Univerzita Tomáše Bati ve Zlíně
Fakulta managementu a ekonomiky

DOCTORAL THESIS

Utilization of advanced analytical quantitative methods for measuring companies' performance

Využití pokročilých analytických kvantitativních metod pro měření výkonnosti podniků

Author:	Ing. Lubor Homolka
Degree Program:	P 6202 Hospodářská politika a správa
Degree Course:	6202V010 Finance
Supervisor:	doc. Ing. Petr Klímek, Ph.D.
Year:	2015

Published by Tomas Bata University in Zlín.
Full version of the Doctoral Thesis available at the TBU Library

Klíčová slova

finance, výkonnost, kreditní riziko, statistická analýza, machine learning, složené modely

Keywords

finance, performance, credit risk, statistical analysis, machine learning, emsemble models

ISBN 978-80-

ABSTRAKT

Oblast hodnocení výkonnosti podniků je aktuálním tématem diskutovaným jak v akademické obci, tak v praxi. Vzhledem k povaze studovaného problému, který spadá do oblasti společenských věd, neexistuje jeden jediný správný metodologický přístup k řešení dané problematiky. Tato práce přináší kvantitativní pohled.

Na začátku práce je provedena analýza současného stavu. Identifikace klíčových poznatků důležitých pro pochopení problematiky z ekonomických disciplín je uvedena v samostatné kapitole. V této kapitole jsou uvedeny rozdílné definice a perspektivy (např. účel informací vycházející z rozdílných účetních systémů) uváděné v literatuře, které je nutné zohlednit zejména v první fázi tvorby analytického modelu. Další část uvádí základní principy a metody moderní datové analýzy. Tato část obsahuje odkazy na teorii pravděpodobnosti, teorii informace, statistickou analýzu v její klasické i bayesovské podobě, moderní nástroje strojového učení a umělé inteligence. Konceptuální model pro tvorbu analytického modelu je uveden na začátku kapitoly.

Rozšíření teoretických a praktických přístupů k oblasti hodnocení výkonnosti podniků bylo hlavní motivací pro napsání práce. Hlavním cílem práce je vytvoření konceptuálního modelu, který identifikuje důležité fáze tvorby analytického modelu a jejich obsah. K dosažení tohoto cíle byly stanoveny dva dílčí cíle. Prvním dílčím cílem je zodpovězení tří vědeckých otázek. Druhý dílčí cíl spočívá v demonstraci pokročilých metod na reálných finančních datech.

V následující části práce je vymezena podstata problému z pohledu metodického. Jsou identifikovány převládající názory na podstatu neurčitosti a typy výzkumu ve společenských vědách. Dále jsou uvedeny metodické možnosti a omezení výzkumu a jejich implikace pro předkládanou práci.

Další kapitola je věnována demonstracím principů a metod uvedených v předchozích kapitolách. Hlavní důraz je kladen na popis a vizualizaci neurčitosti odhadu parametrů a výstupů modelů. Složené modely (ensemble modeling) a neuronové sítě dosáhly nejvyšší klasifikační úspěšnosti.

V posledních kapitolách jsou shrnuty dosažené výsledky, přínosy pro vědu a praxi a jsou navrhnuty oblasti dalšího výzkumu.

ABSTRACT

Companies performance evaluation field is an actual topic in academic and practice. There is no single one and correct approach how to treat the studied phenomena. It's because of the nature of social sciences where this topic belongs to. This thesis presents a quantitative perspective.

Important findings from the economics perspective and related fields of study are presented in beginning. This part points to various definitions, systems and objectives (i.e., purpose of accounting figures in different accounting systems) which appear in literature and which need to be considered in the initial phase of model building process. Core principles and methods of modern data analysis are reviewed in the following chapter. This part refers to various disciplines, such as probability theory, information theory, statistical analysis (from both standard and Bayesian perspective), machine learning and artificial intelligence techniques. The conceptual model is presented in the beginning of the chapter.

The motivation for writing this thesis was to extend the theoretical ground of corporate performance modelling research and to provide guidance in the analytical model building process. The thesis goal and objectives are stated in the beginning of the text. The main goal is to develop the conceptual model which suggests the general approach to the model-building process of the analytical model. To accomplish the main goal, two objectives were stated. The first objective is to answer three research questions. The second objective is to demonstrate state-of-the-art analytical methods on real financial data.

The following chapters deal with methodological aspects. Prevailing opinions on uncertainty and types of research conducted in social sciences are reviewed. Possibilities and restrictions emerging from methodological setup and their implications for the thesis are described.

Result's chapter contains demonstrations of principles and analytical methods described in previous chapters. The main emphasis is put on the description and visualisation of parameters' and outcomes' uncertainty. Demonstrations reveal that ensemble models and neural network model outperforms other models.

The last chapters discuss achieved results and highlight contributions to theory, academia and practice. Recommendations for further development conclude the thesis.

ACKNOWLEDGEMENT

What a journey it was! I was lucky to meet so many inspirational, knowledgeable and supportive people during my study. It is impossible to mention all of you in this acknowledgement. But, honestly, I do not even think it is necessary. What happened matters more than what is written here and you know how much you mean to me.

Foremost, I am thankful to my family for their support and trust you have put in me.

I would like to thank my supervisor, doc. Ing. Petr Klímek, Ph.D for his guidance. Besides my supervisor, I am especially thankful to prof. Ing. Jaroslav Belás, Ph.D for his advices related to both professional and personal life.

My sincere thanks belong to my colleagues at Department of Statistics and Quantitative Methods. Special credit belongs to Libor for his support in difficult early years and for his tremendous work with LaTeXtemplate. I would like to thank to prof. Dr. Ing. Drahomíra Pavelková for creating perfect working environment and for giving me a chance to participate in various projects from the very first day.

I have learnt a lot from our community of R and statistics enthusiasts. I have read hundreds of blog posts which shaped my perspectives about the subject. I cannot imagine this thesis would exist without the immense effort of lectures at coursera.org, threads on stack overflow or Cross Validated. Thank you all.

I cannot forget to mention my mates at (legendary) 201. This place became my second home for the last 5 years. Although all members changed since I have entered the room, that unique atmosphere of friendship prevailed.

Děkuji, ďakujem, ස්තූතියි , danke, gracias, cám ơn , thank you all.

CONTENTS

LIST OF FIGURES, TABLES AND ABBREVIATIONS

List of Figures

List of Tables

List of Abbreviations

AUC	Area Under Curve
BIS	Basel Committee on Banking Supervision
CI	Confidence Interval
CFROI	Cash Flow Return on Investment
CM	Credit Metrics Model
EBITDA	Earnings Before Interest, Taxes, Depreciation and Amortisation
EVA	Economic Value Added
ER	Error Rate
FDP	Financial Distress Prediction
FL	Fuzzy Logic
GAAP	Generally Accepted Accounting Principles
GLMM	Generalized Linear Mixed Models
GOF	Goodness of Fit
HLT	Hosmer-Lemeshow Test
HPD	Highest Posterior Density region
IFRS	International Financial Reporting Standards
LDA	Linear Discriminant Analysis
MSE	Mean Square Error
ML	Machine Learning
MM	Modigliani Miller Model
MSCM	Multiple Single-Company Models
NN	Neural Network
NOPLAT	Net Operating Profit Less Adjusted Taxes
QDA	Quadratic Discriminant Analysis
RBF	Gaussian Radial Basis Function
RF	Random Forest
RI	Residual Income
RL	Relational Lending
ROA	Return on Assets
ROC	Receiver Operating Characteristics
ROE	Return on Equity
RONA	Return on Net Assets
SC	Supervised Classification
SMOTE	Synthetic Minority Over-sampling TEchnique
SOF	Substance Over Form
SOM	Self Organising Map
SVM	Support Vector Machine
UC	Unsupervised Classification
WACC	Weighted Cost of Capital

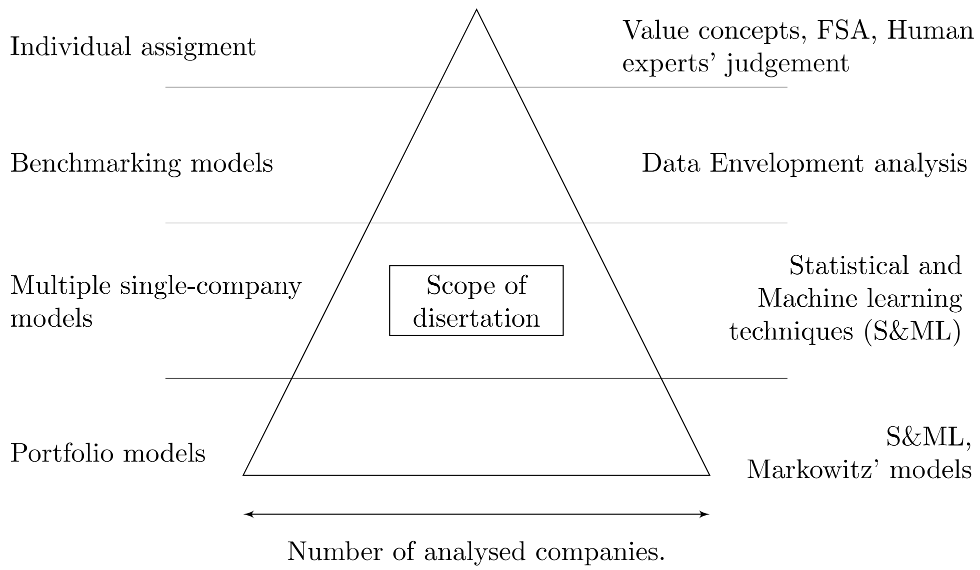

Fig. 1: *The scope of the thesis is defined by the size of the portfolio of analysed companies and methods used. Source: Own*

1 INTRODUCTION

The building element of every market-oriented economy is a company. Other subjects who are operating on the market, such as banks, funds or individual investors follow various aims. Some try to allocate their resources, minimise risks or increase efficiency by potential merges and acquisitions. There are several ways how to decide about their actions they should take to achieve these aims. On the one extreme, subjects rely on subjective factors, such as personal experience. Proponents of the other extreme seek to identify objective rules usually created by mathematics or statistics. There are tasks for which near-to-extreme approaches are suitable. Valuation of start-ups requires expert's insight and detailed analysis. On the other hand, high frequency trading systems demands strong computational requirements as they process billions of operations in a minute. Many tasks involve the assessment of the company's performance.

Company performance is a term which has to be analysed within some domain. In the credit risk domain performance is connected to a company's ability to generate enough funds to meet its obligations. In the long term investing, performance indicates whether the company achieves high and sustainable economic results. There are other perspectives, such as production or research performance. These performance indicators are reflected in the financial performance, which is considered as the top-level indicator of economic (financial) health of the company.

This dissertation does not cover all topics mentioned above. Instead, it focuses on multiple single-company models (MSCM). This category of models might be of interest for companies evaluating credibility of their business partners or for individual investors. A family of MSCM is a special case of portfolio models which does not explicitly account for inter-correlations between companies. This is the distinguishable feature from models which have stemmed from Markowitz's idea of correlated entities. The purpose of models presented in this thesis is to assess and predict performance (expressed by various means) of a particular company of interest.

The distinction between single company analysis and analysis of a set of companies is not as clear as it might look. An analyst who is concerned about performance of a single company builds up their model on the ground of some theoretical framework, such as Economic Value Added. Their model is tailored to capture unique details of the company. This evaluation suits only to one particular subject at a given time. After the performance is assessed its interpretation exhibits a low external validity because of company-specific adjustments. If the portfolio of companies would consist of only individually-evaluated companies, overall performance would suffer in situations when new information (such as price shock or unexpected export declination) is introduced to the market. This will be the effect of a low degree of generalisability.

If the companies' performance in their portfolio is assessed by some predefined set of rules (such as statistical or machine-learning procedures) learned from the data and supported by economic reasoning, performance behaviour of particular one company will not be described as precisely as if the single-company oriented model was used. However, this kind of evaluation compensates low internal validity in terms of external validity. Negative and positive effects of new information are more likely to cancel out as the model is more robust (if the model is correctly specified).

Both economics and finance are social sciences. Economic agents behave *irrationally* and cannot be analysed as entities in natural sciences. Many empirical models and economic theories rely on *ceteris paribus* condition which seems to be unjustified in reality. Yet, some theories and models can produce relevant outputs. In the following text I present a quantitative perspective on the financial modelling. It should help the reader to identify strong and weak characteristics of currently employed approaches. Also, the thesis pinpoints the diverse scope of the published literature and problems which make comparison of results and general understanding of the subject difficult.

2 CURRENT STATE – ECONOMIC PERSPECTIVE

This chapter presents several dimensions of the studied topic from the economic perspective. It aims to provide a clear overview of the studied framework at reasonable detail. Mathematical abstraction was suppressed for readiness of the text. Technical details can be found in chapter 3.

In accordance with the first research question, classification of tasks related to performance measurement of companies is made. In the following chapters economic theories related to corporate finance, accounting and domain-specific theories are discussed. These chapters provide top-level information background which is necessary to consider in many analytical tasks. The majority of theories and empirical evidence comes from the authors from countries with a developed capital and financial market. Such analysis often relies on availability of market data which might have a different information value, meaning and importance than in developing markets.

This chapter demonstrates the complexity of studied problem. Various assumptions and methodological stands which are used in literature might result in serious problems if they are not treated properly. Aforementioned aspects represent big problems for synthesis of information obtained through literature review. Also, it is unclear how underlying economic principles affect reported performance of analytical tools as this is not a concern of majority of papers.

2.1 Classification of Tasks

From the general point of view, tasks can be classified into categories:

- relation understanding (association, inferential statistics)
- determination of variable importance
- classification (discrete outcome)
- regression (continuous outcome)

The ultimate reason for performing any from the aforementioned tasks is to provide user relevant information for decision-making purposes. The task of *relation understanding* between variables is challenging because of the stochastic nature of studied problem. Rigour identification of cause and effect relationship requires fully experimental settings which is virtually impossible in social sciences. This identification problem can be split into two parts. The first relates to design of the study which is described in more details in in chapter 5.2. The second relates to analytical methods used in the process of estimation.

It can be concluded that in social sciences the causal effect can be only suggested by the researcher. It is convenient to apply regression technique to assess the strength of the oriented relation between variables if temporal occurrence of independent (x) variable predeceases dependent (y). In some cases instrumental variables can be included in the model. The effect of inclusion of the variable should not significantly affect the original relation. Other forms of relationship are analysed instead. Correlation is a measure of strength of linear relation. Association is a broader term which includes metrics of non-linear relations or relations between non-metric variables.

From the methodological perspective it is important to distinguish between statistical and a *substantial* significance of results. Using some analytical methods (such a hypothesis testing introduced in chapter 3.2) analyst can assess whether the evidence for/against the claim is strong enough or not in light of observed uncertainty. Results are considered as significant if they are very unlikely to be observed if the null hypothesis would be true. Substantial importance refers to the real-word meaning for particular decision maker. As an example of unclear distinction between two types of significance can be considered Bottazzi, Grazzi, Secchi, and Tamagni (2011, p.391). They find that indicator Leverage turns to be significant in the probability of the default model one year before the event of default occurs. This is demonstrated in Table 3 which contains parameter estimates accompanied with stars which indicate whether the corresponding p-value is lower than 0.01. Their probit model was estimated in the period of 1999-2002 and only the 2002 estimate was statistically significant.

Assessment of *variable importance* is crucial in constructing a regression or classification model. In descriptive models variables are usually pre-selected by the analyst according to the theory. Variable importance process seeks to identify variables and their position in a model (transformations and interactions) which improve model's performance. This selection can be made from the economic perspective or computationally (see chapter 3.1.4).

Classification and regression tasks require the creation of a model (pattern identification) on historical data which are useful in unseen, hold out samples. This can be out-of-time or spaces sample. As will be described later, huge problems represent transferability of results. These models are usually tailored to particular situations (current portfolio) or are closely related to the economic system.

The implication of the text above is that analytical techniques cannot uncover all relations among variables unless data design is in convenient form. It is possible to identify that companies which adopt some managerial techniques (such as Balance Scorecard) tend to outperform other companies. However, unless all other factors are fixed, analysts cannot conclude that the reason for higher performance is the managerial technique per se or just a fact that better companies use the tool. Moreover, this estimated effect can be statistically significant (estimated effect is around two times higher than corresponding estimation error) but the effect can be too small to have a practical impact (higher classification accuracy). In some tasks analysts can be interested in identification of important variables. Traditionally, many variables and their transformations and interactions were computed and used in the analysis. Computational advances allow identification of interactions, non-linearities or even reduce problem's complexity automatically. For example, it is possible to compute weighted composite indicator "returns" from several returns indicators through principal component analysis. Many tasks are post-hoc analysis which aim to understand what happened in the past. However, if model should perform on unseen data, different modelling strategies are usually employed.

2.2 Analytical Techniques for Performance Measurement

According to Çelik (2013) models can be divided from developing strategy into two categories. To the first group belong models which grow upon a sound theoretical justification. This family of models is named *theory-based models*. Second category is called *non-theory-based models*. Models of this type are created from the statistical and machine learning perspective with only a negligible influence from the economic theory. Çelik (2013) presents large literature review of models. He lists most frequent explanatory variables (financial multiples) and provides prediction accuracy summaries achieved in 122 research papers from 1960–2011. In this thesis methods are divided in similar way, but in more details are treated only non-theory based models:

1. Statistical models

Models which treat uncertainty in terms of probability are described in Chapter 3.3.

2. Machine learning and artificial intelligence models.

   Models that seek for hidden underlying data-structure and patterns. Although these models also uses probabilities in the process of identification and interpretation of results, probability itself is not the main concept. These models are, however, closely related to statistical methods. As an example can be use of series of logistic regressions (statistical methods) to create neural network (machine learning concept). This class of models is described in chapter 3.4.

2.3 Field of Study

According to the purpose, models and analysis can be divided into following fields (adopted and expanded from Bahrammirzaee (2010)):

1. Risk management – especially with focus on credit risk modelling
2. Portfolio management and investing
3. Prediction and planning
4. Other

The health of the company does not have a clear definition. The definition is usually closely related to the field of study. In the credit risk assessment field, the company is considered as health if the performance allows the company to meet its obligations in time and in the full amount. Although company's health reflects many dimensions (going concern principle, prospects of future growth, etc.) and can be observed on many levels (Figure 3). Financial health is considered as a top-level indicator which reflects state of other dimensions, which might be difficult to measure directly. Opposite state to healthy company is distressed company. This distress can be either short or long termed. From the economic perspective, early identification of distress resulting can be difficult to identify because of the flexibility of accounting principles or by managerial discretionary operations (chapter 2.5).

2.3.1 Risk Management

Risk occurs in situations when the subject is exposed to the situation with unknown outcome. Situations in which the outcome can be quantified in terms of the probability of occurrence and the magnitude is called uncertainty. Risk management seeks to minimize the difference between expectation and reality. Under the term risk therefore belongs both positive and negative outcome. If returns expected by analytical model turn to be lower than actual, observed returns, analyst underestimated the reality. The outcome in this situation is an additional profit. But from the perspective of risk management this represent a mistake. Risk management is used to predict a future as well. Traditionally, portfolio risk profile is described in terms of variance and Value at Risk at predefined level α for some horizon. Important step in the risk management is to develop *prospective* analysis which tells what to do to prevent losses, not only what the losses/gains will be in the future.

There is a large number of research papers about financial distress modelling and early-warning systems development. W.-Y. Lin, Hu, and Tsai (2012) consider bankruptcy prediction and credit scoring as two major research problems in accounting and finance domain. All of aforementioned fields involve process of classification and often employ the same (or very similar) classification methods and techniques. Therefore, they can be deemed as a members of the same discipline from the quantitative perspective.

According to Sun, Li, Huang, and He (2014, p.41) financial distress prediction (FDP) aims to "predict whether or not company will fall into financial distress based on current data, through mathematical, statistical, or intelligent models". Authors do not distinguish between financial distress prediction, bankruptcy prediction, business failure prediction or financial failure discrimination in they literature review. This is, however, common in the literature. Philosophov, Batten, and Philosophov (2008) emphasized different stages of financial distress and consequent bankruptcy. They proposes different design study to tackle this problem. Traditionally, companies are divided only into two subsets: non-defaulted and defaulted. This division omits bankruptcy trajectory and therefore decrease chances of correct prediction. Sun et al. (2014) summarise recently published original papers about financial distress modelling. They review and synthesise views about definition of financial distress, points to the best performing analytical methods, highlights sampling schemas and feature-space selection approaches. Their findings favour complex models from the perspective of predictive accuracy. They also note that usefulness of such models in real-world setting is limited because of the absence of clear interpretation of parameters which is often required by managers. Problem can also arise from abstract nature of model-estimation process. They suggest that models which can include non-financial models and expert supplied data were not developed enough compared to standard models which use financial ratios as a main source of information.

Credit risk modelling is the key topic discussed by practitioners and scholars during several last decades. Its importance was magnified recently. Malfunction of the credit risk policies is considered by many as a starter of financial crises in 2007. Banking portfolio and credit rating systems represent an important and innovative discipline. Portfolio models are used to gauge creditworthiness of the client and potential impact of failure on the whole portfolio. According to Majumder (2006) and Nouy (1999) risk management of banks exercise "universally accepted strategies" to:

1. establishing sound, well-defined credit-granting criteria,
2. adopting pricing strategy for new contracts and contract renewal,
3. fixing credit exposure limits to individual borrower,
4. developing a system for monitoring individual credit exposures
5. assessing the economic an regulatory capital

Despite the popularity and large body of literature, it is challenging to adopt results and recommendations from other sources. It is because of the different definitions used for terms health or default. In the context of credit risk modelling several definitions have emerged. Following definitions of E. Altman and Hotchkiss (2006), we distinguish between four generic terms used in literature almost interchangeably. *Failure* is commonly used when comparing any performance metrics which exhibit lower figures than expected ones (both in short and long period). *Legal failure* is a situation defined by some legislative document. When company cannot fulfil creditors' or stakeholders inquires, *business failure* situation occurs. *Insolvency* and *technical insolvency* are situations of insufficient amount of liquid assets. Insolvency is usually the immediate cause of bankruptcy. *Bankruptcy* according to Altman occurs when the fair valuation of assets does not correspond (in negative way) with amount of liabilities in long term. *Default* is the situation when any violation of credit agreement occurs.

Previous terms are commonly used within the context of corporate finance. Even in the banking industry, there is no unanimously accepted definition. In the Basel Committee on Banking Supervision (BIS, 2005) documents some extent of vagueness exists in the definitions. The definition used in Basel II utilises two conditions when meeting one suffices to declare the default situation. The first one refers to *subjective* “the bank considers that the obligor is unlikely to pay [in full]”. BIS recognise the risk of the latent, thus unobservable, conditions in the definition. These conditions are undertaken by risk-management assessment which incorporates subjective judgements. The second condition is *objective* and is stated as “the obligor is past due more than 90 days on any material credit obligation”. Violation of these conditions is identifiable by the set of predefined rules, such an average due time in last time-period, which can be controlled automatically, without human intervention. Company’s default might be the first signal to more serious difficulties or even to bankruptcy. As BIS (2005) add, “In practice, bank databases frequently do not include all historical defaulted facilities but only those facilities that ultimately resulted in a loss.”

According to Ross, Westerfield, and Jordan (2008) some signals of financial distress can be considered as actions which management practices to avoide declaration of bankruptcy. These actions are selling major assets, reducing capital investments to research and development, excessive exchange of debt for equity and negotiation with creditors and banks. They add that companies with more debt are more likely to face distress earlier. On the other hand, this gives them them more time for private workouts and reorganisation when compared to their competitors.

Sun et al. (2014) point to official reasons for bankruptcy in particular countries (i.e., Iranian companies are considered as distressed when their negative retained earnings exceeds 50% of their equity). Sun et al. (2014) also propose relative definition of distress which is connected to phase of company’s life cycle. According to Keasey, Pindado, and Rodrigues (2014) financial distress is recognised when three conditions are met simultaneously. First condition checks whether the size of funds generated from operational activities (proxied by EBITDA) can cover financial expanses for two consecutive years. To distinguish between short-term or seasonal decrease in cash flows from operational activities, financial structure of the company is analysed in the second condition. If the solvency ratio (net worth/total debt) and net worth decline between two consecutive years company is assigned as distressed. Although several distress definitions were proposed, majority of them rely on qualitative indicators. Homolka, Doležal, and Novák (2014) used definition closely related to managerial failure. This definition reflects negative public perception which arise when company is said that does not to fulfil its obligations. This happens when company is officially listed for bankruptcy in publicly available database. It should be emphasised that distress can have different intensity ranging from delayed covenant fulfilment to inability of repaying obligations. Moreover, the same conditions do not affect all companies the same way (large companies can liquidate more assets and meet obligations easier).

Companies which face economic distress often undergo informal reorganisation. As signs of informal reorganisation can be considered liquidation of assets or negotiations with creditors over existing debts. These changes project into financial statements and an early warning systems can be built upon them. It is, however, difficult to tell whether such changes are discretionary (management decides selling inventories as a part of new production strategy) or whether they are a consequence of financial distress (inventories are sold to improve cash-flow position). Gupta, Gregoriou, and Healy (2014) demonstrated that using both legal- and finance-based definitions of default to classify company as defaulted results in higher accuracy of classification model.

US legal system recognises bankruptcy when company is filled under Chapter 7/11. The same chapter is used for successful companies which use bankruptcy as exit-strategy. Distress situation are usually observed through market- or legal-proxies. The first group is connected to capital structure theory and costs of capital (Gupta et al., 2014).

Two distinctive approaches to credit risk modelling can be recognised (Majumder, 2006). The first group is called *micro approach* and aims to understand firm-specific credit risk. This path of analysis is also called equity based approach and was pioneered by Black and Scholes (1973) and extended by Merton (1974). The second type of models can be called portfolio-specific or rating based models. This approach can be traced to 1997 when CreditMetrics model was presented by JP Morgan's Risk Management Research division (Morgan, 1997). CreditMetrics Model (CM) recognises the difficulty of direct estimation of credit risk resulting from sparse occurrence of the default situation. CM model the risk indirectly by creating a *construct* which address of contract value volatility as a function of changes in creditor's quality. Both types of models require market information. This restricts their proper applications to specific economic environment with high volume of trades and large investors base. In other environment models accounting data are used as a main source of information.

Related field is credit scoring. Marqués, García, and Sánchez (2013) defines credit scoring as *a set of decision models and their underlying methods that help lenders determine whether credit should be approved to an applicant*. Result from the scoring is usually (Zhong, Miao, Shen, & Feng, 2014) alphanumerical label which describes the ability and willingness of rated company to meet its obligations. This rating grade is usually transformed into probability of default computed by internal rating system. These grades are than used in lending process (interest rates) and regulatory capital calculation. According to Moro and Fink (2013) lending process in can be divided into four categories:

1. financial statement lending
2. asset based lending (reflects quality of collateral)
3. credit scoring models (based on analytical models)
4. relationship lending (RL)

Only scoring models directly provide credit score. Outcome of remaining approaches is a score sheet. This sheet is usually further analysed by official guidelines to achieve objectivity. Also, first three approaches rely heavily on financial and other public data whereas relationship lending involves non-financial information in addition. RL uses private information and relies on trustworthy relationship between both sides. This type of commitment helps to reduce agency problems (problems between stockholders and management), lowering moral hazard and asymmetrical information. It also affects transactional costs and costs related to monitoring and control. (Moro & Fink, 2013). Given the absence of complete historical records and the fact that subjective factors affect performance of SME companies (such a personal characteristics of owners) more than large and well-established companies, *judgemental rating* based on experts' assessment seems to be the most convenient approach in SME lending process. (Angilella & Mazzù, 2015).

2.3.2 Portfolio Management and Investing

Portfolio management seeks to identify optimal asset allocation. This allocation should reflect time horizon and risk tolerance Bahrammirzaee (2010). Other concerns are portfolio performance benchmark, estimation of sensitivity to external (i.e, market and political risks) shocks, hedging and overall risk estimation. Overall portfolio risk is usually provided by calculation Value at Risk based metrics. Historically, portfolio models gain attention in US with the development of computational power. Idea of shared risk is due Markowitz (1952). In economies with developed financial systems CAPM model is used to estimate risk of individual company. It is made by estimation of β value which summarises the risk added by the investment if is included to the market portfolio.

All of aforementioned tasks require analysis of shared risk. This topic is too broad to be covered in this thesis as it is concerned only about multiple single-company models.

Before the model was presented market-risk tools were used, such a Value at Risk (VaR) or Expected shortfall (ES). These methods were successfully implemented and worked because of high liquidity on the market and frequently changing prices.

This type of analysis belongs to the MSCM framework only partially (it is located on the bottom of the pyramid in Figure 1)

2.3.3 Other

Ngai, Hu, Wong, Chen, and Sun (2011) provide extensive literature review on fraud detection in corporate and banking & insurance industry. Market risk is not considered in the thesis as it is mainly related to the changes in prices of financial instrument (stocks, bonds). Operational risk is a company-specific risk and its modelling therefore cannot be modelled by automatised procedures.

2.3.4 Source of Data

It is possible to divide information sources into three:

1. general level information
2. company-specific
3. expert knowledge

General level information describe macroeconomic and sectoral behaviour. Usually, aggregated values are analysed and further used in projections of long term cycles and trends. Hernandez Tinoco and Wilson (2013) studied importance of accounting, market and macroeconomic variables on bankruptcy prediction on longitudinal sample of 3020 non-financial listed British companies. They found that macroeconomic indicators Retail Price Index and the United Kingdom Short Term (3-month) Treasury Bill Rate Deflated improve classification accuracy only marginally. Macroeconomic indicators have higher importance in models which contain several sectors. It is reasonable to expect that in single-sector analysis companies will react similarly on the macroeconomics shocks. In multi-sectoral analysis hierarchical models can estimate sectoral and general effect.

Information on company level have to be treated in the wider context. It is due to Jensen and Meckling (1976) and their agency theory. Authors provide theoretical reasoning how company governance differs in situations where owners are also managing their company and when they delegate this task to managers (agents). Later (Jensen & Ruback, 1983) develops "hypothesis about conflict of interests" which explains empirically verified evidence that managers which owns high proportion of issued stocks often act inefficiently and their discretionary decision might lead to poor operational performance. Some authors (Lu, Wei, & Chang, 2015) include governance indicators such as Ratio of cash flows to voting rights, Management proportion (whether Chief Executive Officer (CEO) is also a board chair), Cross Holding (whether company owns securities issued by other listed company) and Director Ownership (percentage of director shareholdings) to their analysis to distinguish between management-ruled companies and other. H. Platt and M. Platt (2012) find positive association between number of independent board member and financial health. They do not find difference in proportions of CEOs who are simultaneously board members in healthy and bankrupted companies.

Expert knowledge eliciting received a lot of attention because collaboration of both domain expert and data analyst is present throughout the whole model development process. This process is more technical and will be described in more details in chapter 3.2.3.

2.4 Economic Theories Related to Financial Modelling

Financial theories should give analyst sound background in understanding of studied problem, assist in identification of important factors and pointing to assumptions and limitations.

Scott (1981) finds connection between bankruptcy theories and empirical models. He starts with *gambler's ruin theory* which explains behaviour of company when it meets loss. A natural solution is to sell assets. In Scott's *perfect-access model*, which was presented in 1976, company is allowed to sell assets but also either debt or equity on efficient market. Both theories rely on mean-variance paradigm introduced by Markowitz in portfolio theory. Variation in earnings indicator reflects higher risk which is compensated by higher mean of such an indicator. Under gambler's theory this indicator reflects risk of stockholders equity which contains liquidation values. In case of Perfect-access model this indicator estimates market value equity. Gambler's theory starts with the total size of capital K and with some empirically estimated probability of loss during some period. Company is defaulted when the $K \leq 0$. There were several attempts to employ this theory but results have been disappointing. (Scott, 1981)

Large body of research was dedicated to capital structure analysis and cost of capital. Seminal work of Modigliani and Miller (1958) (MM) shaped thinking of researchers for decades. They concluded that market value of the company is depends only income which company generates. One of implications is that capital structure is irrelevant. As the best theoretical solution is *corner* solution – company with 100% debt financing. Following implication is related to weighted cost of capital (WACC). WACC is, according to standard corporate finance theory, related to debt/equity ratio. According to MM theorem debt/equity does not affect WACC if conditions for theory are met (efficient market theory). The third statement concerns dividend policy and its relation to retained profit.

Dempsey (2014, p. 279) studied Modigliani and Miller prepositions. He concludes that "... focus on the Modigliani and Miller propositions as the foundation of corporate management has led to a stylized representation of corporate financial decision making that is far removed from reality".

Some authors (Keasey & Hudson, 2007; Dempsey, 2014; Coleman, 2014) point to the differences between “academic research” and practitioner activities. In Keasey and Hudson (2007) summarise the history of critical thinking in financial research. He argues that theoretic of financial science are working with the new evidence in a wrong way. Instead of considering new observation which contradicts the theory as a reason for refuting the idea of the theory, such observation is called “anomaly” or a “new fact”. Than, the researchers try to extend the framework of old theory to accommodate this “new fact”.

Melnyk, Stewart, and Swink (2004) point to problems of knowledge transferability from academia to practise. He claims that practitioner discuss different metrics because each side persuade different goal. Academics are more concern about clear definitions and valid outcomes (closely related to research questions and hypothesis), practitioners are more sensitive to time and cost needs. Thus, generalisability of findings is of less importance to them.

2.5 Accounting

Another discipline which affects the quality of model is the accounting system and principles. Accounting history has evolved in two accounting basis. Major differences can be found in two perspectives. The first one is recognition in time, the second lies in priorities put on accounting information quality.

1. Accrual accounting

   According to *A dictionary of accounting* (2010, p. 12) “accrual accounting is a system of accounting in which revenue is recognized when it is earned and expanses are recognized as they are incurred.” Most of accounting systems (including Czech Accounting Standards, CAS) use accrual base.

2. Cash based accounting

   A dictionary of accounting (2010, p. 81) defines cash-flow accounting as “a system of accounting that records only cash payments and receipts relating to transactions made by a business, rather than when the money is earned or when expenses are incurred.” Cash based method focuses on cash incomes and expenditures thus describes the current situation without any assumptions about future business realisation. Reliability of the method is high - it accounts only actual transactions with known values. The major method’s pitfall is low degree of *faithful representation* characteristic. As an exampled can serve violation of the *matching principle* in case of assets depreciation over time. The only entry to the book is when the asset is paid, not used over time.

Another distinction can be found in the way accounting principles are created. International Financial Reporting Standards belong to the *principle-based* accounting. Generally Accepted Accounting Standards (GAAP) US is *rule-based* approach.

The conceptual framework of IFRS Mackenzie, Coetsee, Njikizana, Chamboko, and Colyvas (2011) defines two fundamental characteristics of financial reporting information:

- Relevance – Information provided by financial reports has to be relevant to the decision makers.

- Faithful representation – Information content is true and fair from mistakes (misstatements).

These characteristics must be accompanied by other qualitative features if the financial information should be useful. These are:

- comparability – Comparability should govern that financial information is comparable within the accounting system and does not change significantly over time.
- consistency – It is expected that adopted accounting methods are used over long period.
- verifiability – Informed analysts who analyse financial reports independently should reach very similar conclusion.
- timeliness – Information are not out-dated and are still relevant for today's decisions.
- understandability – Reports should be clear enough to be understood for analyst with some background of finance and economy.

Definitions can be found in Mackenzie et al. (2011) with reference to other concepts which were not included in the final proposal. Interestingly, term *high quality* - which can be "achieved by adherence to objective and qualitative characteristics of financial accounting" (Mackenzie et al., 2011, p.13) was not accepted.

Czech GAAP is directly oriented towards tax assessment and other legislation. Major systems, IFRS and US GAAP are more concern about the process of transaction, not only about the legal form. This is referred as Term Substance Over Form (SOF). SOF principle can be demonstrated on financial leasing example or on recognition of revenues. Even though company does not own leased assets, company benefits from its usage. Under IFRS this assets become part of the lessee assets. Accrual accounting (Czech GAAP) recognise revenues even if the corresponding cash-flow is not confirmed. SOF principle, according to IFRS, governs that accounting is reliable. These differences can lead into different importance of economic variables used by other authors.

It is assumed that the most important shareholders' source of information are accounting documents. According to Agency Theory (Jensen & Meckling, 1976) company can be viewed as a path of connected entities which make efficient contracts to maximize shareholders value. Shareholders use accounting documents to analyse connections and contracts to mitigate redundancy processes. As Hassab Elnaby, Mohammad, and Said (2010, p. 56) state "in some instances managers use their accounting discretion to achieve predetermined results to the point where financial information no longer reliably represents a company's underlying economic condition." Literature refers to these practises as to *accounting choices*. But, the concept of accounting choice is ambiguous. According to Fields, Lys, and Vincent (2001, p. 256) term "accounting choice" (AC) is *any decision whose primary purpose is to influence (either in form or substance) the output of the accounting system in a particular way, including not only financial statements published in accordance with GAAP, but also tax returns and regulatory filings*. This definition was criticised by Francis (2001) who disagree with the broadness of the definition. Francis restricts decision makers only to managers responsible for the choices. Other subjects, such as auditor with alternative standpoint on appropriateness of accounting method used by company, does not subject to this definition.

2.5.1 Accounting Choices

Flexibility of accounting systems allows better capturing of the reality. On the other side, it creates room for inappropriate adjustments or even intentional misconduct in selection of accounting

Tab. 1: *Motives for Earnings Management and resulting rewards (Mulford & Comiskey, 2002).*

Category	Rewards
Share-price effects	Higher share prices
	Reduced share-price volatility
	Increased corporate valuation
	Lower cost of equity capital
	Increased value of stock options
Borrowing cost effects	Improved credit quality
	Higher debt rating
	Lower borrowing costs
	Less stringent financial covenants
Bonus plan effects	Increased profit-based bonuses
Political cost effects	Decreased regulations
	Avoidance of higher taxes

methods. There is no generally accepted definition which would cover all aspects of accounting choices. Literature frequently refer to the accounting choices as to *Earning Management*. Several definitions of Earnings Management can be found in the literature. Schipper quoted in (Fields et al., 2001, p. 260) defines EM as "implementation that impairs an element of decision usefulness or implementation that is inconsistent with intent of the standard". According to Fields et al. (2001) this definition and approach does not properly address the encumbrance of managers' intents thus should be replaced by Watts's definition introduced in Watts and Zimmerman (1990). Watt's defines EM as a situation when managers exercise their discretions over the accounting numbers with or without restrictions. Watt divides the discretions' objectives in firm value maximizing and opportunistic. Healy and Wahlen (1999) define EM as occurring when "managers use judgement in financial reporting and in structuring transactions to alter financial reports to either mislead some stakeholders about the underlying economic performance of the company, or to influence contractual outcomes that depend on reported accounting numbers".

Earnings management appears in other fields of study in which analytical tools are employed. Fraud detection (Perols & Lougee, 2011; Brazel, Jones, & Zimbelman, 2009) or analysis of accounting conservatism (Badertscher, Collins, & Lys, 2012) are some of them. Motives for exercising EM are reviewed in Table 1.

2.5.2 Discretionary operations

The large body of literature of accounting research deals with the discretionary accounting choices. The topic is crucial with respect to aforementioned trade-off between reliability and relevance.
Accruals are considered as a proxy for degree of discretionary accounting. Analysis of accruals can be projected into three perspectives:

1. informational – discretionary managerial actions reveal private, implicit information and expectations about the future,

2. opportunistic – these actions are exercised to bend true economic performance to meet investors' forecasts and expectations,

3. contracting – the accounting is manipulated to improve company's position in supplier-customer relationship. The advantage is then (at least in short-time period) projected in lower contracting costs.

The first incentive is made for the purpose of higher relevance and representational faithfulness. Opportunistic and contracting perspectives decrease the ability of future predictions by incorporating bias to the reported figures. Fields et al. (2001) divide the consequences of AC into three main categories:

1. agency costs – costs assignable to contractual issues, such a debt covenants managing
2. information asymmetry – this consequence arise when some subject operates with better (in terms of quality) information than other (less well informed) parties.
3. other externalities – hidden consequences, such worsening of a partnership(s) or trustworthiness of company.

Badertscher et al. (2012) address the topic of accruals with connection to cash flow prediction.

2.5.3 Definition and importance of Earnings

Earnings are the net income or profit of a business. This accounting figure might be considered as the cornerstone for analysing company's performance (i.e., as part of the EPS indicator) and also as a determining variable of total tax liability. Different accounting/reporting standards (i.e. International Accounting Standard 33 or Statement of Standard Accounting Practice 6) can have different composition of the top aggregated value. (*A dictionary of accounting*, 2010).

The discussion about earnings grown in the last two decades. According to DeFond (2010) discussion was driven mainly because of Security and Exchange Commision (SEC) critique on auditing a rating agencies and because of "high profile accounting frauds" which resulted in the 2002 to the acceptance of Sarbanes-Oxley Act.

Dechow, Ge, and Schrand (2010) in their seminal work discuss different proxies for earnings quality, including "persistance, accruals, smoothness, timeliness, loss avoidance, investor responsiveness and other external indicators". Connection between earnings quality and earnings management was analysed by many authors. Lo (2008) states that absence of only small influence of earning's management does not guarantee quality of earnings. He also make connection between quality of accruals and accounting figures.

Burgstahler and Dichev (1997) hypothesised that practise of avoiding annual loses by managers' discretionary acts can be detected by inspections of earnings distribution. If earnings would not be affected by other factors than "objective" company's performance, there would not be any breaks and the distribution would be smooth. In the Figure 2 distributions of earnings suggest that on the operational level companies report more symmetrical and smooth earnings than in the distribution of net income. This suggests that the financial and extraordinary part of net earnings contain discretionary operations.

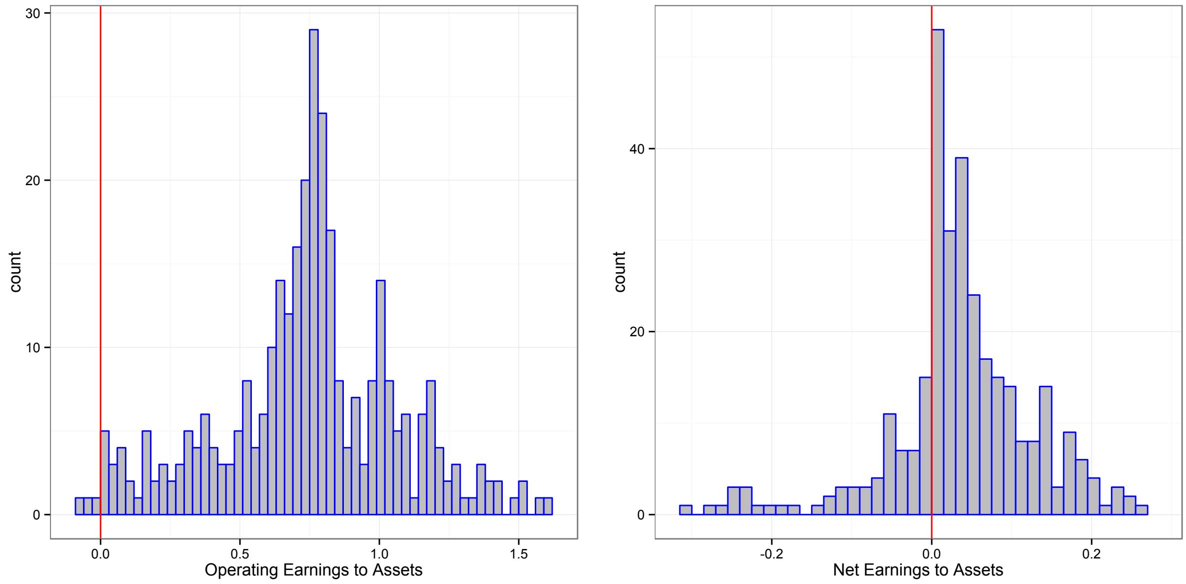

Fig. 2: *Two plots demonstrate distributions of operating earnings (left) and net earnings. Earnings were deflated by assets to achieve comparability of results across companies with different sizes. Source: Own*

2.6 Corporate Finance Theory

Field of corporate finance has received large attention in the 20th century. These theories were primarily concerned about:

1. Measuring reality of corporate performance
2. Valuation and capital budgeting
3. Measuring and hedging risk,
4. Capital structure, dividend policy and long-term investing
5. Short-term financing
6. special topics

This list is adopted from the Ross et al. (2008).

Analysts can use internal or external data for their decisions. Internal data usually comes from the financial statements. Although legislative differ in economic regimes, three types of document are widely available:

- balance sheet
- profit & loss statement
- cash-flow statement

R. T. Lee (2013) distinguishes between metrics and measures. Both are values used for analytical purposes. Measures are values (such a revenues or space capacity) measured directly by standardised and objective approaches. Metrics are more subjective values and are derived from measures (price to earning ratio). Subjectivity is mostly linked with its interpretation. Lee suggests 3C rule for constructing appropriate metric:

1. clear – clearness governs its applicability and guidance how to obtain better results. Its purpose and assumptions must be clearly articulated.

2. concise – with rising metrics' complexity declines transparency of individual measure contribution.

3. correctly modelled – term "per" might suggest a direct causal relation while measures are only correlated.

R. T. Lee (2013) recommends not to mix financial and operational measures to create financial metrics. Financial metrics usually focus on financial health, efficiency, effectiveness, and growth. Operational metrics typically focus on process performance, capability and capacity.

Two major approaches to measuring economic performance are *standard* and *value based* approach.

2.6.1 Standard Approach

Economics multiples were heavily used in financial literature since the first modelling attempts. As Brealey, Myers, and Allen (2003) suggest, financial ratios help as to ask right questions, not answer them. Brealey et. al suggested five classes of ratios to measure:

1. Leverage
2. Liquidity
3. Efficiency
4. Profitability
5. Market based value

Leverage ratios describe debt repay prospects - how much debt managers used with respect to company's ability to pay for them. It is advised to use accounting value of equity instead of a market value. Market valuation contains values of intangible assets (investors take into account reputation, promotion abilities, staff training, etc.) which are not (easily) saleable and cannot be used in obligation fulfilling. As an example of leverage ratio based solely on balance sheet data can be used debt ratio:

$$\text{Debt ratio} = \frac{\text{long-term debt + leases}}{\text{long-term debt + leases + equity}}$$

Short-term leverage analysis is also important especially when company faces cash-flow problems. This can be assessed by using income variables which describes company's earnings power:

$$\text{Times-interest-earned} = \frac{\text{EBIT + depreciation}}{\text{interest}}$$

Liquidity ratios describe company's ability to find assets which can be used in transactions. Liquid assets' values are more reliable because their valuation might be almost continuous (value of money on bank account) or can be valued frequently (other operating assets). As Brealey et al. (2003) notes, this brings one large drawback - values and volumes change very often and monitoring must be done accordingly. Several metrics were introduced to analyse liquidity, such a current ratio, acid-test ratio or cash ratio:

$$\text{Cash ratio} = \frac{\text{cash + short-term securities}}{\text{current liabilities}}$$

Efficiency ratios are used to estimate how efficiently the firm uses its current and fixed assets.

$$\text{Sales-to-Asset ratio} = \frac{\text{sales}}{\text{average total assets}}$$

Sales-to-Asset ratio is one of the general metrics of efficiency. It is recommended to work with scaled assets (averaged over two years) because it's used with connection to flow measure (cumulative value over time period, here sales). This ratio answers how much generate one unit of assets employed. However, this ratio cannot describe effectiveness of current and fixed assets, how was sales achieved - was capital used to produce hight volume products or only to produce few pieces of highly priced assets? Several multiples were designed to answer these detailed questions, how to measure use of inventory (Days in inventory ratio) or how often customers pay their bills (Average collection period).

Profitability metrics estimate earning-generation power. Profitability metrics can be considered as a crucial metrics for investment decision because it allows comparison of earning potential and corresponding risk. Brealey et al. (2003) uses profitability metrics in two ways - as a margin ratios:

$$\text{Net profit margin} = \frac{\text{EBIT} + (\text{tax + interest tax shield })}{\text{sales}}$$

or as a return ratios:

$$\text{Return on equity } = \frac{\text{earnings available for common stockholders}}{\text{average equity}}$$

For a return multiples Damodaran (2012) measuring accounting returns, Damodaran points to different use of ROE, ROCE and ROC, and ROA.

- ROE metric measures how much return is being created by equity holders. This metric is compared to cost of equity or to growth rates (used in dividend discount and FCFE models).

- ROCE and ROIC measure operational earnings therefore can be used in FCFF model. Returns rates should be compared to overall cost of capital.

- If net income or operational income are used to compute ROA then metrics is not directly comparable to either cost of equity or capital. In this case ROA is not recommended by Damodaran for valuation purposes. This does not apply when earnings before interest and taxes is used. In Altman's model (1968) ROA is used to depict true productivity abstracting from any tax or leverage factors. Altman justifies its use because assets' earnings power is an ultimate condition for company's existence. It must be appropriate for corporate failure studies, too. Rate of return for valuation is usually limited only to returns from core, operational activities. As a recommended indicator for Value concept used by McKinsey, Company, and Koller (2010) is ROIC. Return on net assets (RONA) is used for estimating Economic Value Added (EVA).

When company do not internalise dividends payout ration can be appropriate profitability metric.

$$\text{Payout ratio} = \frac{\text{dividends}}{\text{earnings}}$$

This ratio connects profitability ratios valued "from inside" to "outside" profitability ratios (market based). A ratio which sets how much much are investors willing to pay for one unit of earnings is called P/E ratio. Dividends can be tightened to stock price (Dividend yield ratio) to measure performance to equity holders. Stock price to one book value per share (Market-to-book) is another ratio serving similar purpose, especially when dividends are not paid.

2.6.2 Value Based Analysis

Although valuation is an important part of financial modelling, only relevant topics to the MSCM framework are considered. Models which require large number of company-specific adjustments, such as Economic Value Added, are not detailed.

Value is a long-term measure of company performance which contains information useful for all stakeholders, not only to shareholders. According to McKinsey et al. (2010), companies which aims to maximize value create more employment opportunities, increase both employees and customers satisfaction and behave better in terms of social responsibility. The major problem with value-based concepts is their uniqueness in application. Although general rules are applied, details need to be tailor to specific company to capture different accounting valuations, depreciation techniques, non-operating assets, etc. Holler (2009, p. 1) states "Only measures based on official financial data, publicly accessible, and on objective and consistent methodologies can be used effectively by and communicated to the financial community, shareholders, and other stakeholders, to monitor management performance assess investment and evaluate corporate value creation." The lack of detailed information strongly undermine possibility of wide application. These methods can be successfully used in countries with developed financial systems. In this economic environment financial data are being published quarterly and undergo detailed revisions from the market participants. This also raises a question of applicability and transferability of findings from authors operating in these systems to other economic systems.

As the valuation concepts were developed in aforementioned economic systems, financial data takes important role in almost all concepts. As the simplest indicator of growth in corporate value can be considered change in stock price. Stock price change is a measure of added shareholder wealth, but it does not necessary convey information about created value by virtue of managers operations. This addition to stakeholders wealth can be driven by more general condition, such a industry or global economy sentiment. Thus, metrics based on market variables should be considered only as an indirect way of estimating value.

Several techniques for estimating value of corporation have emerged from both academics and practitioners. Holler (2009) founds differences between indicators which should be used if the general principles of value theory would be followed and actual managers' opinions on ideal performance indicators. According to valuation theory, future cash flows determine changes in stock prices. Managers, though, usually consider net earnings as a key performance indicator. This is not surprising findings, but as Tsay, Lin, and Wang (2009) claims, the strength of earnings response coefficients (which aim to describe relation between earnings and returns) is only weak; only a small part of variation in stock prices can be explained by accounting earnings (authors point to value-irrelevant components of earnings and imperfection of earnings measurement).

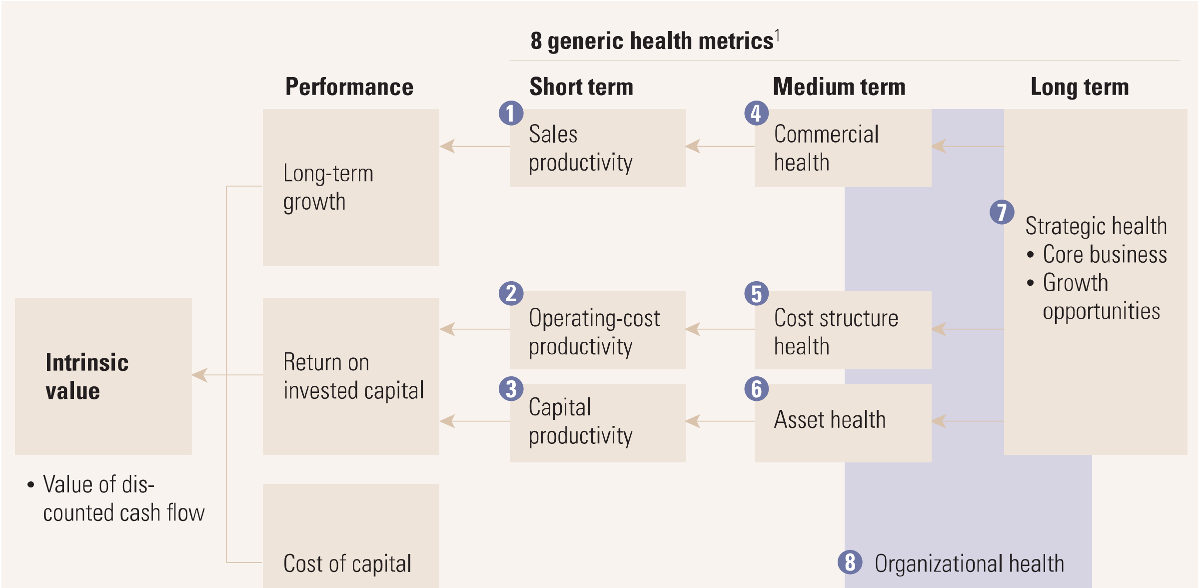

Fig. 3: *Scheme connects intrinsic value of the company with its health indicators. These indicators measure value-drivers which generate cash flows at cost of capital. Source: Dobbs and Koller (2005)*

Earnings need to be analysed further to reveal value-generating potential. Positive earnings bring positive effects to shareholders only when cost of equity is lower than the net earnings. When the cost of capital (both equity and debt) is subtracted from net we got more meaningful indicator, residual income (RI). Magni (2009) presents different indicator named residual income. Latter residual income is defined as the difference between income and opportunity costs. The logic behind its constitution remains the same, but opportunity cost make it more open to comparisons with similar businesses and therefore can be used in benchmarking tasks as well. RI is usually adjusted to capture industry level or general systematic risk. Then, the industry excess rate is computed. Importantly, RI is not a perfect measure as it is affected by non-cash accruals and is vulnerable to changes and differences in accounting principles (Holler, 2009). From the idea of RI has stemmed several other concepts. One of the most popular was introduced in 1989 by Stern Stewart & Company and is called Economic Value Added (EVA). This method seeks to identify assets needed for running core business, it's price reflected in weighted capital costs and profit from core business. S. Lee and Kim (2009) summarises findings on performance of EVA models (both positive and negative) and points to other concepts which modify or extend original EVA (such a Refined Economic Value Added or Market Value Added). Holler (2009) introduces models which can address problem of different depreciation methods used by companies. These models are Cash Residual Income and Cash Economic Value Added.

The core idea of value creation is that company utilises investors' capital to govern future cash flows in future at rates of return exceeding the cost of capital. Combination of revenues, return on invested capital and cost of capital creates value. This is also known as a conservation principle - anything what doesn't increase cash-flows doesn't create value. (McKinsey et al., 2010) Following this definition, earnings are not considered as the main driver. High value of return rate is not sufficient condition to long-term value creation. High returns have to be supported by stable revenue growth. This value driver combination generates stable cash flows. Total amount of cash flow is an important indicator and is relatively easy to calculate. But, its value must be treated carefully because it does not convey an information about cost of capital used to generated cash flows. Cost of capital is usually measured by weighted average cost of capital (WACC).

According to Brealey et al. (2003) profitability ratios are used to measure the firm's return on its investment. While it might not be obvious in countries with underdeveloped financial markets, investors decide on their investments based on spread between financial market return and company-specific investment return (reflected in an equity premium). Brealey introduces rate-of return rule which states that investors accept investments which exceeds their opportunity cost of capital (while accounting for risk in discount rate). While some stakeholders might be interested in overall performance of a company, some might be interested only on their equity share.

When computing return performance of the company, earnings must be linked to underlying assets to reflect interest of particular stakeholder. Net income is a profit to equity holders (therefore used for Return on Assets ratio) while NOPLAT (used in Return on invested capital) is a profit available to all investors (McKinsey et al., 2010).
Return on invested capital (ROIC) is a measure of earnings from core operations to invested capital (described below). It is used in two forms, as a return to invested capital or only to the change in invested capital from specified time period (so called incremental capital.)

$$\text{ROIC} = \frac{\text{NOPLAT}}{\text{Invested capital}} \tag{2.6.1}$$

Net operating profit less adjusted taxes (NOPLAT) can be defined in terms of Free Cash Flow (FCF) as:

$$\text{NOPLAT} = \text{FCF} + \text{Net Investment}$$

$$\text{NOPLAT} = \text{FCF} + \text{Investment in Invested Capital} - \text{ Noncash Operating Expenses}$$

NOPLAT can be expressed from income statement in three steps:

1. Operating profit should not contain interest. Interests are considered as financial expanses.
2. After tax operating profit should be cleaned from those earnings which were created by other assets not considered in invested capital.
3. Actual taxes should be recalculated to theoretical one which would be in situation of all-equity financing.

RONA is a ratio of net operating profit after taxes to net operating assets, NOPAT/NOA.)

Invested capital (IC) used in (2.6.1) represents total capital required to fund operations. When deriving IC no connection to the external or internal financial sources are considered. Invested capital can be derived as:

$$\text{Operating Assets} - \text{operating Liabilities} = \text{Invested Capital} = \text{Debt} + \text{Equity}$$

Invested capital consists of operating working capital, fixed and intangible assets and net long-term operating assets. Long term operating assets are long terms assets excluding non-operating assets, such a deferred taxes, non-consolidated subsidiaries, etc. Intangible assets are problematic assets not only because of their valuation essence, but also from accounting viewpoint (difference in IFRS and domestic accounting standards as can be seen in goodwill).

Free cash flow (FCF) is the cash flow after taxes which is available to all investors (not only to equity holders). FCF is independent of financing (interest and tax shield is considered as a financing cash flow) and non-operating items (McKinsey et al., 2010):

$$\text{FCF} = \text{NOPLAT} + \text{Noncash Operating Expanses} - \text{Investments in Invested Capital}$$

alternatively:

$$\begin{aligned}\text{FCF} = {} & \text{EBIT}\,(1 - \text{Tax Rate}) + \text{D\&Amortization} - \Delta\,(\text{Net Working Capital}) - \\ & - \text{Capital Expenditure}\end{aligned}$$

Tab. 2: *List of ratios used by Beaver (1966)*

Group of Ratios	Number of Ratios
Cash-Flow	4
Net Income	4
Debt to Total Assets	4
Liquid Assets to Total Assets	4
Liquid Assets to Current Debt	3
Turnover	11

Tab. 3: *Bayesian decomposition provided by Beaver (1966)*

Event	Prior	Likelihood	Joint Probability	Posterior Probability
Fail	$P(F)$	$P(R\|F)$	$P(R,F)$	$P(F\|R)$
Not Fail	$P(\bar{F})$	$P(R\|\bar{F})$	$P(R,\bar{F})$	$P(\bar{F}\|R)$

2.7 Genesis of Analytical Models

Beaver's model (1966) can be considered as a first of analytical models which used financial ratios and simple statistical tools. In his empirical work Beaver points to crucial assumption which are still valid today. He states on page 71: *"The usefulness of ratios can only be tested with regard to some particular purpose"*. Here Beaver refers to economic purpose (such a credit-worthiness or loan application procedure). From methodological perspective following two statements: *"The emphasis upon financial ratios does not imply that ratios are the only predictors of failure."* (page 72) and *"The primary concern is not with predictors of failure per se but rather with financial ratios as predictors of important events-one of which is failure of the firm. Further, the primary concern is not with the ratios as a form of presenting financial-statement data but rather with the underlying predictive ability of the financial statements themselves."* Here Beaver emphasises the difference between descriptive and predictive feature of models. This problem is closely related to multicollinearity problem resulting in loss of descriptive power while prevailing the predictive power. Another methodological statement refer to usefulness of results (page 72): *"Strictly speaking, inferences drawn from this study apply only to firms that are members of the population."* and *"The answer (on usefulness) can only be provided by a replication on other populations. However, the findings here can provide some evidence regarding the extent to which such a replication would be worthwhile"*. These refer to an extent in which results can be used (in terms of statistical terminology closely related to the sampling process).

Beaver (1966) used longitudinal data design of length 5 years before failure (defined as (p. 77):*"Failure is defined as the inability of a firm to pay its financial obligations as they mature. Operationally, a firm is said to have failed when any of the following events have occurred: bankruptcy, bond default, an overdrawn bank account, or nonpayment of a preferred stock dividend."*) His sample consisted of initial 158 pair-matched companies and ended with 117 (54 failed + 63 non-failed) companies. These companies were described using 30 ratios shown in table 2.

These ratios were selected as initial. After their discriminative power was estimated (large differences in group mean values indicated high discriminatory power) only 6 ratios (one from each group) were used. These ratios served as an proxy for likelihood of failure. What makes Beaver's article important from statistical point is the section *Likelihood ratios*. Bayesian analysis is presented in its essential nature in the Table 3.

Probability of Failure is denoted as $P(F)$ and not-failure as $P(\bar{F})$. Variable R stands for ratio. According to Beaver likelihoods can be read directly from histograms. Although Beaver did not write it explicitly, this method is an elementary machine learning method Naïve Bayes (described in chapter 3.3.1).

In his seminal work E. I. Altman (1968) defended using financial ratios as a tool for assessing business enterprise performance. Altman agreed that relying on "rules of thumb" regarding recommended values for ratios is outdated practise. On the other hand, if financial ratios are treated appropriately they can reveal hidden information about the investigated phenomena. His model objective was to predict bankruptcy (but this term was soften in the paper to "business failure" with no deeper interpretation). Altman pointed to the problem that the most financial models relying solely on financial ratios which are analysed as a set of univariate cases (analysing several ratios separately). He writes on the page 591:
"In almost every case, the methodology was essentially univariate in nature and emphasis was placed on individual signals of impending problems. Ratio analysis presented in this fashion is susceptible to faulty interpretation and is potentially confusing. An appropriate extension of the previously studies, therefore, is to build upon their findings and to combine several measures into a meaningful predictive model. The question becomes, which ratios are most important in detecting bankruptcy potential, what weights should be attached to those ratios, and how should the weights be objectively established. "

It should be highlighted that Altman from the beginning describes his model as *predictive model*, therefore he does not justify weights values in economic terms (as would be in descriptive model). Although Altman is usually referred as the pioneer of application Discriminant Analysis in finance, in his paper E. I. Altman (1968) can be found several references to other authors (mostly for loan application problems). Altman used multiple discriminant analysis (its linear form) which is described in chapter 3.3.2. Sample consisted from 33 bankrupted and non-failed companies during the period of 1946-1965. Non-bankrupted companies were selected by pair-matched design. Initially, 22 financial ratios were selected but only 5 appeared in the final model. Variable with the highest discriminatory power used by Beaver (1966) - Cash flow to debt - was not used by Altman, because (p. 594) *"of the lack of consistent appearance of precise depreciation data"*. Variables X_1 = Working Capital / Total Assets, X_2 = Retained Earnings / Total Assets, X_3 = ROA, X_4 = Market Value of Equity / Book Value of Total Debt and X_5 = Sales / Total Assets were used instead. Altman's Z-score model takes form of:

$$Z = 0.012X_1 + 0.014X_2 + 0.033X_3 + 0.0006X_4 + 0.999X_5 \tag{2.7.1}$$

Several adjustments for the Z-score model were proposed. Namely, variable X_5 is removed in some models to mitigate industry effects. E. Altman and Hotchkiss (2006) present validation of scores by computing mean values and estimated Z-scores and compare them to U.S. Bond Rating (Bond Rating Equivalent method). This methods allowed to develop Z-score for emerging markets. Other reperametrisation allowed better capturing of industry specifics and other factors (such a estimation of going concern principle in audit-oriented literature) (Carcello & Hermanson, 1995).

2.7.1 Review of Altman's Model

Altman's model is arguably the most influential model in the history of corporate defaults modelling. For this reason this section presents review of applicability of Altman's model and results from replicated studies.

Deakin (1972) has replicated both Beaver's and Altmans' model and offered some suggestions for improvement. One of these improvements were better companies-matching scheme. Then, he made a series of Spearman correlations to gauge similarity between financial multiples of non-defaulted and defaulted companies. Differences in capital structure changes (increase in amount of preferred stocks and other debts while lowering retained earnings) resulted from this analysis. As Deakin (1972) note, this form of financing is rather long-term than short-term oriented and is more likely projected in plant a equipment investment than in raising liquid asset position. Moreover, this form of capital bears higher costs. Deakin highlights an absence of this phenomena in Beaver's (1966) paper. Scott (1981) reviews Deakin's own model and concludes that this model has high classification accuracy in 3-5 year prediction horizon.

Grice and Ingram (2001) analyse usability of Altman's model in the current practice. They state that commercial banks use models developed using DA in the periodic loan review process and also in managing portfolio. Due to its clear structure and interpretation of results it is used in operational and auditing processes. Grice also replicated original Altman's (1968) model and tested three additional hypothesis about its usefulness. He brings clear evidence that results (in terms of proportion of correctly assessed companies) do not remain stable over period spanning 30 years. They use both original and recalibrated coefficients of Altman's model. Results obtained on data from 1988–1991 show significantly lower predicting accuracy. Secondly, they find that the original model does not perform well for both non-manufacturing and manufacturing companies (only large manufacturing companies were considered by Altman). Accuracy in manufacturing industry reaches 69.1 % and in non-manufacturing sector only 57.8%. Lastly, they results suggest that Altman's model is useful the same way for predicting financial stress conditions as for bankruptcy prediction.

2.7.2 Contemporary Development

Authors continued to develop model using logistic regression (both logit and probit) in 1970's. Important paper which addressed default modelling problems was due Zmijewski (1984). Zmijewksi highlights flaws in sampling techniques used in current distress prediction models. In his paper two major admonitions to non-random sampling are presented. *Choice-based sample bias* occurs when a probability of an observation being in the sample depends on the values of dependent variable. This scheme leads into reality over/under estimating because created sample do not correctly describe proportions of classes in the population. In context of discriminant analysis, constant term is affected. On the other hand, constant term serves as an threshold value for deciding of classification and usually subjects to optimization (through cost matrix). As he adds, in the most specifications of probability type models (such a logit model) the constant and other coefficients are asymptotically biased. The second problem is *complete data sampling*. Statistical methods expect data obtained by exogenous random sampling scheme - observations are randomly drawn with no respect to dependent or independent variables. Zmijewski (1984) finds (by employing correlation analysis) that firms with high default probabilities have low probability of having complete data. Zmijewski concludes that an impact of these biases were not substantial in general (descriptive power), but the prediction rates on the individual group classification were significantly affected.

In the 80's new approach, named *machine learning* (ML), become extensively used. Models arising from concepts of decision trees or cluster analysis allowed faster learning, improved parameter tuning and better model-to-data adjustments. These methods did not relied on asymptotic assumptions to same extent as statistical methods did. Work of Efron (1979) meant breakthrough in computational science as the bootstrap technique was introduced.

The first survival model which received attention of academia was hazard model estimated by Shumway (2001). In this paper Shumway criticises *static* models and suggests longitudinal data design. He claims that single-period models exhibit inconsistent estimates. He tests financial multiples for their predictive power. Half of originally significant ratios turned to be statistically insignificant in the hazard model. He identifies as statistically significant market size, stock returns and idiosyncratic return variability. Philosophov et al. (2008) developed Bayesian forecasting rules utilizing financial ratios and “maturity schedule” variable. His model which was identified on longitudinal data outperformed in terms of predictive performance Altman’s Z-score (on 1-3 prediction horizon).

As the necessity for ”highly-optimised”solutions rose, standard ML techniques become insufficient. New field of *Artificial Intelligence* (AI) was proposed as an answer. With AI methods more abstract solutions were introduced. Increasing complexity yield in loss of model’s description feature. Milestone in financial modelling was achieved with implementation of artificial intelligence techniques. Concept of neural nets was introduced in 1940‘ but its first simple implementation appeared twenty years later. This delay was caused by methodological problem which was solved by developing backpropagation algorithm. Initially, there was no method available which would allow learning hidden layer and only single layer neural network was possible to learn. This kind of network can only classify linearly separable cases, which are rare in practice. Odom and Sharda (1990) introduced single-layer neural network model in financial modelling. They fit LDA model on the same covariates used in original Altman model. Although their model was identified on dataset (with 71 companies), it took around 24 hours to fit three models (authors adjusted training set to assess robustness: model 1 with 50% healthy / 50% bankrupted, model 2 with 80/20 and model 3 with 90/10 proportion) with only five neurons. They concluded that NN tend to outperform LDA model as the original dataset becomes imbalanced. Their final model classified 77.78% bankrupted companies correctly whereas LDA only 59.26%. Application of neural networks was systematically reviewed by Atiya (2001). Self Organizing Map (SOM) is a NN extension which allows projection on topological map and allows better interpretation. Chen, Ribeiro, Vieira, and Chen (2013) present two-stage application which allows reflecting time changes using bankruptcy trajectory. Huang, Zhu, and Siew (2006) introduced Extreme Machine Learning approach which increases learning speed of feed-forward neural network. Nowadays, papers oriented on NN modelling are more technically oriented (estimation efficiency, parameter optimisation, over-fitting prevention) or NN models are for models comparisons.

Concept of highly flexible classifier Support Vector Machine (SVM) was introduced by Vapnik (1995). Some authors (Y.-C. Lee, 2007) found better classification accuracy of SVM in credit risk (bond) analysis than LDA, case-based reasoning (CBR) or back-propagation neural networks (BPN). Zhong et al. (2014) compares learning efficiency of NN and SVM and agrees with previous results that suggest that in credit scoring (in both retail and corporate area) SVM tend to over-perform NN back-propagation model in terms of accuracy.

Neural networks and SVM are used in the time series modelling as well. Coakley and Brown (2000) present classification scheme of tasks related to various research questions and list notable papers. For assessing relationship between variables in time-series context, NN were used in conjecture with Box-Jenkins methodology on stock market valuation, cost estimation and in managerial forecasting studies. Classification tasks involved fraud detection, bank failure, credit risk and bankruptcy prediction and bond rating. NN were also used in cluster analysis and in its special approach developed by Kohonen (1982). Efficiency of NN can be improved by adopting evolutionary algorithms.

Finally, Bahrammirzaee (2010) W.-Y. Lin et al. (2012) review 130 journals papers from the period of 1995 to 2010 is reviewed and the performance of the most commonly used machine-learning techniques in financial domain is presented. He concludes that notes that although AI tend to outperform traditional techniques, it cannot be considered as absolute.

3 CURRENT STATE – ANALYTICAL METHODS

This chapter is dedicated to analytical methods and relevant topics which are being used in the financial modelling. It should provide theoretical background which is needed to understand all applications in the Main Empirical Results chapter. Final conceptual model (in Figure 4) is included in this chapter to emphasise connections with sub-chapters upon which the model was created.

General strategies and important aspects which need to be considered before the model is created are discussed in chapter 3.1. Chapter 3.2 on probabilistic (statistical) methods presents scope of methods widely used in financial modelling and highlights major advantages, disadvantages and points to often misconceptions. This chapter also introduces Bayesian viewpoint on uncertainty and confronts it with traditional frequentist approach. In chapter 2.1 division of models according their tasks in economic field was made.

We can recognise two types of models:

1. descriptive models
2. predictive models

Even if the type of model is the same (i.e. logistic regression), different aims would require different evaluation criteria and modelling strategies which are used in model-building process. Descriptive models aims to clearly describe underlying pattern which is analysed. Usually inferences about parameters of interests and corresponding uncertainty estimates are made. Input covariates, their interactions and transformations are well-defined by the theory or expert's knowledge. Predictive models are judged by their predictive performance rather than their interpretability. Generally, model building process for both types of models can be described as in Figure 4, but the content of blocks differ.

Predictive models can be divided into two classes - *classification* and *regression* models. Main goal of classification models is to assign instances to qualitative groups. Classification can be either supervised or unsupervised. Supervised classification (SC) starts with instances whose class-labels are known (i.e., company was in a group of healthy companies last year). Final model should classify new instances to these well-defined classes with the highest accuracy. Unsupervised classification (UC) seeks to find common patterns in instances with no knowledge about their membership to any particular class. This approach is helpful in exploratory analysis because it can identify groups of similar instances, companies. Some authors prefer to use term *clustering* for unsupervised and classification for supervised classification. Regression models estimate continuous outcome.

Classification establishes a transformation between feature space and classes (Bandyopadhyay & Saha, 2012). We distinguish between *non-probabilistic* and *probabilistic* classifiers. Non-probabilistic classifier maps instances with covariates from space X onto output space Y. Output space Y contains discrete values, classes. An output y of probabilistic classifier is called *score*. Score can is a value which can be either scaled (and can be interpreted as a probability) or unscaled. Altman's Z-score is an example of unscaled score. Score is then being compared to some threshold value T to classify instance to class. Process of threshold identification is described in chapter 3.1.2.

There are several approaches to classification techniques. All of them try to uncover underlying patterns which have high discriminatory power. Hair (2010) distinguishes between two types of patterns. If instances are similar given their behaviour pattern rather by similar absolute values of measured variables *covariance-based models* are employed. Covariance measure strength and orientation of linear relation among variables. If absolute values have higher importance (e.g., small, medium and large company), *distance-based* methods should be used instead.

Balcaen and Ooghe (2006) divide (FDP) models into six categories. Single variable analysis, risk index models, multivariate discriminant analysis and conditional probability models. Ravi Kumar and Ravi (2007) were concerned by both statistical methods and "intelligent techniques". They divide techniques used in FDP to statistical techniques, neural networks, case-based reasoning, decision trees, operational research, evolutionary algorithms, rough sets techniques and "other" techniques. They also described hybrid methods of various analytical techniques. Following types of analytical methods are generally distinguished in the literature:

1. statistical learning
2. rule-based
3. distance-based
4. neural network models
5. other non-linear models

Statistical learning models use probability theory as the main tool. This class contains regression analysis tools (i.e., logistic regression) and classification techniques based on analysis of differences of probability distributions across classes (i.e., Naïve Bayes). These techniques are described in chapter 3.3. Logic-based methods can be divided into rule-based classifiers and decision trees (chapter 3.4.1). These models have good interpretability and can achieve high predictive accuracy. Several extensions, such as random forests were proposed. Distance based models are based on concept of closeness, which is computed by some distance measure such as Euclidean distance. Neural Networks (chapter 3.4.2) represent family of models which are capable to capture non-liner dependencies. As a typical representative of the last class can be named Support Vector Machine (chapter 3.4.3).

In the following text main focus will be put on

- Statistical Methods – generalised linear models, survival analysis, discriminant analysis
- Machine Learning methods – Support Vector Machine,
- Artificial Intelligence – Neural Networks

Before concrete models will be presented, general modelling strategies will be discussed.

3.1 General Modelling Strategies

This chapter presents general topics which appear throughout the modelling process. First chapter is dedicated to training sample. Importance of sampling techniques is highlighted. In chapter 3.1.2 general scheme used for assessing accuracy for both regression and classification tasks is presented. In chapter 3.1.3 fundamental trade-off between flexibility and simplicity and its consequences on model interpretability and prediction accuracy are discussed. Appealing problem which arises in high dimensional tasks is described as well. Chapter 3.1.4 points to feature (variable) selection and presents methods for relevant variable determination. This section concludes with chapter devoted to ensemble modelling.

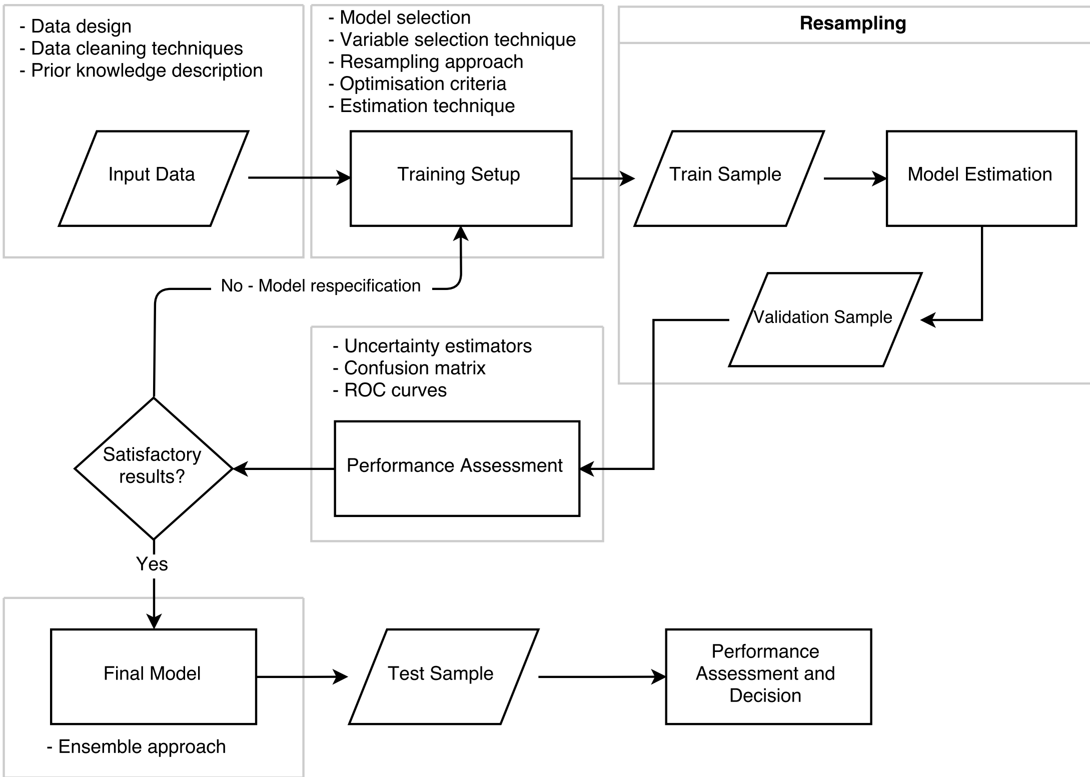

Fig. 4: *Conceptual model points to the most important phases in model building process. Source: Own processing*

3.1.1 Training Sample

Training sample consists of data which is used in model learning phase. Before the sample is created, data design (chapter 5.2) has to be selected. First step of training sample creation involves variable selection. This topic is discussed in more details in chapter 3.1.4. After variables are identified data has to be cleaned. This step involves identification or influential outliers and missing values. Several techniques for outlier detection were designed. Outliers are generally considered as data generated by another mechanism than the rest of the data (Hawkings, 1980). This mechanism can be important to the analysed problem and outliers should not be removed from the sample without closer inspections Aggarwal (2013). Missing values should be treated the same way. In some financial applications models are accustomed to cope with outliers (i.e., robust modelling approach). Outliers represent big risk because they cannot be predicted easily and usually bear high effect, such an unexpected fall in portfolio returns. Problem with missing values can be solved by complete case removal or partial removal. For example, partial removal technique would allow computing correlation matrix with different number of subjects. Subject with missing value in variable x_1 would not be included in pair-correlations $\operatorname{cor}(x_1, X_k)$ but would be in pairs which does not contain x_1. Another option is to use any *imputation* technique. Such techniques usually use nearest-neighbour algorithm or regression model to estimate missing value.

Outliers should be either eliminated (trimming) from the analysis or transformed. If the outliers are suspected to be a result of a noise rather than some systematic cause, Winsorisation is a convenient transformation technique. This transformation replaces outliers with some value which is usually set as a high (or low) quantile of analysed variable. Variable scaling should be considered for two reasons. First lies in the model interpretability (i.e., expected sales if company does not have any assets would equal to intercept value in simple regression model). The second reason is computational efficiency. Some methods, especially those which require Monte Carlo simulations, might fail to find a solution even in simple estimation tasks. Also, scaling improves computational times in many applications.

Another problem emerges in classification tasks with important minor (rare) class. Analytical techniques tend to adjust to the majority of data. This can be prevented by imposing weights (cost) on minor class. As a solution to undersampling Synthetic Minority Over-sampling TEchnique (SMOTE) can be used. SMOTE which is due Chawla, Bowyer, Hall, and Kegelmeyer (2002) artificially increases number of rare events by nearest-neighbour algorithm.

Whole training set is used in exploratory analysis and in some descriptive models. With computational advances resampling techniques become popular and are used routinely in practice nowadays. We can distinguish following resampling schemas:

- Bootstrap
- Sample split

Bootstrap (Efron, 1979) is commonly used in assessment of parameters' estimation accuracy. Core principle lies in random sampling (with repetition) of subjects from the train sample. It's therefore possible that particular subject appears in the training sample multiple times or does not appear at all. Formal model is estimated on all training sub-samples and parameters of corresponding models is further analysed. Distribution of parameters and descriptive statistics (usually quantiles) are usually computed. This approach is presented in Figure 26. Bootstrap allows estimation of standard errors which would be difficult to find by conventional (analytical) way. As an example can serve estimation of confidence interval of Value at Risk computed by Extreme Value Theory (L. Homolka, 2013). Bootstrap does not require any distribution assumption but it allows computing of confidence intervals as in equation 3.2.2.

General idea of split sample techniques lies in splitting training set into two partitions. Model is identified on the first part and on the second its performance is validated. Such model will be tested on unseen data stored in test (holdout) sample. In the *Validation Split* approach sample is divided into two partitions. Usually partition of 80% of data serves as a training set and remaining 20% are used for validation. *Leave-One-Out Cross Validation* is a special case of previous. Validation sample contains only one observation (James, Witten, Hastie, & Tibshirani, 2013) which is estimated by the remaining data. This procedure should repeated as many times as possible (maximum is n times) which might be infeasible for computationally-heavy methods. *K fold cross validation* technique splits training set into k parts (folds). Model is identified on $k-1$ folds and tested on the remaining fold. Potential drawback occurs in classification task with minority class. As the number of folds increases, observations from minor class become rarer in the test fold. Repeating cross validation procedure with only a few folds can fix the problem. Models which are learnt on sub-samples differ because values of splitting key (which subjects to include) are selected by random. It is possible to fit more than one formal model. Usually these models differ in the complexity (number of variables) or in some parameter value (shrinkage). Summarising statistics (Figure 36) can be computed for all models and appropriate model can be selected and applied on test sample.

In many applications instances are categorised into several classes. Some models, such as Naïve Bayes or J.48 are eligible for this task directly. However, not all models are able to analyse ordinal dependent variable (such as rating classes). Ordered logistic model can incorporate such information after some adjustments. Many models (e.g., J.48) treat classes only as nominal values. If selected method is a binary classifier (i.e., can differentiate only between two classes), different learning strategy has to be used. These strategies directly affect training set. First strategy is called *One-vs-All*. If K denotes number of classes and $1 \leq k \leq K$ than *k-th* classifier labels class k as 0 and all other classes as 1 and standard binary classification can be performed. After all K different binary classifiers are identified, class with highest probability of membership (or similar metric) is assigned to the subject. Second approach is *One-vs-One* which requires computing $\binom{K}{2}$ classifiers. This approach directly compares one group to another. Less frequent approach Error-Correcting Output Codes which was created by Dietterich and Bakiri (1995) transforms predicted class label into binary string. For each position in the string classifier is learnt. Classification is made by computing closeness of predicted string. This can be done by Hamming or any other convenient distance methods.

3.1.2 Performance Evaluation Tools

Assessing accuracy of classification or regression model is step which is present throughout the process of the model identification. Best model is usually built in iterative process which involves optimisation of parameters and complexity (chapter 3.1.3). Also, performance measure should be used in in decision-making step as it contains information about uncertainty and possibilities of the model. Arguably most common indicator of accuracy in regression modelling is *mean squared error*:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3.1.1}$$

where y_i is observed value and $\hat{y}_i = \hat{f}(x)$ is the estimate. MSE also serves as an optimisation criteria for identifying parameter of function $\hat{f}(x)$ which minimise the sum of squares. MSE is a convenient indicator for models with numerical outcomes, but is not suited for discrete-classification problems.

Classification performance evaluation received great attention especially as the field of machine-learning emerged. Hernández-Orallo, Flach, Ferri, and Elkan (2012) track the development and point to seminal works. They present an unified view on performance metrics for predictive models. Classification models can be divided into *non-probabilistic* and *probabilistic*. Results obtained from non-probabilistic classifiers are directly presented in the contingency table. Class label with no additional information indicating strength of membership is an outcome. Probabilistic models results in numerical variable which can be transformed to $(0,1)$ interval to have clearer interpretation (hence the name probabilistic). After the model created classification results are summarised in *confusion matrix*. This matrix contains all possible combination of predicted and observed classes. Values on the diagonal represent correctly classified observations. James et al. (2013) suggest as a simplest classification indicator *error rate* (ER), sometimes called hit rate:

$$\mathrm{ER} = \frac{1}{n}\sum_{i=1}^{n} I(y_i \neq \hat{y}_i) \tag{3.1.2}$$

Under settings with equal misclassification costs sum of off-diagonal subjects is minimised implicitly. In situations with different misclassification costs error rate is not suitable indicator of performance (Provost & Fawcett, 2001). Also, if proportions of classes are imbalanced and error rate is used as an optimisation criteria, selected model will achieve high hit rate because model will assign subjects to majority class. It is therefore possible to achieve high

Tab. 4: *Classification table for binary classifier and generally accepted terminology. Predicted value by classifier is denoted as $\hat{y}$. True class is y. Value 0 means negative, 1 positive.*
Source: Own processing

	reality $y=0$	$y=1$	
$\hat{y}=0$	True Negative	False Negative	Negative classifications
$\hat{y}=1$	False Positive	True Positive	Positive classifications
	Real negative cases	Real positive cases	Total sample size

classification accuracy (correctly classified to all companies) with model which does not have any discriminatory power. Accuracy of 90% can be achieved if minority class presents only in 10% of all companies.

In many real world applications misclassification costs differ. In such situations, it is possible to assign costs and gains (profit) for all combinations which appear in confusion table. Table which summarises costs is called *cost matrix* although it can contain benefits from right classification. In the optimisation step confusion matrix is multiplied by cost matrix and the total sum (total profit) serves as an optimisation criteria. If different costs are used, final confusion matrix might get significantly worse in terms of correct hits. This is magnified if minority class has the highest misclassification cost. Cost values should be carefully estimated by both credit specialist and analysts. An example of cost-matrix is at Table 5. Cost matrix can be designed from several perspectives:

- economic
- managerial
- computational

From the economic perspective costs and gains are the mean values of losses suffered and profit gained from an average contract. This involves only values which appear in accounting statements. In banking industry costs are calculated as credit exposure times the loss given default (LGD) value. LGD reflects value of collateral and residual value for which defaulted security can be sold. Additional administrative and legal costs might be considered as well. Managerial perspective reflects expectations, true and opportunistic costs. Expectation represents total profit which would is set in the beginning of the period. Defaults of particular subjects bear cost of unfulfilled expectations plus direct costs. Managers value missed opportunity as well (Nayak & Turvey, 1997). Cost matrix therefore contains virtual values which does not meet values in the accounting. Last approach is used for model tuning purposes. Here, cost matrix is set to force learning algorithm not to commit predefined mistakes by imposing extremely high costs.

Economic or managerial cost matrix is not used often in empirical papers. This support Lessmann, Baesens, Seow, and Thomas (2015) who point to the lack of available information in the literature. Authors usually present classifier's sensitivity to other parameters and class imbalance effect.

It is possible to learn model from computational perspective and transform results to have economic interpretation. After the model(s) is/are identified, proportions in training confusion table(s) are computed. Expected profit can be computed by multiplying number of predicted good companies by expect proportions of incorrectly classified defaulted companies (α error rate). This is demonstrated in chapter 6.8.

Model is identified by minimising costs or errors. In case of probabilistic models decision rule which labels instances according to numerical values has to be selected. This rule consists of set of $k-1$ threshold values (T). Number of classes is denoted as k. For example, Altman's model sets threshold 1.81 to separate "distressed zone" from "grey zone" and further threshold 2.99 to specify "safe" zone. Process of identification optimal threshold value is demonstrated at Figure 33.

Terminology which is used to describe classification outcomes in the literature is showed in Table 4. Performance indicators which are derived from the table are listed below:

- Accuracy = (TN + TP)/all
Proportion of correctly classified cases.

- Sensitivity, Recall, Power = TP/(TP + FN)
Proportion of correctly classified positives from all positive cases.

- Specificity = TN/(TN + FP)
Proportion of correctly classified negatives from all negative cases.

- Precision = TP / (TP + FP)
Proportion of true positive cases from predicted positives.

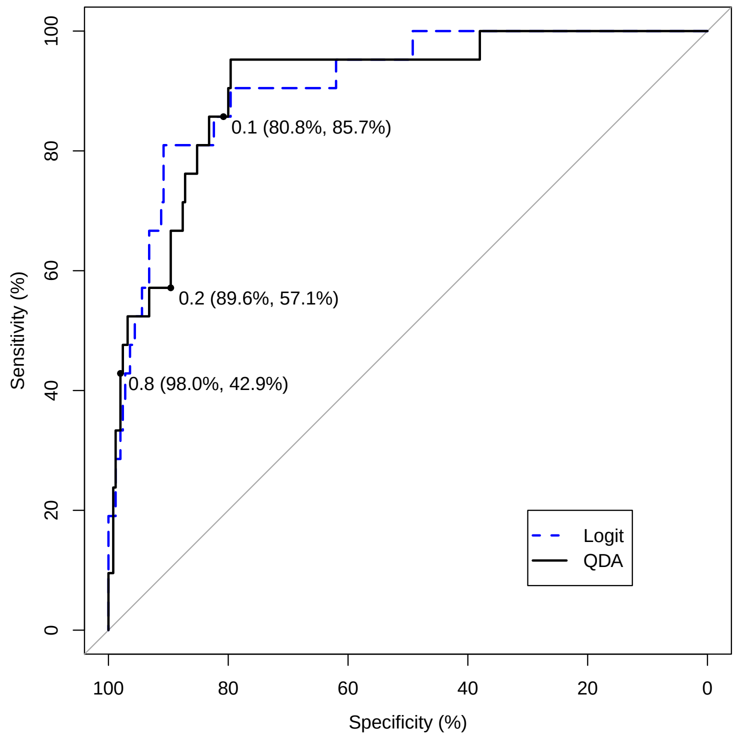

Fig. 5: *Performance of two classifiers (logistic regression and quadratic discriminant analysis) was inspected by comparing two ROC curves. After logit model is learnt, threshold value has to be appropriately selected. If this threshold would be set on 0.1, then all companies with estimated probability of being a member of class 1 (defaulted) higher than 0.1 would be considered as defaulted. It means, that 80.8% healthy companies would be labelled correctly and 85.7 % of defaulted companies would be recognised as defaulted by model. Source: Own processing*

Arguably the most used combination of classification measures is False Positive rate (1-Specificity) and True Positive rate (Sensitivity). This relation is graphically in Receiver Operational Characteristic curve plot (ROC). Diagonal line in Figure 5 indicates random guess model. Further away from this line ROC curve is (to the top-left corner) the better classification performance is achieved. ROC plot depicts all possible classification combinations with respect to some threshold value. ROC is a convenient tool for comparison of several classifiers (Figure 32 compares performances of four classification algorithms).

Performance of a single classifier can be done by inspection of ROC plot or by computing Area Under Curve (AUC) statistics. AUC is computed as an integral of the ROC curve over all threshold values. Classifier which achieves highest value of AUC is considered as the best performing. But, this indicator is flawed in the way that it computes performance on thresholds which are of a poor use in practise. Corner and near-to-corner thresholds are taken into computations as well. Despite of this drawback AUC is used in majority of classification oriented papers in finance. For probabilistic models several metrics were developed. Mean Squared Error (MSE), often referred as to Brier score is the most frequently used.

On imbalanced data-sets κ (kappa) statistics is a convenient indicator which allows projection of expected accuracy (computed from confusion matrix in the same way as expected counts in contingency table) into the result. κ statistics is computed as:

$$\kappa = \frac{\text{observed} - \text{expected accuracy}}{1 - \text{expected accuracy}} \tag{3.1.3}$$

Negative κ value indicates that classifier performed worse than random classifier. This value can occur even if achieved accuracy is very high. Values higher than 0.5 indicate well performing classifier.

Quality of classifiers with probability score can be assessed by some metrics developed in information theory. The well-known log-loss for binary classification acknowledges whether object was classified correctly given strong (i.e., $P(y = 1|x) = 0.8$) or weak evidence.

$$\text{log-loss} = -\frac{1}{n}\sum_{i=1}^{n} y_i \log p_i + (1 - y_i)\log(1 - p_i) \tag{3.1.4}$$

Probabilistic outcomes can be validated visually by comparing empirical distributions of probabilities of default. If the problem is perfectly separable and the model can identify such decision rule all distressed companies would have higher estimated probability of being in class of distressed.

In the end of the analysis predictive scores should be verified for classification consistency. Figure 6 shows that defaulted companies are present across all values of estimated probability of default. Distributions would not overlap if the model would perfectly separate both classes.

3.1.3 Model Complexity

In the last chapter resampling methods were presented. Typically several models with different complexity are created on the training set. Higher complexity of estimated model $\hat{f}(x)$ can be achieved by adding higher polynomials or interactions between variables. Models with higher complexity tend to outperform simpler models on the training set, because they can adjust better to the data. Performance on test sample usually turns into bad results as the model is too tied to test data. This is so-called bias-variance problem.

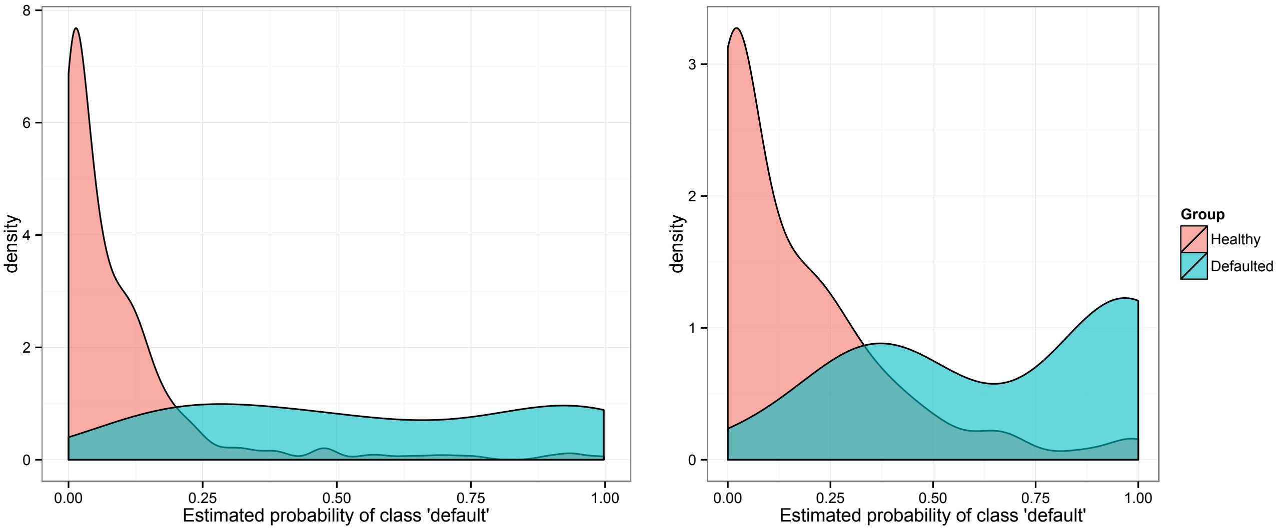

Fig. 6: *In each plot estimated densities are printed. Results achieved on both training (left panel) and test set show that logit model assigns lower probability values to healthy companies than to defaulted. Note that the likelihood scale is not the same in both panel. This can be interpreted as increase in uncertainty as the distributions become flatter. Source: Own processing*

Mean Square Error is an indicator of model's quality. MSE can be decomposed into three components, *bias*, *variance* and variance of irreducible error ε. Models with high flexibility (e.g., model with high degree polynomials) differ largely from sample to sample when they are re-estimated. This is caused by their heigh ability to adjust. Estimated model's parameters are unstable and can result in misleading predictions if the pattern identified on test data differ even slightly from pattern in training set. This problem is called *variance* as there is a big variability in estimated models (their coefficients). *Overfitting* is a situation when the model's complexity is unnecessary high. *Bias* problem means that model is not able to capture complex relation because of undue simplicity. Ultimate goal is to identify ideal model complexity which minimises test MSE, as shown at right panel of Figure 7. This complexity can be found on re-sampled models. In the Figure 36 total cost and accuracy was used instead of MSE to identify optimal number of hidden nodes in Neural Network model.

Curse of dimensionality is another important problem related to the complexity of the model. Increasing complexity complicates the whole analysis from the computational as well as from the methodological perspective. As the number of parameters (or nodes in decision trees) becomes larger, model's complexity increases non-linearly. Additional parameter will require higher number of observations to be estimated correctly. In high-dimensional problems traditional measures of distance are loosing their intuitive meaning and, sometimes, relevance. Nearest-neighbour classification can serve as an example (it is easier to find two similar companies based on only two variables rather than on three variables). Also, chance of observing spurious correlations among variables increases. This represents problem mainly for approaches which try to identify covariance based patterns.

3.1.4 Variable Selection

Variables should be selected to cover all important dimensions of studied problem. For exploratory, descriptive and models created to test the theory, variables should be selected with prospect that corresponding coefficients will have clear interpretation. In this case variables

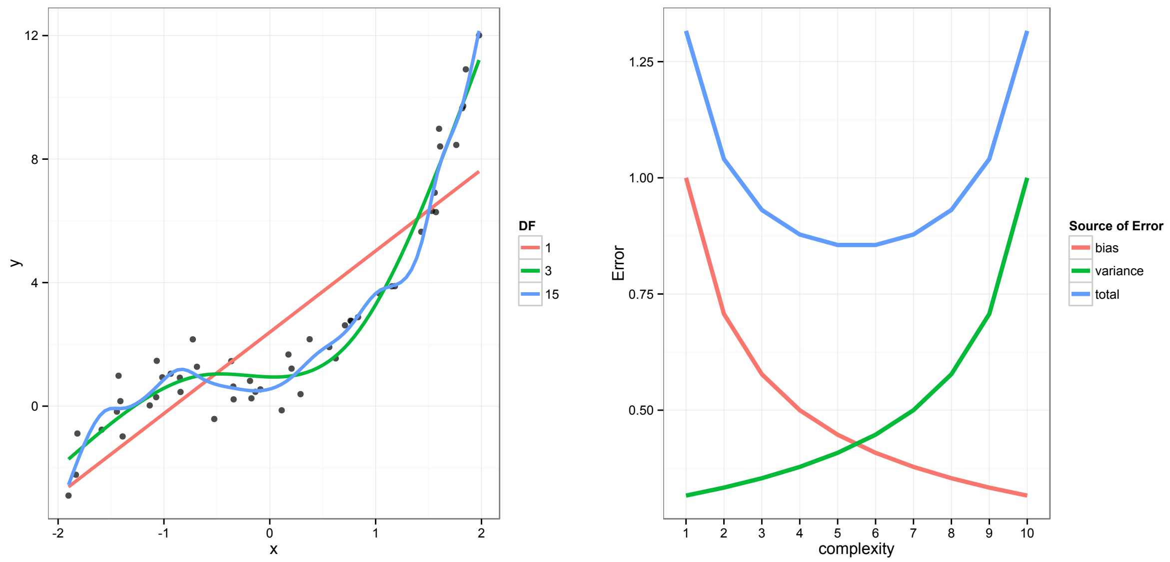

Fig. 7: *Three spline models are fitted in the left panel. These fits differ in complexity which is expressed in terms of degrees of freedom (DF). Simple spline (red) is not flexible and fails to recognise non-linear relation between variables. Highly flexible (blue) model adjusts overly to the data upon which is created. Green spline seems to be ideal model. Right panel shows general scheme of total error (test MSE) decomposition. Problem of variance increases as complexity increases (overfitting). Simple models (low complexity) cannot capture problem properly. This error comes from the bias component. In this example complexity of ideal model is around 5. Source: Own processing*

should reflect domain expert's knowledge. If the prediction of some quantity (probability of default) or behaviour (transition matrices between rating grades or shared risk in portfolio which varies over time) is the model's purpose, variable selection is more difficult. Let's consider rating assessment – this task requires analysis of several independent variables with shared properties (return on assets indirectly refer to structure of assets). Appropriate variables complex models are usually selected by algorithmic approaches.

Although advances in computer science allow processing of highly dimensional data, it is advisable to only work with a subset of all potentially useful variables. Variable (feature) selection is a family of procedures designed for finding optimal set of variables with respect to some objective (or objective functions). Guyon and Elisseeff (2003) divides objectives into three groups:

1. importance and variable understanding
2. predictive performance improvement
3. learning method effectiveness (computational speed, costs).

If an inference or description is the main aim of the analysis *interpretability* of parameters is crucial. In this case researchers usually try to identify small subset of variables to avoid problems of over-fitting or multicollinearity. In predictive modelling multicollinearity does not represent as important problem as users are not concerned with standard errors (which are inflated) to the same extent as analysts in the previous case. However, according to Kuhn and Johnson (2013) multicollinearity can results in unstable models and consequently in deteriorating predictive abilities. Problem of instability of coefficients is demonstrated in Figure 25. Predictive performance is influenced by its complexity and corresponding bias-variance trade-off discussed in chapter 3.1. Variable selection can reduce the variance component. Lastly, computational efficiency is amongst the biggest challenges in the era of Big Data. Despite advances of parallel, batch or distributed computing, dimensionality reduction is still an important task.

The result of feature selection is a set of *relevant* variables. Kohavi and John (1997) defines relevance in classification setting but the idea can be extended. Variable is relevant if knowing its value will positively result in some prediction indicator or it will improve understanding of behaviour of analysed variable.

High dimensional problems can suffer from several problems. Analyst face *identification* problem when there is insufficient number of observation compared to number of parameters to be estimated. Finally, elimination of *irrelevant* variables can overcome problem with collinearity and also with curse-of-dimensionality problem.

Variable selection using computational intelligence can be divided into two branches - filters and wrappers (Angelis, Felici, & Mancinelli, 2006). Filters are methods of identifying appropriate feature with no reference to selected (analytical) technique. Firstly, evaluation function and stopping rule is set. In the next step search algorithm is selected. After the rule is achieved selected variables are utilised in some analytical technique (i.e., logistic regression). Angelis et al. (2006) present and discuss main approaches (methods based on: consistency, information theory, correlation and combinatorics) for setting evaluation functions in more detail. Although filter approach can be successfully applied on large datasets (it is not as computation heavy as the wrapper approach) it does not usually over-perform wrapper methods. Wrapper method needs search algorithm and analytical technique to work simultaneously. It is an iterative process which can be time consuming on large datasets. Main wrapper types are decision trees based (with popular algorithms ID3, C4.5 and ID4.5), context based and methods developed for Bayesian classifiers.

Angelis et al. (2006) note that association and classification rules are the most important features of data-mining tools. They consider association rules as a set of rules that allows identification of factors' combinations (values in feature space) which are very likely to occur simultaneously (usually measured in terms of probability).

In many situations feature space is large because some data is cheap to acquire. It is therefore important to identify those variables which contain information for particular task. Angelis et al. (2006) and Engelbrecht (2007) urge system developers to use feature selection in data-mining application as it brings lower computational efforts, better model performance, clearer rules (i.e., shorter classification trees), cost-cutting related to data acquisition and maintenance. W.-Y. Lin et al. (2012) provided literature review on feature selection with connection to classifier performance. They found strong evidence that "selection of features to reduce dimensionality promotes the performance of single classifier beyond what is possible without considering feature selection". These findings are consistent with Ravi Kumar and Ravi (2007).

If the task is to select variables with highest exploratory power by some automatised procedure, three different strategies are usually used:

- dimensionality reduction – Principal Component Analysis is a standard tool which identifies *components* which describe largest proportion of total variance. Components are connected to variables via component loadings. Dimensional reduction is showed in Figure 8.

- step analysis – techniques of forward and backward selection. In forward selection after variables is added some measure of goodness-of-fit is computed (AIC). In backward selection starting point is full model. Variables are being removed as the model still present good results.

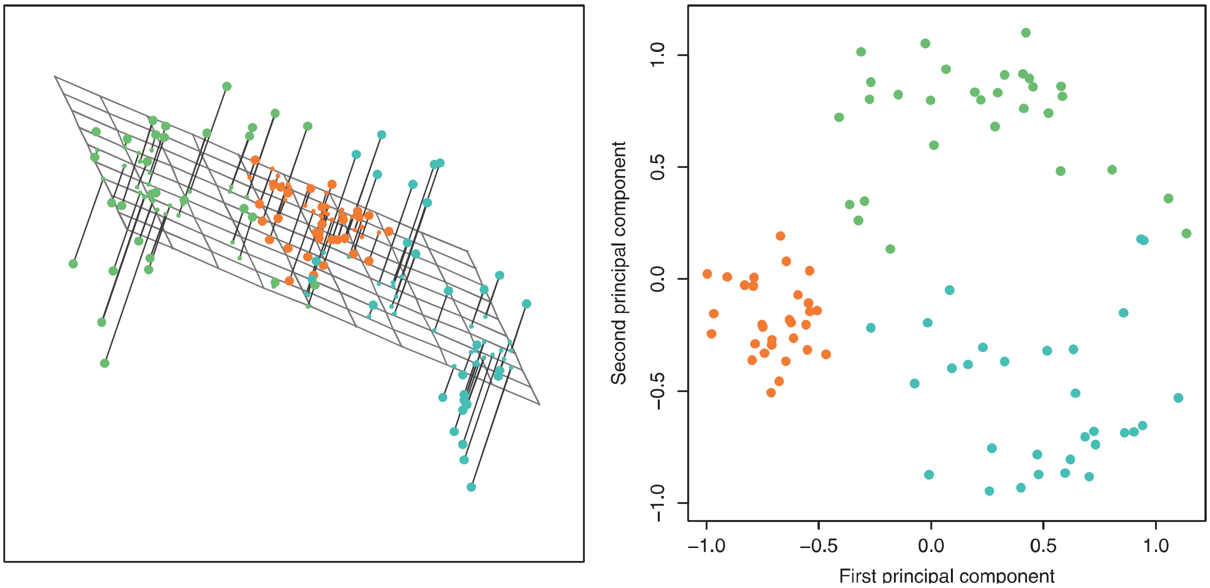

Fig. 8: *In the left panel initial 3 dimensional problem is plotted. After PCA was performed problem was reduced to 2 dimensions. Some information is lost during the process. This information is expressed in terms of variance. Source: James et al. (2013, p. 380)*

- shrinkage methods - regularization parameter minimizes values of less important variables.

Two well known shrinkage methods were developed in early 90s. The first is called *ridge* regression. Linear regression models find its coefficients by minimising some loss function. Usually, this cost function is quadratic loss function (RSS):

$$\mathrm{RSS} = \sum_{i=1}^{n} \left(y_i - b_0 - \sum_{j=1}^{p} b_j x_{ij} \right)^2 \tag{3.1.5}$$

where n is number of observations and p is number of covariates. Ridge regression adds to the loss function *shrinkage penalty*:

$$RSS + \lambda \sum_{j=1}^{p} b_j^2 \tag{3.1.6}$$

second method is called *lasso*. It minimizes loss defined as:

$$RSS + \lambda \sum_{j=1}^{p} |b_j| \tag{3.1.7}$$

problem of setting λ values is usually solved by estimating function on cross-validation samples and computing MSE value.

3.1.5 Combination of Results

This chapter provides overview of methods which allows merging of several models into final one. *Bagging* and *boosting* belong to the group of homogeneous ensemble techniques because they improve performance of a single classifier by resampling.

According to W.-Y. Lin et al. (2012) general idea of combining classifiers can be expressed as:

$$P(y|x) = \sum_{k=1}^{K} w_k P(y|x,k) \tag{3.1.8}$$

where K is the total number of classifiers, $P(y|x,k)$ is the probability or likelihood of subject belonging to particular class (y) given covariates x in re-sampled classifier k.

The idea of *bagging* (bootstrap aggregation) is to average predictions obtained from bootstrapped samples. After B bootstrap samples is generated B predictions are computed and stored in vector $\hat{f}^b$. If the outcome of the model is non-probabilistic, vector $\hat{f}^b$ contains only nominal values. Rule of *majority voting* is used. Individual is assigned to class which has received most votes – the most frequent category in vector $\hat{f}^b$. If the classifier provides probabilistic outcome, mean value of the vector of predictions and selected threshold would decide about the class label. According to James et al. (2013) classification residuals obtained from *out-of-bag* sample (data not used in model building) can be considered as a valid estimate of test error for bagged model. Formally, probabilistic bagging is written in equation (3.1.9).

$$\hat{f}_{\text{bag}}(x) = \frac{1}{B} \sum_{b=1}^{B} \hat{f}^b(x) \tag{3.1.9}$$

Another technique is *boosting*. The idea behind boosting is to combine results from several weak classifiers to create one with better predictive or descriptive performance. Hastie, Tibshirani, and Friedman (2009) define weak classifier as a classifier which performance is slightly better than random guess. Under this approach the same model (such as LDA) is estimated repeatedly. Starting parameters and weights are modified according their relevance. This method is usually described as a close relative to bagging but, as Hastie et al. (2009) state, boosting is fundamentally different. In bagging, bootstrap samples are drawn independently. In boosting, the information (which helps in tuning function in the following steps) is created sequentially. Boosting does not involve bootstrap techniques. Boosting algorithm is described in Hastie et al. (2009) details, here only the boosted model is presented:

$$\hat{f}_{\text{boost}}(x) = \frac{1}{B} \sum_{b=1}^{B} \lambda \hat{f}^b(x) \tag{3.1.10}$$

parameter λ is a shrinkage parameter used for adjusting bias/variance problem. Hastie considers *AdaBoost.M1* algorithm as the most popular boosting algorithm.

As Lessmann et al. (2015) note, this type of ensemble modelling received more attention than heterogeneous techniques. Prominent place in heterogeneous ensemble modelling belongs to *stacking* procedure. This approach allows combination of diverse models into one output. Several classifiers (so called 0 level models) are created in the first phase. High predictive accuracy on unseen data can be achieved either by selecting best performing models or by selecting diverse models. It is expected that collection of diverse model would create more robust system which does not have as high predictive accuracy under normal conditions. After the approach is selected, level 1 model is used to determine weights for outcomes of level 0 models. This is usually done by decision tree model or any other simple model. This approach is demonstrated in chapter 6.12.

In many real-world applications more than one model (or approach) is used to find a solution. From the temporal perspective (when we combine outcomes) three combination methods can be distinguished.

- cascade combination – only outcomes from the first step are used in the second. (First step cleans data and second performs classification.)

- serial combination – several independent method are employed in the first step. In the following steps results from the models and initial data are used. (Cluster analysis is used to classify subjects. In the second step some inferential analysis is made). Example of serial analysis can be found in chapter 6.5. Probabilities of default for all companies were obtained from GLMM model in the first step. These values as well as other variables were used in the survival model.

- hybrid combination – this is a special case of combinations which occurs when analysis is performed iteratively. (Neural network is identified by back-propagation algorithm).

3.2 Probabilistic Approaches

Statistics is a methodological approach which uses probability as a main tool for understanding uncertainty. From the two centuries long discourse has emerged two approaches to statistics – standard (frequentist) and Bayesian. Both approaches can perform the same kind of analysis. However, interpretation of results is different because of the underlying definitions.

3.2.1 Frequentist Statistical Analysis

As a standard approach to probability is considered the frequentist approach. Probability P is viewed as a frequency (proportion) of positive results to all observations $n \rightarrow \inf$. Under this approach researcher assesses uncertainty of phenomena based on sampling distribution. If the distribution is obtained by random sampling and is large enough, it is possible to generalise findings about quantity of interest. This can be mean value of random variable μ. Main aim of an inference is to estimate population characteristics (μ) based on sample statistics ($\bar{x}$). This process is called *inference*.

Standard model assumes that population characteristics θ is a fixed quantity. Observed data Y_i from the sampling distribution $i = 1, \ldots, N$ are generated from some distribution D with parameter θ. This model can be written as

$$Y(\theta) \sim \mathrm{D}(\theta) \tag{3.2.1}$$

to highlight that data is a function of parameter θ. Let θ be an object of interest (such as a mean value of ROA in rating class AAA). Variable D denotes data which was collected to describe studied phenomena (sample of n observation). From the observed data statistics t is computed (continuing in the ROA mean value example, t is a sample mean value $\bar{x}$). Now, researcher has two options - to continue sampling from other sources (and satisfy assumption of independence of samples) and compute enough sample statistics t to make direct inference on θ; or to assume, that the sampling distributions of t converges to some underlying distribution (such as a normal distribution) and estimate the interval from this distribution. If the second approach is selected, than the interval of θ (for mean value) will be:

$$P\left(t - z_{1-\frac{\alpha}{2}} \frac{s_t^2}{\sqrt{n}} \leq \theta \leq t + z_{1-\frac{\alpha}{2}} \frac{s_t^2}{\sqrt{n}}\right) = 1 - \alpha \tag{3.2.2}$$

Interval computed from equation 3.2.2 is called *confidence interval* (CI) and describes how many times $(1-\alpha)$ the true value θ will appear in the confidence interval computed on independent samples. This is confusing interpretation, because it does not say anything about the "most probable" estimate of θ - all numbers contained in interval are considered as the best estimates of θ.

Although standard techniques are considered as objective in that sense that they do not utilise any prior information and rely only on the data in hand, Robert (2007) disagrees. He points to different optimisation criteria for identifying parameter's value (such as criteria in variable selection process) or in other loss functions.

3.2.2 Bayesian Analysis

Bayesian analysis is a competing approach to standard frequentist analysis. Contrary to standard approach, parameters are considered as random variables. In standard model (3.2.1) data was generated by some parameter, in Bayesian approach observed data is considered as conditioned on the parameter θ:

$$Y|\theta \sim \mathrm{D}(\theta) \tag{3.2.3}$$

Random variable θ is distributed according to some probability distribution which has to be set before data is observed. This allows analysts to include domain expert's knowledge or results from previous research (in terms of *prior distribution*) to the analysis. Although many methodologists recognise this as an advantage, for many this is only way how to "scientifically bend" analysis towards expected results.

After data x is collected sample distribution $p(x|\theta)$ can be estimated. Parameter θ is the parameter of underlying distribution according which the data is generated. In the next step prior distribution $p(\theta)$ of the parameter θ is specified. These two distributions constitute complete Bayesian model. From these two distribution *joint distribution* can be computed as:

$$p(\theta,x) = p(x|\theta)\,p(\theta) \tag{3.2.4}$$

which describes relationship between both distributions on the whole common domain. The cornerstone of Bayesian analysis is the *posterior* distribution of parameter θ (in continuous case)

$$\begin{aligned} p(\theta|x) &= \frac{p(\theta,x)}{\int p(\theta,x)\,\mathrm{d}\theta} \\ &= \frac{p(x|\theta)\,p(\theta)}{\int p(x|\theta)\,p(\theta)\,\mathrm{d}\theta} \end{aligned} \tag{3.2.5}$$

Finally, *posterior predictive* distribution of new values $\tilde{x}$ is computed as:

$$p(\tilde{x}|x) = \int p(\tilde{x}|\theta)\,p(\theta|x)\,\mathrm{d}\theta \tag{3.2.6}$$

Predictive distribution is therefore mean of $\tilde{x}$ values weighted by the likelihood of posterior distribution and is derived from Bayes rule. After the posterior distribution $p(\theta|x)$ is found, value of θ which maximizing this distribution is considered as s solution (also known as the *most probable a posteriori* - MAP solution).

From the equation 3.2.5 impact of prior knowledge expressed in as distribution $p(\theta)$ on the outcome (posterior distribution) can be addressed. In case of absence of prior information, it is possible to use non-informative prior distribution. Despite the name, such distribution contains small amount of information which is reflected in posterior distribution. Other type of prior distribution used mostly for computational purposes are weakly informative priors. These are used to regularize posterior distribution.

Bayesian analogy to confidence interval is credible interval (CrI) or credible set. A credible set is defined as a subset C of the parameter space θ:

$$1-\alpha \leq P(C|x) = \int_C p(\theta|x)\,\mathrm{d}\theta \tag{3.2.7}$$

for continuous variable and for discrete:

$$1-\alpha \leq P(C|x) = \sum_C p(\theta|x)$$

Bayesian interval makes direct statement about the parameter sub-space and computed on some probability level, there is no reference to replications. Let's consider both Confidence interval (frequentist) and Bayesian CrI for mean parameter μ at the same α. Recall that in classic approach μ is a fixed quantity, in Bayesian approach is unknown random variable. Interpretation of CI is as follows: if we were able to replicate the research under the same conditions on the very large number of data-sets, about $1-\alpha\%$ of them would contain the true parameter value μ. CrI would be interpreted as: the probability of θ being in interval C is at least $1-\alpha$ given observed data y.

As Carlin and Louis (2000) show, sometimes it's impossible to set C exactly to equal $(1-\alpha)$. The next step in estimating C under Bayesian framework is *highest posterior density region* (HPD). This HPD is defined as:

$$C = \{\theta \in \Theta : p(\theta|x) \geq k(\alpha)\} \tag{3.2.8}$$

where $k(\alpha)$ is the largest constant which satisfies $P(C|x) \geq 1-\alpha$. The difference between HPD and CI is presented in Figure 9.

Important topic is hypothesis testing. This procedure is used for falsifying null hypothesis by setting contrasting alternative hypothesis. Both Neyman-Pearson testing approach and Bayesian testing are methods for evaluating the weight of evidence for the hypotheses. Bayesian analysis is a method for determining weights of evidence (Christensen, Johnson, Branscum, & Hanson, 2011). Standard approach evaluates how probable observed evidence is in light of alternative hypothesis. Hypothesis testing is often reduced to presenting *p-value* and its comparison to selected confidence level.

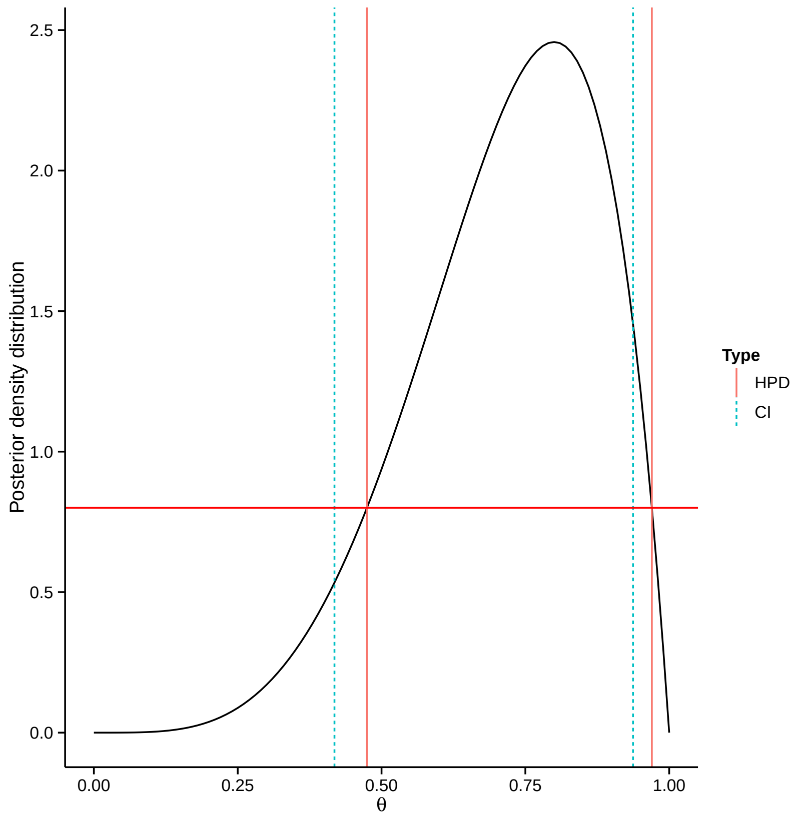

Fig. 9: *Highest posterior density region is shown in solid red line. Notice the likelihood of estimated quantities of interest (θ values) have the same likelihood (on vertical axis). This is not true for confidence intervals which are drawn in dotted lines. Both intervals share the same α value but differ in location. Source: Own processing*

Incorporation of prior knowledge is considered as the most distinctive feature between both approaches, although other exist. For example Bayesian approach does not refer to the *mean representative*. Frequentist confidence intervals have counterpart in credible regions although interpretation differs. There is also a distinction in answer being asked. Bayesian approach is able to answer $P(\theta|\text{Data})$ whereas standard approach reports $P(\text{Data}|\theta)$ in hypothesis testing process.

3.2.3 Prior Information Identification

As Robert (2007) comments, the most critical and criticized part of Bayesian modelling is the selection of proper prior distribution. This difficulty arises not only from the mathematical notation, which might be difficult for managers to use, but because priors directly affect the outcomes (credible intervals, parameter estimates, bayes factors, ...). Inappropriate selected prior distribution can also make computation difficult or even infeasible. Prior distribution should arise from historical analysis of studied phenomena. We therefore have to assume that *apriori* information (recall problem related to Knight's approach to uncertainty in 5.1) exist. If this is not a case, we have to use either *subjective* estimate (expert's opinion) or use methods of *empirical bayes*. Empirical Bayes treats the uncertainty related to prior distribution in pragmatic way. This approach relies "on the conjugate prior modelling, where the hyperparameters are estimated from the observations." (Robert, 2007, p. 459) Because there are two of three components (likelihood and prior distribution) directly read from the data, results cannot be longer considered as a

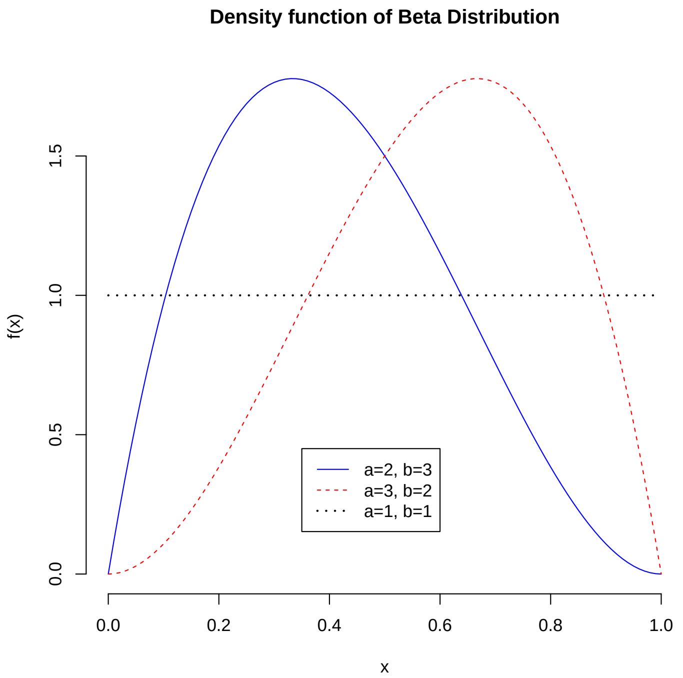

Fig. 10: *Prior distributions are used to capture expert's knowledge. In case of the solid line we expect lower chance that company fails. Even if we do not have any prior information Bayesian framework can be still used. Prior distribution is then set as "non-informative" what is showed in dotted line. Source: Own*

Bayesian. However, this approach allows better prior estimation and can be useful for operational decisions.

Critics of selecting priors can be traced back to the times of Laplace (Robert, 2007). Bayes's response was justifiable as he referred to physical models and therefore there were solid theoretical background available. Using Bayesian framework will always carry this stigma. *There is no universally agreed (right) way how to describe prior distribution.* Jeffreys (1946) proposed *uninformative* priors as a cure in the the situations when no prior information is available. Robert (2007) sorts prior distribution identification strategies into:

1. subjective determination (computation involves Monte Carlo sampling)
2. conjugate priors - prior distribution is the same family as the posterior (solution can be analytical derived),

The simplest way of assessing prior distribution is to estimate region of possible values of interest and corresponding probabilities. In chapter 6.1 prior distribution which reflects previous results is demonstrated. As Carlin and Louis (2000, p. 37) write "prior elicitation issues tend to be application- and elicitee-specific, meaning that general purpose algorithms are typically unavailable.".

This chapter concludes with brief introduction of *Fuzzy* logic (FL) and rough set theory. Although this approach is not probabilistic, it plays an important role in description of vague information. Although concept of variables with ambiguous meaning was discussed before seminal paper of Zadeh (1965), Zadeh's work presents first systematic approach. As Dubois and Prade (1998, p. 3) write, "Fuzzy sets are acknowledged as a major tool in information engineering

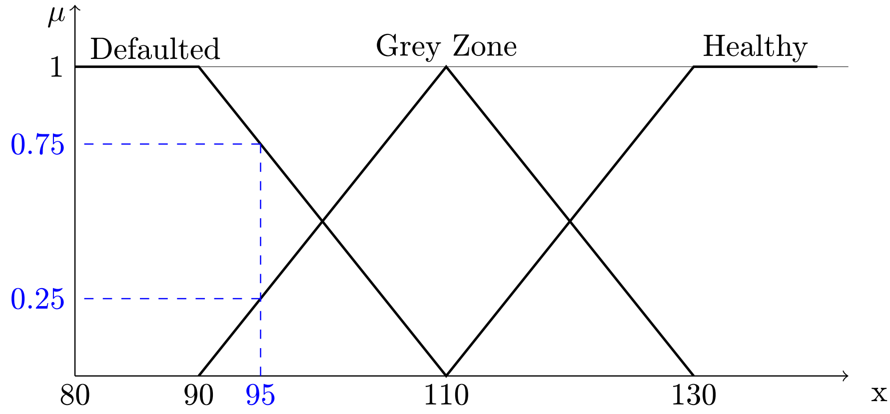

Fig. 11: *Triangular membership functions compute degree of membership for three fuzzy sets. If client's credit score assessed by bank achieve only* $x = 95$ *points, company is almost "defaulted", but not for 100%. Source: Own processing*

for the purpose of bridging the gap between human-originated formalized knowledge, and numerical data." Fuzzy theory reacts to the standard boolean logic of two states. Let's consider test which determines whether company is defaulted or not. Standard logic allows only two statements: company is or is not defaulted. FL allows company to be partially in both states simultaneously. Company can be "nearly" defaulted. Initially, set of states has to be defined. Let A be a set named "Healthy company". Degree of *membership* to this set is computed through membership function $\mu_A(x)$, where x is observed covariate. Several membership functions were proposed in the literature. Situation of nearly defaulted company describes Figure 11.

Reasoning in figure 11 can be reverted. If manager thinks that company does not have good prospects but there is still a chance for improvement, than expected results of analytical test should be around some value $x = 95$. This process can validate agreement between analytical model and domain expert knowledge.

Fuzzy logic provides framework for developing IF THEN rules. Process of creating rule based system consists of three steps. During the first step (*fuzzification*) membership functions assign membership values to observed quantities (x). Seconds step is *inferencing*. This step involves identification of the logical rule which leads to outcome fuzzy set. Last step involves *defuzzification* which computes crisp value with desired economic meaning (can be expressed in natural units). Simple fuzzy rule framework was extended to capture real-world conditions. Fuzzy rules have practical implementation in fuzzy controllers. Amongst the most popular belong Mandami and Takagi-Sugeno controllers which utilise table of all feasible combinations of input and output sets.

Close discipline to Fuzzy sets is theory of Rough sets. Both sets can capture vague information, but in rough sets set lower boundary represent values with full certainty whereas upper boundary represent full uncertainty. (Engelbrecht, 2007)

3.3 Statistical Models

This chapter presents the most important methods which are common in the field of study. Models presented in this chapter are defined from the standard perspective, although some references to Bayesian analysis occur. This chapter presents both cross-sectional (static) and dynamic models. General overview on regression analysis with emphasis on logistic model is provided.

3.3.1 Naïve Bayes Classifier

Naïve Bayes (NB) classifier is a statistical method developed around 1950. This method works with densities of independent variables which are assumed to be different for different classes (otherwise the variable does not have discriminatory power). This method requires proper identification of distribution likelihood $f(X)$. Assuming p covariates are used, joint likelihood for class k is obtained as:

$$f_k(X) = \prod_{j=1}^{p} f(x_{kj}) \tag{3.3.1}$$

After joint likelihood for particular subject $f_k(X_i)$ is computed across all classes, highest achieved value determines class.

3.3.2 Discriminant Analysis

Application of discriminant analysis was thoroughly discussed in chapter 2.7.1. Following from the previous chapter density function of covariates $f(X)$ is computed. Tuning parameter π_k is a prior parameter which adjusts for imbalance of classes. This parameter can be computed from the training data or can entered by expert.

$$\mathrm{P}(Y = k|X = x) = \frac{\pi_k f_k(X)}{\sum_{l=1}^{K} \pi_l f_l(X)} \tag{3.3.2}$$

For optimal classification accuracy $f(X)$ is required to follows multivariate Gaussian distribution with vector of mean values (μ) of l classes and common covariance matrix Σ. Multivariate Gaussian density is defined as:

$$f(X) = \frac{1}{(2\pi)^{(p/2)} |\Sigma|^{(1/2)}} \exp\left(-\frac{1}{2}(x-\mu)^{\mathrm{T}} |\Sigma|^{-1} (x-\mu)\right) \tag{3.3.3}$$

where p is a number of predictors. Optimal performance of LDA is achieved if covariance between predictors in both groups is identical. This represent problem in real-world settings. It is also assumed that difference between groups can be computed as a distance between mean vector of covariates. Outcome of LDA model is usually accompanied with these mean vectors and corresponding coefficients. Variable has discriminatory power if the mean values differ in both classes. Taking logarithm of 3.3.2 leads to classification rule (after some algebraic operations) which leads to class labelling.

$$\gamma_k(X) = x^{\mathrm{T}}\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^{\mathrm{T}}\Sigma^{-1}\mu_k + \log(\pi_k) \tag{3.3.4}$$

An individual is assigned to class k for which γ_k is largest. Note how the last term $\log(\pi_k)$ affects classification. This is empirically demonstrated in Table 9. When the covariance matrix is common to all classes we refer to the model as Linear Discriminant Analysis. If different covariance matrices are allowed we refer to Quadratic Discriminant Analysis (QDA). The name quadratic comes from the quadratic form of decision rule:

$$\gamma_k(X) = -\frac{1}{2}(x-\mu_k)^{\mathrm{T}}\Sigma_k^{-1}(x-\mu_k) + \log(\pi_k) \tag{3.3.5}$$

James et al. (2013) points to bias-variance problem in method selection. QDA is more flexible as it allows different covariance structures in individual classes. Especially on small samples this might result in over-learned parameters as they cannot be considered as proper population estimates. In case of large data sample when the variance is not a primer problem QDA tends to outperform LDA. In high dimensional problem QDA can be computationally heavy as the number of estimated parameter increases and it might be difficult to invert covariance matrices for all groups.

3.3.3 Regression Analysis

Regression is a set of procedures which are designed to analyse dependency between set of predictors (X) and response variable (Y). This relation is summarised through regression function:

$$f(x) = \mathrm{E}[Y|X = x] \tag{3.3.6}$$

The ultimate goal is to identify underlying function $f(\cdot)$ and corresponding regression coefficients. Identification of simple deterministic models can be done by standard algebraical way. If stochastic part is present parameters can only be estimated. Several estimators (identification techniques) were proposed in the literature. Selection of estimator should reflect nature of the relation (linearity in parameters), hierarchical design, presence of missing values and other aspects. Regression model can be formally written in several ways. First way implicitly states that only stochastic part is distributed according to some predefined distribution D.

$$Y \sim \mathrm{D}(\theta) \tag{3.3.7}$$

Secondly, model can be written as a function:

$$Y = f(X) + \varepsilon \tag{3.3.8}$$

with closer specification of the stochastic part $\varepsilon \sim (\theta)$. To highlight that estimated parameters and observed values x influence model fit, response variable Y can be described in conditional way as

$$Y|x_1,\ldots,x_k \sim \mathrm{D}(\theta(\beta,\gamma,x_1,\ldots,x_k)) \tag{3.3.9}$$

Model in equation 3.3.9 contains only two vectors of parameters β and γ but can be extended for other parameters. Values x_0 which correspond to coefficient β_0 equal to 1 and therefore are usually omitted. Sometimes, coefficient β_0 is replaced by α. Vector β contains $\beta_0,\ldots,\beta_k$ coefficients which correspond to covariates $x_0,\ldots,x_k$. Vector γ can describe behaviour of random component in hierarchical model. This can be a mean value, variance or other higher moments. Simple (univariate regression) regression model would be written as:

$$Y|x_1 \sim N\left(\mu(\beta_0,\beta_1),\sigma^2,x_1\right)$$

Regression analysis can handle various types of dependent variables. In traditional settings dependent variable is a continuous type. In many applications, however, counts or dichotomous responses are analysed. Generalized Linear Models (GLM) has to be employed to handle such variables. GLM model consists of three parts. First part (*random component*) describes conditional distribution of dependent variable, usually from exponential family. Second part is a *linear predictor*, which is a function $\eta = \beta_0 + \ldots + \beta_k x_k$. Last component is a *link function* $g(\cdot)$.

The mean parameter of dependent variable is identified as a linear combination of covariates (η)transformed by link function. This function can be non-linear, such as in case of logistic function. Inverse link is preferred to standard link because of its easier interpretation. This inversion works as:

$$g(\mu) = \eta = \beta_0 + \ldots + \beta_k x_k$$

In the previous equation mean value of dependent variable μ is not directly expressed. Inverse link function $g^{-1}(\cdot)$ is therefore used as:

$$\mu = g^{-1}(\eta) \tag{3.3.10}$$

Many link functions are available. Starting with the standard mean link, inverse-mean or log link function, just to name the most frequently used (Fahrmeir, Hennevogl, & Tutz, 1994).

Standard regression estimates conditional mean (μ) on the level of x. We can rewrite standard model:

$$Y = \beta_0 + \ldots + \beta_k x_k + \varepsilon$$

to the form highlighting the mean link as:

$$\mu = g^{-1}(\beta_0 + \ldots + \beta_k)$$

and finally in form of complete description:

$$Y \sim D(\mu(\beta_0 + \ldots + \beta_k), \gamma_l) \tag{3.3.11}$$

Prominent place in GLM model belongs to logistic regression model. Logistic model deals with dependent variable which follows binomial distribution. Binomial dependent variable can be described by probability of positive result π.

$$\mathrm{E}(y|x) = \mathrm{P}(y = 1|x) = \pi \tag{3.3.12}$$

Instead of predicting class label $\{0,1\}$ probability π is estimated. There are several functions available to properly estimate $\pi(X)$. These function are called sigmoid because of their shape. Goleţ (2014) suggests application of several functions from the most common logit, probit and complementary log-log to more robust functions allowing better handling of extreme observations. However, his reported results were not significantly (in statistical sense) better than simple logit results. Robust approaches seems to be more promising in inferential statistics then in predictive tasks.

In the right window of Figure 12 probability of an outcome being labelled as member of class 1 is described by logistic curve. Steeper curve between classes indicates better fit. Logistic model can be used for both regression (prediction and effect identification) and classification tasks. In classification tasks ideal threshold is found by optimising some classification performance indicator (chapter 3.1.2). If misclassification costs exist, threshold value is shifting away from the position $\mathrm{P}(Y = 1|\eta) = 0.5$. Identification of this threshold is demonstrated in chapter 6.9.

As noted earlier, logistic regression does not work with classes directly. Instead it works with probabilities and odds. Parameters of estimated model are in log odds and the interpretation is not intuitive. Also, because of the non-linear transformation, unit increase of independent variable does not have constant effect on dependent variable.

$$\text{odds} = \frac{P(Y = 1)}{P(Y = 0)} \tag{3.3.13}$$

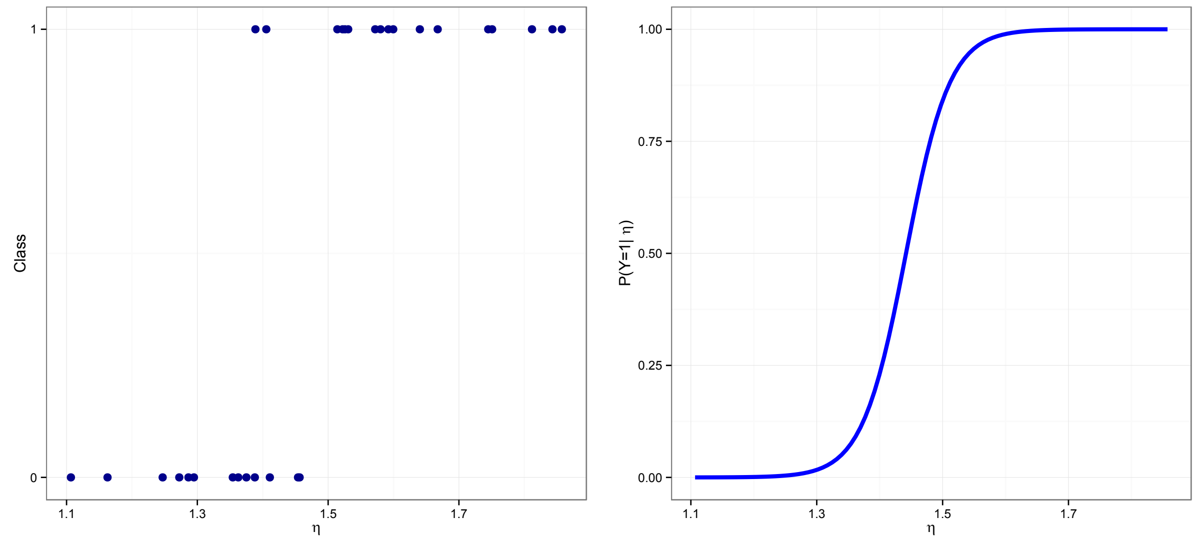

Fig. 12: *Relations between dichotomous variable (classes 0 and 1) and linear combination of covariates η is depicted in the left plot. Right plot presents logistic curve which assesses probability that subject with covariates η belongs to class 1. Source: Own processing.*

and the the odds ratio (OR) as:

$$\mathrm{OR} = \frac{\frac{P(Y=1|x=1)}{1-P(Y=1|x=1)}}{\frac{P(Y=1|x=0)}{1-P(Y=1|x=0)}} \tag{3.3.14}$$

If the OR=1 than the association between binary outcome variable (Y) and covariate x is not present. If OR is higher than 1 then the odds for $Y = 1$ are higher for subject with $x = 1$. In the context of performance analysis we are interested effect of variable (x) on odds being classified into group of good $(Y = 1)$ or bad companies. It is straightforward to estimate log of OR instead of OR as the domain of OR is limited to 0 to infinite. This transformation is known as *logit* of distribution. Let's X be a vector of covariates $X = \{x_1, x_2, \ldots, x_p\}$ for an individual observation. Using notation of conditional probability of success given vector of covariates $\pi(X) = P(Y = 1|X)$ leads to logistic model:

$$\begin{aligned} \ln\left(\hat{Y}\right) &= \mathrm{logit}\left(\pi\left(X\right)\right) \\ &= \ln\left(\frac{\pi\left(X\right)}{1-\pi\left(X\right)}\right) \\ &= \alpha_0 + \alpha_1 x_1 + \cdots + \alpha_p x_p \end{aligned} \tag{3.3.15}$$

where $\hat{Y}$ refers to odds ratio. Underlying logit transformation is a link function.

Logistic (sigmoid) function is defined as:

$$\begin{aligned} \pi\left(\hat{Y}\right) &= \frac{\exp\left(\alpha X\right)}{1+\exp\left(\alpha X\right)} \\ &= \frac{1}{1+\exp\left(-\alpha X\right)} \end{aligned} \tag{3.3.16}$$

As mentioned in chapter 3.3.3 it is convenient to use inverse link functions. Here the logit function (equation 3.3.15) is the inverse of logistic function (equation 3.3.16).

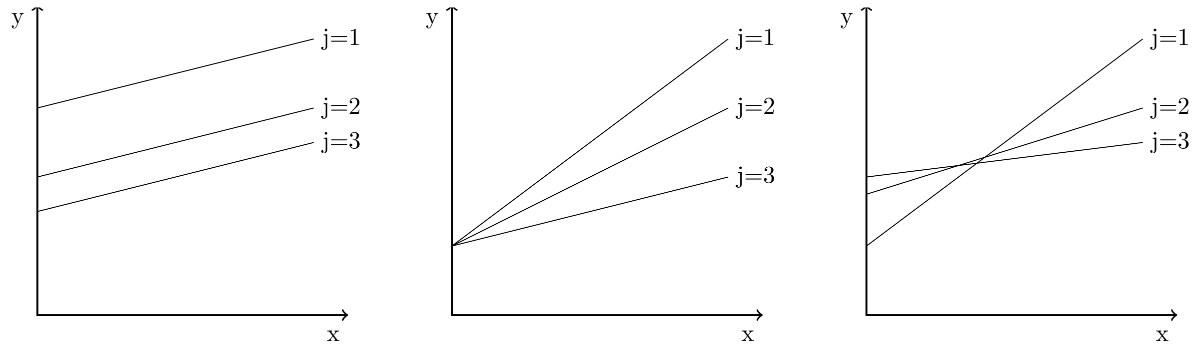

Fig. 13: *From left to right are presented: Varying intercepts model, varying slopes models and varying intercepts and varying slopes model. Source: Own processing*

This function can be estimated using maximum log-likelihood method. It is to use this methods because it's optimisation criteria can be used as a comparative fit indicator (after adjustments). When there are more than two qualitative outcomes (logistic regression can be extended to handle ordinal dependent variable) of Y are possible we need to adopt series of logistic regressions and finding several thresholds. Let's consider 3 item rating scale. Two cutting points separating classes $1|2$ and $2|3$ are needed as suggested in chapter 3.1.1.

Hierarchical Models

When analysed object belongs to some well-defined group it is advisable to use *hierarchical* modelling design. Company can be a member of macroeconomic group (industry) or it can be divided into microeconomic entities (branches). It might be researcher's aim to estimate how each hierarchical level affects performance of individual company. Jeff Gill (2013) defines *multilevel models* as group of models which account for different levels of aggregation. In some disciplines terms hierarchical and *nested* models are used more often, but the idea of decomposition is the same. Time-dimension extension to these models is extremely important; it allows for correct manipulation of autocorrelated data (this might occur when the assumption of independence of data is violated) which is present in longitudinal design. Let's subscript j denotes group of $1 \dots J$. Varying intercepts and constant slope model is described as

$$y_{ij} = \beta_{0j} + \beta_1 x_{ij} + \varepsilon_{ij} \tag{3.3.17}$$

Constant intercept and varying slopes model:

$$y_{ij} = \beta_0 + \beta_{1j} x_{ij} + \varepsilon_{ij} \tag{3.3.18}$$

and the varying intercepts and varying slopes is:

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \varepsilon ij \tag{3.3.19}$$

Graphical demonstration of aforementioned models is showed at Figure 13. Type of model should reflect initial conditions (is the risk profile the same for all companies in the beginning of the study) and individual development.

This setting is described in terms of fixed and random effects. According Jeff Gill (2013) *fixed effects* are coefficient that are assumed to have one (mean) value. It is meaningful to make inference about this value (effect of industry on returns). *Random effects* has their own distribution and are subject specific (thus we do not make generalisation statements about one value describing random effect). Model displayed in the right window of Figure 13 is also called random-intercept and random-slope model. When both random and fixed effects are present in model, we refer it as to *mixed* model. Formally, we describe mixed model for all subjects in class j as:

$$y_j = X_j\beta + Z_j b_j + \varepsilon_j \tag{3.3.20}$$

where y_j is a vector of dependent variable, X_j is a data matrix of fixed covariates with corresponding β coefficients. Z_j is a random effect design matrix for group j. Random coefficients are stored in vector b_j.

As standard models, hierarchical models are capable of modelling various types of dependent variables (such a binary or count numbers). We can rewrite previous models for subject i as:

$$\theta_{ij} = g^{-1}\left(X_{ij}\beta + Z_{ij}b_{ij}\right) \tag{3.3.21}$$

called them Generalised Linear Mixed Models (GLMM). Logistic model with binary dependent variable, which is used in many classification applications in finance, can be written as:

$$P\left(Y_{ij} = 1|\beta, b, X, Z\right) = \text{logit}^{-1}\left(X_{ij}\beta + Z_{ij}b_{ij}\right) \tag{3.3.22}$$

This model demonstrated in chapter 6.3.2.

3.3.4 Survival Analysis

Survival analysis, also known as a *event* history analysis, is used in situations where the *time until some event occurs* is modelled. This approach is suitable for analysis of *censored variables*. Values of censored variable do not appear in all possible times in the examined period. This absence has to be caused by some reason which is unrelated to the event. *Right censoring* occurs when subject enters the analysis and terminates before the study finishes. Company can merge and loose its legal status and therefore does not appear in the next fiscal year sample. This absence is not caused by bankruptcy (event) therefore we label it as right-censored variable. *Left censoring* occurs when event has already occurred but the time is unknown. This kind of censoring is rare in practice. Other perspective on left-censoring comes from two-phase study design. If the subject fails in the first phase and do not participate in the second, its complete removal would be erroneous (it might create a space for a selection bias). If the censoring occurs within two control-periods and no other information is available, *interval censoring* should be applied.

If the occurrence time is known exactly, *continuous-time* survival model is employed, otherwise we use *discrete-time* model. Discrete-time models are used more frequently because financial statements are released annually or quarterly.

In case of continuous-time survival model dependent variable is the *hazard rate*. Let's define T as a random variable "time of event occurrence". Value t is any feasible value of T. Survival function in (3.3.23) describes probability of surviving (not occurring of event).

$$S(t) = P(T > t) \tag{3.3.23}$$

If the T is a continuous random variable then $S(t)$ is strictly decreasing function. This function is a complement to cumulative distribution function $1 - F(t)$. Several approaches were designed to estimate such a distribution. It can be done parametrically – $F(t)$ can be approximated by many distributions, such a Weibull, gamma or Gompertz – or non-parametrically. Popular non-parametrical approach is Kaplan-Meier estimator.

If T is a discrete variable (control points are in set to discrete times, i.e. quarters) $S(t)$ is decreasing step-function. $\hat{S}$ is estimated as a proportion of survived to all subject who entered the analysis.

Hazard function describes rate of non-surviving (event occurrence). Hazard function is written in (3.3.24):

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t \mid T \ge t)}{\Delta t} \tag{3.3.24}$$

Hazard function is a rate rather than a probability. Nominator is probability, denominator is Δ of time (thus hazard value can exceed 1). As Δt approaches 0 hazard rate becomes instant probability of event occurrence – *potential* occurrence of event at time t given survival up to time t. Special model occurs if the $h(t)$ is constant λ over the time. This model is called *exponential*. If the λ changes over time we use other models (such a Weibull, LogNormal, etc.). Hazard function of continuous random variable T can be transformed into hazard function and vice versa.

$$S(t) = \exp\left[-\int_0^t h(u)\mathrm{d}u\right]$$

$$h(t) = -\left[\frac{\mathrm{d}S(t)/\mathrm{d}t}{S(t)}\right]$$

Hazard function is sometimes named *risk function* and function $H(t)$ cumulative risk. Cumulative risk tells how many events are expected to occur up to time t. Natural extension to the model is considering time-invariant covariates w which influence on rate of event. We can extend equation (3.3.24) as:

$$\begin{aligned} h_i(t, w_i) &= \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t \mid T \ge t, w_i)}{\Delta t} \\ &= h_0(t) \exp\left(\gamma^T w_i\right) \end{aligned} \tag{3.3.25}$$

Function $h_0(t)$ is called *baseline hazard* or *baseline risk*. Effect of k-th covariate $\exp(\gamma_k)$ is interpreted as an influence of covariate on ratio of hazards ($h_i(t, w_i)$ on $h_0(t)$). The limitation of this approach is that covariates can be only time-independent, therefore constant over time. As an example can serve economic sector or geographical location. Let k-th variable denotes binary characteristic of the company (stability of economic sector) and estimated $\exp(\hat{\gamma}_k) = -0.25$. Change from baseline $w_k = 0$ (stable) to $w_k = 1$ (unstable sector) will result in increase of instantaneous risk of an event-occurrence relatively to companies operating in stable industry by 0.25. On the Figure 15 Weibull distribution is used to approximate survival curve and hazard function. This approximation allows easier estimation of unobserved quantities. Analyst can also influence the shape by including his or her knowledge expressed in terms of prior distribution of Weibull parameters.

Cox model is not suited for endogenous time-varying covariates. Internal (endogenous) variables should be analysed on the subject level. This can be done by techniques introduced in chapter 3.3.3. Class of Joint Models models, which consists of two-step process, was proposed to work with this type of covariate. During the initial stage endogenous variables are analysed using hierarchical design. The outcome from the first stage is used in the survival analysis as an adjusted covariate. Hazard rate h_i for subject i is

$$h_i(t|H_i(t), w_i) = h_0(t) \exp\left[\gamma^T w_i + f\{\eta_t(t), b_i, \alpha\}\right] \tag{3.3.26}$$

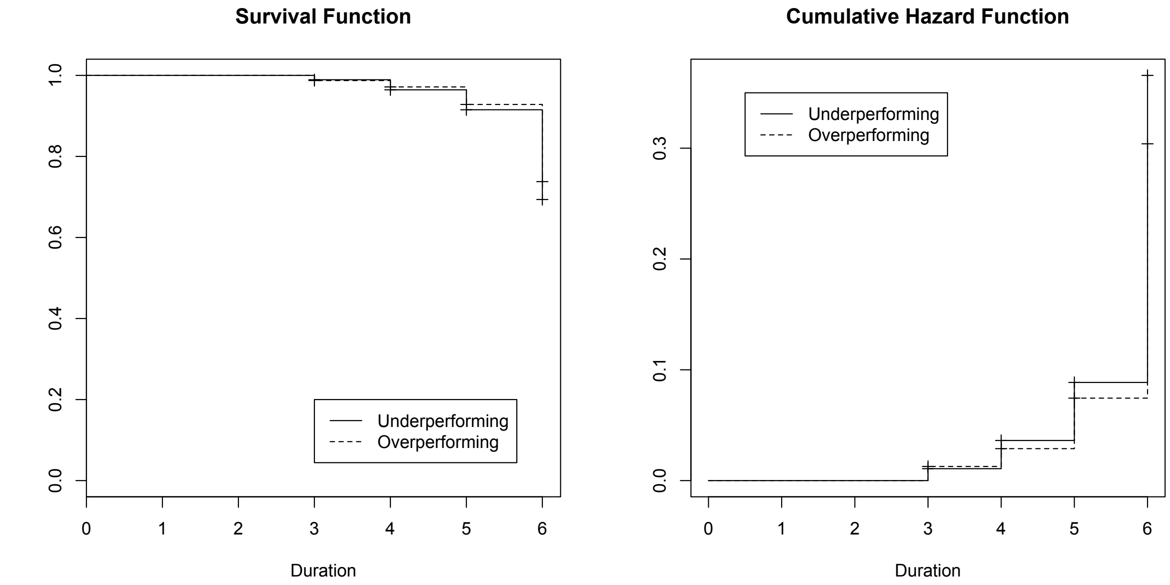

Fig. 14: *Left plot shows decreasing survival curve for two groups of companies. As expected, companies that overperform the market have higher survival rate and, conversely, lower hazard rate (right plot). Source: Own processing.*

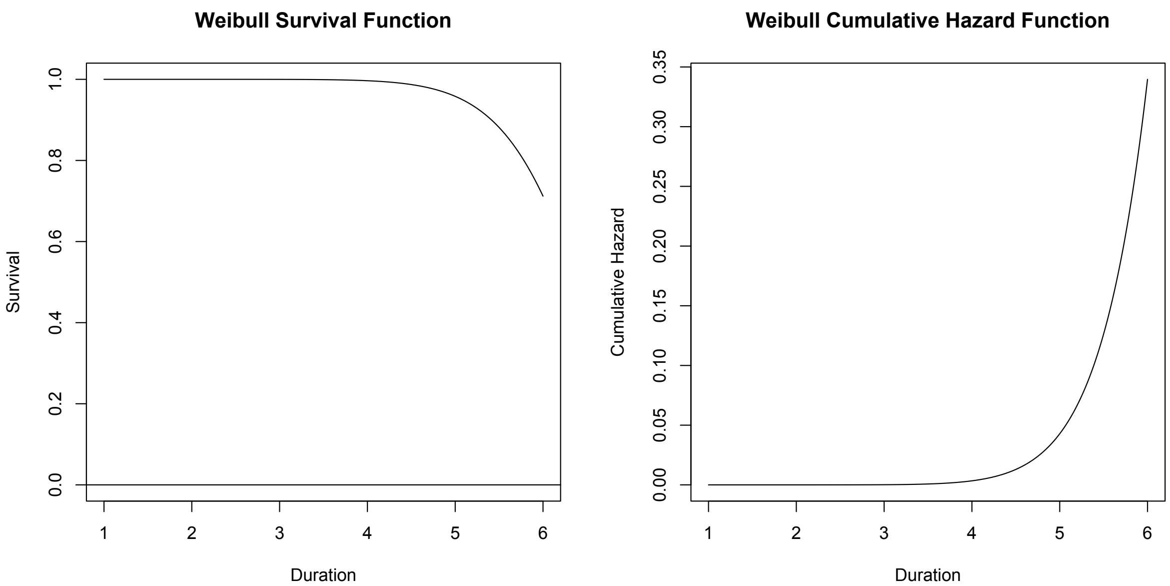

Fig. 15: *There are several distribution which are used for approximation of survival function. The best one can be found by comparing log-likelihood values or by underlying properties of the distribution, such a robustness to extreme values. Source: Own processing.*

and consists of baseline hazard $h_0(0)$, exogenous time-invariant covariates w_i with corresponding coefficients γ. This hazard is modelled throughout recorded history H_i of longitudinal outcome y_i from time $0 \leq s < t$. In the equation 3.3.26 $f\{\cdot\}$ is the density function describing the subject-specific variable behaviour. Value of α indicates strength of association between endogenous variable and hazard. Values b_i are random effects.

$$\eta_t(t) = g\left[E\{y_i(t)\,|b_i\}\right] = x_i^T(t)\,\beta + z_i^T(t)\,b_i \tag{3.3.27}$$

Illustration: Analyst might be interested in association between z-scores computed according to Altman's model from equation 2.7.1 and a hazard rate of default. She will compute z-scores (y_{it}) for all available years and she will call it history H_i. Because these scores are subject-specific, she has to treat them as time-varying endogenous variables. She will fit hierarchical model to adjust for individual behaviour. This will be done by fixed and random model which results in estimates of coefficients b. Association between adjusted variable and a hazard rate is expressed as α. Z-values will be used to compute probability of default using logistic regression specified by $f\{\cdot\}$. If $\alpha = -0.3$ then observing z-value=1 will indicate that risk decreased by 30% compared to company with $z = 0$.

3.4 Machine Learning Tools and Intelligent Systems

Algorithmic models which are inspired by biological processes and natural intelligence are named intelligent systems (Engelbrecht, 2007). Engelbrecht includes to the family of intelligent algorithms "neural networks (NN), evolutionary computation, swarm intelligence, artificial immune systems and fuzzy system. If these techniques are accompanied with logic, deductive reasoning, expert systems, case-based reasoning or machine learning systems" they form part of the field of *Artificial Intelligence* (Engelbrecht, 2007, p. 3). As Marqués et al. (2013, p. 1385) state "evolutionary computing methods can process imprecise and complex data and find hidden patterns which cannot be identified by standard techniques". This can be done mainly because these methods can learn or adapt to changes, they have good abilities for generalisation of findings, making a simple abstraction of the problem (i.e., graphical inspection of neural net), and finding associative patterns in data. Thus, these techniques were successfully applied in classification, variable selection and parameter optimisation tasks (Marqués et al., 2013).

Marqués et al. (2013) provide literature review on evolutionary computing on credit scoring methods and concludes that genetic algorithms and genetic programming have been broadly used for scoring models.

3.4.1 Decision Trees

Decision tree is a rule based non-parametric approach to classification and regression. This approach gained a lot of attention because of clear representation of induced knowledge and good performance. Several learning algorithms were proposed in the literature. They mostly differ by techniques used in rule identification. Although there are many advantages of decision trees approach (representation, no assumptions about data data or distribution, computational efficiency), disadvantages exist, too. Major concern is over-fitting. All trees start with the initial node and at least two branches which lead to less informative nodes. It is possible to learn created a tree which perfectly identifies all instances (in classification) or data points (regression) in training sample. Tree pruning is a technique which cuts the depth of the tree to achieve better performance on unseen data. Conventionally, pruning is done iteratively by comparison of error

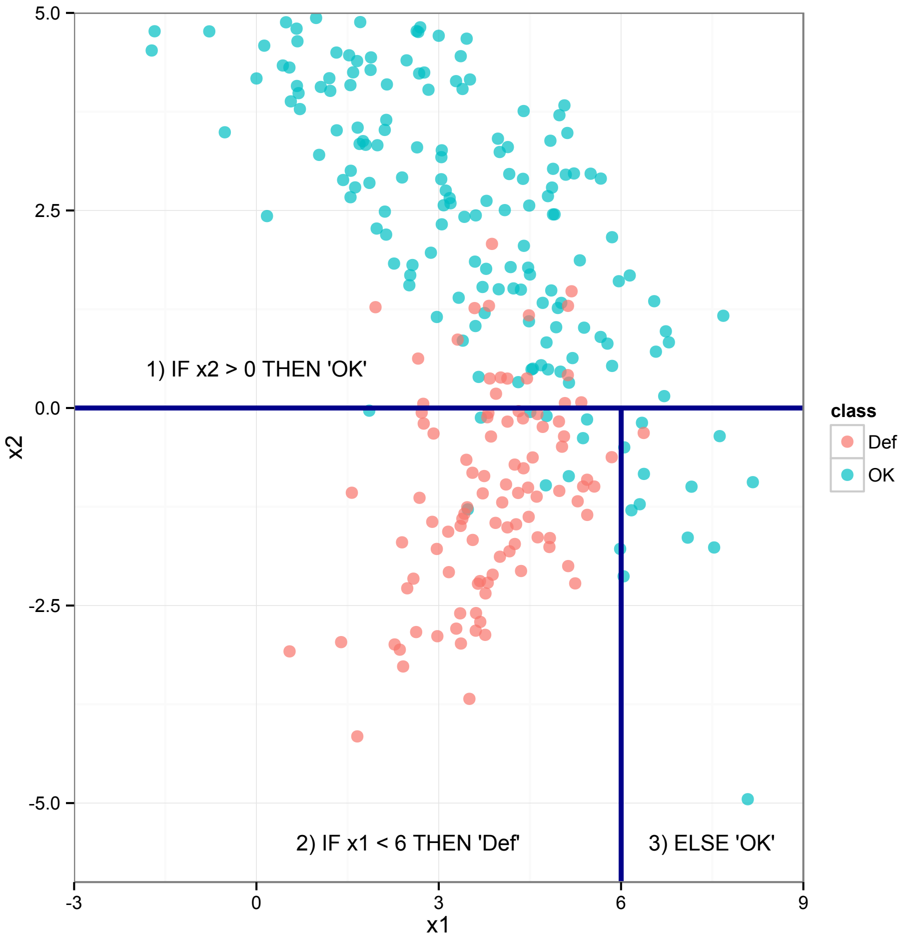

Fig. 16: *Classification tree. Two rules were applied in this example. First has split feature space by covariate X2. Second rule divided remaining region into two regions by setting boundary on covariate X1. Source: Own*

criteria change (such as MSE in equation 3.1.1) induced by increasing model's complexity to predefined threshold.

Classification trees partition feature space into smaller areas of higher density of particular class. These areas are defined in terms of IF-THEN rules. Figure 16 visualises decision tree with two numerical variables and one outcome.

Kuhn and Johnson (2013) point to several rule identification techniques. *Gini Index* (GI) is a simple approach which computes probabilities that class k occurs in the region (p_k). In each node Gini Index (for binary classification) is computed as:

$$GI = p_1(1-p_1) + p_2(1-p_2) \tag{3.4.1}$$

GI takes low value if the region for which is GI computed contains one dominant class. This identification technique

Core principle of C4.5 method (also known as J48) has emerged from information theory. According this theory information can be measured in bits and can be found by computing *information statistics* (InS) derived by Shannon (1948):

$$\text{InS} = -\{p\log_2 p + (1-p)\log_2(1-p)\} \tag{3.4.2}$$

InS works for binary classification and p value is a probability of class 1. Although this number can be chosen subjectively, usually represents proportion op class 1 on the whole sample. InS is used to determine whether to create a rule and therefore split the feature or not. This is provided by optimisation techniques based on *information gain*. C4.5 algorithm was extended by boosting procedure and adjustment in informational heuristic and is called C5.0

Bagging and boosting are employed with decision trees. *Random forest* model is inspired by bagging technique which weights bootstrapped outcomes. These models are identified on different covariates, which are selected randomly at each node split. Random forests are convenient approach in applications with many covariates.

3.4.2 Neural Networks

Artificial neural networks (NN) is an approach to both supervised and unsupervised learning. NN gained its popularity because of the ability to capture highly non-linear relations between variables. Although NN model can be visualised in an oriented graph, practical interpretation of weights and nodes is not viable in most cases. NN does not make any assumption regarding distribution of covariates. Because of this internal complexity and unclear interpretation of middle steps, NN is often referred as to *black-box* model.

Neural network is an approach which was inspired by neurobiology. Terminology used in the context of NN is affected by this origin. Model receives signals x and passes them through weights (synapses) w to processing units, neurons. Neuron processes signals by *activation function*. Activation function is usually a sigmoid function (logit), but it can be step or ramp function. If the outcome of the processed signal is binary value, neural network can be interpreted as a set of logical rules. There are several types of NN models. Engelbrecht (2007, p. 7) points to the five most important:

- single layer NN
- multilayer feedforward
- temporal NN
- self organising NN
- combined supervised and unsupervised NN

Single neuron NN can be used only in cases of linearly separable problems. Multiple neuron and multiple layer models were proposed to capture non-linear patterns. Feedforward NN is a standard approach which passes information from the receptors through neurons to the final neuron which decides outcome value. Neuron sends processed signal, but does not receive feedback (other than error outcome error) from following neurons (hence feedforward). Recurrent neural network which sends processed information at some point to the previously used neuron (through context layer) was proposed.

Neural networks are learned iteratively. Under standard settings, each iteration consists of two steps. During the *feedforward pass* parameters of activation function (weights) are identified. In the second step outcome error is sent through all neurons backwards and weights are adjusted. This approach is called *backpropagation*. As the complexity of the model rise, standard identification techniques become inefficient. Evolutionary algorithms (such as genetic algorithms) were successfully applied.

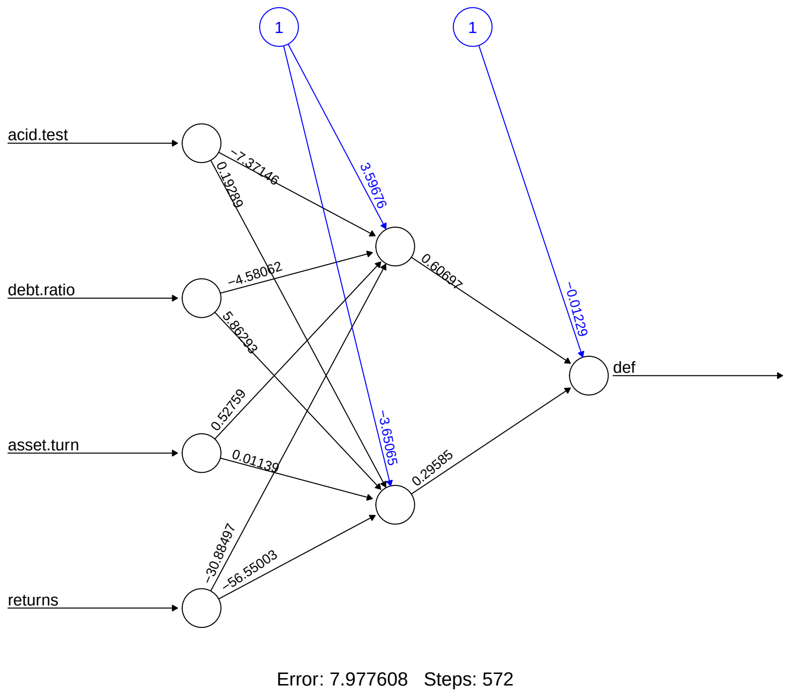

Fig. 17: *Neural network with one hidden layer. Four financial ratios represent input signals, two nodes with the middle layer represent neurons. Blue nodes are intercept values. Signals are processed by neurons. Weights are coefficients of activation functions represent importance of signal for particular neuron. Default is estimated from values computed by two neurons. Source: Own processing.*

Another extension of NN is Self Organizing Map (SOM) approach. SOM was originally designed as an unsupervised machine learning technique, but it can be used as a supervised technique as well. This technique transforms input variables from a input space and transform them into low-dimensional topological map. SOM is a neural network technique with neurons connected by lateral inhibition connections. Network of neurons is learned in several iterations. There are several ways how to create organised map, but, as the most commonly used is considered competitive learning (with winner-takes-all neurons). SOM received popularity because it allows visualisation of complex relation by topographical map of output neurons. Topological map is in Figure 18.

3.4.3 Support Vector Machine

Support Vector Machine (SVM) is a learning algorithm which gained large attention because of its high flexibility and computational efficiency. SVM can be used in both supervised and unsupervised tasks. Ideal classification model is found by identification of largest possible margin M between two classes. This margin is constructed around boundary, usually (hyper)plane or some other convenient non-linear function. Margin is found by minimising training error. This error ε_i is computed as a distance from the boundary to the incorrectly classified object. If the observation lies inside of the margin (on the wrong side) ε is ≤ 1. If the observation lies even further, $\varepsilon > 1$. Optimisation task can be written as:

$$y_i\left(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + \varepsilon_i\right) \geq M\left(1 - \varepsilon_i\right)$$

Constraint on on sum of errors $C \leq \sum_{i=1}^{n} \varepsilon_i$ can be imposed. If the problem is perfectly separable as in left panel of Figure 19, $C = 0$. In other cases C can be used as an tuning parameter which

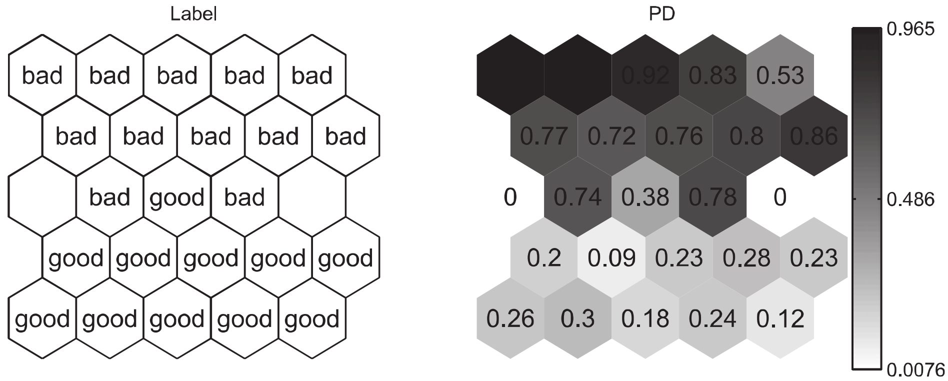

Fig. 18: *Left panel shows final classification according to probability of default regions (right panel). Note that close hexagonal parts have similar values of probabilities. Each part is defined by weights which correspond to variables. Source: (Chen et al., 2013, p. 390)*

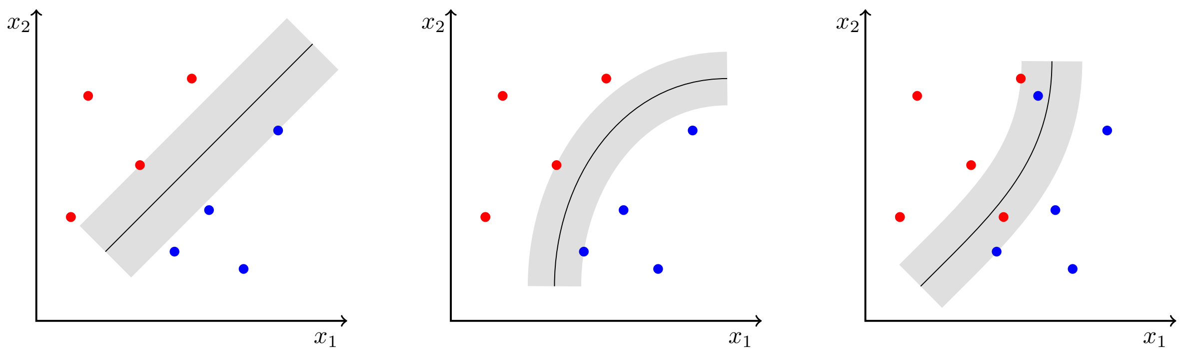

Fig. 19: *First two panels show perfectly separable problem. SVM with linear and polynomial kernels were found by maximising width of the margin showed as grey region. In the last panel classification boundary was found by allowing misclassification. Width of the region can be adjusted by parameter C. Source: Own processing*

sets the sensitivity to misclassification. SVM is very flexible approach because it can extend its feature space using *kernels*. This enlarged feature space can create non-linear classification boundary. James et al. (2013) show that estimation of the boundary can be done computing inner products of the observations. Moreover, only support vectors (set of observations S which lie on the margin) have non-zero inner products. Therefore, computation is not as intensive because not all points are used. SVM classifier can be expressed as

$$f(x) = \beta_0 \sum_{i \in S} \alpha_i K \langle x_i, x_j \rangle \tag{3.4.3}$$

where $K\langle\rangle$ is a kernel function. Most frequently used kernels are linear: $x_i^{\mathrm{T}} x_j$, polynomial: $\left(\gamma x_i^{\mathrm{T}} x_j + r\right)^d$ with non-negative parameter γ, Gaussian radial basis (RBF): $\exp\left(-\gamma |x_i - x_j|^2\right)$ and sigmoid: $\tanh\left(\gamma x_i^{\mathrm{T}} x_j + r\right)$.

Although it is possible to directly estimate class label, posterior probability of membership can be derived by Platt's method described and improved in H.-T. Lin, Lin, and Weng (2007). Posterior probabilities indicates strength of classification assessment. In the Figure 20 simple SVM model is trained on two covariates, returns and acid test. Colours in contour plot indicates probability of being classified as class 1.

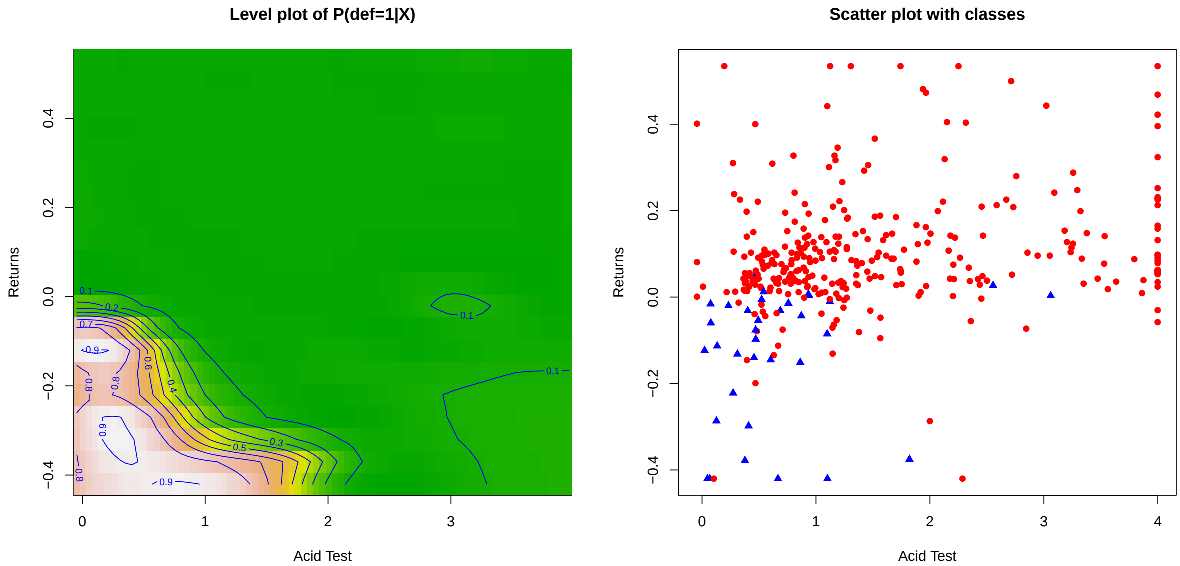

Fig. 20: *Level plot in the left panel shows classification probabilities for combinations of acid test and returns values. Non-linear boundary and posterior probabilities were found for Gaussian Radial Basis Function (RBF) kernel model. Scatter plot on the right presents observed values and corresponding classes. Triangles represent defaulted companies. Source: Own processing*

3.4.4 Clustering

In *cluster analysis* framework similarity measures adopted from mathematics (e.g., Euclidean distance) are widely used. According to Bandyopadhyay and Saha (2012) four types of clustering algorithms can be employed. Namely, hierarchical (single linkage clustering algorithm), partitional (K-means), density and grid based algorithms. Hierarchical clustering starts with randomly chosen instance. It's closest neighbour is then considered as a member of the same cluster. This bottom-up method is called agglomerative clustering. Top-down method considers all points to be member of one cluster. In the next step the large cluster is divided to create more sub-clusters with similar characteristics. This approach is called divisive clustering. In partitional clustering a reference point is set in the centre of data cloud and then connections to closest points are made. The most widely used methods are K-means or K-medoids algorithm and its derivatives (e.g., C-means which incorporate theory of fuzzy sets). James et al. (2013) point to the problem of optimal number of cluster in terms of bias/variance in both approaches, while in hierarchical clustering its easier to find cutting point.

Cluster analysis is employed in exploratory studies or as a part of ensemble. Cluster analysis is also used in missing value analysis.

4 DISSERTATION GOAL AND OBJECTIVES

This thesis aims to extend theoretical and practical framework of companies' performance measurement field by proposing modelling strategies which involve state-of-the-art analytical techniques. This will be done by extensive literature review and by utilisation of advanced analytical methods used in academic research and in practise.

Main **goal** is to develop conceptual model which describes general approach to model-building process of analytical model. Following objectives has to be accomplished to support importance of steps depicted in the model.

Objective 1 – Research Questions
First objective is to answer three research questions stated as:

1. *Which topics belong to the performance modelling framework?*
   What are the most important tasks and being analysed? Framework from the economic and analytic perspective will be defined.

2. *Which analytical techniques related to the studied phenomena were used and which were the most successful?*
   What are the most prominent researchers and practitioners and how have they contributed to the current state of knowledge? Why the financial modelling become highly mathematical and computational extensive field of study? Also, relatively small discussion is dedicated to the underlying nature of finance. Is the theory of finance close to physics, with general rules (market as the best optimisation approach), or should be treated only as time and space ranged problem with no attempts on generalisability (qualitative approach)?

3. *What can we learn from the previous research?*
   There are several obstacles in learning from others' works. Some of them are transferability, objectivity or irreproducibility of results due to the poor description of methods or other non-standard adjustments. Answering this question should help us to avoid committing old mistakes and re-entering blind alleys.

Objective 2 – Demonstrations
Second objective is to demonstrate state-of-the-art analytical techniques on real or, if not available, on generated financial data. Demonstrations will present various techniques and research designs contained in the conceptual model. This should justify their importance in the model and show their practical implementation.

5 Methodological Approach

This chapter presents the methodological background. Also, it should provide strong support to arguments and justify the usage of employed methods.

Admittedly, we face uncertainty which comes from both the stochastic nature of the studied phenomena and from our limited knowledge. In this chapter, questions related to uncertainty and developing knowledge will be presented. Topics of nature of randomness and how to measure uncertainty will also be addressed in the chapter 5.1.

An overview of the research design used in the quantitative analysis is described in this chapter 5.2. This chapter points to the limited scope of tasks in social sciences and explains why causal relations cannot be identified as in natural sciences.

5.1 Paradigms in Finance

There is an ongoing discussion about appropriateness of methodological approaches in finance. Several schools of thoughts emerged throughout the history of the economics. They mainly differ in the perception of reality. Following Kuhn (rephrased) definition of a paradigm as a *set of procedures used to solve specific problems and take theories to their logical conclusions* and division provided by Pickard (2007) we will consider these three paradigms used in methodology of science:

- positivism – reality can be understood through experiments. It is possible to remove objectivity from the analysis.
- postpositivism – recognise that reality cannot be understood completely because of imperfections of those who study it. It relies on external validation (by the process of falsification).
- interpretivism – there are many realities which are time and space bounded. Results cannot be generalised because reality is changing.

The research paradigm implies research methodology and its therefore crucial point in any scientific work. To properly identify a paradigm used by a researcher three dimensions need to be examined closely:

- ontological questions about the nature of reality,
- epistemological questions related to how we can gain knowledge,
- methodological questions related to the means of acquiring knowledge.

Ontological stand describes the nature of the reality. It sets the connection between the reality and its agents (involved subjects, such as a company operating on a financial market). *Realists* (positivism) believe that agents can be filtered out and the the reality can be fully analysed and described. It's because reality follows some general principles, rules, which can be identified in a very similar way as the natural laws (i.e., physics laws). Votaries of *critical realism* (postpositivism) believe, as realists, that there is only one reality. This reality, however, cannot be fully described. Any finding subjects to interpretation of the researcher. As Creswell (2012) points, validity of the results come from the researchers who reproduce initial findings, not from the researcher. This is caused by imperfections of agents which obfuscate the underlying principles.

The last ontological stand is *relativism* (interpretivism). Here, several realities of the studied phenomena exist and therefore any claim is valid only for a particular context and time. This branch of research is also called qualitative research.

Schinckus (2009) reviews methodological stances of the most prominent philosophers of economic theory and argues in favour of *econophysics*. Econophysics can be considered as a new approach stemming from the empirical economy and theoretical and statistical physics. Schinckus makes clear distinction between standard economic analysis (which he, unjustly, reduces to framework vaguely called "Gaussian" framework) and econophysics. This distinction is made on either analytical techniques both sides use and, more importantly, in the way the uncertainty and risks are treated.

Knightian uncertainty (based on the work of (Knight, 1964)) results from the the inability to understand the complexity of the real situation. This complexity, however, should not be confused with contingency. Agents can handle contingency by developing new knowledge (reference to epistemological stand of developing reality). If the probabilities (form of knowledge) about some phenomena are known, we are referring to risk. But, if there is no knowledge, we face pure uncertainty. According to Knight there are two types of probability situations. *A priori* computation when there is a "logical treatment" of information about a situation. The second is *statistical* calculation. The first can be characterised by large homogeneity of instances (such a toss of a coin), the latter involves situations, when the the only way of computing probabilities is on the data historical data (therefore no logical treatment can be used here). Under Keynesian framework uncertainty is dependent on the degree people understand the reality. Contrary to Knight, Keynes does not connect knowledge and uncertainty as he does not believe in "external uncertainty" which exists apart from the agents' knowledge and behaviour. If we want to handle uncertainty we have to develop generally accepted stances - conventions. Lastly, Hayekian uncertainty (based on the work of Friedrich von Hayek) refers to either *internal* and *external* uncertainty. Internal rises from the lack of knowledge, external from the (natural) laws of the studied phenomena. Hayek (1942, p.288) states that "The reason of the difficulty which the natural scientist experiences in admitting the existence of such an order in social phenomena is that these orders cannot be stated in physical terms, that if we define the elements in physical terms no such order is visible, and that the units which show an orderly arrangement do not (or at least need not) have any physical properties in common... It is an order in which things behave in the same way because they mean the same thing to man."

Based on the discussion in the previous paragraphs tools used for handling uncertainty can be various. It always depends on the way how the researcher treats uncertainty. For many economic phenomena we can make only statistical computations (in the Knight's meaning). Physical laws cannot be directly transferred to social sciences (Hayek), but in the terms acquiring knowledge some techniques from the natural sciences might be useful (econophysics approach).

5.2 Research Design

For the purpose of this thesis term *research design* covers *study, data* and *sampling design*, methods and techniques used for conducting research. From the perspective of study design we recognise:

- experimental design – all factors are controlled by the researcher
- pseudo-experimental design – only some factors are controlled

- observational study design – form of pseudo-experimental design with paired subjects which is used when treatment effect is the object of enquiry.

Observational study design fixes some subject's characteristics by finding an appropriate counterpart. This matching is made by means of *propensity score*, which is usually computed by logistic regression (Rosenbaum, 2010). It is also often used when numbers of classes are highly unbalanced. Problems with sparse data can be partially solved by using appropriate analytical methods (i.e., zero-inflated models). Observation design (the old term paired design) can help as well. In the context of failure, the modelling analyst faces the problem of a small number of failed companies. As Beaver (1966) states paired design (and consequent paired analysis) can mitigate effects of company size and industry. This approach brings one important drawback - it is not possible to draw inferences about single company. This approach is suitable for purposes when some general fact is examined (riskier companies tend to have higher values of indicator A).

We can distinguish between these data designs:

- cross sectional – a set of observations collected at one time
- longitudinal – observation are collected repeatedly from the same individuals

Sampling design defines how data was collected. Two major classes are recognised:

- probabilistic sampling – subjects which create samples are selected randomly from the population
- non-probabilistic sampling – the sample is not randomly drawn from the population

Sampling design is an important topic influenced by many economic and methodological needs. Sampling design has to be examined clearly in the beginning of the analysis, as the results can be substantially biased if improper design is selected.

If experimental design with random sampling is correctly employed, then results have high *internal validity*. It is the only case when *causal inference* can be done.

Successful inferences can be only achieved upon the methodologically sound procedures in the long run. An expert has to warrant that selected instruments and employed methods lead to *valid* conclusions. Some authors argue that *validity* is the feature of the preposition, not the data itself. However, Hair (2010, p. 7) defines validity as "the degree to which a measure accurately represents what it is supposed to." In both theoretical and empirical economics validity "tends to be equated with the accuracy or 'truthfulness' of results". (MacPhail, 1998, p. 119). Moreover, term validity is usually avoided and replaced by terms which have statistical meaning, such unbiased or statistically significant. In many fields of social sciences term validity has received closer attention. For the purpose of this thesis we will consider:

1. *Construct* validity – whether the variables measure what they were designed for. It is expected that variables with the same or similar theoretical meaning will behave in a similar way (*convergent* validity). Also, variables which were designed to measure different dimensions (liquidity and profitability) should be clearly recognisable from themselves (*discriminant* validity).

2. *Content* validity – this validity is relevant to the validity of the results. Content validity assures that all relevant components (i.e., variables and their relations) were considered and are reflected in the results.

3. Predictive validity – results have out-of-sample predictive power. This concept is related to *external* and *internal* validity. Results obtained from the research with high external validity can be generalised. Results with high internal validity are usually obtained on the experimental design where some factors are filtered out to allow a more precise estimation of the effect. In the real world settings these factors cannot be fixed and therefore results fail when they are employed outside of the original experimental space.

5.3 On Defence of Quantitative Finance

As mentioned earlier, financial modelling involves analysis of irrational agents and very complex systems of interactions. Indisputably, expert's knowledge might represent the most important part of the overall analysis. Automatised systems can provide some insights which are too abstract to identify. Marqués et al. (2013) summarise findings from several authors on benefits resulting from the usage of automatised scoring systems over "subjective methods":

1. consistency of recommendations which eliminate biases introduced by human experts,
2. logical rules which can better/faster accommodate new policies,
3. performance of models can be easily tracked, analysed and adjusted,
4. reduction of costs implied from the risk evaluation,
5. savings in time and effort

Although the last two can be viewed less important when considering possible losses introduced by malfunctioning systems, benefits of the first three are clear. Consistency is an important feature of the system because it allows human experts who make the final decision stable and comparable results over individuals and time. The need for variability of rules can be driven by need of meeting new policies, trading strategies, or just to generate various retrospective/perspective scenarios. Bahrammirzaee (2010) emphasises the non-linear behaviour which is modelled in financial applications. The necessity of addressing this kind of behaviour gave rise to artificially intelligent techniques (AI) over traditional statistical models and expert judgements.

5.4 Implications for Thesis

My methodological stance is postpositivism. This thesis is therefore restricted to quantitative methods and thus qualitative techniques are omitted. This is reflected in methodology and methods (statistical analysis and machine learning) which I use in my argumentation. Also, many models in demonstrations do not aim to analyse the reality and estimate effects of variables. The main purpose is to adjust the model to data to mimic the data generating process closely. In such cases research design does not have to be the same as in research which aims to derive generalising conclusions (i.e., random sampling cannot be achieved in bankruptcy models because bank's clients are not selected randomly from all potential clients).

In this dissertation I make no assumption about “Gaussian” reality as suggested by Schinckus (2009). My findings in the most cases aim to exhibit high predictive power and do not serve as a vehicle for further generalisations; they are not searching for true, intrinsic principles which are, in my opinion, too complex to be appropriately captured by automatised analytical methods.

6 MAIN EMPIRICAL RESULTS OF THE THESIS

This chapter contains empirical results which should confirm importance of all steps which appear in the conceptual model. This model is presented on page 35. Due to the nature of conducted research, main results lie in answering research questions and in proposing conceptual model. All research questions were answered in chapters 2 and 3. Empirical results cannot be considered solely as the main result of the thesis. Empirical results demonstrate importance of findings obtained from the first chapters.

Methods and techniques will be illustrated on the accounting data which was exported from database Albertina (1/2015) if it's not otherwise stated. Data sample consists of 446 companies. Selected companies operated or still operate in the rubber and plastic industry during time period 2008-2013. Companies were not selected by random. The most important selection criteria was the availability of financial data (ebitda, financial expanses, net worth and total debt) throughout the analysed period. Some variables were computed by probabilistic imputation (financial interests were computed by multiplying random value from normal distribution $N(0.05, 0.01)$ and size of liabilities). Financial statements summarising the full period fiscal year were considered only.

All of the aforementioned adjustments create a space for biased results. It also violates assumptions for inferential statistics. However, this data set serves as a training and testing set for classification, prediction and decision making tasks. These tasks do not require random sampling. Moreover, in practise, analysed companies are not randomly selected too.

Values of five multiples were computed. These variable are representatives of five dimensions described in chapter 2.6.1.

- Liquidity: Acid test = (current assets-inventory)/short term liabilities
- Leverage: debt ratio = debt/total assets
- Efficiency: working capital turnover = sales / working capital
- Profitability: operational income / total assets

Definition of financial distress given by Keasey, Pindado, and Rodrigues (2014) was followed. Variable EBITDA used in the original definition was replaced by EBIT to enlarge number of potentially distress companies (this adjustment resulted in doubling the number of distress companies). Data set 2011 contains values from year 2011 and 2012 as default definition involves annual changes in some variables.

First chapter is dedicated to application Bayesian analysis. Expert's knowledge expressed in terms of prior distribution is used to extend analysis which was performed on collected data. Following chapter deals with sensitivity analysis. It demonstrates how LDA model results change as expected proportion of majority class changes.

All proposed models were built as cost sensitive. Throughout the text predicted values of outcome variable y will be denotes as $\hat{y}$. Computational perspective described in chapter 3.1.2 was used to design cost matrix in Table 5. Cost values were estimated to minimise type I error. Values in table inform that it takes on average 4 profitable contracts to cover losses from one type I misclassification. This error (type I) occurs when company is assigned as healthy ($\hat{y} = 0$) but ends distressed $y = 1$. If managerial perspective was used, virtual benefit resulting from early identification of risky transaction (valued 15) would not be as high as in the table. Values in table were estimated by author. In chapter which aim to find best performing algorithm general schema will be followed:

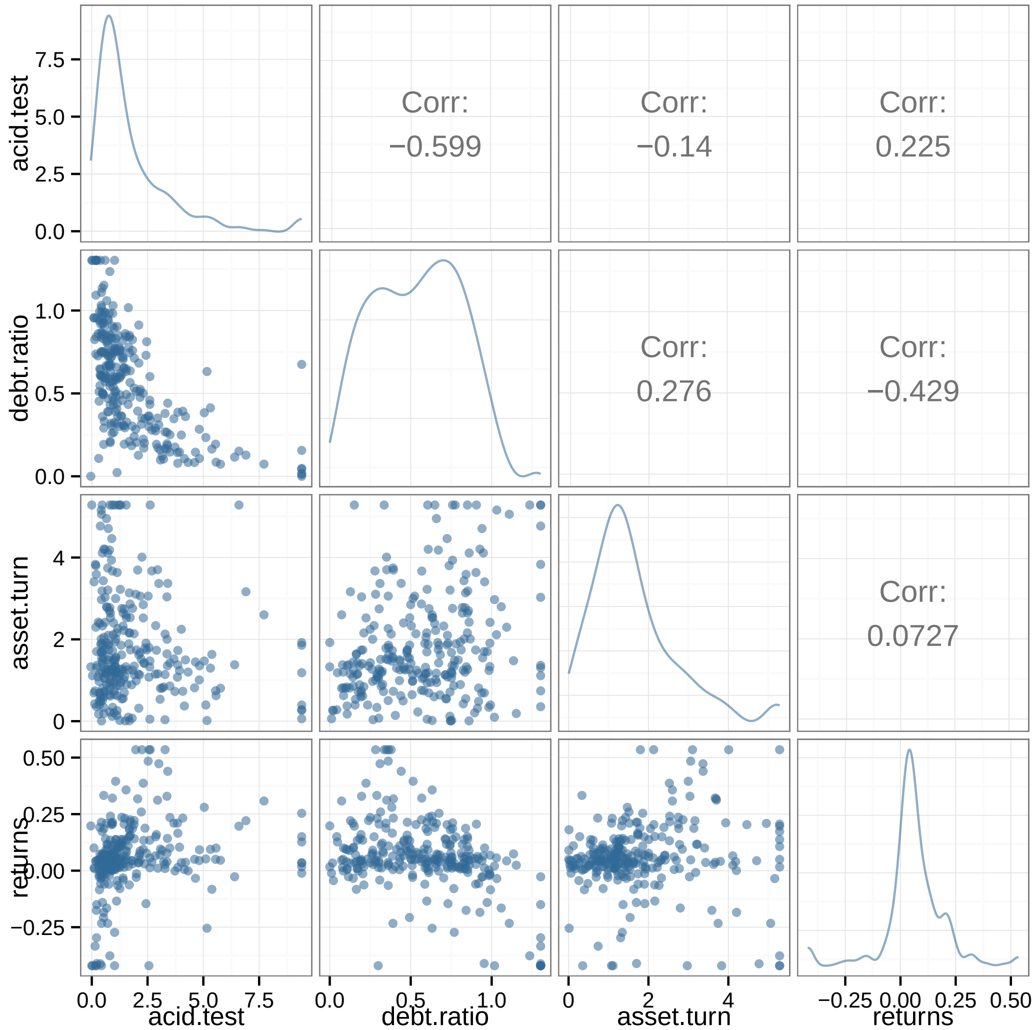

Fig. 21: *Exploratory analysis of 2012 data. Correlation coefficients are meaningful only if the relation between variables is linear. From empirical distribution function shown at the main diagonal can be concluded that variables are not normally distributed. This is not required, though. Source: Own processing*

Tab. 5: *Cost matrix summarises combinations of situations. Predicted value $\hat{y}$ and real value y indicated whether company is in healthy conditions (0) or is distressed (1). Source: Own processing*

	non-defaulted $y = 0$	defaulted y
$\hat{y} = 0$	5	-20
$\hat{y} = 1$	-5	15

1. Model definition – This step involves selection of variable and starting parameters.
2. Setting objective function – Objective function Q reflecting cost matrix (tab. 5) is used as a main criteria. In some demonstrations other metrics are used as a supplement to demonstrate how this selection affects model behaviour.
3. Model fitting – Fitting procedure involves optimisation over predefined space of parameters. Threshold value T is reported only if it helps in interpretation.
4. Classification – prediction of probabilities or direct class assignment $\hat{y} = 1$ is made on sample from year 2012.
5. Performance assessment – classification performance of best model will be reported in confusion table. Other metrics and plot will be used if needed.

In all analysis data from 2011 serve as a train sample. Data from 2012 are unseen in model-building stage and are only used in evaluation stage.

Table 6 describes compositions of throughout the analysed period. Minority class represents less than 10% in all years.

Tab. 6: *Number of defaulted companies (Default=1) according adjusted definition of financial distress. Note that 21 defaulted companies are labelled in 2012. These companies met conditions for default assignment on 2012 and 2013 data. Source: Own processing.*

Year	Defaulted	Count	Proportion
2008	0	251	
2008	1	19	7.04%
2009	0	286	
2009	1	17	5.61%
2010	0	325	
2010	1	19	5.52%
2011	0	316	
2011	1	34	9.71%
2012	0	250	
2012	1	21	7.75%

6.1 Bayesian analysis

This chapter demonstrates application of Bayesian inference on two binomial problem. Results are compared to standard approach. Different interpretation of results is emphasised.

Investor evaluates companies according to rule which results in dichotomous outcome. This rule can test whether company's performance beat market predictions in the last two year or whether company met their obligations to public institutions, etc. Let's consider two groups of companies $k = 1,2$ where $k = 2$ indicates bad companies. Bad company can be defaulted company or company which does not create value to stakeholders.

If the company i failed the test then $y_{ik} = 1$ otherwise 0. Historical proportion of companies in group k which failed in the test is denoted as p_k. This value is computed as a sample proportion. Analyst wants to generalise her findings, therefore she is interested whether population characteristics π_k differ. Let's denote this difference of proportions as $\gamma = \pi_2 - \pi_1$.

Analyst is interested whether proposed test can distinguish between groups of companies. Binomial random variable is a convenient random variable for analysing proportions. If the $\gamma \geq 0$ than analyst can conclude that the proportion of failed companies in bad group is higher then in the good group. Test can be employed in practices as it can distinguish between the groups.

From the historical records analyst concludes that from 200 good companies only 16 companies failed in the test. This means that $p_1 = 0.08$. From the $n_2 = 120$ bad companies 18 failed (p_2 = 0.15). Point estimated of γ is 0.07.

If the standard (frequentist) approach would be selected than two-proportional test with would be employed. Test outcome $\chi^2 = 3.168$ with 1 degree of freedom suggests that null hypothesis should be rejected in favour of alternative hypothesis ($\gamma > 0$) on the standard $\alpha = 0.05$ level as the p-value = 0.038. Lower value of 95% confidence interval is 0.0011. If it is true that $\gamma = 0$) than only 3.8% of all repeated analysis would found such data as analyst observed in this analysis and "worse". By worse is meant difference of 7 and even lower. This reflects the alternative hypothesis. Analysts knows that if the analysis would be repeated many times 95% of all computed confidence intervals would contain the true value of γ. If she believes that she is not amongst those 5% whose results were corrupted by sampling error then she concludes that γ is at least 0.0011.

Analyst can approach the problem from Bayesian perspective as well. She needs to specify the distribution which describes proportions. This will be the same distribution as in previous case. Because she wants to minimise her subjective influence on the results, she designs prior distribution as uniform. This means that all possible π values are equally likely to be true. Prior distribution on both π_k are set to $\beta(1,1)$. This prior was visualised in Figure 10.

Although this problem can be solved analytically as the prior (Beta distribution with parameter a=1, b=1) is conjugated to posterior distribution, model will be estimated by Monte Carlo procedure (Gibbs sampler).

Diagnostic plot in the left panel of Figure 22 is an important tool for assessing stability of proposed solution. This plot shows whether sampling procedure converges to some solution or not. This should be verified by computing independent chain(s). It usually takes some time to find region close to the optimum. Chain values are obtained on initial values which can be either generated or estimated by analyst. Such values can be far from optimal values and this influence the chain (chain gets very volatile) and this might negatively affect final results. This can be prevented, as it was done in this example, by computing burn-in sample in the initial phase of the parameter identification. In this demonstration 1000 samples were discarded and estimation was made on remaining 5000 samples. This plot presents only last 2000 samples. Samples can be considered as stationary as the mean value as well as the variance does not change. Presence of autocorrelation was not found too. Estimated values of proportions and γ can be considered as correctly computed.

Highest posterior density regions were computed on both chains and results are showed in Table 7. It can be concluded that probability that γ is in the interval $\langle 0.001, 0.148 \rangle$ equals to 0.95 for the first chain. HPD interval computed on the second chain contains negative value. Estimated difference in proportions is around 0.071 (distribution is almost symmetrical so mean and medium values have the same weight). Standard error of this estimate is 0.03821 (not written in the table). Difference of 7.1% might be substantially significant (important for decision) to the analyst. Corresponding uncertainty, however, suggests that this observed difference can be attributed to sampling error. Systematic influence cannot be rejected, though.

In the following text two binomial problem will be expanded by updating prior information. Consider situation when domain expert handle analyst his experience, which was collected by

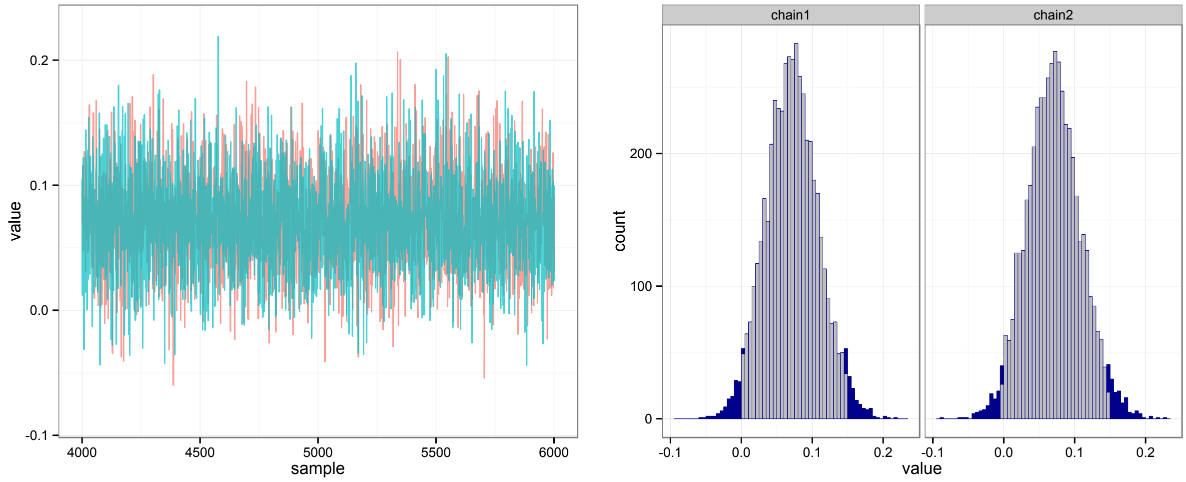

Fig. 22: *Left panel presents last 2000 iteration of two Markov Chains. This plot shows that two independent chains are converging. Solutions can be considered as stable. Right panel presents histogram of γ. Values in blue regions lie outside of the 95% empirical HPD interval. Source: Own processing*

Tab. 7: *Results of two chains show similar values. Column hpd.025* shows lower bound of HPD interval. Credible interval is delimited by *q.025* and *q.975* columns. Source: Own processing

Chain	hpd.025	hpd.975	mean	median	q.025	q.975
chain1	0.001	0.148	0.072	0.071	0.001	0.149
chain2	-0.003	0.147	0.071	0.07	-0.001	0.149

means of previous analysis. Expert suggests that it is not reasonable to expect all π values to be equally likely to be true. This expert has an experience with portfolio which consisted of 60 good companies and the proportion on that portfolio was 6.67%(four companies). In the portfolio was 18 bad companies and three of them failed the same test (p=16.67).

Parameters of Beta distribution can be estimated to match his previous experience as using equations (Kruschke, 2011).

$$a = mn \quad \text{and} \quad b = (1-m)n$$

where m is the expected proportion and n sample size which which determined value m. Prior distribution for group of good companies is beta$(4,56)$ and for bad companies beta$(3,15)$. Three prior distributions are showed in Figure 23. Higher variance of the "bad" group prior reflects that less information was used in its determination.

Table 7 presents final results which reflect both historical data and expert's knowledge. Difference in proportions is now higher (around 7.5%). Standard error of this estimate decreased slightly to 0.035.

Incorporating expert's opinion changed results. Now, evidence seems to be more conclusive in favour of the classification rule. Proposed rule should be applied in classification between

Tab. 8: *Point estimates as well as the intervals changed. Neither of hpd or quantile intervals contain null after informative prior distribution was used. Source: Own processing*

Chain	hpd.025	hpd.975	mean	median	q025	q975
chain1	0.011	0.147	0.075	0.074	0.009	0.146
chain2	0.009	0.145	0.075	0.074	0.011	0.149

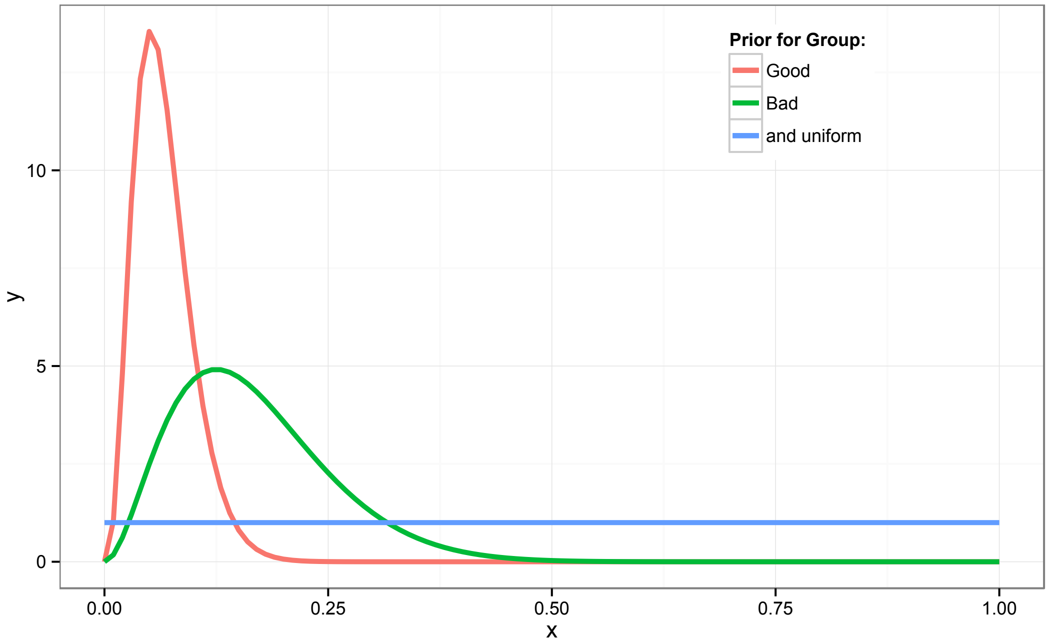

Fig. 23: *Three beta prior distributions describe expert's knowledge. Note that axes do not have the same scale. Source: Own processing*

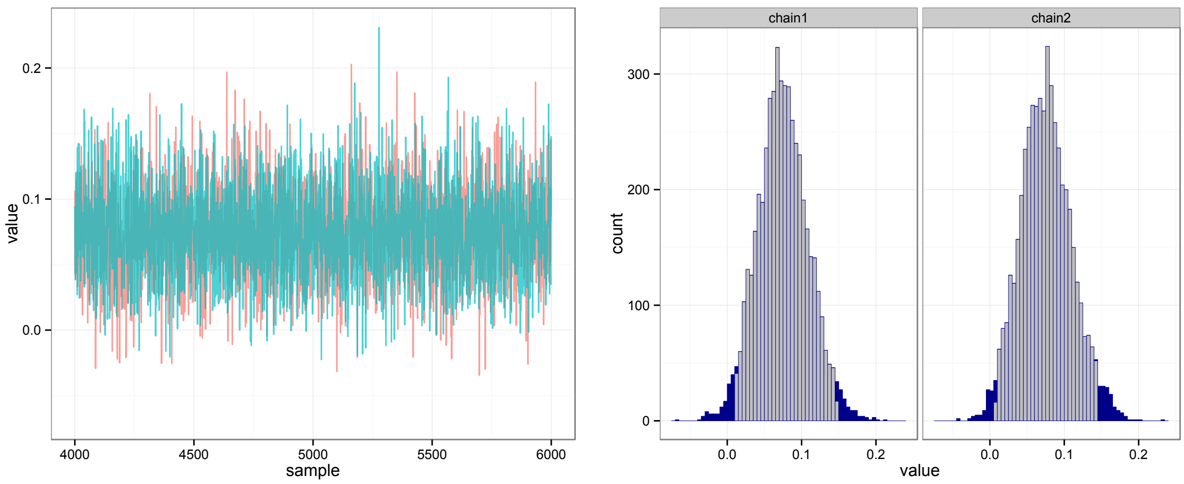

Fig. 24: *Both chains converged as in the previous analysis. Histogram in the right panel have similar shape but location (expected value) shifted to the right. It indicates that the difference in proportions is higher. Source: Own processing*

good and bad companies as it has discriminatory power. Companies which fail in the test are more likely to be defaulted than companies that pass.

6.2 Discriminant Analysis

In this chapter Linear Discriminant Model will be used to demonstrate how to include expert's knowledge in classification task. LDA was described in more details in 3.3.2. Although LDA approach stems from classic probability theory, it allows incorporation of prior information (information known before the data was obtained). Analyst can include expectation about class proportions. If the economic outlook is negative, proportion of defaulted companies can be adjusted accordingly. Increased value of parameter π will results in higher posterior probabilities and more companies would be classified as defaulted. If there is no prior information about the proportion, parameter's value can be estimated from the sample. Prior probabilities can be therefore considered as a tuning parameter which can significantly influence results. Special treatment is given in observational studies which balance classes to that extent that model does not mimic the true economic reality (i.e., original Altman's model). Such models might have good predictive predictive power, though. In the following example LDA was learned on the whole 2011 dataset. In the final comparison, performance of cross-validated model is presented.

```
Prior probabilities of groups:
         0          1
0.90285714 0.09714286

Group means:
  acid.test debt.ratio asset.turn    returns
0 1.5750527  0.5499797   1.860837  0.1031426
1 0.6774244  0.9157720   1.928767 -0.1322076
```

Mean values in particular groups are values of μ in equation 3.3.2. Importance of variables for model can be assessed by t-statistics or, if the variables' variances differ, by Wilcoxon Rank Sum. Although biggest differences were found in conditional means of acid test and returns, variance of those variables has to be considered. Also, multicollinearity affects the coefficient. Different group variances of acid test variable explain lower importance which is reflected in the model's coefficients.

```
Coefficients of linear discriminants:
                    LD1
acid.test   0.038
debt.ratio  1.798
asset.turn  0.000
returns    -5.758
```

Positive coefficients indicate that change in corresponding variable increases probability of being classified in the defaulted class. Value of expected proportion of majority group does not affect coefficients. Posterior probabilities, however, change. This is showed at Table 9.

Outcome of the LDA can be either probability as showed earlier, or score. In the following analysis optimal thresholds have to be found. This step is visualised in Figure 33 for logistic model.

Tab. 9: *Sensitivity of LDA posterior classification probabilities. As the certainty supplied by analyst grows (prior proportions showed in the second header-row are further from (0.5, 0.5)) probability of being in major group declines. Source:Own*

	P(Y=0|X)		
Subject	(0.95, 0.05)	(0.85, 0.15)	(0.5, 0.5)
1	0.997	0.990	0.943
2	0.993	0.976	0.878
3	0.993	0.977	0.885
4	0.996	0.986	0.927
5	0.996	0.987	0.933
6	1.000	1.000	0.999
7	0.990	0.968	0.843
8	0.997	0.990	0.946
9	0.992	0.974	0.867
10	0.985	0.952	0.777

6.3 Generalised Linear Models

This chapter demonstrates application of family of models described in 3.3.3. GLM models will be identified on cross-sectional and longitudinal data design.

6.3.1 Cross Sectional Analysis

Logistic regression model was computed in all available years. We consider following model formally defined at equation 3.3.16:

$$P(y_i = 1|X_i, b) = \frac{1}{1 + \exp^{-z_i}}$$

where $z_i = b_0 + b_1 \cdot \text{acid.test}_i + \cdots + b_4 \cdot \text{returns}_i$. All companies are recorder only once and the only dynamic element in the model is the default variable, which implicitly computes annual changes of variables. Estimated regression coefficients are shown in 10. Coefficient cannot be interpreted directly as they are not on a linear scale. Unit change does not have constant effect on probability of company being in defaulted class. Consider results from 2012. If returns increase by 10% from 1% to 11% probability of being in class decreases:

$$\frac{1}{1 + \exp^{3.23+5.08\cdot 0.11}} - \frac{1}{1 + \exp^{3.23+5.08\cdot 0.01}} \doteq -0.014$$

by 1.4%. Now, consider returns increase from 0.11% to 0.21%:

$$\frac{1}{1 + \exp^{3.23+5.08\cdot 0.21}} - \frac{1}{1 + \exp^{3.23+5.08\cdot 0.11}} \doteq -0.0088$$

This return increase is not as important for class-assignment as rise of returns from 1% to 11%. Probability of being in class 1 (defaulted) has decreased only by 0.8%. Resulting probability values are small because of the highly unbalanced class proportions. Such results do not have direct economic interpretation (they cannot be considered as *probability of default* indicator in Basel regulatory standards). Though, they are useful in classification tasks because they are compared with appropriately selected threshold.

Tab. 10: *Values of regression coefficients. Intercept value and returns are statistically significant coefficients* throughout all periods. Source: Own processing*

Year	Intercept	acid.test	debt.ratio	asset.turn	returns
2009	-2.731*	-0.291	0.831	-0.188	-5.813*
2010	-2.735*	-0.423	1.672	-0.500*	-6.028*
2011	-2.063*	-0.850	1.402	-0.146	-11.313*
2012	-3.219*	-0.905	2.134	-0.047	-5.055*

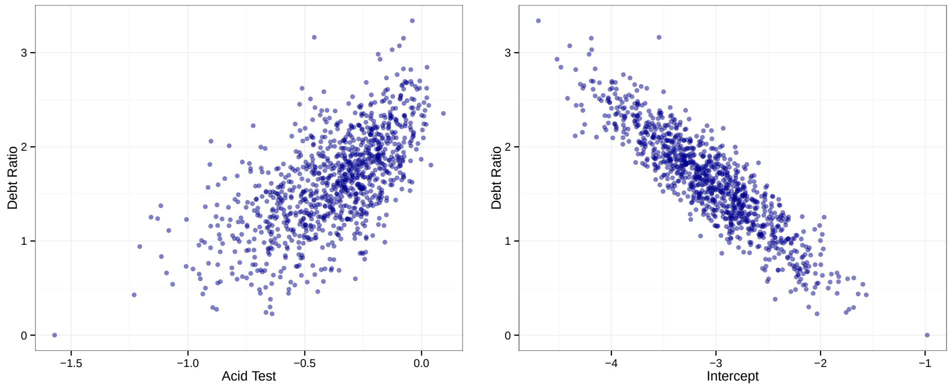

Fig. 25: *Bi-variate analysis of bootstrapped coefficients reveals that coefficients' values vary significantly. This instability is the reason for low interpretable value of coefficients. Source: Own processing*

Negative sign of regression coefficient indicates that association between probability of being in a defaulted group is negatively associated with corresponding variable. Negative value of coefficient "returns" means that increase in returns will result in decreasing probability of default class membership. All coefficient have expected sign.

Presence of multicollinearity might lead into misleading judgements about estimates and corresponding errors. Some correlations exhibit large values (see Figure 21) which results in high variance in estimated coefficients. To demonstrate such a problem Figure 25 was created. Figure reveals high instability of coefficients which might represent problem if the main model's purpose is to describe variables' effects. An implication of this empirical evidence is that analyst cannot attribute interpretative value to a single estimate unless correlation matrix does not contain any significant pair correlations. Relationship between observed pairs of estimated coefficients show that low coefficient value is balanced by high value of another. The overall effect is, however, the same. This is an important consequence which provides sound support to usability of forecasting models without "economically meaningful" coefficient values.

It is convenient to describe uncertainty of estimated coefficients by standard errors, confidence intervals and outcome of statistical test, such a t-test. These standards test can be successfully applied if underlying conditions are met. Uncertainty can be estimated by employing bootstrap technique, too. Bootstrapped estimates were computed and are showed at Figure 26. Histograms reveal that sampling distributions are not symmetrical for all parameters. Corresponding bootstrapped confidence intervals would capture this asymmetrical behaviour better than traditional approach which would be symmetrical around point estimate which are in Table 10.

We can consider two types of prediction of $P(y = 1|x, b)$. First is a point estimate which can be obtained by multiplying coefficients showed in Table 10 by corresponding covariates. This value would suffer from lack of information about the uncertainty of the estimate. Second approach is to analytical estimation of predictive intervals. In some situations (such as interval for Value at Risk) finding and solving analytical form of prediction intervals can be cumbersome. In

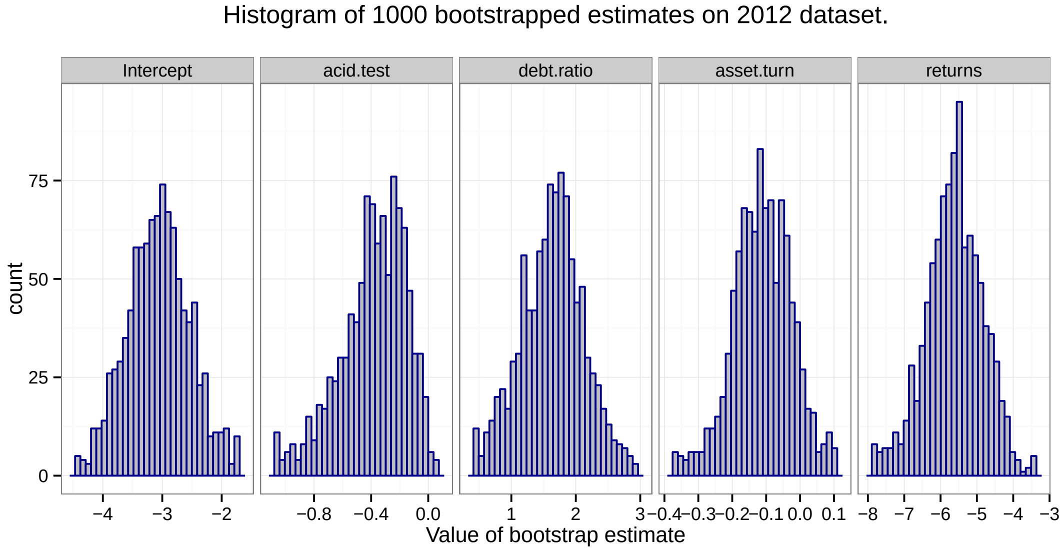

Fig. 26: *Bootstrapped estimates of 1000 replications were computed. Histograms show both central tendency, which is close to previously estimated values in Table 10 and corresponding uncertainty. This uncertainty is expressed as the shape and range of x values. Source: Own processing*

Tab. 11: *Descriptive statistics of bootstrapped predictions. Values of prediction intervals describe uncertainty of the estimated values. Prediction interval of 90% is delimited by q05 and q95. This interval contains 90% of all estimated values. It might be useful to use technique highest posterior density region to identify such an interval in case of asymmetrical prediction boundaries. Source: Own processing*

ICO	year	q01	q05	q25	q50	q75	q95	q99
543551	2008	0.07	0.08	0.09	0.10	0.11	0.12	0.13
543551	2009	0.03	0.03	0.04	0.04	0.05	0.06	0.07
543551	2010	0.03	0.04	0.04	0.05	0.06	0.06	0.07
543551	2011	0.04	0.05	0.06	0.06	0.07	0.08	0.09
543551	2012	0.01	0.02	0.02	0.02	0.03	0.04	0.04

this example bootstrapped intervals are computed instead. Regression functions obtained from replicated samples are used to provide point estimates. These point estimates are than analysed by both descriptive statistics and visually. A convenient way for description such a data is a fanplot.

Fanplots in Figure 27 summarise evolution of probabilities of being classified as defaulted company. Note how uncertainty changes throughout the years and subjects as well. Analyst can be more sure about the classification as the width of fanplot is narrower. Estimated probabilities from fanplot are showed in Table 11.

Results are plotted as time series object to demonstrate variability of estimations. Analysts cannot directly conclude whether probability of default increases or not. It might be the case when general level (intercept) add or removes some amount of probability value. This is also reflected in threshold values which which differs from year to year. For example, exceeding probability 0.05 would be strong evidence for being a member of class *def* in 2009 but not enough in 2010. Results are only collection of static models. In the next section logit model will be learned on panel data which will allow inclusion of past data and the subject-specific dynamic characteristics. Corresponding probability estimates are comparable across years.

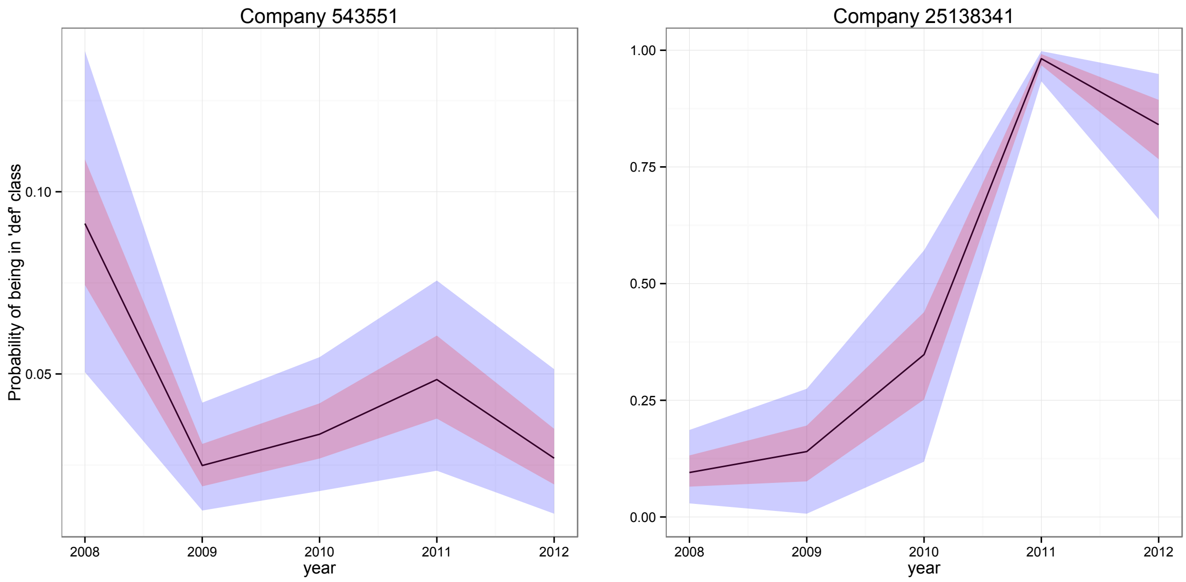

Fig. 27: *Fanplot summarises predictive bootstrapped predictive intervals. Quantiles 10, 30, 50, 70 and 90 delimit colour zones. Source: Own processing*

Tab. 12: *Values of GLMM regression coefficients. Bound Optimization by Quadratic Approximation was used to estimate parameters. All variables were identified as statistically significant. Source: Own processing*

Intercept	acid.test	debt.ratio	asset.turn	returns
-10.203	-0.604	0.709	-0.482	-0.854

6.3.2 Hierarchical Structure Analysis

Previous demonstration was to show logistic regression framework on cross-sectional data. Here, extended framework is proposed. This extension allow better handling of auto-correlated values. Technical details were given in chapter 3.3.3. Random slope (in our example reflects time) and intercept model is fitted as it is assumed that companies do not start with the same risk profile and risk trajectories might differ as well.

Because of computational demands covariates were scaled prior fitting the model. Observed regression coefficient and their interpretation is therefore not the same as in logistic model on cross-sectional data. Table 12 presents fixed-effect coefficients. These coefficients apply for all all subjects.

Figure 28 inspects relations between variables and estimated probability of default. Line in the plot is computed threshold from previous year. All companies which are above the line are assigned as defaulted. Note that some companies with positive returns are expected to get defaulted in the following year.

Previous models did not estimated individual effects directly. Individual behaviour is captured in random effects component showed in Table 13.

Coefficients in Table 13 are subject-specific. They indicate that performance of company 554812 is getting better as the probability of defaults decreases in time.

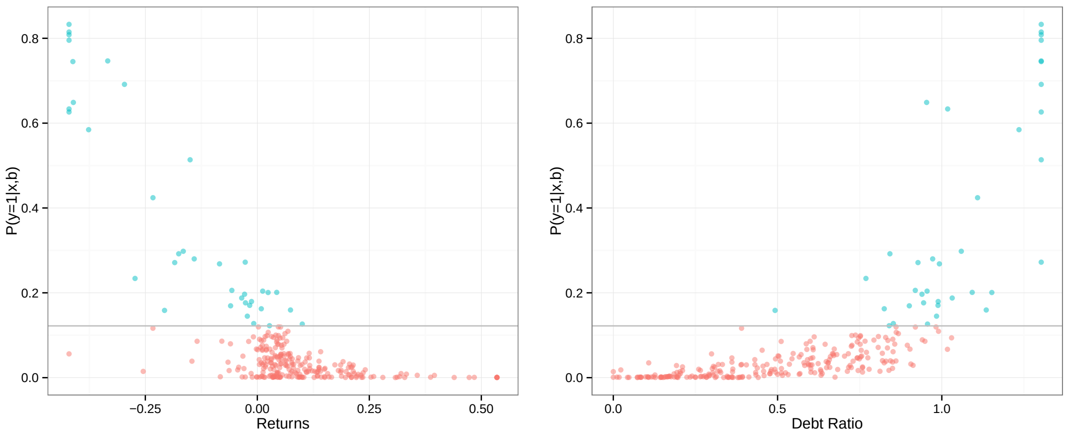

Fig. 28: *Both scatter plots demonstrates that the orientation of relations between variables is as expected by the theory. Source: Own processing*

Tab. 13: *Random effect from GLMM logit model. The same two companies were selected as in cross-sectional logit model. Source: Own processing*

Company ID	Intercept	Slope (year)
543551	-0.0251	0.0005
554812	-0.00982	-0.00324

Fig. 29: *Probability values in the left panel are virtually zero. For defaulted company estimates are closer to 1 than in cross sectional logit model. Source: Own processing*

6.4 Survival Analysis

Analysis of censored data is demonstrated in this chapter. Investor is trying to assess whether some knowledge about the company is relevant for her risk analysis. It might be indication whether company had to liquidate large proportion of assets unexpectedly or some other, from financial statements directly unobservable, management-oriented reason. This information divides companies into two groups. In the first group companies are supposed to be less risky than in the second group. She wants to avoid investing into financially distressed company. If the analyst projects such a criteria into historical data she finds this table:

```
                as.factor(group)=0
 time n.risk n.event entered censored survival std.err low 95% CI upp 95% CI
    1    239       1      44       12    0.996 0.00418      0.988      1.000
    2    270       1      24       22    0.992 0.00555      0.981      1.000
    3    271       4       0       53    0.977 0.00910      0.960      0.995
    4    214      12       0      202    0.923 0.01761      0.889      0.958

                as.factor(group)=1
 time n.risk n.event entered censored survival std.err low 95% CI upp 95% CI
    1     64       1      12        1    0.984  0.0155      0.954      1.000
    2     74       2      10        3    0.958  0.0239      0.912      1.000
    3     79       7       0       15    0.873  0.0376      0.802      0.950
    4     57      30       0       27    0.413  0.0604      0.311      0.551
```

In the initial year of her analysis (2009) 239 companies were according her rule classified as Group 0. From these companies only one has defaulted that year. To the next year twelve companies left her historical records (from other reasons that financial distress), 44 new companies entered her analysis (i.e., start-ups). Column survival contains cumulative proportions of financial distress companies in the group 0 up to time t. Last three columns show standard error of survival estimate with 95% confidence interval.

From the comparison of both tables can be understood that survival rate is lower in group 1. Following table presents results from Cox-Proportional test.

```
                     coef exp(coef) se(coef)     z Pr(>|z|)
as.factor(group)1  2.3294   10.2719   0.2854 8.162 3.33e-16 ***

                  exp(coef) exp(-coef) lower .95 upper .95
as.factor(group)1     10.27    0.09735     5.871     17.97

Concordance= 0.766  (se = 0.031 )
Rsquare= 0.169   (max possible= 0.809 )
Likelihood ratio test= 72.59  on 1 df,   p=0
Wald test            = 66.62  on 1 df,   p=3.331e-16
Score (logrank) test = 101.5  on 1 df,   p=0
```

Formal definition of this model can be found in (3.3.25). Being in group 1 is associated with 10.27 time higher risk become financially distressed. It should be noted that this model estimates only one value of risk which is time-invariant. Risk for group 1 is therefore 10 times

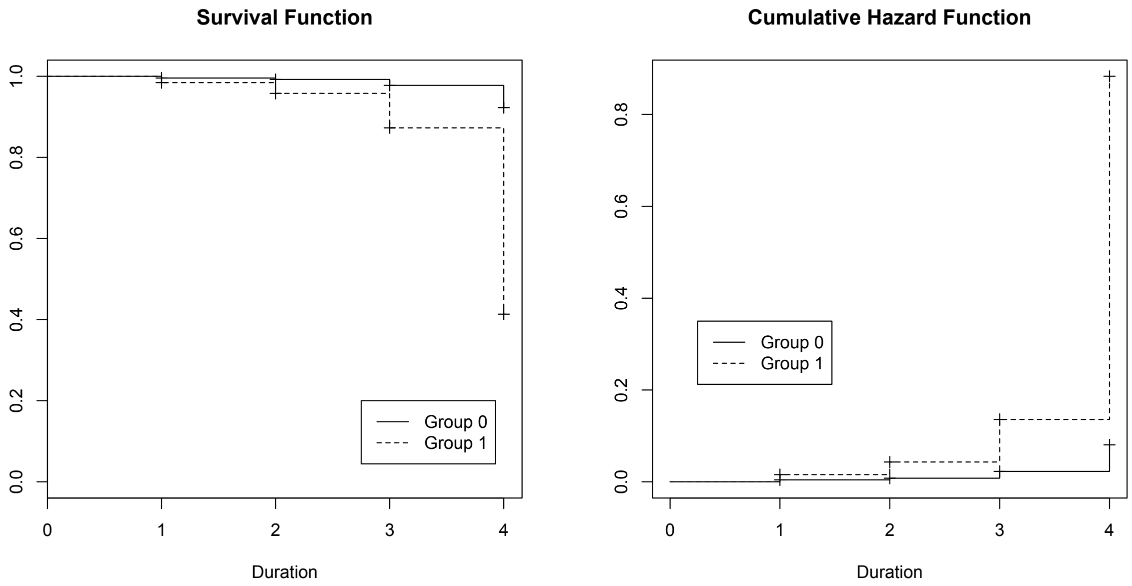

Fig. 30: *Survival function and Cumulative Hazard function suggest that group 1 members more likely become financially distressed. Source: Own processing*

higher in the first year as in the following years. This risk effect is statistically significant as the confidence interval of transformed instantaneous hazard rate is non-negative (5.871 – 17.97) and corresponding p-value is ≤ 0.05. Whether 10 times higher risk is substantially significant is up to the analysts, but such a high value indicates grouping companies according this criteria has some explanatory value.

6.5 Joined Modelling

Chapter 3.3.4 introduced notion of Joined Models. We will be particularly interested in (3.3.26). In this analysis, we will evaluate importance of variable z. This variable is the probability estimate from the GLMM model showed presented in chapter 6.3.2. Variable z is endogenous time-varying variable. To make interpretation more intuitive, z values were scaled. Original mean value was 0.0664 and standard deviation 0.2174.

```
Data Descriptives:
Longitudinal Process                Event Process
Number of Observations: 1538       Number of Events: 33 (8.4%)
Number of Groups: 393
```

On overall, 1538 records were analysed. There were 393 groups - different companies. An average default event rate was 8.4%. Analysis of underlying longitudinal process shows that, on general, likelihood of default situation is increasing (year=0.0398). This results is not statistically significant on standard $\alpha = 0.05$ level.

```
Coefficients:
Longitudinal Process
              Value Std.Err z-value p-value
(Intercept) -0.1110  0.0727 -1.5265  0.1269
year         0.0398  0.0234  1.6994  0.0892
```

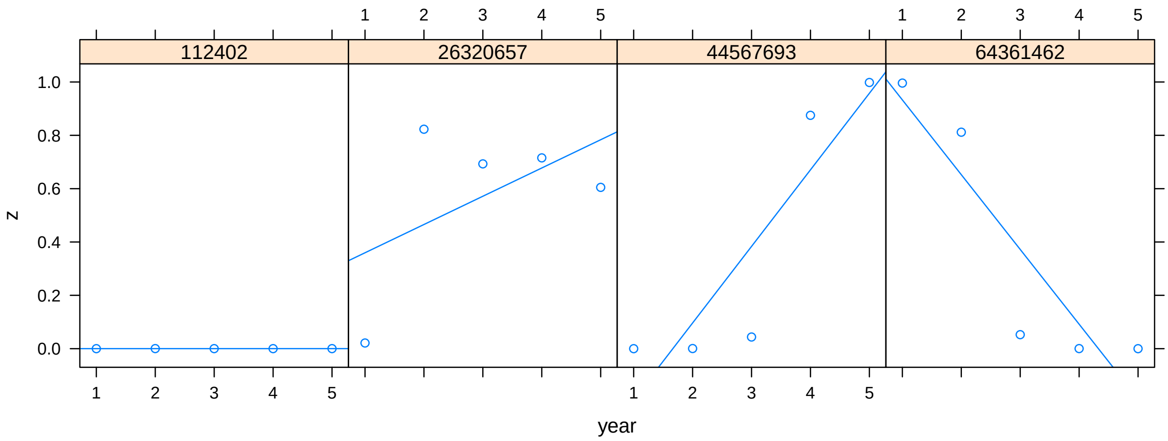

Fig. 31: *Values z on the vertical axe represent estimated values from logistic GLMM step. These values are used in longitudinal part of joined model. In settings of this demonstration we expect linear trend of z values. This is not a case of the third company. It is possible to employ higher polynomial fit if required. However, given only 5 observations, over-fitting with higher polynomial model is very likely. Own processing*

Results from survival model shows that a company with an average z score faces risk of $\exp(-14.0560) = 7.862429\text{E}-07 \sim 0$. However, unit change in standardised z value increases risk of default $\exp(-1.3053) = 3.688796$ times.

```
Event Process
              Value Std.Err z-value p-value
(Intercept) -14.0560  2.1423 -6.5612 <0.0001
Assoct        1.3053  0.1543  8.4617 <0.0001
log(shape)    1.8063  0.2189  8.2522 <0.0001
```

Comparison of two companies with scores $z_1 = 0.25$ and $z_2 = 0.15$ reveals that

$$\frac{\exp(\frac{0.25-0.0664}{0.2174})}{\exp(\frac{0.15-0.0664}{0.2174})} = \frac{2.326875}{1.468945} = 1.5840 \tag{6.5.1}$$

risk of being defaulted for first company is 1.5 times higher than for the first group. Because of non-linearity unit change does not result in the same change of risk (see similar in chapter 6.3.1).

6.6 Support Vector Machine

This chapter demonstrates application of SVM method. This highly flexible classification approach requires careful identification to overcame problem of over-fitting. SVM with RBF and quadratic will be used. Cost parameter C and kernel parameter has to be added by analysts or can be estimated by some optimisation technique, such as grid search.

Cost function was used as an optimisation criteria for determination SVM parameters. Two fold cross-validation was repeated 40 times was performed to identify the best performing model. Several kernels were estimated too, but their performance was not as good as *Radial Basis Function Kernel*.

Tab. 14: *Performance statistics of overall classification. Source: Own processing*

	Logit	LDA	QDA	SVM
AUC	0.909	0.901	0.904	0.913
Log-loss	0.211	0.217	0.262	0.209

```
  C   sigma  Accuracy    total.cost  Accuracy SD  total.cost SD
  1  0.25   0.9109286   723.1875    0.007199056  28.47655
  1  0.50   0.9097143   719.2500    0.007182170  29.16670
..
  2  0.25   0.9129286   725.1250    0.009675436  39.89682
  2  0.50   0.9136429   727.8125    0.009621888  39.91164
  2  0.75   0.9164286   738.1875    0.009975428  40.58312
  2  1.00   0.9160000   736.9375    0.011334149  46.28950
..
  3  0.75   0.9165714   735.5625    0.012683375  52.67523
  3  1.00   0.9186429   743.8750    0.012993848  53.95717
  3  1.25   0.9168571   737.7500    0.012822739  53.88043
  3  1.50   0.9149286   731.4375    0.011993066  50.81110
..
  4  0.50   0.9144286   725.6250    0.013048803  54.13517
  4  0.75   0.9179286   738.6875    0.013462540  55.99923
  4  1.00   0.9184286   740.8750    0.013788078  57.41684
  4  1.25   0.9157857   731.7500    0.013230270  55.11266
  4  1.50   0.9139286   726.1875    0.013217767  55.17087
```

Best performing model can be selected according to several metrics. Accuracy and κ were computed to demonstrate that final model (Cost=1 and $\sigma = 1$) would not be selected if κ would have served as the decision criteria. In final comparison, model with Cost = 7 and $\sigma = 1.1$ was used as it is expected that results will be more robust and there will be less miss-classifications on unseen data set.

6.7 Classification Performance Analysis

In this chapter classification performance techniques will be demonstrated. It is recommended strategy to fit several models and either merge the results or compare them and choose the best-performing. Tools described in chapter 3.1.2 will be used here. To demonstrate these techniques cross-sectional logistic model, LDA, QDA and SVM with linear kernel were used. Models were trained on 2011 dataset. Results summarise performance on 2012 data. From the Figure 32 we cannot directly conclude which of the classifiers is the best performing. Values of Area Under Curve and Log-loss suggest that the best performing is the SVM algorithm. AUC is the highest from all classifiers and therefore has the best overall classification performance (with respect to True Positive and False Positive rates). Log-loss indicates that values which led to binary classification were more informative (estimated probabilities were closer to 0 if the subject's true label was 0) then in case of other classifiers.

Table 15 presents misclassification table after thresholds T_{logit} and T_{SVM} from year 2011 were applied on 2012 data.

Performance statistics from the misclassification table are showed in Table 16.

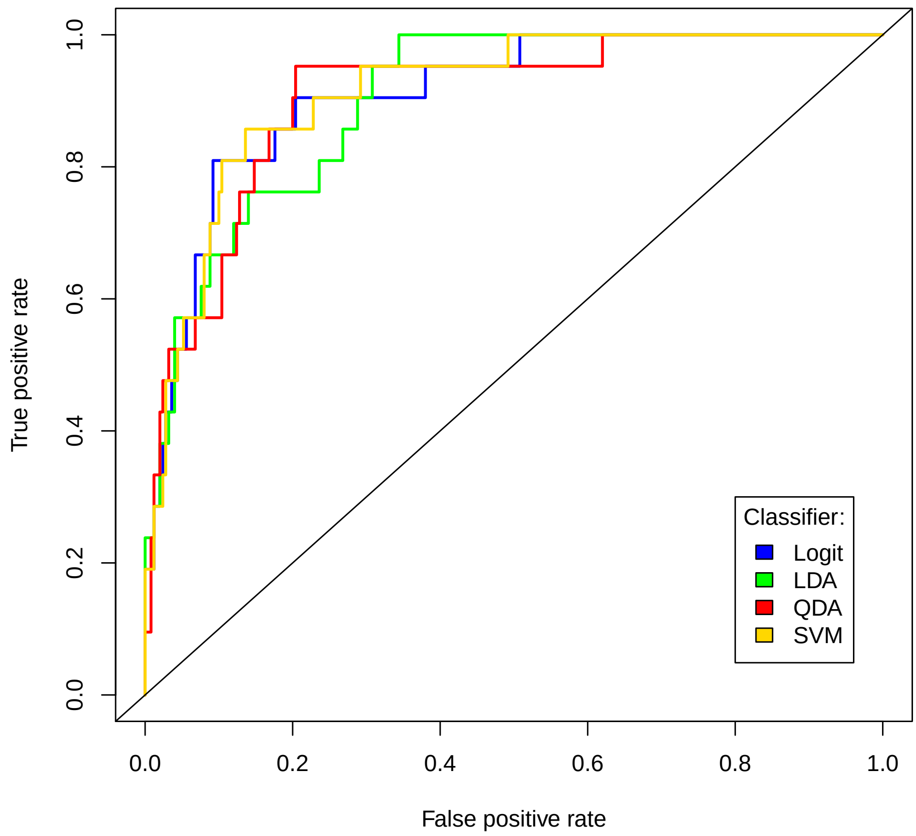

Fig. 32: *ROC curves for Logit, LDA, QDA and SVM classifiers do not suggest superior classifier. Source: Own processing*

Tab. 15: *Two classification tables are showed below. Left corresponds to Logistic regression. Values in brackets indicate number of classified observation on 2011 dataset. Source: Own processing*

	Logit		SVM	
	non-defaulted $y=0$	$y=1$	$y=0$	$y=1$
$\hat{y}=0$	224(297)	4(8)	245(315)	12(12)
$\hat{y}=1$	26(19)	17(26)	5(1)	9(22)

Tab. 16: *Table shows classification performance measures. Comparison Logit and SVM in year 2011 and 2012. Statistics show high values of Sensitivity, which is affected by cost matrix which gave high penalty on this type of misclassification. In 2012 values of Precision are 39.5% and 36.2% which can be interpreted that almost 60% of companies assigned as distressed are in reality healthy. Source: Own processing*

	Logit		SVM	
	2011	2012	2011	2012
Accuracy	0.923	0.889	0.914	0.875
Sensitivity	0.765	0.810	0.824	0.810
Specificity	0.974	0.982	0.980	0.982
Precision	0.578	0.395	0.538	0.362

6.8 Two Step Analysis

Credit specialists receives accounting data of 271 companies which applied for credit line for the period of 2012-2013. She estimates that there will be 257 healthy companies and 14 defaults (see right panel of Table 15). Credit line will not be granted to these 14 companies. At the time of the decision, values in 17 are the only available data about the performance of analytical model (SVM classifiers).

Tab. 17: *Confusion matrix for SVM classifier. Values in brackets show row-proportions. Only 4% of companies which were assigned as healthy became defaulted. Source: Own*

train.results	$y = 0$	$y = 1$	
$\hat{y} = 0$	315 (96%)	12 (4%)	327 (100%)
$\hat{y} = 1$	1 (4%)	22 (94%)	23
	316	34	350

Table 17 simplifies the analysis because it presents results of cross-validated model on the whole *train* sample (results are optimistic). Estimated values differ with those in Table 24. Those values present average misclassification on validation sets. Estimates showed in latter table capture underlying uncertainty better, but do not have as clear interpretation as required for the purpose of this chapter.

Taking into account error rate from the past she can estimates that $257 \times 96\% \doteq 246$ companies will be classified correctly. Remaining 11 companies will turn defaulted. The best estimate of the loss and profit per contract is an average loss/profit from the past. This is because she does not know which company will fail in the next year. If only direct profits and costs were considered (values are in table 18), expected profit would be computed as $246 \times 5 + 11 \times (-20) = \text{USD}1010$. This profit, however, does not include information about 5 companies which were incorrectly classified as defaulted. If opportunistic costs would be considered profit would decrease to $1010 + 5 \times (-5) = \text{USD}985$.

Expected economic profit for the next forecasting period is 1010 USD. In the end of the period results in 15) are available. It can be concluded, that from out of the 257 approved credit lines 245 were performing well and 12 defaulted. Total profit is $245 \times 5 + 12 \times (-20) = 985$. Expected profit was higher than actual.

This methodology was used for all models used in the thesis. Results are summarised in Table 24. Note, that SVM model with cross-validated error rate expected higher proportion of incorrectly classified companies to 7.38%, which resulted in lower expected profit.

6.9 Visualisation of Cost Function

If cost function is set as an optimisation function, maximal value has to be found. Positive value indicated profits. If the function is visualised it can provide information about sensitivity of cost

Tab. 18: *Cost matrix with only values which have impact on accounting figures. No opportunistic cost were set. Source: Own processing*

	$y = 0$	$y = 1$
$\hat{y} = 0$	5	-20

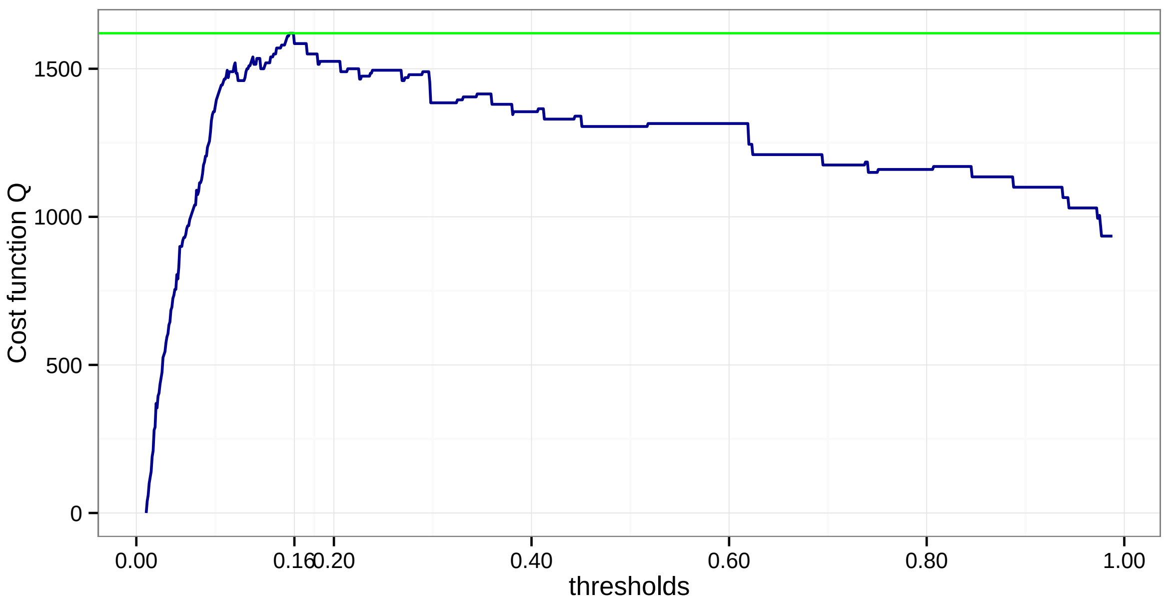

Fig. 33: *Cost function suggests that optimal value of threshold T if the criteria is* `argmax` *is 0.122. Source: Own processing*

Tab. 19: *Classification table shows results from 2011 and 2012. If this analysis would have been done in early 2013 (after 2012 financial statements are available), analyst would see the left panel and part of the right panel. She would predict that there will be 224+4 healthy companies and 26+17 defaulted companies. At last in the early of 2014 she would recognise (after statements are released) that 4 of predicted healthy companies were defaulted and that from 43 companies predicted to gone defaulted only 17 actually defaulted. Source: Own processing*

	2011		2012	
	non-defaulted $y = 0$	$y = 1$	$y = 0$	$y = 1$
$\hat{y} = 0$	297	8	224	4
$\hat{y} = 1$	19	26	26	17

on threshold value. Analyst has to decide whether decrease in profit should be traded for more robust solution (lower threshold value).

Reading from Figure 33 threshold value $T = 0.16$ which corresponds to Q=1620 on the train set. This threshold is used in the next year model.

6.10 Classification Tree

In this section regression trees techniques introduced in chapter 3.4.1 will be demonstrated. The main objective of this demonstration is pointing to:

- classification performance of various learning techniques,
- parameter sensitivity and consequences of cost-matrix application,
- estimation of uncertainty and fit quality,
- benefits resulting from graphical representation.

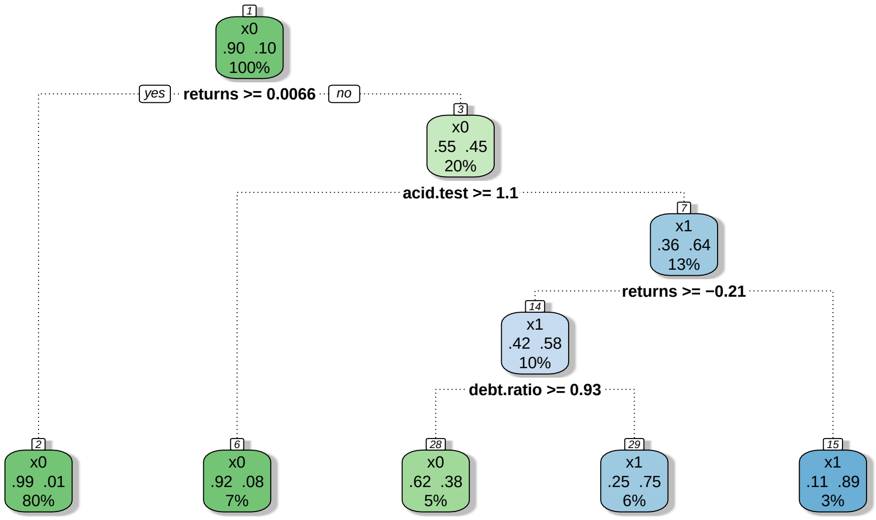

Fig. 34: *Regression tree set* `returns` *as the variable with the highest discriminatory power. Left branch groups companies with returns higher than 0.0066. These companies will be labelled as 0 (upper box number). Colour of the box indicates majority and strong of the splitting rule - the lighter the colour is the weaker the rule is. Note, that variable returns appears twice in the model. Source: Own processing*

Recursive partitioning model was learned on 2011 data. In the Figure 34 classification on 2012 data is showed. Graphical outcome of C4.5 (J48) algorithm is showed below. This models was created with equal misclassification costs. Results were obtained on 10 fold cross validation samples. First node of the model represents the most informative node (J48 computes impurity function derived from information theory) which partitioned space according to value of variable returns. Application of this rule results in 281 correctly classified good companies at cost of 3 incorrectly classified defaulted companies. Further inspection reveals that debt ratio was used in two consequent nodes. This division seems to indicate over-fitting problem as there is no clear reason for such branching. Output was truncated as it resulted in 11 level decision-tree. One of the advantages of decision tree approach is direct interpretation of classification steps.

```
J48 pruned tree
------------------

returns <= 0.006494
|   debt.ratio <= 0.489598: 0 (18.0)
|   debt.ratio > 0.489598
|   |   debt.ratio <= 0.736165: 1 (11.0)
|   |   debt.ratio > 0.736165
|   |   |   acid.test <= 0.310385: 1 (10.0/1.0)
|   |   |   acid.test > 0.310385
|   |   |   |   returns <= -0.286596: 1 (4.0)
returns > 0.006494: 0 (281.0/3.0)
```

Performance on training set achieved high classification accuracy (93.14 % all companies were assigned correctly). On the test sample this statistics dropped slightly (91.51%) but the structure of misclassification is changed considerably. Results are showed in 20. When the cost matrix was considered (misclassification cost of defaulted company was was set for this example 20 times higher than misclassification of healthy company) final model looks as:

Tab. 20: *Classification tables contain results from two C4.5 algorithm on train (left) and test sample. First values correspond to numbers of companies classified using model with indifferent misclassification matrix. Values in brackets correspond to model with adjusted cost matrix. Source: Own processing*

	Train		Test	
	non-defaulted $y = 0$	$y = 1$	$y = 0$	$y = 1$
$\hat{y} = 0$	305 (283)	11 (33)	231 (210)	19 (40)
$\hat{y} = 1$	13 (0)	21 (34)	4 (3)	17 (18)

```
returns <= 0.012089
|   debt.ratio <= 0.533187
|   |   debt.ratio <= 0.489598: 0 (8.08)
|   |   debt.ratio > 0.489598: 1 (7.03)
|   debt.ratio > 0.533187: 1 (226.31/8.43)
returns > 0.012089
|   debt.ratio <= 0.788314: 0 (79.07)
|   debt.ratio > 0.788314
|   |   returns <= 0.050045
|   |   |   asset.turn <= 2.433137: 1 (17.22/3.16)
|   |   |   asset.turn > 2.433137: 0 (2.11)
|   |   returns > 0.050045: 0 (10.19)
```

Variable returns was again considered as the most important, number of nodes and corresponding values changed. Classification performance can be found at Table 20. Application of cost matrix introduced in Table 5 It might not be always best way to point to importance of one class (defaulted). If bagged ensemble (chapter 3.1.5) is created, weight to unsuccessfully classified objects are increasing and more emphasis is put on developing rules which can classify the object correctly. After ten-step bagging procedure on training set over-fitted model was created. All instances perfectly classified. Final model looks as:

```
returns <= 0.006494
|   debt.ratio <= 0.489598: 0 (15.11)
|   debt.ratio > 0.489598
|   |   asset.turn <= 4.981156
 >pruned
|   |   asset.turn > 4.981156
|   |   |   acid.test <= 0.077302: 1 (2.99)
|   |   |   acid.test > 0.077302: 0 (26.02/2.12)
returns > 0.006494
|   asset.turn <= 2.29919
|   |   asset.turn <= 2.296981
|   |   |   returns <= 0.026909
 >pruned
|   |   |   returns > 0.026909: 0 (22.84)
|   |   asset.turn > 2.296981: 1 (8.87)
|   asset.turn > 2.29919: 0 (134.68)
```

and the performance on test sample was correctly classified 233 good companies and 16 defaulted, whereas 17 and 5 healthy and defaulted companies were classified incorrectly.

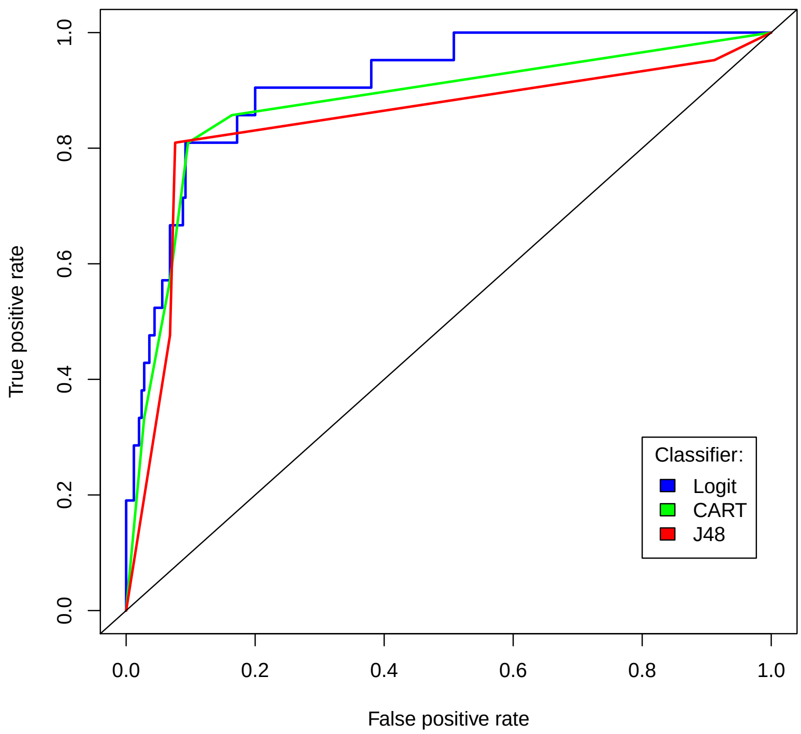

Fig. 35: *ROC of Logit, CART and J48 reveals that on overall the best performing algorithm is Logit model. Source: Own processing*

Tab. 21: *Confusion Matrix for Logit, CART, J48. Technique J48 reports highest total gain from classification if the cost matrix from Table 5 is used. Predicted values are $\hat{y}$. Source: Own*

	Logit		CART		J48	
	y = 0	1	0	1	0	1
$\hat{y} = 0$	224	4	226	4	231	4
1	26	17	24	17	19	17
Total Gain	1165		1185		1235	

Based on the Figure 35 it seems that Logit model is the best performing. This is supported by highest AUC value (0.909 compared to CART=0.875 and J48=0.843).

Previous analysis was made to show overall performance with all possible thresholds. We proceed to analysis which involves cost matrix defined at Table 5. Optimal (cost `argmax`) threshold was found on training sample (year 2011). This threshold was used for classification on 2012 data. Results are showed in Table 21. As the best performing classifier is assigned J48 because it maximizes cost function.

6.11 Neural Networks

Neural networks were described in details in chapter 3.4.2. As the major problem NN application represent over-fitting. If the model is too simple (contains only a few neurons) it cannot describe the complex nature of the problem – which is the main appeal for employing this advanced method.

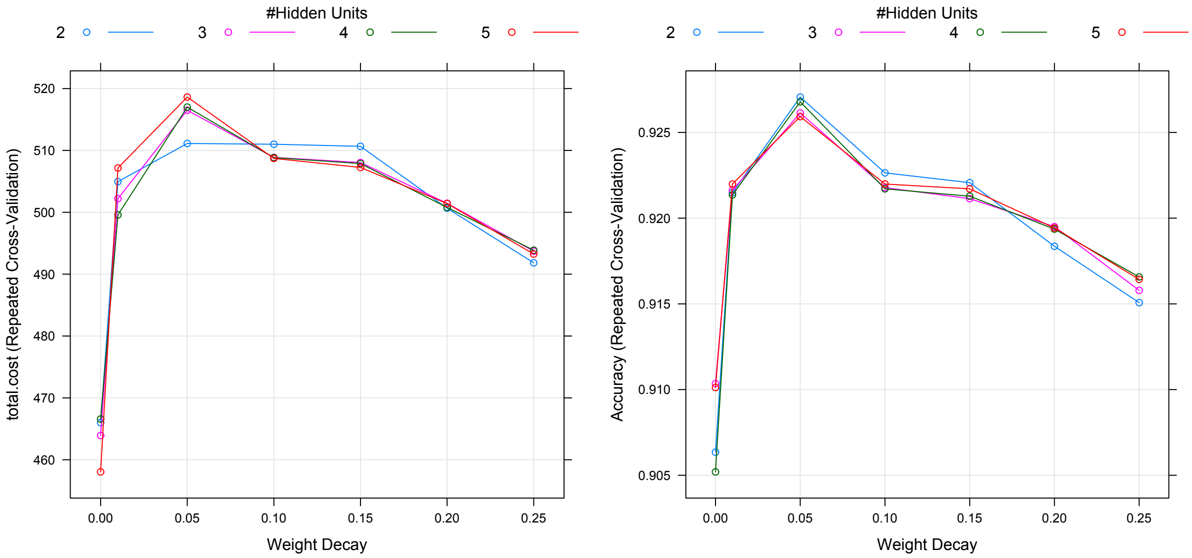

Fig. 36: *Lines in both plots represent the average value of performance indicator which was achieved by model with complexity given by number of neurons and shrinkage λ value. Source: Own processing*

In the next example identification of optimal parameter and design settings will be showed. Single hidden layer NN is learnt on the different numbers of neurons and values of regularisation parameter λ. To highlight the importance of optimisation criteria on final model, cost function which is used throughout the demonstrations and accuracy indicator is used. Results are visualised at figure 36.

To validate results bootstrapped performance indicators were computed. Bootstrap (described in chapter 3.1.1) technique creates sub-samples on training set. Value of total profit from classification is not as high as in other methods because is computed on smaller dataset. This does not represent a problem as this step serves only for parameter identification. Total gain will be computed on test sample, which is not available in the phase of model building.

Analyst now have to decide which model is the best performing and have the best prospects for successful classification on the test sample. Based on the maximal values in both figures model with $\lambda = 0.05$ should be selected. Model with 5 nodes would be selected if total cost criteria would be considered. This is the model with highest complexity. Model with only 3 internal nodes performs almost identically. To reduce over-fitting risk simpler model should be selected. For demonstration purposes both models will be evaluated on test sample.

Application of SMOTE technique (chapter 3.1.1) extends this chapter. Neural network model will be learned on data set which consists of original data and artificially created. Parameter identification is shown at Figure 37. Performance of three models will be tested on test sample. First model (NN1) has 5 nodes with $\lambda = 0.05$, second model (NN2) has 3 neurons and the same λ. Both models were learnt on the original data. Last model (NN3) comes from the SMOTEd training set. Model NN3 has 3 nodes parameter $\lambda = 0.01$. Performance on training set is showed at Figure 39.

Table 22 reveals that the best performing model on the training set (NN1) does not provide the best performance on the test sample. SMOTEd model has the best performance as it correctly classified 71.4% of all defaulted companies. On the other hand, it has highest rate of false positives (highest opportunity costs).

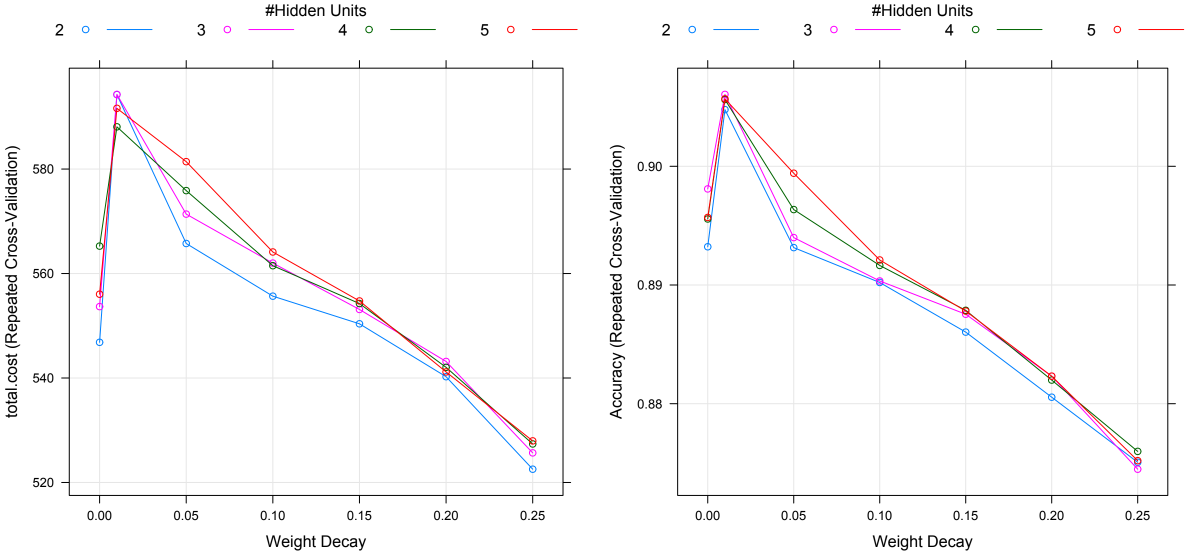

Fig. 37: *From the perspective of total costs/gains model with three neurons is preferred too. The best performance is achieved with λ set on 0.01. Source: Own processing*

Tab. 22: *Confusion Matrix three Neural Network models. Predicted values are $\hat{y}$. Source: Own*

	NN1		NN2		NN3	
	y = 0	1	0	1	0	1
$\hat{y} = 0$	237	11	238	11	230	6
1	13	10	12	10	20	15
Total Gain	1050		1060		1155	

Tab. 23: *Confusion Matrix for three ensembles. Source: Own*

	Ensemble 1		Ensemble 2		Ensemble 3	
	y = 0	1	0	1	0	1
$\hat{y} = 0$	232	5	238	9	224	2
1	18	16	12	12	26	19
Total Gain	1210		1130		1210	

6.12 Ensemble Techniques – Stacking

In this chapter three heterogeneous ensembles will be created and compared. The first ensemble will consist of a statistical model (logit) and a representative of tree-based algorithm (J48). Second ensemble will consist of boosted J48 tree algorithm and machine learning tool (SVM). Last ensemble consists of the same classifiers as the second one but number of defaulted companies on training set would be enlarged by SMOTE technique. Whole analysis was made in Weka software with *MetaCost* algorithm which allows cost-sensitive analysis. Both level 0 models were learned on 3 folded cross-validation sample. As a level-1 model was selected CART model. SMOTE technique raised proportion of defaulted companies to 24.4% from 9.71%.

First ensemble (logit + J48) correctly classified 91.58% instances. Corresponding κ equals 0.5375 which indicates good fit compared to expected accuracy. Second ensemble (SVM + LDA) reached higher classification accuracy 91.88% and $\kappa = 0.565$.

Performance of ensembles does not increase overall performance much compared to a single-model analysis. There is also virtually no difference in classification performance between ensembles. Observed differences should be attributed to sampling variation rather than by to some systematic cause.

6.13 Variable Importance

Importance of covariates can be identified with respect to some learning algorithm, or by its general discriminatory power. Second approach was used in LDA model when t statistics were computed to identify whether conditional mean values differ. In the Figure 38 relative importance is presented.

In highly dimensional problems feature selection techniques are commonly used. These techniques use various optimisation criteria, such as decrease of R^2 or change of Akaike Information Criteria.

6.14 Final Comparison of Models' Performances

In the last chapter overall performance of selected classifiers is analysed. All models are cost-sensitive. Cost function reflects costs defined in Table 5. For the best parameter identification model were tuned on 3 fold cross-validation sampling techniques which was repeated 40 times. Three folds were chosen as a reasonable split rule because of the presence of minority class. Figure 39 presents point estimate (mean value) and confidence interval. Each mean value was obtained on approximately one third of the sample. In the Figure 33 optimal threshold was found by maximising the cost function on the whole sample. No resampling technique was used. Total cost was therefore about three times higher than in the Figure 39. Variation of total

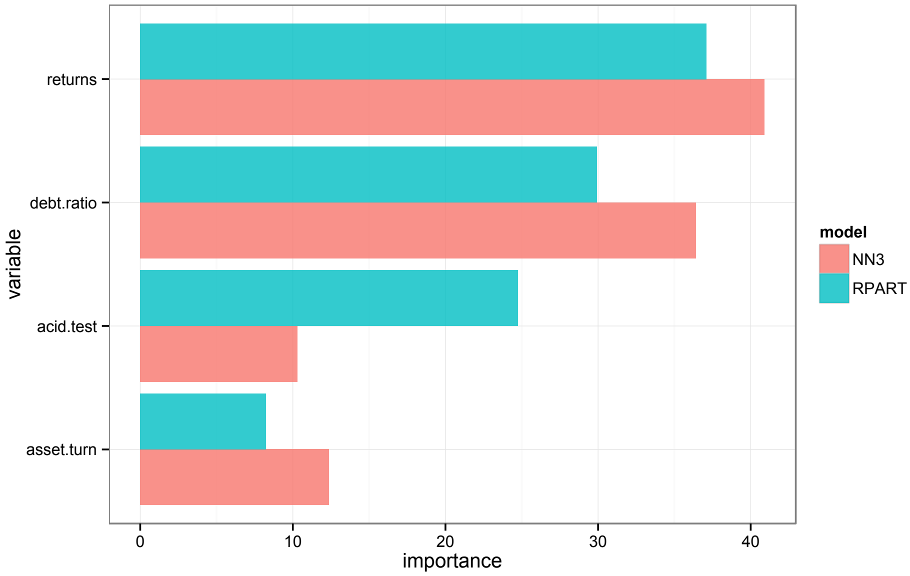

Fig. 38: *Standardised values of importance reveal that importance of variables differ in classification tree model and neural network model. Source: Own processing*

cost function values achieved on re-sampled samples is visualised in intervals. Intervals were computed as a symmetrical 95% confidence intervals around the mean value using normal distribution quantiles and estimated standard errors. Based on the Figure 39 NN3 model should be selected. This model, however, has very high value of variance. This indicates a presence of a high risk related to the performance on unseen data. Highest uncertainty was achieved in Quadratic Discriminant Model. Although it was not formally tested, it can be concluded that none of the classifiers outperform other classifiers. All confidence intervals are overlapping.

6.14.1 Financial Plan

The ultimate reason for building an analytical model is to provide a decision-maker an information. Here, analytical models will be used for financial planning. Company provides credit lines to companies. Risk profiles of 271 companies are assessed in the beginning of the planning period. Based on the risk profile of each applicant credit line is either approved or declined. Expected profit can be computed after some adjustments described later. Although in the following example credit-risk application is demonstrated, this example can be transformed into investment application. Company's market value decrease can be considered instead of the credit default. Maximum value of expected loss resulting from this decrease can be fixed by risk-management policies. On the other hand, an average profit from good investment can be estimated from historical records.

Classification performance indicator on re-sampled training sets revealed error rates for all available classifiers. Analysts therefore have some information about model's error rate. Alpha error rate indicates misclassification of companies which were assigned by model as "healthy" but turned to be "defaulted". Beta error rate does not enter financial plan as it represents opportunity costs. It is expected that profit from average successful contract worth 5.000 EUR. Erroneous assignment on defaulted company costs 20.000 EUR. Expected profit is calculated as:

$$\begin{aligned} \text{Expected profit} = 5000 \times (\text{Expected positives} - \text{Expected alpha (counts)}) - \\ - 20000 \times \text{Expected alpha (counts)} \end{aligned} \tag{6.14.1}$$

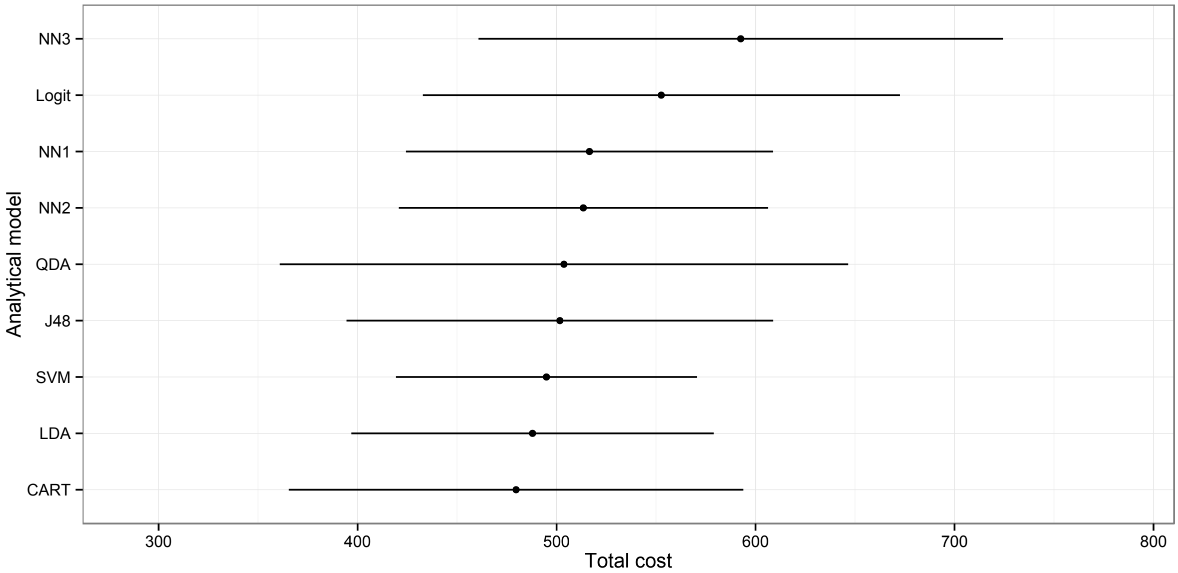

Fig. 39: *Dot in the middle of each interval represents mean value of the total cost function on cross-validated samples. Wider intervals point to higher uncertainty of predictive correctness. Source: Own processing*

Where *Expected alpha (counts)* corresponds to a number of defaulted companies which were labelled as good companies.

After all models are created, decision which one to select has to be made. Results are summarised in Table 24. Note, that in the time of the decision column *Real profit* is not available to analysts.

Number of predicted good companies vary across classifiers. Ensemble III expects that there will be only 226 good companies. This would represent 16.67% rate of defaulted companies. Historically, this rate never exceeded 9.71%. This model is therefore very conservative as it tries to eliminate alpha misclassification. On the other hand, SVM model assigned 257 companies (this corresponds to 5.16% default rate). This is a suitable model for risk-takers who grant more credit line in the faith of better loss recovery. Linear Discriminant Model (Altman-type model) is among riskier models.

During the identification phase misclassification rates were computed. It can be concluded that from the long-term perspective application of NN1 model results in 6% alpha error rate (6% of assigned good companies turn to be defaulted). Classifiers with lowest expected alpha rate have the highest opportunity costs.

Expected profit from classification should not be the only one decision criterion. Second Neural Network (NN2) model would be selected according to maximum value of profit. Given the large uncertainty related to classification performance (see Figure 39) NN2 cannot be recommended, though. Differences in predicted profits vary only negligibly in the light of corresponding uncertainty. Although there are no estimates for Ensemble III in the aforementioned Figure (due to computational problems), Ensemble III seems to be a better option due to low expected alpha rate.

If the NN2 model would be selected, real profit would exceed expected profit by 100.000 EUR. Real profit would be, however, lower than Ensemble III profit. Ensemble III resulted in the highest profit 1.080.000 EUR (the same profit would be achieved by Ensemble II). Notably, simple J48 algorithm achieved the third highest profit. This should be attributed to a sample

Tab. 24: *The second column shows how many good companies were assigned by classifiers. Third column contains proportion of defaulted companies which were assigned as healthy but defaulted on the train sample. The fourth column summarises expected number of companies which will default but are assigned as healthy. Expected profit is computed according to equation 6.14.1. Column Test profit reports classification results on test sample. Last column computes opportunity costs (5000 EUR / misclassification). Values with * are estimated by author. Source: Own processing*

Model	Expected positives	Expected alpha	Expected alpha (counts)	Expected profit	Real profit	Opportunity costs
NN1	248	6.00%	15	865	965	-65
NN2	249	6.00%	15	870	970	-60
NN3	236	6.10%	15	805	1030	-100
Logit	246	9.74%	24	630	1005	-65
GLMM	-	-	-	-	1010	-120
SVM	257	7.38%	19	810	985	-25
J48	235	7.30%*	18	725	1075	-95
CART	230	5.18%	12	850	1050	-120
LDA	253	6.93%	18	815	965	-45
QDA	235	8.74%	21	650	975	-115
Ensemble I	237	8.00%*	19	710	1060	-90
Ensemble II	247	6.70%	15	860	1080	-90
Ensemble III	226	4.82%	11	855	1080	-130

variation effect rather than the systematic behaviour. Classification performance of J48 was not among the best performing classifiers on verification samples.

Hierarchical logistic model (GLMM) was not identified on re-sampled training samples because of the computational capacities. This model, however, presents valuable information about risk-profile development for all companies. This was demonstrated in separate chapter 6.3.2. Ensemble models were estimated by different methodology which does not provide comparable outcomes as in the Figure 39).

7 MAIN RESULTS OVERVIEW

This section summarises key findings. The motivation for writing this thesis was to extend theoretical ground for corporate performance modelling research by proposing modelling strategies that involve state-of-the-art analytical techniques. The main goal was to develop a conceptual model which provides the general approach applicable in analytical model building process. To achieve this goal, two objectives were set. **First**, to answer three research questions. **Second**, to demonstrate principles and models on real data.

Research questions stated in chapter 4 were answered.

First research question: *Which topics belong to the performance modelling framework?* was answered in chapter 2. Literature was classified in four distinct fields which require different analytical approach. The most important was identified as predictive modelling which involves classification of discrete outcomes and regression. These generic tasks are most frequently used for risk management and portfolio modelling. In the risk management class, credit risk plays an important role. Other topics, such as companies health modelling and prediction are frequently discussed in literature as well. Portfolio modelling involves analysis of shared risks and identification of inter-correlated behaviour of subjects. This topic was not analysed in depth because this thesis focuses on multiple single-company models. MSCM seeks to identify general patterns of behaviour which can be used for various tasks without inter-correlated structure. Throughout the review financial performance of the company was usually considered as synonym to corporate performance and company's health. Company is considered as healthy if its prospects for survivor, future performance and stability are high.

Which analytical techniques related to the studied phenomena were used and which were the most successful? is the second research question. Literature review was conducted and seminal works and models were identified. These models were demonstrated in the empirical results section. From literature review also emerged a fundamental shift in thinking about modelling strategies. Sophisticated *black-box* models which lack descriptive power were employed in the last two decades. This was caused by fast development of machine learning and artificial intelligence techniques. Other aspects of modelling significantly improved over time. Single-period (static) models are replaced by longitudinal-data based models. There is a strong demand on scalability as the problem of Big Data becomes actual. Many authors claim that systems such as Neural Networks or Support Vector Machine outperform standard models. This was supported only partially in the demonstration section (chapter 6.14). From obtained results cannot be concluded that there is some superior combination of data design and analytical method.

Last research question, *What can we learn from the previous research?* aimed at identifying any possible sources of problems which can reduce information value of findings. This reduction can affect transferability of results. Different economic environments and accounting systems were identified as two of the most important. Important consequences of differences in accounting theory were described in chapter 2.5. Modern economy theories which might guide analysts in model development processes were presented in chapter (2.4). Identification of earnings management (EM) represents a big challenge in future research. Consequences of EM were discussed in devoted chapter. Two approaches to corporate finance (standard and value-based) were considered. Identification of performance indicators was based on the standard approach. Returns were identified as the most important health indicator in all models which were created.

Second objective was accomplished by creating several analytical models in chapter 6. General modelling strategies which play imminent role in model identification and tuning were used throughout the whole chapter. An example of Bayesian approach to analytical task which involves expert's knowledge was provided in the beginning of the chapter. Modern probabilistic methods such as Generalised Linear Mixed Model, Joint Modelling approach to Survival Analysis demonstrated its usefulness in both explanatory and predictive tasks. Sophisticated methods from the machine learning and artificial intelligence field, such as Neural Networks, Support Vector Machine, ensemble techniques and decision trees, achieved very good predictive performance.

After two objectives were accomplished, all necessary information for designing a conceptual model were obtained. Conceptual model is presented in Figure 4 on page 35. This model depicts all important steps in the model-building process. Detailed methodological discussion should help any analyst to create more robust models, to understand model's limits, and to get better awareness of corresponding uncertainty.

8 THESIS CONTRIBUTIONS

Performance modelling is an appealing field of study to both theory and practice. Main findings and contributions result from multidisciplinary strategy applied in both theoretical and practical part. This thesis was not planned to be an empirical study. It does not provide the reader with one best solution. Instead, work points to seminal papers, methodological perspectives and most successful techniques.

Conceptual model, which was proposed in Figure 4, provides both academics and practitioners guidance on how to build a model and what to consider in intermediating steps. These steps were described in corresponding chapters.

8.1 Theoretical Contribution

This thesis presents up-to-date perspective on performance measurement from the quantitative perspective. The literature review revealed various inconsistencies among empirical research papers which can be attributed mainly to different economic environments in which research was conducted and to used definitions.

The thesis employs methods and techniques from standard frequentist statistics, Bayesian analysis and machine learning. During the model building process elements of information theory were used. Special care was given to estimation of corresponding uncertainty and model tuning. Also, a novel approach (to best author's knowledge this approach was not used before in the studied field) joined modelling which combines survival analysis and Generalized Linear Mixed Models was presented.

Part of the text was dedicated to the elementary concepts of probability theory which are related to description of uncertainty and to assessment of evidence strength. As some authors argue (works of Andrew Gelman and Deborah Mayo), such concepts are often misinterpreted or misused in academic writings. Competing approaches (frequentist and bayesian) were introduced.

Topics covered in the thesis were presented in both domestic and foreign conferences. Several methodological lectures about uncertainty and statistics with demonstrations in finance were given in UTB and Ton Duc Than University in Saigon.

8.2 Practice

Proposed conceptual model in Figure 4 can be used as a general modelling strategy. This thesis presents state-of-the-art analytical techniques. Also, graphical outcomes are suggested as a convenient tool for representing uncertainty.

All analyses and outcomes in form of data samples and source codes are available on author's Github account and are publicly available. Analysis were programmed in the open-source software R. Analysis were designed to be reproducible.

9 FUTURE RESEARCH DIRECTIONS

Performance modelling will remain an important topic in the future. Although breakthroughs which would change the whole area of research (such as Altman's model) are not expected to appear, many sub-disciplines will be improved significantly. Computational problems and experts' knowledge processing are recognised by author as two major research directions.

The amount of data grows exponentially. This is mostly due to advances in information technology. Term "Big data" which was coined in 1997 gave rise to a new field of computer science. Big data problem does not however, involve only volume of raw data. It also covers problems with increasing speed and variety of generated data. Traditional relational databases are being replaced by non-relational databases (document base). As the number of variables increases, the risk of spurious correlations increase as well. Different sampling methods (batch) are used in initial modelling stage. Demand for fast (or even on-line) processing will require higher computational efficiency. This computational shift will require new methods and redefinitions of standard approaches.

Transformation of knowledge elicited from experts to a form which can be processed is another important challenge. Subjective information can be supplied to expert's systems even today, but only in a limited way. Bayesian analysis allows expressing prior information in a form of probabilistic distribution. Fuzzy and rough sets can capture vague information but its processing is difficult in more sophisticated techniques.

10 CONCLUSION

The topic of performance modelling can be viewed from qualitative, quantitative or mixed perspective. Quantitative approach relies on domain expert's knowledge. Proponents of this approach apply a different set of methods and techniques than quantitative to understand reality. Different sources of information are used as well. Among the various advantages of this approach, the possibility of early identification of problems which did not occur in the past can be considered. These can be revealed by intuition. Intuition is something what cannot be achieved in quantitative analysis as it relies on patterns identified during training process.

On the other hand, qualitative approach results in user-objective model (after analytical model is created final user can not influence its performance) which is created by a transparent set of rules. Such model can be used for scenario or sensitivity analysis. For most of the models, clear descriptions of the uncertainty relating to parameters and performance estimates are viable. Automated procedures can process high volume of data and can be updated in a short period of time.

Performance modelling is a complex field of study. It consists of subjects who do not behave rationally. Uncertainty, which comes both from the limited knowledge and economic environment conditions, is an integral part of the problem. Neither of the aforementioned approaches is capable of capturing the whole scope of the problem. In author's opinion only a balanced combination of the two approaches can lead to better understanding of reality and to successful implementation of analytical model.

References

Monographs

Aggarwal, C. C. (2013). *Outlier analysis*. New York: Springer.

Altman, E. & Hotchkiss, E. (2006). *Corporate financial distress and bankruptcy: predict and avoid bankruptcy, analyze and invest in distressed debt*. Wiley Finance. New Jersey: Wiley.

Bandyopadhyay, S. & Saha, S. (2012). *Unsupervised classification: similarity measures, classical and metaheuristic approaches, and applications*. SpringerLink : Bücher. Berlin: Springer verlag.

Brealey, R., Myers, S., & Allen, F. (2003). *Principles of corporate finance* (7th). New York: McGraw-Hill.

Carlin, B. & Louis, T. (2000). *Bayes and empirical bayes methods for data analysis, second edition*. Texts in Statistical Science. Boca Raton: Taylor & Francis.

Creswell, J. W. (2012). *Qualitative inquiry and research design: choosing among five approaches*. Thousand Oaks: SAGE Publications.

Damodaran, A. (2012). *Investment valuation: tools and techniques for determining the value of any asset, university edition*. New Jersey: Wiley.

Engelbrecht, A. (2007). *Computational intelligence: an introduction*. Chichester: Wiley.

Fahrmeir, L., Hennevogl, W., & Tutz, G. (1994). *Multivariate statistical modelling based on generalized linear models*. Springer Series in Statistics. New York: Springer.

Hair, J. (2010). *Multivariate data analysis*. New Jersey: Prentice Hall.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference and prediction* (2nd ed.). New York: Springer.

Hawkings, D. M. (1980). *Identification of outliers*. London: Chapman and Hall.

Holler, A. (2009). *New metrics for value-based management: enhancement of performance measurement and empirical evidence on value-relevance*. Unternehmensführung & Controlling. Wiesbaden: Gabler Verlag.

Christensen, R., Johnson, W., Branscum, A., & Hanson, T. (2011). *Bayesian ideas and data analysis: an introduction for scientists and statisticians*. Chapman & Hall/CRC Texts in Statistical Science. Boca Raton: Taylor & Francis.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: with applications in R*. Springer Texts in Statistics. New York: Springer.

Jeff Gill, A. J. W. (2013). *The sage handbook of multilevel modeling*. Thousand Oaks: SAGE Publications Ltd.

Knight, F. (1964). *Risk, uncertainty and profit*. Reprints of economic classics. New York: Augustus M. Kelley.

Kruschke, J. K. (2011). *Doing bayesian data analysis: a tutorial with r and bugs*. Oxford: Academic Press.

Kuhn, M. & Johnson, K. (2013). *Applied predictive modeling*. SpringerLink. New York: Springer.

Mackenzie, B., Coetsee, D., Njikizana, T., Chamboko, R., & Colyvas, B. (2011). *Wiley interpretation and application of international financial reporting standards 2011 (wiley ifrs: interpretation & application of international financial reporting standards)*. Wiley.

McKinsey, Company, & Koller. (2010). *Valuation: measuring and managing the value of companies*. Wiley Finance. New Jersey: Wiley.

Mulford, C. W. & Comiskey, E. E. (2002). *The financial numbers game: detecting creative accounting practices*. New York: Wiley.

A dictionary of accounting. (2010). Oxford: Oxford University Press.

Pickard, A. (2007). *Research methods in information*. London: Facet.

Robert, C. (2007). *The bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Texts in Statistics. Berlin: Springer.

Rosenbaum, P. R. (2010). *Design of observational studies*. Springer Series in Statistics. New York: Springer.

Ross, S., Westerfield, R., & Jordan, B. (2008). *Fundamentals of corporate finance* (8th ed., Standard ed.). Finance, insurance and real estate. New York: McGraw-Hill/Irwin.

Vapnik, V. N. (1995). *The nature of statistical learning theory*. New York: Springer.

Articles

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, *23*(4), 589–609. doi:10.2307/2978933

Angelis, V. d., Felici, G., & Mancinelli, G. (2006). Feature selection for data mining. *Data Mining and Knowledge Discovery Approaches Based on Rule Induction Techniques*. Massive Computing, (6), 227–252.

Angilella, S. & Mazzù, S. (2015). The financing of innovative SMEs: A multicriteria credit rating model. *European Journal of Operational Research*, *244*(2), 540–554. doi:10.1016/j.ejor.2015.01.033

Atiya, A. (2001). Bankruptcy prediction for credit risk using neural networks: A survey and new results. *IEEE Transactions on Neural Networks*, *12*(4), 929–935. doi:10.1109/72.935101

Badertscher, B. A., Collins, D. W., & Lys, T. Z. (2012). Discretionary accounting choices and the predictive ability of accruals with respect to future cash flows. *Journal of Accounting and Economics*, *53*(1–2), 330–352. doi:10.1016/j.jacceco.2011.11.003

Bahrammirzaee, A. (2010). A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems. *Neural Computing and Applications*, *19*(8), 1165–1195. doi:10.1007/s00521-010-0362-z

Balcaen, S. & Ooghe, H. (2006). 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. *The British Accounting Review*, *38*(1), 63–93. doi:10.1016/j.bar.2005.09.001

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, *4*, pages.

BIS. (2005). Working paper no. 14. *Studies on the Validation of Internal Rating Systems*, 63.

Black, F. & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, *81*(3), 637–654.

Bottazzi, G., Grazzi, M., Secchi, A., & Tamagni, F. (2011). Financial and economic determinants of firm default. *Journal of Evolutionary Economics*, *21*(3), 373–406. doi:10.1007/s00191-011-0224-6

Brazel, J. F., Jones, K. L., & Zimbelman, M. F. (2009). Using nonfinancial measures to assess fraud risk. *Journal of Accounting Research*, *47*(5), 1135–1166. doi:10.1111/j.1475-679X.2009.00349.x

Burgstahler, D. & Dichev, I. (1997). Earnings management to avoid earnings decreases and losses. *Journal of Accounting and Economics*, *24*, 99–126. doi:10.1016/S0165-4101(97)00017-7

Carcello, J. V. & Hermanson, D. R. (1995). Temporal changes in bankruptcy-related reporting. *Auditing*, *14*(2), 133–143.

Çelik, Ş. (2013). Micro credit risk metrics: a comprehensive review. *Intelligent Systems in Accounting, Finance and Management*, *20*(4), 233–272. doi:10.1002/isaf.1344

Coakley, J. R. & Brown, C. E. (2000). Artificial neural networks in accounting and finance: modeling issues. *Intelligent Systems in Accounting, Finance & Management*, *9*(2), 119–144. doi:10.1002/1099-1174(200006)9:2¡119::AID-ISAF182¿3.0.CO;2-Y

Coleman, L. (2014). Why finance theory fails to survive contact with the real world: A fund manager perspective. *Critical Perspectives on Accounting*, *25*(3), 226–236. doi:10.1016/j.cpa.2013.02.001

Deakin, E. B. (1972). A discriminant analysis of predictors of business failure. *Journal of Accounting Research*, *10*(1), 167–179. doi:10.2307/2490225

DeFond, M. L. (2010). Earnings quality research: advances, challenges and future research. *Journal of Accounting and Economics*, *50*(2–3), 402–409. doi:10.1016/j.jacceco.2010.10.004

Dechow, P., Ge, W., & Schrand, C. (2010). Understanding earnings quality: a review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics*, *50*(2–3), 344–401. doi:10.1016/j.jacceco.2010.09.001

Dempsey, M. (2014). The Modigliani and Miller Propositions: The History of a Failed Foundation for Corporate Finance? *Abacus*, *50*(3), 279–295. doi:10.1111/abac.12030

Dietterich, T. G. & Bakiri, G. (1995). Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research*, *2*, 171–178. doi:10.1613/jair.105

Dobbs, R. & Koller, T. (2005). Measuring long-term performance. *McKinsey on Finance*, *4*(16), 1–6.

Dubois, D. & Prade, H. (1998). An introduction to fuzzy systems. *Clinica Chimica Acta*, *270*(1), 3–29. doi:10.1016/S0009-8981(97)00232-5

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, *7*(1), 1–26. doi:10.1214/aos/1176344552

Fields, T. D., Lys, T. Z., & Vincent, L. (2001). Empirical research on accounting choice. *Journal of Accounting and Economics*, *31*(1–3), 255–307. doi:10.1016/S0165-4101(01)00028-3

Francis, J. (2001). Discussion of empirical research on accounting choice. *Journal of Accounting and Economics*, *31*(1–3), 309–319. doi:10.1016/S0165-4101(01)00017-9

Goleţ, I. (2014). Symmetric and asymmetric binary choice models for corporate bankruptcy. *Procedia - Social and Behavioral Sciences*. Challenges and Innovations in Management and Leadership 12th International Symposium in Management, *124*, 282–291. doi:10.1016/j.sbspro.2014.02.487

Grice, J. S. & Ingram, R. W. (2001). Tests of the generalizability of altman's bankruptcy prediction model. *Journal of Business Research*, *54*(1), 53–61. doi:10.1016/S0148-2963(00)00126-0

Gupta, J., Gregoriou, A., & Healy, J. (2014). The Correct Use of Hazard Models: A Comparison Employing Different Definitions of SMEs Financial Distress. *SSRN Scholarly Paper*, (ID 2457917).

Guyon, I. & Elisseeff, A. (2003). An introduction to variable and feature selection. *An Introduction to Variable and Feature Selection*, *3*, 1157–1182.

Hassab Elnaby, H. R., Mohammad, E., & Said, A. A. (2010). Nonfinancial performance measures and earnings management. *Advances in Management Accounting*, *18*, 55–79. doi:10.1108/S1474-7871(2010)0000018006

Hayek, F. A. v. (1942). Scientism and the study of society. part i. *Economica*. New Series, *9*(35), 267–291. doi:10.2307/2549540

Healy, P. & Wahlen, J. (1999). A review of the earnings management literature and its implications for standard setting. *Accounting Horizons*, *13*(4), 365–383. doi:10.2308/acch.1999.13.4.365

Hernandez Tinoco, M. & Wilson, N. (2013). Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis*, *30*, 394–419. doi:10.1016/j.irfa.2013.02.013

Hernández-Orallo, J., Flach, P., Ferri, C., & Elkan, C. (2012). A unified view of performance metrics: Translating threshold choices into expected classification loss. *Journal of Machine Learning Research*.

Homolka, L. (2013). Extreme value approach value approach for estimating value at risk metrics with respect to basel ii. *International Journal of Mathematics and Computers in Simulation*, *7*(2), 171–178. doi:10.1016/j.jacceco.2005.04.005

Homolka, Doležal, J., & Novák, P. (2014). Longitudinal design as an extension to standard banking approaches to companies' financial performance assessment and prediction. *Proceedings of the 11th International Scientific Conference*, (11), 236–243.

Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*. Neural NetworksSelected Papers from the 7th Brazilian Symposium on Neural Networks (SBRN '04)7th Brazilian Symposium on Neural Networks, *70*(1–3), 489–501. doi:10.1016/j.neucom.2005.12.126

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. doi:10.1613/jair.953

Chen, N., Ribeiro, B., Vieira, A., & Chen, A. (2013). Clustering and visualization of bankruptcy trajectory using self-organizing map. *Expert Systems with Applications*, *40*(1), 385–393. doi:10.1016/j.eswa.2012.07.047

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical*, *Vol. 186*(No. 1007), pages.

Jensen, M. C. & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, *3*(4), 305–360. doi:10.1016/0304-405X(76)90026-X

Jensen, M. C. & Ruback, R. S. (1983). The market for corporate control: The scientific evidence. *Journal of Financial Economics*, *11*(1–4), 5–50. doi:10.1016/0304-405X(83)90004-1

Keasey, K. & Hudson, R. (2007). Finance theory: A house without windows. *Critical Perspectives on Accounting*, *18*(8), 932–951. doi:10.1016/j.cpa.2006.06.002

Keasey, K., Pindado, J., & Rodrigues, L. (2014). The determinants of the costs of financial distress in SMEs. *International Small Business Journal*. doi:10.1177/0266242614529317

Kohavi, R. & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*. Relevance, *97*(1–2), 273–324. doi:10.1016/S0004-3702(97)00043-X

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*(1), 59–69. doi:10.1007/BF00337288

Lee, R. T. (2013). Designing effective metrics. *Journal of Corporate Accounting & Finance*, *24*(5), 69–73. doi:10.1002/jcaf.21876

Lee, S. & Kim, W. G. (2009). EVA, refined EVA, MVA, or traditional performance measures for the hospitality industry? *International Journal of Hospitality Management*, *28*(3), 439–445. doi:10.1016/j.ijhm.2009.01.004

Lee, Y.-C. (2007). Application of support vector machines to corporate credit rating prediction. *Expert Systems with Applications*, *33*(1), 67–74. doi:10.1016/j.eswa.2006.04.018

Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, *247*(1), 124–136. doi:10.1016/j.ejor.2015.05.030

Lin, H.-T., Lin, C.-J., & Weng, R. C. (2007). A note on Platt's probabilistic outputs for support vector machines. *Machine Learning*, *68*(3). doi:10.1007/s10994-007-5018-6

Lin, W.-Y., Hu, Y.-H., & Tsai, C.-F. (2012). Machine learning in financial crisis prediction: a survey. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, *42*(4), 421–436. doi:10.1109/TSMCC.2011.2170420

Lo, K. (2008). Earnings management and earnings quality. *Journal of Accounting and Economics*, *45*(2–3), 350–357. doi:10.1016/j.jacceco.2007.08.002

Lu, Y.-C., Wei, Y.-C., & Chang, T.-Y. (2015). The effects and applicability of financial media reports on corporate default ratings. *International Review of Economics & Finance*. Corporate Financing and Banks in Greater China, *36*, 69–87. doi:10.1016/j.iref.2014.11.008

MacPhail, F. (1998). Moving Beyond Statistical Validity in Economics. *Social Indicators Research*, *45*(1-3), 119–149. doi:10.1023/A:1006989612799

Magni, C. A. (2009). Splitting up value: a critical review of residual income theories. *European Journal of Operational Research*, *198*(1), 1–22. doi:10.1016/j.ejor.2008.09.018

Majumder, D. (2006). Inefficient markets and credit risk modeling: why merton's model failed. *Journal of Policy Modeling*, *28*(3), 307–318. doi:10.1016/j.jpolmod.2005.10.006

Markowitz, H. (1952). Portfolio Selection*. *The Journal of Finance*, *7*(1), 77–91. doi:10.1111/j.1540-6261.1952.tb01525.x

Marqués, A. I., García, V., & Sánchez, J. S. (2013). A literature review on the application of evolutionary computing to credit scoring. *Journal of the Operational Research Society*, *64*(9), 1384–1399. doi:10.1057/jors.2012.145

Melnyk, S. A., Stewart, D. M., & Swink, M. (2004). Metrics and performance measurement in operations management: dealing with the metrics maze. *Journal of Operations Management*, *22*(3), 209–218. doi:10.1016/j.jom.2004.01.004

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

Modigliani, F. & Miller, M. H. (1958). The Cost of Capital, Corporation Finance and the Theory of Investment. *The American Economic Review*, *48*(3), 261–297.

Morgan, J. (1997). Creditmetrics(tm) — technical document.

Moro, A. & Fink, M. (2013). Loan managers' trust and credit access for SMEs. *Journal of Banking & Finance*, *37*(3), 927–936. doi:10.1016/j.jbankfin.2012.10.023

Nayak, G. N. & Turvey, C. G. (1997). Credit Risk Assessment and the Opportunity Costs of Loan Misclassification. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*, *45*(3), 285–299. doi:10.1111/j.1744-7976.1997.tb00209.x

Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, *50*(3), 559–569. doi:10.1016/j.dss.2010.08.006

Nouy, D. (1999). Credit risk modelling: current practices and applications.

Odom, M. & Sharda, R. (1990). A neural network model for bankruptcy prediction. In *, 1990 IJCNN International Joint Conference on Neural Networks, 1990* (pp. 163–168 vol.2). doi:10.1109/IJCNN.1990.137710

Perols, J. L. & Lougee, B. A. (2011). The relation between earnings management and financial statement fraud. *Advances in Accounting*, *27*, 39–53. doi:10.1016/j.adiac.2010.10.004

Philosophov, L. V., Batten, J. A., & Philosophov, V. L. (2008). Predicting the event and time horizon of bankruptcy using financial ratios and the maturity schedule of long-term debt. *Mathematics and Financial Economics*, *1*(3-4), 181–212.

Platt, H. & Platt, M. (2012). Corporate board attributes and bankruptcy. *Journal of Business Research*, *65*(8), 1139–1143. doi:10.1016/j.jbusres.2011.08.003

Provost, F. & Fawcett, T. (2001). Robust Classification for Imprecise Environments. *Machine Learning*, *42*(3), 203–231. doi:10.1023/A:1007601015854

Ravi Kumar, P. & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques – a review. *European Journal of Operational Research*, *180*. doi:10.1016/j.ejor.2006.08.043

Scott, J. (1981). The probability of bankruptcy: A comparison of empirical predictions and theoretical models. *Journal of Banking & Finance*, *5*(3), 317–344. doi:10.1016/0378-4266(81)90029-7

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal, The*, *27*(3), 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x

Shumway, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *The Journal of Business*, *74*, 101–124. doi:10.1086/jb.2001.74.issue-1

Schinckus, C. (2009). Economic uncertainty and econophysics. *Physica A: Statistical Mechanics and its Applications*, *388*(20), 4415–4423. doi:10.1016/j.physa.2009.07.008

Sun, J., Li, H., Huang, Q.-H., & He, K.-Y. (2014). Predicting financial distress and corporate failure: a review from the state-of-the-art definitions, modeling, sampling, and featuring approaches. *Knowledge-Based Systems*, *57*, 41–56.

Tsay, R. S., Lin, Y.-M., & Wang, H.-W. (2009). Residual income, non-earnings information, and information content. *Journal of Forecasting*, *28*(6), 487–511. doi:10.1002/for.1104

Watts, R. L. & Zimmerman, J. L. (1990). Positive accounting theory: a ten year perspective. *The Accounting Review*, *65*, 131–156. ArticleType: research-article / Full publication date: Jan., 1990 / Copyright © 1990 American Accounting Association. doi:10.2307/247880

Zadeh, L. (1965). Fuzzy sets. *Information and Control*, *8*, 338—353.

Zhong, H., Miao, C., Shen, Z., & Feng, Y. (2014). Comparing the learning effectiveness of BP, ELM, I-ELM, and SVM for corporate credit ratings. *Neurocomputing*, *128*, 285–295. doi:10.1016/j.neucom.2013.02.054

Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, *22*, 59–82. doi:10.2307/2490859

LIST OF PUBLICATIONS

Articles

Belás, J. & Homolka, L. (2013a). The development of customers' satisfaction in Slovak banking sector during the crisis. *In Conference Proceedings Enterprise and the Competitive Environment*, *2*, 1–22.

Belás, J. & Homolka, L. (2013b). The development of customers' satisfaction in the banking sector of Slovakia in the turbulent economic environment. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, *61*(7), 221–230.

Homolka, L. (2012a). Extreme Value Theory and bootstrap approach for robust value at risk estimation under Basel II. *Proceedings of the 1st WSEAS International Conference on Finance, Accounting and Auditing*, *12*, 68–74.

Homolka, L. (2012b). Substantive relevance and accuracy of information provided by methods of statistical analysis in financial decision-making process. *Sborník příspěvků 8. ročníku Mezinárodní Baťovy konference pro doktorandy a mladé vědecké pracovníky DOKBAT 2012*, (8), 80–85.

Homolka, L. (2013a). Does distribution of accruals change with different payment morality? empirical study from the Czech transportation industry. *World academy of science, engineering and technology*, (82), 536–539.

Homolka, L. (2013b). Extreme value approach value approach for estimating Value at Risk metrics with respect to Basel II. *International Journal of Mathematics and Computers in Simulations*, *7*(2), 171–178.

Homolka, L., Belás, J., & Doležal, J. (2015). Subsample analysis in structural equation modelling. Example from customer satisfaction in banking industry. *In Proceedings of the 7th International Scientific Conference*, *7*, 400–409.

Homolka, L., Doležal, J., & Novák, P. (2014). Longitudinal design as an extension to standard banking approaches to companies' financial performance assessment and prediction. *Proceedings of European Financial Systems*, *2*, 236–244.

Homolka, L. & Hubáčková, V. (2011). Possibilities of application of statistical methods for evaluating regional differentiation of foreign direct investment in the Czech Republic in 2008. *In Proceedings of the 5th International Scientific Conference Finance and the Performance of Firms in Science, Education and Practice*, (5), 171–178.

Homolka, L., Knápková, A., & Pavelková, D. (2015). Plastics cluster members and their competitors - DEA benchmarking study. *In Proceedings of the 7th International Scientific Conference*, *7*, 409–414.

Huspeninová, K. & Homolka, L. (2012). Virtual enterprise: benefits and its quantification. *In Proceedings of the 8th European Conference on Management, Leadership and Governance*, *8*, 497–504.

Jirčíková, E., Pavelková, D., Bialic-Davendra, M., & Homolka, L. (2013). The age of clusters and its influence on their activity preferences. *Technological and Economic Development of Economy*, *19*(4), 621–637.

Knápková, A., Homolka, L., & Pavelková, D. (2014a). Využití balanced scorecard a vliv jeho využívání na finanční výkonnost podniků v ČR. *E+M*, *17*, 146–160.

Knápková, A., Homolka, L., & Pavelková, D. (2014b). Využití ekonomické přidané hodnoty a vliv jejího využívání na finanční výkonnost podniků v ČR. *Trendy ekonomiky a managementu*, *19*, 18–26.

Knápková, A., Pavelková, D., & Homolka, L. (2014). Benchmarking: can it increase the company financial performance? *Proceedings ICFE 2014 - The International Conference on Finance and Economics*, *3*, 295–305.

Nguyen, H., Homolka, L., Molnár, Z., & Macdonald, R. (2011). Knowledge management as a solution for the shortage of competent employees in SMEs at the developing country (Case study: Vietnam). *Journal of Systems Integration*, *3*(1), 38–46.

Otrusinová, M., Homolka, L., & Pastuszková, E. (2015). Influence of statistical methods on efficiency of financial audit in public administration. *In Proceedings of the 7th International Scientific Conference*, *7*, 1133–1142.

Paseková, M., Homolka, L., Strouhal, J., & Müllerová, L. (2011). Accounting principles versus taxation: how behave Czech SMEs? (on case of impairment issues). *Proceedings of the 6th IASME/WSEAS International Conference on Economy and Management Transformation*, (6), 50–60.

Pavelková, D., Bialic-Davendra, M., Vejmělková, E., & Homolka, L. (2013). Clusters' activities and economy stage of development: evidence from V4 and advanced economies. *Ekonomický časopis*, *61*(2), 187–205.

Strouhal, J., Paseková, M., & Homolka, L. (2012). Bad debt provisions and impairment: how behave Czech SMEs? *Sborník z mezinárodní vědecké konference: Transakční náklady českých ekonomických subjektů v insolvenčním řízení, možnosti jejich snižování a zlepšení statistiky insolvenčních řízení. 1*, 76–81.

Svoboda, J. & Homolka, L. (2012). Can aggregate performance metric be used to predict future returns? *In Proceedings of the 18th International Business Information Management Association*, *18*, 73–79.

Monographs

Belás, J. & et al. (2013). *Řízení úvěrového rizika SME*. Žilina: Georg.

CURRICULUM VITAE

Personal Information

Full name:	Lubor Homolka
Residence:	M. Alše 450, Zlín
Email:	homolka@fame.utb.cz

Education and Qualification works

now	Ph.D. candidate	TBU - Finance (in Czech)
		Ton Duc Than University, Ho Chi Minh City, Vietnam 2015 (Jan – Mar)
		Universidad de León, Spain 2013 (Oct – Dec)
2010	Master's degree	TBU - Finance (in Czech)
		The Project of Operational Rating Model Creation for UniCredit Bank CZ
2008	Bachelor's degree	Tomas Baťa University (TBU) - Economics and Management (in Czech)
		Visual Data Mining and Its Practical Application

Professional Experience

2015 (Jan – Mar)	internship at **BAOVIET Securities, Research Department**
	- Ho Chi Minh Stock Index Analysis
	- development of prediction model (VAR) in R and Shiny
2014 (Feb – Dec)	part time job at **SAB Finance, a.s.**
	- responsibility for data analysis
	- collaboration over monthly financial reports
	- functionalities testing of a new managerial information system
2011–2015	**Tomas Baťa University**
	- consultancy and lectures on methodology and statistics
	- publishing of research papers and book chapters
	- supervision of bachelor qualification works
	- journal and conference papers reviewing
before 2011	**Temporary jobs**
	- offering catch-up courses on statistics and computer skills

Extracurricular Courses

2014	Six sigma and lean management
2012	Summer school of Statistics – Methods for Industry with Minitab
2009	Project Management, IPMA level D Certificate

Language Skills

English	advanced
Czech	native

Participation in Projects

1. Operational Programme Education for Competitiveness co-funded by the European Social Fund (ESF) and national budget of the Czech Republic - Human Resources Development in the field of Measurement and Management of Companies, Clusters and Regions Performance, No. CZ.1.07/2.3.00/20.0147

2. IGA - UTB: Optimization of internal rating model parameters of commercial banks in the small and medium enterprises

3. IGA - UTB: Creating a Model for the Performance Measurement and Management of Enterprises No. 402/09/1739

4. GaCR - IAS/IFRS Usage in Small and Medium-sized Enterprises and its Influence on Performance Measurement No. 402/09/0225

5. IGA - UTB: Development and performance evaluation of clusters, cluster policies and cluster members using the principles of benchmarking