	-1.513412	-1.038241
lncf	-1.808679	.3189729	-5.67	0.000	-2.433854	-1.183503
lnliqr	-2.226527	.304285	-7.32	0.000	-2.822915	-1.63014
lncurr	-2.822681	.3684012	-7.66	0.000	-3.544734	-2.100628
lnltdta	.1483109	.0621966	2.38	0.017	.0264078	.270214
ebitdar	-.0547336	.0159708	-3.43	0.001	-.0860358	-.0234313
_cons	8.635484	1.248496	6.92	0.000	6.188477	11.08249
3	(base outcome)					
4						
roa	.4605399	.0487104	9.45	0.000	.3650693	.5560106
roe	-.1018576	.0160608	-6.34	0.000	-.1333362	-.0703789
eqta	8.097681	1.679835	4.82	0.000	4.805264	11.3901
lnta	-1.425148	.5396985	-2.64	0.008	-2.482938	-.3673586
lnintcov	2.491839	.1733325	14.38	0.000	2.152113	2.831564
lncf	1.312629	.5232541	2.51	0.012	.2870694	2.338188
lnliqr	1.722581	.327991	5.25	0.000	1.079731	2.365432
lncurr	.9511114	.4107909	2.32	0.021	.1459761	1.756247
lnltdta	-.2159039	.0677846	-3.19	0.001	-.3487592	-.0830486
ebitdar	-1.067788	.2275411	-4.69	0.000	-1.51376	-.6218152
_cons	-12.10431	1.694033	-7.15	0.000	-15.42455	-8.784064
5						
roa	.8797732	.0740136	11.89	0.000	.7347092	1.024837
roe	-.194657	.0223564	-8.71	0.000	-.2384747	-.1508392
eqta	2.595909	3.664416	0.71	0.479	-4.586214	9.778033
lnta	-.4280212	1.207879	-0.35	0.723	-2.79542	1.939378
lnintcov	4.161439	.2675087	15.56	0.000	3.637132	4.685747
lncf	.6923759	1.185547	0.58	0.559	-1.631254	3.016005
lnliqr	3.394567	.6853617	4.95	0.000	2.051282	4.737851
lncurr	.4596287	.8285468	0.55	0.579	-1.164293	2.083551
lnltdta	-.4427664	.1248282	-3.55	0.000	-.6874251	-.1981077
ebitdar	-3.525994	.8148474	-4.33	0.000	-5.123065	-1.928922
_cons	-26.60738	3.870043	-6.88	0.000	-34.19252	-19.02224

Table 3-C Model 7 (OLR, non-random, 5 categories)

Ordered logistic regression

Number of obs	=	3,518
LR chi2(10)	=	6323.97
Prob > chi2	=	0.0000
Pseudo R2	=	0.7021

Log likelihood = -1341.5781

rating	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
roa	.4216598	.0186372	22.62	0.000	.3851317	.458188
roe	-.0461488	.0042403	-10.88	0.000	-.0544597	-.037838
eqta	15.4409	.6221944	24.82	0.000	14.22142	16.66038
lnta	-1.016424	.1651812	-6.15	0.000	-1.340173	-.692675
lnintcov	1.191601	.0606794	19.64	0.000	1.072671	1.31053
lncf	.9067322	.1547262	5.86	0.000	.6034743	1.20999
lnliqr	1.508619	.1539043	9.80	0.000	1.206972	1.810266
lncurr	.8399439	.1873313	4.48	0.000	.4727813	1.207106
lnltdta	-.1279205	.0332742	-3.84	0.000	-.1931366	-.0627043
ebitdar	.0175648	.0098324	1.79	0.074	-.0017063	.0368358
/cut1	-.0770474	.6232127			-1.298522	1.144427
/cut2	6.812134	.6365603			5.564499	8.05977
/cut3	15.85777	.7474834			14.39273	17.32281
/cut4	25.34595	.9478124			23.48827	27.20363

Table 3-D Model 8 (OLR, random, 5 categories)

Ordered logistic regression			Number of obs	=	3,274
			LR chi2(10)	=	5688.92
			Prob > chi2	=	0.0000
Log likelihood = -1225.3263			Pseudo R2	=	0.6989

rating	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
roa	.4193747	.0198184	21.16	0.000	.3805314	.458218
roe	-.0500249	.0054069	-9.25	0.000	-.0606222	-.0394276
eqta	14.68374	.6686125	21.96	0.000	13.37328	15.99419
lnta	-1.083233	.1730766	-6.26	0.000	-1.422456	-.7440088
lnintcov	1.186723	.0632337	18.77	0.000	1.062787	1.310659
lncf	.9885724	.1630977	6.06	0.000	.6689068	1.308238
lnliqr	1.537436	.1579884	9.73	0.000	1.227784	1.847087
lncurr	.8566806	.1933362	4.43	0.000	.4777487	1.235613
lnltdta	-.1423019	.0348389	-4.08	0.000	-.2105848	-.074019
ebitdar	.0170748	.0089245	1.91	0.056	-.0004169	.0345664
/cut1	-.4255852	.6639604			-1.726924	.8757533
/cut2	6.116953	.6669957			4.809665	7.42424
/cut3	15.34776	.7786476			13.82164	16.87388
/cut4	24.5734	.9775083			22.65752	26.48928

Table 3-E Model 9 (MLR, non-random, 3 categories)

Multinomial logistic regression			Number of obs	=	3,222
			LR chi2(20)	=	5068.91
			Prob > chi2	=	0.0000
Log likelihood = -754.60888			Pseudo R2	=	0.7706

rating	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
2						
roa	-.4663421	.0499148	-9.34	0.000	-.5641733	-.3685108
roe	.0534334	.008952	5.97	0.000	.0358878	.070979
eqta	-25.07236	1.500817	-16.71	0.000	-28.01391	-22.13082
lnta	2.116132	.3146352	6.73	0.000	1.499458	2.732806
lnintcov	-1.393496	.1229975	-11.33	0.000	-1.634567	-1.152425
lncf	-1.968311	.2996561	-6.57	0.000	-2.555626	-1.380996
lnliqr	-2.48541	.3071375	-8.09	0.000	-3.087389	-1.883432
lncurr	-2.950921	.3678114	-8.02	0.000	-3.671818	-2.230024
lnltdta	.2024991	.0609323	3.32	0.001	.083074	.3219243
ebitdar	-.0643623	.0128113	-5.02	0.000	-.089472	-.0392526
_cons	8.965982	1.194144	7.51	0.000	6.625503	11.30646
3	(base outcome)					
4						
roa	.8127013	.0699252	11.62	0.000	.6756504	.9497522
roe	-.2712486	.0275504	-9.85	0.000	-.3252464	-.2172509
eqta	.1292672	1.801071	0.07	0.943	-3.400768	3.659302
lnta	-1.331151	.5350695	-2.49	0.013	-2.379868	-.2824343
lnintcov	2.521068	.1770816	14.24	0.000	2.173995	2.868142
lncf	1.262611	.5210652	2.42	0.015	.2413424	2.283881
lnliqr	1.553258	.3308599	4.69	0.000	.9047846	2.201732
lncurr	1.495498	.4310791	3.47	0.001	.6505982	2.340397
lnltdta	-.2586712	.069951	-3.70	0.000	-.3957727	-.1215698
ebitdar	-.8746674	.2226558	-3.93	0.000	-1.311065	-.43827
_cons	-9.091825	1.712185	-5.31	0.000	-12.44765	-5.736005

Table 3-F Model 10 (MLR, random, 3 categories)

Multinomial logistic regression	Number of obs	=	3,004
	LR chi2(20)	=	4586.88
	Prob > chi2	=	0.0000
Log likelihood = -673.61807	Pseudo R2	=	0.7730

rating	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
2						
roa	-.5397647	.0630849	-8.56	0.000	-.6634087	-.4161206
roe	.0905741	.0148789	6.09	0.000	.061412	.1197361
eqta	-22.16638	1.585417	-13.98	0.000	-25.27374	-19.05902
lnta	2.115563	.343245	6.16	0.000	1.442816	2.788311
lnintcov	-1.378329	.1299323	-10.61	0.000	-1.632992	-1.123666
lncf	-2.028952	.3300618	-6.15	0.000	-2.675861	-1.382042
lnliqr	-2.310107	.3208395	-7.20	0.000	-2.938941	-1.681274
lncurr	-3.289145	.396834	-8.29	0.000	-4.066925	-2.511364
lnltdta	.1609138	.0655593	2.45	0.014	.03242	.2894077
ebitdar	-.0621959	.0138194	-4.50	0.000	-.0892814	-.0351105
_cons	7.516076	1.25711	5.98	0.000	5.052186	9.979965
3	(base outcome)					
4						
roa	.8168897	.0739023	11.05	0.000	.6720439	.9617354
roe	-.2782015	.0292878	-9.50	0.000	-.3356046	-.2207984
eqta	-.2189209	1.910185	-0.11	0.909	-3.962815	3.524973
lnta	-1.659557	.5744295	-2.89	0.004	-2.785418	-.5336956
lnintcov	2.74301	.1970567	13.92	0.000	2.356785	3.129234
lncf	1.496148	.5564281	2.69	0.007	.405569	2.586727
lnliqr	1.94331	.3469483	5.60	0.000	1.263304	2.623316
lncurr	1.470273	.4440946	3.31	0.001	.5998638	2.340683
lnltdta	-.2707047	.0726431	-3.73	0.000	-.4130825	-.1283269
ebitdar	-1.070727	.2365387	-4.53	0.000	-1.534335	-.6071197
_cons	-7.932538	1.728754	-4.59	0.000	-11.32083	-4.544243

Table 3-G Model 11 (OLR, non-random, 3 categories)

Ordered logistic regression	Number of obs	=	3,222
	LR chi2(10)	=	4765.17
	Prob > chi2	=	0.0000
Log likelihood = -906.48046	Pseudo R2	=	0.7244

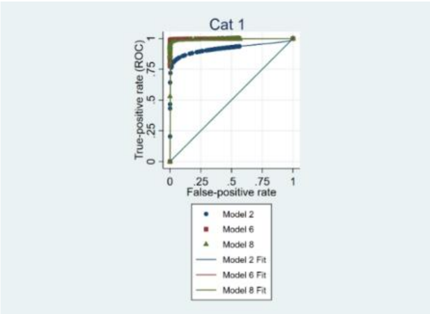
rating	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
roa	.5972161	.032393	18.44	0.000	.5337269	.6607052
roe	-.117428	.0093734	-12.53	0.000	-.1357996	-.0990564
eqta	14.45239	.80132	18.04	0.000	12.88183	16.02295
lnta	-1.610103	.2129214	-7.56	0.000	-2.027421	-1.192784
lnintcov	1.428202	.0799767	17.86	0.000	1.27145	1.584953
lncf	1.443458	.2008166	7.19	0.000	1.049865	1.837052
lnliqr	1.660559	.199447	8.33	0.000	1.26965	2.051468
lncurr	1.619918	.2405473	6.73	0.000	1.148454	2.091382
lnltdta	-.1772948	.0410855	-4.32	0.000	-.2578209	-.0967687
ebitdar	.0439109	.0125309	3.50	0.000	.0193508	.068471
/cut1	5.11407	.7823995			3.580595	6.647545
/cut2	15.64529	.9255537			13.83124	17.45934

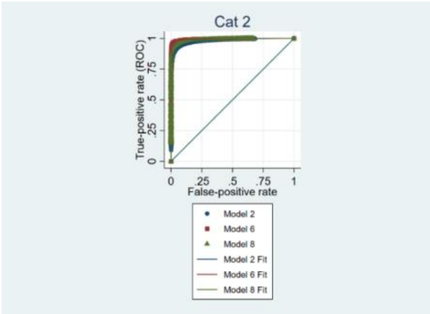
Table 3-H Model 12 (OLR, random, 3 categories)

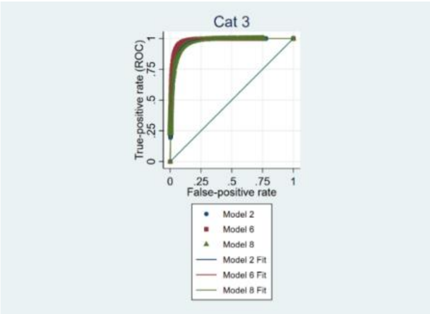
Ordered logistic regression	Number of obs	=	3,004
	LR chi2(10)	=	4331.91
	Prob > chi2	=	0.0000
Log likelihood = -801.09988	Pseudo R2	=	0.7300

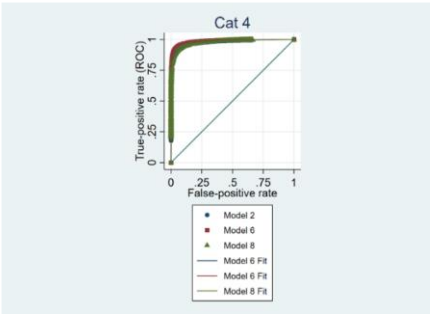
rating	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
roa	.6562592	.0359928	18.23	0.000	.5857147	.7268038
roe	-.1432285	.0103027	-13.90	0.000	-.1634215	-.1230355
eqta	13.36132	.8323957	16.05	0.000	11.72985	14.99278
lnta	-1.67366	.2272175	-7.37	0.000	-2.118998	-1.228322
lnintcov	1.493245	.0862979	17.30	0.000	1.324104	1.662385
lncf	1.506382	.2156548	6.99	0.000	1.083707	1.929058
lnliqr	1.772456	.2081745	8.51	0.000	1.364442	2.180471
lncurr	1.734623	.2531471	6.85	0.000	1.238464	2.230782
lnltdta	-.187392	.0439263	-4.27	0.000	-.273486	-.101298
ebitdar	.044036	.0132251	3.33	0.001	.0181153	.0699567
/cut1	4.229193	.8162787			2.629316	5.82907
/cut2	15.37176	.9674602			13.47558	17.26795

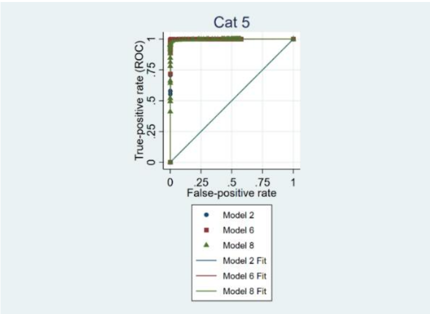
Appendix 4


Cat 1
True-positive rate (ROC)
0 .25 .5 .75 1
False-positive rate
Model 2
Model 6
Model 8
Model 2 Fit
Model 6 Fit
Model 8 Fit


Cat 2
True-positive rate (ROC)
0 .25 .5 .75 1
False-positive rate
Model 2
Model 6
Model 8
Model 2 Fit
Model 6 Fit
Model 8 Fit


Cat 3
True-positive rate (ROC)
0 .25 .5 .75 1
False-positive rate
Model 2
Model 6
Model 8
Model 2 Fit
Model 6 Fit
Model 8 Fit


Cat 4
True-positive rate (ROC)
0 .25 .5 .75 1
False-positive rate
Model 2
Model 6
Model 8
Model 6 Fit
Model 6 Fit
Model 8 Fit


Cat 5
True-positive rate (ROC)
0 .25 .5 .75 1
False-positive rate
Model 2
Model 6
Model 8
Model 2 Fit
Model 6 Fit
Model 8 Fit

Appendix 5

Time (years)	1	2	3	4
	Sector			
1	.0058	.0099	.0076	.0000
2	.0101	.0192	.0094	.0030
3	.0144	.0278	.0117	.0090
4	.0187	.0358	.0131	.0090
5	.0216	.0450	.0145	.0090
6	.0262	.0569	.0168	.0150
7	.0308	.0706	.0200	.0242
8	.0354	.0819	.0224	.0321
9	.0397	.0927	.0243	.0369
10	.0448	.1050	.0275	.0402
11	.0492	.1148	.0291	.0402
12	.0541	.1264	.0315	.0437
13	.0605	.1432	.0405	.0492
14	.0664	.1692	.0480	.0513
15	.0707	.1931	.0534	.0534
Time (years)	**Legal**			
1	.0031	.0306	.0071	.0000
2	.0044	.0306	.0138	.0000
3	.0071	.0306	.0201	.0000
4	.0102	.0306	.0257	.0000
5	.0161	.0306	.0304	.0000
6	.0243	.0327	.0373	.0000
7	.0340	.0327	.0452	.0000
8	.0430	.0370	.0516	.0000
9	.0507	.0370	.0579	.0000
10	.0618	.0402	.0647	.0000
11	.0714	.0402	.0699	.0000
12	.0777	.0435	.0768	.0071
13	.0884	.0501	.0873	.0071
14	.1024	.0569	.1001	.0229
15	.1087	.0818	.1118	.0229
Time (years)	**Size**			
1	.0053	.0041	.0023	.0000
2	.0110	.0083	.0053	.0032
3	.015	.0143	.0082	.0064
4	.0182	.0198	.0105	.0064
5	.0223	.0253	.0159	.0097
6	.0277	.0330	.0212	.0196
7	.0340	.0424	.0260	.0295
8	.0392	.0504	.0333	.0329
9	.0455	.0553	.0419	.0397
10	.0503	.0619	.0526	.0500
11	.0565	.0656	.0574	.0575
12	.0630	.0711	.0654	.0656
13	.0702	.0835	.0744	.0783
14	.0790	.0984	.0895	.1014
15	.0838	.1135	.1021	.1258

Appendix 6

Table 6-A Weibull model

		Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb							
	_rcs1	.6303861	.054594	11.55	0.000	.5233837	.7373884
	_cons	-7.396833	.4703291	-15.73	0.000	-8.318661	-6.475005

Table 6-B PH(2) model

		Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb							
	_rcs1	1.123976	.2261959	4.97	0.000	.6806403	1.567312
	_rcs2	.0207503	.0087744	2.36	0.018	.0035528	.0379477
	_cons	-9.860817	1.226673	-8.04	0.000	-12.26505	-7.456582

Table 6-C PH(3) model

		Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb							
	_rcs1	1.119542	.3297984	3.39	0.001	.4731489	1.765935
	_rcs2	.0064808	.0384585	0.17	0.866	-.0688965	.0818582
	_rcs3	.0183107	.0578292	0.32	0.752	-.0950324	.1316537
	_cons	-9.840718	1.570866	-6.26	0.000	-12.91956	-6.761876

Table 6-D PH(4) model

		Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb							
	_rcs1	1.159798	.4101065	2.83	0.005	.3560041	1.963592
	_rcs2	.0207709	.0707393	0.29	0.769	-.1178755	.1594173
	_rcs3	-.0450502	.1718981	-0.26	0.793	-.3819641	.2918638
	_rcs4	.0652372	.1586738	0.41	0.681	-.2457577	.3762322
	_cons	-9.983716	1.822385	-5.48	0.000	-13.55552	-6.411908

Table 6-E PH(5) model

		Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb							
	_rcs1	.9584063	.4391262	2.18	0.029	.0977348	1.819078
	_rcs2	-.091766	.0982927	-0.93	0.351	-.2844162	.1008843
	_rcs3	.3368306	.2457342	1.37	0.170	-.1447996	.8184609
	_rcs4	-.5339933	.3357616	-1.59	0.112	-1.192074	.1240873
	_rcs5	.4689203	.2888055	1.62	0.104	-.097128	1.034969
	_cons	-9.336677	1.846062	-5.06	0.000	-12.95489	-5.718462

Table 6-F PH(6) model

		Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb							
	_rcs1	1.097961	.4943391	2.22	0.026	.1290745	2.066848
	_rcs2	.0322903	.1490014	0.22	0.828	-.2597471	.3243276
	_rcs3	-.2612932	.4017819	-0.65	0.515	-1.048771	.5261847
	_rcs4	.8619785	.5716172	1.51	0.132	-.2583707	1.982328
	_rcs5	-1.581232	.6723966	-2.35	0.019	-2.899105	-.2633593
	_rcs6	1.420608	.5462602	2.60	0.009	.3499581	2.491259
	_cons	-9.769841	2.023879	-4.83	0.000	-13.73657	-5.80311

Appendix 7

Table 7-A Weibull model

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
industry						
2	.5533101	.2567024	2.16	0.031	.0501826	1.056438
3	-.6492085	.556748	-1.17	0.244	-1.740414	.4419975
4	-.8244569	1.020887	-0.81	0.419	-2.825359	1.176445
legal						
2	.3152413	1.087847	0.29	0.772	-1.8169	2.447382
3	-.0242182	.3647704	-0.07	0.947	-.7391551	.6907186
4	.2074794	1.059841	0.20	0.845	-1.869772	2.284731
size						
2	-.1673749	.2770986	-0.60	0.546	-.7104782	.3757283
3	-.0193697	.4019965	-0.05	0.962	-.8072684	.768529
4	.3439641	.7547141	0.46	0.649	-1.135248	1.823177
_rcs1	.6414438	.0686601	9.34	0.000	.5068725	.7760152
_cons	-7.839775	.6856906	-11.43	0.000	-9.183704	-6.495847

Table 7-B PH(2) model

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
industry						
2	.5605202	.2567081	2.18	0.029	.0573817	1.063659
3	-.627071	.5560319	-1.13	0.259	-1.716873	.4627314
4	-.8017734	1.021036	-0.79	0.432	-2.802967	1.19942
legal						
1	9.93e-17	1.56e-16	0.64	0.524	-2.06e-16	4.05e-16
2	.3506135	1.087281	0.32	0.747	-1.780419	2.481646
3	-.0115836	.3652604	-0.03	0.975	-.7274808	.7043137
4	.2504851	1.060487	0.24	0.813	-1.828032	2.329002
size						
1	-2.68e-16	4.58e-17	-5.86	0.000	-3.58e-16	-1.78e-16
2	-.1518909	.2773388	-0.55	0.584	-.695465	.3916833
3	.0161797	.4023574	0.04	0.968	-.7724263	.8047858
4	.3721553	.7560634	0.49	0.623	-1.109702	1.854012
_rcs1	1.021153	.2728456	3.74	0.000	.4863852	1.55592
_rcs2	.0164403	.0109728	1.50	0.134	-.005066	.0379466
_cons	-9.767045	1.535906	-6.36	0.000	-12.77736	-6.756725

Table 7-C PH(3) model

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
industry						
2	.5645973	.2568368	2.20	0.028	.0612065	1.067988
3	-.6203298	.5557931	-1.12	0.264	-1.709664	.4690047
4	-.7980476	1.021061	-0.78	0.434	-2.799291	1.203196
legal						
1	2.88e-16	1.36e-16	2.12	0.034	2.16e-17	5.55e-16
2	.3636483	1.087008	0.33	0.738	-1.766849	2.494146
3	-.0062898	.3653078	-0.02	0.986	-.72228	.7097004
4	.2602803	1.060565	0.25	0.806	-1.81839	2.33895
size						
1	-1.14e-16	9.96e-17	-1.15	0.252	-3.09e-16	8.10e-17
2	-.1495619	.2774158	-0.54	0.590	-.6932869	.394163
3	.0257415	.4024987	0.06	0.949	-.7631415	.8146246
4	.382455	.7562129	0.51	0.613	-1.099695	1.864605
_rcs1	.8827026	.3627799	2.43	0.015	.1716672	1.593738
_rcs2	-.0208976	.0483006	-0.43	0.665	-.1155651	.0737699
_rcs3	.0548229	.0753289	0.73	0.467	-.0928189	.2024648
_cons	-9.230525	1.766292	-5.23	0.000	-12.69239	-5.768657

Table 7-D PH(4) model

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
industry						
2	.5685293	.2569087	2.21	0.027	.0649975	1.072061
3	-.6157176	.5554996	-1.11	0.268	-1.704477	.4730415
4	-.7965578	1.021063	-0.78	0.435	-2.797804	1.204688
legal						
1	6.26e-16	2.61e-16	2.40	0.017	1.14e-16	1.14e-15
2	.3808806	1.086957	0.35	0.726	-1.749515	2.511276
3	-.0014986	.3652457	-0.00	0.997	-.717367	.7143697
4	.2689628	1.060525	0.25	0.800	-1.809627	2.347553
size						
1	-2.31e-16	1.61e-16	-1.43	0.151	-5.48e-16	8.48e-17
2	-.1473228	.2774211	-0.53	0.595	-.6910581	.3964126
3	.0360149	.4026007	0.09	0.929	-.7530679	.8250978
4	.3923771	.7560129	0.52	0.604	-1.089381	1.874135
_rcs1	.9347136	.4268978	2.19	0.029	.0980094	1.771418
_rcs2	.0268203	.0942173	0.28	0.776	-.1578421	.2114828
_rcs3	-.1184025	.2285919	-0.52	0.604	-.5664343	.3296293
_rcs4	.1684749	.2153678	0.78	0.434	-.2536382	.590588
_cons	-9.419546	1.962328	-4.80	0.000	-13.26564	-5.573454

Table 7-E PH(5) model

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
industry						
2	.5771588	.256863	2.25	0.025	.0737165	1.080601
3	-.6070248	.5543444	-1.10	0.274	-1.69352	.4794704
4	-.7864262	1.02117	-0.77	0.441	-2.787882	1.21503
legal						
1	8.61e-18	1.09e-16	0.08	0.937	-2.04e-16	2.21e-16
2	.4415604	1.086613	0.41	0.684	-1.688161	2.571282
3	.0164306	.3651587	0.04	0.964	-.6992673	.7321284
4	.3085055	1.060927	0.29	0.771	-1.770874	2.387885
size						
1	-5.34e-16	9.30e-17	-5.74	0.000	-7.16e-16	-3.51e-16
2	-.1357144	.2772098	-0.49	0.624	-.6790357	.4076068
3	.0788501	.4021879	0.20	0.845	-.7094238	.8671239
4	.4351383	.7559862	0.58	0.565	-1.046567	1.916844
_rcs1	.6763176	.3993425	1.69	0.090	-.1063792	1.459015
_rcs2	-.170092	.11158	-1.52	0.127	-.3887847	.0486008
_rcs3	.5173931	.260648	1.99	0.047	.0065324	1.028254
_rcs4	-1.251378	.5135865	-2.44	0.015	-2.257989	-.2447669
_rcs5	1.346933	.533249	2.53	0.012	.3017841	2.392082
_cons	-8.691531	1.773837	-4.90	0.000	-12.16819	-5.214875

Table 7-F PH(6) model

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
industry						
2	.5824482	.2568803	2.27	0.023	.078972	1.085924
3	-.6026824	.5539108	-1.09	0.277	-1.688328	.4829628
4	-.7873736	1.021146	-0.77	0.441	-2.788784	1.214036
legal						
1	1.84e-16	1.85e-16	1.00	0.318	-1.78e-16	5.46e-16
2	.4422316	1.085762	0.41	0.684	-1.685822	2.570285
3	.0193468	.3649031	0.05	0.958	-.6958502	.7345438
4	.3177369	1.061057	0.30	0.765	-1.761897	2.39737
size						
1	-2.14e-16	4.67e-17	-4.59	0.000	-3.06e-16	-1.23e-16
2	-.1351025	.277197	-0.49	0.626	-.6783986	.4081936
3	.0880003	.4015291	0.22	0.827	-.6989822	.8749828
4	.4482701	.7557245	0.59	0.553	-1.032923	1.929463
_rcs1	.7032258	.4268019	1.65	0.099	-.1332905	1.539742
_rcs2	-.0857021	.1521876	-0.56	0.573	-.3839843	.2125801
_rcs3	.0134819	.4493236	0.03	0.976	-.8671762	.8941399
_rcs4	.7233714	.7048382	1.03	0.305	-.6580862	2.104829
_rcs5	-1.759777	.8055432	-2.18	0.029	-3.338612	-.1809412
_rcs6	1.795005	.6311543	2.84	0.004	.5579651	3.032044
_cons	-8.745298	1.830364	-4.78	0.000	-12.33275	-5.157851

Appendix 8

Table 8-A Weibull model

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
2.industry_b	-.6377852	.0653733	-9.76	0.000	-.7659145	-.5096559
2.legal_b	.7599181	.2117105	3.59	0.000	.3449732	1.174863
2.size_b	.2914748	.0633654	4.60	0.000	.1672809	.4156687
_rcs1	.9463965	.0266963	35.45	0.000	.8940726	.9987203
_cons	-11.21048	.3170751	-35.36	0.000	-11.83194	-10.58903

Table 8-B PH(2) model

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
2.industry_b	-.642346	.065378	-9.83	0.000	-.7704845	-.5142074
2.legal_b	.7555694	.2117148	3.57	0.000	.340616	1.170523
1.size_b	2.82e-43	1.36e-44	20.80	0.000	2.56e-43	3.09e-43
2.size_b	.2962976	.0633729	4.68	0.000	.172089	.4205062
_rcs1	1.389318	.133176	10.43	0.000	1.128298	1.650338
_rcs2	.0259917	.0075086	3.46	0.001	.011275	.0407083
_cons	-13.65343	.7938176	-17.20	0.000	-15.20928	-12.09758

Table 8-C PH(3) model

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
2.industry_b	-.6495686	.0653753	-9.94	0.000	-.7777019	-.5214354
1.legal_b	-2.35e-15	1.33e-16	-17.69	0.000	-2.61e-15	-2.09e-15
2.legal_b	.7431579	.2117247	3.51	0.000	.3281852	1.158131
1.size_b	2.45e-42	1.82e-43	13.44	0.000	2.09e-42	2.81e-42
2.size_b	.3052491	.0633792	4.82	0.000	.1810281	.4294701
_rcs1	.866513	.1489569	5.82	0.000	.5745629	1.158463
_rcs2	-.2250525	.0435701	-5.17	0.000	-.3104484	-.1396567
_rcs3	.4111053	.0731158	5.62	0.000	.267801	.5544097
_cons	-11.26942	.8173266	-13.79	0.000	-12.87135	-9.667485

Table 8-D PH(4) model

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
2.industry_b	-.6571706	.0653553	-10.06	0.000	-.7852645	-.5290766
1.legal_b	2.28e-16	4.44e-17	5.14	0.000	1.41e-16	3.15e-16
2.legal_b	.7010793	.2117528	3.31	0.001	.2860515	1.116107
1.size_b	-7.81e-44	6.47e-44	-1.21	0.228	-2.05e-43	4.88e-44
2.size_b	.3147632	.0633597	4.97	0.000	.1905805	.4389459
_rcs1	1.587449	.2051144	7.74	0.000	1.185432	1.989466
_rcs2	.5443479	.0739922	7.36	0.000	.399326	.6893699
_rcs3	-2.347627	.2405595	-9.76	0.000	-2.819115	-1.876139
_rcs4	2.698232	.2461822	10.96	0.000	2.215724	3.18074
_cons	-14.3895	1.069251	-13.46	0.000	-16.4852	-12.29381

Table 8-E PH(5) model

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
2.industry_b	-.6565548	.0653466	-10.05	0.000	-.7846317	-.5284778
2.legal_b	.6853032	.2117234	3.24	0.001	.2703331	1.100273
1.size_b	5.30e-42	4.57e-43	11.60	0.000	4.40e-42	6.19e-42
2.size_b	.3144274	.0633457	4.96	0.000	.1902721	.4385826
_rcs1	1.485734	.2328253	6.38	0.000	1.029404	1.942063
_rcs2	.0948812	.0691736	1.37	0.170	-.0406966	.230459
_rcs3	.84295	.3289706	2.56	0.010	.1981794	1.48772
_rcs4	-4.330172	.5720056	-7.57	0.000	-5.451282	-3.209061
_rcs5	4.729204	.3890263	12.16	0.000	3.966726	5.491682
_cons	-13.92153	1.167335	-11.93	0.000	-16.20947	-11.6336

Table 8-F PH(6) model

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
2.industry_b	-.6540634	.0653519	-10.01	0.000	-.7821508	-.5259761
1.legal_b	-1.28e-15	9.29e-17	-13.81	0.000	-1.47e-15	-1.10e-15
2.legal_b	.6879783	.211714	3.25	0.001	.2730265	1.10293
1.size_b	2.80e-42	1.58e-43	17.68	0.000	2.49e-42	3.11e-42
2.size_b	.3118251	.0633495	4.92	0.000	.1876622	.4359879
_rcs1	1.827936	.2741193	6.67	0.000	1.290672	2.3652
_rcs2	.3229867	.0804398	4.02	0.000	.1653277	.4806457
_rcs3	-1.604773	.4085616	-3.93	0.000	-2.405539	-.8040067
_rcs4	4.434056	.8605027	5.15	0.000	2.747502	6.12061
_rcs5	-9.455195	1.077831	-8.77	0.000	-11.56771	-7.342685
_rcs6	8.00536	.6484428	12.35	0.000	6.734435	9.276284
_cons	-15.37552	1.343584	-11.44	0.000	-18.0089	-12.74215

Table 8-G Information criteria

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
PHB1	92,933	.	-4847.847	5	9705.694	9752.893
PHB2	92,933	.	-4841.279	6	9694.558	9751.196
PHB3	92,933	.	-4826.325	7	9666.651	9732.728
PHB4	92,933	.	-4779.259	8	9574.518	9650.036
PHB5	92,933	.	-4750.835	9	9519.67	9604.627
PHB6	92,933	.	-4724.008	10	9468.017	9562.413

Appendix 9

Figure 9-A Model 1

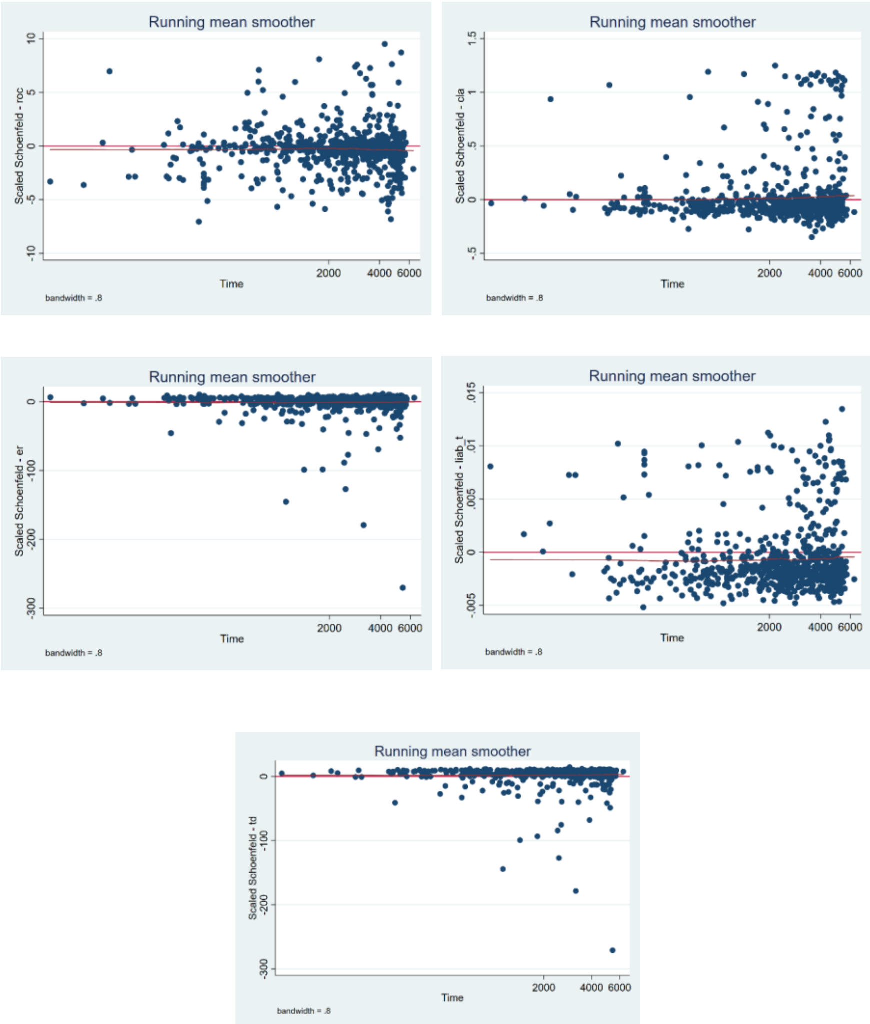
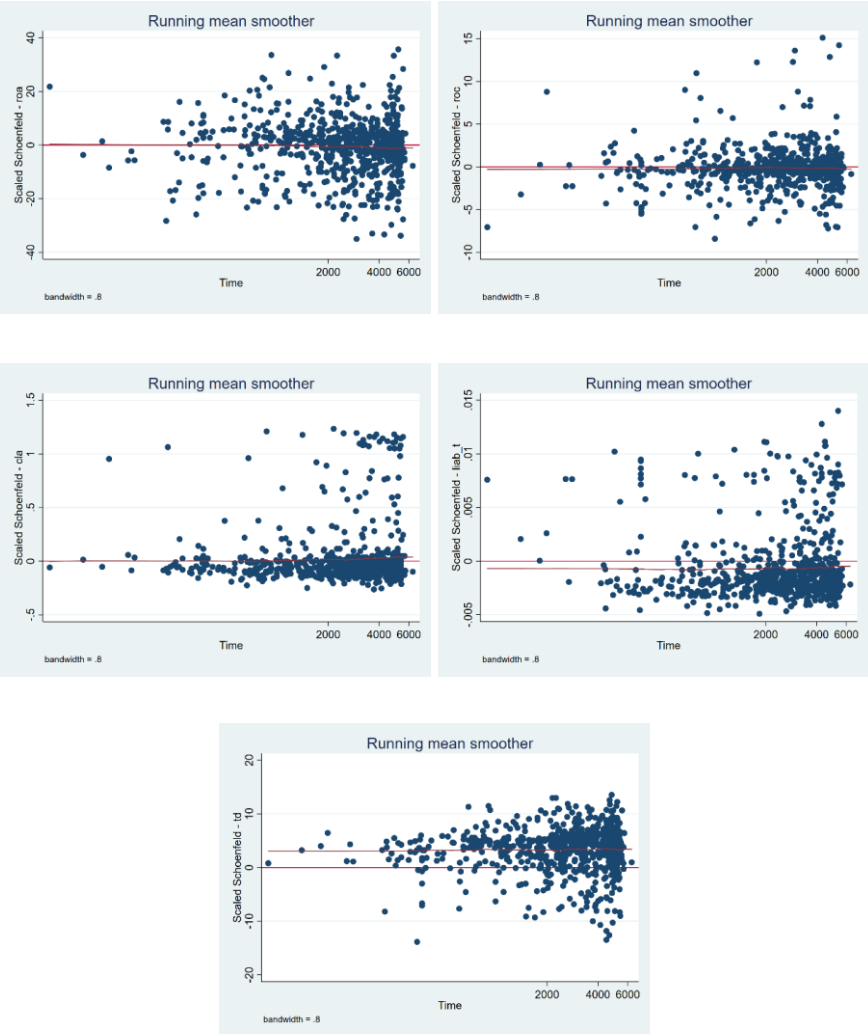


Figure 9-B Model 2



Appendix 10

Table 10-A Weibull model

Log likelihood = -2626.081 Number of obs = 62,107

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
roa	-.8511133	.4344592	-1.96	0.050	-1.702638	.0004112
roc	-.1906631	.1047644	-1.82	0.069	-.3959975	.0146713
cla	.0300309	.0102899	2.92	0.004	.009863	.0501989
er	-1.488546	.6628031	-2.25	0.025	-2.787616	-.1894757
liab_t	-.0005968	.000128	-4.66	0.000	-.0008476	-.000346
td	1.933596	.6646117	2.91	0.004	.6309809	3.236211
cr	-.0833644	.02248	-3.71	0.000	-.1274244	-.0393044
lnta	.153397	.0255173	6.01	0.000	.103384	.20341
rec_to_ca	.9821094	.1506011	6.52	0.000	.6869366	1.277282
_rcs1	1.281273	.0442437	28.96	0.000	1.194557	1.367989
_cons	-16.26865	.7841077	-20.75	0.000	-17.80548	-14.73183

Table 10-B PH(2) model

Log likelihood = -2625.9749 Number of obs = 62,107

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
roa	-.8562889	.4346178	-1.97	0.049	-1.708124	-.0044538
roc	-.1905983	.10474	-1.82	0.069	-.3958851	.0146884
cla	.0299755	.0102911	2.91	0.004	.0098054	.0501456
er	-1.484411	.662656	-2.24	0.025	-2.783193	-.1856289
liab_t	-.0005959	.000128	-4.66	0.000	-.0008467	-.0003451
td	1.929518	.6644513	2.90	0.004	.6272177	3.231819
cr	-.0833463	.0224793	-3.71	0.000	-.1274049	-.0392877
lnta	.1539258	.0255449	6.03	0.000	.1038589	.2039928
rec_to_ca	.9819837	.1506102	6.52	0.000	.6867931	1.277174
_rcs1	1.350585	.1584752	8.52	0.000	1.03998	1.661191
_rcs2	.0091224	.0199399	0.46	0.647	-.0299591	.0482038
_cons	-16.7124	1.25153	-13.35	0.000	-19.16536	-14.25945

Table 10-C PH(3) model

Log likelihood = -2625.7769 Number of obs = 62,107

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
roa	-.8663674	.434922	-1.99	0.046	-1.718799	-.013936
roc	-.1903867	.104713	-1.82	0.069	-.3956204	.0148471
cla	.0299044	.0102907	2.91	0.004	.009735	.0500739
er	-1.478196	.6625536	-2.23	0.026	-2.776777	-.179615
liab_t	-.0005956	.0001279	-4.65	0.000	-.0008463	-.0003448
td	1.925954	.6642939	2.90	0.004	.6239619	3.227946
cr	-.083332	.0224774	-3.71	0.000	-.1273869	-.039277
lnta	.1544585	.0255601	6.04	0.000	.1043616	.2045555
rec_to_ca	.981375	.1506307	6.52	0.000	.6861442	1.276606
_rcs1	1.293613	.1941758	6.66	0.000	.9130358	1.674191
_rcs2	-.052917	.097766	-0.54	0.588	-.2445348	.1387008
_rcs3	.120388	.1931857	0.62	0.533	-.2582489	.499025
_cons	-16.40388	1.394969	-11.76	0.000	-19.13797	-13.6698

Table 10-D PH(4) model

Log likelihood = -2616.5314 Number of obs = 62,107

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
roa	-.9253601	.435379	-2.13	0.034	-1.778687	-.0720329
roc	-.1860006	.1047181	-1.78	0.076	-.3912444	.0192431
cla	.0296415	.0102828	2.88	0.004	.0094876	.0497953
er	-1.446617	.6637939	-2.18	0.029	-2.747629	-.1456049
liab_t	-.000595	.0001279	-4.65	0.000	-.0008456	-.0003443
td	1.919176	.6655875	2.88	0.004	.6146483	3.223703
cr	-.0830469	.0224457	-3.70	0.000	-.1270397	-.039054
lnta	.1561533	.0255986	6.10	0.000	.105981	.2063256
rec_to_ca	.9792462	.1507234	6.50	0.000	.6838338	1.274659
_rcs1	1.682805	.2439505	6.90	0.000	1.204671	2.16094
_rcs2	.621904	.1650157	3.77	0.000	.2984791	.9453289
_rcs3	-2.305252	.5497408	-4.19	0.000	-3.382724	-1.22778
_rcs4	2.928073	.6967917	4.20	0.000	1.562387	4.29376
_cons	-18.55789	1.638807	-11.32	0.000	-21.76989	-15.34589

Table 10-E PH(5) model

Log likelihood = -2600.7356 Number of obs = 62,107

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
roa	-1.010709	.4361879	-2.32	0.020	-1.865621	-.1557964
roc	-.1775339	.104684	-1.70	0.090	-.3827107	.0276429
cla	.0294498	.010276	2.87	0.004	.0093093	.0495904
er	-1.408876	.6651313	-2.12	0.034	-2.71251	-.1052429
liab_t	-.0005888	.0001278	-4.61	0.000	-.0008393	-.0003383
td	1.897453	.6669352	2.85	0.004	.5902836	3.204622
cr	-.0827926	.0223882	-3.70	0.000	-.1266727	-.0389125
lnta	.157765	.0256295	6.16	0.000	.107532	.207998
rec_to_ca	.9814311	.1507887	6.51	0.000	.6858906	1.276971
_rcs1	1.605969	.2695638	5.96	0.000	1.077634	2.134304
_rcs2	.0368831	.1732406	0.21	0.831	-.3026624	.3764285
_rcs3	2.145199	.7672627	2.80	0.005	.6413919	3.649007
_rcs4	-8.264959	1.59505	-5.18	0.000	-11.3912	-5.138717
_rcs5	10.12787	1.612526	6.28	0.000	6.967378	13.28836
_cons	-18.17845	1.747584	-10.40	0.000	-21.60365	-14.75325

Table 10-F PH(6) model

Log likelihood = -2566.0414 Number of obs = 62,107

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
roa	-1.087805	.436315	-2.49	0.013	-1.942967	-.2326437
roc	-.1730393	.104589	-1.65	0.098	-.37803	.0319513
cla	.0293729	.010265	2.86	0.004	.009254	.0494919
er	-1.390576	.6643021	-2.09	0.036	-2.692584	-.0885682
liab_t	-.0005844	.0001278	-4.57	0.000	-.0008349	-.000334
td	1.872435	.6657892	2.81	0.005	.567512	3.177358
cr	-.0829251	.0223363	-3.71	0.000	-.1267035	-.0391467
lnta	.1584555	.0256274	6.18	0.000	.1082267	.2086844
rec_to_ca	.9834533	.1507555	6.52	0.000	.687978	1.278929
_rcs1	2.015897	.3223368	6.25	0.000	1.384128	2.647665
_rcs2	.6863735	.19803	3.47	0.001	.2982418	1.074505
_rcs3	-3.501136	.8835258	-3.96	0.000	-5.232815	-1.769458
_rcs4	10.90185	1.739807	6.27	0.000	7.49189	14.31181
_rcs5	-30.10674	2.988858	-10.07	0.000	-35.96479	-24.24868
_rcs6	30.8749	2.658388	11.61	0.000	25.66455	36.08524
_cons	-20.36908	2.014005	-10.11	0.000	-24.31646	-16.4217

Appendix 11

Table 11-A Model 4

Log likelihood = -2628.585 Number of obs = 64,955

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
roc	-.3681395	.0829548	-4.44	0.000	-.5307279	-.2055511
cla	.029738	.0102565	2.90	0.004	.0096357	.0498404
er	-3.465598	.1745046	-19.86	0.000	-3.807621	-3.123575
liab_t	-.0003597	.0001288	-2.79	0.005	-.0006121	-.0001074
cr	-.0972409	.0235939	-4.12	0.000	-.143484	-.0509978
lnta	.0636264	.0275186	2.31	0.021	.009691	.1175617
rec_to_ca	1.001673	.1496545	6.69	0.000	.7083556	1.29499
2.industry_b	-.8332076	.0832345	-10.01	0.000	-.9963442	-.670071
2.size_b	.3599574	.0874648	4.12	0.000	.1885295	.5313853
_rcs1	1.288641	.0438741	29.37	0.000	1.20265	1.374633
_cons	-13.55233	.4345766	-31.19	0.000	-14.40408	-12.70057

Table 11-B Model 5

Log likelihood = -2617.8608 Number of obs = 62,107

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
roa	-.8475387	.4345109	-1.95	0.051	-1.699164	.0040869
roc	-.1810899	.106566	-1.70	0.089	-.3899555	.0277757
cla	.0321138	.0103177	3.11	0.002	.0118915	.052336
er	-1.474444	.6662877	-2.21	0.027	-2.780344	-.1685441
liab_t	-.0004887	.0001305	-3.75	0.000	-.0007445	-.000233
td	1.932326	.6682032	2.89	0.004	.6226719	3.24198
cr	-.0819099	.0225514	-3.63	0.000	-.1261098	-.0377101
lnta	.1138661	.0272752	4.17	0.000	.0604076	.1673245
rec_to_ca	.9737566	.1507806	6.46	0.000	.6782321	1.269281
2.size_b	.3530968	.0881512	4.01	0.000	.1803235	.52587
_rcs1	1.284528	.0442675	29.02	0.000	1.197765	1.37129
_cons	-16.18096	.786956	-20.56	0.000	-17.72336	-14.63855

Appendix 12

Table 12-A Rating models

Variable	Classification model		Downgrade model	
	Influential variables	Effect on rating grade	Influential variables	Effect on rating downgrade hazard
Total assets	x		✓	+
Return on total assets	✓	+	✓	-
Return on equity	✓	-	✓	+
Liquidity ratio	✓	+	✓	-
Cash flow	✓	+	✓	-
Interest cover	✓	+	✓	-
Long-term debt to total assets	✓	-	x	

Table 12-B Bankruptcy models

Variable	Cox model		Weibull model	
	Influential variables	Effect on bankruptcy hazard	Influential variables	Effect on bankruptcy hazard
Total assets	x		✓	+
Return on total assets	✓	-	✓	-
Return on costs	✓	-	✓	-
Current ratio	x		✓	-
Coverage of long-term assets	✓	+	✓	+
Equity ratio	x		✓	-
Debt ratio	✓	+	✓	+
Liability turnover	✓	-	✓	-
Receivables to current assets	x		✓	+

Table 12-C Corporate characteristics

Sector	**Category**	**Survival probability**
Industry	Industrials Utility Services Agriculture	⇩
Legal form	Joint-stock Limited-liability Cooperatives Other	⇩
Size	Small Large Medium Micro	⇩

List of Figures

List of Tables

References

AGARWAL, R., AUDRETSCH, D. B. (2001). Does Entry Size Matter? The Impact of the Life Cycle and Technology on Firm Survival. *The Journal of Industrial Economics* 49(1): 21–43.

ALTMAN, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance* 23(4): 189–209.

ALTMAN, E. I. (1970). Ratio Analysis and the Prediction of Firm Failure: A Reply. *The Journal of Finance* 25(5): 1169–1172.

ALTMAN, E. I. (2008). Default recovery rates and LGD in credit risk modelling and practice: An updated review of the literature and empirical evidence. In: Jones, S., Henser, D. A. (eds.): *Advances in Credit Risk Modelling and Corporate Bankruptcy Prediction. Quantitative Methods for Applied Economics and Business Research*. Cambridge: Cambridge University Press, 175–206.

ALTMAN, E. I., HALDEMAN, R. G., NARAYANAN, P. (1977). ZETA™ analysis A new model to identify bankruptcy risk of corporations. *Journal of Banking & Finance* 1(1): 29–54.

ALTMAN, E. I., HOTCHKISS, E. (2005). *Corporate Financial Distress and Bankruptcy*. New York: John Wiley & Sons.

ALTMAN, E. I., KATZ, S. (1976). Statistical Bond Rating Classification Using Financial and Accounting Data. In: Schiff, M., Sorter, G. (eds.): *Proceedings of the Conference on Topical Research in Accounting*. New York: New York University School of Business, 205–239.

ALTMAN, E. I., RESTI, A., SIRONI, A. (2001). Analyzing and Explaining Default Recovery Rates. *ISDA Research Report*. London.

ALTMAN, E. I., SABATO, G. (2007). Modelling Credit Risk for SMEs: Evidence from the US Market. *Abacus* 43(3): 332–357.

ALTMAN, E. I., SABATO, G., WILSON, N. (2010). The Value of Non-Financial Information in SME Risk Management. *Journal of Credit Risk* 6(2): 95–127.

ANG, J. S., PATEL, A. K. (1975). Bond Rating Methods: Comparison and Validation. *Journal of Finance* 30(2): 631–640.

AVGOULEAS, E. (2009). The Global Financial Crisis and the Disclosure Paradigm in European Financial Regulation: The Case for Reform. *European Company and Financial Law Review* 6(4): 440–475.

BANK FOR INTERNATIONAL SETTLEMENTS (2001). The Internal Ratings-Based Approach. Supporting Document to the New Basel Capital Accord. [Online], [cit. 23. 6. 2017]. Available at: https://www.bis.org/publ/bcbsca05.pdf.

BELKAOUI, A. (1980). *Industrial Bond Ratings: A Discriminant Analysis Approach*. Working paper. University of Ottawa.

BERK, J. B., DEMARZO, P. M. (2017). *Corporate Finance*. Harlow: Pearson.

BODIE, Z., KANE, A., MARCUS, A. J. (2014). *Investments*. Maidenhead: McGraw-Hill Education.

BOSWELL, T., GUTIERREZ, R., G. (2011). Manipulation of prediction parameters for parametric survival regression models. *The Stata Journal* 11(1): 143–144.

BREALEY, R. A., MYERS, S. C., ALLEN, F. (2014). *Principles of Corporate Finance*. Maidenhead: McGraw-Hill Education.

BUSINESSINFO (2019). Registration of legal form of business. [Online], [cit. 18. 1. 2019]. Available at: https://www.businessinfo.cz/en/psc/start-your-business/registration-of-legal-form-of-business.html.

CANTOR, R., PACKER, F. (1996). Determinants and Impacts of Sovereign Credit Ratings. *Economic Policy Review* 2(2): 37-53.

CAO, L. (2016). *The Altman Z-Score in Edward Altman's Own Words.* CFA Institute Enterprising Investor [Online, blog], [cit. 2. 2. 2016]. Available at: https://blogs.cfainstitute.org/investor/2016/02/02/the-altman-z-score-in-edward-altmans-own-words/.

CECCHETTI, S. G., SCHOENHOLTZ, K. L. (2015). *Money, Banking, and Financial Markets:* New York: McGraw-Hill Education.

CHAN, K., JEGADEESH, N. (2004) Market-Based Evaluation for Models to Predict Bond Ratings. *Review of Pacific Basin Financial Markets and Policies* 7(2): 153–172.

CHAPMAN, J. M. (1940). *Commercial Banks and Consumer Instalment Credit*. New York: The National Bureau of Economic Research.

CHOUDHRY, M. (2010). Structured Credit Products: Credit Derivatives and Synthetic Securitisation. Singapore: John Wiley & Sons (Asia).

CHOUDRY, M. (2012). *The Principles of Banking*. Singapore: John Wiley & Sons Singapore.

CLEVES, M. (2000). Analysis of Multiple Failure-time Data with Stata. *Stata Technical Bulletin* 9.

CLEVES, M., GOULD, W., GUTIERREZ, R. G., MARCHENKO, Y. V. (2010). *An Introduction to Survival Analysis Using Stata*. College Station: Stata Press.

CROTTY, J. (2009). Structural Causes of the Global Financial Crisis: A Critical Assessment of the 'New Financial Architecture'. *Cambridge J Econ* 33(4): 563–580.

CROUHY, M., GALAI, D., MARK, R. (2014). *The Essentials of Risk Management*. New York: McGraw-Hill Education.

CZECH NATIONAL BANK (2022). ARAD data series system. Monetary and Financial Statistics. [Online], [cit. 28. 9. 2022]. Available at: https://www.cnb.cz/docs/ARADY/HTML/index_en.htm.

CZECH STATISTICAL OFFICE (2022). GDP, National Accounts. [Online], [cit. 28. 9. 2022]. Available at: https://www.czso.cz/csu/czso/gdp_national_accounts_ekon.

DE LAURENTIS, G., MAINO, R., MOLTENI, L. (2010). Developing, Validating and Using Internal Ratings: Methodologies and Case Studies. Chichester: Wiley.

DE LEONARDIS, D., ROCCI, R. (2008). Assessing the default risk by means of a discrete-time survival analysis approach. *Applied Stochastic Models in Business and Industry* 24: 291–306.

DE SERVIGNY, A., RENAULT, O. (2004). *Measuring and Managing Credit Risk*. New York: McGraw-Hill.

DLUHOŠOVÁ, D., ČULÍK, M., GURNÝ, P., KRESTA, A., VALECKÝ, J., ZMEŠKAL, Z. (2014). Financial Management and Decision-Making of a Company: Analysis, Investing, Valuation, Sensitivity, Risk, Flexibility. Ostrava: VŠB-TU Ostrava.

DUTTA, S., SHEKHAR, S. (1988). Bond Ratings: A Non-Conservative Application of Neural Network. In: Proceedings of the *IEEE International Conference on Neural Networks.* San Diego: IEEE, 2: pp. 443–450.

EC (2016). Study on the State of the Credit Rating Market. Final Report. [Online], [cit. 25. 7. 2017]. Available at: http://ec.europa.eu/info/system/files/state-of-credit-rating-market-study-01012016_en.pdf.

EC (2017). Regulating Credit Rating Agencies. [Online], [cit. 25. 7. 2017]. Available at: https://ec.europa.eu/info/business-economy-euro/banking-and-finance/financial-supervision-and-risk-management/managing-risks-banks-and-financial-institutions/regulating-credit-rating-agencies_en#current-regulatory-framework.

EFING, M., HAU, H. (2015). Structured Debt Ratings: Evidence on Conflicts of Interest. *Journal of Financial Economics* 116(1): 46–60.

ESMA (2017). Credit Rating Agencies. [Online], [cit. 21. 7. 2017]. Available at: https://www.esma.europa.eu/supervision/credit-rating-agencies/supervision.

ESMA (2022). CRA Authorisation. [Online], [cit. 28. 9. 2017]. Available at: https://www.esma.europa.eu/supervision/credit-rating-agencies/risk.

EUROPEAN BANKING AUTHORITY (2016). Final Report. Guidelines on the application of the definition of default under Article 178 of Regulation (EU) No 575/2013. [Online], [cit. 20. 6. 2017]. Available at: https://www.eba.europa.eu/documents/10180/1597103/Final+Report+on+Guidelines+on+default+definition+%28EBA-GL-2016-07%29.pdf/004d3356-a9dc-49d1-aab1-3591f4d42cb.

EUROPEAN COMMISSION (2009). Regulation (EC) No 1060/2009 of the European Parliament and the Council of 16 September 2009 on Credit Rating Agencies. [Online], [cit. 21. 7. 2017]. Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32009R1060&from=EN.

EUROPEAN COMMISSION (2013). Regulation (EC) No 462/2013 of the European Parliament and the Council of 21 May 2013 on credit rating agencies. [Online], [cit. 20. 7. 2017]. Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32013R0462&from=EN.

FABOZZI, F. J. (2013). *Bond Markets, Analysis, and Strategies*. Harlow: Pearson.

FAIR ISAAC CORPORATION (2017). Credit Basics Demystifying Credit Scoring & Credit Reports. [Online], [cit. 26. 6. 2017]. Available at: https://www.myfico.com/credit-education/credit-report-credit-score-articles/.

FEDERAL DEPOSIT INSURANCE CORPORATION (2017). FDIC: Regulations & Examinations. [Online], [cit. 29. 6. 2017]. Available at: https://www.fdic.gov/regulations/.

FIGLEWSKI, S., FRYDMAN, H., LIANG, W. (2012). Modelling the effect of macroeconomic factors on corporate default and credit rating transitions. *International Review of Economics & Finance* 21: 87–105.

FINGER, C. (1999). Conditional Approaches for Creditmetrics® Portfolio Distributions. *Creditmetrics® Monitor*.

FINLAY, S. (2010). *Credit Scoring, Response Modelling and Insurance Rating: A Practical Guide to Forecasting Consumer Behaviour*. Houndmills, Basingstoke: Palgrave Macmillan.

FISHER, L. (1959). Determinants of Risk Premiums on Corporate Bonds. *Journal of Political Economy* 67: 217–237.

FISHER, R. A. (1936). The Use of Multiple Measurement in Taxonomic Problems. *Annals of Eugenics* 7: 179–188.

FLANNERY, M. J., HOUSTON, J. F., PARTNOY, F. (2010). Credit Default Swaps Spreads as Viable Substitutes for Credit Ratings. *University of Pennsylvania Law Review* 158: 2085–2123.

FRIDSON, M. S., ALVAREZ, F. (2011). *Financial Statement Analysis: A Practitioner's Guide*. Hoboken: Wiley.

FRYDMAN, H., SCHUERMANN, T. (2008). Credit Dynamics and Markov Mixture Models. *Journal of Banking & Finance* 32(6): 1062–1075.

GLENNON, D., NIGRO, P. (2005). Measuring the Default Risk of Small Business Loans: A Survival Analysis Approach. *Journal of Money, Credit and Banking* 37(5): 923–947.

GOLBAYANI, P. FLORESCU, I., CHATTERJEE, R. (2020). A comparative study of forecasting corporate credit ratings using neural networks, support machines, and decision trees. *North American Journal of Economics and Finance* 54: 1–16.

GORDY, M. A. (2000). Comparative Anatomy of Credit Risk Models. *Journal of Banking and Finance* 24(1–2): 119–149.

GOURIEROUX, C., JASIAK, J. (2007). *Econometrics of Individual Risk: Credit, Insurance and Marketing*. Princeton: Princeton University Press.

GRAY, S., MIRKOVIC, A., RAGUNATHAN, V. (2006). The Determinants of Credit Ratings: Australian Evidence. *Australian Journal of Management* 31(2): 333–354.

GUNNVALD, R. (2014). *Estimating Probability of Default Using Rating Migrations in Discrete and Continuous Time*. [Online], [cit. 2. 9. 2017]. Available at: https://www.math.kth.se/matstat/seminarier/reports/M-exjobb14/140908.pdf.

HAIR, J. F., BLACK, W. C., BABIN, B. J., ANDERSON, R. E. (2014). *Multivariate Data Analysis*. Harlow: Pearson.

HAROLD, G. (1938). Bond Ratings as an Investment Guide. An Appraisal of Their Effectiveness. Ronald Press Company.

HARRELL, F. E. (2010). Regression Modeling Strategies: with Applications to Linear Models, Logistic Regression, and Survival Analysis. New York: Springer.

HICKMAN, W. B. (1958). *Corporate Bond Quality and Investor Experience*. Princeton University Press: Princeton.

HODGSON, P. (2016). *Auditor Rotation: Are We Going Round in Circles?* [Online, blog], [cit. 21. 7. 2017]. Available at: https://www.complianceweek.com/blogs/global-glimpses/auditor-rotation-are-we-going-round-in-circles#.WiqJnTCDM-V.

HOSMER, D. W., LEMESHOW, S., MAY, S. (2008). *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*. Hoboken: John Wiley & Sons.

HOSMER, D. W., LEMESHOW, S., STURDIVANT, R. X. (2013). *Applied Logistic Regression*. Hoboken: Wiley.

HOUWELINGEN, J. C., STIJNEN, T. (2014). Cox Regression Model. In: Klein, J. P., van Houwelingen, J. C., Ibrahim, J. G., Scheike, T. H. (eds).: *Handbook of survival analysis*. Boca Raton: CRC Press, Taylor & Francis Group, 5–25.

HSU, F. J., CHEN, M. Y., CHEN, Y. C. (2018). The human-like intelligence with bio-inspired computing approach for credit ratings prediction. *Neurocomputing* 279: 11–18.

HU, J. (2011). *Asset Securitization: Theory and Practice*. Singapore: John Wiley & Sons (Asia).

HU, Y-T., KIESEL, R., PERRAUDIN, W. (2002). The Estimation of Transition Matrices for Sovereign Credit Ratings. *Journal of Banking & Finance* 26(7): 1383–1406.

HUBERTY, C. J., OLEJNIK, S. (2006). *Applied MANOVA and Discriminant Analysis*. Hoboken: John Wiley & Sons.

IBM KNOWLEDGE CENTRE (2017b). Testing Homogeneity of Covariance Matrices. [Online], [cit. 14. 11. 2017]. Available at: https://www.ibm.com/support/knowledgecenter/en/SSLVMB_sub/spss/tutorials/glmm_patlos_homcov.html.

ISDA (2003). Credit Derivatives Definitions. [Online], [cit. 13. 7. 2017]. Available at: http://cbs.db.com/new/docs/2003_ISDA_Credit_Derivatives_Definitions.pdf.

JONES, S., JOHNSTONE, D., WILSON, R. (2015). An empirical evaluation of the performance of binary classifiers in the prediction of credit rating changes. *Journal of Banking & Finance* 56: 72–85.

JORION, P. (2011). *Financial Risk Manager Handbook Plus Test Bank: FRM Part I/Part II.* Hoboken: John Wiley & Sons.

JOSEPH, C. (2013). *Advanced Credit Risk Analysis and Management.* Chichester: Wiley.

KAPLAN, R. S., URWITZ, G. (1979). Statistical Models of Bond Ratings: A Methodological Inquiry. *Journal of Business* 52(2): 231–261.

KARAGOZOGLU, A., JACOBS, M. (2010). *Measuring Credit Risk: CDS Spreads vs. Credit Ratings*. Submitted for presentation at the 2010 FMA Meetings in New York. [Online], [cit. 17. 7. 2017]. Available at: https://www.researchgate.net/publication/228231806_Measuring_Credit_Risk_CDS_Spreads_vs_Credit_Ratings.

KARIMZAD, B. (2015). *Citi Free FICO Scores Are Live – With Something Different.* [Online, blog], [cit. 26. 6. 2017]. Available at: http://www.magnifymoney.com/blog/building-credit/citi-free-fico-scores-live-something-different367796063/.

KELLY, R., BRIEN, E. O., STUART, R. (2015). A Long-Run Survival Analysis of Corporate Liquidations in Ireland. *Small Business Economics* 44(3): 671–683.

KIESEL, F., SPOHNHOLTZ, J. (2017). CDS Spreads as an Independent Measure of Credit Risk. *Journal of Risk Finance* 18(2): 122–144.

KLEIN, J. P., HOUWELINGEN, J. C., IBRAHIM, J. G., SCHEIKE, T. H. ed. (2014). *Handbook of Survival Analysis*. Boca Raton: CRC Press, Taylor & Francis Group.

KOOPMAN, S. J., LUCAS, A., MONTEIRO, A. (2008). The Multi-State Latent Factor Intensity Model for Credit Rating Transitions. *Journal of Econometrics* 142(1): 399–424.

KRISTANTI, F., ISYNUWARDHANA, D. (2018). Survival Analysis of Industrial Sectors in Indonesia Companies. *Jurnal Keuangan dan Perbankan* 22(1): 23–36.

KRZANOWSKI, W. J., HAND, D. J. (2009). *ROC Curves for Continuous Data*. Boca Raton: CRC Press.

LAITINEN, T., KANKAANPÄÄ, M. (1999). Comparative analysis of failure prediction methods: the Finnish case. *The European Accounting Review* 8(1): 67–92.

LAMBERT, P. C., ROYSTON, P. (2009). Further Development of Flexible Parametric Models for Survival Analysis. *The Stata Journal* 9(2): 265–290.

LANE, W., LOONEY, S., WANSLEY, J. (1986). An application of the Cox proportional hazards model to bank failure. *Journal of Banking and Finance* 10: 511–531.

LEOW, M., CROOK, J. (2014). Intensity models and transition probabilities for credit card loan delinquencies. *European Journal of Operational Research* 236(2): 685–694.

LEVINE, A., SARCHESE, N. (2010). Credit Analysis and Analyzing a High-Yield Issuance. In: Maxwell, W. F., Shenkman, M. R. (eds.): *Leveraged Financial Markets. A Comprehensive Guide to High-Yield Bonds, Loans, and Other Instruments*. New York: McGraw-Hill, 113–170.

LONG, J. S., FREESE, J. (2014). *Regression models for categorical dependent variables using Stata*. College Station: Stata Press Publication.

LOUIS, P., VAN LAERE, E., BAESENS, B. (2013). Understanding and predicting bank rating transitions using optimal survival analysis models. *Economics Letters* 119(3): 280–283.

LOUZADA, F., CANCHO, V. D., OLIVEIRA, M. R., BAO, Y. (2014). Modeling Time to Default on a Personal Loan Portfolio in Presence of Disproportionate Hazard Rates. *Journal of Statistics Applications & Probability* 3(3): 1–11.

LUDWIG-MAYERHOFER, W. (2020). Winsorizing and Trimming. *Internet Guide to Stata*. [Online], [cit. 1. 4. 2022]. Available at: https://wlm.userweb.mwn.de/Stata/wstatwin.htm.

MANLY, B. F. J. (2005). *Multivariate Statistical Methods: A Primer*. Boca Raton: Chapman & Hall/CRC.

MAYS, F. E., LYNAS, N. (2011). Credit Scoring for Risk Managers: The Handbook for Lenders. CreateSpace Independent Publishing Platform.

MEDCALC (2018). ROC curve analysis with MedCalc. [Online], [cit. 28. 6. 2023]. Available at: https://www.medcalc.org/manual/roc-curves.php.

MEGGINSON, W. L., LUCEY, B. M., SMART, S. B. (2008). *Introduction to Corporate Finance*. London: Cengage Learning.

MENARD, S. W. (2010). *Logistic Regression: From Introductory to Advanced Concepts and Applications*. Los Angeles: Sage Publications.

MERTON, R. C. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *Journal of Finance* 29(2): 449–470.

MISHKIN, F. S., EAKINS, S. G. (2009). *Financial Markets and Institutions*. Boston: Pearson.

MOKARAMI, M., MOTEFARES, Z. (2013). Corporate Governance and Predicting Bankruptcy of Firms Using Survival Analysis (Case Study of Companies Listed in Tehran Stock Exchange). *Life Science Journal* 10(1): 1089–1095.

MOODY'S ANALYTICS (2012). Public Firm Expected Default Frequency (EDF™) Credit Measures: Methodology, Performance, and Model Extensions. [Online], [cit. 12. 10. 2017]. Available at: https://www.moodys.com/researchdocumentcontentpage.aspx?docid=PBC_143179.

MOODY'S ANALYTICS (2017). History of KMV. [Online], [cit. 11. 7. 2017]. Available at: https://www.moodysanalytics.com/About-Us/History/KMV-History.

MOODY'S CORPORATION (2017). Rating Symbols and Definitions. [Online], [cit. 19. 9. 2017]. Available at: https://www.moodys.com/researchdocumentcontentpage.aspx?docid=PBC_79004.

MOODY'S INVESTORS SERVICE (2006). Measuring Corporate Default Rates [Online], [cit. 5. 10. 2017]. Available at: https://www.moodys.com/sites/products/DefaultResearch/2006200000425249.pdf.

MOODY'S INVESTORS SERVICE (2016). Default and Recovery Rates of European Financial and Non-Financial Corporate Issuers. [Online], [cit. 20. 6. 2017]. [online]. Available at: https://www.moodys.com/researchdocumentcontentpage.aspx?docid=PBC_1055118.

MOODY'S INVESTORS SERVICE (2017). Rating Process Detail. [Online], [cit. 19. 9. 2017]. Available at: https://www.moodys.com/sites/products/ProductAttachments/mis_ratings_process.pdf.

MSCI (2019). The Global Industry Classification Standard (GICS®). [Online], [cit. 18. 1. 2019]. Available at: https://www.msci.com/our-solutions/indexes/gics.

NORUŠIS, M. J. (2012). *IBM SPSS Statistics 19 Advanced Statistical Procedures Companion*. Upper Saddle River: Prentice Hall.

NOVOTNÁ, M. (2012a). The use of different approaches for credit rating prediction and their comparison. In: *Proceedings of the Managing and Modelling of Financial Risks: 6th International Scientific Conference*. Ostrava: VŠB – Technical University of Ostrava: 448–457.

NOVOTNÁ, M. (2012b). Estimating credit rating models by a logistic regression approach. *ECON 12 (Journal of Economics, Management and Business)* 21: 84–93.

NOVOTNÁ, M. (2013). A multivariate analysis of financial and market-based variables for bond rating prediction. *Journal of Economic Computation and Economic Cybernetics Studies and Research* 47(2): 67–83.

NOVOTNÁ, M. (2015). The effect of industry and corporate characteristics on bond rating. *Journal of Applied Economic Sciences* 10(4): 223–233.

NOVOTNÁ, M. (2016). Modelling of Firm Survival in the Czech Transport Sector. In: *Proceedings of the Managing and Modelling of Financial Risks: 8th International Scientific Conference*. Ostrava: VŠB – Technical University of Ostrava: 684–691.

NOVOTNÁ, M. (2017). The Use and Comparison of Survival Models for Corporate Bankruptcies. In: *Proceedings of the Financial Management of Firms and Financial Institutions: 11th International Scientific Conference*. Ostrava: VŠB – Technical University of Ostrava: 644–652.

NOVOTNÁ, M. (2019). Modelling of rating downgrade based on multiple failure-time data. *Ekonomická Revue – Central European Review of Economic Issue* 22(2): 45–56.

NOVOTNÁ, M. (2020). The Impact of Industry and Business Entity Type on Corporate Survival. *Journal of Economic Computation and Economic Cybernetics Studies and Research* 54(1): 97–112.

NOVOTNÁ, M. (2021). Evaluation of Classification Ability of Multiclass Rating Models. In: *Proceedings of the Financial Management of Firms and Financial Institutions*. *13th International Scientific Conference*. Ostrava: VŠB – Technical University of Ostrava: 118–128.

NYE, R. P. (2014). *Understanding and Managing the Credit Rating Agencies: A Guide for Fixed Income Issuers*. London: Euromoney Books.

OECD (2019). Enterprises by business size (indicator). [Online], [cit. 18. 1. 2019]. Available at: https://data.oecd.org/entrepreneur/enterprises-by-business-size.htm.

PAGANO, M., VOLPIN, P. (2010). Credit Ratings Failures and Policy Options. *Economic Policy* 25(62): 401–431.

PARNES, D. (2007). Time Series Patterns in Credit Ratings. *Finance Research Letters* 4: 217–226.

PEREIRA, J. (2014). Survival Analysis Employed in Predicting Corporate Failure: A Forecasting Model Proposal. *International Business Research* 7(5): 9–20.

PETERSON, G. E. (1998). *Measuring Local Government Credit Risk and Improving Creditworthiness*. Prepared for World Bank. [Online], [cit. 13. 7. 2017]. Available at: http://siteresources.worldbank.org/INTMF/Resources/339747-1251914038224/tools.PDF.

PETITT, B. S. P., PINTO, J. E., PIRIE, W. L., GRIEVES, R., NORONHA, G. M. (2015). *Fixed Income Analysis*. Hoboken: Wiley.

PINCHES, G. E., MINGO, K. A. (1973). A Multivariate Analysis of Industrial Bond Ratings. *Journal of Finance* 28(1): 1–18.

REILLY, F. K., BROWN, K. C. (2015). *Analysis of Investments and Management of Portfolios*. Andover: Cengage Learning.

RENCHER, A. C. (2002). *Methods of Multivariate Analysis*. Hoboken: John Wiley & Sons.

REROLLE, J., RIMAUD, C. (2009). Does Independent Credit Research Add Value? *Alternative Intelligence Quotient* 32: 31–37.

RESTI, A. (2002). Replicating Agency Ratings Through Multinomial Scoring Models. In: Ong, M. K. (eds.): *Credit Ratings: Methodologies, Rationale and Default Risk*. London: Risk Books, 213–232.

ROA, M., GARCÍA, A. F., BONILLA, L. (2009). Country Risk Ratings and Financial Crisis 1995 – 2001: A Survival Analysis. *Review of Business* 30: 33–45.

ROYSTON, P. (2006). Explained Variation for Survival Models. *The Stata Journal* 6(1): 83–96.

ROYSTON, P., LAMBERT P. C. (2011). *Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model.* College Station: Stata Press.

S&P (2015). Annual Global Corporate Default Study and Rating Transitions. [Online], [cit. 13. 6. 2017]. Available at: http://www.spratings.com/documents/20184/774196/2015+Annual+Global+Corporate+Default+Study+And+Rating+Transitions/6d311074-5d56-4589-9ef8-a43615a6493b.

S&P FINANCIAL SERVICES (2014). Guide to Credit Rating Essentials. What are Credit Ratings and How They Work? [Online], [cit. 20. 9. 2017]. Available at: https://www.spratings.com/documents/20184/760102/SPRS_Understanding-Ratings_GRE.pdf/298e606f-ce5b-4ece-9076-66810cd9b6aa.

S&P FINANCIAL SERVICES (2017a). Default, Transaction and Recovery: 2016 Annual Global Corporate Default Study and Rating Transitions. [Online], [cit. 20. 6. 2018]. Available at: https://www.spratings.com/documents/20184/774196/2016+Annual+Global+Corporate+Default+Study+And+Rating+Transitions.pdf/2ddcf9dd-3b82-4151-9dab-8e3fc70a7035.

S&P GLOBAL MARKET INTELLIGENCE (2016). S&P Global Ratings Definitions. S&P Global Market Intelligence. [Online], [cit. 19. 9. 2017]. Available at: http://img.en25.com/Web/StandardPoorsRatings/RatingsDirect_Commentary_1695715_Sep-13-2016_15_16.pdf.

S&P GLOBAL RATINGS (2017). Understanding Ratings. [Online], [cit. 19. 9. 2017]. Available at: https://www.spratings.com/en_US/understanding-ratings#fourthPage.

S&P GLOBAL RATINGS (2021). Default, Transition, and Recovery: 2020 Annual Global Corporate Default and Rating Transition Study. [Online], [cit. 10. 1. 2022]. Available at: https://www.spglobal.com/ratings/en/research/articles/210407-default-transition-and-recovery-2020-annual-global-corporate-default-and-rating-transition-study-11900573.

SANGIORGI, F., SPATT, C. S. (2013). Opacity, Credit Rating Shopping and Bias. *SSRN Electronic Journal* 10.2139/ssrn.2021073.

SAUNDERS, A., ALLEN, L. (2010). *Credit Risk Measurement In and Out of the Financial Crisis: New Approaches to Value at Risk and Other Paradigms.* Hoboken: John Wiley & Sons.

SAUNDERS, A., CORNETT, M. M. (2008). *Financial Institutions Management. A Risk Management Approach*. New York: McGraw-Hill.

SCHROEDER, S. K. (2015). *Public Credit Rating Agencies: Increasing Capital Investment and Lending Stability in Volatile Markets*. New York: Palgrave Macmillan.

SEC (2017). Municipal Bonds: Understanding Credit Risk. Investor Bulletin. [Online], [cit. 20. 9. 2017]. Available at: https://www.sec.gov/files/municipalbondsbulletin.pdf.

SINCLAIR, T. J. (2005). *The New Masters of Capital: American Bond Rating Agencies and the Politics of Creditworthiness*. Ithaca: Cornell University Press.

STATACORP. (2017). *Stata 15 Base Reference Manual*. College Station, TX: Stata Press.

STATACORP. (2021). *Stata: Release 17*. College Station, TX: StataCorp LLC.

STOWEL, D. (2010). *An Introduction to Investment Banks, Hedge Funds & Private Equity. The New Paradigm*. San Diego: Academic Press.

SUN, J., LI, J. (2014). Interval Censoring. In: Klein, J. P., van Houwelingen, J. C., Ibrahim, J. G., Scheike, T. H. (eds).: *Handbook of survival analysis*. Boca Raton: CRC Press, Taylor & Francis Group, 369–390.

SURKAN, A. J., SINGLETON, J. C. (1990). Neural Networks for Bond Rating Improved by Multiple Hidden Layers. In: *Proceedings of the International Joint Conference on Neural Networks.* San Diego, California, 2: 157–162.

TABACHNICK, B. G., FIDELL, L. S. (2007). *Using Multivariate Statistics*. Boston: Pearson Allyn & Bacon.

THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION (2013). Directive 2013/14/EU of the European Parliament and the Council of 21 May 2013. [Online], [cit. 25. 7. 2017]. Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32013L0014&from=EN.

U. S. SECURITIES AND EXCHANGE COMMISSION (2017). Municipal Bonds: Understanding Credit Risk. Investor Bulletin. [Online], [cit. 13. 7. 2017]. Available at: https://www.sec.gov/files/municipalbondsbulletin.pdf.

UCLA (2021a). Multinomial Logistic Regression. Stata Annotated Output. [Online], [cit. 10. 2. 2022]. Available at: https://stats.oarc.ucla.edu/stata/output/multinomial-logistic-regression-2/.

UCLA (2021b). How Do I Interpret Odds Ratios in Logistic Regression? Stata FAQ. [Online], [cit. 10. 2. 2022]. Available at: https://stats.oarc.ucla.edu/stata/faq/how-do-i-interpret-odds-ratios-in-logistic-regression/.

WAAGEPETERSEN, R. (2010). A Statistical Modeling Approach to Building an Expert Credit Rating System. *The Journal of Credit Risk* 6(2): 81–94.

WANG, M., KU, H. (2021). Utilizing historical data for corporate credit rating assessment. *Expert systems with applications* 165(1): 1–12.

WILKINSON, J. (2013). *Z Score Model. The Strategic CFO*. [Online], [cit. 11. 7. 2017]. Available at: https://strategiccfo.com/z-score-model.

WINGLER, T. R., WATTS, J. M. (1980). An Analysis of Bond Rating Changes in the Electric Utility Industry. *The Financial Review* 15(4): 1–89.

ZHANG, J., THOMAS, L. C. (2012). Comparisons of linear regression and survival analysis using single and mixture distributions approaches in modelling LGD. *International Journal of Forecasting* 28(1): 204–215.

Index

Micro-Modelling Approaches for Credit Rating and Corporate Survival

Martina Novotná

Summary

Corporate micro-modelling approaches for credit risk measurement rely on fundamentally based models, mostly utilising a company's financial reports. The purpose of these models is to establish scoring systems that serve as effective tools to assess a borrower's creditworthiness. This monograph is focused on measuring and modelling credit risk associated with the counterparty's failure to meet contractual obligations. The main purpose is to provide a theoretical background and then apply and compare econometric models for measuring credit risk.

The application uses a micro-approach to assess individual subjects' credit risk by monitoring basic indicators. The main goal is to apply selected methods of credit risk modelling to real data from Central and Eastern European countries, with a specific focus on Czech companies. The emphasis is on evaluating individual credit risk in the context of credit rating and corporate survival. Hence, the principal contribution lies in the application of real corporate data.

The text is divided into three parts devoted to the main principles of credit risk, the description of econometric approaches, and four empirical studies on credit rating and corporate bankruptcy modelling. The common aim of the application chapters is to identify the main predictive variables of particular credit risk measurements and determine, based on that, the association between credit rating and corporate survival models.

This book is intended for everyone interested in credit risk, particularly rating and corporate survival modelling, mainly for academia and students at all levels of study. The monograph aims to provide complex information on credit risk fundamentals, current trends and rating systems' principles. However, the primary

purpose is the practical application and estimating models using real corporate data.

About the authors

Ing. Martina Novotná, Ph.D.

Martina Novotná is an Assistant Professor at the Department of Finance of the Technical University of Ostrava. She has been working at the Department of Finance, where she has taught courses in financial management for MBA students, introductory and advanced courses in financial markets, trends and innovations in financial markets, banking and ethics in finance, both at undergraduate and postgraduate levels. Her work experience involves teaching activities at Trinity College Dublin and Dublin City University in Ireland. She participated as a junior researcher at the faculty and in several research projects of the Student Grant Competition (SGS), four of which she led as the main investigator. Her research has been published in international scientific journals and proceedings at international conferences.

EDITORS' OFFICE
VŠB-Technical University Ostrava,
Faculty of Economics, 17. listopadu 2172/15
708 00 Ostrava-Poruba, Czech Republic
Assistant Editor: *Markéta FICELOVÁ*

PUBLISHER
VŠB-TUO, IČ 61989100
Faculty of Economics, 17. listopadu 2172/15
708 00 Ostrava-Poruba, Czech Republic

Series on Advanced Economic Issues is published by Faculty of Economics, VŠB-Technical University of Ostrava. It covers a broad set of topics in business/economics disciplines, but mainly current issues in economics, finance, management, business economy, and informatics. Original results of research focusing on any of the topics mentioned above are welcome.
The series addresses researchers, students, and practitioners interested in advanced treatment of all economic disciplines. Manuscripts can be submitted to *saei@vsb.cz*. We kindly ask potential authors to follow the instructions about the structure of the book before they proceed to submission procedure. The text can be written either in *Czech* or *English*. The text's length shouldn't be less 100 pages, when the template is followed. Before publishing, each manuscript must be reviewed at least by two independent reviewers. The reviewing procedure is strictly double–blind. For further information authors may visit www.ekf.vsb.cz/saei.

MICRO-MODELLING APPROACHES FOR CREDIT RATING AND CORPORATE SURVIVAL
Martina Novotná
Vydala VŠB-TUO
1. vydání 2024
Tisk VŠB-TUO
Náklad 100 ks, neprodejné
Počet stran 248

Volume 1
Prokop, L., Medvec, Z., Zmeškal, Z. (2009). *Problematika oceňování nedodané energie v průmyslu,*SAEI, vol. 1. Ostrava: VSB – TU Ostrava.
Volume 2
Hrubý, P. a kol. (2010). *Víceúrovňové modelování podnikových procesů (systém REA),* SAEI, vol. 2. Ostrava: VSB – TU Ostrava.
Volume 3
Kutscherauer, A. a kol. (2010). *Regionální disparity. Disparity v regionálním rozvoji země, jejich pojetí, identifikace a hodnocení*, SAEI, vol. 3. Ostrava: VSB – TU Ostrava.
Volume 4
Hančlová, J. a kol. (2010). *Makroekonometrické modelování české ekonomiky a vybraných ekonomik EU*, SAEI, vol. 4. Ostrava: VSB – TU Ostrava.
Volume 5
Martiník, I. (2015). *Modelování objektově orientovaných programových systémů Petriho sítěmi*, SAEI, vol. 5. Ostrava: VSB – TU Ostrava.
Volume 6
Tichý, T. (2010). *Simulace Monte Carlo ve financích: Aplikace při ocenění jednoduchých opcí*, SAEI, vol. 6. Ostrava: VSB – TU Ostrava.
Volume 7
Doleželová, H., Halásek, D. (2011). *Služby v obecném hospodářském zájmu v EU. Komparace České republiky a Německa*, SAEI, vol. 7. Ostrava: VSB – TU Ostrava.
Volume 8
Machová, Z., Kaštan, M., Janíčková, L., Kotlán, I. (2011). *Tax Burden in OECD Countries: WTI Application*, SAEI, vol. 8. Ostrava: VSB – TU Ostrava.
Volume 9
Tichý, T. (2011). *Lévy Processes in Finance: Selected Applications with Theoretical Background*, SAEI, vol. 9. Ostrava: VSB – TU Ostrava.
Volume 10
Macurová, P. a kol. (2011). *Řízení rizik v logistice*, SAEI, vol. 10. Ostrava: VSB – TU Ostrava.
Volume 11
Dvoroková, K. a kol. (2012). *Ekonometrické modelování konvergence ekonomické a cenové úrovně. Analýza průřezových a panelových dat*, SAEI, vol. 11. Ostrava: VSB – TU Ostrava.
Volume 12
Melecký, M., Melecký, A. (2012). *Aplikované modelování pro potřeby hospodářské politiky*, SAEI, vol. 12. Ostrava: VSB – TU Ostrava.
Volume 13
Halásková, M. (2012). *Veřejná správa a veřejné služby v zemích Evropské unie*, SAEI, vol. 13. Ostrava: VSB – TU Ostrava.

Volume 14
Šebestíková, V. a kol. (2012). *Daňová a sociální optimalizace ve vztahu k nezaměstnanosti v České republice*, SAEI, vol. 14. Ostrava: VSB – TU Ostrava.
Volume 15
Pytlíková, M. a kol. (2012). *Gender Wage Gap and Discrimination in the* Czech R*epublic*, SAEI, vol. 15. Ostrava: VSB – TU Ostrava.
Volume 16
Hučka, M. a kol. (2012). *Správa společností v zemích střední a východní Evropy*, SAEI, vol. 16. Ostrava: VSB – TU Ostrava.
Volume 17
Vrabková, I. (2012). *Perspektivy řízení kvality ve veřejné správě*, SAEI, vol. 17. Ostrava: VSB – TU Ostrava.
Volume 18
Marček, D. (2013). *Pravděpodobnostné modelovanie a soft computing v ekonomike*, SAEI, vol. 18. Ostrava: VSB – TU Ostrava.
Volume 19
Čulík, M. (2013). *Aplikace reálných opcí v investičním rozhodování firmy*, SAEI, vol. 19. Ostrava: VSB – TU Ostrava.
Volume 20
Šnapka, P. a kol. (2013). *Rozhodování a rozhodovací procesy v organizaci (vybrané problémy)*, SAEI, vol. 20. Ostrava: VSB – TU Ostrava.
Volume 21
Šimíčková, M., Drastichová, M. (2013). *Ekonomie udržitelnosti – alternativní přístupy a perspektivy*, SAEI, vol. 21. Ostrava: VSB – TU Ostrava.
Volume 22
Kutscherauer, A., Šotkovský, I., Adamovský, J., Ivan, I. (2013). *Socioekonomická geografie a regionální rozvoj. Regionální analýzy v přístupech socioekonomické geografie k regionálnímu rozvoji,* SAEI, vol. 22. Ostrava: VSB – TU Ostrava.
Volume 23
Kutscherauer, A., Václavková, R., Malinovský, J. a kol. (2013). *Komplementární přístupy k podpoře regionálního a municipálního rozvoje,* SAEI, vol. 23. Ostrava: VSB – TU Ostrava.
Volume 24
Komárková, Z., Frait, J., Komárek, L. (2013). *Macroprudential Policy in a Small Economy,* SAEI, vol. 24. Ostrava: VSB – TU Ostrava.
Volume 25
Komárek, L. a kol. (2013). *Money, Pricing and Bubbles,* SAEI, vol. 25. Ostrava: VSB – TU Ostrava.

Volume 26
Tichý, T. (2013). *Backtesting of VaR Based Models: Methodological Review and Selected Applications,* SAEI, vol. 26. Ostrava: VSB – TU Ostrava.
Volume 27
Fojtíková, L. a kol. (2014). *Postavení Evropské unie v podmínkách globalizované světové ekonomiky*, SAEI, vol. 27. Ostrava: VSB – TU Ostrava.
Volume 28
Dluhošová, D. et al. (2014). *Financial Management and Decision-making of a Company. Analysis, Investing, Valuation, Sensitivity, Risk, Flexibility*, SAEI, vol. 28. Ostrava: VSB – TU Ostrava.
Volume 29
Zmeškal, Z. a kol. (2014). *Aplikace vícekriteriálních dekompozičních metod rozhodování v oblasti podnikové ekonomiky a managementu*, SAEI, vol. 29. Ostrava: VSB – TU Ostrava.
Volume 30
Toloo, M. (2014). *Introduction to Data Envelopment Analysis with Basic Models and Applications*, SAEI, vol. 30. Ostrava: VSB – TU Ostrava.
Volume 31
Janasová, E., Slavata, D., Ardielli, J. (2014). *Specifikace míry kapitalizace vybraného segmentu realitního trhu. Analyticko-statistická studie trhu s byty v Ostravě,* SAEI, vol. 31. Ostrava: VSB – TU Ostrava.
Volume 32
Tichý, T. (2015). *Simulation Monte Carlo In Finance*. *Pricing of (Exotic) Options*, SAEI, vol. 32. Ostrava: VSB – TU Ostrava.
Volume 33
Kresta, A. (2015). *Financial Engineering in Matlab: Selected Approaches and Algorithms*, SAEI, vol. 33. Ostrava: VSB – TU Ostrava.
Volume 34
Macháček, M. (2015). *Soumrak ekonomie? K problému formalizace a krize smyslu společenské vědy*, SAEI, vol. 34. Ostrava: VSB – TU Ostrava.
Volume 35
Macek, R. et al. (2015). *Effective Corporate Taxation in the European Union and the Czech Republic*, SAEI, vol. 35. Ostrava: VSB – TU Ostrava.
Volume 36
Vrabková, I., Vaňková, I. (2015). *Evaluation Models of Efficiency and Quality of Bed Care in Hospitals*, SAEI, vol. 36. Ostrava: VSB – TU Ostrava.
Volume 37
Bazsová, B., Křížová, A., Řeháček, P. (2015). *Teorie organizace. Přístupy, metody, nástroje, softwarová podpora*, SAEI, vol. 37. Ostrava: VSB – TU Ostrava.

Volume 38

Šotkovský, I. (2015*). Socioekonomické struktury v rozvoji regionu: Postavení obyvatelstva Moravskoslezska v prostorové struktuře regionů soudržnosti ČR,* SAEI, vol. 38. Ostrava: VSB – TU Ostrava.

Volume 39

Fojtíková, L., Vahalík, B. (2015). *Zahraniční obchod České republiky pod vlivem globalizačních jevů a procesů,* SAEI, vol. 39. Ostrava: VSB – TU Ostrava.

Volume 40

Melecký, A., Melecký, M. (2015). *Řízení vládního dluhu: strategie, riziko, alokace*, SAEI, vol. 40. Ostrava: VSB – TU Ostrava.

Volume 41

Hančlová, J. et al. (2015). *Optimization Problems in Economics and Finance,* SAEI, vol. 41. Ostrava: VSB – TU Ostrava.

Volume 42

Peterková, J., Ludvík, L. (2015). *Řízení inovací v průmyslovém podniku*, SAEI, vol. 42. Ostrava: VSB – TU Ostrava.

Volume 43

Kresta, A. (2016). *Finanční modelování s aplikacemi v prostředí MATLAB*, SAEI, vol. 43. Ostrava: VSB – TU Ostrava.

Volume 44

Melecký, L., Staníčková, M. (2015). *Soudržnost a konkurenceschopnost vybraných zemí a regionů Evropské unie,* SAEI, vol. 44. Ostrava: VSB – TU Ostrava.

Volume 45

Marček, D. (2016). *Supervizované a nesupervizované učení z dat: Statistický a soft přístup*, SAEI, vol. 45. Ostrava: VSB – TU Ostrava.

Volume 46

Friedrich, V. (2017). *Postojové a hodnoticí škály v marketing a managementu. Vybrané statistické metody a aplikace*, SAEI, vol. 46. Ostrava: VSB – TU Ostrava.

Volume 47

Vrabková, I., Vaňková, I., Bečica, J., Kryšková, Š. (2017). *Příspěvkové organizace. Postavení, úkoly a technická efektivnost*, SAEI, vol. 47. Ostrava: VSB – TU Ostrava.

Volume 48

Czudek, D., Dvořáková, P., Kozieł, M., Krügerová, M., Kubica, J., Závada, V. (2014). *Vybrané právní a ekonomické aspekty islámského finančnictví*, SAEI, vol. 48. Ostrava: VSB – TU Ostrava.

Volume 49

Peterková, J. (2018). *Využití konceptů inovací v průmyslovém podniku*, SAEI, vol. 49. Ostrava: VSB – TU Ostrava.

Volume 50
Velčovská, Š. (2018). *Zhodnocení mezigeneračních postojů českých spotřebitelů ke značkám kvality potravin*, SAEI, vol. 50. Ostrava: VSB – TU Ostrava.
Volume 51
Staníčková, M. (2018). *EU Competitiveness and Resilience: Evidenc-based on Regional Leve*l, SAEI, vol. 51. Ostrava: VSB – TU Ostrava.
Volume 52
Drastichová, M. (2018). *The Theory and Measurement of Sustainable Development*, SAEI, vol. 52. Ostrava: VSB – TU Ostrava.
Volume 53
Zapletal, F. (2018). *Optimization Models for Emissions Management*, SAEI, vol. 53. Ostrava: VSB –U Ostrava.
Volume 54
Jenčová, S. (2018). *Aplikácia pokročilých metód vo finančno-ekonomickej analýze elektrotechnického odvetvia Slovenskej republiky,* SAEI, vol. 54. Ostrava: VSB – TU Ostrava.
Volume 55
Balcar, J. (2018). *Soft Skills on the Czech Labour Market*, SAEI, vol. 55. Ostrava: VSB – TU Ostrava.
Volume 56
Vitali, S. (2018). *Pension Fund Management in a Stochastic Optimization Framework*, SAEI, vol. 56. Ostrava: VSB – TU Ostrava.
Volume 57
Hozman, J. et al. (2018). *Robust Numerical Schemes for Pricing of Selected Options*, SAEI vol. 57. Ostrava: VSB – TU Ostrava.
Volume 58
Martiník, I. (2018). *Petriho sítě s časovými známkami*, SAEI vol. 58. Ostrava: VSB – TU Ostrava.
Volume 59
Kouaissah N., Ortobelli S., Tichý T. (2018). *Portfolio Theory and Conditional Expectations: Selected Models and Applications*, SAEI, Vol. 59. Ostrava: VŠB – TU Ostrava.
Volume 60
Hodula, M. (2019). *Stínové bankovnictví v Evropě: pojetí, regulace a faktory růstu*, SAEI, vol. 60. Ostrava: VŠB – TU Ostrava.
Volume 61
Cassader, M., Tichý, T., Vitali, S. (2019). *Benchmark Tracking Portfolio Problems with Stochastic Ordering Constraints*, SAEI, vol. 61. Ostrava: VSB – TU Ostrava.

Volume 62
Kovářová, E. a kol. (2019). *Společensko-ekonomické předpoklady sociálního vyloučení dětí a mládeže v České republice,* SAEI, vol. 62. Ostrava: VSB – TU Ostrava.
Volume 63
Komárek, L. et al. (2019). *Stabilization Policies: Selected Challenges in the International Context,* SAEI, vol. 63. Ostrava: VSB – TU Ostrava.
Volume 64
Krajňák, M. (2020). *Osobní důchodová daň v České republice se zaměřením na příjmy ze závislé činnosti,* SAEI, vol. 64. Ostrava: VŠB – TU Ostrava.
Volume 65
Seďa, P. (2020). *Teoretické aspekty hypotézy efektivního trhu a empirie na akciových trzích v zemích střední Evropy,* SAEI, vol. 65. Ostrava: VŠB – TU Ostrava.
Volume 66
Janků, J. (2020). *Ekonomické důsledky neinformovanosti voliče: Politicko-rozpočtový cyklus,* SAEI, vol. 66. Ostrava: VŠB-TUO.
Volume 67
Tománek, P. (2022). *Rozpočtové určení daní v ČR,* SAEI, vol. 67. Ostrava: VŠB-TUO.
Volume 68
Wu, X. (2023). *The impact of selected moderators on the relationship between CSR and profitability: Evidence form China,* SAEI, vol. 68. Ostrava: VŠB-TUO.

MICRO-MODELLING APPROACHES FOR CREDIT RATING AND CORPORATE SURVIVAL

Corporate micro-modelling approaches for credit risk measurement rely on fundamentally based models, mostly utilising a company's financial reports. The purpose of these models is to establish scoring systems that serve as effective tools to assess a borrower's creditworthiness. This monograph aims to apply selected methods of credit risk modelling to real data from Central and Eastern European countries, with a specific focus on Czech companies. The emphasis is on evaluating individual credit risk in the context of credit rating and corporate survival. The principal contribution lies in the application to real corporate data.

The monograph targets individuals interested in credit risk, particularly in rating and corporate survival modelling, mainly for academia and students at all levels of study. The text is structured into three key parts. The first part provides an overview of credit risk and rating assessment essentials. The second part describes econometric approaches utilised in the applications. Lastly, the third part focuses on practical applications and encompasses four empirical studies within the framework of credit rating and corporate bankruptcy modelling.

The publication has been supported by the SGS project Financial decision-making of firms and financial institutions under risk II (SP 2019/132).

Published:
VŠB - TECHNICAL UNIVERSITY OF OSTRAVA
FACULTY OF ECONOMICS
17. listopadu 2172/15, 708 00 Ostrava
tel.: +420 597 322 421
e-mail: saei@vsb.cz
ISBN 978-80-248-4728-3 (print)
ISBN 978-80-248-4729-0 (on-line)
DOI 10.31490/9788024847290