Models and Methods for Automated Credit Rating Prediction

Claude Gangolf

Operation Research and Business Informatics *and* E-Science

University of Saarland *and* Luxembourg Institute of Science and Technology

This dissertation is submitted for the degree of

Doctor Rerum Oeconomicarum

June 2016

Defense of the PhD thesis

Disputation Committee:

Chairman	Univ.-Prof. Dr. Dieter Schmidtchen
First rapporteur	Univ.-Prof. Dr. Günter Schmidt
Second rapporteur	Univ.-Prof. Dr. Imed Kacem
Observer	Dr. Stefanie Becker

The defense of the thesis was held on the 31st Mai 2016.

The dissertation was passed with ***magna cum laude***.

This PhD thesis is dedicated to my parents and my sister.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Claude Gangolf
June 2016

Acknowledgements

Abstract

The problem of prediction the rating degrees of bonds is treated in this work. In the financial world, risk assessment of bonds is almost entirely conducted by rating agencies, like Moody's Investors Service (Moody's), Standard and Poor's (S&P) and Fitch's Ratings (Fitch). Using the rating scale of Moody's, there exists 21 rating degrees, with *Aaa* the best degree and *C* the worst degree indicating that the bond has defaulted. Nevertheless, with the beginning of the financial crisis, starting in 2009, criticism is raised concerning their rating procedure. In this way, automated credit rating prediction (ACRP) models are developed to investigate the possibility to obtain an alternative to the rating agencies. The existing ACRP models have some main limits which are depicted in this work. A new ACRP model is developed to overcome some of the identified drawbacks.

First, the benefit of personal financial planning (PFP) tools is discussed and their limitation to give a suggestion how private persons should invest is presented. The requirements for the ideal ACRP model which is able to overcome this limitation by being integrated into a PFP tool are stated. The need of such an integration is shown by a scenario of an exemplary user. Second, the necessary definitions from the financial world are specified. Financial bonds are described in detail and the different types of bonds are explained. As bonds represent a kind of debt instrument for countries and companies, the distinction between bonds issued by countries and those emitted by companies is made. Sovereign bonds are issued by states and corporate bonds are putted into circulation by companies. Additionally, the different basic features of bonds, like coupon rate, maturity and currency notation, are defined. The general rating procedure of the rating agencies is also explained. This allows to differentiate the employed approaches to undertake a risk evaluation if an ACRP model is used or the rating agencies undertake the evaluation.

Third, different classification techniques are globally introduced. In the description of the technique, there is no assumption made concerning a specific use-case. The presented pseudo-codes of the techniques allows their reproduction and their employ at any use-case. The description is not limited to the financial world. The described techniques are: artificial neural

network (ANN), support vector machines (SVM) and support vector domain description (SVDD).

Fourth, three exemplary ACRP models are discussed. The first model is based on ANN and is designed to predict the rating degrees of sovereign bonds [7]. SVM is used in the second model. This model tries to predict the correct rating group of corporate bonds [48]. A rating group is a merging of several rating degrees. The last discussed ACRP model is based on SVDD and it is also developed to predict the correct rating groups of corporate bonds [40]. The majority of ACRP models are limited to the handling of corporate bonds. This is due to the fact that risk assessment of corporate bonds is a more lucrative market than the risk evaluation of sovereign bonds.

Fifth, a new ACRP model has been developed with the idea to rate sovereign and corporate bonds simultaneously. The employed information have to be publicly available so that private investor have easily access to them. The model is based on SVDD and linear regression (LR). In a first step, the bonds are divided into several rating groups with the help of SVDD. Afterwards, in each rating group, LR is used to predict the final rating degrees of the bonds. The mathematical programs of the model are described in detailed and the procedure of the model is explained with the help of an artificial example. A competitive and an empirical analysis is undertaken to investigate the performance of the model.

Sixth, the development of a prototype is tackled to give a proof of concept for the stated requirements for the ideal ACRP model. The new ACRP model is integrated into a PFP tool called LifeCharts. The implementation of the prototype allows to abstract the ACRP model in such a way that private investors can easily handle it and to make it accessible to a greater public. In the final version of the prototype, a portfolio selection tool is integrated to help the investor to take an investment decision. The implemented prototype fulfills nearly all the stated requirements, i.e., in the prototype, the user-given variables are reduced to maintain its usability and sovereign and corporate bonds are handled. The requirements of different maturities and different types of coupon rates are only partially fulfilled.

Finally, several points of interest, which has been identified, are discussed. These identified problems, like the limitation of the new ACRP problems or the limitation in the implementation in the prototype, represent starting points for future research.

Table of contents

List of figures

List of tables

List of Symbols

α_i	Lagrange factors, with $i = 1, \ldots, n$
Δ	Fixed threshold value
δ	Distance between two adjacent rating degrees
Δ_i	Distance between predicted and correct rating degree, with $i = 1, \ldots, \tilde{n}$
ε	Error criterion
ε^D	First non-investment grade rating degree
γ	Weighting exponent
λ	Regulation constant
λ_{hid}	Learning rate in the hidden layer
λ_{out}	Learning rate in the output layer
$\mathbf{A}$	Matrix of attributes
$\mathbf{B}$	Set of rated bonds
$\mathbf{w}$	Normal to the hyperplane
$\mathscr{A}$	Set of all possible attributes combinations
$\mathscr{H}$	Feature space
μ_{hid}	Momentum of the hidden layer
μ_{out}	Momentum of the output layer
v_i	Membership vector of bond i, with $i = 1, \ldots, n$

Ω_d^D	Rating degree d, with $d = 1, \ldots, D$
Ω_d^D	Rating group g, with $g = 1, \ldots, G$
Π_1^C	Set of investment grade rating degrees
Π_2^C	Set of non-investment grade rating degrees
Π_i^G	Set of rating degrees in rating group g, with $g = 1, \ldots, G$
σ	Kernel factor
θ	Threshold value to identify the rating group
$\tilde{A}_i$	Attribute vector of unrated bond i, with $i = 1, \ldots, \tilde{n}$
$\tilde{A}$	Matrix of attributes
$\tilde{a}_{ij}$	Attribute value of unrated bond i, with $i = 1, \ldots, \tilde{n}$ and $j = 1, \ldots, m$
$\tilde{B}_i$	Unrated bond i, with $i = 1, \ldots, \tilde{n}$
$\tilde{B}$	Set of unrated bonds
$\tilde{n}$	Number of testing data points
$\tilde{R}_i^D$	Predicted rating characteristic of unrated bond i, with $i = 1, \ldots, \tilde{n}$
$\tilde{R}_i^D$	Predicted rating degree of unrated bond i, with $i = 1, \ldots, \tilde{n}$
$\tilde{R}_i^D$	Predicted rating group of unrated bond i, with $i = 1, \ldots, \tilde{n}$
$\tilde{R}_P^D$	Computed rating degree of the portfolio
$\tilde{x}_i$	Testing point i, with $i = 1, \ldots, \tilde{n}$
$\tilde{X}$	Testing set
$\tilde{x}_{ij}$	Attribute j of testing point i, with $i = 1, \ldots, \tilde{n}$ and $j = 1, \ldots, m$
υ_i	Lagrange factors, with $i = 1, \ldots, n$
ξ_i	Slack variables, with $i = 1, \ldots, n$
A_g^*	Attribute vector of center B_g^*, with $g = 1, \ldots, G$
A_i	Attribute vector of bond i, with $i = 1, \ldots, n$

a_{ij}	Attribute value of bond i, with $i = 1, \ldots, n$ and $j = 1, \ldots, m$
actual	Predicted result
b	Bias of the hyperplane
B_g^*	Center of hypersphere describing rating group g, with $g = 1, \ldots, G$
B_i	Rated bond i, with $i = 1, \ldots, n$
b_{gj}	Regression factor j for rating group g, with $j = 0, \ldots, m$ and $g = 1, \ldots, G$
C	Trade-off constant
c_{max}	Competitive ratio
D	Number of rating degrees
d_l^g	Starting rating degree of rating group g, with $g = 1, \ldots, G$
d_r^g	Finishing rating degree of rating group g, with $g = 1, \ldots, G$
G	Number of (rating) groups
G_g	Group g, with $g = 1, \ldots, G$
H	Optimal hyperplane
H_1	Auxiliary hyperplane 1
H_2	Auxiliary hyperplane 2
hbias	Bias in the hidden layer
Hidden_Neurons	Number of hidden neurons
m	Number of attributes
n	Number of data points
n_g	Number of data points in group g, with $g = 1, \ldots, G$
obias	Bias in the output layer
Output_Neurons	Number of output neurons

R	Radius of the hypersphere
R_i^C	Given rating characteristic of bond i, with $i = 1,\dots,n$
R_i^D	Given rating degree of bond i, with $i = 1,\dots,n$
R_P^D	User-given maximal rating degree for the portfolio
R_i^G	Given rating group of bond i, with $i = 1,\dots,n$
R_g	Radius of the hypersphere describing group g, with $g = 1,\dots,G$
r_i	Expected return of bond i, with $i = 1,\dots,\tilde{n}$
r_P	Expected return of the portfolio
s_i	Parts of bond i in the portfolio, with $i = 1,\dots,\tilde{n}$
$target$	Given output for one data point
w_{ho}^{new}	Updated weights of the connections between hidden and output layer
w_{ih}^{new}	Updated weights of the connections between input and hidden layer
W_g	Weight vector for group g, with $g = 1,\dots,G$
w_{ho}	Weights of the connections between hidden and output layer
w_{ih}	Weights of the connections between input and hidden layer
x^*	Center of the hypersphere
y_i	Given output of data point i, with $i = 1,\dots,n$
x_i	Data point i, with $i = 1,\dots,n$
X	Dataset set
x_{ij}	Attribute j of data point i, with $i = 1,\dots,n$ and $j = 1,\dots,m$

List of Abbreviations

ACRP   Automated Credit Rating Prediction

ACRP_ANN   ACRP model based on ANN

ACRP_SVDD   ACRP model based on SVDD

ACRP_SVM   ACRP model based on SVM

ALG   Arbitrary ACRP model

ANN   Artificial Neural Network

EURIBOR   Euro Interbank Offered Rate

Fitch   Fitch Ratings

IMF   International Monetary Fund

LC   Linkage Criterion

LIBOR   London Interbank Offered Rate

LR   Linear Regression

Moody's   Moody's Investors Service

New_ACRP   New developed ACRP

OECD   Organization for Economic Cooperation and Development

OPT   Optimal benchmark ACRP model

PFP   Personal financial planning

S&P   Standard and Poors'

SEC   U.S. Securities and Exchange Commission

SVDD   Support Vector Domain Description

SVM   Support Vector Machines

Chapter 1

Introduction

This chapter introduces the terms as bonds, financial risk analysis, credit ratings and rating agencies. Additionally, automated credit rating prediction as an alternative to rating agencies is introduced. Afterwards the three research questions are motivated and formulated. The chapter concludes with the structure of the thesis.

1.1 Preliminaries

In finance, the financial assets can mainly be distinguished between shares and bonds. Shares and bonds are the major parts of assets traded on the market in contrast to other products like futures [69]. Shares offer the holder a membership of the underlying company. Therefore, the return of the investment is directly linked to the financial performance of the company. In the case of distress, the investors have to assume their part of the loss according to their investment in the company. In contrast, a bond is defined as being a debt security which offers the investor contractual payments for a given time period [69]. Beside the contractual payments, bonds have also to fulfill other regulations, like the bond indenture which fixes the rights of the investors [45, 69]. When all the regulations have been complied, then the bonds can be quoted on the market [69]. The process of quoting a bond on the market is known as issuing a bond. An issuer is defined as a company or a state which has undertaken the process of quoting a bond.

There are two main groups of market participants: institutional and private investors. According to the U.S. Securities and Exchange Commission (SEC) [80], bank trusts, insurance companies, mutual and pension funds are defined as institutional investors. All the other participants are considered as private investors. This thesis specifically addresses private investors and focuses on the risk analysis of bonds. Institutional investors have their own

department to monitor and supervise possible financial risk which can occur due to several taken investments and are therefore not taken into account [49]. In contrast, private investors, and especially individual persons, do not have this possibility. Private investors, mostly, take their decision driven by liquidity or psychological consideration without taking a more detailed look to the chosen bond [19]. A first view on the financial risk of a bond can be obtained by looking its credit rating. The credit rating expresses the probability that the issuer is unable to refund his debt and represents only the opinion of the disbursing analyzing company [83]. A major point of interest for a bond holder is to be able to assess the possible risk of losing the invested money. The risk can be read from these credit ratings. At the moment, there are three main analyzing companies called rating agencies. These rating agencies are: Moody's Investors Service (Moody's), Standard and Poor's (S&P) and Fitch's Ratings (Fitch). These rating agencies utilize a scale of credit ratings (see Figure 3.1). According to this scale, a bond with a credit rating Aaa is declared as almost risk-free. In contrast, a bond with a credit rating C states that the issuer is unable to pay the contractual payments to the investors.

The history of private companies which undertake risk analysis of financial assets is very long. At the beginning of the 20 century, the need for risk analysis arose because more and more assets were traded on the financial market. Institutional investors as well as private investors were not able to get an overview of the risk of all the different assets. Additionally, financial bonds became more and more popular. In 1909, Moody's was the first company which undertook risk analysis of financial bonds and published them. In the following twenty years, different companies like Standard Statistics Company, Poor's Publishing Company and Fitch Publishing Company, were founded. After several mergers, Standard and Poor's as well as Fitch Ratings were launched. The typical process of rating agencies to analyze the credit risk is as follows. An analyst collects the public available information about the asset, for example the financial balance sheets of the issuing company. Additionally, a meeting with the company's management is organized. In this meeting, private information concerning the financial situation as well as the outlook of the company are discussed. The public and private information are analyzed according to the model of the rating agency and out of this result, the analyst suggests a rating degree for the asset. A committee of three persons is then responsible to take the final decision. This committee can adopt the suggestion of the analyst or increase or decrease the rating degree. The final decision is communicated to the contracting company and the company can choose if the rating degree is published or not [97, 69]. Another point is that the company has to pay a fee to obtain a risk analysis and at the end, the rating degree. For this reason, there is a possible conflict of interest because on the one side, the rating agency wants to have a satisfied client who is also willing to come

back and on the other side, its own credibility is at stake [5, 10]. Therefore, the investors have to trust the rating agencies to undertake an accurate analysis and publish the correct respective rating degrees.

In business informatics, different automated credit rating prediction (ACRP) models are used to offer to the investors more objective alternatives to the rating agencies. ACRP is defined as follows: quantitative information of rated bonds is used to predict the credit ratings of unrated bonds. Data extracting methods commonly used in the informatics community, are used to develop an ACRP model. The methods can be divided into unsupervised and supervised clustering [28, 61]. In cluster analysis, a dataset is divided into different clusters. A cluster is defined as a group of data points which shares the majority of characteristics [28]. The most common methods which are employed to develop ACRP models are: artificial neural network (ANN), support vector machines (SVM) and support vector domain description (SVDD) [7, 48, 40]. In brief, ANN is based on the human brain. In this logic, ANN utilizes different layers of processing units called neurons. Usually, three layers are used: input -, hidden - and output layer. Similar to the synapses in the human brain, the neurons of the different layers are also inter-connected. In a first step, a set of data points with known output is employed to determine the importance of these connections. Weights on the connections represent their importance. The weights specify which neurons have to be used and in which intensity to identify the right output values. After the optimal weights have been founded, new data points can be introduced into the network and their outputs are predicted.[24]. In contrast, SVM tries to divide the space, in which the data points lie, with the help of hyperplanes. After determining of the hyperplanes, new data points can classify into one of the existing groups due to their position according to the hyperplanes [31, 96]. SVDD was primary defined to describe one group of data points and consider all the others data points as outliers. The single group is defined by a hypersphere which encloses data points with similar characteristics. Afterwards, new data points are classified as outliers or members of the group depending on their place inside or outside the hypersphere. During the last years, SVDD has been extended to handle also multi-class problems [51, 85, 86]. A class is defined as a group of data points belonging together.

Credit rating prediction can be seen as a multi-class problem because bonds of one credit rating represent one class. In this way, different classes exist and the objective is to predict the correct class for unrated bonds. In the literature, there are several ACRP models based on different methods [7, 48, 40]. These models are analyzed and their advantages as well as drawbacks are discussed. Furthermore, to overcome some existing drawbacks of ACRP models, a new ACRP model hs been develop during this PhD work [33]. The new model is

based on a combination of the SVDD model and linear regression (LR). In a first step, the bonds are divided into rating groups with the help of SVDD. A rating group is a merging of credit ratings and afterwards for each rating group, the credit ratings are predicted by using LR.

Until the beginning of this thesis, the performance of ACRP models has only been examined with the help of empirical testing [7, 48, 40]. Nevertheless, empirical testing always depends on the dataset that was used and an analytical guarantee can not be given. As the target group of the new ACRP model are private investors, it is important to indicate to the investors the performance of the new model in a worst-case situation. Therefore, according to [4, 54], the model is evaluated against an optimal benchmark under several assumptions. These assumptions include the number of information of bond attributes that are inserted into the model. The definitions of the different type of risk information used in the assumptions are given in [34]. This type of analysis is called competitive analysis. In competitive analysis, the difference between the optimal benchmark and the analyzed ACRP model is determined. Afterwards, an empirical analysis to determine the average performance of the new model is also undertaken. For this reason, the performance of the new model is compared against the performance of an ACRP model based on ANN. Finally, a prototype of the new model is implemented and integrated into the private financial planning tool LifeCharts to allow private investors to plan their financial situations and to offer them the possibility to undertake a risk analysis of bonds in case an investment is considered. LifeCharts has been developed in the department of my hosting university and a complete description can be found in the book [78]. The prototype is designed in such a way that private investors without any knowledge concerning risk analysis or SVDD and LR can handle the tool. Furthermore a simple portfolio selection tool is added to help the investors to take a real investment decision.

1.2   Motivation and Research questions

Financial bonds are essentially divided into corporate and sovereign bonds. Bonds issued by companies are called corporate bonds. In the same logic, sovereign bonds are issued by states [69]. At the moment, the majority of existing ACRP models is limited to corporate bonds. The reason for this limitation is quite simple. First, the market of corporate bonds is more important than the market of sovereign bonds [69]. Additionally, a company undertaking a risk analysis of the bonds of its competitors, can obtain an assessment of their financial health and in this way, it can obtain a competitive advantage on the market. Second, the needed information to undertake a risk analysis are obtained directly from the companies

if their balance sheets are analyzed. States do not publish balance sheets and therefore, the acquisition of the necessary information is easier in the case of corporate bonds [69]. Nevertheless, investors want to have the possibility to undertake a risk analysis of sovereign and corporate bonds because they do not want to be limited in their investments [49]. In this way, this PhD work focuses on a new ACRP model which handles the two different types of bonds simultaneously. New attributes have to be chosen because the balance sheets can not be used anymore. Nevertheless, the needed information has to be publicly available because private investors will not be willing to purchase attributes describing the risk of bonds from external experts like Bloomberg. For this reason, the needed attributes are computed from the bonds' features retrieved from the financial market, like historical prices. The used bond features are described in the course of this thesis. Thus, the first two research questions can be formulated:

Question 1: Can an ACRP model be developed using only public information?

Question 2: Can an ACRP model be developed which handles sovereign and corporate bonds?

The usability of ACRP models for private investors should always be one main criteria of development. At the moment, ACRP models have been developed in view to demonstrate the feasibility to generate alternatives to the three main rating agencies. However, the focus in the development of ACRP models has never been on private investors. With the investors into the focus, the models should help them to make decisions. The use of an ACRP model has to be comprehensible for investors and the declarations concerning the risk of the bonds have to be easily understandable just like the credit ratings used by the agencies. The investors have to immediately understand which bonds contain a high or a low risk of default. Furthermore, an ACRP model which only illustrates the riskiness of the bonds to the investors, can not be seen as a real decision help. As previously explained, private investors do not make a complete analysis of the bonds before taking the investment decision [19]. Therefore, to become a real decision help, an ACRP model has to help investors to select the right bonds in which they should invest. For this reason, a user-friendly ACRP model should also propose the optimal portfolio of bonds to the investors. Thus, the final research question of this thesis has to be:

Question 3: How does an ACRP model have to be designed to be a decision help for private investors?

1.3   Structure of the thesis

The chapters of this thesis are based on each other and it is recommended to read them in numerical order.

The necessary requirements for the practical usability of ACRP models are discussed in Chapter 2.

Chapter 3 sets up the necessary financial framework. The definitions of bonds and credit ratings as well as the description of risk analysis are given in this chapter.

Chapter 4 offers a general overview of unsupervised cluster analysis, also called classification, and especially, artificial neural network, support vector machines and support vector domain description, are described in detail.

Question 1 is answered in Chapter 5. In this chapter, existing ACRP models are presented. Additionally, their advantages and drawbacks are discussed.

Chapter 6 introduces a new ACRP model and in this way, Question 2 is answered. Additionally, different fields of application beside of the financial risk analysis are introduced in this chapter.

Furthermore, in Chapter 7 the performance of the new ACRP model is investigated by undertaking a competitive and an empirical analysis. Thus, Chapter 7 deals with the theoretical and experimental foundations of the answer of Question 2.

Question 3 is answered in Chapter 8. A prototype of the new developed ACRP model is implemented and integrated into the private financial planning tool LifeCharts. Furthermore, in the final prototype, a portfolio selection tool is added to the ACRP model in such a way that the information obtained from the risk analysis can be directly utilized to identify the optimal portfolio of bonds.

inally, the last chapter sums up the findings of this thesis and proposes several points for future research.

Chapter 2

Requirements for the practical usability of ACRP models

This chapter deals with the requirements for the practical usability of automated credit rating prediction (ACRP) models by private investors. For this reason, personal financial planning is introduced. An example of an user employing a personal financial planning tool to manage his future is discussed. Finally, an ideal ACRP model is formulated.

2.1 Preliminaries

In times of financial distress, private persons become more and more attentive to plan their financial situation. Nowadays, with the different crisis all around the globe, especially the financial crisis, the countries are forced to reduce their expenses. These reductions are necessary to be able to prevent a financial collapse due to excessive debts. Countries often decrease the pension claims to economize money. This reduction has direct consequences for each private person. Therefore, each private person underlies the risk to fall into the poverty among the elderly. At this point, personal financial planning tries to help persons to identify the necessary measures which they should undertake to avoid this risk [50, 41].

To support private persons in their financial precaution, there exists three main types of personal financial planning (PFP) tools: (1) Financial overview, (2) Investment plan and (3) Point of financial freedom. Financial overview tools collect the information about the actual financial situation of the users to offer them an overview of their expenses and incomes [50]. This overview allows the users to determine their savings capacity. However, these tools do not propose the users any future investments. Investment plan tools are more sophisticated.

After the analyzing of the actual financial situation of the users, these tools offers an advice of investments to obtain capital gains according to the risk tolerance level of the users [50, 41]. One main problem in this type of PFP tools is the determination of the correct risk tolerance level. Even if several standardized question can be asked, the determination of the correct risk tolerance level can never be guaranteed [41]. Finally, the last type of PFP tools determines the actual financial situation of the users and asks them to estimate their expected future incomes and outcomes to identify their point of financial freedom. The point of financial freedom is the point in time at which the users are able to manage their life without any additional incomes from work until a given expected end of life [78]. Furthermore, in case of a gap between the needed consumption capitals and the actual savings, these tools gives some advices which measures the users should undertake to increase their savings and to attempt their point of financial freedom.

Excluded the basic overview tools, the two other types of PFP tools gives an advice of investments to the users. Nevertheless, they only distinguish between risky and non-risky assets. Risky assets are defined as assets that do not provides a guaranteed return. In contrast, non-risky assets provide a guaranteed return [69]. As this work focuses on bonds, bond investment in case of personal finance is investigated. Even if bonds are often considered as non-risky assets because of their guaranteed periodic returns, bonds incorporate many different risk. The detailed definition of bonds and their risks are given in Chapter 3. Therefore, the risk of bonds has to be analyzed to offer the users an accurate investment advice. The lack of such a complete analysis of the risk of the assets in PFP tools is due to maintain the simplicity of the tools [50, 41]. To overcome this drawbacks, ACRP models can be used. The ACRP models allows to evaluate the credit risk of bonds. The credit risk is defined as being the risk that the bonds defaults and the investors do not receive anymore the guaranteed payments [69]. A detailed definition of the credit risk is given in Chapter 3. Nevertheless, the existing ACRP models are not really adapted for the practical use by private investors. The focus of their development is the demonstration of the feasibility of alternatives to the rating agencies, like Moody's, S&P and Fitch, which are usually responsible to evaluate the credit risk.

The integration of ACRP models into PFP tools helps to distinguish the assets according to their risk and to offer the users a more accurate investment advice. Without corrupting the usability of PFP tools, the needed modifications of ACRP models to be usable by private persons are investigated in this chapter. Therefore, the following research question is answered: *Which requirements has to fulfill an ideal ACRP model to constitute a benefit to PFP tools?*

These requirements help to develop a prototype of an ACRP model integrated into a PFP tool. The development of the prototype represents the proof of concept of the stated requirements.

Briefly the structure of the thesis is described in view to develop an operable prototype of an integrated ACRP model which would fulfill the stated requirements. After setting the financial framework needed in this work and describing the different methods used by existing ACRP models (Chapter 3 and Chapter 4), the existing ACRP models are specified in detail and their main drawbacks are stated in Chapter 5. In Chapter 6, a new developed ACRP model is presented to fulfill the requirement that an ACRP model should be able to handle simultaneously sovereign and corporate bonds. Afterwards, the theoretical and empirical performance of the new developed ACRP model is investigated in Chapter 7. Finally, the new ACRP model is simplified by eliminating the user-given variables and a portfolio selection tool is integrated to the ACRP model to fulfill the other requirements stated for the ideal ACRP model. Then, the obtained ACRP model is integrated into the PFP tool, LifeCharts, to demonstrate that such an integration yields in benefits and offers users a suggestion to restructure their wealth. A proof of concept is undertaken in Chapter 8.

This chapter is structured as follows. The next section briefly introduces the PFP tool LifeCharts which belongs to the third type of PFP tools, point of financial freedom. An exemplary scenario based on LifeCharts is discussed in Section 3. This scenario helps to show the necessity of ACRP models in PFP tools. Section 4 states the requirements for an ideal ACRP model. Finally, the chapter is summarized in the last section.

2.2   LifeCharts

LifeCharts is developed by the Operations Research and Business Informatics department of the Saarland University by the team of Professor Günter Schmidt. The focus of LifeCharts is to help private persons to get an overview of their financial situation and to plan it. LifeCharts is priority employed to identify and control the risks of retirement through financial planning. The financial aspects of life are represented graphically with the help of two lines to simplify the understanding of the specific situation. The first line starts with the current age and represents all the money that is needed for the future. It is referred to as the required capital savings. At this point, it has to be mentioned that private persons have to indicate an end date which corresponds with the probably estimated end of life. Additionally, they have also to specify their expected incomes and outcomes over the different periods. For example, they can divide their life into two main periods: (1) job and (2) retirement. The second line shows

the existing saving capital at any time of life. The point of financial freedom (FF-point) is defined as the intersection of the two lines. At this point, the private persons have saved enough money to finance the rest of their life without any additional income from work. In the subsequent figure, the two lifelines and the FF-point are represented. The figure is labeled in German because LifeCharts is developed in this language.
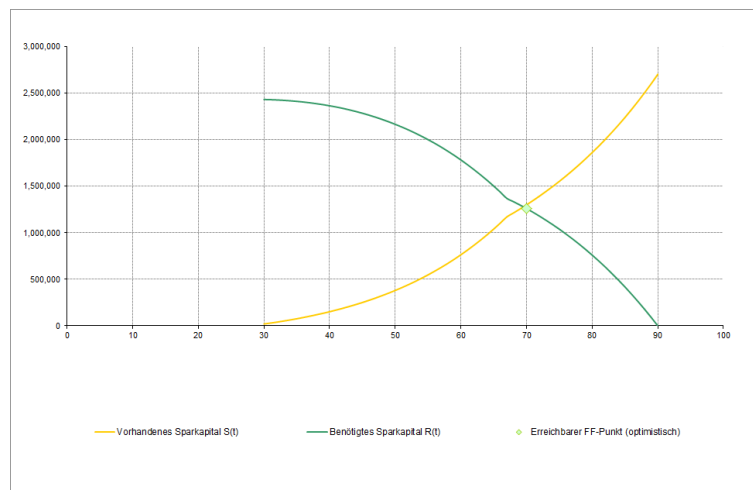
Fig. 2.1 Representation of two financial lifelines and the FF-points taken from LifeCharts

As the persons indicate the expected incomes and outcomes, LifeCharts also takes their current assets and liabilities as well as their revenues and expenses into consideration to predict the future development of their financial properties. In a further step, the personal goal, e.g., the own desired time of financial freedom, can be defined. The required savings capital to achieved this favored FF-point is determined. LifeCharts does not only indicates the possible current gap between the required and the existing savings capital but it also proposes different measures which help to achieve this goal. The theory behind LifeCharts as well as a more detailed description of the tool are given in [78]. Among the different measures suggested by LifeCharts to improve the financial situation, the investment of a portion of the saving capital into several assets is proposed. At this stage, the integration of an ACRP model could help to offer an investment advice. The necessity of such an integration is shown by an example in the next section.

2.3   Exemplary scenario

The exemplary user, called John Smith, is born in 1986 and wants to identify the point in time at which he can manage his life without any additional incomes from work. He expects that he would achieve the age of at least 90 years. To obtain an overlook of his financial

situation, he utilizes the PFP tool, LifeCharts.

First, John has to enter his actual wealth and his actual possible debts into the tool. Concerning the liquid assets, John has the following partition:

- Cash: 4850 €

- Shares: 3070 €

- Other assets: 5400 €

John does not have any real property, but John has invested into an endowment insurance. At this moment, John has accessed to 7283 € from this insurance. John has not any debts, thus he has fully indicated his wealth without any possible actual incomes and outcomes. LifeCharts collects this information of John as represented in the following figures B.1 and B.2 in the appendix.

Second, the actual and expected incomes and outcomes have to be specified. At the time John utilizes LifeCharts, he has a job with a permanent contract. For this reason, he is able to indicate his incomes and outcomes until his retirement at the age of 63 years with certainty.

- Salary: 57713 € / year

- Taxes and dues: 11625 € / year

- Social security: 11754 € / year

- Consumption of daily life: 23300 € / year

- Insurance contribution: 2090 € / year

This first life period is represented in Figure B.3 in the appendix. Afterwards, from the year of his retirement to his expected end time, John estimates his future incomes and outcomes for each year as follows:

- Capital assets: 2088 € / year

- Other incomes: 23069 € / year

- Taxes and dues: 3592 € / year

- Social security: 2667 € / year

- Consumption of daily life: 23300 € / year

- Insurance contribution: 1330 € / year

Figure B.4 in the appendix illustrates the second life period of John. The given listings by John are exemplary for the majority of private persons. In the western society, after the academic path, the majority of the persons searches to find a job with a permanent contract. Thus, the career of John Smith with one life period consisting of the job and one life period consisting of the retirement is plausible. Additionally, each other person can find similar incomes and outcomes in its own life. Under "other incomes", the pension of John Smith can be found.

Afterwards, LifeCharts gives John an overview of his situation until his given end time. The given overview is shown in Figure B.5 in the appendix. The used formulas to compute the existing savings and the needed consuming capital can be found in [78] and are not reproduced in this work. The following savings and consuming capital for John Smith are determined:

- Savings: 20603 €

- Consuming capital: 4293499 €

Note that the given values are valid for the start of the planning. This means that at the beginning of the planning, John has 20603 € saved and that he would need in total 4293499 € to finance his remaining life. Furthermore, John sees that at the moment, he has not enough money to manage his life without any additional incomes from work and that this gap is actually equal to −4272896 €. LifeCharts also indicates that John achieves his FF-point at 73 years. Due to his estimated end of life corresponding to 90 years, the obtained FF-point is not favorable for John.

In the next step, John has the possibility to indicate his desired FF-point. John indicates that he wants to be financially free at 69 years. He specifies the portion which he is willing to invest into risky assets. John has also to estimate the average yearly return of the risky and non-risky assets. Figure B.6 in the appendix represents how LifeCharts assembles this new information from John. LifeCharts uses this additional information to determine a new FF-point under the consideration that John undertakes a restructuring of his wealth. This restructuring takes into account the given portions of risky and non-risky assets. Furthermore, the invested assets have to achieved the estimated average return to ensure that the obtained results remain correct. LifeCharts has determined that after John has restructured his wealth, he can achieved his FF-point at 72 years. The obtained results from the new computations of

the FF-point are illustrated in Figure B.7 in the appendix.

Nevertheless, John wants to achieve his FF-point at 69 years. Even with the suggested restructuring of his wealth, this goal is not achieved. Therefore, John has to increase his savings. LifeCharts determines the required savings and identifies the gap between the required and the existing savings. To achieve his goal, John has to make the effort to save additionally 1583 € per year. The overview of his actual situation and the proposed measures to achieve the desired FF-point are represented in Figure B.8 in the appendix. If John can realize this additional saving then he will also achieve his desired FF-point. The amount of additional saving is the only information given by LifeCharts. It does not give John any suggestions how he can accomplish this additional saving. However, John probably requires some advices as he is not an expert in finance. At this stage, ACRP models can help to overcome this common drawback of PFP tools.

Instead of leaving John along to find the correct measure to save enough money. LifeCharts by the bias of an ACRP model could suggest to John several bonds and offer him the opportunity to enter own bonds in which he could invest. Afterwards, the bonds are analyzed by the ACRP model and the obtained risk evaluation of the bonds is shown to John. He has an actual risk evaluation and can take a better decision in which bonds an investment is lucrative as without any information about their risk [76, 72]. The answer to the question of how to proportionate his wealth into the different bonds is still leave to John. The integrated ACRP model should be able to determine a portfolio out of the available bonds. The comprehensibility of LifeCharts or any other PFP tool do not have to be complicated. The requirements for an ideal ACRP model which is used by private persons, like the exemplary John Smith, are stated in the next section.

2.4   Requirements for an ideal ACRP model

Remember that the existing ACRP models have several drawbacks which prevent developers to integrate them directly into PFP tools (see Section 1). First, ACRP models are developed with a scientific focus to demonstrate the feasibility of alternatives to the rating agencies, like Moody's or Fitch. The practical usability by private investors is not considered [7, 48, 40]. Additionally, the existing ACRP models are limited to one specific type of bonds: sovereign or corporate bonds. However, private investors who are using a PFP tool does not want to be limited to one type of bonds to undertake their investment. Therefore, it is important that the

ideal ACRP model handles sovereign and corporate bonds simultaneously.

Bonds are not only divided into sovereign and corporate bonds but each bond guarantees a periodic payment, called coupon rate, to the investors. The coupon rate can be variable or fix. The difference between fixed and variable coupon rates is explained in detail in the next chapter. The existing ACRP models usually focus on bonds with a fixed coupon rate. Private persons get knowledge about different bonds from the news and use this acquired knowledge in their financial planning [72]. However, they do not differentiate about bonds with fixed or variable coupon rates. Therefore, the ideal ACRP model handles fixed and variable coupon rate bonds such that the users have no restriction by employing their PFP tool.

During the calibration of an ACRP model, a set of rated bonds is used. A detailed description of the calibration of ACRP models is given in Chapter 5. The use of a set of rated bonds limits the ACRP models in their accuracy to predict the credit ratings of bonds with any possible maturity. The maturity of a bond refers to a finite time period at the end of which the bond will cease to exist [69]. In [39], the problem is presented that a calibrated ACRP model can only predict accurately the credit ratings of bonds with maturities which only differ by one to two years from the maturities of the bonds in the set used for the calibration. Private persons do not want to verify if the maturities of their selected bonds corresponds with those of the bonds used by the ACRP models. The ideal ACRP model allows private persons to utilize bonds with any valid maturities. This is important to maintain the known usability of PFP tools by the users [50, 41].

Another problem for the practical use of ACRP models represents the user-given input variables. Each ACRP model is based on a specific classification method, whether it is artificial neural network or support vector machines or support vector domain description [7, 48, 40]. The listed classification methods are explained in detail in Chapter 4. Briefly, for each method, several parameters are set by the user to increase the accuracy of the ACRP model. Private persons can not have full knowledge about each possible method used in an ACRP model to fix the correct input variables. Until now, ACRP models are used by scientists, especially by their own developers. To maintain the comprehensibility of the PFP tools, the integrated ACRP tools have to reduce the number of user-given input variables and to help the users to set the correct values for the remaining input variables. The ideal ACRP model eliminates all user-given input variables and automatically determines the correct values of the needed parameters in a pre-processing step before the risk evaluation is undertaken.

In the case that an ACRP model is integrated into a PFP tool, the representation of the obtained risk evaluation to the users is not sufficient to help them to restructure their wealth. Private persons need an advice in which bonds they should invest. Therefore, the integrated ACRP model should propose the users to construct a portfolio out of the available bonds. In this way, the model asks the users to indicate the maximal risk, which they are willing to take. With this information, the optimal portfolio with maximal return is determined. In the other way, the users indicate the minimal return, which they want to obtain from their investment, and then the optimal portfolio with minimal risk is constructed [65]. In this case, the PFP tool is able to give the users a real investment advice to help them to restructure their wealth and to implement the elaborated financial plan. Hence, the ideal ACRP model also proposes a portfolio selection method to process the obtained risk evaluation and to offer users a real help for an investment decision.

Briefly summarizing the identified requirements which an ideal ACRP model has to fulfill to constitute a benefit to PFP tools:

1. Simultaneous handling of sovereign and corporate bonds

2. Treatment of bonds with coupon rates of different types

3. Handling of bonds with different maturities

4. Elimination of the user-given input variables

5. Integration of a portfolio selection tool

With the ideal ACRP model integrated into a PFP tool, the PFP tool is not only able to offer users the planning of their financial situation but also helps to implement the financial plan by undertaking a bond investment. The proof of concept of the gain by integrating an ACRP model into a PFP tool is undertaken by the development of a prototype. The prototype will represent the implementation of an ACRP model such that it can be used by private users without any previous knowledge of the employed methods.

2.5   Conclusion

First, the importance of personal financial planning is briefly discussed and three different types of PFP tools are identified: (1) Financial overview, (2) Investment plan and (3) Point of financial freedom. PFP tools support private persons to plan their financial future. The main drawback of the majority of the actual PFP tools is that they only undertake a distinction

between risky and non-risky assets. In reality the distinction of the assets according to their risk is more complicated than using only two categories: risky and non-risky. To overcome this drawback ACRP models represents a possible solution.

Afterwards, the PFP tool, LifeCharts, is introduced in more detail and an exemplary scenario is given. In this scenario, an exemplary user of LifeCharts is discussed. In case that a gap between the actual savings and the needed consumption capitals exits, LifeCharts suggests several measures how the user could increase its savings. However, it does not indicate how the user could realize these measures. The evaluation of risk of bonds and the realization of a bond investment can be adopted by an ACRP model. The integration of an ACRP model into a PFP tool increases the expressiveness of the tool. In this case, the PFP tool can make a concrete advice for a bond investment to the user to increase its savings.

Due to the fact that existing ACRP models are developed with the focus to demonstrate the feasibility of alternatives to the rating agencies, they are not appropriated for the practical use by private persons. To constitute a benefit for PFP tools, several requirements are identified which an ideal ACRP model has to fulfill. The requirements guarantee that the PFP tools with the integrated ACRP model maintains its usability. The ideal ACRP model has to handle simultaneously different types of bonds. Additionally, the user-given input-variables defined by the underlying method used by the ACRP model has to be reduced or completely eliminated. Private persons do not have the necessary knowledge about the used methods to determine the correct values of the input variables. Finally, a portfolio selection tool should be integrated into the ACRP model to process the obtained risk evaluation of the bonds and to suggest the users a concrete investment into the available bonds. Concluding, the stated research question is completely answered in this chapter.

Chapter 3

Financial Bonds, credit ratings and information

In this chapter, financial bonds are defined and their importance in the financial world is shown. For each financial product, the risk analysis is an important step before undertaking an investment decision. In this way, the assessment of their obtained credit ratings is appropriated. The meaning of credit ratings and their main drawbacks are discussed. Finally, an overview of the current credit risk analysis is given. This chapter sets the financial theory needed in this thesis.

3.1 Description of bonds and their importance

3.1.1 Financial bonds: Definition and basic features

Out of the different financial products which are handled on the stock market, shares and bonds are important and widely-used products [69]. Shares offer investors the ownership rights of the underlying company. In this way, the owner of the shares does not only profit from the company's gain but has also to endure possible losses. In contrast, a bond is defined in the following way [69]:

Definition 1. *A bond, denoted B, is a financial instrument that allows governments and companies to borrow money. B represents a kind of contract which fixes the obligations, like the payments, of the issuer (government or company) to the investor.*

To clarify the difference between a bond and a normal bank credit, one has to know that a bond always has a special form which is due to the regulations that have to be fulfilled. Furthermore, in contrast to a bank credit, a bond is certified and as a certificated credit it can

be quoted on the market and investors can trade with it. The whole process before the bond is freely tradable on the market is called: issuing [69].

The possibility to issue bonds allows governments or companies to collect the necessary money from investors to realize their projects. Additionally, as the bondholder has a prior claim on the company's earnings and assets compared to common shareholders, a bond investment normally includes less risk than an investment into shares [69].

All bonds are characterized by the following basic features:

Definition 2. *A bond, B, is characterized by the subsequent basic features:*

- *Issuer*

- *Maturity*

- *Par Value*

- *Coupon Rate and Frequency*

- *Currency Denomination*

The **issuer** is the underlying entity which issues the bonds [69]. This entity is very heterogeneous and can include everyone, from individuals to companies and governments, too. According to the bond issuers, the following distinction can be made:

1. Supranational organizations, such as the World Bank or the European Investment Bank

2. Sovereign governments, such as the United States, Germany or Luxembourg

3. Non-sovereign governments, such as the state of Texas in the United States or the region of Madrid in Spain

4. Quasi-government entities, agencies that are owned or sponsored by governments such as the postal services or the rail transports in many countries

5. Companies, in this group of issuers the distinction between financial issuers, like banks, and non-financial issuers is often made

According to the bond issuer, a first risk for investors resulting from a bond investment can be defined.

Definition 3. *The risk of loss resulting from the issuer to refinance its debt is called credit risk. This means that the bond issuer is unable to pay the interest and/or to repay the principal.*

Maturity defines the fixed date when the issuer has to redeem the bond by paying the outstanding principal amount [69]. The tenor, also known as the term to maturity, is the remaining time until the bonds' maturity date. This period is the time span where the investors can expect to receive the coupon payments. Additionally the bondholder knows how long it takes until the investors can count on the payment of the principal. According to the maturity, the bonds can be divided into the following groups:

1. Money market securities: bonds with a maturity of one year or less

2. Capital market securities: the maturity of the bonds is longer than one year

3. Perpetual bonds: bonds with endless maturities

The **par value**, often also called principal value or principal, represents the amount that a bond issuer aggres to repay at maturity [69]. Generally, the price of a bond is always quoted as a percentage of its par value. For example, if the par value of bond B is 1000 € and B is quoted at 90%, then the price of B will be 900 €. According to the quotation of B, the following distinction is made:

- B is traded at par: B is quoted at 100%

- B is traded at premium: B is quoted above 100%

- B is traded at discount: B is quoted below 100%

The interest rate of a bond is called **coupon rate**. This is the agreed rate which the issuer agrees to pay the investor each year until maturity. The coupon is the annual interest payments and is determined by multiplying the coupon rate with the bond's par value [69]. The majority of the traded bonds has a fixed rate of interest. Such bonds are called plain-vanilla bonds [22, 56]. In this case, the coupon payments are unchangeable during the whole period until maturity. On the one hand, a fixed interest gives investors a kind of protection against uncertain payments. On the other hand, if during bond's life, the inflation increases, then the investors would loose money by taking this investment. To overcome this drawback, there are bonds with floating rates of interest. Such bonds are called floating-rate notes or floaters [69]. The coupon rate of a floater typically includes two components: a reference rate and a spread. The spread is generally constant and is expressed in basic points. One basic point equals 0.01%. The reference rate is reset periodically and if the reference rate changes, then the payments of coupon also change. Common reference rates are "London Interbank offered Rate (LIBOR)" or "Euro Interbank offered Rate (EURIBOR)" [71].
A third kind of bond is the zero-coupon bond which does not offer any periodic interest

payment [69, 82]. Instead, zero-coupon bonds are traded at discount, however at maturity date they are redeemed at par. The difference between the discount price and the par value represents the interest rate paid by the issuer. Nevertheless, the focus of this thesis lies on plain-vanilla bonds.

Currency denomination is the definition of the currency in which the bonds is issued [69]. Even if any currency is possible, the majority of the bonds are issued in Euro or US Dollar. The reason for choosing either Euro or US Dollar as the bonds' currency is to make the bonds more attractive. If the currency is not liquid and freely traded or the currency is very volatile relative to the major currencies on the market, then investors warned against investing in this currency because a conversion into their home currency is difficult and the risk of losing real value because of currency change is high [60]. A combination of foreign currency and local currency for the bond issuers are possible by issuing dual-currency bonds. In this case, the coupon payments are made in one currency and the par value is paid at maturity in another currency. Another possibility represents the currency option bonds. These bonds give investors the possibility to choose between two proposed currencies, in which currency they want to receive the coupon payments and the principal. Nevertheless, in this thesis only Euro-denominated bonds are investigated.

The bonds are not only characterized by their basic features but the indenture, the legal contract, which can vary from one bond to another, is also very important. The bond indenture as well as other considerations are described in brief in the next subsection.

3.1.2 Bond indenture and other considerations

Remembering Definition 1, a bond represents a legal contract between the bond issuer and the investors, the bondholders. This legal contract which describes the form of the bond, the obligation of the issuer and the rights of the bondholders is called **bond indenture**. The indenture is always written in the name of the issuer and contains the following features [8, 45, 69]:

1. principal value

2. coupon rate

3. dates of coupon payments

4. maturity date

5. any other important information describing the bond, like the funding sources

The funding source for the coupon payments and the principal payments has to be named to allow the investor to estimate the risk of insufficient funding. Additionally, the indenture also includes information about possible collaterals, credit enhancements and covenants. Assets or financial guarantees underlying the bond above and beyond the issuer's promise to pay are called **collaterals** [69]. To reduce the credit risk of the issuer, provisions are often used. This provisions are entitled **credit enhancements** [69]. There are internal and external credit enhancements. The most popular technique of internal credit enhancement is subordination which means that the claim priority of the bondholders are ordered. The cashflows generated by the assets are allocated according to different classes of seniority. The seniority defines in which order a bondholder has the right to receive the payments. The subordinated tranches function as credit protection for the more senior tranches. Thus, the more senior tranches receive first the payments and afterwards the subordinated tranches. However, in cases of a default, the subordinated tranches would not be paid first and only if the outstanding payments surpass the global amount of subordinate tranches, then the senior tranches will feel the default. The indenture can also specify the rights of the bondholders and the actions that the issuer is obligated to perform or prohibited from performing. These specifications are named **covenants** [64, 69, 17]. Covenants are divided into affirmative and negative ones. Affirmative covenants are typically of administrative nature. For example, an affirmative covenant is that the issuer has to maintain its current lines of business or to maintain its assets. No additionally costs are imposed to the issuers by such covenants. Furthermore, no constraint to manage the issuer's operating business is set. In contrast, negative covenants are often costly and do materially constrain the issuer's potential business decision. Common negative covenants are [69]:

1. Restriction on debt: Regulation of the maximal acceptable debt ratio

2. Negative pledges: Forbiddance to issue bonds which are senior to the actual bonds

3. Restriction on prior claim: Protection for unsecured bondholders

4. Restriction on distribution to shareholders: Restriction of dividends payments and repurchases

5. Restriction on asset disposals: Limitation on the amount of assets that can be disposed during the bond's life

6. Restriction on investments: Blocking risky investments

7. Restriction on mergers and acquisitions: Preventing these actions if the issuer is the underlaying company

The legal and regulatory requirements depend on where the bond is issued. These considerations have an influence on the investor to take the investment decision. A first division is made between domestic and foreign bonds. A domestic bond is issued in the country in which the issuing entity is registered. Whereas bonds issued by entities which are incorporated in another country are called foreign bonds. The Eurobond market was created to overcome the drawback of different regulations for domestic and foreign bonds and to provide investors with a greater offer of bonds [52, 69]. The Eurobond market bypasses the legal, regulatory and tax constraints imposed on bond issuers and investors. Bonds issued and traded on the Eruobond market are called Eurobonds and they are named after their denominated currency, like Eurodollar bonds for Eurobonds denominated in US Dollar. Generally, Eurobonds are less regulated than domestic and foreign bonds because they are issued outside the jurisdiction of one single country. Additionally, only the clearing system knows exactly who the bond owners are. In contrast, for domestic or foreign bonds, the ownership of the bonds is registered with a name and a serial number. Another type of bonds are global bonds which are issued on the Eurobonds market and at least one domestic bond market. The advantage of issuing bonds on different markets is to ensure a sufficient demand for large bond issues and to allow investors to purchase the bonds without any restriction of their location. The strongest effect on the bond price is the currency in which the bond is denominated. The reason is that the market interest rates which influence the bond price are those associated with the currency in which the bond is denominated [69].

Tax considerations are always important for a bond investment decision because the income portion of a bond investment is taxed at the ordinary income tax rate [37, 69]. Additionally, a bond investment can also generate a capital gain or loss if the bond is sold before its maturity. In this case, taxes different from the tax applicable on the interests have to be paid. Naturally, the tax rates are different from one country to another. Nevertheless, investors have in mind the possible taxes which have to be paid. However, there are some bonds which are exempt from taxes. In the United States, the different states often issue tax-exempt municipal bonds. Then, these bonds are exempt from taxes but only in the States which have issued them [69]. Next, a general overview of the bond market is given. First, the different classifications which can be made are shown and before defining the primary and secondary bond markets. Finally, the process of these two markets is explained.

3.1.3   Bond market

The bond market can be classified by different criteria which are common for each bond [69]. The market can be classified by the type of issuer which usually leads to the identification of three bond market sectors. The government and government-related sector, the corporate

sector and the structured finance sector are identified. Supranational organizations, sovereign governments, non-sovereign (local) governments, entities owned or sponsored by governments are included in the government and government-related sector. The corporate sector contains all financial and non-financial companies. The last sector is composed of bonds which are created from the process of securitization. In the process of securitization, private transactions between borrowers and lenders are transformed into securities (bonds) which are traded in the market. A second possible classification of the market can be made by using the credit quality of the bonds. Each bond investor is confronted with the risk of loss resulting from the issuer failing to make full and timely payments of interest and/or principal. This risk is called credit risk. Rating Agencies, like Moody's, described in Chapter 1, Section 1.1, estimate the credit risk of the different bonds and publish their results with the help of credit ratings. The credit ratings are explained in detail in the next section. According to their credit risk, the bonds can be divided into investment grade and non-investment grade bonds and in this logic, two sectors of credit risk can be identify on the bond market. The bonds' maturities can be used to divide the bonds into different sectors. The maturities range from overnight to several years. Furthermore, the currency denomination is used to distinguish different sectors in the bond market. For example, one sector contains all the bonds denominated in Euro and a second sector includes all the bonds signed in US Dollar. Finally, the bond market can also be divided with the help of the type of coupon. As previously seen, different types of coupon exist. The coupon rate can be fixed or variable. For the variable coupon rates, all the bonds with the same reference rate can be regrouped together to determine different sectors. However, as in this thesis, the focus lies only on plain-vanilla bonds, fixed coupon rates are assumed. Naturally, other criteria, such as the home countries of the bonds' issuers, can be used to determine different sectors in the bond market.

For possible investors, the most important distinctions of the market is between primary and secondary bond market. In the primary market, the issuers sell their bonds for the first time ever to investors. In contrast, in the secondary bond market, the existing bonds are traded among different investors. A detailed definition about the first issue of bonds on the primary market can be found in the book [69]. In this thesis, the secondary market is only of interest as private investors can only intervene at this stage to buy or sell bonds. In the secondary market, bonds are directly bought or sold from investors to investors. For investors, a liquid market is important. If liquidity is guaranteed, then bonds are not overestimated or underestimated but their real value is the basis for their price [16, 69]. In this case, investors are able to sell or buy bonds in a very small time lapse. According to [16, 69], liquidity is measured by the difference between the bid and ask price. The smaller this difference is more liquid the market is. The bid price is the price which an investor wants to buy and the

ask price is the price which is asked by an investor to sell the bond. The bonds' liquidity always expresses a part of their credit risk [69]. Therefore, in this thesis, liquidity of the bonds is always analyzed.

3.2 Credit Ratings

3.2.1 Definitions of credit ratings

Rating agencies, like Moody's, S&P and Fitch, publish credit ratings for the issues themselves as well as for the issuer. According to S&P [83], which can be used as an example for the other rating agencies, an issue credit rating is defined as follows:

Definition 4. *An issue credit rating is a forward-looking opinion about the creditworthiness of an obligor with respect to a specific financial obligation, a specific class of financial obligations or a specific financial program. It also takes into account the creditworthiness of guarantors, insurers or other forms of credit enhancement on the obligation.*
In other words, the credit rating expresses the risk of loss resulting from the obligor, the issuer, to be unable to refinance the debt and in this way fail to pay the interests and/or the principals to the investors.

In the same way, the definition of the issuer credit rating is given in the following way:

Definition 5. *An issuer credit rating is a forward-looking opinion about an obligor's overall creditworthiness in order to pay its financial obligations. The opinion focuses on the obligor's capacity and willingness to meet its financial commitments as they come due.*

For the two definitions, it becomes clear that a credit rating is an opinion and that it is not an irrevocable fact. This concept has two main reasons. First, the rating agencies do not want to be responsible if one of their credit ratings is false and in consequence make an investor lose money from an investment. Second, a credit rating just an estimation of the real creditworthiness of an issuer and so a credit rating can never be an indicator describing the real risk, only the opinion about the possible risk of loss [43].
In the remainder of this thesis, issue credit ratings are regarded. Thus, when credit ratings are mentioned, issue credit ratings are meant. The rating agencies utilize an alphanumeric scale to express the credit ratings.

The subsequent figure shows the scale for the three main agencies, Moody's, S&P and Fitch:

Moody's		S&P		Fitch		Rating description
Long-term	Short-term	Long-term	Short-term	Long-term	Short-term	
Aaa	P-1	AAA	A-1+	AAA	F1+	Prime
Aa1		AA+		AA+		High grade
Aa2		AA		AA		
Aa3		AA−		AA−		
A1		A+	A-1	A+	F1	Upper medium grade
A2		A		A		
A3	P-2	A−	A-2	A−	F2	
Baa1		BBB+		BBB+		Lower medium grade
Baa2	P-3	BBB		BBB	F3	
Baa3		BBB−	A-3	BBB−		
Ba1	Not Prime	BB+	B	BB+	B	Non-investment grade speculative
Ba2		BB		BB		
Ba3		BB−		BB−		
B1		B+		B+		Highly speculative
B2		B		B		
B3		B−		B−		
Caa1		CCC+	C	CCC+	C	Substantial risks
Caa2		CCC		CCC		
Caa3		CCC−		CCC−		
Ca		CC		CC		Extremely speculative
		C		C		Default imminent
C		RD	D	DDD	D	In default
/		SD		DD		
/		D		D		

Fig. 3.1 Alphanumeric Scale used by Rating Agencies [69]

Figure 3.1 shows that credit ratings can be divided into two main groups: investment and non-investment grade. For example using the scale employed by Moody's, investment grade regroups all the credit rating in the range of $[Aaa, \ldots, Baa3]$ and all the credit ratings below $Ba1$ are summarized as non-investment grade.

Naturally, there are also other rating models different from those used by the rating agencies. Each major financial institution has an internal rating model to estimate the creditworthiness

of the issuers of the bonds as well as the bonds themselves. The scales used by the institutions also vary from those employed by the rating agencies. The number of ratings used by the financial institutions usually varies from 4 to 7 [88]. Nevertheless, as the credit ratings out of these internal models are private, our focus lies on the credit ratings published by the rating agencies which also are of great importance in the financial world.

3.2.2   Importance of the credit ratings

The importance of credit ratings yields from two main reasons. First, the investors have to obtain an overview of the credit risk of the bonds. As previously described, credit ratings express the estimation of the creditworthiness of the issuer and in this way, the probability that the issuer is willing to pay the interests and the principals. In other words, credit ratings communicate the possible risk of loss, i.e., the credit risk. The rating agencies publish the credit ratings with the help of the alphanumeric scale shown in figure 3.1. This simple notation allows investors to quickly obtain an overview of the credit risk and to select bonds which fulfill their requirements concerning the credit risk.

Second, each company or state aims to obtain a high credit rating because the interest rate which the issuer has to pay depends from the credit rating. with increasing interest rates, the raise of money to finance projects becomes more and more difficult for issuers [69]. Thus, credit ratings are important for the potential investors and the issuers. With higher credit ratings, lower interest rates have to be paid and projects become more affordable. Therefore, credit ratings have a global importance in the financial market. On the one hand, credit ratings help investor to take a first investment decision and on the other hand, they are one of the components which define the interest rates which are applied to the issuers.

3.3   Credit risk analysis

In this section, an introduction to the credit analysis is given. Additionally, the main models of credit analysis are introduced.

3.3.1   Introduction to Credit Analysis

The risk of loss resulting from the issuer to be unable to refund the debt is called credit risk. However, in reality, credit risk has two components: default risk and loss severity. The probability that an issuer fails to make full and timely payments of interest and principal is known as default risk. Nevertheless, in the event of a default, a portion of the bond's value

which is lost for investors is named loss severity. Normally, loss severity is expressed with the help of the recovery rate. The recovery rate is the percentage of the principal amount recovered in the event of a default [69]. In a first phase, investors focus on reducing the default risk and only if the default risk is high, their focus changes to the recovery rate in the second phase. There are some important risks which are directly related to the credit risk [69]. These risks are described as follows:

1. Spread risk: Bonds with higher default risk have to pay a higher interest compared to bonds which are considered as risk-free, like German government bonds or US Treasury bonds. This difference between the interest rates is called spread.

2. Credit migration risk: This is the risk when the issuer's creditworthiness declines and in this way, investors believe that the default risk of the issuer is increasing.

3. Market liquidity risk: If the investors can not transact a bond at the price which is quoted on the market then the market for this bond is illiquid. The bid-ask spread is a good estimator for the market liquidity risk. If the bid-price differs a lot from the ask-rice then a higher spread compared to risk-free bonds is asked from the investors to compensate the illiquid market situation of the bond.

Next, the capital structure of a company is defined. The capital structure is important to analyze the credit risk of the company. The capital structure is the composition and distribution across operating units of a company debt and equity [69]. This includes not only bonds of all seniority ranks but also bank debt, stocks and common equity. Seniority ranking defines the priority of investors to claim the payments from the issuer in case of default. Mainly, the distinction is made between secured and unsecured debt. Bondholders of secured bonds have a direct claim on certain assets and their associated cash flows. Only a general claim on an issuer's assets and cash flow is given to unsecured bondholders [69]. An example of seniority ranking is given in Figure 3.2. The first lien loan is the highest ranked debt in term of priority of repayment. A pledge of certain assets, which can include buildings, equipments, licenses, patents and so on, is referred to the first lien loan. Second and even third lien bonds are possible. Nevertheless, these bonds always rank below the first lien loans. After the secured bonds, the unsecured bonds are situated in the ranking. The highest ranking for unsecured bonds is senior unsecured and the lowest rank is junior subordinate unsecured. The majority of bonds traded on the financial market is rank senior unsecured. The question, that arises, is the following: "Why would investors or issuers buy or sell unsecured subordinated bonds if the probability of repayments is very low in case of a default?"
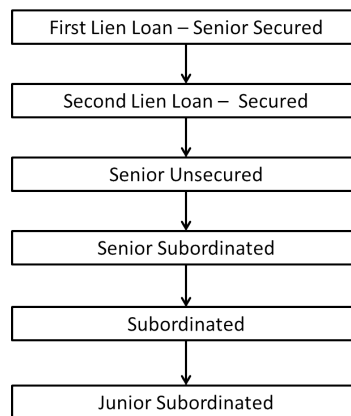
Fig. 3.2 Seniority ranking [69]

Companies issue subordinated bonds for two main reasons. First, issuing subordinated bonds is often seen to be less expensive than issuing equity and subordinated bonds because they do not modify the existing structure of shareholders. Second, subordinated bonds usually do not have as many restrictions as senior bonds. Under restriction, the covenants are meant. Subordinated bonds are bought by investors mainly for the following reason. Subordinated debt incorporates a higher credit risk then senior debt and higher credit risk results in higher returns. Thus, if investors believe that the interest rates paid by the issuers are high enough to compensate the higher risk, then they are also willing to invest into subordinated bonds [69]. Next, recovery rates are briefly discussed. Generally, there is a ranking called "pari passu" in right of payment concerning bonds. This means that all creditors at the same level of the capital structure are treated as one class [69]. Thus, investors of a senior unsecured bond with maturity of 30 years would have the same claims in the case of default as bondholders whose bonds mature in 6 months. In the majority of the cases, defaulted bonds continue to be traded by investors based on the recovery rate. Normally, the bonds still have some recovery value which results either from the liquidation of the defaulted company or from the reorganization of the company. The subsequent table indicates the average corporate recovery rates measured by Moody's.

Bond Type / Region	Europe	Global
Senior Secured Bond	43.99%	51.52%
Senior Unsecured Bond	32.40%	36.95%
Senior Subordinated Bond	36.95%	30.81%
Subordinated Bond	37.06%	30.81%
Junior Subordinated Bond	n.a.	24.71%

Table 3.1 Average corporate recovery rates, 1985-2012

Nevertheless, these recovery rates stated in table 3.1 are not irrevocable because some important points about recovery rates have to be considered [69]:

1. Recovery rates vary widely across industries

2. Recovery rates also vary depending on the credit cycle: If the economy is in a recession then the rates are lower than if the economy is in a boom.

3. Recovery rates are only averages

In reality, the priority of claims is even not absolute. In case of a default, all the different claimants come together. The plan of reorganization is either confirmed by a judge's order or by a vote of the claimants. However, there are usually disputes over the value of the various assets. Due to time constraints, the different claimants normally try to find an agreement, a compromise to accelerate the disbursement of the recovery value of their holding debts. This compromise includes that the priority of claims is modified such that each claimant can agree [69].

The assessment of an issuer's ability to satisfy its debt obligations is the goal of credit analysis. All types of bonds, sovereign or corporate, represent a contract which maintains all important points, like interest rate, frequency and timing of the payments, maturity date and the covenants. Instead of estimating the issuer's willingness to pay, the focus of credit analysis usually lies on the determination of the issuer's ability to pay. The issuer's ability to pay can be identified more accurately than its willingness to pay. The source, as well as, the predictability of the cash flows generated by an issuer to service its debt obligations represent the main consideration in traditional credit analysis [69].

In the next subsection, the main models used in the financial world are introduced. First, two traditional credit rsik models, credit scoring and credit ratings, are briefly described. Afterwards, the structural models are briefly discussed.

3.3.2   Credit risk models

Credit scoring only ranks the issuer's credit riskiness without proving any estimation of its default probability. Credit scoring is usually used to determine the riskiness of small companies or even individuals. The obtained credit scores are of ordinal character. In this way, it can not be distinguished if one issuer is twice better than another one. The following factors play an important role in the determination of the credit score of an issuer [69]:

- Payment history: the question, if one has paid ones bills on time, is important to determine if the issuer is reliable

- Evaluation of the comparison of the amount of debt with the issuer's credit limits

- Review of the credit track: the track record allows to determine if the issuer is considered as a solid creditor

- Determination of the amount of credit accounts: Too many credit amounts influence the credit score negatively

Nevertheless, credit scoring has some drawbacks. The ordinal character of credit scores is a first disadvantage because a real comparison of the credit risk of several issuers is not feasible [69]. Second, the credit scores do not depend on the current economic situation. If the economy is deteriorating then the credit score of an issuer is not automatically changed [69]. The credit score only varies if the financial situation of the issuer has changed. Third, as the majority of investors prefers stable credit scores, the analysts of credit scores put more attention on stability than on accuracy [69].

Credit ratings are usually employed to classify the credit risk of large companies and governments. Credit ratings are also of ordinal character. Apart from the rating agencies, like Moody's, described in the previous chapter, internal ratings are also defined and widely used in the financial world, especially by banks and other financial institutions. Without going into detail, the analysis to obtain credit ratings includes an extensive and detailed investigation about the financial situation, the prevision of the possible future cash flows and the credit history of the issuer [69, 101]. As for the case of credit scoring, the rating agencies attempt to keep their rating stable over time. The idea behind this is to reduce the volatility in the bonds' prices. Nevertheless, the conflict already discussed for credit scoring, between stability and accuracy, remains. The credit ratings are losing a bit of their accuracy if stability over time is aimed. Credit ratings are widely used because of two main reasons [69]. First, the ratings summarize a complex analysis of the issuers and allow to pass the information about the creditworthiness of the issuers in a simple way. Additionally, the stability over time of the ratings reduces the volatility of the bonds' prices and in this way, they allow to trade with prices where the focus is on offer and demand. However, credit ratings also have some weaknesses [69]:

1. The stability over time reduces the correspondence to the issuer's default probability

2. The explicit independence from business cycle; however, the default probability depends on the economic situation

3. The potential conflict of interest due to the policy that the issuer has to pay to obtain a credit rating, resulting in a distortion of the accuracy of the ratings

Next, the **structural models** are briefly discussed. These models are called this way because they are based on the structure of a company's balance sheet. In this way, structural models are primarily used to estimate the creditworthiness of companies. The structure models build on the insights of option pricing theory. The idea can be explained in the following way: *"The investors lend the company K dollars and simultaneously sell the company an insurance policy for K dollars on the value of its assets. If the assets fall below K, the investors take the assets in exchange for their loan. This possibility creates the credit risk [69]."* Due to the fact that structural models are based on option pricing models, they utilize the same assumptions:

1. The market is frictionless and arbitrage free

2. The riskless rate of interest is constant over time

3. The assets' value has a lognormal distribution depending on time T, with mean μT and $\varsigma^2 T$

The first assumption states that the assets are traded without any transaction costs and free of any arbitrage. Additionally, as the market is frictionless, investors can buy and sell the assets at the quoted price without any delay. There is no bid-ask spread. The first assumption also guarantees that the assets' prices are observable over all the time horizon.

In structural models, interest rate risk does not exist. This is given by the second assumption, that the riskless interest rate does not change over the whole time period.

The third assumption fixes the evolution of the assets' value. The assets' value changes according to a lognormal distribution with mean $\mu\%$ per year and volatility $\varsigma^2\%$ per year. Thus, over the time period $[0, T]$, the expected return and volatility of the assets' value equals μT and $\varsigma^2 T$ respectively [69].

The stated assumptions are identical with those of the Black-Scholes option pricing method. In [84] all the details concerning the Black-Scholes pricing model can be found. The structural models use the Black-Scholes model to determine the credit risk. The Black-Scholes model allows to determine the debt of a company. The debt is composed of two components: the present value if default occurs and the value if the default does not occur. The determination of the present value is depending on a standard normal distribution with mean 0 and variable 1. Due to the fact that a probabilistic distribution is assumed, the different risk measures, like probability of default, expected loss and present value of the expected loss, can easily be defined and their formulations are given in [69, pp. 293ff]. The input data of the structural models have to be estimated [68, 69, 77] . An estimation is necessary because a company usually possesses more assets than only bonds and shares. There are two main methods to estimate the input data for models: historical estimation and implicit

estimation. Under historical estimation, one understands the determination of the necessary parameters, like the mean and variance of the prices, out of historical time-series observations of the assets' prices. In contrast, calibration is made in implicit estimation. The underlying option prices from the model are observed on the market and used to estimate the values of the model's parameters. In contrast to the first stated assumption, the prices of assets, like buildings or non-traded investment, are not observable and quoted on the market. Therefore, implicit estimation is only usable to estimate input parameters. Implicit estimation is a very complex procedure and this is also the reason why implicit estimation is only made by some commercial vendors, like Moody's KMV [68]. As implicit estimation is not used in this thesis, the interested reader is referred to [69] to get a detailed description. Beside of the popular useability of structural models, they also have several disadvantages [68, 69, 77]. A first drawback is the complexity of a typical company's balance sheet which has a complex liability structure. A traditional balance sheet includes many different assets and not only bonds and shares. The fixed riskless interest rate is the second main disadvantage. However, interest rate risk is relevant in the financial market. Therefore, to achieve a more realistic analysis, the riskless interest rate should be discarded. Usually, a company's loss distribution has a fat left tail. However, the assumed lognormal distribution implies a thin tail for the company's loss distribution. Thus, the underlying probabilistic distribution is the third big disadvantage. Another drawback is the assumption that the volatility of the assets' prices are constant over time. This means that the evolution of the assets' prices is independent from economic conditions and business cycles. However, prices always evolve with the financial situation of the company and the general business cycle of the economy [59]. The last and the most important drawback of structural models is the assumption about frictionless markets because a large part of a company's assets is not traded in liquid market. The prices of buildings owned and several investments undertaken by companies can not be observed at each point in time [69].

The structural models are used for the following main advantages. Due to the use of option pricing models, the default probability and recovery rate of a company is defined in the same analogy and facilitates the understanding. The possibility to only use current market prices to estimate the necessary input data of the model remains its main advantage [69]. Using current prices has the advantage that the real assets' values are the basis of the model estimation to determine the credit risk [68, 69, 77].

Another type of model to determine the credit risk are the so-called **reduced form models** [68]. These models have been developed to overcome some of the previously mentioned disadvantages of the structural models [69]. The name of these models is based on the fact that they impose their assumptions on the outputs of a structural model, i.e., the probability of default and the loss given default, rather than on the balance sheet structure. This change of perspective allows to include actual market conditions. The assumptions of the reduced form models are as follows [69]:

1. The zero-coupon bonds of a company are traded in a frictionless and arbitrage free market

2. The riskless interest rate is stochastic

3. The state of the economy is describable by a vector of stochastic variables which represent the macroeconomic factors influencing the economy at each point in time

4. The company's default probability for a given time point and given state of economy can be represented by a probabilistic distribution

5. The company's default represents idiosyncratic risk if the vector of macroeconomic state variables are given

6. The percentage loss on the company's debt depends on the economic state variable at the time point at which default occurs

The first assumption is refuted in contrast to the one given for structural models. In reduced form models, only zero-coupon bonds of a company are assumed to be traded on frictionless markets which are arbitrage free. This assumption is more realistic because the majority of the company's liabilities are not traded in frictionless markets or are not traded at all.

The second assumption implies that in reduced form models, interest rate risk is captured. The stochastic mapping of the riskless interest rate enables to connect the interest rate risk to the bonds' prices. In reality, the change of the riskless interest rate has an direct impact on the prices of bonds [9]. For a given bond, if the riskless interest rate is decreasing then the price of the bond is increasing and vice versa.

The third assumption is not very restrictive because the vector of macroeconomic state variables can evolve arbitrarily. The general state of the economy is described by this vector of macroeconomic variables. Unemployment rate, inflation rate and growth rate of the gross domestic product are examples of used macroeconomic variables. This assumption allows to integrate, in a realistic manner, the current state of the economy into the determination of the credit risk.

The time at which a company defaults can be modeled as a Cox process with a default intensity depending on the current state of economy. The Cox process is a generalization of a Poison process and all the details are given in [20]. This modeling of the default time is given in the fourth assumption. The probabilistic distribution defines the probability of default of a company as follows: If the company has not yet defaulted then the default intensity gives the probability of default for the next time point. The main advantage of the fourth assumption is that the probability of default depends on the state of the economy. This allows the probability of default to increase in case of a recession and decrease in case of a healthy economy [69].

The fifth assumption defines the component of the probability of default which does not depend on the state of the economy. The increase of the probability also depends on the action taken by the company's management and the financial situation of the company. This fact is taken into account by the fifth assumption that for a given state of the economy, the default is an idiosyncratic risk.

The last assumption is again not very restrictive. This assumption only imposes that the recovery rate is a percentage of the bonds' face value and that this percentage depends on the current state of economy when a default occurs. This means that investors can expect less recovery from defaulted bonds in a recession than in a healthy economy.

As reduced form models are an improved version of structural models, the valuation of the credit risk is based on the option pricing methodology. In the same logic, the company's debt is composed by two components: the value of debt if default has occurred and the value of debt if default has not yet occurred. The prices are described in an abstract and general way. Nevertheless, in practice, the interest rates as well as the economic state variables have to be clearly specified, such that the credit risk can be accurately determined. The probability of default, expected loss and present value of expected loss also depend on the state variables and the interest rates. The exact formulations can be found in [69, pp. 300ff]. To estimate the necessary input data of the models, implicit estimation can be used again. However, in case of reduced form models, historical estimation is also possible to determine the input data. In historical estimation, the input data of the reduced form model are determined with the help of an hazard rate estimation. The technique behind hazard rate estimation is to determine the probability for a binary event, like default or no default. A detailed description about hazard rate estimation can be found in [93].

Reduced form models have the following advantages [69]. First, due to the stated assumptions, the model's input data is observable and in this way, historical estimation can be used to determine the credit risk measure, like default probability. Second, all the possible credit risk depends on the evolution of the economy. Thus, the determination of the credit risk is

more realistic because the actual business cycle influences the result. Finally, the model is independent of a specific balance sheet structure in contrast to structural models [68].

The main drawback of the reduced form models is the use of hazard rate estimation. In hazard rate estimation, past observations are used to predict the future. To obtain an accurate model, the model has to be properly formulated and back tested. This chapter will be completed with a conclusion.

3.4   Conclusion

In this chapter, bonds are defined as well as the market in which they are traded. The difference between primary and secondary bond market is explained and an introduction into credit ratings is given. Additionally, it is shown why credit ratings are important in the financial world and the reason of their wide use is given. Furthermore, the credit risk analysis is introduced and different models are briefly discussed. The differentiation into traditional models and structural / reduced form models is made. Credit scoring and credit ratings are traditional models to determine the credit risk. However, credit scoring is only used for individuals and small companies. Structural and reduced form models are the most accurate predictors concerning credit risk. Credit ratings are the most used in the financial world. However, the lack of accuracy is due to the fact that rating agencies tend to keep ratings relatively stable over time to reduce the risk of price volatilities. Credit ratings is widely used due to the fact that credit ratings are applicable to companies and countries. Reduced form models perform better than structural models due to several reasons [69, 68]. First, the performance is increased with the help of the flexibility of hazard rate estimation. Second, reduced form models are able to incorporate changes in business cycle. Finally, they are not dependent on one specific structure of a company's balance sheet [69].

Chapter 4

Classification methods

This chapter deals with different methods of classification. First, classification is defined in general. Afterwards, a general description of three main methods is furnished: artificial neural networks (ANN), support vector machines (SVM) and support vector domain description (SVDD). The description is kept quite general, which means that the indicated formulas for the several methods refer to an unspecific dataset. Under unspecific dataset, it is understood that the data points do not have to be of a specific nature like bonds but they can vary from cars to houses or any other item.

4.1 Classification

In classification, also called supervised clustering, the outputs are known and the dataset should be divided into different clusters. A widely used definition of clusters is as follows [28]:

Definition 6. *A cluster is defined as a set of data points which share many characteristics. Additionally, the data points of one specific cluster are very different in comparison to the data points of another cluster*

The optimal number of clusters can be determined with the knowledge over the given outputs. Furthermore, the information about the desired output is used during the clustering process [98, 61]. The clustering process can be summarized as follows. The procedure of a cluster analysis contains different steps. First of all, the attributes of the data points have to be selected to compute the similarities or dissimilarities between the data points and to determine the clusters/groups. After this decision is taken, the selection of the clustering method has to be made. This step (selection of the method) is crucial because each method usually produces a different result. The selection of the data points' attributes is very important

because false assumptions on the variables lead to wrong and uninterpretable clusters. In [87], it is shown that the used attributes should always be independent to achieve significant clusters. Additionally, too many attributes will also reduce the quality of the result. The selected attributes should always have a degree of uniqueness [87, 28] . With a high degree of collinearity between attributes, their uniqueness can not be guaranteed anymore [28]. In this way, the following advice can be given: if the correlation between two variables is above 0.90 then one variable should be omitted [87, 28]. The only common existing rule which exists is that the number of data points always have to be greater than the selected attributes. Nevertheless, a general rule explaining the ratio between number of data points and number of attributes does not exist. A widely used acceptance is that if m is the number of attributes then the number of data points should be approximately 2^m [28]. By output, it is meant the membership to a cluster of the data points. This additional information is used to influence the clustering method to recover the clusters in the dataset and finally, to obtain a classification rule. This rule is then used to assign new data points into one cluster. The benefit of classification is that a deeper insight of the dataset is achieved as the relation between the data points and their clusters' assignment is better interpretable [26]. Additionally, in classification, the known desired output is often used to optimize or define the distance function. This way, the best appropriated decision function can be identified which allows to obtain additional information about the dataset [26].

Different methods are developed and applied to execute classification. The spectrum of employed methods which are discussed in this thesis spans from artificial neural networks [24] to support vector machines [29]. All these methods have in common that they require a so-called training set. The following definition describes the training set.

Definition 7. *Let x_i, for $i = 1, \ldots, n$, be a data point characterized by its m attributes denoted by x_{ij}, for $j = 1, \ldots, m$. y_i, for $i = 1, \ldots, n$, represents the known output for data point x_i. Thus, the training set, X, is the set of data points characterized by their attributes and their known output. Mathematically, the training set is expressed as follows: $X = \left\{ x_i = (x_{ij}, y_i), \forall i = 1, \ldots, n; \quad j = 1, \ldots, m \right\}$*

The training set is used to determine the parameters of the algorithm in such a way that the desired clusters are recovered. Usually, the dataset containing new unclassified data points is called testing set and is defined as follows.

Definition 8. *Let $\tilde{x}_i$, for $i = 1, \ldots, \tilde{n}$, be new unclassified data points. Thus, $\tilde{x}_i$ is only described by its attributes denoted by $\tilde{x}_{ij}$, for $j = 1, \ldots, m$ and no information about the output is given. The testing set, $\tilde{X}$, is defined as the set regrouping all the new unclassified data points which is expressed mathematically in the following way: $\tilde{X} = \left\{ \tilde{x}_i = (\tilde{x}_{ij}), \forall i = 1, \ldots, \tilde{n}; \quad j = 1, \ldots, m \right\}$*

The trained method can then be used to classify the data points from the testing set into one of the existing clusters. Three different methods are described in detail in the next sections: artificial neural network, support vector machines and support vector domain description.

4.2 Artificial Neural Networks

Artificial neural networks (ANN) are widely used in the scientific community. The structure of an ANN is described and afterwards, the procedure of ANN is explained with the help of the most popular ANN technique, called feed-forward back-propagation.

4.2.1 Structure of ANN

Artificial neural networks imitates the human brain. The human brain is composed of millions of neurons and the neurons communicate witch each other with the help of synapses. The idea of neurons and synapses is used to develop an ANN. A neuron is defined as a single computation unit of a layer. Three different layers are distinguished:

- Input layer

- Hidden layer

- Output layer

The following definitions provides the characteristics of the different layers [91, 55].

Definition 9. *The input layer communicates with the external world through its neurons. Each neuron in the input layer represents one attribute of the data points. Thus, in the input layer, there are as many neurons as attributes are used to characterize the data points.*

Definition 10. *The obtained results of the network are communicated to the external world with the help of the output layer. The number of neurons in the output layer is equal to the number of desired classes/outputs which are defined by the training set. Thus, each output neuron represents one specific output.*

Definition 11. *The hidden layer lies between the input and output layer. It is possible that there is more than one hidden layer. The number of neurons in the hidden layer is unknown and is variable at the start of the training process. The number of hidden neurons is defined by the user. The hidden layer can be defined by using feature extraction methods on the input.*

Finally, the definition of an artificial neural network can be given.

Definition 12. *An artificial neural network (ANN) is composed of at least two layers, the input and output layers. Normally, at least one hidden layer is contained in an ANN. An example of an ANN is given in the following figure.*
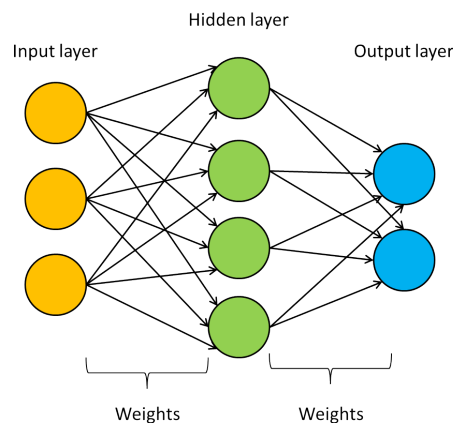
Fig. 4.1 Structure of an artificial neural network [55]

Each connection between two neurons is described by its weight. The weight defines the importance of the connection in the network. Furthermore, a neuron is said to be active if a threshold value is exceeded. However, the training of the threshold values is complicated and therefore an additional weight is integrated and connects a so-called bias neuron or simply bias. The bias regroups the threshold values and the training is made simultaneously with the other weights in the network. Before explaining in detail the optimal weights, the neural networks can be differentiated according to their used connections [24].

- Feed forward network: The connections are only permitted to neurons of the next layer.

- Feed forward network with shortcut connections: Similar to the normal feed forward network, the connections are not only directed towards the following layer but they are permitted to be directed towards other subsequent layers.

- Direct recurrence network: Based on the normal feed forward network, the additional permission is added such that all the neurons are connected to themselves.

- Indirect recurrence network: Again the starting point is a normal feed forward network, additional connections are set between the neurons and their preceding layer.

- Lateral recurrence network: A feed forward network with additional connections between the neurons of the same layer.

- Complete interconnection network: In this case, each neuron is allowed to be connected to every other neuron. However, in this case, it is also possible that each neuron can be an input neuron.

Here, the focus lies on the most used ANN, the feed forward networks [24, 55, 42]. Furthermore, only one hidden layer will be used in the introduced ANN. The training procedure of the ANN is described by using the back propagation rule. Afore, the transfer or activation function has to be defined.

Definition 13. *A transfer function, also called activation function, transforms the neuron's input into its output. In other words, the transfer function transmits the information of one neuron to the neurons of the next layer.*

The most used transfer function is the sigmoid function and is defined as follows:

$$f(x) = \frac{1}{1 + \exp(-x)} \tag{4.1}$$

The transfer between input and hidden layer as well as hidden and output layer can take place with different functions. However, for the two transmissions, the same transfer function is usually used [55]. Additionally, the input of one single neuron has to be defined [55, 42].

Definition 14. *The input of one neuron is a function depending on the transmitted attributes of one data point and the weights connected to the neuron. The most used function is the weighted sum which corresponds to the summation of the multiplications of the output of each neuron in the previous layer with the corresponding weight which connects the neuron to the neuron for which the input is searched.*

Herewith, the training of an ANN to determine the optimal weights is described and in the next subsection the whole procedure of an ANN is explained in detail with the help of an example. The first step is to go through the network, from the input layer to the output layer. Therefore, the weights are initialized with random values. Afterwards, the output for each data point is computed by introducing the data point's attributes into the network. Through the transfer function, the attributes are transformed and transmitted to the different layers (input, hidden and output layer) after finally attaining the output layer. This direct transmission is called feed forward. After the predicted outputs for each data point are determined, the error between the predicted and the given output has to be calculated for each data point. At the stage of the output layer, the error is easily asserted; the prediction has to coincide with the given output. Then, the error is propagated back to the different preceding layers (hidden and input) and the weights are updated to improve the performance

of the network [95, 75]. This step is called back propagation of the error or, in short, back propagation. After several iterations of the combination, feed forward succeeded by back propagation, the weights are trained and the neural network predicts in the best possible way the output of the data points. Finally, the outputs of new data points can be predicted with the help of the trained network.

4.2.2 Procedure of ANN

The bias of each layer is integrated into the weight vector [55, 42]. The pseudo-code of the functions, *FeedForward* and *BackPropagate* are given in Algortihm 2 and Algorithm 3. The pseudo-codes are inspired by [63]. The pseudo-code of the feed-forward ANN is as follows.

Algorithm 1 Feed-Forward Neural Network

1: **procedure** ANN
2: *Input:*
3:     X Training set
4:     i_{max} maximal number of iterations
5:     $Hidden\_Neurons$ number of hidden neurons
6:     λ_{hid} learning rate in the hidden layer
7:     λ_{out} learning rate in the output layer
8:     μ_{hid} momentum of the hidden layer
9:     μ_{out} momentum of the output layer
10: *Output:*
11:     w_{ih} weights of the connections between input and hidden layer
12:     w_{ho} weights of the connections between hidden and output layer
13: *Do:*
14:     Assign random values to the weights
15:     Iterate unil i_{max} is reached
      Iterate over all data points of X
      FeedForward()
      BackPropagate()
16:     Output the obtained weights w_{ih} and w_{ho}

Algorithm 2 FeedForward - Function

1: **procedure** FEED - FORWARD FUNCTION
2: *Input:*
3:      *Hidden_Neurons* number of hidden neurons
4:      w_{ih} weights of the connections between input and hidden layer
5:      w_{ho} weights of the connections between hidden and output layer
6:      x_i one data point
7: *Output:*
8:      *actual* predicted result
9: *Do:*
10:      Iterate over all hidden neurons
            For each k in $\{0,\dots,Hidden\_Neurons-1\}$ compute the sum
            $sum(k) = \sum_{j=0}^{m-1}(x_{i(j+1)}w_{ih}(j,k)) + w_{ih}(m,k)$
            Apply the transfer function on the sum: $hidden(k) = \frac{1}{1+\exp\{-sum(k)\}}$
11:      Iterate over all output neurons
            For each k in $\{0,\dots,Output\_Neurons-1\}$ compute the sum
            $sum(k) = \sum_{j=0}^{Hidden\_Neurons-1}(hidden(j)w_{ho}(j,k)) + w_{ho}(Hidden\_Neurons,k)$
            Apply the transfer function on the sum: $actual(k) = \frac{1}{1+\exp\{-sum(k)\}}$
12:      Output the predicted result, *actual*

Afterwards, the whole procedure of a neural network to determine the optimal weights is explained in detail. *X* is the training set containing *n* data points. Each data point is characterized by its *m* attributes, x_{ij}, for $j = 1,\dots,m$ and the given output y_i. As previously explained, the number of attributes defines the number of input neurons, so *m* input neurons are necessary. However, as the network uses bias factors, $m+1$ neurons and respective weights are defined. Nevertheless, as seen in the Functions 2 and 3, the weights vectors are starting from 0 instead of 1 and thus $m-1$ is the last computing neuron and *m* stands for the bias neuron. Furthermore, the given outputs of the data points have to be transformed. The number of different possible given outputs determines the maximum number of output neurons, noted *Output_Neurons*. Afterwards, a vector of zeros and ones is assembled to identify the correspondence of a data point to one output neuron. For example, lets assume that there are 4 data points with the following given outputs: $y_1 = 1$, $y_2 = 2$, $y_3 = 2$ and $y_4 = 1$.

Algorithm 3 BackPropagate - Function

1: **procedure** BACK - PROPAGATION FUNCTION
2: *Input:*
3: 　　 *Hidden_Neurons* number of hidden neurons
4: 　　 w_{ih} weights of the connections between input and hidden layer
5: 　　 w_{ho} weights of the connections between hidden and output layer
6: 　　 x_i one data point
7: 　　 *target* the given output of data point, x_i
8: *Output:*
9: 　　 w_{ih}^{new} updated weights for the connections between input and hidden layer
10: 　　 w_{ho}^{new} updated weights for the connections between hidden and output layer
11: *Do:*
12: 　　 "Compute the error made in the output layer"
　　　　 Iterate over all output neurons
　　　　 For each k in $\{0,\dots,Output\_Neurons-1\}$ compute the error
　　　　　 $erro(k) = (target(k) - actual(k))(actual(k)(1-actual(k)))$
13: 　　 "Compute the error made in the hidden layer"
　　　　 Iterate over all hidden neurons
　　　　 For each k in $\{0,\dots,Hidden\_Neurons-1\}$ compute the error
　　　　　 $errh(k) = (\sum_{l=0}^{Output\_Neurons-1} erro(l)w_{ho}(k,l))(hidden(k)(1-hidden(k)))$
14: 　　 "Update the weights for the output layer"
　　　　 Iterate over all output neurons
　　　　 For each k in $\{0,\dots,Output\_Neurons-1\}$ update the weights
　　　　　 $IF(erro(k)\neg 0)Then$
　　　　　　 For each l in $\{0,\dots,Hidden\_Neurons-1\}$
　　　　　　　 $dwho = \mu_{out}(w_{ho}^{new}(l,k) - w_{ho}(l,k))$
　　　　　　　 $w_{ho}(l,k) = w_{ho}^{new}(l,k)$
　　　　　　　 $w_{ho}^{new}(l,k) = w_{ho}(l,k) + \lambda_{out}erro(k)hidden(l) + dwho$
　　　　　　 Update the bias
　　　　　　　 $obias = \mu_{out}(w_{ho}^{new}(Hidden\_Neurons,k) - w_{ho}(Hidden\_Neurons,k))$
　　　　　　　 $w_{ho}(Hidden\_Neurons,k) = w_{ho}^{new}(Hidden\_Neurons,k)$
　　　　　　　 $w_{ho}^{new}(Hidden\_Neurons,k) = w_{ho}(Hidden\_Neurons,k) + \lambda_{out}erro(k)$
　　　　　　　　　　　　　　　 $+obias$
15: 　　 "Update the weights for the hidden layer"
　　　　 Iterate over all hidden neurons
　　　　 For each k in $\{0,\dots,Hidden\_Neurons-1\}$ update the weights
　　　　　 $IF(errh(k)\neg 0)Then$
　　　　　　 For each j in $\{0,\dots,m-1\}$
　　　　　　　 $dwih = \mu_{hid}(w_{ih}^{new}(j,k) - w_{ih}(j,k))$
　　　　　　　 $w_{ih}(j,k) = w_{ih}^{new}(j,k)$
　　　　　　　 $w_{ih}^{new}(j,k) = w_{ih}(j,k) + \lambda_{hid}errh(k)x_{i(j+1)} + dwih$
　　　　　　 Update the bias
　　　　　　　 $hbias = \mu_{hid}(w_{ih}^{new}(m,k) - w_{ih}(m,k))$
　　　　　　　 $w_{ih}(m,k) = w_{ih}^{new}(m,k)$
　　　　　　　 $w_{ih}^{new}(m,k) = w_{ih}(m,k) + \lambda_{hid}errh(k) + hbias$
16: 　　 Output the updated weights and bias, w_{ho}^{new} and w_{ih}^{new}

Thus, 2 output neurons have to be used and the transformed outputs for each data point is as follows:

Data points	Output neuron 1	Output neuron 2
x_1	1	0
x_2	0	1
x_3	0	1
x_4	1	0

The transformed output for a given data point is described in a vector called as *target*. Then, the procedure of a neural network is explained by going through the different steps. First, random values are assigned to the weights vectors, w_{ih} and w_{ho}. Nevertheless, two instances of each weight vector are used and differentiated by adding the upper-script *new*. Thus, the second instances of w_{ih} are noted as w_{ih}^{new}. In the beginning the two instances are initialized with the same values. The two instances of weight vectors are used to undertake the optimization of the weights.

Starting the detailed description of the training, each single data point is introduced in the network and passed through it in the forward direction. For example, the computations are described for data point x_i. Each neuron of the input layer takes over one attribute of the data point. This means that the first attribute x_{i1} is introduced in the network by the first input neuron and so on. Then, the input of the neurons in the hidden layer has to be determined. Lets compute, for example, the input of the kth hidden neuron which is given by the following formula:

$$sum(k) = \sum_{j=0}^{m-1} x_{i(j+1)} w_{ih}(j,k) + w_{ih}(m,k) \tag{4.2}$$

The first term on the right side in the formula 4.2 is the so-called weighted sum, which is nothing else than the sum of the multiplication of the weights with the attributes of x_i. The second term represents the bias factor which is added to the weighted sum and the result is the input of the kth hidden neuron. Afterwards, with the help of the transfer function, the output of the hidden neuron is defined. The output is calculated as follows:

$$hidden(k) = \frac{1}{1 + \exp\{-sum(k)\}} \tag{4.3}$$

The computations given by 4.2 and 4.3 have to be executed for each neuron in the hidden layer. After the output for each hidden neuron is identified, the input of the neurons in the output layer can be determined. Again, the computations are explained on the exemplary kth output neuron. The computations are similar to those that are used to determine the input of

the hidden neurons. The input of the kth output neuron is determined as follows:

$$sum(k) = \sum_{j=0}^{Hidden\_Neurons-1} hidden(j)w_{ho}(j,k) + w_{ho}(Hidden\_Neurons,k) \qquad (4.4)$$

The meaning of the two terms on the right hand side of the formula is identical to the one for the formula 4.2. Afterwards, the transfer function determines the output of the output neuron and especially the result of the network's computation is issued to the external world with the help of the transfer function. The output is given as follows:

$$actual(k) = \frac{1}{1 + \exp\{-sum(k)\}} \qquad (4.5)$$

After the output for each neuron in the output layer is computed, the predicted output is determined for x_i. However, the predicted output is possibly different from the given one. Therefore, the error made in each stage (output and hidden layer) has to be determined and the weights have to be updated to reduce the error. The error made in the output layer is the difference of the predicted and the given output for each neuron. The subsequent formula identifies the exactly error for the kth output neuron:

$$erro(k) = (target(k) - actual(k))(actual(k)(1 - actual(k))) \qquad (4.6)$$

The first term of the multiplication is the difference of the prediction and the right output. The second term represents the inverse form of the transfer function. As with the function 4.5, the input of the output neuron is transfered to the external world, the error has to be transfered back to the neurons in the output neurons from the external world to be able to adjust the weights [13]. Lets assume, an error has occurred. Then, iterating over the hidden neurons, the adjustment of the weights is given in the following way:

$$dwho = \mu_{out}(w_{ho}^{new}(l,k) - w_{ho}(l,k)) \qquad (4.7)$$

$$w_{ho}(l,k) = w_{ho}^{new}(l,k) \qquad (4.8)$$

$$w_{ho}^{new}(l,k) = w_{ho}(l,k) + \lambda_{out}erro(k)hidden(l) + dwho \qquad (4.9)$$

The formula 4.7 is only used if a momentum, μ_{out}, is employed. A momentum is utilized in ANN to influence the speed of adjustment. The momentum adds the difference between the actual and the previous values of the weights to the adjustment and forces in such a way that the network increases or decreases the change. If the difference is positive, then the change increases as it is assumed that the weights are still far off their optimum. In contrast,

if the difference is negative, then the change decreases as the weights are near their optimum. Additionally, the momentum allows to reduce the risk to be trapped in a local optimum [74]. The second formula in the adjustment step, 4.8, stores only the actual weights in the second instance. Thus, during the next iteration, the second instance of the weight vector always contains the weights before the update is made. In the last formula, 4.9, the update of the weights is actually undertaken. The learning rate, λ_{out}, indicates the speed of change in the value of the weights. A higher value of the learning rate means that the adjustment of the weights is more important than if the learning rate has a small value. Nevertheless, if the learning rate is too high, there is a risk that the optimum values can not be found because the change of the weights makes the values exceed the optimum instead of reaching it [55]. In the opposite way, a too small value of the learning rate slows down the speed of the network in such a way that it can not be used for large datasets. Afterwards, the bias factor in the output layer also has to be updated. The adjustment is undertaken by the following formula:

$$obias = \mu_{out}(w^{new}(Hidden\_Neurons,k) - w(Hidden\_Neurons,k)) \qquad (4.10)$$

$$w_{ho}(Hidden\_Neurons,k) = w_{ho}^{new}(Hidden\_Neurons,k) \qquad (4.11)$$

$$w_{ho}^{new}(Hidden\_Neurons,k) = w_{ho}(Hidden\_Neurons,k) + \lambda_{out}erro(k) + obias \qquad (4.12)$$

To update the bias of the k$^{\text{th}}$ output neuron, the momentum and the learning rate are again utilized. The principle is the same as for the adjustment of the weights. First, the difference of the actual and previous bias is computed, 4.10, then the actual bias is stored in the second instance for the next update iteration 4.11 and finally, the bias is adjusted, 4.12. The impacts of the momentum and the learning rate are the same as mentioned for the update of the weights.

After the weights of the connections between the hidden and output layer are adjusted, the error which is made in the hidden layer, has to be identified. The error for the k$^{\text{th}}$ hidden neuron is calculated as follows:

$$errh(k) = \left(\sum_{l=0}^{Output\_Neurons-1} erro(l)w_{ho}(k,l) \right) (hidden(k)(1-hidden(k))) \qquad (4.13)$$

To compute the error, the error of the output layer is multiplied with the weights connecting the neurons of the output to those of the hidden layer. Thus, the error made in the output layer is propagated back to the hidden layer. This explains also the name of the method which is "Back - Propagation". Now, as the error is determined, the weights connecting the input to the hidden layer can be updated. The adjustment for the k$^{\text{th}}$ neuron is made by iterating over

all input neurons and is given by the subsequent formula:

$$dwih = \mu_{hid}(w_{ih}^{new}(l,k) - w_{ih}(l,k)) \tag{4.14}$$

$$w_{ih}(l,k) = w_{ih}^{new}(l,k) \tag{4.15}$$

$$w_{ih}^{new}(l,k) = w_{ih}(l,k) + \lambda_{hid}errh(k)x_{i(l+1)} + dwih \tag{4.16}$$

And the update of the bias factor of the hidden layer is given as follows:

$$hbias = \mu_{hid}(w_{ih}^{new}(m,k) - w_{ih}(m,k)) \tag{4.17}$$

$$w_{ih}(m,k) = w_{(}^{new}m,k) \tag{4.18}$$

$$w_{ih}^{new}(m,k) = w_{ih}(m,k) + \lambda_{hid}errh(k) + hbias \tag{4.19}$$

The same explanations to describe the steps to update the weights and bias of the output layer are valid to describe the formulas to adjust the weights and bias of the hidden layer. The adjustment is made in the same principle. To determine the optimal weights, it is not only necessary to adjust them by each pass by a new data point of the training set, X, but the whole training with all the data points is repeated several times. The maximum number of reputation, i_{max}, is a user-given input value at the beginning of the training process. However, one has to mention that during the training, only the achievement of the local optimum is guaranteed. There is no guaranty that the determined weights are globally optimal [91]. Finally, the neural network is trained and can be used to predict the output of new data points with unknown outputs. According to the introduced notations, $\tilde{X}$ represents the set of new data points. The subsequent pseudo-code furnishes the method to utilize a trained ANN.

Algorithm 4 Apply Neural Network

1: **procedure** APPLYING ANN
2: *Input:*
3:     $\tilde{X}$ Data set of new data points
4: *Output:*
5:     *prediction* the predicted output
6: *Do:*
    Iterate over all data points of $\tilde{X}$
        FeedForward()
        Determine the index which corresponds to the maximum in the vector *actual*
7:     Output the predicted output *prediction*

Thus, the use of a trained ANN is very simple and is described as an example for the new data point, $\tilde{x}_i$. First, $\tilde{x}_i$ is introduced into the network and is then processed by traversing the input, hidden and output layer according to the feed forward rule. The formulas 4.2, 4.3, 4.4 and 4.5 are applied to the data points. The result of the "Feed-Forward function" is the vector of predicted outputs of $\tilde{x}_i$, noted by *actual*. The vector *actual* contains a certain kind of membership of $\tilde{x}_i$ to each output neuron. Keeping in mind, that each output neuron represents one of the different existing outputs which are defined and given by the training set, the prediction is completed by determining the index of the output neuron which delivered the maximal value based on the attributes of $\tilde{x}_i$. Then, the index determines the exact output/class to which the data point belongs.

To conclude the section, the main disadvantages and advantages are listed [91]. The main advantages are:

1. Neural networks can be developed and trained without great knowledge of statistics

2. Complex non-linear relations between dependent and non-dependent attributes can be implicitly detected by neural networks

3. The ability to detect all possible interactions between the attributes of the data points is owned by neural networks

4. Besides back propagation, different other training techniques exist to develop neural networks [32, 99]

However, neural networks are not the perfect analysis method and have several drawbacks:

1. Neural networks are often considered as black boxes because a causal relationship between input and output is difficult to identify

2. The results of neural networks can not easily be published in such a way that other scientists can directly repeat the tests

3. The required computational resources for developing and training neural networks are significantly greater than for other methods, like logistic regression

4. Over-fitting is a great problem of neural networks which means that a network represents the training universe well but has poor performance when it is applied on new data points

5. The development of neural networks for a specific case is, in the main part, an empirical affair because the optimum neural network does not exist

This section has introduced ANN. The structure and the procedure of the widely used feed forward neural networks are given. Additionally, the different computations to train a neural network are explained and illustrated with their respective formulas. In the next section, the support vector machines are described in detail.

4.3   Support Vector Machines

Another method to divide data points into different clusters are support vector machines (SVM). Based on the research of ANN, SVM have been developed to overcome some of the disadvantages of neural networks. The problem of over-fitting and the restriction to the local optimum are the main focus in the development of SVM. The whole theory of SVM is explained in detail.

4.3.1   Theory of SVM

The idea behind SVM is able to define hyperplanes separating the different groups/clusters in a dataset. In other words, if the data points are only characterized by two attributes, so if they lie in the 2D-space, then the clusters are separated with the help of lines. The following figure shows the separating lines for 3 groups.
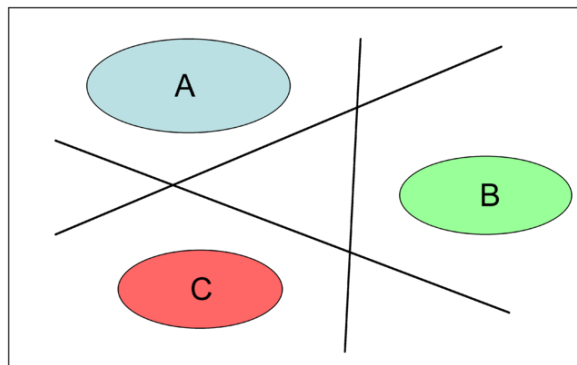
Fig. 4.2 Example of SVM with 3 groups [2]

Even if there are SVM to divide a dataset into several groups, this subsection is restricted to the case of binary SVM. A binary SVM is used to divide a dataset into two different groups. Thus, a hyperplane which lies between these two groups is searched and SVM tries to find this optimal separator. In the 2D-space, the searched hyperplane is identical to a line.
On a more theoretical level, one has to remember that the training dataset is given by X and the data points are noted as x_i for $i = 1, \ldots, n$ with n number of data points. Each data point is characterized by its attributes, x_{ij} for $j = 1, \ldots, m$ with m maximum number of attributes,

and its given output y_i. Naturally, the given output represents the membership to one of the two existing groups. y_i is either 1 or -1 depending on which group the data point x_i belongs to. According to the given output y_i, the data points of the training set can be divided into two different groups and the hyperplane separating these two groups is searched. The hyperplane is denoted by H. For a better understanding, lets name the data points with $y_i = 1$, the positive data points and those with $y_i = -1$ the negative data points. Let now suppose that one separating hyperplane H is found, then the data point x_i which lies on H fulfills the following equation [11, 12, 21]:

$$\mathbf{w}x_i + b = 0 \tag{4.20}$$

with $\mathbf{w}$ is normal to H and $\frac{|b|}{\|\mathbf{w}\|}$ represents the distance from the hyperplane to the origin. The subsequent definition defines the margin [11, 12, 21].

Definition 15. *The margin of a separating hyperplane H is defined by the sum of the shortest distance between H and the positive data points and the shortest distance between the negative data points and H.*

The goal of SVM is to determine the optimal hyperplane which separates the positive and negative data points. The optimal hyperplane is defined as follows [11, 12, 21]:

Definition 16. *The separating hyperplane with the largest margin is defined as the optimal hyperplane. Furthermore, this hyperplane is unique.*

By defining the optimal hyperplane, SVM tries to separate the dataset in a linear way into two groups. Afterwards, new data points can be classified into one of the two existing groups by determining on which side the new data points lie according to the hyperplane. As linearity is a key point in SVM, the distinction between linearly separable and non-separable cases has to be made.

4.3.2 Linear SVM

With the help of linear SVM, a dataset is divided linearly into two groups. Nevertheless, the following differentiation has to be made:

1. The groups in the dataset are clearly separate and can be divided by a hyperplane.

2. The groups in the dataset are not clearly separable and for the determination of the separating hyperplane, it is allowed that some data points lie on the wrong side of the hyperplane.

Separable case

The linearly separable case is illustrated in Figure 4.3 in the 2-dimensional space.
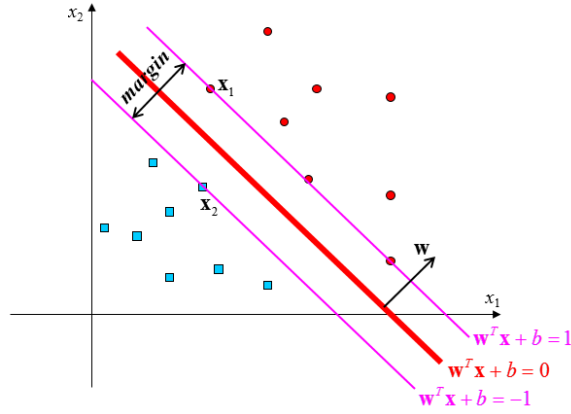
Fig. 4.3 Linear separable case in the 2D space [12]

The data points of the training set X has to satisfy the following constraints:

$$x_i\mathbf{w} + b \geq 1 \text{ for } y_i = 1 \tag{4.21}$$

$$x_i\mathbf{w} + b \leq -1 \text{ for } y_i = -1 \tag{4.22}$$

$$\forall i = 1,\ldots,n$$

The two stated constraints, 4.21 and 4.22, can be summarized as one combined constraint:

$$y_i(x_i\mathbf{w} + b) - 1 \geq 0 \quad \forall i = 1,\ldots,n \tag{4.23}$$

The constraint identifies the side according to the hyperplane of a data point and in this way, the group's membership of this data point. To determine the optimal separating hyperplane, two auxiliary hyperplanes are defined [11, 12, 21]. Starting with the positive data points ($y_i = 1$), the first auxiliary hyperplane H_1 is characterized by the data points which are the same as in constraint 4.21. Thus, the following equation defines this hyperplane:

$$x_i\mathbf{w} + b = 1 \tag{4.24}$$

Additionally, the second auxiliary hyperplane H_2 is characterized by the negative data points which equate constraint 4.22 and is defined by the subsequent equation:

$$x_i\mathbf{w} + b = -1 \tag{4.25}$$

As the margin can also be defined as the distance between two groups, the group of positive and the group of negative data points, the two auxiliary hyperplanes, H_1 and H_2, allow to determine the margin. Therefore, the margin is given by: $\frac{2}{\|\mathbf{w}\|}$. As in SVM the optimal hyperplane H is defined as the hyperplane with the maximum margin, it is enough to minimize $\|\mathbf{w}\|$ to determine the optimal hyperplane [11, 12, 21]. The optimal hyperplane is obtained by solving the following mathematical program:

<div align="center">

Mathematical Program 4.1:
SVM - Primal Problem Formulation - Separable Case

</div>

$$\begin{aligned} &\min && \|\mathbf{w}\|^2 \\ &\text{subject to} && y_i(x_i\mathbf{w}+b)-1 \geq 0 \quad \forall i = 1,\ldots,n \end{aligned}$$

Those data points which equalize constraint 4.23, that means the data points which lie on H_1 respectively H_2, are called support vectors.

To solve Mathematical Program 4.1, the Lagrangian formulation of the problem is employed. There are two main reasons to utilize this formulation. First, The constraint 4.23 is replaced by constraints on the Lagrange factors which are easier to handle. Second, in the reformulation of the problem, only the data points appear in forms of dot products between vectors. The rule concerning the Lagrange factors states that if the constraint can be expressed to be greater or equal than 0, then the constraint is multiplied by a positive Lagrange factor and subtracted from the objective function to form the Lagrangian [12, 21]. Thus, using n Lagrangian factors, noted α_i, the primal Lagrangian is obtained:

$$L_P \equiv \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i y_i(x_i\mathbf{w}+b) + \sum_{i=1}^{n} \alpha_i \tag{4.26}$$

Now, L_P has to be minimized to obtain the optimal hyperplane H. Before defining the dual Lagrangian, the following proposition is stated without given its proof. Its proof can be found in [12].

Proposition 1. *A linear constraint defines a convex set and a set of n linear constraints defines the intersection of n convex sets. The intersection of convex sets is always a convex set [12].*

Additionally, the subsequent definition is stated [44].

Definition 17. *Let S be a convex set in a real vector space and let $f : S \to \mathbb{R}$ be a function. Then f is called convex if:*

$$\forall x_1, x_2 \in S, \quad \forall t \in [0,1]: \qquad f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)$$

The L_P represents a convex quadratic problem since, all the data points satisfying the constraints form a convex set and the objective function itself is also a convex function [11, 12, 21]. Thus, solving the primal or the dual Lagrangian is identical [11, 12, 21]. The dual Lagrangian is obtained by taken the partial derivatives according to $\mathbf{w}$ and b of the primal Lagrangian 4.26:

$$\mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i x_i \tag{4.27}$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \tag{4.28}$$

The fact that these two derivatives are equalities, they can be substituted into the primal Lagrangian to obtain their dual form.

$$L_D \equiv \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{n} \alpha_i \alpha_k y_i y_k x_i x_k \tag{4.29}$$

However, instead of minimizing, the dual Lagrangian has to be maximized according to the Lagrange factors α_i to determine the optimal hyperplane H. Finally, the following mathematical program can be state:

<div align="center">

Mathematical Program 4.2:
SVM - Dual Problem Formulation - Separable Case

</div>

$$\text{max} \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{n} \alpha_i \alpha_k y_i y_k x_i x_k$$

$$\text{subject to} \quad \alpha_i \geq 0 \quad \forall i = 1, \ldots, n$$

With the help of the partial derivate 4.27, the normal of the optimal hyperplane can be identified. Additionally, all the data points with $\alpha_i \geq 0$ gives the solution to constraint 4.23. Thus, these data points are the support vectors. All the other data points have $\alpha_i = 0$ and lie either on the side of H_1 or H_2. For these data points, constraint 4.23 is a strict inequality. To ensure that the results of the dual Lagrangian are also the optimal values for the primal Lagrangian, the obtained Lagrange factors, α_i, have to satisfy the Karush-Kuhn-Tucker (KKT) conditions [30]. In theory and practice of constrained optimization, a central role is played by the KKT conditions. As previously stated, the problem for SVM is convex and this implies that the KKT conditions are necessary and sufficient to guarantee that $\mathbf{w}$ and b are the optimal solutions for the problem [66, 30]. For the primal problem, 4.26, the KKT

conditions are as follows:

$$\frac{\partial}{\partial \mathbf{w_j}} L_p = \mathbf{w_j} - \sum_{i=1}^{n} \alpha_i y_i x_{ij} = 0 \quad j = 1, \ldots, m \tag{4.30}$$

$$\frac{\partial}{\partial b} L_p = -\sum_{i=1}^{n} \alpha_i y_i = 0 \tag{4.31}$$

$$y_i(x_i \mathbf{w} + b) - 1 \geq 0 \quad i = 1, \ldots, n \tag{4.32}$$

$$\alpha_i \geq 0 \quad \forall i \tag{4.33}$$

$$\alpha_i(y_i(x_i \mathbf{w} + b) - 1) = 0 \quad \forall i \tag{4.34}$$

In contrast to the normal $\mathbf{w}$ representing the distance of the hyperplane to the origin, b is not directly given. However, b can be identified with the help of one support vector and the constraint 4.23 of the mathematical program 4.1. Lets suppose that data point x_i is a support vector then b is computed in the following way:

$$b = \frac{1}{y_i} - x_i \mathbf{w} \tag{4.35}$$

Thus, all the parameters characterizing the optimal hyperplane are defined and the SVM is completely trained. Finally, the trained SVM is ready to be applied on new data points. Next, the non-separable case is discussed.

Non-separable case

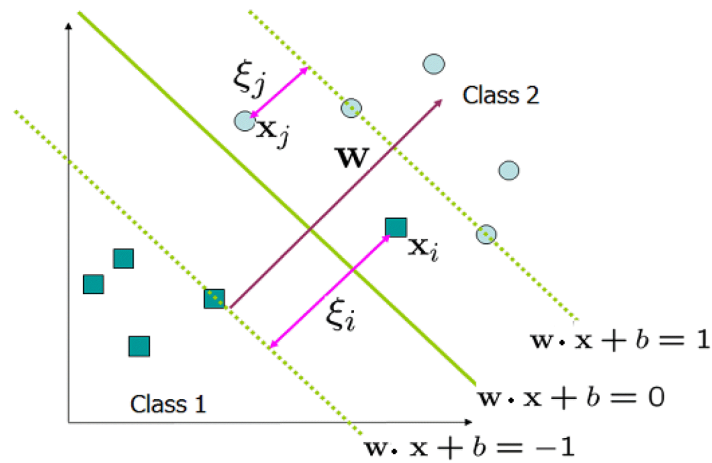The non-separable case for 2 dimensions is illustrated in the following figure.

Fig. 4.4 Example of the non-separable case

In the non-separable case, the negative and positive data points from the training set X can not be clearly divided into two groups. In the non-separable case, there always are some data points which lies on the wrong side of the hyperplane. Therefore, if the mathematical program given for the separable case 4.2 is employed, then no feasible solution would be found because the objective function 4.29 grows arbitrarily large [11, 12, 21]. To overcome this drawback, the two inequalities, 4.21 and 4.22, defining, H_1 and H_2, have to be relaxed. The relaxed versions of these inequalities are as follows:

$$x_i\mathbf{w} + b \geq 1 - \xi_i \quad \text{for} \quad y_i = 1 \tag{4.36}$$

$$x_i\mathbf{w} + b \geq -1 + \xi_i \quad \text{for} \quad y_i = -1 \tag{4.37}$$

$$\xi_i \geq 0 \quad \forall i = 1, \ldots, n \tag{4.38}$$

where ξ_i are slack variables. Thus, if data point x_i lies on the false side of the hyperplane, then an error has occurred and the corresponding slack variable ξ_i is increased to be greater than unity. Therefore, the sum of the slack variables $\sum_{i=1}^{n} \xi_i$ represents the upper bound on the number of training errors [11, 12, 21]. Hence, the objective function also has to be changed by adding a cost function of the errors: $\frac{\|\mathbf{w}\|^2}{2} + C\sum_{i=1}^{n} \xi_i$, where C is a user-given parameter. A higher value of C corresponds to a higher penalty to errors [11, 12, 21]. Nevertheless, the advantage of these modifications is that the same objective function in the dual Lagrangian is applicable. The constraints only change a little in comparison to the separable case and are as follows:

$$0 \leq \alpha_i \leq C \tag{4.39}$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \tag{4.40}$$

Combining everything, the mathematical program for the linear non-separable case can be stated:

<div align="center">

Mathematical Program 4.3:
SVM - Dual Problem Formulation - Non-Separable Case

</div>

$$\max \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{n} \alpha_i\alpha_k y_i y_k x_i x_k$$

$$\text{subject to} \quad 0 \leq \alpha_i \leq C \quad \forall i = 1, \ldots, n$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0$$

Furthermore, after solving Mathematical Program 4.3, the optimal normal $\mathbf{w}$ is computed with Formula 4.27. To determine the parameter b defining the distance between the origin and the optimal hyperplane, the KKT conditions of the primal Lagrangian are needed. The primal Lagrangian is given in the subsequent way [11, 12, 21]:

$$L_P = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i(y_i(x_i\mathbf{w}+b)-1+\xi_i) - \sum_{i=1}^{n}\upsilon_i\xi_i \qquad (4.41)$$

with υ_i are the additional Lagrange factors. The KKT conditions are as follows:

$$\frac{\partial}{\partial \mathbf{w_j}}L_P = \mathbf{w_j} - \sum_{i=1}^{n}\alpha_i y_i x_{ij} = 0 \qquad (4.42)$$

$$\frac{\partial}{\partial b}L_P = -\sum_{i=1}^{n}\alpha_i y_i = 0 \qquad (4.43)$$

$$\frac{\partial}{\partial \xi_i}L_P = C - \alpha_i - \upsilon_i \qquad (4.44)$$

$$y_i(x_i\mathbf{w}+b)-1+\xi_i \geq 0 \qquad (4.45)$$

$$\xi_i \geq 0 \qquad (4.46)$$

$$\alpha_i \geq 0 \qquad (4.47)$$

$$\upsilon_i \geq 0 \qquad (4.48)$$

$$\alpha_i(y_i(x_i\mathbf{w}+b)-1+\xi_i) = 0 \qquad (4.49)$$

$$\upsilon_i\xi_i = 0 \qquad (4.50)$$

Hence, the parameter b is determined with the help of one support vector by applying the following formula:

$$b = \frac{1-\xi_i}{y_i} - x_i\mathbf{w} \qquad (4.51)$$

In this way, all the necessary parameters, $\mathbf{w}$ and b, are defined and so the optimal hyperplane for the non-separable case is determined. Next, the generalization of SVM to the non-linear case is undertaken.

4.3.3   Non-linear SVM

Real world problems which can be linearly divided into two groups are a minority. The majority of real world problems are non-linear separable and in this way, a modification of the linear SVM has to be undertaken. The subsequent figure illustrates the non-linear case for 2 dimensions.
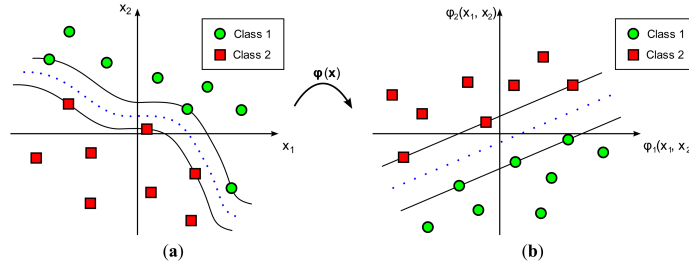
Fig. 4.5 Example of non-linear SVM [73]

The idea behind non-linear SVM is to map the data points into another space, the so-called feature space in which the data points can be divided into two groups with the help of linear SVM [11, 12, 21]. The feature space $\mathscr{H}$ usually has a higher dimensional as the initial space, $\mathbb{R}^n$, [11]. The considerations to use a higher dimensional space are that the probability of linear separability of the dataset is greater. Therefore, the following mapping function is introduced [12, 21]:

$$\begin{aligned} \Phi \quad & :\mathbb{R}^n \to \mathscr{H} \\ & x_i \mapsto \Phi(x_i) \end{aligned} \tag{4.52}$$

The advantage of SVM is that all the necessary computations between the data points to determine the optimal hyperplane are dot products. These dot products are replaced by the dot products of the mapped data points to define the non-linear SVM. The dot product of the mapped data points can be seen as a new function, called Kernel function. Nevertheless, the mapping function Φ is usually unknown and therefore another definition of the kernel function is used [92, 18].

Definition 18. *A kernel function $K(\cdot,\cdot)$ is a function which satisfies the Mercer's condition. The Mercer's condition says that there exists a mapping Φ and an expansion*

$$K(x_i, x_{i'}) = \sum_j \Phi(x_i)_j \Phi(x_{i'})_j$$

if and only if, for any $g(x_i)$ such that $\int g(x_i)^2 dx_i$ is finite then

$$\int K(x_i, x_{i'}) g(x_i) g(x_{i'}) dx_i dx_{i'} \geq 0.$$

The Mercer's condition is not easy to verify but there are several well-defined and widely used kernel functions [12].

Kernel	Formula
Polynomial	$K(x_i, x_{i'}) = (x_i x_{i'} + 1)^p$ with p degree of the polynomial
Radial Basic Function	$K(x_i, x_{i'}) = \frac{\exp(-\|x_i - x_{i'}\|^2)}{2\sigma^2}$ with σ a real number greater than 0
Hyperbolic	$K(x_i, x_{i'}) = \tanh(\kappa x_i x_{i'} - \delta)$ with κ and δ real numbers

Table 4.4 Example of Kernel functions

By substituting the dot products of the data points by the kernel functions in the dual Lagrangian 4.29, the modified form of the Lagrangian for the non-linear SVM is obtained:

$$L_D = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{n} \alpha_i \alpha_k y_i y_k K(x_i, x_k) \tag{4.53}$$

The constraints stated for the non-separable case remain the same for non-linear SVM. Thus, the optimal separating hyperplane is identify by the following mathematical program.

Mathematical Program 4.5
Non-linear SVM - Dual Problem Formulation

$$\max \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{n} \alpha_i \alpha_k y_i y_k K(x_i, x_k)$$
$$\text{subject to} \quad 0 \le \alpha_i \le C \quad \forall i = 1, \dots, n$$
$$\sum_{i=1}^{n} \alpha_i y_i = 0$$

The normal $\mathbf{w}$ and the distance to the origin b are determined in a similar way. Remember the formula to identify the normal $\mathbf{w}$:

$$\mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i \Phi x_i \tag{4.54}$$

The data points mapped into the higher dimensional feature space $\mathscr{H}$ have to be used because the hyperplane is defined in this space. In the same way, for one given data point x_i with $\alpha_i > 0$ the distance to the origin b is given by the formula:

$$b = \frac{1 - \xi_i}{y_i} - \Phi(x_i)\mathbf{w} \tag{4.55}$$

Integrating the definition of the normal **w** given by Formula 4.54, into the formula of b, the computation of b is reduced to the use of the Kernel function.

$$b = \frac{1 - \xi_i}{y_i} - \sum_{k=1}^{n} \alpha_k y_k K(x_i, x_k) \qquad (4.56)$$

After the distance to the origin is determined, the SVM is completely trained and is applicable to new data points $\tilde{x}_i$ with $i = 1, \ldots, \tilde{n}$. $\tilde{n}$ is the number of new data points. The classification of the new data points is made with the help of the following decision function [11, 12, 21]:

$$f(\tilde{x}_i) = \sum_{k=1}^{n} \alpha_k y_k K(x_k, \tilde{x}_i) + b \quad \forall i = 1, \ldots, \tilde{n} \qquad (4.57)$$

The sign of the decision function 4.57 for each new data point allows to define its group's membership. If the sign of the decision function is negative then the data point would belong to the group defined by $y_i = -1$ and in the same way, if the sign of the function is positive, then the data point would belong to the group characterized by $y_i = 1$.

Remember that all the explications and formulas given in these subsections are for binary SVMs, that means that they can only identify two different groups. Several modifications are undertaken to allow SVM the handling of multi-class problems [1, 3, 31, 94]. The main idea is to combine at least as many binary SVMs as possible groups exist. The two main methods to train a multi-class SVM are: one-against-all and one-against-one [25]. In one-against-all, one binary SVM is trained in the following way that one class corresponds to the first group and all the remaining classes represent the second group. Only as many binary groups are needed as there are groups. In contrast, in one-against-one, one binary SVM is trained by taking one class as the first group and a second class as the second group. So, using this method, the number of binary SVMs is exceeding the number of groups. The decision of the membership for a new data point in multi-class SVM is made by votes. For each binary SVM, a decision is obtained for the new data point and according to the decision, the respective group obtains a vote. Afterwards, the group with the maximal number of votes is the winning group and the new data points is affiliated to this group/class [25]. Next, the main benefits and drawbacks of SVM are discussed.

4.3.4   Benefits and drawbacks of SVM

Just as every other classification technique, SVM has advantages and disadvantages. First, the main benefits are briefly discussed before the major drawbacks are reviewed [12, 92]. The key benefit is the uniqueness of the solution. As the determination of the optimal

hyperplane is a convex optimization problem, the global optimum is always achieved. In contrast, remember that for ANN, only a local optimum is guaranteed. The achievement of the global optimum implies that the solution is unique [12, 92]. Another advantage is that SVM is a robust classifier. The robustness is obtained by the fact that SVM is less vulnerable to outliers than ANN and the risk of over-fitting is very small [12, 92]. A third benefit is the great flexibility of SVM. With the use of a kernel function, SVM is no longer limited dividing a dataset linearly into groups, but non-linear separators are usable as well [12, 92]. These non-linear separators are naturally linear separating hyperplanes in the higher dimensional feature space. Additionally, the mapping into this higher dimensional feature space is made implicitly with the help of the kernel function. Nevertheless, the theory of the kernel function - the Mercer's condition 18 is a good example for this - is robust. Thus, the robust theoretical basis represents another advantage [92, 18]. Finally, the last and main benefit of SVM is the fast evaluation of new data points. In the decision function only the support vectors, training points with $\alpha_i > 0$, are employed to classify a new data point into one of the possible groups [12, 92].

The key drawback concerning SVM is that the training and calibration is very slow. The long training time is due to the fact that the dot product or the kernel function of one data point with each of the remaining data points out of the training set has to be computed to determine the Lagrange factors α_i for $i = 1, \ldots, n$ which characterize the optimal hyperplane. Even if the black box effect is not so extensive than in the case of ANN, the partition, obtained by non-linear SVM of a dataset, is not easily explicable because the linear separation is obtained in the higher dimensional feature space. Hence, the feature space is usually unknown and therefore a relationship between the attributes of the data points and their group's membership is not feasible [12, 92]. This difficulty to explain the partition of the dataset is another disadvantage of SVM. Finally, the last main drawback is the complexity to integrate domain knowledge into SVM. The concept of SVM makes it almost impossible to incorporate domain knowledge into the training process [12, 92]. The training and calibration of SVM relies completely on the data points. However, in some specific cases, additional knowledge from an expert, the so called domain knowledge, is available. A user of SVM wants to include this knowledge to improve the accuracy and to correct the errors made during the training. Nevertheless, this incorporation is almost infeasible and therefore represents a big disadvantage of SVM [1, 12].

After this section has introduced support vectors machines and reviewed their main benefits and drawbacks, the next section will focus on a further development of SVM: support vector domain description (SVDD).

4.4 Support Vector Domain Description

Support vector domain description (SVDD) can be seen as a further step in the development of SVM. However, other influences, especially the description of the data and the detection of outliers, are also integrated in the development of a SVDD. First, the theory of SVDDs is explained. The section concludes with the advantages and disadvantages of SVDDs.

4.4.1 Theory and generalization of SVDD

Theory of SVDD

In SVDD, a closed boundary around the target dataset is determined. The closed boundary is represented by a hypersphere which is characterized by its center x^* and radius $R > 0$. The objective is to minimize the radius of the hypersphere and simultaneously to require the hypersphere to contain all or most of the training data points x_i for $i = 1, \ldots, n$ [86, 85]. The following figure illustrates the 2 dimensional case.
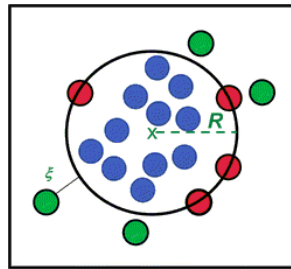
Fig. 4.6 Example of SVDD applied in 2D-space

A very large hypersphere to enclose a dataset due to some data points which are dissimilar in contrast to the remaining data points reduces the descriptive power of the hypersphere [86, 85]. To avoid a large hypersphere, some data points are allowed to be outside of the hypersphere. In the same analogy as for the linear non-separable case for SVM, slack variables ξ_i are introduced to handle these data points [85, 86]. The optimal hypersphere is the hypersphere with the minimal radius and enclosing most of the training points.

Mathematical Program 4.6 allows to determine this hypersphere. C is a user given constant which represents the trade-off between the volume of the hypersphere and the number of rejected data points [86, 85]. The radius of the hypersphere R is unknown and therefore Mathematical Program 4.6 is not solvable. In this way, the mathematical program has to be reformulated with the help of the Lagrange factors.

<div align="center">

Mathematical Program 4.6
SVDD - Primal Problem Formulation

</div>

$$\min \quad R^2 + C \sum_{i=1}^{n} \xi_i$$

$$\text{subject to} \quad \xi_i \geq 0 \quad \forall i = 1,\ldots,n$$

$$\|x_i - x^*\|^2 \leq R^2 + \xi_i \quad \forall i = 1,\ldots,n$$

Incorporating the constraints into the objective function of the mathematical program, the primal Lagrangian is obtained:

$$L_P = R^2 + C \sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} \alpha_i \left(R^2 + \xi_i - \|x_i\|^2 - 2x_i x^* + \|x^*\|^2 \right) - \sum_{i=1}^{n} \upsilon_i \xi_i \tag{4.58}$$

with $\alpha_i \geq 0$ and $\upsilon_i \geq 0$ Lagrange factors. By taking the partial derivatives according to all the unknowns variables, i.e. the radius and center of the hypersphere as well as the slack variables, the new constraints are identified:

$$\frac{\partial}{\partial R} L_P = 0 \Leftrightarrow \sum_{i=1}^{n} \alpha_i = 1 \tag{4.59}$$

$$\frac{\partial}{\partial x^*} L_P = 0 \Leftrightarrow x^* = \sum_{i=1}^{n} \alpha_i x_i \tag{4.60}$$

$$\frac{\partial}{\partial \xi_i} L_P = 0 \Leftrightarrow C - \alpha_i - \upsilon_i = 0 \tag{4.61}$$

Due to the fact that the Lagrange factors α_i and υ_i are positive, the third constraint 4.61 can be reformulated without the Lagrange factors υ_i in the following way [86, 85]:

$$0 \leq \alpha_i \leq C \quad \forall i = 1,\ldots,n \tag{4.62}$$

Using the obtained information from the constraints to rewrite the primal Lagrangian and to obtain its dual form, the dual Langrangian is given in the subsequent formula:

$$L_D = \sum_{i=1}^{n} \alpha_i x_i x_i - \sum_{i=1}^{n} \sum_{k=1}^{n} \alpha_i \alpha_k x_i x_k \tag{4.63}$$

The dual Lagrangian is only composed by the data points. In this way, the dual Lagrangian is solvable and can be used to determine the optimal hypersphere. The following mathematical program allows to determine the optimal hypersphere.

Mathematical Program 4.7
SVDD - Dual Problem Formulation

$$\max \quad \sum_{i=1}^{n} \alpha_i x_i x_i - \sum_{i=1}^{n} \sum_{k=1}^{n} \alpha_i \alpha_k x_i x_k$$

$$\text{subject to} \quad \sum_{i=1}^{n} \alpha_i = 1$$

$$0 \le \alpha_i \le C \quad \forall i = 1, \dots, n$$

As SVDD is based on SVM, its optimization problem, given in Mathematical Program 4.7, is also convex. After the solving of the mathematical program, the support vectors are identified. All data points with $\alpha_i > 0$, is considered as a support vector. The support vectors are sufficient to define the hypersphere and in this way to describe the dataset. Nevertheless, the data point x_i, with $\alpha_i = C$, is considered to be a bounded support vector because it lies outside the hypersphere [85, 86]. This data point can be viewed as an outlier because its Lagrange factor α_i contradicts the third constraint 4.61. All the data points with $0 < \alpha_i < C$ are considered as unbound support vectors. With the help of the support vectors, the center x^* as well as the radius R of the obtained hypersphere are computed. According to the second constraints 4.60 the center of the hypersphere is a linear combination of the data points and the Lagrange factors and is identified as follows:

$$x^* = \sum_{i=1}^{n} \alpha_i x_i \tag{4.64}$$

As the data points, which are unbound support vectors, lie on the boundary of the hypersphere, the distance between the center of the hypersphere and one unbound support vector is equal to its radius R. Let the data point x_k be an unbound support vector, then the radius of the hypersphere is given by the following formula:

$$R^2 = x_k x_k - 2 \sum_{i=1}^{n} \alpha_i x_i x_k + \sum_{i=1}^{n} \sum_{l=1}^{n} \alpha_i \alpha_l x_i x_l \tag{4.65}$$

Thus, the hypersphere is completely defined and the analysis of new data points can be undertaken. The distance between one new data point $\tilde{x}_i$ and the hypersphere's center is compared to the hypersphere's radius. The decision function reads as follows [86, 85]:

$$f(\tilde{x}_i) = \tilde{x}_i \tilde{x}_i - 2 \sum_{i=1}^{n} \alpha_i x_i \tilde{x}_i + \sum_{i=1}^{n} \sum_{l=1}^{n} \alpha_i \alpha_l x_i x_l - R^2 \tag{4.66}$$

The interpretation of the decision function 4.66 is similar to those for SVM. If the sign of the decision function is negative then the new data point belongs to the described class by the training set. Otherwise, if the sign of the decision function is positive, then the new data point is outside the hypersphere and so it is an outlier.

The described method only computes a hypersphere around the data points if they are already spherically distributed even if the most outlying data points are ignored. However, in real world problems, such datasets do not exist and therefore this method can not be applied to accurately detect outliers. Next, the necessary modification of the stated method is given.

Generalization of SVDD

The generalization of SVDD is achieved in two main steps. First, the description power of SVDD becomes more flexible and the employed datasets are not restricted to spherically regrouped data points. Second, SVDD is extended to handle also multi-class problems. At the moment, SVDD is only limited to describe one single group/class and all the remaining data points are considered as outliers. The remaining data points are not further investigated.

Starting with the more flexible description power of SVDD. As SVDD is primarily based on the idea of SVM, the same advantages apply. Thus, the advantage of writing the problem into the dual Lagrangian is that only dot products of the data points are used. To obtain a more flexible description, the data points are mapped into a higher dimensional feature space $\mathcal{H}$. In $\mathcal{H}$, it is possible to describe the mapped dataset as a hypersphere [86, 85]. The dot products of the mapped data points are regrouped into a kernel function. The dual Lagrangian 4.63 is reformulated with the help of the kernel function:

$$L_D = \sum_{i=1}^{n} \alpha_i K(x_i, x_i) - \sum_{i=1}^{n} \sum_{k=1}^{n} \alpha_i \alpha_k K(x_i, x_k) \tag{4.67}$$

The constraints concerning the variables α_i remain the same as for the previous case [86, 85]. Mathematical Program 4.8 has to be solved to obtain the hypersphere in $\mathcal{H}$ [86, 85]. The formula of the radius as well as the decision function have to be modified by integrating the kernel function [86, 85]. Let x_k be a support vector. The radius is computed as follows:

$$R^2 = K(x_k, x_k) - 2\sum_{i=1}^{n} \alpha_i K(x_i, x_k) + \sum_{i=1}^{n} \sum_{l=1}^{n} \alpha_i \alpha_l K(, x_i x_l) \tag{4.68}$$

<div align="center">

Mathematical Program 4.8
Kernelized SVDD - Dual Problem Formulation

</div>

$$\max \quad \sum_{i=1}^{n} \alpha_i K(x_i, x_i) - \sum_{i=1}^{n} \sum_{k=1}^{n} \alpha_i \alpha_k K(x_i, x_k)$$

$$\text{subject to} \quad \sum_{i=1}^{n} \alpha_i = 1$$

$$0 \leq \alpha_i \leq C \quad \forall i = 1, \dots, n$$

Let $\tilde{x}_i$ be a new data point, then the decision function is given in the following way:

$$f(\tilde{x}_i) = K(\tilde{x}_i, \tilde{x}_i) - 2 \sum_{i=1}^{n} \alpha_i K(x_i, \tilde{x}_i) + \sum_{i=1}^{n} \sum_{l=1}^{n} \alpha_i \alpha_l K(x_i, x_l) - R^2 \quad (4.69)$$

If the sign of the decision function is negative then the new data point $\tilde{x}_i$ belongs to the hypersphere and receives the description of the class. Otherwise, the new data point is considered as an outlier.

The most used kernel function is the radial basic function also called Gaussian kernel and is given in the subsequent formula [86, 85]:

$$K(x_i, x_{i'}) = \frac{\exp(-\|x_i - x_{i'}\|^2)}{2\sigma^2} \quad (4.70)$$

with σ being the kernel factor; a user given constant. This kernel function is widely used because of one main property: $K(x_i, x_i) = 1$. Hence, the decision function can be rewritten as follows:

$$f(\tilde{x}_i) = 1 - 2 \sum_{i=1}^{n} \alpha_i K(x_i, \tilde{x}_i) + \sum_{i=1}^{n} \sum_{l=1}^{n} \alpha_i \alpha_l K(x_i, x_l) - R^2 \quad (4.71)$$

with the reformulated radius:

$$R^2 = 1 - 2 \sum_{i=1}^{n} \alpha_i K(x_i, x_k) + \sum_{i=1}^{n} \sum_{l=1}^{n} \alpha_i \alpha_l K(, x_i x_l) \quad (4.72)$$

The next step of generalization is to expand SVDD to enable the handling of multi-class problems. SVDD when applied in supervised cluster analysis, implies that a dataset X has to be described by several groups/classes. The different groups are defined by the dataset as each data point also has an attribute expressing its membership to a group. In this way, for each group, a hypersphere is determined. Let G be the number of different groups and $\tilde{x}_i$ a new data point without any group membership. Then, a widely used decision function to

determine the group membership of $\tilde{x}_i$ is given as follows [86, 51]:

$$f(\tilde{x}_i) = \arg \min_{g=1,\dots,G} K(\tilde{x}_i, \tilde{x}_i) - 2 * \sum_{i=1}^{n_g} \alpha_i K(x_i, \tilde{x}_i) + \sum_{i=1}^{n_g} \sum_{l=1}^{n_g} \alpha_i \alpha_l K(x_i, x_l) - R_g^2 \qquad (4.73)$$

with n_g number of data points in group g and R_g the radius of the hypersphere describing group g, for $g = 1, \dots, G$. Nevertheless, using this decision function, the following problem can occur: if a data point $\tilde{x}_i$ does not belong to any of the determined hyperspheres and the distances from $\tilde{x}_i$ to two different centers is equal, then no explicit assignment can be made [40]. This situation is illustrated in the subsequent figure.

$d(\tilde{x}_i, G1) \quad \tilde{x}_i \quad d(\tilde{x}_i, G2)$

Group 1   Group 2
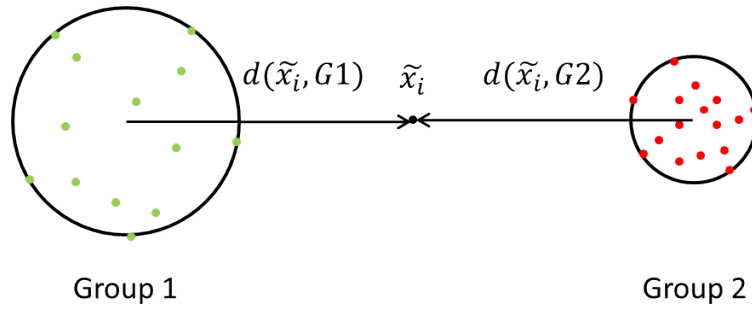
Fig. 4.7 Illustration of the problem of an equal distant data point to two groups

Let $d(\tilde{x}_i, G_g)$ be the distance between the data point $\tilde{x}_i$ and the center of the hypersphere describing group g. The mentioned problem can be overcome by taking into account the distribution of the groups. Taking the example shown in figure 4.7, group 1 is sparser than group 2 and in this way, the data point $\tilde{x}_i$ belongs rather to group 1 than to group 2. Therefore, a new decision function is introduced by integrating the distribution of the groups [40]:

$$f(\tilde{x}_i) = \arg \min_{g=1,\dots,G} \varphi(\tilde{x}_i, g) \qquad (4.74)$$

where $\varphi(\tilde{x}_i, g)$ is computed in the following way:

$$\varphi(\tilde{x}_i, g) = \begin{cases} \lambda \left(\dfrac{1 - \frac{d(\tilde{x}_i, G_g)}{R_g}}{1 + \frac{d(\tilde{x}_i, G_g)}{R_g}} \right) + \gamma & \text{if } 0 \le d(\tilde{x}_i, G_g) \le R_g \\[4mm] \gamma \left(\dfrac{R_g}{d(\tilde{x}_i, G_g)} \right) & \text{if } d(\tilde{x}_i, G_g) \ge R_g \end{cases} \qquad (4.75)$$

with $\lambda + \gamma = 1$. Thus, γ is defined by the value of λ which is a user-given regulation constant. Suggested values for λ lie in the range of 0.8 and 1 [40]. With the help of this decision function, each new data point $\tilde{x}_i$ can be assigned to exactly one group. In the next subsection, the advantages and disadvantages of SVDD are discussed.

4.4.2   Benefits and drawbacks of SVDD

As SVDD is based on the idea of SVM, it also inherits some of the drawbacks of SVM. A first disadvantage is that the training phase is very slow due to the fact that each data point is computed against the remaining of the training set X with the help of dot products or the kernel function. A second drawback is the incorporation of knowledge into SVDD because the concept of SVDD and SVM are similar [1, 12, 85, 86]. Finally, another drawback is that if the polynomial kernel function is employed, then SVDD is unable to define a tight description of the data [85, 86].

Naturally, SVDD inherits the main advantages of SVM. First, undertaking an analysis of a dataset with the help of SVDD, no probability density estimation is conducted but the description of the groups is made with the help of the hypersphere, fully described by the data points [85, 86]. Second, as the hypersphere is completely defined by the support vectors, i.e., data points with $0 < \alpha_i < C$, the affiliation of new data points is fast. New data points have only to be compared to the support vectors by dot products or a kernel function to determine their group membership [85, 86]. Third, with the use of a kernel function, the description defined by SVDD becomes flexible and is not limited to a sphere in the initial space of the data points. The hypersphere determination in the higher dimensional feature space allows to define complex boundaries in the initial space and so the description is better adopted to the data [85, 86]. Fourth, the parameter C indicating the trade-off between the radius of the hypersphere and the rejected data points out of the training set, does not really have a great influence on the finding of a good solution [85, 86]. Fifth, the error of the target group can be estimated immediately by calculating the fraction of data points which become support vectors. Using the RBF kernel function, which is the best performing kernel for SVDD, the kernel factor σ can be set in such a way that the fraction of support vectors equals to the error [85, 86]. Finally, the key advantage is that the solving of the mathematical problem determining the hypersphere either in the initial or the feature space, results in a unique solution. The achievement of the global optimum and in this way the determination of the hypersphere with the optimal radius is thus guaranteed [85, 86].

4.5   Conclusion

This chapter introduces supervised cluster analysis, also called classification. First, the general understanding of supervised cluster analysis is given. After the introduction to classification, three main methods are explained: artificial neural networks, support vector machines and support vector domain description. A detailed description of each method is furnished such that their implementation into any programming language and their application

are easily feasible. Different classification tasks, from the domain of medicine to water management, can be solved with these methods [6, 14, 27, 53, 100].

Nevertheless, the focus of this thesis lies on credit ratings and especially the determination or the prediction of credit ratings of bonds. Therefore, in the next chapter, three different automated credit rating prediction (ACRP) models are presented. One of the three introduced classification methods is employed in the ACRP models. Additionally, these three ACRP models allow to give an overview of the work done in the field of credit rating prediction.

Chapter 5

Automated credit rating prediction models from the literature

In this chapter, the focus lies on the existing automated credit rating prediction (ACRP) models described in the literature. This focus on the stability of credit ratings issued by rating agencies like Moody's has divulged a main problem during the financial crisis when several ratings did not agree with the real financial situation. This is the main reason why the necessity of ACRP models emerges. First, the main notations are introduced and afterwards three different ACRP models representing the main work in this field are described. The approach of each model is explained and their drawbacks as well as their advantages are discussed.

5.1   Notations

The following list introduces the main notations used in this chapter and the subsequent ones.

- G: number of rating groups

- D: number of rating degrees

- m: number of attributes

- Ω_d^D: rating degree d, with $d = 1, \ldots, D$

- Ω_g^G: rating group g, with $g = 1, \ldots, G$

- δ: distance between two rating degrees

Rated Bonds

- n: number of rated bonds

- B_i: rated bond i, with $i = 1, \ldots, n$

- $\mathbf{B}$: set of rated bonds, $\{B_1, \ldots, B_n\}$

- a_{ij}: attribute value j of bond B_i, with $i = 1, \ldots, n$ and $j = 1, \ldots, m$

- A_i: attribute vector of bond B_i, with $i = 1, \ldots, n$

- $\mathbf{A}$: matrix of the rated bonds' attributes

- R_i^D: given encoded rating degree of B_i

- R_i^G: rating group of B_i based on R_i^D

- R_i^C: rating characteristic of B_i based on R_i^D with $R_i^C = 1$ investment grade and $R_i^C = 0$ non-investment grade

Unrated Bonds

- $\tilde{n}$: number of unrated bonds

- $\tilde{B}_i$: unrated bond i, with $i = 1, \ldots, \tilde{n}$

- $\tilde{\mathbf{B}}$: set of unrated bonds, $\{\tilde{B}_1, \ldots, \tilde{B}_{\tilde{n}}\}$

- $\tilde{a}_{ij}$: attribute value j of bond $\tilde{B}_i$, with $i = 1, \ldots, \tilde{n}$ and $j = 1, \ldots, m$

- $\tilde{A}_i$: attribute vector of bond $\tilde{B}_i$, with $i = 1, \ldots, \tilde{n}$

- $\tilde{\mathbf{A}}$: matrix of the unrated bonds' attributes

- $\tilde{R}_i^D$: predicted rating degree of $\tilde{B}_i$

- $\tilde{R}_i^G$: predicted rating group of $\tilde{B}_i$

- $\tilde{R}_i^C$: rating characteristic of $\tilde{B}_i$ with $\tilde{R}_i^C = 1$ investment grade and $\tilde{R}_i^C = 0$ non-investment grade

5.2 Modelling sovereign credit ratings: Neural networks versus ordered probit [7]

In [7], the prediction of credit ratings of sovereign bonds is analyzed. The access to international capital markets and the terms of that access for the different countries can be determined with the help of sovereign credit ratings. According to the rating guide of Moody's [90], attributes describing the capacity and willingness of sovereign issuers to service their debts are selected. Political indicators are not used as attributes but are indirectly included in the other attributes. The political stability is taken into account during the evaluation of the creditworthiness of a country [57]. The employed attributes are given in the following table.

	Name	Description
a_{i1}	External Debt / Export	Total external debt relative to exports for the previous year
a_{i2}	Fiscal Balance	Average annual central government deficit or surplus relative to gross domestic product (GDP) for the previous three years
a_{i3}	External Balance	Average annual current account balance relative to GDP for the previous three years
a_{i4}	Rate of Inflation	Average annual consumer price inflation rate for the previous three years
a_{i5}	GDP per Capita	GDP for the previous year
a_{i6}	GDP Growth	Average annual real GDP growth on a year-over-year basis for the previous four years
a_{i7}	Development Indicator	International Monetary Fund (IMF) country classification for the current year (1 = industrial, 0 = not industrial)
a_{i8}	Standard & Poor's	1 = rating assigned by this agency, 0 = no rating obtained
a_{i9}	Moody's Investor Service	1 = rating assigned by this agency, 0 = no rating obtained
a_{i10}	IBCA (now Fitch Investors Service)	1 = rating assigned by this agency, 0 = no rating obtained
a_{i11}	Thomson Bank Watch	1 = rating assigned by this agency, 0 = no rating obtained
a_{i12}	Duff and Phelps	1 = rating assigned by this agency, 0 = no rating obtained
a_{i13}	Fitch Investors Service	1 = rating assigned by this agency, 0 = no rating obtained

a_{i14}	Japanese Bond Research Institute, Nippon Investors Service and Japan Credit Rating Agency	1 = rating assigned by these agencies, 0 = no rating obtained
a_{i15}	Dominion Bond Rating Service and Canadian Bond Rating Service	1 = rating assigned by these agencies, 0 = no rating obtained
a_{i16}	Africa / Middle East	1 = country being rated is in this region, 0 = otherwise
a_{i17}	Asia / Pacific Central	1 = country being rated is in this region, 0 = otherwise
a_{i18}	Eastern Europe	1 = country being rated is in this region, 0 = otherwise
a_{i19}	Latin America	1 = country being rated is in this region, 0 = otherwise
a_{i20}	Western Europe and North America	1 = country being rated is in this region, 0 = otherwise

Table 5.1 Employed attributes and their description

Thus, 20 attributes are used to predict the rating degree of a country and in this way of a bond. Furthermore, the attributes can be divided into three main categories. The first seven attributes ($a_{i1} - a_{i7}$) describe the financial situation of the evaluated country. Thus, these attributes are from the macroeconomic category. The relationship between these attributes and the rating degree is given by [89]. Lower rating degrees are observed in combination with higher levels of external debt, higher rates of inflation and a history of default on foreign currency debt. In the same logic, higher levels of fiscal and external balance, higher levels of income, higher rates of GDP growth and being classified as an industrial country by the International Monetary Fund (IMF) result in higher rating degrees.

The next eight attributes ($a_{i8} - a_{i15}$) are agency indicators and state which rating agency has rated the issuing country. Additionally, the last five attributes ($a_{i16} - a_{i20}$) indicate the region in which the country is located. These two sets of attributes are important to discover the inter-agency rating behavior [89]. A look at the chosen attributes reveals that it is not the creditworthiness of the bonds that is investigated but the financial situation of the issuing states to determine the rating degrees.

The developed ACRP model is based on artificial neural networks (ANN). Different configurations of networks are explored to define the best configuration. The analyzed configurations include the multi-layer perception, generalized feed-forward, radial basis function and modular networks. A detailed description of the different types of networks is given in [42]. At the beginning of the examination, the radial basis function network and the modular network are eliminated because of difficulties in the convergence [7]. The two remaining types of net-

works, multi-layer perception and generalized feed-forward, have a quite similar architecture. The same notation to describe the different components of a neural network, as in Chapter 4, are used. The complexity to determine the number of hidden neurons *Hidden_Neurons* and the number of hidden layers in a first step, remains. In a further step, the learning rate and the momentum have to be selected. Nevertheless, it was shown that more than one hidden layer does not improve the performance of an ANN [46]. Thus, only one hidden layer is utilized and the number of hidden neurons has to be determined. According to the rule of thumb suggested by [70], the different values for the learning rates and momentum rates are identified. The learning rate of the connections between the input and hidden layer λ_{hid} equals 1.0 and the rate between the hidden and output layer λ_{out} corresponds to 0.1. The values of the two momentum rates, μ_{hid} and μ_{out}, are the same and are equal to 0.7. The number of maximal iterations i_{max} is taken out of the following set: $\{1000, 2000, 3000, 4000, 5000\}$. After several tests with different numbers of hidden neurons, the generalized feed-forward network has shown the best performance out of the two remaining types of networks. Hence, the generalized feed-forward network is the basis of the ACRP model by [7]. This type of network is quite identical to the ANN introduced in Chapter 4, given by the pseudo-code 1. Two different models are developed: one based on classification and another based on regression. In the classification layout, the output layer contains as many neurons as rating degrees are used. In [7], 16 different rating degrees are used based on the credit ratings in the range of *AAA* to *Ba3*. In contrast, in the regression layout, the output layer is given by one single neuron which produces a number in the range $[0, 16]$. In this case, the rating degree is estimated by taking the integer part of the output and adding 1 [7]. The optimal number of hidden neurons is determined by trial and error and is 22 for the classification model and 25 for the regression model. Before discussing the accuracies of the two models, the ordered probit model is explained. This type of model is used as benchmark in the conducted empirical analysis by [7].

Ordered probit is a generalization of the popular probit method. The probit method is a type of regression where the output can only take two different values, like 0 and 1 [38]. In ordered probit, the method is generalized in such a way that the output can represent several groups / classes [38]. The different classes are differentiated by well-defined threshold values. For example, the 16 different rating degrees can be expressed with the help of 15 threshold values, ε_i with $i = 1, \ldots, 15$.

Let $Out(i)$ be the output of the ordered probit model, then the following formula can be used to predict the rating degrees of the bonds:

$$\tilde{R}_i^D = \begin{cases} 1 & 0 \leq Out(i) \leq \varepsilon_1 \\ 2 & \varepsilon_1 < Out(i) \leq \varepsilon_2 \\ \vdots & \\ 16 & Out(i) > \varepsilon_{15} \end{cases} \quad \forall i = 1, \ldots, \tilde{n} \tag{5.1}$$

In [89], a model is described to predict rating degrees for sovereign bonds. At this moment, it has been the most appropriate technique to develop an ACRP model. Additionally, the ordered probit model is deterministic. In this way the same output is always produced [89].

The accuracies of the introduced ACRP models are determined with different performance criteria. These criteria are [7]:

- Percentage of correct classifications

- Percentage of correctly classified within one rating degree

- Percentage of correctly classified within two rating degrees

- Percentage of correctly classified within three rating degrees

The subsequent table summarizes the obtained accuracies.

Performance criterion	Ordered probit model	ANN classification model	ANN regression model
% correctly classified	31.8	40.4	34.6
% correct within one degree	59.2	63.6	68.9
% correct within two degrees	77.1	80.4	87.3
% correct within three degrees	89.8	87.6	96.7

Table 5.2 Accuracies of the different ACRP models [7]

The table shows that the ACRP model based on neural networks outperforms in each category the ordered probit model. Nevertheless, only if the predicted rating degree is allowed to differ by at least one degree from the correct rating degree then the ACRP model is effectively applicable and has an accuracy of at least 63.6%. The accuracy of the analyzed ACRP models does not exceed 40.4% if the correct rating degree has to be predicted.

To conclude this section, the drawbacks and advantages of the described ACRP model are given. First, the ACRP model demonstrates that the prediction of rating degrees of sovereign bonds with the help of publicly available information is feasible. Second, according to the shown accuracies, the ACRP model is not usable by the public but can be seen as a tool to support analysts in the decision taking process. This decision taking support by the ACRP model makes the rating process more objective and more transparent. The main drawback of the model is that the prediction is based on an analysis of the financial situation of the issuing countries and not on the characteristics describing the creditworthiness of the bonds. The focus on the issuing countries allows to obtain trusted information from third party sources like the IMF, the Organization of Economic Cooperation and Development (OECD) and the International Financial Statistics. However, the information provided by these parties is not always up to date because it is collected only once a year [101]. Hence, the current financial situation is not really taken into consideration if the analysis of the credit risk is undertaken. Another disadvantage is that the necessary information is not easily accessible because different sources have to be explored to collect all the necessary data. If the bonds' characteristics were used directly to predict their rating degrees, this would have two main advantages. First, the information could be retrieved directly from the financial market and in this way they would be easily accessible. Second, the received information would always be up to date and contain the latest news on the financial health of the bonds [101].

5.3 Credit rating analysis with support vector machines and neural networks: a market comparative study [48]

In [48], credit ratings of corporate bonds constitute the focus. The obtaining of a credit rating usually means additional costs for the company. Companies have to contact a rating agency and pay a fee to cover the costs of the agency. After signing a contract with a rating agency, the agency assigns a financial risk analyst to undertake a deep analysis of the company's risk status and finally the rating committee of the agency decides about the assignment of a credit rating to the company [97, 69]. To offer companies a more cost-competitive alternative, ACRP models have been developed. Nevertheless, the development of ACRP models for corporate bonds usually makes the assumption that the used attributes extracted from public financial statements, such as financial ratios, contain the necessary information about a company's credit risk. Furthermore, the selected attributes are often combined with historical rating degrees given by the rating agencies to extract the expertise of the agencies

in evaluating companies' credit risk in the best way [48]. ACRP models are typically used to help users to better understand the bond-rating process as the subjective component is eliminated from the process. Typical financial information furnished by a company to a rating agency are: annual reports from the past years, the latest quarterly reports, income statement and balance sheet, most recent prospectus for debt issues and other statistical reports [97, 69]. In [48], the subsequent table introduced the 21 selected attributes.

	Name / Description
a_{i1}	Total assets
a_{i2}	Total liabilities
a_{i3}	Long-term debts / total invested capital
a_{i4}	Debt ratio
a_{i5}	Current ratio
a_{i6}	Times interest earned (EBIT / interest)
a_{i7}	Operating profit margin
a_{i8}	(Shareholders' equity + long-term debt) / fixed assets
a_{i9}	Quick ratio
a_{i10}	Return on total assets
a_{i11}	Return on equity
a_{i12}	Operating income / received capitals
a_{i13}	Net income before tax / receive capitals
a_{i14}	Net profit margin
a_{i15}	Earnings per share
a_{i16}	Gross profit margin
a_{i17}	Non-operating income / sales
a_{i18}	Net income before tax / sales
a_{i19}	Cash flow from operating activities / current liabilities
a_{i20}	Cash flow from operating activities / (capital expenditures + increased in inventory + cash dividends) for the last 5 years
a_{i21}	(Cash flow from operating activities - cash dividends) / (fixed assets + other assets + working capital)

Table 5.3 Employed attributes by [48]

Two sets of companies are used to undertake empirical tests of the developed ACRP model which is based on support vector machines (SVM). The first sets contain 74 companies from Taiwan and the second set includes 265 corporations from the United States.

Out of the whole 21 available attributes the following four sets are created:

- set TW1 uses the following attributes: a_{i1}, a_{i2}, a_{i3}, a_{i4}, a_{i6} and a_{i7}

- set TW2 utilizes the following attributes: a_{i1}, a_{i2}, a_{i3}, a_{i4}, a_{i6}, a_{i7}, a_{i8}, a_{i10}, a_{i11}, a_{i12}, a_{i13}, a_{i14}, a_{i15}, a_{i16}, a_{i18} and a_{i21}

- set US1 employs the following attributes: a_{i1}, a_{i2}, a_{i3}, a_{i4} and a_{i7}

- set US2 uses the following attributes: a_{i1}, a_{i2}, a_{i3}, a_{i4}, a_{i7}, a_{i8}, a_{i10}, a_{i11}, a_{i12}, a_{i13}, a_{i14}, a_{i15}, a_{i16} and a_{i17}

Hence, TW1 and US1 represents simplified models of the market. In TW1 and US1, only a small number of attributes is used to predict the rating degrees of the bonds in contrast to the more complex models TW2 and US2. Furthermore, the differentiation into a simplified and a complex model makes it possible to determine if a small number of attributes is sufficient to predict the rating degrees of bonds or not. The financial information which represents the basis of the attributes is collected from the Securities and Futures Institute [79]. Next, the used ACRP model and its employed benchmark are explained.

The proposed ACRP model by [48] is based on SVM. The radial basis function is used as kernel function to map the attributes into a higher dimensional feature space $\mathscr{H}$ and to separate the bonds according to their credit risk. Remember, the radial basis function is as follows:

$$K(B_k, B_{k'}) = \exp(-\frac{\|A_k - A_{k'}\|^2}{2\sigma^2}) \quad \forall k, k' = 1, \ldots, n \tag{5.2}$$

The parameter C which represents the penalty if during the training process some bonds are classified on the wrong side of the hyperplane, is set to 1000. The kernel factor σ equals 4.47 [48]. According to [47], several methods are examined from one-against-one to more complex methods like directed acyclic graph (DAG) which is based on one-against-one. As in Chapter 4, a voting strategy is employed to determine the right group membership for each bond. Each group represents one rating degree and in total 5 different rating degrees are used [48]. For companies from Taiwan the following rating groups are used: *AAA*, *Aa*, *A*, *Baa* and *Ba*. The rating groups for corporations from the US are: *Aa*, *A*, *Baa*, *Ba* and *B*. The intermediate degrees are regrouped. For example, the rating group *Aa* is composed of the rating degrees *Aa*1, *Aa*2 and *Aa*3. The developed ACRP model is compared to a benchmark model based on ANN.

The ANN used in the benchmark ACRP model has the following configuration [48]. The network has one hidden layer and the back-propagation method is utilized to train the network. The network is similar to the one described in Chapter 4 and given by the pseudo-code 1. The number of hidden neurons *Hidden_Neurons* is fixed to the value: $\frac{m + Output\_Neurons}{2}$. The number of output neurons *Output_Neurons* is equal to the number of different existing rating groups i.e. 5. The number of attributes m varies from 5 to 16 according to the set that is employed. In [48], the values of the learning rates (λ_{hid} and λ_{out}) as well as of the momentum

rates (μ_{hid} and μ_{out}) are not given. The stated results can not directly be reproduced.

In the same way as for the prediction of rating degrees of sovereign bonds, the accuracies are determined with similar performance criteria [48]:

- Percentage of correct classification

- Percentage of correct classified within one rating group

The subsequent table illustrates the obtained accuracies for the ACRP model and the benchmark model.

Set	ACRP Model	Benchmark Model
TW1	79.73	75.68
TW2	77.03	75.68
US1	78.87	80.00
US2	80.00	79.25

Table 5.4 Accuracies for the ACRP model

A clear statement concerning the simplified and the complex model can not be made. For the Taiwanese companies, the simplified set, TW1, has given higher accuracies than the complex set, TW2. In contrast, for US companies, the complex set, US2, has produced the best accuracy, only the benchmark model performs slightly better on the simplified set, US1. Nevertheless, the difference, even for the Taiwanese companies, is so small that a set of 5 or 6 attributes is sufficient to predict the rating degrees of the bonds [48]. The percentage of correctly classified bonds within one rating group is only given for the ACRP model. This performance criterion is not applied to the benchmark model [48]. The following table indicates the obtained accuracies.

Set	Accuracy of the ACRP model
TW1	91.89
TW2	92.24
US1	97.74
US2	98.44

Table 5.5 Accuracies of the ACRP model within one rating degree

If the predicted rating group can differ by one group, the complex sets, TW2 and US2, produce slightly better results than the simplified sets, TW1 and US1. However, the difference is so small that even with the simplified sets good results are obtained. Furthermore, more than

90% are received which demonstrates the good performance of SVM based ACRP model.

The main benefit is that the developed ACRP model demonstrates that SVM can be used to
develop accurate ACRP models. Additionally, it is shown that the number of attributes can
be delimited to 5 or 6 well-determined attributes [48]. Furthermore, the use of SVM allows
to lift the veil of opacity a little bit because SVM is not such a black box method like ANN
[12, 92]. A main drawback of the analyzed ACRP model is that the credit risk analysis is
based on attributes describing the financial situation of the issuing companies. The same
comment as for the ACRP model in the previous section can be made, that the obtained
information is not really up to date at the moment of the credit risk analysis and does not
include the actual news about the financial health of the bonds. Additionally, particularly
the set of Taiwanese corporations is very small with 74 different companies, only the set of
US corporations is a little more significant due to the total number of 265 companies [48].
Furthermore, the same comment concerning the access to the needed information can be
reformulated. Even if all the attributes are collected from one single provider, in this case
Securities and Future Institute [79], the service is not free of charge. In this way, the ACRP
model with the proposed attributes can not be used by the public.

5.4 A Corporate Credit Rating Model Using Support Vector Domain Combined with Fuzzy Clustering Algorithm [40]

In [40], the goal is to predict the rating degrees of corporate bonds. A hybrid method is
used to develop the ACRP model: fuzzy clustering and support vector domain description.
The ACRP model and especially fuzzy clustering is explained after the used attributes are
introduced. According to the traded market, two sets of bonds are defined: a Korean dataset
and a Chinese dataset.The empirical analysis of these two sets is the basis for the evaluation
of the ACRP model. Table 5.6 describes the used attributes.
All the utilized attributes describe the financial situation of the issuing companies. In con-
trast to the employed attributes in [48], only economic variables are used as attributes [40].
Nevertheless, in [48] it was shown that the additional information about rating agencies and
geographical location does not significantly improve the accuracy of the prediction. There is
a difference between the Korean and Chinese dataset concerning the employed attributes. The
attributes, years after foundation a_{i6} and inventory assets to current assets a_{i11}, are excluded
in the Chinese dataset.

	Name	Description
a_{i1}	Shareholders' equity	A firm's total assets minus its total liabilities
a_{i2}	Sales	Sales
a_{i3}	Total debt	Total debt
a_{i4}	Sales per employee	Sales / number of employees
a_{i5}	Net income per shares	Net income / the number of issued shares
a_{i6}	Years after foundation	Years after foundation
a_{i7}	Gross earning to total asset	Gross earning / total asset
a_{i8}	Borrowings-dependency ratio	Interest cost / sales
a_{i9}	Financing cost to total cost	Financing cost / total cost
a_{i10}	Fixed ratio	Fixed assets / (total assets-debts)
a_{i11}	Inventory assets to current assets	Inventory assets / current assets
a_{i12}	Short-term borrowings to total borrowings	Short-term borrowing / total borrowing
a_{i13}	Cash flow to total assets	Cash flow / total assets
a_{i14}	Cash flow from operating activities	Cash flow from operating activities

Table 5.6 Used attributes and their descriptions [40]

The two datasets are divided into 4 different groups: $A1$, $A2$, $A3$ and $A4$. Thus, only four rating degrees are differentiated. The proposed ACRP model is designed to differ only rating groups which is a merging of rating degrees. The prediction of all possible existing rating degrees of the bonds is not feasible.

The first step of the ACRP model is the application of the fuzzy clustering algorithm (FCM) to reduce the dataset before determining the hypersphere with the help of support vector domain description (SVDD). Lets briefly explain the theory of FCM. The idea behind FCM is that no strict membership to a group / class is defined but for each group, the membership is determined [62]. For example, bond B_1 would be split into the four defined groups $A1$, $A2$, $A3$ and $A4$ and the following membership could be found: $v_1 = (0.13, 0.07, 0.67, 0.13)$. The membership vector v_1 indicates that bond B_1 belongs rather to group A_3. The example shows one main property of the membership vector which is that the sum of all components equals 1. Thus, for each bond B_i its membership vector v_i holds the following property: $\sum_{g=1}^{G} v_{gi} = 1$ for each $i = 1, \ldots, n$. The used method in the ACRP model is called kernel-based fuzzy attribute C-means clustering algorithm (KFAMC) which was introduced by [62]. First, the algorithm of KFAMC is given and then KFAMC is explained.

Algorithm 5 Kernel-based Fuzzy Attribute C-Means Clustering

1: **procedure** KFAMC
2: *Input:*
3:     **B** dataset
4:     t_{max} maximal number of iterations
5:     G number of groups
6:     γ weighting exponent, with $1 \leq l \leq 10$
7:     ε error criterion
8:     ζ constant, with $\zeta > 0$
9: *Output:*
10:     U matrix of membership
11: *Do:*
12:     Initialize the matrix of membership, U^0
13:     Iterate over t
    Update the cluster centroids and set of attribute measure and the weight matrix
        The centroids are updated with the following formula:

    (a)
$$B_g^{t+1} = \frac{\sum_{i=1}^{n} w((v_{gi}^t)^{\gamma/2}(1-K(B_i,B_g^t))^{1/2})((v_{gi}^t)^{\gamma}K(B_i,B_g^t))B_i}{\sum_{i=1}^{n} w((v_{gi}^t)^{\gamma/2}(1-K(B_i,B_g^t))^{1/2})((v_{gi}^t)^{\gamma}K(B_i,B_g^t))}$$

        The subsequent formula updates the component of the matrix of membership:

    (b)
$$v_{gi}^{t+1} = \frac{(w((v_{gi}^t)^{\gamma/2}(1-K(B_i,B_g^t))^{1/2}(1-K(B_i,B_g^{t+1})))^{-1/(\gamma-1)}}{\sum_{g=1}^{G}(w((v_{gi}^t)^{\gamma/2}(1-K(B_i,B_g^t))^{1/2}(1-K(B_i,B_g^{t+1})))^{-1/(\gamma-1)}}$$

        with $w(x) = \frac{1}{\zeta+x^2}$
        Compute the weighted objective function, Q^{t+1}, :
$$Q^{t+1} = \sum_{g=1}^{G}\sum_{i=1}^{n} w((\mu_{gi}^{t+1})^{\gamma/2}(1-K(B_i,B_g^{t+1}))^{1/2}((v_{gi}^{t+1})^{\gamma}(1-K(B_i,B_g^{t+1})))$$

14:     Stop the iteration if $|Q^t - Q^{t-1}| < \varepsilon$ or t_{max} is achieved
15:     Output the matrix of membership

KFAMC can also be applied at any other dataset X instead of the set of bonds **B**. The weighting exponent γ determines the level of the groups' fuzziness. In this way, larger value of γ results in smaller values of the components of the membership vector for each bond [62]. At the limit value $\gamma = 1$, the components' value converges to 0 and 1. In this way, the partition of the bonds becomes explicit. In practice $\gamma = 2$ is utilized [62]. The error criterion is normally fixed at the value of $1.0e^{-5}$ [62]. The set of rated bonds **B** is divided with the help of the KFAMC algorithm, which is given in the pseudo-code 5, into the existing rating groups defined by the given rating degrees. First, the centroids of the groups, B_g^t for $g = 1, \ldots, G$, and the matrix of membership U^0 are randomly initialized. Then, during an iterating, the centroids and the corresponding membership of each bond B_i to a group are updated with the given formulas 13(a) and 13(b). Finally, if the difference of the actual and previous value of objective function, Q^t and Q^{t-1}, is smaller than the error criterion ε then the algorithm stops and the set of bonds, is completely divided into the different rating

groups. If the error criterion can not be reached at the end of the iterations, e.g., the maximal number of iterations t_{max} then the algorithm also stops and outputs the matrix of membership. A bond has a membership to each rating group. In the ACRP model a threshold θ is included to exclude one part of the dataset [40]. The threshold is employed in the following way: if $v_{gi} \geq \theta$ then $v_{g'i} = \begin{cases} 0 & g' \neq g \\ 1 & g' = g \end{cases}$. The value of the threshold, θ, is set to 0.8 or 0.9 [40]. All the bonds with unique group affiliation are filtered out and all the other bonds are considered as possible support vectors, which are called representative bonds. The subsequent figure illustrates the idea behind the reduction of the dataset.
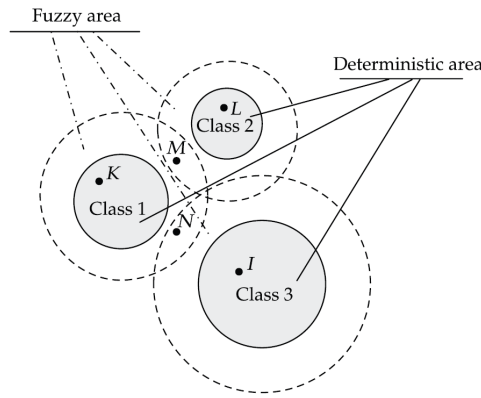

Fig. 5.1 Idea behind the reduction of the dataset [40]

Afterwards, for each rating group, the hypersphere describing the group is determined by taking into consideration the membership to the group. Let be $\mathbf{B}'$ the reduced set of rated bonds, then B'_i for $i = 1, \ldots, n'$ represents the bonds in this reduced set. The hypersphere for rating group Ω^G_g determined by the fuzzy version of support vector domain description (FSVDD) is determined by the following mathematical program:

<div align="center">

Mathematical Program 5.7
FSVDD - Dual Problem Formulation

</div>

$$\max \quad \sum_{i=1}^{ni} \alpha_i K(B'_i, B'_i) - \sum_{i=1}^{n'} \sum_{k=1}^{n'} \alpha_i \alpha_k K(B'_i, B'_k)$$

$$\text{subject to} \quad \sum_{i=1}^{n'} \alpha_i = 1$$

$$0 \leq \alpha_i \leq v_{gi} C \quad \forall i = 1, \ldots, n'$$

The user-given constant C is, as described in Chapter 4, the trade-off between the volume of the hypersphere and the number of bonds excluded of the hypersphere. The only difference with Mathematical Program 4.8 is that the value of the Lagrange factors α_i depends on the corresponding group's membership of bond B_i. The computation of the hyperspheres for the G groups are identical to the normal SVDD. Finally, the decision function 4.75 described in Chapter 4, is used to determine the final and unique group's affiliation. For new unrated bonds, only the comparison with the representative bonds during the computation of the decision function is needed and KFAMC is not applied any more. KFAMC is only executed during the training of the ACRP model to reduce the computation time [40].

The ACRP model of [40] is compared to benchmark models based on SVM (*ARCP_SVM*) and ANN (*ACRP_ANN*). The model *ACRP_SVM* is similar configured than the model described in the previous section and introduced by [48]. The configuration of *ACRP_ANN* is quite identical to the model introduced by [7]. The accuracies of the ARCP model and the benchmark models is defined as the percentage of correctly classified bonds [40]. The obtained accuracies are illustrated in the following table.

Model	Korean dataset	Chinese Dataset
proposed ACRP	72.12%	73.70%
ACRP_SVM	70.23%	71.26%
ACRP_ANN	62.78%	67.19%

Table 5.8 Accuracies of the different ACRP models

No information about any user-given parameters, excluding the threshold value θ, is made in [40]. Only the radial basis function is characterized as the kernel function. Nevertheless, the kernel factor σ is not defined as well as the constant C. Additionally, the used dataset is not publicly available and the authors have not given access to the dataset to other researchers. In this way, a reproduction of the furnished results is not feasible. A confirmation of their accuracy is undertaken with the help of the set of bonds employed in [48]. During this test, different values for the parameters are utilized but the results could not be reproduced [15]. The best detected accuracy equals 50% [15]. Therefore, the global performance of this ACRP model has to be contested.

The main benefit of the proposed ACRP model is that the computation time is reduced by applying the hybrid method: KFAMC and FSVDD. Furthermore, the hypersphere methodology seems to be a better description of rating groups than the division of the dataset by hyperplanes [40]. However, one main disadvantage is that the used attributes are again describing the financial situation of the issuing companies instead of the issued bonds. Additionally, the

ACRP model is only constructed to predict the rating groups of the bonds. Furthermore, the exact group composition is not given by [40]. Another main drawback is that the performance of the KFAMC depends extremely on the initialization of the initial centroids [62]. If the initial centroids are badly chosen then it could happen that the algorithm does not converge and can not output any reasonable membership matrix for the bonds. A unique initialization technique does not exist and so the process to initialize the centroids is a complex task [62].

5.5   Conclusion

In this chapter, three ACRP models from the literature are introduced. The three representative ACRP models are based on different methods: ANN, SVM and SVDD. The given results illustrate that the more recent methods, like SVDD and SVM, are more accurate in the prediction task than the older methods, like ANN. Except for the first presented ACRP model based on ANN [7], all the ACRP models are only able to predict the rating group of the bonds [48, 40]. The prediction of rating degrees is a more complicated task because the partition based on rating degrees is finer than the partition based on rating groups. The introduced ACRP models are always designed for only one specific type of bonds, corporate or sovereign. The majority is focused on corporate bonds [48, 40]. Companies could buy an ACRP model to estimate the financial situation of their main competitors. Additionally, the employed attributes are describing either the financial situation of the issuing states or corporations. In this way, the acquisition of the needed information is complicated and sometimes not free of charges. This Chapter has answered Research Question 1 because, excluding the datasets used in [40], all three ACRP models use attributes based on publicly available information to predict the rating degrees of bonds.

Chapter 6

Automated credit rating prediction model based on support vector domain description and linear regression

In this chapter, a new developed ACRP model[1] is presented. The idea behind the development of the ACRP model is to allow the simultaneous prediction of the rating degrees of corporate and sovereign bonds. The needed information should be publicly available and free of charges. Additionally, the information have to be easily accessible to private investors. First, the model is generally described and a small artificial example is given. Third, it is briefly introduced different field of research where the proposed model or part of it can be reused, even if the proposed model is explicitly developed under the use case of predicting rating degrees of bonds. Finally, a conclusion finalize the chapter.

6.1 Description of the new developed ACRP model

The ACRP model is based on an hybrid method: support vector domain description (SVDD) composed with linear regression (LR). First, the main steps of the model are briefly explained before the formal description is given in all the necessary detail such that a reproduction of the ACRP model is feasible. Finally, this section is completed with a small artificial example to illustrate the procedure of the model.

---

[1]The ACRP model has been presented at the International Conference on Control, Decision and Information Technologies and has been published in the proceedings [33]. The main idea behind the ACRP model has been delivered by myself. Robert Dochow has helped me to formulate the model in an understandable way and to write the paper.

6.1.1   Description of the main steps

The notations introduced in Chapter 5 remain valid. Before the new ACRP model can be really applied, some pre-processing of the dataset has to be made. Assume a rating scale always consists of $d = 1, \ldots, D$ origin ordinal credit ratings. For example, Moody's and S&P utilize a rating scale with $D = 21$ different credit ratings. The credit ratings have to be encoded into rating degrees with the help of the following encoding rule:

$$\Omega_d^D = 1 + (d-1)\delta \quad \forall d = 1, \ldots, D \tag{6.1}$$

Remember δ is the user-given value which determines the distance between two adjacent credit ratings. For example, the credit rating, $Aa2$, is the third entry of the rating scale used by Moody's. Thus, $d = 3$ represents the credit rating $Aa2$. Using $\delta = 0.25$, the corresponding rating degree would be: $\Omega_3^D = 1 + (3-1)0.25 = 1.5$. The attributes are standardized with the help of the min-max method. The min-max method is defined in the following way:

Definition 19. *Let a_{ij} be the attribute j of bond i with $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Then, attribute a_{ij} is standardized to a'_{ij} according to the min-max method:*

$$a'_{ij} = \frac{a_{ij} - \min_{i=1,\ldots,n} \left(a_{ij} \right)}{\max_{i=1,\ldots,n} \left(a_{ij} \right) - \min_{i=1,\ldots,n} \left(a_{ij} \right)} \quad \forall i = 1, \ldots, n \tag{6.2}$$

In the following, it is always talk of standardized attributes. The ACRP model requires the execution of five steps to determine the rating degrees of unrated bonds:

1. Building of rating groups

2. Calibration to the rating groups via SVDD

3. Calibration to the rating degrees via LR

4. Finding the rating group

5. Finding the rating degree

The steps 1-3 can be summarize as the model building and calibration using SVDD and LR. In these steps, only the training set $\mathbf{B}$ is employed. The prediction of the rating group as well as the rating degree of the unrated bonds out of the set $\tilde{\mathbf{B}}$ are described in the steps 4-5.

In Step 1, the rating groups are built. With the help of the set of rated bonds **B** several groups are defined based on their existing rating degrees. Based on the literature [58], the groups are used to undertake a cluster analysis to identify a finer segmentation of the bonds. This leads to a lower risk of false classification due to the fact of overlapped groups [58].

To access the new defined rating groups, SVDD is used to calibrate the groups. The calibration consists in defining for each rating group a hypersphere which encloses all the bonds belonging to this group. The objective is to determine the hypersphere with the minimal possible radius. The definition of the hypersphere is guarantee by the computation of a weight vector composed by the Lagrange factors (see Chapter 4). The bonds, characterized by the weight vectors, are called representative bonds. This process represents Step 2.

In Step 3, the objective is to expand the rating groups to all possible rating degrees. Therefore, the relationship between the bonds' attributes, $A_1, \ldots, A_n$, and their rating degrees, $R_1^D, \ldots, R_n^D$ is determined with the help of LR. For each rating group, the obtained regression factors are used to determine the rating degrees of the unrated bonds.

With the help of the decision function defined by the representative bonds, determined in Step 2, the unrated bonds are associated to one of the existing rating groups. In this way, the bonds, $\tilde{B}_1, \ldots, \tilde{B}_{\tilde{n}}$, obtain their unique rating group affiliation. This first classification of the unrated bonds represents Step 4.

Finally, in Step 5, the rating degrees of the unrated bonds are predicted. For each group, the regression factors are employed to determined the rating degrees based on the bonds' attributes.

6.1.2 Formal description of the new ACRP model

In this subsection, the different steps of the ACRP model are explained in detail by going through them.

Step 1: Building of the rating groups

In the set **B** each bond B_i with $= i, \ldots, n$ is characterized by its vector of attributes A_i and its rating degree R_i^D. The form of the attributes' vector is as follows: $A_i = (a_{i1}, \ldots, a_{im})$. The rating degrees are aggregated to G rating groups. G is always set smaller than D to avoid the risk of miss-classification due to overlapping [58]. For example, a possible group building is

to split the set of bonds into $G = 3$ rating groups: prime to high grade, upper medium grade and below lower medium grade. The rating group R_i^G is the group affiliation of bond B_i. In other words, R_i^G can be considered as a simplified rating degree of bond B_i. The obtained rating groups are part of the input of Step 2.

Step 2: Calibration of the rating groups via SVDD

Even if the mathematical programs for SVDD are introduced in Chapter 4, in this section the formulas are rewritten to adopt them to the used notations. To be able to determine the group affiliation of unrated bonds, the representative bonds for each group have to be defined. The objective of SVDD consists to determine a hypersphere with the minimal possible radius which describes one rating group. The hypersphere is characterized by its center B_g^* with the corresponding attributes' vector A_g^* and its radius R_g, for each $g = 1, \ldots, G$. Thus, for each rating group Ω_g^G, the optimal hypersphere is searched. The primary problem of SVDD is given as follows:

<div align="center">

Mathematical Program 6.1
SVDD- Primary Problem Formulation

</div>

$$
\begin{aligned}
&\min && R_g^2 + C \sum_{k=1}^{n_g} \xi_{g,k} \\
&\text{subject to} && \xi_{g,k} \geq 0 \quad \forall k = 1, \ldots, n_g \\
& && R_g^2 \geq \|A_k - A_g^*\|^2 - \xi_{g,k} \quad \forall k = 1, \ldots, n_g
\end{aligned}
$$

with n_g the number of bonds belonging to rating group Ω_g^G. $\xi_{g,k}$ are slack variables to take into account possible outliers which remains outside the hypersphere. C represents a user-given constant which indicates the penalization of ignoring outliers in the determination of the minimal radius.

Nevertheless, the analyzed datasets are not spherically distributed, even if the most outlying bonds are ignored. In this way, the generalized version of SVDD is used, see Chapter 4, Section 4. The kernel function allows to implicitly map the bonds' attributes into some higher dimensional feature space. With the choose of a suitable feature space, a spherical distribution of the data can be assumed. The new ACRP model utilizes the radial basic function as kernel function. The employed kernel function is defined in 5.2. For each rating group, the kernel matrix P_g is computed in the following way:

$$
P_g(k, k') = \exp \left(-\frac{\sum_{j=i}^{m} \left(a_{kj} - a_{k'j} \right)^2}{2\sigma^2} \right) \quad \forall k, k' = 1, \ldots, n_g \tag{6.3}
$$

with σ_g the kernel factor which represents a user-given constant. Instead of using α_k according to the notation used in Chapter 4, $W_{g,k}$ is introduced as the Lagrange factor and represents the entries of the searched weight vector W_g. The optimal hypersphere is determined with the help of the following mathematical program.

<div align="center">

Mathematical Program 6.2
SVDD - Dual Problem Formulation

</div>

$$\text{max} \qquad L_g = 1 - \sum_{k=1}^{n_g} \sum_{k'=1}^{n_g} W_{g,k} W_{g,k'} P_g(k,k')$$

$$\text{subject to} \quad \sum_{k=1}^{n_g} W_{g,k} = 1$$

$$0 \leq W_{g,k} \leq C \quad \forall k = 1,\ldots,n_g$$

Each bond with non-zero weight factor, $W_{g,k} > 0$, is called as representative bonds. Only the representative bonds are needed to characterize the hypersphere as well as to determine the group affiliation of the unrated bonds. Furthermore, the objective function L_g of the Mathematical Program 6.2 can be rewritten using the matrix form:

$$L_g = 1 - W_g^T P_g W_g \tag{6.4}$$

The determination of the weight vectors only requires the bonds belonging to the analyzed rating group. Thus, the computation of the weight vectors is independently and for all groups, the weights vectors can be identified with the help of one single objective function. Finally, the Mathematical Program 6.2, which is only valid for one rating group, can be rewritten in the subsequent form.

<div align="center">

Mathematical Program 6.3
Mathematical Formulation of Step 2

</div>

$$\text{max} \qquad f_1 = \sum_{g=1}^{G} L_g$$

$$\text{subject to} \quad \sum_{k=1}^{n_g} W_{g,k} = 1 \quad \forall g = 1,\ldots,G$$

$$0 \leq W_{g,k} \leq C \quad \forall g = 1,\ldots,G \text{ and } k = 1,\ldots,n_g$$

The optimization of this mathematical program determines all the necessary weight vectors and finally gives access to the different rating groups.

Step 3: Calibration of the rating degrees via LR

In this step, the relationship between the bonds' attributes and their rating degrees is determined. For each group, regression factors which indicates these relationships are identified with the help of LR. Let β_{g0} be the attribute independent factor and β_{gj} with $j = 1, \ldots, m$ be the attributes' dependent factors. For rating group Ω_g^G the subsequent quadratic function permits to determine the regression factors:

$$Q_g = \sum_{k=1}^{n_g} \left(R_k^D - \beta_{g0} - \sum_{j=1}^{m} \left(a_{kj}\beta_{gj} \right) \right)^2 \tag{6.5}$$

As the regression factors are computed with only the knowledge of the bonds of the analyzed rating group, their determination is also independent. In this way, the computation of the regression factors for all the rating groups is obtained with one single objective function. The solving of the following mathematical program outputs the searched regression factors.

<div align="center">

Mathematical Program 6.4
Mathematical Formulation of Step 3

</div>

$$\min \qquad f_2 = \sum_{g=1}^{G} Q_g$$

$$\text{subject to} \quad \beta_{gj} \in \mathbb{R} \quad \forall g = 1, \ldots, G \text{ and } j = 0, \ldots, m$$

All the mathematical programs mentioned in the steps 2 and 3 can be solved with any quadratic programming tool, like Matlab or VB.net.

Step 4: Finding the rating group

The main input of this step is the set of unrated bonds $\tilde{\mathbf{B}}$. Furthermore, the weight vectors computed in Step 2 as well as the set of rated bonds are also utilized. After the ACRP model is completely trained and calibrated during the first three steps, the evaluation of unrated bonds is undertaken. First, the rating group $\tilde{R}_i^G$ of the unrated bond $\tilde{B}_i$ has to be identified. Several computations are undertaken to be able to determine the rating group of the unrated bonds. Starting to compute the respective distance between the unrated bonds and each bond in the different rating groups. These distances are given in the following formula:

$$K_g(i,k) = \exp\left(-\frac{\sum_{j=1}^{m} \left(\tilde{a}_{ij} - a_{kj} \right)^2}{2\sigma_g^2} \right) \tag{6.6}$$

with $k = 1, \ldots, n_g$ and $i = 1, \ldots, \tilde{n}$. Second, for each rating group, the radius of the hypersphere have to be computed. Therefore, one representative bond of the group is selected. The selected representative bond is denoted by B_g'. The distance between B_g' an all the remaining bonds of its rating group is determined:

$$radK_g(k) = \exp\left(-\frac{\sum_{j=1}^m \left(a_{gj}' - a_{kj}\right)^2}{2\sigma_g^2}\right) \tag{6.7}$$

These preliminary computations allows now to identify the distances of the unrated bonds to the rating groups as well as the radiuses of the hyperspheres. The distance between the unrated bond $\tilde{B}_i$ and rating group Ω_g^G is given by the subsequent formula:

$$D_g(i) = 1 - 2\sum_{k=1}^{n_g} W_{g,k} K_g(i,k) + \sum_{k=1}^{n_g}\sum_{k'=1}^{n_g} W_{g,k} W_{g,k'} P_g(k,k') \tag{6.8}$$

With the help of the matrix form, Formula 6.8 can be rewritten as follows:

$$D_g(i) = 1 - 2K_g^T W_g + W_g^T P_g W_g \tag{6.9}$$

with $i = 1, \ldots, \tilde{n}$ and $g = 1, \ldots, G$. The radius of the hypersphere describing rating group Ω_g^G is computed in the following way:

$$R_g = 1 - 2\sum_{k=1}^{n_g} W_{g,k} radK_g(k) + \sum_{k=1}^{n_g}\sum_{k'=1}^{n_g} W_{g,k} W_{g,k'} P_g(k,k') \tag{6.10}$$

Formula 6.10 is also rewritten in matrix form:

$$R_g = 1 - 2radK_g^T W_g + W_g^T P_g W_g \tag{6.11}$$

with $g = 1, \ldots, G$. The distribution of the rating groups is taken into consideration. Therefore, decision function 4.75 introduced in Chapter 4, Section 4 is used. The decision function is given as follows:

$$\varphi_g(i) = \begin{cases} \lambda\dfrac{1 - \frac{D_g(i)}{R_g}}{1 + \frac{D_g(i)}{R_g}} + \gamma & 0 \le D_g(i) \le R_g \\ \gamma\dfrac{R_g}{D_g(i)} & D_g(i) > R_g \end{cases} \tag{6.12}$$

with $i = 1, \ldots, \tilde{n}$ and $g = 1, \ldots, G$. λ and γ are user-given regulator constant with $\lambda + \gamma = 1$. The consideration of the groups' distributions is important because in case of equal distance

to two groups $\tilde{B}_i$ will be affiliated to the sparser rating group. To facilitate the further manipulation of the decision function, an affiliation vector for each bond is introduced. The affiliation vector is defined in the following way.

$$V_i(g) = \begin{cases} 1, & \varphi_g(i) = \max_{g=1,\dots,G} \varphi_g(i) \\ 0, & \varphi_g(i) \neq \max_{g=1,\dots,G} \varphi_g(i) \end{cases} \tag{6.13}$$

The interpretation of the affiliation vector is as follows: 1 stands for true and 0 for false. Hence, the index, corresponding to the entry 1, gives the rating group R_g^G of the analyzed bond. It is obviously that for each $i = 1,\dots,\tilde{n}$, $\sum_{g=1}^{G} V_i(g) = 1$ is always true.

Step 5: Finding the rating degree

The final step requires the following input data: the set of unrated bonds, the affiliation vectors, computed in Step 4 and the regression factors determined from Step 3.

The affiliation vector allows to select the right regression factors to computed the rating degree of $\tilde{B}_i$. The rating degree $\tilde{R}_i^D$ is identified by using the following formula:

$$\tilde{R}_i^D = \sum_{g=1}^{G} V_i(g) \left(\beta_{g0} + \sum_{j=1}^{m} \beta_{gj} \tilde{a}_{ij} \right) \tag{6.14}$$

with $i = 1,\dots,\tilde{n}$. With the help of Formula 6.14, a risk profile of the unrated bonds is determined and the bonds can be classified into the set of rated bonds. In the next subsection, the new ACRP model is applied on an artificial example to illustrate its procedure.

6.1.3 Example with artificial data

The new ACRP model is applied on artificial bonds. Let **B** comprise 5 bonds. The bonds have the following characteristics.

$$
\begin{array}{c}
i \\
1 \\
2 \\
3 \\
4 \\
5
\end{array}
\begin{pmatrix}
a_{i1} & a_{i2} \\
2 & 3 \\
3 & 8 \\
10 & 2 \\
12 & 1 \\
13 & 2
\end{pmatrix}
,\qquad
\begin{array}{c}
i \\
1 \\
2 \\
3 \\
4 \\
5
\end{array}
\begin{pmatrix}
R_i^D \\
1.00 \\
1.25 \\
3.50 \\
3.75 \\
3.75
\end{pmatrix}
$$

Furthermore, assume that there exist one unrated bond $\tilde{B}_1$ with the following observed attributes.

$$\begin{array}{c|c} \tilde{a}_1 1 & \tilde{a}_1 2 \\ \hline 11 & 2 \end{array}$$

The small number of bonds is due to the fact that the presented example is solvable with Excel Solver. A credit rating prediction should be initiated to identify the credit risk of $\tilde{B}_1$. Before the rating degree of $\tilde{B}_1$ can be determined, the ACRP model has to be built and calibrated.

Step 1: Building of the rating groups

Based on the given rating degrees, the rating groups are defined. Due to the small number of rated bonds, two different rating groups are determined. The following merging of rating degrees for the set **B** is made:

$$B_i \in \begin{cases} \Omega_1^G & R_i^D < 1.5 \\ \Omega_2^G & R_i^D \geq 1.5 \end{cases} \qquad \forall i = 1, \dots, n$$

Hence, the bonds have obtained the following rating groups:

$$\begin{array}{c} i \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{pmatrix} R_i^G \\ 1.00 \\ 1.00 \\ 2.00 \\ 2.00 \\ 2.00 \end{pmatrix}$$

Step 2: Calibration of the rating groups via SVDD

The rating groups are calibrated with the help of SVDD. For the two rating groups, the kernel factor σ is set equal to 2. First, the mutual distances between each bond in one rating group is computed. Using Formula 6.3, the following kernel matrices are obtained:

$$P_1 = \begin{pmatrix} 1.00 & 0.78 \\ 0.78 & 1.00 \end{pmatrix} \quad , \quad P_2 = \begin{pmatrix} 1.00 & 0.54 & 0.32 \\ 0.54 & 1.00 & 0.78 \\ 0.32 & 0.78 & 1.00 \end{pmatrix}$$

These kernel matrices are used to solve the Mathematical Program 6.3. Thus, $f_1 = 0.45$ and the subsequent optimal weight vectors are obtained:

$$W_1 = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} \quad , \quad W_2 = \begin{pmatrix} 0.5 \\ 0.0 \\ 0.5 \end{pmatrix}$$

The representative bonds are defined by the optimal weight vectors. For the first rating group W_1 specifies that both bonds have representative character with equal impact. In contrast, for rating group Ω_2^G, only two out of the three bonds are necessary to describe the bonds. The bond B_4 has no influence in the determination of the descriptive hypersphere.

Step 3: Calibration of the rating degrees via LR

The ACRP model is completely defined and calibrated when the relationship between the bonds' attributes and their rating degrees is identified. For this reason, the regression factors have to be computed by solving the Mathematical Program 6.4. The obtained solution is that f_2 equals 0.004 and the regression factors are given in the following table.

$$\beta_1 = \begin{pmatrix} 0.37 \\ 0.26 \\ 0.01 \end{pmatrix} \quad , \quad \beta_2 = \begin{pmatrix} 2.63 \\ 0.09 \\ 0.00 \end{pmatrix}$$

Table 6.7 Regression factors

Finally, the ACRP model can be applied to the unrated bond $\tilde{B}_1$ to evaluate its credit risk.

Step 4: Finding the rating group

One representative bond is selected for each rating group and noting them by B'_g, with $g = 1, 2$. According to the respective weight vectors, the first bond in each group is chosen as the representative bond B'_g. Thus, out of the whole set of bonds, B_1 and B_3, are selected. The subsequent table recalls their attributes.

B'_g	a'_{g1}	a'_{g2}
1	2	9
2	10	2

The distances between $\tilde{B}_1$ and each bond of the several rating groups is computed with the help of Formula 6.6.

K_1	K_2
0.0000	0.3679
0.0000	0.4169
	0.0183

Based on the distance matrix K_g for $g = 1, 2$ and the kernel matrices P_g for $g = 1, 2$, Formula 6.8 determines how distant the bond $\tilde{B}_1$ is from the different rating groups.

D_1	D_2
1.89	1.28

Before determining the radiuses of the two descriptive hyperspheres, the mutual distances between the selected representative bond B'_g and all the remaining bonds of its rating group Ω_g^G have to be computed. The distances are obtained by applying Formula 6.7.

$radK_1$	$radK_2$
1.00000	1.00000
0.41686	0.02075
	0.00012

Afterwards, the respective radiuses are computed by Formula 6.10.

R_1	r_2
0.47	0.66

Finally, the decision function 6.12 is computed and the affiliation vector 6.13 is set up. The following results are obtained:

$$
\begin{array}{c|c}
\varphi_1 & \varphi_2 \\
\hline
0.04 & 0.08
\end{array}
\quad , \quad
V_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}
$$

The affiliation of $\tilde{B}_1$ to Ω_2^G is directly given by V_1. $\tilde{B}_1$ obtains rating group $\tilde{R}_1^G = 2$.

Step 5: Finding the rating degree

The affiliation vector V_1 and the regression factors β_1 and β_2 allows to predict the rating degree of $\tilde{B}_1$. By computing Formula 6.14, the prediction of the credit risk for the unrated bond equals $\tilde{R}_1^D = 3.61$. The interpretation of the obtained result is as follows:

- $\tilde{B}_1$ and B_3 seem to have similar attributes but $\tilde{B}_1$ is riskier than B_3.

- $\tilde{B}_1$ has less risk than the two remaining bonds of rating group Ω_2^G.

6.2   Exemplary field of application of the new model

This section should demonstrate that the entire or parts of the new developed ACRP model can also be reused in other field of research. Even if the focus of development of the ACRP model is the prediction of the rating degrees of financial bonds, the concept is mainly based on classification methods. Therefore, parts or the entire model can also be employed to solve other problems which do not lie in the financial domain. As the department of the supporting research institute, ERIN, has its main research focus on environmental topics, this section handles the possibilities of application of the model to these topics.

Classification is also used in biology to solve complex problems, like the prediction of membrane protein types [14]. Without going into detail but a cell is enclosed by the plasma membrane. The most important cells activities are carried out by the membrane proteins. Therefore the prediction of the proteins is a major task to improve the treatment of different disease. Due to the fact that the several membrane proteins have relative clear discriminating factors, the application of SVM is already tested to predict the different proteins [14]. In the same way, the second step of the new ACRP model, which consists of the use of SVDD, could be reused in such types of problems.

Another possible application is the classification of genes. A false structuring in the DNA sequence can causes severe diseases, like cancer [6]. Therefore, the detection of false structured sequences in the genes has a great advantage in the early diagnosis of various diseases. SVM as well as SVDD have already been successfully used to solve this problem [6, 100]. Thus, the new ACRP model can also be adapted to fulfill the requirements to classify accurately DNA sequences. Nevertheless, one problem in the classification of DNA sequences as well as membrane proteins is that they are typically encoded with the help of letters. Either, a re-encoding have to be used such that the normal kernel function are usable by the classification methods (SVM or SVDD), or new so-called string kernels are employed. Specific Mismatch string kernels to classify proteins are introduced in [27]. As the kernel function in the new ACRP model is not fixed, string kernels can also be integrated into the model.

Beside of the stated problems of classifications of proteins and DNA sequences, other fields also offer challenges which can be tackled by the new ACRP model. In medicine, computer-based methods are more and more employed to make a diagnostic. The diagnostic of the brain is usually made with the help of an electroencephalogram (EEG). The signals from an EEG can be classified to extract more knowledge from them [53]. Before the use of

classification methods, like SVM, the wavelet transform is used to extract the needed features which represents the attributes of the input data. After the preprocessing steps, SVM is applied to obtain a partition of the different signals [53]. At this stage, after some minor modifications, the new ACRP model can be employed.

Finally, this section illustrates that the new ACRP model is not only limited to the prediction of rating degrees of financial bonds but can also be employed in other fields of research.

6.3   Conclusion

In this chapter, a new ACRP model which is based on a hybrid method (SVDD and LR) is introduced. The ACRP model executes five steps to evaluate the credit risk of unrated bonds and to predict their rating degrees. The first three steps represents the building and calibration of the model. Afterwards the trained model can be used to evaluated the credit risk of unrated bonds. Therefore, the last two steps are executed. An artificial example is used to explain the computations undertaken in the different steps of the ACRP model.

In the new ACRP model, the bonds' attributes are computed from the features retrieved from the financial market, like their historical prices and their coupon rates. No balance sheets or other information from the issuers are needed, but the prediction of the rating degrees is only based on the creditworthiness of the bonds. In this way, sovereign and corporate bonds can be simultaneously rated by the new ACRP model. Thus, the described model answers Research Question 2. The employed cardinal rating scale allows to get a deeper insight into the credit risk because each analyzed bond obtains its individual rating degree.

Finally, it is also shown that beside the main focus of the new ACRP model, the prediction of rating degrees, the model is also applicable in other field of research. Due to the fact that the department of the supporting research institute, ERIN, has its focus on environmental topics, several possible topics from biology and medicine are presented. In this way, ERIN has the opportunity to reuse the new ACRP model after some modifications to their main research focuses.

Chapter 7

Competitive and empirical analysis of the new ACRP model

In this chapter, the performance of the model is investigated. First, the framework for the competitive analysis is described. The framework includes the definition of different types of risk information as well as the used performance measure [1]. Second, mis-classification guarantees for the different analyzed worst-case scenarios are determined. The competitive analysis is also employed to establish a first classification of the different introduced ACRP models. Third, an empirical analysis is undertaken to assert the performance of the ACRP model on real bonds. Finally, the obtained results are summarized in the conclusion.

7.1 Framework for the competitive analysis

Several types of risk information are introduced and formally defined. The relations between the different types of information are established to set up an hierarchy.

7.1.1 Types of risk information

A credit rating incorporates usually three different types of risk information. The notation introduced in Chapter 5 remains valid. The following types of risk information can be distinguished:

---

[1] The framework for competitive analysis has been presented on the conference Risk Information Management, Risk Models and Applications in Berlin and has been published in the proceedings [34]. The idea to undertake a competitive analysis was raised by Günter Schmidt. The realization has been made by myself with some help of Robert Dochow.

1. Rating characteristic

2. Rating group, Ω_g^G

3. Rating degree, Ω_d^D

Setting the distance between to adjacent rating degrees δ equal to 0.25, the subsequent figure illustrates the three mentioned types of risk information and offers a view on their relations.

Rating characteristic	investment												non-investment							
Rating group, Ω_g^G	1			2				3						4			5			
Rating degree, Ω_d^D	1	1.3	1.5	1.8	2	2.3	2.5	2.8	3	3.3	3.5	3.8	4	4.3	4.5	4.8	5	5.3	5.5	5.8	6
Moody's rating	Aaa	Aa1	Aa2	Aa3	A1	A2	A3	Baa1	Baa2	Baa3	Ba1	Ba2	Ba3	B1	B2	B3	Caa1	Caa2	Caa3	Ca	C

Fig. 7.1 Types of risk information

Before, the different types of risk information are defined, the prediction process of an arbitrary ACRP model is abstactly described. The prediction process of an ACRP model is given by its rating function $f(\cdot)$. The rating function depends on the employed methods. For example, the rating function $f(\cdot)$ of the new ACRP model introduced in Chapter 6 depends on the two functions: 6.12 and 6.14.

Type of risk information, rating characteristics, described if the bond is of investment or non-investment grade. For example, for Moody's rating scale, investment grade is defined for all the credit ratings between *Aaa* and *Baa3*. All the bonds with a credit rating below *Ba*1 are considered as non-investment grades. Rating characteristics is a binary information. 0 indicates non-investment grade and 1 specifies investment grade. The formal definition of type of information 1 is given as follows:

Definition 20. *Let $\tilde{B}$ be one unrated bond with $\tilde{A}$ its vector of attributes. Rating characteristics is formally defined by:*

$$\mathbb{R} \to \{0,1\} \tag{7.1}$$
$$\tilde{A} \mapsto f_C(\tilde{A}) = \tilde{R}^C \tag{7.2}$$

Definition 21. *Assuming that $\tilde{A}$ only incorporates type of information 1, then the sets of possible rating degrees are given as follows:*

1. $\Pi_1^C = \left\{ \Omega_1^D, \dots, \Omega_{(\varepsilon^D - 1)}^D \right\}$ *if the bonds are of investment grade*

2. $\Pi_2^C = \left\{ \Omega_{\varepsilon^D}^D, \dots, \Omega_D^D \right\}$ *if the bonds are of non-investment grade*

with ε^D the first rating degree of non-investment grade.

Example 1. *Moody's rating scale contains* 21 *different rating degrees and the first degree of non-investment grade* ε^D *equals* 11. *With this information, the two sets of possible rating degrees can be stated with the help of the well-known credit ratings:*

1. *Investment grade:* $\Pi_1^C = \left\{ \Omega_1^D, \ldots, \Omega_{10}^D \right\} = \{Aaa, \ldots, Baa3\}$

2. *Non-Investment grade:* $\Pi_2^C = \left\{ \Omega_{11}^D, \ldots, \Omega_{21}^D \right\} = \{Ba1, \ldots, C\}$

Type of information 2 expresses the rating group affiliation of $\tilde{B}$. Bonds in the same rating group are considered to have a similar credit risk. Type of information 2 is also of ordinal character. This type of information is formally stated in the following definition:

Definition 22. *Given $\tilde{A}$ the vector of attributes of $\tilde{B}$. Rating group is defined as follows:*

$$\mathbb{R} \to \{1, \ldots, G\} \tag{7.3}$$

$$\tilde{A} \mapsto f_G(\tilde{A}) = \tilde{R}^G \tag{7.4}$$

The sets of possible rating degrees corresponding to type of information 2 are defined.

Definition 23. *Let d_l be the starting rating degree and d_r the finishing rating degree of a rating group. If types of information 1 and 2 are known than the following sets are defined. Let d_l^1 equals 1 and d_r^1 some degree d, with $d > d_l^1$. In this case, the first set is given by: $\Pi_1^G = \left\{ \Omega_{d_l^1}^D, \ldots, \Omega_{d_r^1}^D \right\}$. For each additional set, the following recursive rule is employed:*

$$d_l^g = d_r^{g-1} + 1, \quad \forall g = 1, \ldots, G. \tag{7.5}$$

In this way, the additional sets have the following look: $\Pi_g^G = \left\{ \Omega_{d_l^g}^D, \ldots, \Omega_{d_r^g}^D \right\}$, with $g = 2, \ldots, G$. It is obviously that d_r^G is always equal to the last rating degree D.

Example 2. *Using Moody's rating scale, 5 different rating groups can be identified. With $G = 5$, the obtained groups can receive the following risk label:*

1. Π_1^G: *very little risk*

2. Π_2^G: *little risk*

3. Π_3^G: *medium risk*

4. Π_4^G: *high risk*

5. Π_5^G: *very high risk*

The following starting and finishing rating degrees are fixed: $d_l^1 = 1$, $d_r^1 = 3$, $d_r^2 = 7$, $d_r^3 = 13$, $d_r^4 = 16$ and $d_r^5 = 21$. Then, the obtained sets are as follows:

1. $\Pi_1^G = \{\Omega_1^D, \ldots, \Omega_3^D\} = \{Aaa, \ldots, Aa2\}$

2. $\Pi_2^G = \{\Omega_4^D, \ldots, \Omega_7^D\} = \{Aa3, \ldots, A3\}$

3. $\Pi_3^G = \{\Omega_8^D, \ldots, \Omega_{13}^D\} = \{Baa1, \ldots, Ba3\}$

4. $\Pi_4^G = \{\Omega_{14}^D, \ldots, \Omega_{16}^D\} = \{B1, \ldots, B3\}$

5. $\Pi_5^G = \{\Omega_{17}^D, \ldots, \Omega_{21}^D\} = \{Caa1, \ldots, C\}$

Finally, type of information 3 expresses the exact predicted rating degree $\tilde{R}^D$ of $\tilde{B}$. Basically, the rating degree is also of ordinal character. However, due to the fact that some ACRP models outputs cardinal rating degrees, a relaxation to a cardinal rating scale is undertaken. A mapping to the well-known ordinal rating scale used by the rating agencies is always feasible. In this way, the relaxation can be made without loss of generality. The formal definition of type of information 3 is given in the subsequent way.

Definition 24. *Rating degree $\tilde{R}^D$ is defined in the following way if $\tilde{A}$, the vector of attributes of $\tilde{B}$ is given:*

$$\mathbb{R} \to [1, \ldots, (D-1)\delta] \tag{7.6}$$
$$\tilde{A} \mapsto f_D(\tilde{A}) = \tilde{R}^D \tag{7.7}$$

Figure 7.1 illustrates the case for Moody's rating scale with $\delta = 0.25$. Now all the three types of risk information are formally described and the respective sets of possible rating degrees are defined.

7.1.2 Relations between the different types of information

As the three different types of risk information are defined, one important question remains: *Can an ACRP model, which focuses on type of information 3, also offer the two other types of information to investors?* Starting with the analysis of the relations if type of information 3 is known.

Proposition 2. *Type of information 1 can be extracted from type of information 3.*

Proof. ε^D represents the first non-investment rating degree, also called threshold rating. According to ε^D, type of information 1 can be extracted from type of information 3 as follows:

$$V_1^{3 \mapsto 1} = \begin{cases} 0 \text{ if } \tilde{R}^D \geq \varepsilon^D & \text{"non-investment grade"} \\ 1 \text{ if } \tilde{R}^D < \varepsilon^D & \text{"investment grade"} \end{cases} \tag{7.8}$$

$\square$

The subsequent proposition states the relation between type of information 2 and 3.

Proposition 3. *Type of information 2 can be deduced from type of information 3.*

Proof. A merging of rating degrees has to be undertaken to restore the different desired rating groups. Let $\rho_1^D, \ldots, \rho_{(G-1)}^D$ be the distinctive rating degrees. Type of information 2 is extracted from type of information 3 with the help of these distinctive rating degrees in the following way:

$$V_1^{3 \mapsto 2} = \begin{cases} 1 & \text{if } \tilde{R}^D < \rho_1^D \\ 2 & \text{if } \rho_1^D \leq \tilde{R}^D < \rho_2^D \\ \ldots & \\ G & \text{if } \tilde{R}^D \geq \rho_{(G-1)}^D \end{cases} \tag{7.9}$$

$\square$

Concluding, an ACRP model focusing on type of information 3 is able to offer to investors all the remaining types of risk information: rating characteristics and rating group. To complete the establishing of an hierarchy, the relation between the types of information 1 and 2 has to be investigated.

Proposition 4. *Type of information 2 can reproduce type of information 1.*

Proof. If type of information 2 is known, then the bonds are divided into different rating groups. In this way, the following two cases are possible:

   i. Each rating group consists of investment **or** non-investment grade bonds

   ii. There is one rating group which includes investment **and** non-investment grade bonds

Case i.:
The last group containing bonds with rating characteristic investment grade has the index $\bar{g}$. As each ACRP model is calibrated with the help of a set of rated bonds, rating group

$\Omega_{\bar{g}}^G$ is known. In this way, the relation between type of information 2 and 1 is given by the following formula:

$$V_1^{2 \mapsto 1} = \begin{cases} 0 & \text{if } \tilde{R}^G \in \Omega_g^G, \quad \forall g \in \{(\bar{g}+1), \ldots, G\} \\ 1 & \text{if } \tilde{R}^G \in \Omega_g^G, \quad \forall g \in \{1, \ldots, \bar{g}\} \end{cases} \qquad (7.10)$$

<u>Case ii.:</u>

Let g' be the index of the rating group including investment and non-investment grade bonds. Thus, all the bonds in the rating groups, Ω_g^G with $g = 1, \ldots, (g'-1)$, are of investment grade and all the bonds in the rating groups Ω_g^G, with $g = (g'+1), \ldots, G$, are of non-investment grade. $\Omega_{g'}^G$ is the only rating group where a division of the bonds is not feasible and a precise analysis has to be undertaken. The separation into investment and non-investment grade bonds in rating group $\Omega_{g'}^G$ is made with the help of the threshold ε^G. This threshold is identifiable by applying the ACRP model only on $\Omega_{g'}^G$. In this way, the relevant rating group can be split into two subgroups:

$$\Omega_{g'}^G = \begin{cases} \text{"investment grade"} & \text{if } \tilde{R}^G < \varepsilon^G \\ \text{"non-investment grade"} & \text{if } \tilde{R}^G > \varepsilon^G \end{cases} \qquad (7.11)$$

Finally, type of information 1 is deducible from type of information 2 and the relation is given as follows:

$$V_1^{2 \mapsto 1} = \begin{cases} 0 & \text{if } \tilde{R}^G \in \Omega_g^G \quad \forall g = \{(g'+1), \ldots, G\} \text{ or } \tilde{R}^G \in \Omega_{g'}^G \text{ and } \tilde{R}^G > \varepsilon^G \\ 1 & \text{if } \tilde{R}^G \in \Omega_g^G \quad \forall g = \{1, \ldots, g'\} \text{ or } \tilde{R}^G \in \Omega_{g'}^G \text{ and } \tilde{R}^G < \varepsilon^G \end{cases} \qquad (7.12)$$

$\square$

In this way, all the relations between the different types of risk information are established. The description of the framework for the competitive analysis is completed with the definitions of the performance measure and the optimal benchmark algorithm.

7.1.3 Performance measure and benchmark model

Let $\tilde{A}$ be the vector of attributes of $\tilde{B}$. In addition, let $\mathscr{A}$ be the set of all possible attributes combinations for one single bond. The set $\mathscr{A}$ is restricted by allowing the attributes incorporating more and more informations. In competitive analysis, the predicted result is compared to an optimal benchmark model, noted as OPT [4, 54]. In this thesis, OPT is defined in the following way:

Definition 25. *The benchmark model OPT is considered as the adversary player of the analyzed ACRP model, noted as ALG. Thus, ALG will always try to minimize the damage which OPT can perpetrate. In contrast, OPT always wants to maximize this damage. The damage is defined as the difference between the rating degree obtained by ALG and OPT.*

The attributes incorporate in different stage the several introduced types of risk information. In this way, the attributes introduce the credit risk into the ACRP model. The subsequent definition established the used performance measure.

Definition 26. *Let $\tilde{A} \in \mathscr{A}$ the vector of attributes of bond $\tilde{B}$. The deviation between ALG and OPT is calculated as follows:*

$$c(\tilde{A}) = |ALG(\tilde{A}) - OPT(\tilde{A})| \quad \forall \tilde{A} \in \mathscr{A} \tag{7.13}$$

However, in competitive analysis, the worst-case is always considered. The competitive ratio is defined in absolute terms and similar to [4, 36, 67].

Definition 27. *The competitive ratio c_{max} is defined by the maximal possible deviation between ALG and OPT. The following formula allows to compute c_{max}:*

$$c_{max} = \max_{\tilde{A} \in \mathscr{A}} c(\tilde{A}) = \max_{\tilde{A} \in \mathscr{A}} |ALG(\tilde{A}) - OPT(\tilde{A})| \tag{7.14}$$

This definition of the competitive ratio is used due to the fact that the division of rating degrees can differ even if their deviation are equal. For example, set $\delta = 0.25$ and use the following rating degrees: $\Omega_1^D = 1$, $\Omega_3^D = 1.5$, $\Omega_4^D = 1.75$ and $\Omega_6^D = 2.25$. the subsequent absolute and relative ratios are obtained:

- absolute ratios: $\Omega_3^D - \Omega_1^D = 0.5$ and $\Omega_6^D - \Omega_4^D = 0.5$

- relative ratios: $\frac{\Omega_3^D}{\Omega_1^D} = \frac{1.5}{1} = \frac{3}{2}$ and $\frac{\Omega_6^D}{\Omega_4^D} = \frac{2.25}{1.75} = \frac{9}{7}$

Nevertheless, one single rating degree is always between the regarded rating degrees: Ω_2^D respectively Ω_5^D. In this way, the absolute ratios represent correctly the deviation and also show that the two deviations are identically. After the performance measure and the optimal benchmark model are defined, the framework for the competitive analysis is completely stated. In the next section, the competitive analysis is conducted and the theorem defining competitive ACRP models is given.

7.2 Competitive Analysis

The competitive analysis is undertaken under the following setting: with each investigated worst-case scenario more information are assumed to be known. The following three scenarios are identified:

1. No additional information are given and the ACRP model (ALG) has to predict the rating characteristic of $\tilde{B}$

2. The rating characteristic is known and ALG has to determine to which rating group $\tilde{B}$ belongs

3. The rating characteristic and the rating groups are known and the correct rating degree of $\tilde{B}$ has to be determined

A first simplified view on the problem leads to the following conclusion. ALG predicts the first rating degree Ω_1^D and the correct one determined by OPT is Ω_D^D. Thus, the competitive ratio equals: $c = |\Omega_1^D - \Omega_D^D| = D - 1$. This competitive ratio c represents the upper bound. The setting of ALG focuses to achieved OPT. Additionally, it is assumed that a minimal information which can be treated by every ACRP model is always given. Each ACRP model is built and calibrated with the help of a set of rated bonds. During the training phase, this minimal information is inputted into the ACRP model. The following assumption is stated: *ALG will always miss the optimal result in the lowest possible way if an error occurs.*

7.2.1 Scenario 1

The first scenario consist to investigate the worst-case performance of ALG if no additional information is known and ALG should predict the rating characteristic of $\tilde{B}$. As OPT is the adversary player, the declaration of OPT will always be the exact opposite of the prediction made by ALG. In this way, the following lemma can be stated.

Lemma 1. *Under the assumption of scenario 1, the global maximal deviation between ALG and OPT equals:*
$$c_{max}^1 = \max\left\{(D - (\varepsilon^D - 1))\delta, (\varepsilon^D - 1)\delta\right\}.$$

Proof. Scenario 1 includes two different cases which can occur:

i. $\tilde{B}$ is predicted by ALG to be of investment grade

ii. $\tilde{B}$ is predicted by ALG to be of non-investment grade

For the two cases, the set of possible rating degrees are defined in Definition 21.

Case i.:

In this case, ALG predicts $\tilde{B}$ to be of investment grade. Thus, out of the set Π_1^C, ALG assigns a rating degree to $\tilde{B}$. As ALG always tries to minimize the damage to OPT, ALG assigns the last investment grade rating degree $\Omega_{(\varepsilon^D-1)}^D$ to the unrated bond $\tilde{B}$. In contrast, OPT identifies $\tilde{B}$ to be of non-investment grade and assigns a rating degree out of the set, Π_2^C to it. OPT tries to maximize the damage between itself and ALG, so OPT designates the last existing rating degree Ω_D^D to $\tilde{B}$. Thus, the maximal deviation is computed in the following way:

$$c_{max}^{(i)} = \max_{\tilde{A} \in \mathscr{A}} |ALG(\tilde{A}) - OPT(\tilde{A})| \tag{7.15}$$

$$= |(1 + ((\varepsilon^D - 1) - 1)\delta) - (1 + (D-1)\delta)| \tag{7.16}$$

$$= (D - (\varepsilon^D - 1))\delta \tag{7.17}$$

Case ii.:

$\tilde{B}$ is predicted by ALG to be of non-investment grade and OPT identifies $\tilde{B}$ to be of investment grade. In the same logic as for Case i., ALG assigns to $\tilde{B}$ the first non-investment grade rating degree $\Omega_{\varepsilon^D}^D$ to try to minimize the damage to OPT. In contrast, OPT assigns the first existing rating degree Ω_1^D to maximize the damage. The subsequent formula allows to compute the maximal deviation:

$$c_{max}^{(ii)} = \max_{\tilde{A} \in \mathscr{A}} |ALG(\tilde{A}) - OPT(\tilde{A})| \tag{7.18}$$

$$= |(1 + (\varepsilon^D - 1)\delta) - (1 + (1-1)\delta)| \tag{7.19}$$

$$= (\varepsilon^D - 1)\delta \tag{7.20}$$

The maximum over the two cases determines the global maximal deviation for Scenario 1 and is given by:

$$c_{max}^1 = \max\left\{ c_{max}^{(i)}, c_{max}^{(ii)} \right\} \tag{7.21}$$

$\square$

Example 3. *Moody's rating scale is used to illustrate the computations to determine the global maximal deviation under the assumption of Scenario 1. In Example 1, the threshold rating degree ε^D is identified and equals 11. Let set the value indicating the distance between to adjacent rating degrees δ equal to 0.25. In all, $D = 21$ different rating degrees are employed by Moody's. First, computing the deviation if $\tilde{B}$ is predicted as investment grade*

by ALG.

$$
\begin{aligned}
c_{max}^{(i)} &= |1 + ((\varepsilon^D - 1) - 1)\delta - (1 + (D - 1)\delta)| \\
&= |1 + ((11 - 1) - 1)0.25 - (1 + (21 - 1)0.25)| \\
&= |3.25 - 6| \\
&= |-3.75| = 3.75
\end{aligned}
$$

Afterwards, the second case is analyzed, e.g., ALG has predicted $\tilde{B}$ as non-investment grade.

$$
\begin{aligned}
c_{max}^{(ii)} &= |1 + (\varepsilon^D - 1)\delta - (1 + (1 - 1)\delta)| \\
&= |1 + (11 - 1)0.25 - (1 + (1 - 1)0.25)| \\
&= |3.5 - 1| \\
&= |2.5| = 2.5
\end{aligned}
$$

After taking the maximum over $c_{max}^{(i)}$ and $c_{max}^{(ii)}$, the global maximum deviation for Scenario 1 is identified:

$$
c_{max}^1 = \max\left(c_{max}^{(ii)}, c_{max}^{(ii)} \right) = 3.75
$$

Assume, $\tilde{B}$ has a rating degree of 1 and ALG has mis-predicted its rating degree, then the global maximum deviation indicates that the false predicted rating degree of $\tilde{B}$ never exceed the rating degree: $1 + 3.75 = 4.75$. According to Moody's rating scale, B3 is never exceed in the case that $\tilde{B}$ has a rating degree of Aaa. in the appendix, the example is illustrated in Figure A.1.

7.2.2 Scenario 2

The assumption of Scenario 2 is that the distinction between investment - and non-investment grade is incorporated in the bonds' attributes. Assume that $\Omega_{g'}^G$ is the rating group which includes investment grade and non-investment grade bonds.

Lemma 2. *Under the assumption of Scenario 2, the global maximal deviation between ALG and OPT is given as:*
$c_{max}^2 = \max\{A_1, A_2, B_1, B_2\}$, *with:*
$A_1 = ((\varepsilon^D - 1) - d_r^1)\delta$
$A_2 = d_r^{g'-1}\delta$
$B_1 = (D - d_r^{g'})\delta$
$B_2 = ((d_r^{G-1} + 1) - \varepsilon^D)\delta$

Proof. As the rating characteristic is known, two different cases can occur:

    i. $\tilde{B}$ has rating characteristic investment grade

    ii. $\tilde{B}$ has rating characteristic non-investment grade

Case i:

As $\Omega_{g'}^G$ represents the mixed rating group, there exists g' possible rating groups containing bonds with rating characteristic investment grade. Due to the fact that the maximal deviation is search, only the rating groups on the extremities, Ω_1^G and $\Omega_{g'}^G$ are of interest. Therefore, the following sub-cases can emerge:

- $\tilde{B}$ is assigned to rating group Ω_1^G by ALG

- $\tilde{B}$ is affiliated to rating group $\Omega_{g'}^G$ by ALG

Sub-Case i-1:

According to Definition 23, ALG predicts a rating degree out of Π_1^G. ALG assigns the last rating degree d_r^1 in group Ω_1^G to $\tilde{B}$ in the attempt to minimize the damage to OPT. OPT, as the adversary player, identifies $\tilde{B}$ in group $\Omega_{g'}^G$ and designates the last investment grade rating degree to it. In this way, the following formula computes the deviation:

$$A_1 = \max_{\tilde{A} \in \mathscr{A}} |ALG(\tilde{A}) - OPT(\tilde{A})| \tag{7.22}$$

$$= |d_r^1 \delta - (\varepsilon^D - 1)\delta| \tag{7.23}$$

$$= ((\varepsilon^D - 1) - d_r^1)\delta \tag{7.24}$$

Sub-Case i-2:

In the same logic, the deviation for the second sub-case is given as:

$$A_2 = \max_{\tilde{A} \in \mathscr{A}} |ALG(\tilde{A}) - OPT(\tilde{A})| \tag{7.25}$$

$$= |d_l^{g'} \delta - d_l^1 \delta| \tag{7.26}$$

$$= (d_l^{g'} - d_l^1)\delta \tag{7.27}$$

$$= ((d_r^{g'-1} + 1) - d_l^1)\delta \tag{7.28}$$

Concluding, the maximal deviation for Case i. is equal to: $c_{max}^{(i)} = \max\{A_1, A_2\}$.

Case ii:

There exist $G - g' + 1$ rating groups composed of bonds with rating characteristic non-investment grade. The intermediate groups can be eliminated and the rating groups $\Omega_{g'}^G$ and Ω_G^G are only investigated. Therefore the following sub-cases have to be analyzed:

- ALG predicts rating group $\Omega_{g'}^G$ to $\tilde{B}$

- ALG predicts rating group Ω_G^G to $\tilde{B}$

Sub-Case ii-1:

As ALG assigns a rating degree according to $\Omega_{g'}^G$ to $\tilde{B}$, OPT designates the last existing rating degree to $\tilde{B}$. Thus, the deviation is as follows:

$$B_1 = \max_{\tilde{A} \in \mathscr{A}} |ALG(\tilde{A}) - OPT(\tilde{A})| \tag{7.29}$$

$$= |d_r^{g'}\delta - d_r^G\delta| \tag{7.30}$$

$$= (D - d_r^{g'})\delta \tag{7.31}$$

Sub-Case ii-2:

Similar to the previous sub-case, the deviation is determined in the following way:

$$B_2 = \max_{\tilde{A} \in \mathscr{A}} |ALG(\tilde{A}) - OPT(\tilde{A})| \tag{7.32}$$

$$= |d_l^G\delta - \varepsilon^D\delta| \tag{7.33}$$

$$= (d_l^G - \varepsilon^D)\delta \tag{7.34}$$

$$= ((d_r^{G-1} + 1) - \varepsilon^D)\delta \tag{7.35}$$

Concluding, the maximal deviation for Case ii. is obtained by taken the maximum over the two stated sub-cases: $c_{max}^{(ii)} = \max\{B_1, B_2\}$.

Finally, under the assumption of scenario 2, the global maximal deviation is as follows:

$$c_{max}^2 = \max\left\{c_{max}^{(i)}, c_{max}^{(ii)}\right\} \tag{7.36}$$

$\square$

Example 4. *Reusing the $G = 5$ rating groups introduced in Example 2 and Moody's rating scale to illustrate Scenario 2. The distance between two adjacent rating degrees δ is set to 0.25. The threshold rating degree ε^D equals 11. The third rating group is composed of investment and non-investment rating degrees: $\Pi_3 = \{8, \ldots, 13\} = \{Baa1, \ldots, Ba3\}$. Applying Lemma 2, the following deviations are determined:*

- $c_1 = ((\varepsilon^D - 1) - d_r^1)\delta = (10 - 3)0.25 = 1.75$

- $c_2 = ((d_r^{g'-1} + 1) - d_l^1)\delta = ((7+1) - 1)0.25 = 1.75$

- $c_3 = (D - d_r^{g'})\delta = (21 - 13)0.25 = 2$

- $c_4 = ((d_r^{G-1} + 1) + 1) - \varepsilon^D)\delta = ((16+1) - 11)0.25 = 1.5$

Finally, the global maximum deviation is identified by taking the maximum over c_1, c_2, c_3 and c_4.

$$c_{max}^2 = \max\{1.75, 1.75, 2, 1.5\} = 2$$

Supposing, that $\tilde{B}$ has rating degree Aaa and ALG has mis-predicted the rating degree of $\tilde{B}$. Then, the maximal deviation states that the false predicted rating degree of $\tilde{B}$ by ALG never exceed the following rating degree: $1 + 2 = 3$ which equals Baa2. Figure A.1 in the appendix illustrates Scenario 2.

7.2.3   Scenario 3

In Scenario 3, types of information 1 and 2 are known by ALG and OPT. This means that the bonds' attributes incorporate the necessary information to allow a distinction of the bonds according to their rating characteristics and their affiliation to the rating groups. The number of rating degrees in rating group Ω_g^G is given by: $N_g = d_r^g - d_l^g + 1$, with $g = 1, \ldots, G$.

Lemma 3. *The global maximal deviation between ALG and OPT, under the assumptions of Scenario 3, is equal to:*
$c_{max}^3 = \max\{A1, B1, B2, C1, C2\}$, *with:*
$A1 = \max_{g=1,\ldots,G} \;\; {}_{g \neq g'} \left\{ \frac{d_r^g - d_l^g}{2}\delta, (\lceil \frac{d_r^g - d_l^g}{2} \rceil)\delta \right\}$
$B1 = (\lceil \frac{(\varepsilon^D - 1) + d_l^{g'}}{2} \rceil)\delta$
$B2 = \frac{(\varepsilon^D - 1) + d_l^{g'}}{2}\delta$
$C1 = (\lceil \frac{d_r^{g'} - \varepsilon^D}{2} \rceil)\delta$
$C2 = \frac{d_r^{g'} - \varepsilon^D}{2}\delta$

Proof. The following two cases have to be investigated due to the fact that types of information 1 and 2 are known:

  i. $\tilde{B}$ belongs to rating group Ω_g^G, with $g \neq g'$

  ii. $\tilde{B}$ belongs to rating group $\Omega_{g'}^G$

<u>Case i:</u>
In this case, the rating groups contains either investment grade or non-investment grade bonds but the mixed rating group is not taken into account. ALG always issues a rating degree which is in the middle of the rating group because OPT always assigns a boundary rating degree. Two sub-cases can occur: (1) N_g is odd or (2) N_g is even.

Sub-Case i-1:

If N_g is odd, then the rating degree d_m^g which is the center of Ω_g^G exists. OPT always assigns either d_l^g or d_r^g, so a boundary degree to try to maximize the damage. The possible choices of OPT represent a symmetric choice. The deviation is:

$$c_1 = |d_m^g \delta - d_r^g \delta| \tag{7.37}$$

$$= (d_r^g - d_m^g)\delta \tag{7.38}$$

$$= (d_r^g - \frac{d_r^g + d_l^g}{2})\delta \tag{7.39}$$

$$= \frac{d_r^g - d_l^g}{2}\delta \tag{7.40}$$

Sub-Case i-2:

Now, N_g is even. The rating degree $d_{m'}^g$ lies approximately at the center of Ω_g^G. If ALG assigns $d_{m'}^g$ to $\tilde{B}$, then OPT will issue the boundary rating degree which is the farthest away from $d_{m'}^g$. Without loss of generality, assume that the finishing rating degree d_r^g is selected. ALG tries to react on this situation without leaving the center of Ω_g^G. Therefore, ALG tries to minimize the damage by assigning $d_{m'}^g + 1$. OPT also reacts on this situation and issues the opposite boundary degree d_l^g. Due to symmetry, the deviation is as follows:

$$c_2 = |d_{m'}^g \delta - d_r^g \delta| \tag{7.41}$$

$$= (d_r^g - d_{m'}^g)\delta \tag{7.42}$$

$$= (d_r^g - \frac{d_r^g - d_l^g + 1}{2})\delta \tag{7.43}$$

$$= (\lceil \frac{d_r^g - d_l^g}{2} \rceil)\delta \tag{7.44}$$

Concluding, the maximal deviation for Case i is obtained by taking the maximum over all rating groups excluding the mixed rating group:

$$c_{max}^{(i)} = A_1 = \max_{\substack{g=1,...,G \\ g \neq g'}} \{c_1, c_2\} \tag{7.45}$$

Case ii:

$\Omega_{g'}^G$ includes investment - as well as non-investment grade bonds. The following two sub-cases can occur:

    i. $\tilde{B}$ is of investment grade

    ii. $\tilde{B}$ is of non-investment grade

Sub-Case ii-1:

In a similar logic, ALG always issues a rating degree in the middle of the rating group to try to minimize the damage in comparison to OPT. The last admissible rating degree is $\varepsilon^D - 1$ and d_{m-i} is in the middle of the remaining part of the group. If d_{m-i} does not represent a real rating degree, then the maximum deviation equals:

$$B_1 = |d_{m-i}\delta - (\varepsilon^D - 1)\delta| \tag{7.46}$$

$$= ((\varepsilon^D - 1) - d_{m-i})\delta \tag{7.47}$$

$$= ((\varepsilon^D - 1) - \lceil \frac{(\varepsilon^D - 1) - d_l^{g'}}{2} \rceil)\delta \tag{7.48}$$

$$= (\lceil \frac{(\varepsilon^D - 1) - d_l^{g'}}{2} \rceil)\delta \tag{7.49}$$

Otherwise, the maximum deviation is as follows:

$$B_2 = |d_{m-i}\delta - (\varepsilon^D - 1)\delta| \tag{7.50}$$

$$= ((\varepsilon^D - 1) - d_{m-i})\delta \tag{7.51}$$

$$= \frac{(\varepsilon^D - 1) - d_l^{g'}}{2}\delta \tag{7.52}$$

Sub-Case ii-2:

Now, the first permitted rating degree is ε^D and d_{m-ni} is in the middle of the non-investment part of the group. If d_{m-ni} does not represent a real rating degree, then the maximum deviation is as follows:

$$C_1 = |d_{m-ni}\delta - \varepsilon^D\delta| \tag{7.53}$$

$$= (d_{m-ni} - \varepsilon^D)\delta \tag{7.54}$$

$$= (\lceil \frac{d_r^{g'} - \varepsilon^D}{2} \rceil - \varepsilon^D)\delta \tag{7.55}$$

$$= (\lceil \frac{d_r^{g'} - \varepsilon^D}{2} \rceil)\delta \tag{7.56}$$

Otherwise, the maximum deviation equals:

$$C_2 = |d_{m-ni}\delta - \varepsilon^D \delta| \tag{7.57}$$

$$= (d_{m-ni} - \varepsilon^D)\delta \tag{7.58}$$

$$= (\frac{d_r^{g'} + \varepsilon^D}{2} - \varepsilon^D)\delta \tag{7.59}$$

$$= \frac{d_r^{g'} - \varepsilon^D}{2}\delta \tag{7.60}$$

Concluding, the maximal deviation for Case ii is given by:

$$c_{max}^{(ii)} = \max\{B_1, B_2, C_1, C_2\} \tag{7.61}$$

Finally, the global maximal deviation, under the assumptions of Scenario 3, is equal to:

$$c_{max}^3 = \max\left\{c_{max}^{(i)}, c_{max}^{(ii)}\right\} \tag{7.62}$$

$\square$

Example 5. *Moody's rating scale with the $D = 21$ rating degree is reused. The $G = 5$ rating degrees introduced in Example 2 are also employed for this example. δ, the distance between two adjacent rating degree, is set to 0.25. The threshold rating degree ε^D is equal to 11 and the index of mixed rating group equals $g' = 3$. The starting and finishing rating degrees of the different rating groups are given in Example 2: $d_l^1 = 1$, $d_r^1 = 3$, $d_r^2 = 7$, $d_r^3 = 13$, $d_r^4 = 16$ and $d_r^5 = 21$. First, the deviations for the rating groups, excluded $\Omega_{g'=3}^G$ are computed:*

- $c_{g=1} = \frac{d_r^1 - d_l^1}{2}\delta = \frac{3-1}{2}0.25 = 0.25$

- $c_{g=2} = (\lfloor \frac{d_r^2 - d_l^2}{2} \rfloor)\delta = (\lceil \frac{7-4}{2} \rceil)0.25 = 2*0.25 = 0.5$

- $c_{g=4} = \frac{d_r^4 - d_l^4}{2}\delta = \frac{16-14}{2}0.25 = 0.25$

- $c_{g=5} = (\frac{d_r^5 - d_l^5}{2}\delta = \frac{21-17}{2}0.25 = 0.5$

Second, the deviations for the the mixed rating group, $\Omega_{g'=3}^G$, are determined:

- $c_{g'=3}' = \frac{(\varepsilon^D - 1) - d_l^{g'}}{2}\delta = \frac{10-8}{2}0.25 = (10-9)0.25 = 0.25$: *investment grade*

- $c_{g'=3}'' = \frac{d_r^{g'} - \varepsilon^D}{2}\delta = \frac{13-11}{2}0.25 = 0.25$: *non-investment grade*

Hence, the global maximum deviation for Scenario 3 is equal to:

$$c_{max}^3 = \max\{c_{g=1}, c_{g=2}, c_{g=4}, c_{g=5}, c'_{g'=3}, c''_{g'=3}\}$$
$$= \max\{0.25, 0.5, 0.25, 0.5, 0.25, 0.25\}$$
$$= 1$$

The obtained global maximum deviation can be interpreted in the following way. If $\tilde{B}$ has rating degree Aaa and an error has occurred during the prediction by ALG, then the false predicted rating degree of $\tilde{B}$ will never exceed the rating degree: $1 + 0.5 = 1.5$ which equals Aa2 on Moody's rating scale. For a better understanding, this example is represented in Figure A.3 in the appendix.

7.2.4 Competitive ACRP models

The analysis of the three scenarios allows to define competitive ACRP models. Before stating the theorem of competitive ACRP model, the definition of full ACRP models is given.

Definition 28. *An ACRP model is called full ACRP model if and only if the model is able to handle all three mentioned types of risk information:*

1. *rating characteristics*

2. *rating groups*

3. *rating degrees*

Next, competitive ACRP models are defined by the following theorem.

Theorem 1. *An ACRP model is competitive if a guarantee concerning the misclassification error for each scenario can be provided. Additionally, each full ACRP model is also competitive.*

Proof. As a guarantee for each scenario has to be provided, the ACRP model has to be able to handle all three types of risk information. In the opposite case, no guarantee for the scenario based on the missing type of information can be delivered. Therefore, the model has to be a full ACRP model. The previous stated lemmas, 1, 2 and 3, are applicable and give the respective guarantees for each scenario. Additionally, the relations between the different types of informations, stated in the propositions, 2, 3 and 4 are the basis of the reasoning. Thus, the lemmas and propositions proves the whole theorem. □

7.2.5   Comparison of different ACRP models

The competitiveness of the four ACRP models introduced in the previous chapters is analyzed. Recall quickly the four ACRP models. The first ACRP model is based on artificial neural network and is denoted by *ACRP_ANN*. The model is introduced by [7] and is described in detail in Section 5.2. *ACRP_ANN* focuses on the third type of information, rating degrees, and tries to predict directly the correct rating degree. During the prediction process, the remaining two types of information, rating characteristics and rating groups, are not taken into account.

The basis of the second ACRP model are support vector machines and is denoted by *ACRP_SVM*. The detailed description of *ACRP_SVM* is given in Section 5.3 and is introduced by [48]. The focus of this models are the rating groups. Type of information 2 is the highest risk information furnished by *ACRP_SVM*. Even if type of information 3 is available, the model would ignore this information during its prediction process.

Support vector domain description is used in the third ACRP model and is developed by [40]. The ACRP model is noted by *ACRP_SVDD* and is described in detail in Section 5.4. This model also focuses only on type of information 2. *ACRP_SVDD* predicts the rating group of unrated bonds and is not able to offer a deeper insight of the credit risk by using type of information 3.

Finally, the last analyzed ACRP model is based on support vector domain description combined with linear regression and is developed during this thesis. The full description of the model is made in Section 6.1 and is noted by *New_ACRP*. *New_ACRP* is able to handle all three types of risk information. The subsequent table summarizes on which types of risk information the different ACRP models focus.

	Type of risk information 1	Type of risk information 2	Type of risk information 3
ACRP_ANN			X
ACRP_SVM	X	X	
ACRP_SVDD	X	X	
New_ACRP	X	X	X

Table 7.1 Comparison of ACRP algorithms based on the types of information

The partition of the four introduced ACRP models into competitive and non-competitive models is made by applying Theorem 1. The competitiveness of the models is summarized in the following table.

	Competitive	Non-competitive
ACRP_ANN		X
ACRP_SVM		X
ACRP_SVDD		X
New_ACRP	X	

Table 7.2 Competitiveness of ACRP models

Table 7.2 shows that only *New_ACRP* can be considered as a competitive model. It is the only model which utilizes all three types of information to predict the rating degrees. *ACRP_SVM* and *ACRP_SVDD* only try to predict the right rating group. Therefore, they are by definition non-competitive. Even if *ACRP_ANN* tries to predict the right rating degrees, it is not competitive. The non-competitiveness of *ACRP_ANN* is given by the missing of a mis-classification guarantee for type of information 2.

Concluding, the use of all three types of information increase the accuracy of ACRP models. The competitiveness of ACRP models allows to set up a guarantee of mis-classification in the worst-cases and helps to increase the trustability of models by investors [2].

7.3 Empirical analysis of the new ACRP model

First, the setting of the empirical analysis is described. The new developed ACRP model, *New_ACRP* is tested against two benchmark models. The first benchmark model is taken from the literature and is based on artificial neural network. The same configuration of the ACRP model as described in [7] is used. This ACRP model is denoted by *ACRP_ANN*. Its configuration is as follows:

Learning rate in the hidden layer, λ_{hid}	1.0
Momentum in the hidden layer, μ_{hid}	0.7
Learning rate in the output layer, λ_{out}	0.1
Momentum in the output layer, μout	0.7
Maximal number of iterations, i_{max}	3000 or 10000
Number of hidden Neurons, $Hidden\_Neurons$	22 or 25

Table 7.3 Configuration of *ACRP_ANN* taken from [7]

A second ACRP model which figures as benchmark is a linear regression and is denoted by *ACRP_LR*. The dataset is obtained by collecting the basic features and historical prices of

---

[2]The competitive analysis of ACRP models is published in [35]. The work was realized by myself and the idea behind the undertaken competitive analysis was raised by Robert Dochow and myself.

real bonds listed on the stock exchange in Frankfurt. According to [39], the collected bonds fulfill the following requirements:

- The maturity of the bonds lies around 10 years

- The coupon rate of the bonds are around 4%

- The bonds are listed in *EUR*

- The bonds have to be listed for at least 2 years

Additionally, the bonds have to be rated to be added to the dataset. The "Frankfurter Börse" displays the Moody's rating degrees of the listed bonds. In the case, that Moody's has not rated one bond, then the rating degree of S&P or Fitch is searched and is mapped into Moody's rating scale. After the collection of the bonds, 410 bonds are in the dataset. The size of the dataset lies in the middle according to the used dataset in the literature [7, 48, 40]. Two different settings are investigated.

1. The whole dataset is used

2. The dataset is harmonized and the number of bonds for each rating degree is equal

7.3.1 Setting 1

In Setting 1, the whole dataset is used and it is divided into a training and testing set. The partition between training and testing set is as follows: 77% of the bonds, which are $n = 276$ bonds, represent the training bonds and the remaining $\tilde{n} = 134$ bonds, so 33%, are the testing bonds. The percentage of bonds in the training respectively testing set is similar to the partition founded in the literature [7, 48, 40].

Based on the competitive analysis, the accuracies of the investigated models are determined by computing the absolute distance between predicted rating degree $\tilde{R}_i$ and the correct ones R_i^D, with $i = 1, \ldots, \tilde{n}$. The subsequent formula allows to compute this distance:

$$\Delta_i = |\tilde{R}_i^D - R_i^D|, \qquad \forall i = 1, \ldots, \tilde{n} \tag{7.63}$$

Two different cases are analyzed to determined the accuracies. First, the predicted rating degree has to be the correct one. Second, the predicted rating degree should lie within one degree of the correct one. Due to the fact, that *New_ACRP* employs a cardinal rating scale, the two cases are modeled by setting $\Delta = 0.23$ for case 1 and $\Delta = 0.46$ for case 2. The

accuracy of a model is determined by computing the percentage of correct predicted rating degrees. The subsequent formula defines if the prediction is correct or not.

$$
\mathbb{1}_i = \begin{cases} 1 & \Delta_i <= \Delta \\ 0 & \Delta_i > \Delta \end{cases} , \qquad , \forall i = 1, \ldots, \tilde{n} \tag{7.64}
$$

The accuracy is given by:

$$
Acc = \frac{1}{\tilde{n}} \sum_{i=1}^{\tilde{n}} \mathbb{1}_i \tag{7.65}
$$

Remember that *New_ACRP* needs different input variables. First, δ, the distance between two adjacent rating degrees, is fixed to 0.25. Second, the trade-off constant C is set to 1. The two remaining input variables are fixed during the testing. In a first test, the regulator constant λ is set to 0.85 and the kernel factor σ is varied. Under this test setting, the best accuracy of *New_ACRP* is obtained for $\sigma = 0.190$ and $\sigma = 0.185$:

Correct rating degree	Rating degree within one degree	σ
56.72%	78.36%	0.190
55.97%	79.11%	0.1985

Table 7.4 Global accuracies of *New_ACRP*

After the determination of the best values of σ, λ is varied. The tests show that the value of λ has no impact on the accuracy. Afterwards, the obtained accuracies are compared with those received for the benchmark models. Starting with the basic benchmark model, *ACRP_LR*, the subsequent accuracies are given:

Correct rating degree	Rating degree within one degree
26.87%	58.96%

Table 7.5 Global accuracies of *ACRP_LR*

For the benchmark model, *ACRP_ANN*, four different configurations are tested. First, the maximal number of iterations i_{max} is fixed to 3000 and the number of hidden neurons, *Hidden_Neurons*, is set to 22. In the second test, i_{max} is increased up to 10000 and *Hidden_Neurons* remains the same. The third test is as follow: $i_{max} = 3000$ and *Hidden_Neurons* $= 25$. In the last test, *Hidden_Neurons* remains 25 and i_{max} is again increased up to 10000. The following table illustrates the obtained accuracies for each of the four test runs.

	Correct rating degree	Rating degree within one degree
Test 1	33.58%	67.16%
Test 2	30.60%	63.43%
Test 3	37.31%	64.93%
Test 4	33.58%	61.19%

Table 7.6 Global accuracies of *ACRP_ANN*

According to the tables 7.4, 7.5 and 7.6, *New_ACRP* outperforms significantly the two used benchmark models, *ACRP_ANN* and *ACRP_LR*. To get a deeper insight view, the accuracies of *New_ACRP* for each rating degree are illustrated in the following two figures.
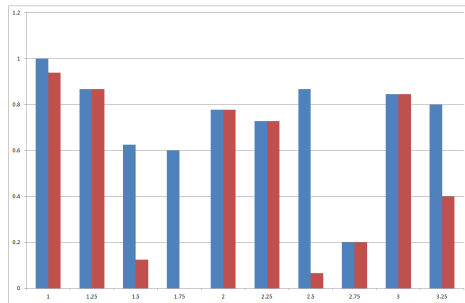
(a) Accuracies for each rating degree with sigma = 0.19, red: correct rating; blue: within one rating degree
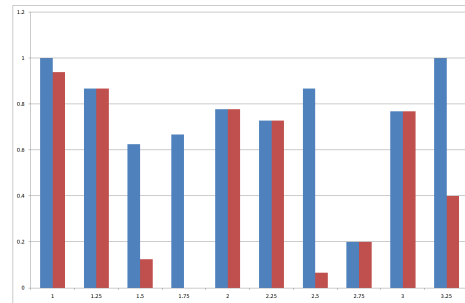
(b) Accuracies for each rating degree with sigma = 0.185, red: correct rating; blue: within one rating degree

If the predicted rating degrees can differ by one degree from the correct one, then an accuracy of at least 60% for each rating degree is achieved. Therefore, *New_ACRP* offers a good estimation of the bonds' credit risk to investors. Additionally, the accuracy of the prediction of the correct rating groups are investigated. The subsequent table indicates the percentage of accuracy for each of the three rating groups:

Rating Group	Accuracy
1	91.1%
2	74.0%
3	60.7%
All	78.36%

Table 7.7 Accuracies of each rating group

The accuracy of any rating groups is less than 60%. According to the different types of risk information introduced in the Section 1 of this chapter, the used attributes incorporate type of information 2.

7.3.2 Setting 2

In Setting 2, the impact of an harmonized dataset on the accuracy is analyzed.The smallest number of bonds for one single rating degree in the used dataset equals 15. For this reason, the number of bonds for all the other rating degrees is also restricted to 15. For each rating degree, 15 bonds are randomly selected to build the new dataset. In this way, the harmonized dataset comprise 150 bonds. 10 bonds of each rating degree represent the training bonds. The remaining bonds are the testing bonds. The test remains the same as for Setting 1. Starting with *New_ACRP*, the regulator constant λ has again no influence on its accuracy. Therefore, λ is fixed to 0.85 and the best result is obtained with $\sigma = 0.19$. Afterwards, the tests with the two benchmark model, *ACRP_LR* and *ACRP_ANN*, are undertaken. The obtained global accuracies for the three ACRP models are given in the following table.

Model	Correct rating	Within one degree
New_ACRP	54%	76%
ACRP_LR	28%	56%
ACRP_ANN		
Test 1	26%	56%
Test 2	32%	62%
Test 3	28%	60%
Test 4	32%	54%

Table 7.8 Global Accuracies of each ACRP model

According to Table 7.8, *New_ACRP* outperforms again significantly the benchmark models. Comparing the accuracies obtained under Setting 1 and Setting 2, a real improvement by using a harmonized dataset is not determined. The number of bonds for each rating degree has not to be similar to obtain accurate prediction results. This is an important ascertainment because in the practical use, investors often collect a different amount of bonds for each rating degree. Afterwards, the impact of the harmonized dataset on the accuracies of *New_ACRP* for each rating degree is investigated.
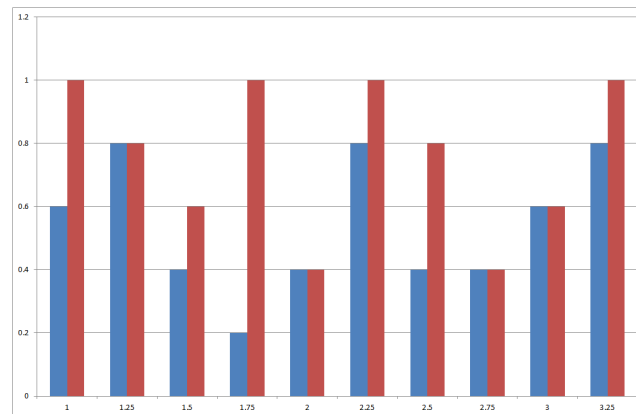
Fig. 7.3 Accuracies for each rating degree, blue: correct rating; red: within one degree

The accuracies slightly decrease and in this way, the following assumption can be made:

With a greater amount of bonds, the prediction of the ACRP model becomes more accurate.

Finally, the accuracies concerning the rating groups are analyzed. The obtained accuracies are summarized in the following table.

Group	Accuracy
1	73.33%
2	85.00%
3	73.33%
All	78.00%

Table 7.9 Accuracies of each rating group for an harmonized dataset

The accuracy of each rating group does not differ significantly if using an harmonized or a non-harmonized dataset. Concluding, an equal number of bonds in each rating group or for each rating degree has not to be assumed to guarantee accurate credit rating predictions. However, it is shown that the whole number of bonds should be sufficient and overcome 150 to improve the accuracy of the models.

7.4 Conclusion

This chapter gives an additional answer to Research Question 2 by undertaking a competitive and an empirical analysis of *New_ACRP*. First, the needed framework to undertake a competitive analysis concerning ACRP models is introduced. This is necessary because a competitive analysis of ACRP models is completely missing at the moment. Afterwards,

the framework is applied and the analytically obtained guarantees are given. It is shown that *New_ACRP* is the only competitive ACRP model. For each of the three types of risk information a mis-classification guarantee in the worst-case can be stated.

After the competitive analysis, *New_ACRP* is tested by using a dataset of real bonds. Two different settings are investigated: (1) the whole dataset of 410 bonds and (2) an harmonized dataset of 150 bonds. The accuracies obtained for *New_ACRP* are compared to those for two benchmark models. The first benchmark model, *ACRP_LR*, is based on linear regression and the second benchmark model, *ACRP_ANN*, is based on artificial neural network. *ACRP_ANN* is taken from the literature, [7]. The results of the empirical analysis have shown that *New_ACRP* has significantly outperforms *ACRP_LR* and *ACRP_ANN* on the two datasets. Additionally, the analysis has asserted that the number of bonds for each rating degree does not have to be equal. It is also shown that with an increasing number of employed bonds, the accuracy of the model is improved.

Chapter 8

Prototype: Implementation of the ACRP model

This chapter describes the development of a prototype of the new ACRP model. The ACRP model is integrated into the personal financial planning (PFP) tool, LifeCharts, to reach the target group: private investors. The integration of the prototype is described in detail. In addition, the use of the prototype is explained. Finally, the chapter is concluded.

8.1   Development of the prototype

The new developed ACRP model has been completely implemented using the programming language VB.net according the mathematical programs given in Chapter 6 to undertake the empirical analysis in Chapter 7. Nevertheless, this first implementation of the new ACRP model does not fulfill all the requirements, stated in Chapter 2, to be usable by private investors. Before, describing the implementation of the new ACRP model and its integration into LifeCharts, some explanations are given. First, the programming language VB.net was chosen because LifeCharts is developed in Microsoft Excel. MS Excel is used to reach a greater public because it is widely used and its main functionalities are known. Second, remember the stated requirements for an ideal ACRP model. These requirements are as follows:

1. ACRP model has to handle sovereign and corporate bonds simultaneously

2. ACRP model has to reduce the user-given input variables which are specific to the utilized methods

3. ACRP model has to handle bonds with different maturities

4. ACRP model has to handle bonds with different types of coupon rates

5. ACRP model has to incorporate a portfolio selection tool

Remember that LifeCharts suggests to invest a portion of the saving capital in view to achieve the desired point of financial freedom. A brief description of LifeCharts is given in Chapter 2. Saving books offered by traditional banks do not provide sufficient returns due to the zero-interest rate policy carried out by the European Central Bank. Therefore, the proposed portion of saving capital should be invested into more risky investments, like bonds or shares. In case of a possible bond investment, the new ACRP model could be used to obtain a risk evaluation of the selected bonds. At this stage, the ACRP model will be integrated.

Now, the implementation is described in detail. Due to the composition of the new ACRP model, the first requirement is directly fulfilled. The new ACRP model has been developed to handle sovereign and corporate bonds simultaneously. However, in its actual implementation, the new ACRP model needs the right values for the input variables: σ, δ and λ. Private persons do not have the necessary knowledge about the methods used in the ACRP model (support vector domain description (SVDD) and linear regression (LR)) to determine these values. In order, to fulfill the second requirement of the ideal ACRP model, the new ACRP model has to be modified to be usable by private investors. According to the empirical analysis undertaken in Chapter 7, the regulator constant λ has no impact on the accuracy of the model and it can be set to 0.85 in advance. The chosen value of λ is also proposed by [40]. The distance between two adjacent rating degrees δ is set to 0.25, once for all. This means that a bond with a predicted rating degree around 1 has very little risk and a bond with a predicted rating degree greater or equal to 6 has very high risk. The problem is the fixing of the value of the kernel factor σ because this value has a great impact on the accuracy of the model. The optimal value of σ depends of the given bonds in the training set. Remember that the training set is used to build and calibrate the model. However, the training set is usually a user-given input variable. In the view to simplify the model to make it usable by private persons, the training set is given and represents a suggestion of bonds in which the users can invest. The fixing of the training set has two main advantages. First, the users do not have the task to find bonds with similar characteristics concerning the coupon rates and the maturities to generate an own training set. Second, with a fixed training set, σ is also fixable in advance. To sum up, all the input variables depending of the used underlying methods (SVDD and LR) are eliminated. Thus, the second requirement of an ideal ACRP model is also fulfilled.

A graphical user interface (GUI) has been developed on the basis of Excel to integrate the ACRP model in the optic of LifeCharts. The main page of the integrated ACRP model is

illustrated in Figure B.11 in the appendix. In Figure B.11, it can be seen that users can also add own bonds to obtain a risk evaluation. To simplify the entering of own bonds, the users have only to indicate the "Wertpapierkennnummer" (WKN) or "International Securities Identification Number" (ISIN) of their selected bonds. However, one restriction concerning the choice of own bonds is imposed. The own bonds must have similar characteristics concerning maturity and coupon rate as the bonds in the training set. This is necessary to guarantee an accurate risk evaluation [39]. However, this restriction has also two main advantages. First, the training bonds have a maturity of around 5 years. The fixing of the maximal maturity allows to fulfill nearly completely the third requirement of the ideal ACRP model. Second, the selected own bonds by the users must have the same type of coupon rates. In this way, the problem of handling bonds with different types of coupon rates is overcome. The fourth stated requirement for the ideal ACRP model is partially fulfilled. With the stated restriction, the users have no great drawback concerning the choice of bonds. The taken choice of 5 year of maturity is plausible because private investors usually reevaluate their financial situation after 5 years. Only fixed coupon rate bonds are in the training set because this type of bonds represents the majority of the issued bonds on the market. The training bonds are also stored by their WKNs and can also be seen on the main page (see Figure B.11 in the appendix). With the help of the WKNs or ISINs, the ACRP model automatically retrieves the following features of the bonds from the web page "ariva.de":

- Historical prices over 2 years

- Ask - and bid prices

- Maturity

- Coupon rate

- Rating degree, in case of training bonds

The download of the features for all the bonds is undertaken at the same time to reduce the computation time. Afterwards, the ACRP model computes the necessary attributes. The needed attributes are as follows:

- Value-at-Risk (VaR)

- Conditional Value-at-Risk (CVaR)

- Duration

- Modified Duration

- Convexity

- Liquidity

- Spread between the bonds and the German Government Bond

According to the computations given in Chapter 6, the rating degrees of the bonds are predicted. Furthermore, for each bond, its mean return over the taken period of the historical prices is determined. The mean return gives the expected performance of the bonds. The mean return, the predicted rating group and the predicted rating degree of the bonds are transfered back to the new developed GUI in LifeCharts and displayed to the users. The output of the ACRP model in LifeCharts is represented in Figure B.12 in the appendix. The rating group affiliation represents a coarser risk analysis of the bonds. 5 different rating groups are defined in the model. The rating groups and their risk interpretation are as follows:

1. Rating group 1: very little risk

2. Rating group 2: little risk

3. Rating group 3: medium risk

4. Rating group 4: high risk

5. Rating group 5: very high risk

Finally, the rating degree represents the individual risk evaluation of the bonds given by the model. The rating degree are expressed in a cardinal rating scale. Therefore, each bond obtains its own rating degree.

After the users have obtained a risk evaluation of the bonds, they have two choices. First, they can use this new information, i.e., the predicted rating degrees and mean returns, to build a portfolio by themselves. In this case, the work of the model is done and all the remaining steps have to be undertaken by the users. Second, investors want that the model computes the optimal portfolio of bonds under several constraints. Therefore, a portfolio selection tool is integrated into the ACRP model. In this way, the last and fifth requirement for the ideal ACRP model is fulfilled. The basic idea of the new integrated portfolio selection tool is the Markowitz portfolio selection theory [65]. In [65], the optimal portfolio is given by maximizing the expected return of the portfolio under the constraint of not exceeding the given risk. The risk is computed by the standard deviation. However, the new ACRP model does not compute the standard deviation but determines the rating degree of each bond to

obtain a risk evaluation. Therefore, instead of the standard deviation, the rating degrees are used in the new designed portfolio selection tool. According to [23, 81], the expected rating degree of the obtained portfolio is computed by the following formula:

$$\tilde{R}_P^D = \sum_{i=1}^{\tilde{n}} s_i \tilde{R}_i^D \tag{8.1}$$

s_i, for $i = 1, \ldots, \tilde{n}$, is the proportion of each bond in the portfolio. $\tilde{n}$ is the total number of bonds and $\tilde{R}_P^D$ represents the expected rating degree of the portfolio. Similar to the Markowitz theory, the rating degree of the portfolio should not exceed a given rating degree by the users. The new designed portfolio selection problem is given by the following mathematical program:

<div align="center">

Mathematical Program 8.1

</div>

$$\begin{aligned} \max \qquad & r_P = \sum_{i=1}^{\tilde{n}} s_i r_i \\ \text{subject to} \quad & \tilde{R}_P^D \leq R_P^D \\ & \sum_{i=1}^{\tilde{n}} s_i = 1 \\ & 0 \leq s_i \leq 1 \quad \forall i = 1, \ldots, \tilde{n} \end{aligned}$$

with, r_P the expected return of the portfolio and r_i, with $i = 1, \ldots, \tilde{n}$, the mean returns of the bonds. The first constraint of Mathematical Program 8.1 indicates that the expected rating degree of the portfolio should not exceed the user-given degree. The second constraint means that the sum of the proportions of the bonds included in the portfolio always equals 100%. Thus, if the portfolio contains only one single bond in which the users have to invest, then s_i equals 100%, for this bond B_i. Finally, the last constraint forbids short-selling. In this way, only bonds can be sell which are in the portfolio. Users can not speculate on a decrease of the bond price to buy it back in the case that users take it into its portfolio without owning it legally. To maintain the usability of LifeCharts, the users have only to indicate the maximal rating degree R_P^D after all the bonds were rated. In this way, they can use the obtained rating degrees of the bonds to determine an appropriated maximal rating degree for the portfolio. After R_P^D is given by the users, the model computed according to the Mathematical Program 8.1 the optimal portfolio. The obtained portfolio is transfered to LifeCharts and is displayed in two different views. First, on the main page of the integrated ACRP model, the different proportions are given for each bond (see Figure B.13 in the appendix). Second, a more detailed view on the portfolio is furnished to the users if the button "Detaillierte Ansicht

des Potfolios" is pushed. In the detailed view, only the bonds included in the portfolio are displayed, as well as the portfolio's characteristics: expected return and rating degree. Furthermore, a pie-chart illustrates the partition of the bonds in the portfolio to facilitate the understanding. An explanation also tells the users the benefits of this proposed bond investment on the achievement of their desired FF-point. The detailed portfolio view is represented in Figure B.14 in the appendix.

The obtained prototype has also been tested with the help of different scenarios: missing entries and wrong entries of users. These tests have allowed to eliminate the last errors due to the integration. Additionally, several members of the university department have made some practical tests to check the handling of the prototype. The understanding of the results of the prototype and especially, the procedure to handle the ACRP model are mentioned laudable by the testing persons. Finally, all the stated requirements for the ideal ACRP model are fulfilled by the prototype.

8.2   Exemplary scenario

The exemplary user, John Smith, is used to go through the developed prototype. In Chapter 2, John Smith has used LifeCharts to determine his FF-point. LifeCharts has identified that an increase of his savings is necessary to achieve his desired FF-point at 69 years. As seen in Chapter 2, John obtains the information concerning the necessity to save an additional amount of money from LifeCharts, but the difference is now that the prototype proposes John to realize this measure by undertaking a bond investment. Figure B.9 in the appendix illustrates this case that John has the possibility to choose a risk analysis of bonds. A short description of the benefits of the integrated ACRP model is also given to help John to take a decision if a bond investment is interesting for him or not.

After John has selected to undertake a bond investment, he starts the integrated ACRP model by pushing on the button named: "Starten". A more detailed description of the integrated ACRP model is given. The description explains how the model works and what are the requirements which own bonds have to fulfill. Additionally, the outcomes of the model are asserted such that John can understand the obtain results without any problems. Furthermore, a note with the additional components which the model needs is displayed. The shown description and note can be seen in the figures B.10a and B.10b in the appendix.

John should read these instructions to have the necessary understanding before starting the risk evaluation of the bonds. In Figure B.11 in the appendix, John directly sees the set of bonds which is proposed by the prototype. However, he has also the possibility to indicate own bonds. In this example, assume that John has heard about bonds in the news and he has also found their WKN on the webpage of the stock exchange in Frankfurt. John indicates the WKNs for the selected bonds in the designated place. This place is the list denoted by "Anleihen, die Sie noch zusätzlich bewertet haben wollen:". Only to mention, own bonds have not to be indicated to use the risk evaluation of the prototype. John starts the risk evaluation of the bonds by pushing the button "Bewertung starten". Afterwards, the prototype undertakes all the necessary steps to obtain the needed information, like historical prices, coupon rates, bid- and ask prices. The computations are made in the background so no additional interaction with the prototype is needed. After the model has done all the computations, the obtained results are displayed to John. Figure B.12 in the appendix shows that the return, the rating group and the rating degree of each bond is given.

With this information, John is already in a better position to take a decision because he knows the credit risk of the bonds. He has two possibilities to employ this new acquired knowledge. First, he undertakes the selection of the bonds by himself. This means that he has to portion the amount of money which he wants to invest and that he has to select the appropriated bonds without any additional help from the prototype. Second, John can use the built-in function of the prototype to construct the optimal portfolio. In this case, John indicates the maximal rating degree which he is willing to accept for the obtained portfolio. John can used the obtained risk evaluation of the bonds to set the maximal rating degree. The portfolio selection tool of the prototype is activated with a positive answer from John of the question if the model should determine the optimal portfolio out of the bonds. 1.3 is the maximal rating degree, set by John.

The model maximize the return of the portfolio under the constraint that the resulting portfolio should not exceed the given maximal rating degree. All the necessary computations are undertaken by the model in the background such that the usability of the prototype is maintained. After the optimal portfolio is computed, the percentage of amount which John should invest in each bond is displayed. In Figure B.13 in the appendix, the obtained partition can be seen for the employed set of bonds. A more detailed view on the resulting portfolio is proposed to John by pushing the button "Detaillierte Ansicht des Potfolios". In this view, only the bonds constituting the portfolio are displayed. Additionally, a pie-chart illustrates the partition of the bonds in the portfolio to transmit this information in a simple way. The

portfolio's characteristics, its return and rating degree, are also indicated. Its return equals: 0.1115 % and its rating degree is: 1.3. Thus, the given maximal rating degree is not exceeded and John can expect to rise his savings by 1.77 € per year with the investment into the proposed portfolio. However, one has to mention that the periodic payments guaranteed by the coupon rates are not taken into account to compute the average return or to estimate the increase of his wealth. These payments will additionally increase his wealth. Finally, the benefits of this investment are briefly explained to John. John gets the information that with the proposed investment, he can reduce his yearly saving amount to achieve his desired FF-point due to the fact of the expected return and the coupon payments. Or, he can achieve his FF-point at an earlier point in time if the total amount of 1583 € per year is invested. The detailed view of the portfolio with all the additional information is represented in Figure B.14 in the appendix.

After the detailed description of the optimal portfolio is given, a global summary is offered to John. As the prototype represents the integration of an ACRP model into Lifecharts, the global summary indicates the results of the evaluations of LifeCharts and the ACRP model. The example illustrates the benefit which consists the integration of an ACRP model into a PFP tool. With the integrated ACRP model, the PFP tool, in this example LifeCharts, is able to help the users to realize the proposed financial plan. In this way, the PFP tool becomes a real decision help for users.

8.3   Conclusion

This chapter answers Research Question 3. The developed prototype integrates the new ACRP model into the PFP tool, LifeCharts. The prototype represents the proof of concept mentioned in Chapter 2. First, the detailed implementation and integration of the ACRP model into LifeCharts is presented. Remember that to overcome the problem of handling bonds with different maturities, the following assumption is taken: *After at least* 5 *years, investors should reevaluate their financial situation*. This assumption also schedules the re-balancing of the portfolio. After 5 years, a new set of proposed bonds is integrated into the prototype and in this way, a new optimal portfolio is determined. At the moment, there is a limitation concerning bonds with different types of coupon rates and the developed prototype is only able to handle bonds with fixed coupon rates. All the other identified requirements are completely fulfilled by the prototype.

Second, the benefits of the prototype are demonstrated with the help of the exemplary user, John Smith, who was introduced in Chapter 2. John uses the prototype to obtain a risk evaluation of the bonds and to determine the optimal portfolio. In this way, the prototype is explained from the perspective of an user. The given example shows the advantage of the prototype to the basic version of LifeCharts. Instead of only indicating several measures to the user, the prototype proposes to realize the needed increase of the savings by a bond investment. To undertake the suggested bond investment, the prototype leads the user through the steps of risk evaluation and portfolio building. The user obtains the necessary explanation which effects such an investment would have to the chosen FF-point. Thus, the prototype represents a more complete PFP tool in comparison to PFP tools without an integrated ACRP model.

The main advantage of the integration of the ACRP model is to reach the target group: private persons. In this way, the used ACRP model in the prototype represents a real and usable alternative to the well-known rating agencies like Moody's. Nevertheless, the integration has also uncovered several drawbacks. First, by fixing the kernel factor σ, the model is deprived of its flexibility. A last disadvantage is that the prototype only indicates the optimal portfolio to investors but it does not allow them to really make this investment. A trading tool is missing. Investors only have the possibility to employ an external trading tool to finalize the proposed investment.

Chapter 9

Conclusion

9.1 Summary

First,the advantages of financial planning are briefly discussed. It is shown that financial planning helps private persons to prevent the risk of poverty among the elderly. Afterwards, a scenario of an exemplary user is discussed to illustrate that the integration of automated credit rating prediction (ACRP) models into financial planning tools would represent a benefit. In this scenario, the user has a gap between his savings and his needed consumption capitals. To close this gap, the used financial planning tool only suggests to increase the savings without given the necessary information how he can realize this increase. With the help of an ACRP model integrated into the financial planning tool, the user should have the possibility to obtain a risk evaluation of bonds and obtain a suggestion how he should invest into the available bonds. To maintain the usability of the financial planning tool, five requirements for an ideal ACRP model are stated:

- ACRP model has to handle sovereign and corporate bonds simultaneously

- ACRP model has to reduce the user-given input variables which are specific to the utilized method

- ACRP model has to handle bonds with different maturities

- ACRP model has to handle bonds with different types of coupon rates

- ACRP model has to incorporate a portfolio selection tool

Second, the necessary definitions of the financial domain are given. The general rating procedure of the main and well-known rating agencies, like Moody's Investors Service (Moody's),

Standard and Poor's (S&P) and Fitch's Ratings (Fitch) is introduced. Their procedure is very closed and one reason is that they use private information obtained by meetings with decision makers to evaluate the financial situation of companies or states. With the start of the financial crisis in 2009, the rating agencies try to make their rating procedure more transparent but there still exit lacks. Furthermore, the bond market and the different types of bonds are defined. In this work, the focus lies on plain-vanilla bonds, i.e., bonds without any additional options, like a call or put option or float-able coupon rates. The distinction between corporate and sovereign bonds is also made. Corporate bonds are issued by companies and bonds issued by countries are called sovereign bonds. Additionally, it is shown that the credit risk of bonds is already fully or at least partially included in the price of the bonds. This ascertainment allows to develop ACRP models.

In a next step, different classification methods, like artificial neural networks (ANN), support vector machines (SVM) and support vector domain description (SVDD), are presented. For each of the introduced methods, the pseudo-code is given to simplify their implementation into any programming language and to ease the understanding. The methods are described in an abstract way such that they are not limited to one single use-case but the whole area of applications is shown. After the general description of the several methods the focus is turn to their application in the financial sector and especially concerning the prediction of bonds' rating degrees. Three different ACRP models developed by one of the introduced methods is described in detail. These ACRP models allow to answer Research Question 1. This question handles about the possibility to develop ACRP models based on public available information. The mathematical programs for each ACRP model is given and their advantages and drawbacks are discussed. It is shown that the majority of the ACRP models is designed to predict the rating degrees or rating groups of corporate bonds. For corporate bonds, the financial balance sheet of the issuing company is publicly available and can be used to evaluate the financial situation of the company. In this way, the credit risk of the bonds is indirectly determined. Additionally, the majority of the ACRP models only try to predict the correct rating group. It has to be mentioned that a rating group is defined as the merging of several rating degrees. A rating group represents a coarser evaluation of the risk. Furthermore, there exists no ACRP model which is able to handle sovereign and corporate bonds simultaneously and to predict their rating degrees.

To overcome this missing link and allow an investor to use one single ACRP model to obtain an risk evaluation of corporate and sovereign bonds, a new ACRP model has been developed. The new ACRP model is based on SVDD and LR. SVDD is used to identify a first coarse

partition of the bonds. The bonds are divided according 5 different rating groups. The description of the groups is as follows:

- Rating group 1: very small risk

- Rating group 2: small risk

- Rating group 3: medium risk

- Rating group 4: high risk

- Rating group 5: very high risk

Afterwards, in each rating group, LR allows to determine the final rating degree. One big advantage of the new method is that the predicted rating degrees are given in a cardinal rating scale. In this way, each bond obtains its individual rating degree and investors can better differentiate the bonds according to their risk as using the standard ordinal alphanumerical rating scale introduced by the rating agencies. A step-by-step description of all the formulas and mathematical programs is given to facilitate the understanding and the reproduction of the ACRP model. After the theoretical description of the model, a competitive and empirical analysis is undertaken to investigate its theoretical and empirical behavior. Before undertaking the competitive analysis, the framework is defined. Three different types of risk information are stated: (1) rating characteristics, (2) rating group and (3) rating degree. Additionally, three worst-cases scenario based on the different types of information are defined. Theorem 1 defines a competitive ACRP model. Consequently, a comparison between the three introduced ACRP models from the literature and the new developed model according their competitiveness is made. The new developed ACRP model can only be considerate as competitive because for each worst-case scenario, a mis-classification guarantee can be given. Afterwards, its empirical behavior is investigated. During the empirical analysis, the new developed model is tested against two ACRP models, one based on ANN and another based on LR. The first benchmark model is taken from the literature [7]. The new developed model has shown a significantly better performance in predicting the rating degrees of the bonds as the other two models. In this way, Research Question 2 is completely answered. The development and the analysis of the new ACRP model has shown that it is feasible to develop ACRP models which predicts simultaneously the rating degree of corporate and sovereign bonds.

Finally, a prototype is developed to allow private investors to employ the new developed ACRP model. For this reason, the model is integrated into a personal financial planning tool,

called LifeCharts. During the implementation of the prototype, the stated requirements for an optimal ACRP model are attempted to be respected by the prototype. First, the number of inputs, which investor have to enter, is reduced. The integrated model proposes a set of bonds to the investors. This set represents, on the one hand, the training set of the model and on the other hand, the set of possible bonds in which investors can make an investment. The advantage of fixing a set of bonds is that the kernel factor σ can also be set in advance and does not represents a user-given input variable. Nevertheless, investors have also the possibility to get a risk analysis of own bonds. To facilitate the handling of the model, the input variables, λ and δ, are also set in advance. This simplification of the model is undertaken to allow investors without any prior knowledge about the theoretical structure of the ACRP model to handle it. Second, the investors have to maintain several restrictions if they select own bonds. The first restriction is that the own bonds have to have a similar maturity. This restriction allows that the ACRP model fulfills partially the stated requirement that an ACRP model has to handle simultaneously bonds with different maturities. The second restriction limits the own bonds to fixed coupon rates. This restriction allows to bypass the problem of having to handle bonds with different types of coupon rates. After the risk evaluation of the bonds , the model is able to determine the optimal portfolio. The optimal portfolio is defined in the following way: maximization of its return under the constraint that the resulting rating degree does not exceed the user-given maximal rating degree. In the end, the integrated ACRP model proposes a portfolio and explains to the investors the advantages of this investment and the direct consequences on their point of financial freedom (FF-point). Remember that the FF-point is the point in time at which a person has saved enough money to finance its life without any additional income from work. The development of the prototype answer Research Question 3. It illustrates that private persons are able to use ACRP models. Finally, the prototype proves that the integrated ACRP model fulfills all the stated requirements for an ideal ACRP model.

9.2   Future Research

Several open questions for future research are identified. First, the empirical analysis has shown that LR is the weakest part used in the new developed ACRP model to predict the rating degrees. In [7], logistic regression is given as the state of the art method for financial predictions. The investigation of the accuracy of the new developed ACRP model in the case if LR is substituted by logistic regression can be undertaken. Probably, logistic regression allows to increase the accuracy of the new developed ACRP model.

Another point of interest is to enlarge the model to different types of bonds. At the moment, the model is limited in two ways. First, according to the works of [39] the bonds' features, like maturity and coupon rate, have to be similar to guarantee an accurate prediction. Second, the model is limited to plain-vanilla bonds. In the future, the model should be extended to be able to handle all kind of bonds, like callable bonds or floaters. The problem of different maturities can be faced with an additional preprocessing step. In this step, the bonds are divided into several groups according to their maturity. Thus, investors would obtain a more flexible ACRP model to undertake a risk analysis.

The development of the prototype has also uncovered some drawbacks. First, by fixing a set of bonds in advance, the model has lost its flexibility. To regain its flexibility, the training bonds should be detect automatically from the user-given bonds. In this way, the model has to determine the number of bonds and the right features which are similar to the user-given bonds. Afterwards, the needed information of the bonds are retrieved automatically from the Internet. After the building of the training set, the model has to identify the best value for the kernel factor σ. By determining the minimal error between the predicted rating degrees and the given rating degrees of the training bonds for different values of σ, the optimal σ is determined. The rating degrees of the bonds given by investors are predicted with the help of this optimal σ. In this way, investors obtain a flexible ACRP model to evaluate bonds. The model adapts always itself automatically to the updated attributes of the bonds and also uses the optimal kernel factor σ. By solving these two mentioned research questions, the prototype would completely fulfill all the requirements stated for the ideal ACRP model.

Finally, the possibility to evaluate shares instead of bonds should also be investigated in the future. If the new developed ACRP model would also be able to handle shares, then the model becomes complete. In this case, investors are not only limited to bonds but can also undertake a risk analysis of shares. With this link between the risk evaluation of bonds and shares, the new developed ACRP model offers private investors a real risk assessment tool. Due to the integration of the model into the personal financial planning tool LifeCharts, a real decision help for private investors can be provided. In a first step, LifeCharts evaluates the financial situation of investors and identifies the additional saving efforts which has to be induced. In a second step, LifeCharts allows investors to select a risk evaluation of bonds and shares. Finally, the optimal portfolio including bonds and shares is constructed by the model and shown to the investors. At the end of this future research, a real decision help for investments is obtained.

References

[1] Abdelhamid, D., Chaouki, B. M., and Abdelmalik, T.-A. (2011). A fast multi-class svm learning method for huge databases. *IJCSI International Journal of Computer Science Issues*, 8:544–550.

[2] Aisen, B. (2006). A comparison of multiclass svm methods.

[3] Angulo, C., Parra, X., and Catala, A. (2003). K - SVCR: A support vector machine for multi-class classification. *Neurocomputing*, 55:57–77.

[4] Antreich, K. J., Graeb, H. E., and Wieser, C. U. (1994). Circuit analysis and optimization driven by worst-case distances. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 13:57–71.

[5] Baker, H. K. and Mansi, S. A. (2002). Assessing credit rating agencies by bond issuers and institutional investors. *Journal of Business Finance and Accounting*, 29:1367–1398.

[6] Ben-Hur, A., Ong, C. S., Sonnenburg, S., Schoelkopf, B., and Raetsch, G. (2008). Support vector machines and kernels for computational biology. *PLoS Computional Biology*, 4:1–10.

[7] Bennell, J. A., Crabbe, D., Thomas, S., and ap Gwilym, O. (2006). Modelling sovereign credit ratings: Neural networks versus ordered probit. *Expert Systems with Applications*, 30:415–425.

[8] Black, F. and Cox, J. C. (1976). Valuing corporate securities: Some effects of bond indenture provisions. *The Journal of Finance*, 31:351–367.

[9] Black, F., Emanuel, D., and William, T. (1990). A one-factor model of interest rates and its application to treasury bond options. *Financial Analysts Journal*, 46:33–39.

[10] Bolton, P., Freixas, X., and Shapiro, J. (2009). The credit ratings game. *Economics Working Papers*, -:1–45.

[11] Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, -:144–152.

[12] Burges, C. J. (1998). A tutorial on support vector machine for pattern recognition. *Data Mining and Knowledge Discovery*, 2:121–167.

[13] Buscema, M. (1998). Back propagation neural networks. *Substance use and misuse*, 33:233–270.

[14] Cai, Y.-D., Ricardo, P.-W., Jen, C.-H., and Chou, K.-C. (2004). Application of svm to predict membrane protein types. *Journal of Theoretical Biology*, 226:373–376.

[15] Chekavska, I. (2015). Can support vector domain description combined with fuzzy clustering algorithm, implemented for sovereign credit rating, show better accuracy in comparison with artificial neural networks? Master's thesis, Saarland University.

[16] Chordia, T., Sarkar, A., and Subrahmanyam, A. (2005). An empirical analysis of stock and bond market liquidity. *Review of Financial Studies*, 18:85–129.

[17] Clifford W, S. J. and Warner, J. B. (1979). On financial contracting: An analysis of bond convenants. *Journal of Financial Economics*, 7:117–161.

[18] Courant, R. and Hilbert, D. (1989). *Methods of Mathematical Physics*. Wiley.

[19] Coval, J. D., Hirshleifer, D. A., and Shumway, T. (2005). Can individual investors beat the market? *HBS Finance Working Paper No. 04-025*, -:1–39.

[20] Cox, D. R. (1955). Some statistical methods connected with series of events. *Royal Statistical Society*, 17:129–164.

[21] Cristianini, N. and Shawe-Taylor, J. (2012). *An Introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge University Press.

[22] Datta, S., Iskandar-Datta, M., and Patel, A. (1997). The pricing of initial public offers of corporate straight debt. *The Journal of Finance*, 52:379–396.

[23] Deng, G., McCann, C. J., and ONeal, E. (2010). What does a mutual funds average credit quality tell investors? *Journal of Investing*, 19:58–65.

[24] Du, K. L. (2010). Clustering: A neural network approach. *Neural Networks*, 23:89–107.

[25] Eichelberger, R. K. and Sheng, V. S. (2013). Does one-against-all or one-against-one improve the performance of the multiclass cclassification? *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence*, -:1609–1610.

[26] Eick, C. F., Zeidat, N., and Zhao, Z. (2004). Supervised clustering - algorithms and benefits. *16th IEEE International Conference on Tools with Artificial Intelligence*, -:774–776.

[27] Eskin, E., Weston, J., Noble, W. S., and Leslie, C. S. (2002). Mismatch string kernels for svm protein classification. In *Advances in neural information processing systems*, pages 1417–1424.

[28] Everitt, B. S., Landau, S., Leese, M., and Stahl, D. (2001). *Cluster Analysis*. John Wiley & Sons.

[29] Finley, T. and Joachims, T. (2005). Supervised clustering with support vector machines. *Proceedings of the 22nd International Conference on Machine Learning*, -:217–224.

[30] Fletcher, R. (1987). *Practical Methods of Optimization*. John Wiley & Sons.

[31] Franc, V. and Hlavac, V. (2002). Multi-class support vector machine. *16th International Conference on Pattern Recognition*, 2:236–239.

[32] Gabriel, L. and Pinzolas, M. (2002). Neighborhood based levenberg-marquardt algorithm for neural network training. *IEEE Transactions on Neural Networks*, 13:1200–1203.

[33] Gangolf, C., Dochow, R., Schmidt, G., and Tamisier, T. (2014). Svdd: A proposal for automated credit rating prediction. In Kacem, I., Laroche, P., and Róka, Z., editors, *International Conference on Control, Decision and Information Technologies (CoDIT)*, IEEE Conference Procedings, pages 48–53.

[34] Gangolf, C., Dochow, R., Schmidt, G., and Tamisier, T. (2015). Automated credit rating prediction and the treatment of information. In Kremers, H. and Susini, A., editors, *RISK Information Management, Risk Models and Applications*, volume 7 of *Lecture Noten in Information Sciences*, pages 69–76. CODATA Germany.

[35] Gangolf, C., Dochow, R., Schmidt, G., and Tamisier, T. (2016). Automated credit rating prediction in a competitive framework. *RAIRO - Operations Research*. In Press.

[36] Gilchrist, A., Langford, N. K., and Nielsen, M. A. (2005). Distance measure to compare real and ideal quantum processes. *Physical Review A*, 71:1–15.

[37] Gordon, R. H. (1986). Taxation of investment and savings in a world economy. *The American Economic Review*, 76:1086–1102.

[38] Greene, W. H. (2012). *Econometric Analysis*. Pearson Education.

[39] Grice, J. S. and Dugan, M. T. (2001). The limitations of bankruptcy prediction models: Some cautions for the researchers. *Review of Quantitative Finance and Accounting*, 17:151–166.

[40] Guo, X., Zhu, Z., and Shi, J. (2012). A corporate credit rating model using support vector domain combined with fuzzy clustering algorithm. *Mathematical Problems in Engineering*, 2012:1–20.

[41] Hanna, S. D., Waller, W., and Finke, M. (2008). The concept of risk tolerance in personal financial planning. *Journal of Personal Finance*, 7:96–108.

[42] Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation*. Prentice Hall PTR.

[43] Hilscher, J. and Wilson, M. I. (2015). Credit ratings and credit risk: Is one measure enough? *AFA 2013 San Diego Meetings Paper*, -:1–58.

[44] Hiriart-Urruty, J.-B. and Lemarechal, C. (2004). *Fundamentals of Convex analysis*. Springer.

[45] Ho, T. S. and Singer, R. F. (1982). Bond indenture provisions and the risk of corporate debt. *Journal of Financial Economics*, 10:375–406.

[46] Hornik, K., Stinchcombe, M., and White, H. (1990). Universal approximation of an unknown mapping and its derivates using multilayer feedforward networks. *Neural Networks*, 3:551–560.

[47] Hsu, C. and Lin, C. (2001). A comparision of methods for multi-class support vector machines. Technical report, National Taiwan University.

[48] Huang, Z., Chen, H., Hsu, C.-J., Chen, W.-H., and Wu, S. (2004). Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision Support Systems*, 37:543–558.

[49] Hull, J. C. (2015). *Risk Management and Financial Institutions*. WILEY.

[50] Jr., K. B., Ciccotello, C. S., and Jr., H. D. S. (2002). Issues in comprehensive personal financial planning. *Financial Services Review*, 11:1–9.

[51] Khazai, S., Safari, A., Mojaradi, B., and Homayouni, S. (2012). Improving the svdd approach to hyperspectral image classification. *IEEE GEOSCIENCE AND REMOTE SENSING LETTERS*, 9:594–598.

[52] Kim, Y.-C. and Schulz, R. M. (1992). Is there a global market for convertible bbond? *The Journal of Business*, 65:75–91.

[53] Kousarrizi, M. R. N., Ghanbari, A. A., Teshnehlab, M., Aliyari, M., and Gharaviri, A. (2009). Feature extraction and classification of eeg signals using wavelet transform, svm and artificial neural networks for brain computer interfaces. *International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing*, -:352–355.

[54] Koutsoupia, E. and Papadimitriou, C. (2000). Beyond competitive analysis. *SIAM Journal on Computing*, 30:300–317.

[55] Kriesel, D. (2007). *A Brief Introduction to Neural Networks*. -.

[56] Lee, I., Lochhead, S., Ritter, J., and Zhao, Q. (1996). The costs of raising capital. *Journal of Financial Research*, 19:59–74.

[57] Lee, S. H. (1993). Relative importance of political instability and economic variables on perceived country creditworthiness. *Journal of International Business Studies*, -:801–812.

[58] Lee, Y.-C. (2007). Application of support vector machines to credit rating prediction. *Expert Systems with Applications*, 33:211–226.

[59] Lemmon, M. (1998). Consumer confidence and asset prices: Some empirical evidence. *The Review of Financial Studies*, 41:257–275.

[60] Levich, R. M. and Thomas, L. R. (1993). The merits of active currency risk management: Evidence from international bond portfolios. *Financial Analysts Journal*, 49:63–70.

[61] Li, X. and Ye, N. (2006). A supervised clustering and classification algorithm for mining data with mixed variables. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 36:396–406.

[62] Liu, J. and Xu, M. (2008). Kernelized fuzzy attribute c-means clustering algorithm. *Fuzzy Sets and Systems*, 159:2428–2445.

[63] Madhusudanan, A. (2009). Implementation of a neural network.

[64] Malitz, I. (1986). On financial contracting: The determinants of bond covenants. *Financial Management*, 15:18–25.

[65] Markowitz, H. (1952). Portfolio selection. *The journal of finance*, 7:77–91.

[66] McCormick, G. (1983). *Non-Linear Programming: Theory, Algorithms and Applications*. John Wiley & Sons.

[67] Mukherjee, S., Banerjee, M., Chen, Z., and Gangopadhyay, A. (2008). A privacy preserving technique for distance-based classification with worst case privacy guarantees. *Data & Knowledge Engineering*, 66:264–288.

[68] Navneet, A., Bohn, J. R., and Zhu, F. (2005). Reduced form vs. structural models of credit risk: A case study of three models. *Journal of Investment Management*, 3:1–39.

[69] Petitt, B. S., Pinto, J. E., and Pirie, W. L. (2015). *Fixed Income Analysis*. John Wiley & Sons.

[70] Principe, J. C., Euliano, N. R., and Lefebvre, W. C. (2000). *Neural and Adaptive Systems: Fundamentals Through Simulations*. John Wiley & Sons.

[71] Rates, G. (2016). Euribor and Libor.

[72] Rudolph, S., Savikhin, A., and Ebert, D. S. (2009). Finvis: Applied visual analytics for personal financial planning. *IEEE Symposium on Visual Analytics Science and Technology*, -:195–202.

[73] Ruiz-Gonzales, R., Gomez-Gil, J., Gomez-Gil, F. J., and Martinez-Martinez, V. (2014). An svm-based classifier for estimating the state of various rotating components in agro-industrial machinery with a vibration signal acquired from a single point on the machine chassis.

[74] Rumelhart, D., Hinton, G., and Williams, R. (1986). Learning representations by backpropagating errors. *Nature*, 323:533–536.

[75] Rumelhart, D. E. and McClelland, J. L. (1986). *Parallel distributed processing: explorations in the microstructure of cognition*. MIT Press Cambridge.

[76] Savikhin, A., Lam, H. C., Fisher, B., and Ebert, D. S. (2011). An experimental study of financial portfolio selection with visual analytics for decision support. *IEEE 44th Hawaii International Conference on System Sciences*, -:1–10.

[77] Schaefer, S. M. and Strebulaev, I. A. (2008). Structural models of credit risk are useful: Evidence from hedge ratios on corporate bonds. *Journal of Financial Economics*, 90:1–19.

[78] Schmidt, G. (2011). *Persoenliche Finanzplanung*. Springer-Verlag Berlin Heidelberg.

[79] Securities and Institute, F. (2016). Market data.

[80] Securities, U. S. and Commision, E. (2015). Definition of institutional investors.

[81] Service, M. I. (2013). Moody's bond fund rating methodology.

[82] Smith, D. J. (2011). *Bond Math: The Theory behind the formulas*. John Wiley and Sons.

[83] Standard and Poors (2012). Defintion of credit ratings.

[84] Szymon, B., Hardle, W. K., and Lopez-Cabrera, B. (2013). Black-scholes option pricing model. *Statistics of Financial Markets*, -:59–78.

[85] Tax, D. M. and Duin, R. P. (1999). Support vector domain description. *Pattern Recognition Letters*, 20:1191–1199.

[86] Tax, D. M. and Duin, R. P. (2004). Support vector data description. *Machine Learning*, 54:45–66.

[87] Tonks, D. G. (2009). Validity and the design of market segments. *Journal of Marketing Management*, 25:341–356.

[88] Treacy, W. F. and Carey, M. (2000). Credit risk rating systems at large us banks. *Journal of Banking & Finance*, 24:167–201.

[89] Trevino, L. and Thomas, S. (2001). Local versus foreign currency ratings: What determines soverign transfer risk? *Journal of Fixed Income*, 11:65–76.

[90] Truglia, V. J. (1999). Moody's sovereign ratings: A rating guide. Moody's Investors Service Global Credit Research, special comment.

[91] Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology*, 49:1225–1231.

[92] Vapnik, V. (1995). *The nature of Statistical Learning Theory*. Spri.

[93] Veraverbeke, N. (2006). Hazard rate estimation. *Encyclopedia of Statistical Sciences*, 5:–.

[94] Wang, Z. and Xue, X. (2014). Multi-class support vector machine. *Support Vector Machines Applications*, -:23–48.

[95] Werbos, P. J. (1974). *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. PhD thesis, Havard University.

[96] Weston, J. and Watkins, C. (1998). Multi-class support vector machines. Technical report, Department of Computer Science, Royal Holloway, University of London.

[97] White, L. J. (2010). Markets: The credit rating agencies. *The Journal of Economic Perspectives*, 24:211–226.

[98] Ye, N. and Li, X. (2001). A machine learning algorithm based on supervised clustering and classification. *Lecture Notes in Computer Science*, 2252:327–334.

[99] Zhang, J.-R., Zhang, J., Lok, T.-M., and Lyu, M. R. (2007). A hybrid particle swarm optimization - back propagation algorithm for feedforward neural network training. *Special Issue on intelligent Computing Theory and Methodology*, 185:1026–1037.

[100] Zhao, Y. and Wang, Z. (2010). Support vector data description for finding non-coding rna gene. *Journal of Biomedical Engineering*, 27:779–784.

[101] Ziebart, D. A. and Reiter, S. A. (2010). Bond ratings, bond yields and financial information. *Contemporary Accounting Research*, 9:252–282.

Appendix A

Illustration of the different worst-case scenarios

Scenario 1

In the following figure, the two sub-cases and the global worst-case are presented with the help of Moody's rating scale. The defined encoding rule, 6.1 is used to map the credit ratings into the rating degrees with δ equal to 0.25.

Ratings	Aaa	Aa1	Aa2	Aa3	A1	A2	A3	Baa1	Baa2	Baa3	Ba1	Ba2	Ba3	B1	B2	B3	Caa1	Caa2	Caa3	Ca	C
Rating degrees	1	1.25	1.5	1.75	2	2.25	2.5	2.75	3	3.25	3.5	3.75	4	4.25	4.5	4.75	5	5.25	5.5	5.75	6

Investment grade — Non-investment grade

Case (i):
OPT => Aaa
ALG1 => Ba1

Case (ii):
OPT => C
ALG1 => Baa3

Worst Case:
OPT => Aaa
ALG1 => Ba1
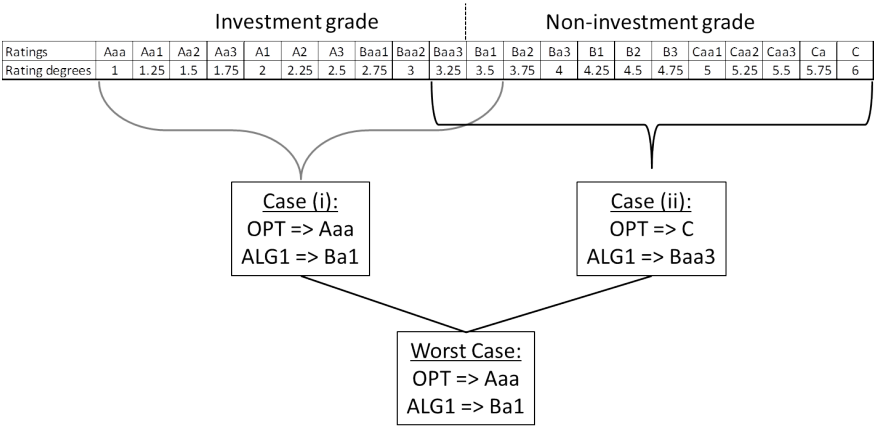
Fig. A.1 Illustration of the worst-case under the assumption of Scenario 1

Scenario 2

The two sub-cases and the global worst-case are represented with the help of Moody's rating scale in the subsequent figure. As in the previous representation, δ is fixed to 0.25.
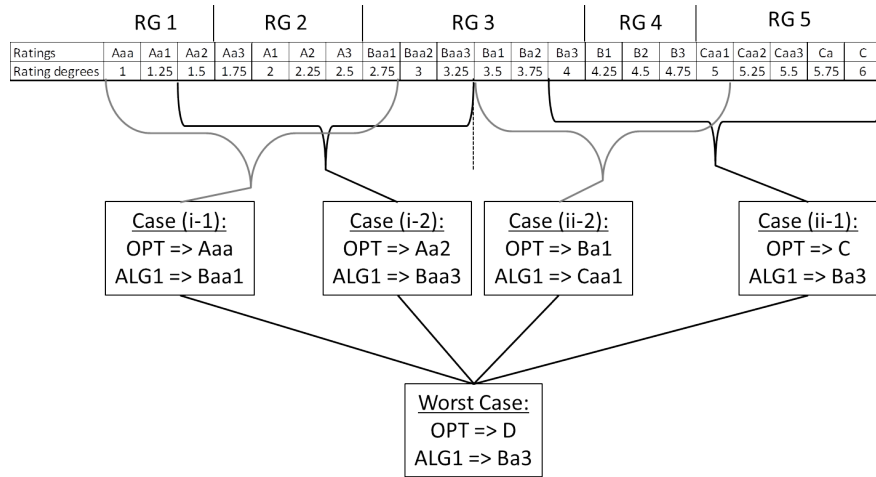
Fig. A.2 Illustration of the worst-case under the assumption of Scenario 2

Scenario 3

In the following figure, the two sub-cases and the global worst-case are presented with the help of Moody's rating scale.
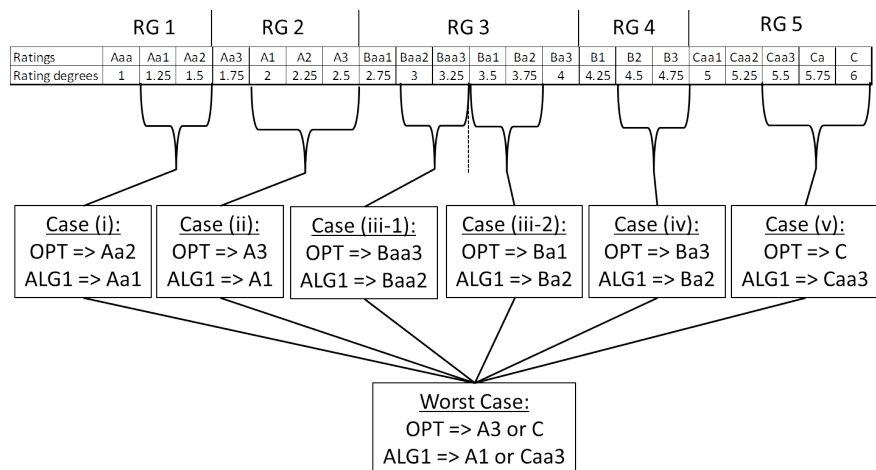
Fig. A.3 Illustration of the worst-case under the assumptions of Scenario 3

Appendix B

Integration of the ACRP model into LifeCharts

B.1   Exemplary scenario in Chapter 2
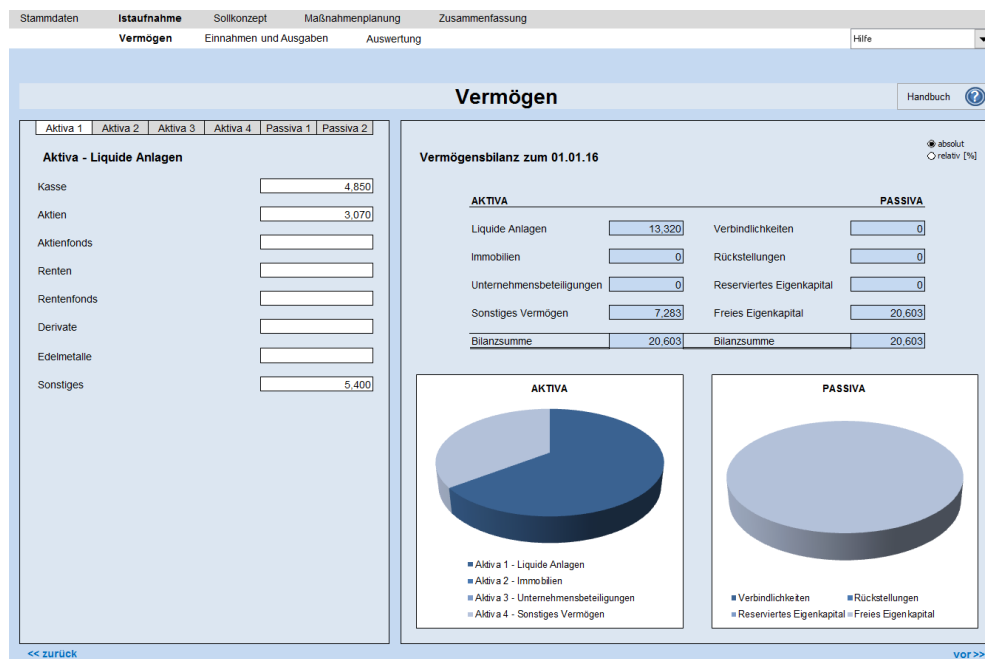
Fig. B.1 Indication of the liquid assets of the user, like cash or shares
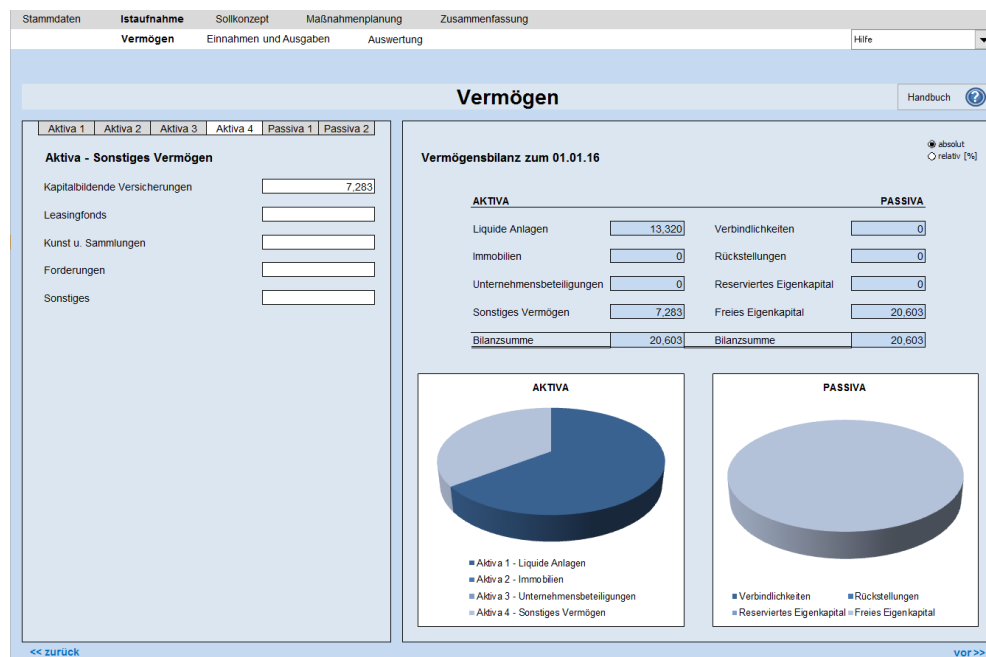
Fig. B.2 Indication of other assets of the user, like endowment insurances
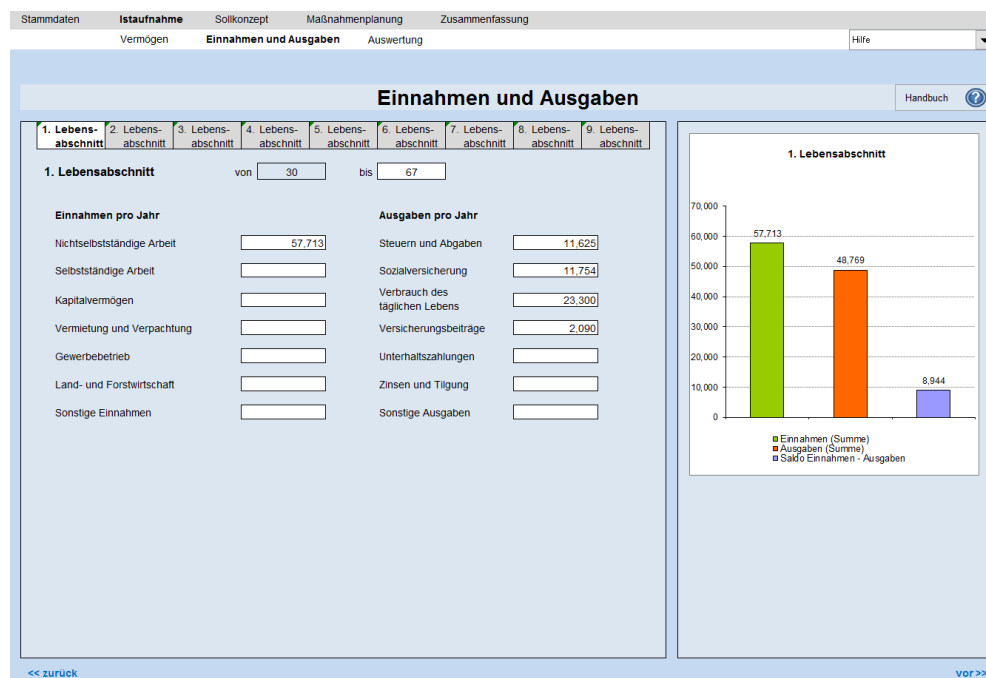
Fig. B.3 Outcomes and incomes of the user for the first life period
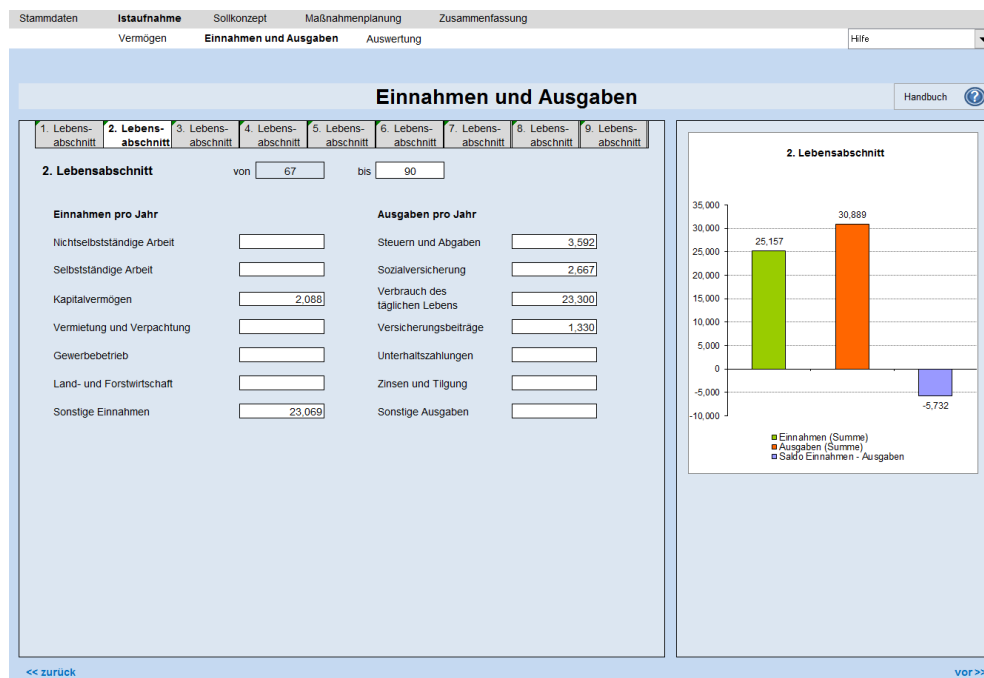
Fig. B.4 Outcomes and incomes of the user for the second life period
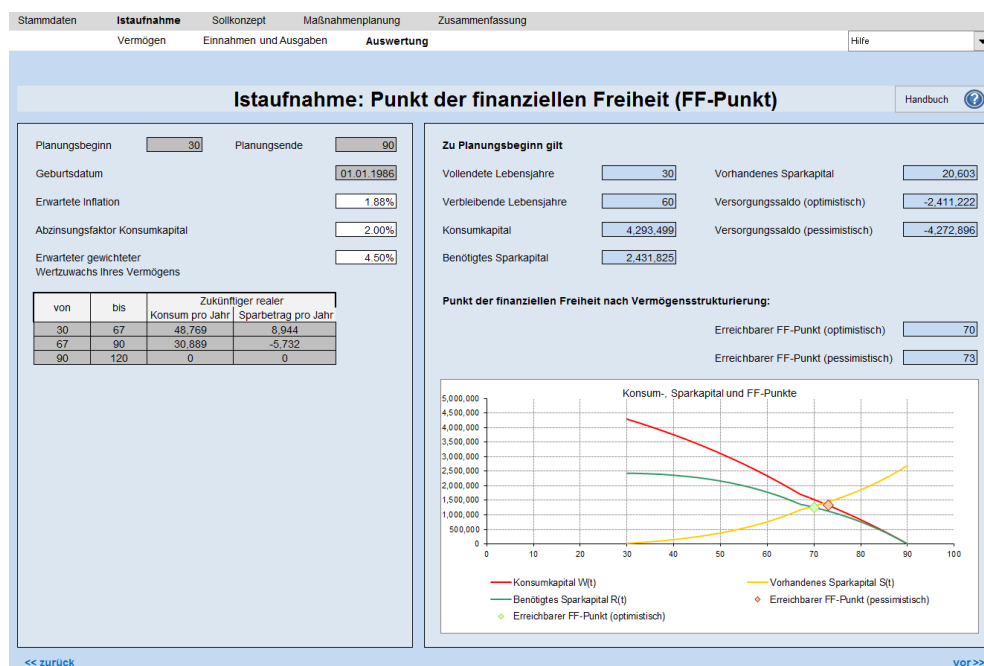
Fig. B.5 Overview of the financial situation of the user and illustration of the determined FF-point

Fig. B.6 User information about the restructuring of the wealth

Fig. B.7 The new determined FF-point under the consideration of the restructuring of the user's wealth
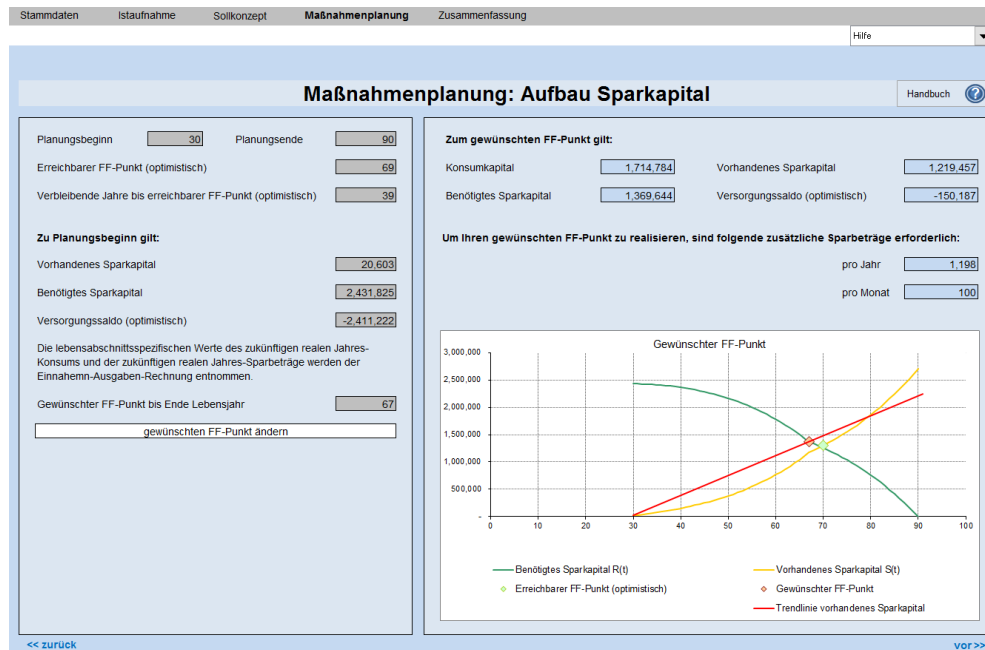
Fig. B.8 The proposed measures to achieve the desired FF-point

B.2 Exemplary scenario in Chapter 8
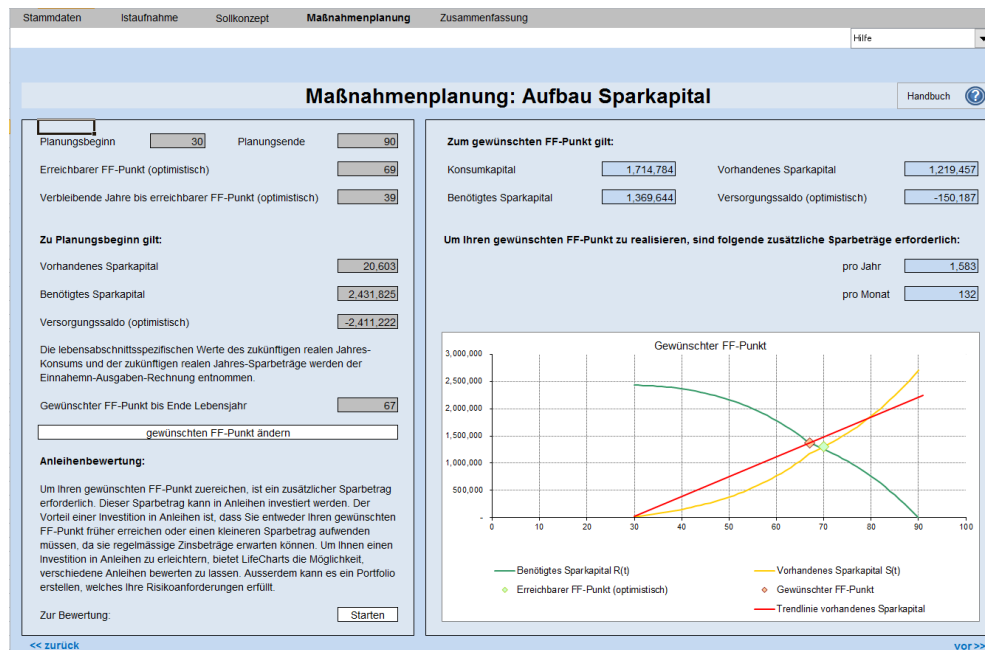
Fig. B.9 At this point, investors can choose to undertake a bond risk analysis
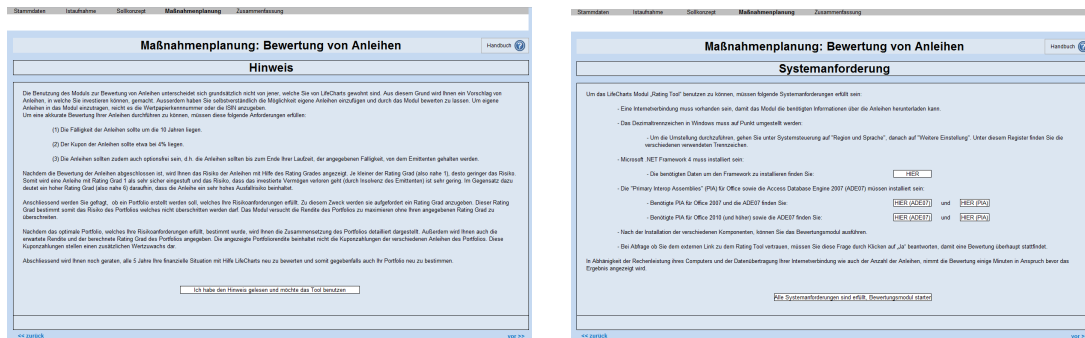
(a) The shown detailed description   (b) The shown system requirement

Fig. B.10 Description and system requirement of the integrated ACRP model

Fig. B.11 Main page of the integrated ACRP model

Fig. B.12 Representation of the predicted rating degrees of the bonds in the integrated ACRP model
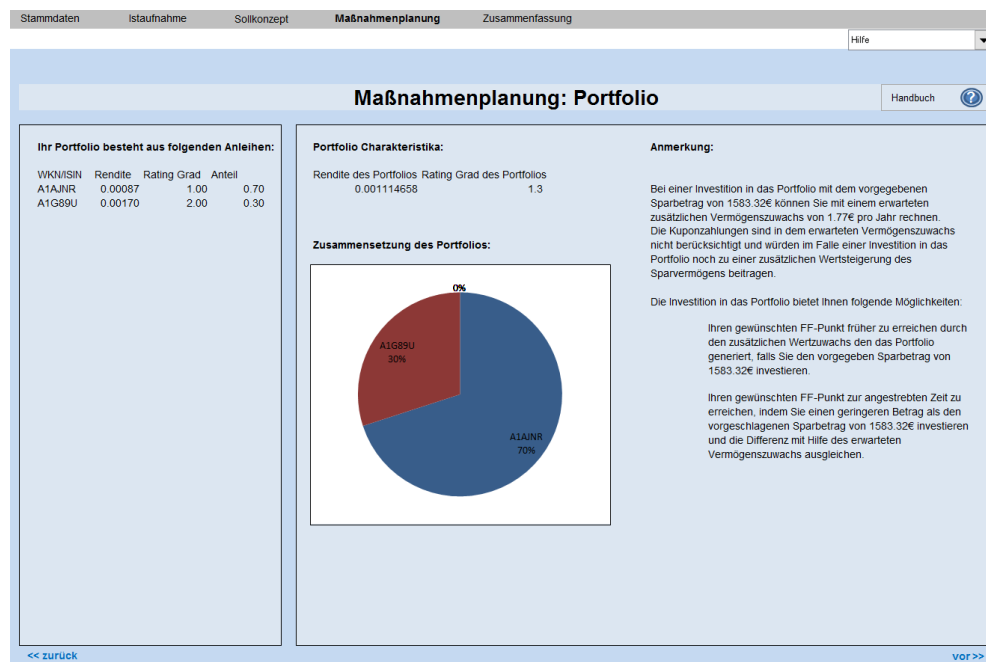

Fig. B.13 The given proportions for each bond in the portfolio

Fig. B.14 Detailed view on the obtained portfolio