BOND RATING CLASSIFICATION

A Probabilistic Fuzzy Approach

Master Thesis Informatics & Economics
Faculty of Economics
Erasmus University Rotterdam

Mark-Alexander Sprengers
E-mail: masprengers@hotmail.com

Supervised by:
Dr.ir. Jan van den Berg
and
Ludo Waltman MSc.

16th February 2006

Abstract

This report is devoted to Bond Rating Classification (BRC). BRC has become increasingly important over the last few decades. Firms, governments, and individuals often make use of debt financing. Generally, the acquired debt is used by these entities to be able to 'grow' in an earlier stage than in the situation where they have to wait for retained earnings. Having disposal of the extra capital in an early stage is seen as an advantage for the entities lending the money, including those issuing bonds.

Nevertheless, a disadvantage sticks together with this advantage, namely an increase in the level of risk, since the event of lending money brings obligations with itself. Therefore, it is of great importance to the investors (like the bond holders) to be informed about the capability of the issuer to keep to these obligations.

In this report, a relatively new technique termed 'probabilistic fuzzy systems' (PFSs) is applied to induce an accurate classification model for Bond Rating Classification. In an out-of-sample test, the classifications (i.e., the ratings) as made by this model are compared to the actual ratings as assigned by Standard & Poor's and Moody's. In addition, the performance of the PFSs' model is compared to the performance of the model induced by means of multiple discriminant analysis (MDA), the most commonly used classification model for BRC.

The motivation for trying PFSs for solving the problem of BRC is two-fold. First, PFSs have the important advantage that they are better interpretable than models constructed with MDA. Second, because of their non-linear properties, it is hoped and expected that PFSs yield models with a better performance.

Both the MDA model and PFS proved to be good classifiers for our BRC problem, and showed similar results in a leave-one-out cross-validation. With regards to the interpretability of the constructed PFS model, it has to be mentioned that our PFS did not bring forward an easy interpretable model, although some relations could be derived from the rules made by our PFS.

Keywords: Bond Rating Classification, Probabilistic Fuzzy Systems, Multiple Discriminant Analysis, Financial Ratios, Artificial Intelligence

Acknowledgements

I hereby want to thank Ludo Waltman MSc. for all his help during the writing process of this master thesis. Without this kind gesture, probably I would not be as satisfied with my thesis as I am now.

I also want to thank my supervisor Dr.ir. Jan van den Berg for his assistance and the helpful suggestions he made regarding my thesis.

Contents

1 Introduction **7**
- 1.1 Motivation . . . 7
- 1.2 Goal . . . 8
- 1.3 Methodology . . . 8
- 1.4 Structure . . . 9

2 The bond rating process **10**
- 2.1 What is a bond ? . . . 10
  - 2.1.1 Background to a bond . . . 10
  - 2.1.2 Motivations for a bond issue . . . 11
  - 2.1.3 Different types of bonds . . . 12
  - 2.1.4 Characteristics of a bond . . . 13
  - 2.1.5 Valuation of a bond . . . 13
- 2.2 Bond rating . . . 14
  - 2.2.1 History of bond rating . . . 14
  - 2.2.2 Importance of bond ratings . . . 15
  - 2.2.3 The rating process . . . 17
- 2.3 Literature review on bond rating models . . . 19
  - 2.3.1 Statistical bond rating models . . . 20
  - 2.3.2 Artificial intelligence bond rating models . . . 22
  - 2.3.3 Conclusions . . . 23

3 Methodology **25**
- 3.1 Multiple discriminant analysis . . . 25
  - 3.1.1 Technical background: Multiple discriminant analysis . . . 26
  - 3.1.2 Multiple discriminant analysis application for bond rating prediction . . . 30
- 3.2 Probabilistic Fuzzy Systems . . . 31
  - 3.2.1 Technical background: Probabilistic Fuzzy Systems . . . 31
  - 3.2.2 Strengths and weaknesses of Probabilistic Fuzzy Systems . . . 39
  - 3.2.3 Probabilistic Fuzzy Systems application for Bond Rating . . . 40

4 Experimental Setup **41**
- 4.1 Financial Setup . . . 41
  - 4.1.1 Data set . . . 41
  - 4.1.2 Financial Statement Analysis . . . 42

4.1.3 Financial variables 43
4.1.4 Variable selection: classification and regression trees 45
4.2 Architecture of the models 46
4.2.1 Multiple discriminant analysis 46
4.2.2 Probabilistic fuzzy systems 47

5 Results and analysis **48**
5.1 Setup of the models 48
5.2 Multiple discriminant analysis 49
5.3 Probabilistic fuzzy systems 51
5.4 Discussion 55

6 Conclusions and Future Research **57**
6.1 Conclusions 57
6.2 Future Research 58

A Appendix A: Figures, charts and tables **64**
A.1 The rating process 64
A.2 General fuzzy inference system model 65

B Appendix B: Data set and accompanying histograms **66**
B.1 List of insurance companies 66
B.2 Descriptive analysis 70
B.3 Histograms 71

C Appendix C: Balance sheet and profits & losses accounts **73**

D Appendix D: Training & testing of a data set **75**
D.1 Cross validation methods 75

E Appendix E: Classification and regression trees **77**

F Appendix F: Probabilistic fuzzy systems **81**
F.1 Membership functions 81
F.2 Probabilistic fuzzy rules 83

List of Figures

2.1 Bar diagram showing the one-year default rates by alpha-numerical ratings, 1983-1999 16

3.1 Scatter plot showing two groups. 26
3.2 Visualization of two possible linear combinations of two independent variables. 26
3.3 Scatter plot of two groups using x or y to discriminate between both groups. 27
3.4 The best linear combination of x and y to discriminate between both groups. 27
3.5 Triangular MF 32
3.6 Trapezoidal MF 32
3.7 Gaussian MF 32
3.8 Generalized Bell MF 32
3.9 Block diagram for a FIS 34
3.10 Three-rule Mamdani FIS deriving an overall output from two crisp inputs service and food 35

4.1 Bond rating histogram of the insurance companies data set 42

5.1 Membership functions of variable X_1 for rules $1-6$ 54
5.2 Misclassification histogram for the PFS model made by set 3 and the MDA model made by set 2. 56

A.1 Flow chart of an on request rating process 64
A.2 Three-rule Takagi-Sugeno FIS deriving an overall output from two crisp inputs service and food 65

B.1 Extensive bond rating histogram of the insurance companies data set . 71
B.2 Histogram with the countries of origin of the companies in the data set 72

E.1 Unpruned tree built by the CART algorithm of the insurance companies data set 78
E.2 Pruned tree built by the CART algorithm of the insurance companies data set 79
E.3 Pruned tree built by the CART algorithm of the insurance companies data set 80

F.1 Membership functions of variable X_2 for rules $1-6$ 81

F.2 Membership functions of variable X_3 for rules $1-6$ 82
F.3 Membership functions of variable X_8 for rules $1-6$ 82
F.4 Membership functions of variable X_{14} for rules $1-6$ 83
F.5 Representation of rule 1 84
F.6 Probabilities belonging to rule 1 84
F.7 Representation of rule 2 85
F.8 Probabilities belonging to rule 2 85
F.9 Representation of rule 3 86
F.10 Probabilities belonging to rule 3 86
F.11 Representation of rule 4 87
F.12 Probabilities belonging to rule 4 87
F.13 Representation of rule 5 88
F.14 Probabilities belonging to rule 5 88
F.15 Representation of rule 6 89
F.16 Probabilities belonging to rule 6 89

List of Tables

2.1 Rating scale for the three largest rating agencies. 15

3.1 The accuracy matrix used to determine the accuracy of a discriminant analysis model. 28
3.2 The accuracy matrix based on the relative frequency with which each group appears in the data. 28
3.3 The MDA accuracy matrix based on the relative frequency with which each group appears in the data. 30
3.4 Nine independent variables for the MDA model introduced by Belkaoui (1983). 30

4.1 Total selection of financial variables and financial ratios examined in order to select the used financial ratios 44
4.2 Fourteen selected financial input variables and ratios 45

5.1 Independent variables selected for the three sets of data. 49
5.2 Test results for set 1, 2 and 3 based on leave-one-out cross-validation. 50
5.3 Accuracy matrix of the MDA model based on 14 independent variables (Set 1) 50
5.4 Accuracy matrix of the MDA model based on 7 independent variables (Set 2) 51
5.5 Accuracy matrix of the MDA model based on 5 independent variables (Set 3) 51
5.6 Results of the PFS models from a ten-fold cross-validation 52
5.7 Results of the PFS models from a leave-one-out cross-validation . . . 52
5.8 Accuracy matrix of the PFS model with 6 rules, based on 5 independent variables (Set 3) 53
5.9 Results of the MDA and PFS models from a leave-one-out cross-validation 55

B.1 Complete list of insurance companies and their ratings 66
B.2 Summary statistics per variable. (Year 2004) 70
B.3 Accuracy matrix based on the relative frequency of each rating class within the insurance data set 71

C.1 Standardized balance sheet. 74
C.2 Standardized profits & losses account. 74

Chapter 1

Introduction

This Chapter gives an introduction to this report and explains the motivation for choosing "Bond rating classification, a probabilistic fuzzy approach" as the subject of this thesis. In section 1.1, the motivation and origin for the topic of this report is discussed. After formulating the motivation, the goal of this report will be described in section 1.2. In section 1.3, the methodologies used in this report will be described shortly. Finally, section 1.4 gives an overview of this report, where every chapter will be discussed briefly.

1.1 Motivation

Evaluation of the financial health condition of companies is a frequently conducted activity since the beginning of the twentieth century. This activity is obviously linked with the depression in 1930, when corporate bankruptcies reached numbers never seen before. Corporate bankruptcy is an event which has a big impact on management, shareholders, employees, creditors, customers and other stakeholders. It will cause financial losses to most of the aforementioned parties. These events also have a negative influence on a nation, both socially and economically (Altman (1968)). Above all, more and more companies decide to make use of debt financing. In a general sense these companies use the acquired debt to 'grow' in an earlier stage than compared to when they would have to wait for their retained earnings. A company can issue bonds to acquire debt, which actually are nothing more than debt securities which can be purchased by individuals. A bond issue leads to a mutation in the debt-to-equity ratio of a company, which will have consequences for the financial health condition of the issuing company. Bondholders are highly interested in the degree of risk belonging to their particular bonds, mainly because of the aforementioned changing financial health conditions. To meet this interest, agencies like Standard and Poor's, Moody's, and Fitch started rating bonds since the beginning of the twentieth century (More a less simultaneously with the demand for bond ratings out of the market). Their ratings form support on the investment quality of debt securities.

Although these bond ratings given by the three rating agencies are well accepted and understood, nobody knows how these ratings come about. It is known that the agencies make use of financial ratios and managerial quality measurements, but which ones and by which proportion is not transparent. Despite lacking transparency, these ratings are of high importance to individuals, companies and eventually for every nation.

In the past few decades extensive research has been done on bonds and especially on bond rating prediction. Several bond rating prediction models have been introduced, which have been useful to the issuing companies, the investors and even to the rating companies. Many statistical and artificial intelligence (AI) models have been examined, even though multiple discriminant analysis (MDA) has been used and described most frequently. In general, more and more AI models outperform these MDA models, as they are newer and more sophisticated. Nevertheless MDA remains a popular technique for bond rating prediction, which for the greater part is debt to the amount of research done on this subject and to the simplicity in use of this technique.

1.2 Goal

The objective of this report is: "Constructing an accurate bond rating classification (BRC) model based on a probabilistic fuzzy system (PFS)."

For many years MDA models are used as standard tools for BRC, despite the fact that many techniques outperform it when predicting corporate bond ratings (Huang et al. (2004)). This is mainly due to two facts. Firstly, because MDA models are easy to apply for BRC. Secondly, extensive research conducted on the application of MDA models for BRC makes these models comprehensible and commonly used.

This report compares the popular MDA technique with an application of PFS for BRC (Belkaoui (1983)). PFS is an AI technique, described extensively in Chapter 3, and uses supervised learning.

The main goal for this report is to construct a highly accurate BRC model based on a PFS. The secondary goal is to construct a BRC model that is better interpretable in comparison with models described in studies previously conducted. This model could give better insight into the variables which are of importance in determining the variable to be explained. The knowledge that is acquired from this report will hopefully help to give directions for future research. For example, it may help individuals or companies valuing their bonds. An extensive literature review on BRC models and an application of a PFS for BRC will provide an accurate and an interpretable BRC model.

1.3 Methodology

For this report a PFS is constructed to perform a classification of bond ratings. The building and testing of the model consists of four stages. The first stage can be viewed as the preparation phase, in which the data set and the PFSs will be made up for the BRC problem. During the second stage, the PFS will be trained by means of the training data. Testing of the PFS is done in the third stage, therefore the constructed PFS will be feeded with test data and the predicted bond rating classes will be compared with the actual classes. In the last stage the results are examined and tested on significance.

This report will also describe other techniques to predict bond ratings comprehensively. From these techniques, the MDA model is the most popular one and the most extensively described in the articles written on the bond rating prediction (See among them Huang et al. (2004)). That is why this technique will act as a benchmark technique for PFS examined in this report.

The data set used in this report consists of 161 insurance companies from the United States of America, Europe and some from South-Africa. All of these companies have a rating of at least 'B' by Standard and Poor's rating. All the financial information on these companies is public, and downloaded from Thomson One Banker, a web-site which provides company information[1].

1.4 Structure

In addition to this introductory chapter, this report contains five chapters. In Chapter 2, the financial background for this report is sketched. A thorough understanding of the reasons and stimulating circumstances under which corporate distress occurs will be created. Section 2.2 gives a literature review about the history of BRC models. This chapter will be concluded with some limitations and suggested improvements of the discussed techniques.

Chapter 3 is the methodology chapter and is divided into two sections. The first section describes the MDA model. The second section is the examination of the PFSs. This section starts with the technical explanation on fuzzy sets, fuzzy systems and PFSs, and proceeds with the application of a PFS for our BRC problem.

Chapter 4 describes the experimental setup. This chapter starts by introducing the data set of the insurance companies that will be used. Details about this data set will be given. Besides that, this chapter will give a short commentary to the composition of the balance sheet. The financial ratios that are important for this research, will be clarified and the framework of the research will be mapped.

The results of both the MDA model and the PFS will be described in Chapter 5. Positive and negative characteristics concerning these techniques solving this problem will be discussed. Also this chapter will show a table with the accuracy of both models with respect to the BRC problem.

Chapter 6 recapitulates findings of the report and attaches a conclusion to it. This chapter will also suggest new topics for future research.

In fact this report can be subdivided into three parts, in where the five chapters are placed. The first part, consisting of Chapter 2 and 3, gives some financial background to corporate bond rating and introduces the models that will be used in this report. The second part, containing Chapter 4, describes the experimental setup. Results, conclusions and suggestions for future research will be discussed in the third and last part. (Chapter 5 and 6)

[1] http://banker.analytics.thomsonib.com/

Chapter 2

The bond rating process

The main objective of this chapter is to make the bond rating process comprehensible. To reach this objective, first the basic principles behind bonds need to be exemplified. Section 2.1 starts with providing some background information on bonds. After having determined the definition of a bond, section 2.2 will carry on with the subject. This section initially reviews the history of bond ratings, which on its turn is followed by subsection 2.2.2, where the importance of bond ratings is expounded. Subsection 2.2.3 describes all possible steps to be taken in the rating process. The last section, section 2.3, concludes this chapter with a literature review on bond rating models. This section is important, given the fact that our to develop model will be partly based on the findings of previous bond rating models.

2.1 What is a bond ?

2.1.1 Background to a bond

Currently an increasing number of individuals, firms and governments make use of the capital market to secure economical growth. To accomplish this economical growth many of the aforementioned groups prefer financial leverage through debt financing. In this way these groups are able to use capital assets in an earlier stage than in the situation where they have to wait for retained earnings. The capital used for debt financing originates from savings, and function as loans in contrast to savings which are put aside without exploiting it. When a firm decides to use debt financing to establish economical growth, it actually raises money for working capital or capital expenditure by selling bonds, bills, or notes to institutional investors and/or individuals.

Another way of raising capital is called equity financing. This way of raising capital is established by issuing shares of stock in a public offering to individuals and/or institutional investors. The owners of these new shares receive ownership interests in the organization who issued these shares. The scope of this report is amplified to debt financing, with in particular to corporate bonds.

Governments or firms who decide to increase their debt-to-equity ratio can issue bonds. Before describing the characteristics of a bond and mentioning the motivation of entities to issue bonds, it is helpful to define the exact meaning of a bond.

Basically a bond is a loan from one entity to another entity, so there are two parties of interest when a bond is issued. The party who issues a bond is called the issuer or obligor and receives the loan from the bondholder. The bond market is known as an over-the-counter market, in contrast to the stock market which makes use of exchanges. Bonds are divided into smaller pieces to make them more tradable. The issuer of a bond obliges itself to periodically pay an amount of interest to the bondholder and to redeem the principal, or face value, at the maturity date. For this reason bonds belong to the fixed income securities group. The following two definitions of a bond were found on the internet. The first definition is given by the financial dictionary Investopedia (2005).

"A debt investment with which the investor loans money to an entity (company or government) that borrows the funds for a defined period of time at a specified interest rate."

The second definition of a bond was found in the financial glossary of MainstayInvestments (2005) and is:

"A debt security issued by a company, municipality or government agency. A bond investor lends money to the issuer and, in exchange, the issuer promises to repay the loan amount on a specified maturity date. In addition, the issuer usually provides the bondholder periodic interest payments."

2.1.2 Motivations for a bond issue

A considerable number of factors can motivate companies or governments to issue bonds. The main motive is to raise capital to secure economical growth in an earlier stage. Belkaoui (1983) quoted a few other reasons next to this main motive, among which the politics of corporate control, maintaining an adequate debt-to-equity ratio, expectations about the term structure of interest, and its capacity to absorb debt. This capacity to absorb debt is an important factor for the issuer and especially for the bondholder. This capacity is based on the ability to repay the fixed, periodically, interest amount and the principal at the maturity date. The higher the risk belonging to a bond the more suspicious a potential bondholder will be. The amount of fixed interest will always be determined by the degree of risk. A bond indenture[1], which in fact is a legal contract between issuing firms or governments, bondholders and the trustee representing the bondholders, counteracts situations were interest payments will be missed or situations where redeem of the principal will end up in a misery.

[1] Definition of an indenture: A contract between an issuer of bonds and the bondholder stating the time period before repayment, amount of interest paid, if the bond is convertible (and if so, at what price or what ratio), if the bond is callable and the amount of money that is to be repaid. (Source of information: Investopedia (2005))

2.1.3 Different types of bonds

There are all kinds of bonds, the variations can be found in the term-to-maturity[2], type of issuing organization, type of repayment, mode of interest payment, and the nature of the claim. In fact every bond is different, but in order to look at bonds in a general sense this paragraph will only look at the difference between the type of issuers.

According to Fabozzi and Mann (2005) the three largest issuers of debt are domestic corporations, municipal governments, and the federal governments and its agencies. In the scope of this report we mainly look at corporate bonds, although the general idea behind the municipal and government bonds will be described subsequently.

Municipal bonds are debt securities issued by states, cities, districts or counties in order to finance their capital expenditures. The majority of the investors of municipal bonds consist of individual investors. Most of these bonds are considered as relatively low risk bonds. Municipal bonds exists in two forms, namely "General obligation" (GO) bonds and "revenue bonds". Projects financed by GO bonds do not produce revenue and have maturities of 10 years and more. According to Fabozzi and Mann (2005) they are backed by full faith, credit, and taxing power of the governmental unit issuing them. Project financed by revenue bonds do generate revenues, in contrast to GO bonds. Examples of projects financed by revenue bonds can be commercial stadiums which generate money by earning on entrance fees.

Government bonds have the smallest risk among all types of bonds. They offer a fixed interest rate and have term-to-maturities comparable to corporate bonds. Government bonds are marketable securities or non-marketable securities. Marketable government bonds can be traded on the exchanges or on the over-the-counter markets. Non-marketable government bonds are exposed to regulations initiated by the issuer, governments, which limits the bondholder. These bonds can only be redeemed back to the government.

Corporate bonds are legal agreements between corporations and individuals or institutional investors. These bonds are secured by the assets and/or credits of the issuing companies. Money raised by issuing companies is frequently used for capital expenditures, in other cases the money is used for one of the reasons discussed in the previous paragraph. Risks connected to these corporate bonds can vary enormously. The degree of risk is totally dependent on chance of default of the bond. This default risk of a bond is the risk that the issuer is unable to pay the interest payments or principal on his debt obligation. An issuer which falls into bankruptcy will automatically default on a bond, although this does not mean that a company which defaulted on a bond has gone bankrupt. Section 2.2 discusses the problem of determining the level of risk to which a certain bond is allocated. When an issuer goes bankrupt, bondholders can make a claim on the assets of that company. The same procedure is applicable to equity holders of the firm which has gone bankrupt, although bondholders get priority over equity holders.

[2]The number of years during which the borrower has promised to meet the conditions of the debt(which are contained in the bond's indenture). Source of information: Fabozzi and Mann (2005)

2.1.4 Characteristics of a bond

Every bond has its own characteristics, although some of these characteristics are basically the same for each bond. The subsequent list comprises of the most basal characteristics of a bond:

Two entities Each bond is a legal agreement between two entities. On the one hand the issuer and on the other hand the bondholder. Both entities are obliged to some regulations which are determined in advance.

Term-to-maturity One of these regulations is the term-to-maturity. Before a bond is issued the termination date of the bond is prescribed. When the term-to-maturity is longer than one year the bond is considered as a long-term debt issue, all others are considered as short-terms.

Coupon The definition for coupon, given by Fabozzi and Mann (2005), is the periodic payment made to the owner during the life of a bond. Except for zero-coupon bonds, every bond receives interest payments. Interest on these coupon bonds is usually paid semiannually. Most often the interest rate of coupon bonds is fixed, sometimes it is floating. The level of interest is largely dependent on the extent of risk to which a bond is exposed to.

Principal The principal of a bond is the original amount invested by the bondholder. Another term used for the principal is the face value, which is amount paid to the bondholder at maturity.

Claim on assets Bond holders have a claim on the assets of an issuer, if situations occur in which an issuer is unable to meet his obligations. At the same time equity holders have this right, nevertheless bondholders have priority in this.

2.1.5 Valuation of a bond

Holding a corporate bond is according to Brealey and Myers (2003) equivalent to lending money with no chance of default, but at the same time giving stockholders a put option on the firm's assets. Furthermore they mentioned that owning a corporate bond is also equivalent to owning the firm's assets, but giving a call option on these assets to the firm's stockholders.

Generally speaking the value of a bond is determined by the combination of the coupon percentage and the current market interest rate. The coupon interest rate is fixed and based on the chance of default of the issuer. Sections 2.2 and 2.3 will describe how to determine the level of default risk. The market interest rate, by nature is floating. The present value of a bond can be calculated by the following equation (van Aalst et al. (1997), page 158).

$$PV = \sum_{t=1}^{T} \frac{r_c \cdot F}{(1+r_f)^t} + \frac{F}{(1+r_f)^T}, \tag{2.1}$$

where r_c is the coupon interest rate, r_f is the current market interest rate, t is the time period between two interest payments, F is the face value and T is the term-to-maturity of the bond. This equation shows that if the market interest rate increases, the present value of a bond will automatically decrease. In such a situation the owner of the bond will probably swap his bond with another financial investment. On the other hand, if the market interest rate decreases the value of a bond will increase. Now other investors presumably want to buy these bonds.

2.2 Bond rating

This section is devoted to the event of bond rating. Belkaoui (1983) and Tan (2000) quoted that according to Standard & Poor's, "a bond rating is an opinion of the general creditworthiness of an issuer with respect to a particular debt security or other financial obligation, based on relevant risk factors." Belkaoui (1983) recapitulated this definition and quoted that a bond rating is intended to indicate how likely it is that the issuer will be able to meet principal and interest payments on time, therefore a bond rating is intended to measure the default risk. The following three subsections will discuss the history, importance and process of bond rating.

2.2.1 History of bond rating

The history of corporate bond rating goes back to the beginning of the twentieth century[3]. In that period, bond rating originated in the United States of America, largely through the efforts of Roger Babson, Freeman Rutney and John Moody. In 1909 John Moody published the first edition of the "Analysis of Railroad Investments". This journal basically introduced the first company ratings by its newly proposed rating scale. Rutney was related to Poor's Publishing Company, that was founded in 1916, which in fact started with the idea of selling that kind of financial information. Actually, Moody's journal was a reaction to the financial situation of that period. The need for financial debt providers was enormous, in order to realize an economical growth. The credit rating system of Moody's gave debt providers the possibility to compare different companies. Fitch Investor Service was another agency which became active in this bond rating industry. The three rating agencies developed their rating process over the years. Especially since the "Great Depression of 1930", these rating agencies take the cyclical developments of the economy into account ,which results in a more conservative way of rating companies. Standard & Poor's introduced a fee structure for its services in March 1968. From that moment they asked for money in exchange for their services. Moody's rapidly followed with this *rating-for-a-fee* structure, which resulted in the situation that other, new rating agencies came into existence.

Today the most influencing corporate bond rating companies are Fitch, Moody's Investor Services and Standards & Poor's Corporation(S&P). Fitch Investor Services essentially only rates banks, in contrast to the other two who rate all types of organizations. All three of them assess the relevant factors, which comprise of quantitative and qualitative factors relating the creditworthiness of a company. These assessments normally result into a rating which is reflected into a letter. Table 2.1[4] represents these rating letters which the rating agencies use.

[3]This subsection in based on Belkaoui (1983), van der Ent (1992) and van der Ent and Smant (1994)
[4]Source of information:
www.moodys.com, www.standardandpoors.com, and www.fitchratings.com

Rating scales			
S&P	Fitch	Moody's	explanation
AAA	AAA	Aaa	Exeptionally strong
AA+	AA+	Aa1	Very strong
AA	AA	Aa2	Very strong
AA-	AA-	Aa3	Very strong
A+	A+	A1	Strong
A	A	A2	Strong
A-	A-	A3	Strong
BBB+	BBB+	Baa1	Good
BBB	BBB	Baa2	Good
BBB-	BBB-	Baa3	Good
BB+	BB+	Ba1	Moderately weak
BB	BB	Ba2	Moderately weak
BB-	BB-	Ba3	Moderately weak
B+	B+	B1	Weak
B	B	B2	Weak
B-	B-	B3	Weak
CCC+	CCC	Caa	Very weak
CCC	CC		Very weak
CCC-	C		Very weak
CC	DDD	Ca	Distressed
C	DD		Distressed
D	D	D	Defaulted

Table 2.1: Rating scale for the three largest rating agencies. The explanation is based on the payment capacity. Therefore a BB+ rating assigned by S&P means that companies with that rating have a moderately weak payment capacity.

It is hard for rating agencies to conceptualize a model which is applicable to every organization, because of the regional and industrial differences. If they would use the same model for each industry or region, it would almost be impossible to compare ratings. For example, the financial and non-financial information on a bank differs substantially from the information on an industrial. For this reason they use different models for different industries and regions. As mentioned before, these rating scales provide a relative rank ordering of creditworthiness. Further details on the chances of default, which are narrowly linked with this creditworthiness, will be discussed in the following paragraph.

2.2.2 Importance of bond ratings

Subsection 2.1.1 ventilated about the fact that increasingly more individuals, firms and governments make use of debt financing. Generally, the acquired debt is used by these entities to be able to grow in an earlier stage than in the situation where they have to wait for retained earnings. Having capital in an earlier stage is frequently seen as an advantage for the issuers. Nevertheless an enormous disadvantage comes with this advantage, namely an increase in the level of risk. As discussed in the previous section, the event of lending money brings obligations with itself. For that reason it is of great importance to both the issuer and the bond holder to be informed about the capability of the issuer to keep promise to these obligations. The three rating agencies express this capability in letters, which indicate the level of risk, in other words the probability that an issuer defaults on a bond. According to Moody's Investor Services the default rates, belonging to their rating classes for the period 1983-1999, for the lower rating classes are much higher than for the rating classes Aaa-Baa3. (Figure 2.1 illustrates this, source of information: Moody's (2000) pages 15 and 26.)

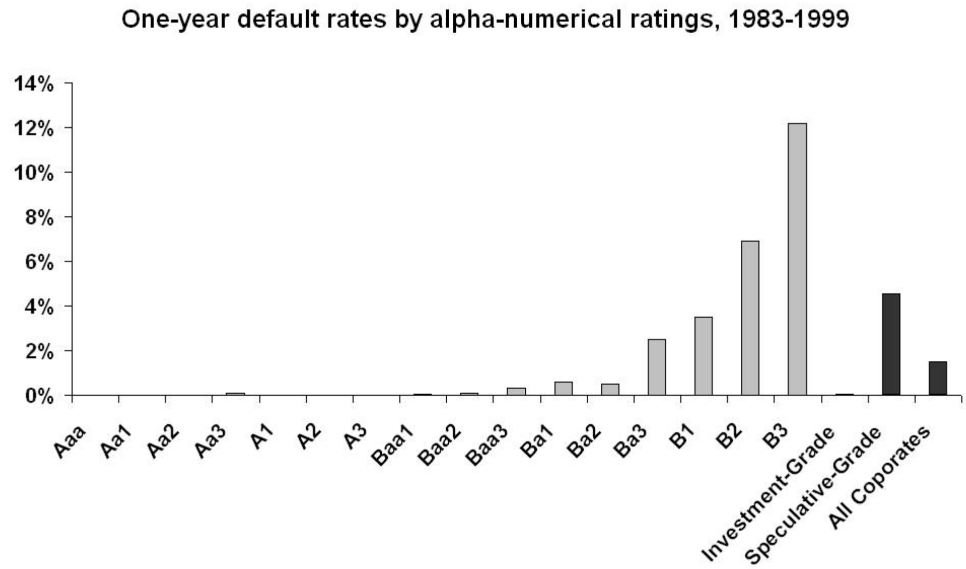

Figure 2.1: Bar diagram showing the one-year default rates by alpha-numerical ratings, 1983-1999. The lower the rating class the higher the default rates. Rating classes Aaa-Baa3 are assigned to the Investment-Grade group, the other lower rating classes are assigned to the Speculative-Grade group.

The results shown in 2.1 confirm the fact that bond ratings can be helpful in the risk evaluation process of a bond.

According to Belkaoui (1983), at least the following six arguments show the importance of bond ratings:

Bond quality Bond rating agencies try to give a judgement on the future of a company. They tend to keep this judgement conservative and based on the future, past and present status of a organization. This judgement can be seen as an indicator of the probability of default.

Default probability This argument is narrowly linked with the aforementioned argument. It says that bond ratings are useful because they have proved to be good predictors of bond defaults.

Bond yield to maturity Bond ratings have proved to be inversely correlated with bond yield maturities The definition of Yield to maturity given by Investopedia (2005) is: “The rate of return anticipated on a bond if it is held until the maturity date”. This fact can be explained in two ways. The first explanation tells that ratings actually identify the coupon rate. The second explanation is based on the fact that both bond ratings and bond yield are determined by the same underlying economic factors.

Beta The beta[5] of an organization, which in fact indicates the relative risk in comparison with the market, is considered as the folding ruler of risk that stock investors face. Bond ratings also measure the risk involved with an organization, and for that reason both the betas and the bond ratings of organizations are related.

Market impact Bond rating can have a market impact, but at the same time the market can have impact on bond ratings. Evidence pointed out that the time between the realization of new market information and the interconnected bond rating change takes at least six months.

Usefulness of bond ratings Bond ratings are useful to all parties of interest. Issuers who receive the rating of their bonds, immediately have the joint coupon rate determined by the rating company. Investors and banks receive an evaluation of the relative risk which is connected to the rated bonds.

2.2.3 The rating process

This subsection is dedicated to the bond rating process[6]. The rating processes as practised by the rating agencies nowadays are considerably mysterious, so far they only share the global idea behind their rating process. It is known that their bond rating processes are based on quantitative and qualitative analysis.

An analysis can be done on request or without request, in some cases by one agency in other cases by more agencies at the same time. The following two paragraphs distinguish between the on request and without request rating processes. Both paragraphs discuss all the steps taken by the rating agencies.

Bond rating on request

In the situation where an issuer requests a rating, the issuer has to pay a fee for the services which the rating agencies provide. In this situation the rating agencies can and make use of public and non-public information, what usually lead to reliable ratings.

The rating process involves two entities, namely the issuer and the rating agency. Figure A.1(see Appendix A) displays the rating process, and also visualizes the interaction between these two entities. The steps taken by figure A.1 will be discussed individually in the following enumeration.

[5]The definition of beta given by Grinold and Kahn (1999) is : “The sensitivity of an portfolio or asset to a benchmark. For every 1 percent return to the benchmark, we expect a beta return to the portfolio or asset

[6]Sources of information: Belkaoui (1983), van der Ent and Smant (1994), van der Ent (1992), Tan (2000) and Standard and Poor’s (2005)

1. An organization requests for a rating by one of the rating agency. In some cases this organization wants to know what the rating may be if it decides to issue bonds. In other cases its interest goes to the effect on the ratings of earlier issued bonds, when the organization is planning to issue additional bonds.

2. The issuing organization has to satisfy to the restrictions set by the rating agency. The requirements of the rating agencies are announced to the issuer during their first meeting. Rating companies, for example, will not rate issues smaller than 10 million dollar, or companies which are younger than 5 years (van der Ent and Smant (1994)).

3. The rating agency will consider the request forms contributed by the issuer. Rating agencies will always be in the position to reject rating requests. Reasons for rejection can diverge, although the main idea laying behind a rejection is the fact that rating agencies want to be entirely sure about the rating they finally announce.

4. The rating agency will assign an analytical team to assess the issuer, if the request is accepted. The selection of the team-members is based on their specialism.

5a. The analytical team will gather all relevant information. This information comprises of internal and external information, which are swept together by unilateral research done by the rating company and by way of meetings between the two entities. Financial statements of the past 5 years (Belkaoui (1983)), comparisons with similar companies and analysis of capital spending are examples of information which the analytical teams are looking for.

5b. An important factor in the information gathering process is the willingness to cooperate by the issuer. In some cases the issuer is asked to deliver a presentation to the rating agency. The better the information, the more accurate the rating will be.

6. After having done the analysis, the analytical team will prepare a presentation for the rating committee. This extensive presentation comprises of all relevant information, by virtue of which the rating committee can define the proper rating.

7a. The issuer will be informed by the rating committee. If the issuer is in disagreement with the proposed rating or the issuer has new relevant information, it can appeal for reconsideration by the rating committee.

7b. The rating committee will determine the rating of the issue and will inform the issuer. If the issuer disagrees with this rating, the rating committee can consider a revision.

8. The rating agency will announce the final rating officially.

9. When the rating agency announced the rating, the issue of the bonds can start. At this point all the potential investors are provided with objective judgements

about the long term debt, and the issuer is informed about the potential risk. The issuer can roughly determine the coupon rate belonging to the bonds to be issued.

10. The process does not stop after the aforementioned 9 steps. The new rating will be supervised on an on-going basis.

Up until now the rating agencies did not make their rating procedures transparent, therefore the aforementioned step-by-step plan can exclusively be seen as an rough indication of their rating procedures.

Bond rating without request

Although most of the rating processes done by the rating agencies are performed on request, some are accomplished without request. The without request rating process does not differ much from the process described in the previous paragraph. The fact that the issuer is not obliged to pay a fee, is the main difference. Another segment from the rating process that can vary is the way of discovering relevant information. In some situation, rating agencies are obliged to exclusively use public information. Even though many of the companies, which are supervised by a rating company, are willing to cooperate, given that these companies are always availed if they get a favorable rating. The without request rating process normally starts at step 4 in figure A.1 appendix A. The right half of the figure is being used solely if the rated company is not willing to cooperate.

2.3 Literature review on bond rating models

So far this chapter has discussed all aspects of about bonds, bond rating agencies and the way these agencies rate bonds. As mentioned in section 1.4, the goal of this report is to construct an accurate bond rating prediction model using Probabilistic Fuzzy Systems (PFS). Before constructing this model, it is important to survey the manifold prediction models described in previous studies. Substantial literature can be found on bond rating prediction, most probably due to the secretiveness around the rating procedures that the rating agencies practise. What is known, is the fact that these rating agencies make use of financial ratios[7], which are quantitative factors and qualitative factors. According to Ang and Patel (1975), these qualitative factors are subjective judgements concerning the managerial quality of an organization, the quantitative factors concern the value of the intangible assets, and the ability to satisfy to the financial commitments made between the issuer and the bondholder. However, most of the bond rating prediction models found in the literature utilize only quantitative historical data. The financial information used in the literature for the construction of the bond rating models diverge greatly. This also is the case for the financial ratios derived from the financial information gathered for the bond rating prediction.

A large number of attempts have been made in the past to predict bond ratings with quantitative models. All of these attempts tried to clarify the ratings based on

[7] Beaver (1966) quoted: “A financial ratio is a quotient of two numbers, where both numbers consist of financial statement items”

quantitative historical data. The fact that these models do not utilize qualitative data most probably is the deficiency in their rating capability. The methods used in prior research can be categorized into statistical and artificial intelligence methods. The following two paragraphs will show the different previous research separated into the two categories.

Before starting the discussion on the statistical bond rating models, first several studies considering bond quality and the existence of a relationship between bond ratings and historical records of bond defaults will be evaluated. Harold (1938) was the first to examine the behavior of bond ratings. He compared the performance of corporate bonds rated by the rating agencies from 1929 to 1936[8]. Harold concluded that investors should reconsider the bases of bond ratings, because of the fact that the rating agencies now and then rate the same bonds differently. Another study in the field of bond rating was conducted by Hickman (1958) and he investigated straight corporate bonds offered from 1900 through 1943. He compared nine prospective measures of bond quality with four measures of investor experience (Belkaoui (1983)). The first of Hickman's convincing finding is that indicators, like rating agencies, market ratings and legal lists, of prospective bond quality proved to be useful. Second, bonds selected by agency ratings were found to be more stable than bonds meeting a fixed market rating standard. The third finding was related to the fact that business cycles and the difficulty to predict trends often cause errors in the rating process. Atkinson and Simpson (1967) continued on the study done by Hickman. They compared the corporate bond quality of postwar bonds with prewar bonds and concluded that defaults of bonds outstanding decreased from an average of 1.7% to an average of 0.1%.

2.3.1 Statistical bond rating models

According to Maher and Sen (1997) and Huang et al. (2004), the seminal work in the area of bond rating prediction was done by Fisher (1959). Fisher implemented ordinary least squares (OLS) to explain the variance of a bond's risk premium, where he defined this risk premium as the difference between the market yield to maturity and the corresponding rate of interest. Studies done by Horrigan (1966) and West (1970) also utilized OLS to predict corporate bond ratings. Horrigan's study describes the ability to predict bond ratings with accounting data. He tried to predict the top six rating classes as described by Moody's and Standard and Poor's based on six variables[9]. Horrigan was able to predict 58% of Moody's ratings correctly and 52% of Standards and Poor's ratings. West criticized Horrigan's model, and suggested that the model of Fisher was theoretically and empirically better. With his implementation of the Fisher model, which made use of four variables, West was able to classify 62% correctly. Another study conducted by Pogue and Soldovsky (1969), made use of a regression model with a dichotomous dependent variable to predict to which rating a bond should be assigned. Their model was able to predict 50 out of 53 correctly, from the experimental sample, and 8 out of 10 bonds from the holdout sample.

[8]This paragraph and the next section is based on: Belkaoui (1983), Huang et al. (2004), Ang and Patel (1975), Kaplan and Urwitz (1979), van der Ent and Smant (1994), Pinches and Mingo (1973a), Gentry et al. (1988) and Maher and Sen (1997)

[9]Subordination, total assets, working capital/sales, net worth/total debt, sales/net worth and operating profit/sales

Pinches and Mingo (1973b, 1975) take on a different view as they experimented with multiple discriminant analysis (MDA) in order to predict corporate bond ratings. MDA is a statistical technique capable of classifying observations into a priori known groupings based on the characteristics of the observations. They tried to increase the classification accuracy by applying the MDA technique, which should better suit the ordinal nature of bond-rating data. Pinches and Mingo first examined 35 variables with factor analysis which potentially could be used within their model. Their MDA model was determined by means of six selected variables and showed prediction accuracies of 65% and 56% for holdout samples in the period 1967-1968 and 1969 (Ang and Patel (1975)).

Until this point, all studies compared their predicted results with the bond ratings given by the rating agencies. They assumed that the ratings given by these rating agencies were correct. Ang and Patel (1975) doubted this assumption, which resulted in a study with a twofold purpose. The first purpose was to compare the statistical bond rating models, proposed by Horrigan (1966), West (1970), Pogue and Soldovsky (1969) and Pinches and Mingo (1973b, 1975), on their ability to duplicate the ratings determined by Moody's. The second purpose was to compare the ability of Moody's, and all other the bond rating methods, to predict financial distress over different times periods. The study pointed out that most of the statistical models do good work at much lower costs than rating agencies, at least when the objective is to forecast short-term probability of financial loss.

Kaplan and Urwitz (1979) commented on the regression and discriminant models discussed in the previous paragraphs. Bonds convey ordinal information[10], and therefore regression models are less applicable because they treat the dependent variable as if it is on an interval scale. MDA models, on the other hand treat bond ratings as classifying bonds into separated categories. Kaplan and Urwitz (1979) utilized multivariate probit analysis in order to take advantage of the ordinal nature of bond rating and showed 68% prediction accuracy for their model. However the regression model showed 71% accuracy for the same data set, implied that the regression model seemed to be more robust. Nevertheless, another much-discussed study done by Belkaoui (1983) preferred to use MDA for the bond rating prediction problem. Belkaoui mentioned several arguments why he chose for an MDA application. His first argument was based on the fact that multivariate probit and MDA models are better applicable than regression models to the ordinal scale which bonds convey. Another argument brought forward is that regression models are more robust than multivariate probit models. These arguments motivated Belkaoui to indicate that MDA as the most appropriate model for BRC (Kaplan and Urwitz (1979)). Belkaoui showed 62.8% prediction accuracy for the experimental example and even 65.9% for the test data set.

Several new studies on statistical models for bond rating prediction have been conducted, since Belkaoui's work, which most often showed prediction accuracies between 50% and 70%. Various financial variables were used to predict bond ratings, although most of them were related to measures of size (Huang et al. (2004)), financial leverage, long-term capital intensiveness, return on investment, short-term capital intensiveness, earnings stability and debt coverage stability. Unfortunately, it did not significantly improve the performance of the previous prediction models. The follow-

[10]Because a bond with an AAA rating is more secured than a bond with an AA rating and so on

ing section discusses some relatively new techniques for bond rating prediction, based on artificial intelligence.

2.3.2 Artificial intelligence bond rating models

Schutzer (1987) cited the following definition of artificial intelligence (AI): "Artificial intelligence is a field of study concerned with designing and programming machines to accomplish tasks that people accomplish using their intelligence". In other words, AI tries to use intelligence like human beings do, by using computers or other machines which are able to cope with an enormous level of possibilities.

Recently, AI techniques have done their entrance in the field of bond rating prediction. Inductive learning, case based reasoning, artificial neural networks, rule-based expert systems, self-organizing maps and many other machine learning techniques have been examined in previous studies. An advantage of these AI techniques is the fact that these techniques extract knowledge from an input data set, which is used to construct different models to represent the data set.

Inductive learning is one of the AI techniques utilized for bond rating prediction, which indeed extract knowledge from a training data set in order to construct a predicting model. Induction algorithms such as ID3 and Classification and Regression Trees (CART) (Quinlan (1986), Breiman et al. (1984)) bring forth decision tree based prediction models, which are able to assign observations to one of the bond rating classes. According to Shin and Han (2001), Shaw and Gentry (1990) have concluded that the performance of their inductive learning model was better than the performance of the models using probit and logit analysis, which both are statistical techniques. The better performance was attributed to the fact that inductive learning does not make use of parametric and structural assumptions, where statistical models do use these assumptions.

Artificial neural networks (ANNs) are the most frequently used AI technique for bond rating prediction. In many of the previous research done on AI methods for bond rating prediction, the authors compared ANNs with statistical or with other AI techniques. These ANNs or simply neural networks offer another suitable classification possibility for the bond rating prediction problem. The first attempt to use ANNs for bond rating prediction was done by Dutta and Shekhar (1988). Their ANN showed a prediction accuracy of 83.3% in discerning "double-A" from "non double-A" rated bonds using a three-layer ANN, for both implementing 6 and 10 variables. They used a linear regression model as the benchmark technique, which did not surpass 50% prediction accuracy. Another paper devoted to ANNs for bond rating was done by Singleton and Surkan (1990), which used a data set consisting of bonds of 18 Bell Telephone companies divested by American Telephone and Telegraph Company (AT&T) in 1982. Their model had to discern between two classes, namely between the bond rating class "Aaa" and the combination of the classes "A1, A2, A3" by Moody's. The performance of the ANN was compared to the performance of a multiple discriminant analysis (MDA) and proved to be more accurate in predicting bond rating classes. The best network showed a prediction accuracy of 88% on the test data set. Kim et al. (1993), Moody and Utan (1995), Maher and Sen (1997) and Kwon et al. (1997) showed other, more recent studies on the performance of ANNs for bond rating prediction. All of their studies pointed out the fact that neural networks perform really well in the world of bond rating prediction, and outperform statistical models in most of the cases. Although ANNs show relatively high prediction accuracy, these neural networks are not easily comprehensible and are often seen as black boxes.

Another AI method seen in previous studies report on the use of case-based reasoning (CBR), which make use of the natural form of knowledge in contrast to ANNs. According to Shin and Han (2001); "the basic principle underlying CBR is the fact that human expert use analogical reasoning to solve complex problems and to learn from problem-solving experiences." Shin and Han (2001) proposed a CBR approach to bond rating prediction, which made use of the nearest-neighbor algorithm to retrieve similar past cases and inductive learning for case indexing. Their experiment showed that their CBR model outperformed ID3 and MDA, prediction accuracies were 75.5%, 59% and 60% respectively.

An application of self-organizing maps (SOMs) for credit rating prediction is described in the Master thesis written by Tan (2000). SOMs make use of an advanced clustering algorithm, as a result of which SOMs are able to compress a complex and voluminous data set into a two-dimensional map. The SOMs constructed by Tan (2000) showed a prediction accuracy about 80% for ratings with a maximum error of at most two notches. The model constructed for this thesis used a selection of financial, quantitative ratios. Qualitative ratios were not taken into account for the construction of the model. The model showed that the SOM algorithm was able to cluster the observed companies based on pure quantitative, financial ratios. A deficiency in the study done by Tan (2000) is the fact that the semi-supervised learning with the self-organizing map was not compared with a normal-supervised learning technique, in order to get a better insight into the special characteristics of SOMs and its performance.

2.3.3 Conclusions

Most of the bond rating prediction techniques used in the past have been taken into consideration in the previous two sections. As mentioned before, these techniques can be subdivided into two types of approaches, namely into the statistical and into the artificial intelligence approach for bond rating prediction. Section 2.3.1 indicated that statistical approaches, which make use of quantitative financial variables, are able to provide us with a relatively simple model that show an approximation of the complex and subjective bond rating process. Many of the outlined studies expounded different important financial variables, and showed prediction accuracies between 50 and 70 %.

Recent applications of artificial intelligence techniques used for bond rating prediction are sketched in section 2.3.2. The models produced by these AI techniques have stronger embedded learning capabilities and are evidently more complex than the statistical models seen before. As a result, most of these models show higher prediction accuracies, however they are often more difficult to comprehend. ANNs, the most often described AI technique for the bond rating prediction problem, showed the best prediction accuracies. Unfortunately, these ANNs are difficult to see through and are frequently seen as black boxes. Mainly because of that reason, the literature on ANNs for bond rating prediction strictly focussed on the prediction performance of the constructed models.

The main goal for this report was to construct an AI model for the bond rating prediction problem which shows high prediction accuracies. A secondary goal was to construct a better understandable bond rating classifier than the AI models seen before. Probabilistic fuzzy systems (PFSs) is the selected AI technique which possesses the characteristics to construct a system which is able to follow our goal. The technical background of PFS will be sketched in chapter 3, the application of the PFS for bond rating prediction will be outlined in chapter 4.

As mentioned in the previous paragraphs the PFS model hopefully takes away the black-box effect, at least for the greater part. Another respectable element of the prediction model is the prediction accuracy. The produced model will be exposed to a training set, followed by a test set, in order to examine the prediction performance. The PFS model will be compared to a statistical model, which in fact will act as a benchmark model. The statistical technique used in this report is multiple discriminant analysis (MDA). The reason for choosing this MDA as the benchmark technique was rather obvious, because this technique simply is the most common used and most examined statistical technique for bond rating prediction. Besides that, this technique proved to be robust and has showed respectable prediction accuracies for this classification problem. The technical background and the implementation of the MDA model will be described in Chapter 3 and 4, in the way as for the PFS model.

Chapter 3

Methodology

This chapter describes the techniques used in this research. The first section discusses the statistical model multiple discriminant analysis (MDA). This MDA model functions as the benchmark model within our research. "Probabilistic Fuzzy Systems" (PFSs) will be described in section 3.2, and are used as the main technique to predict bond rating classes within this research. First the technical background of this AI technique will be described. Subsequently a sketch of a PFS model application for the bond rating problem will be given.

3.1 Multiple discriminant analysis

This section will be dedicated to a MDA approach for bond rating prediction. As mentioned before in Chapter 2, MDA models are heavily used for BRC. Belkaoui (1983) mentioned two main reasons for the fact that MDA wins over other statistical models. The first reason is based on the fact that bonds convey ordinal classes, as a result of which MDA and multivariate probit models are better appropriate than regression techniques. The second reason in fact is an indirect reason, actually Belkaoui (1983) reported that regression models are assumed to be more robust than multivariate models, which imputes that MDA models are most appropriate for bond rating prediction.

Actually, (multiple) discriminant analysis models have always been very popular in the world of risk assessment of companies. Altman (1968) introduced the well known Z-score model, which is seen as the standard tool for bankruptcy prediction ever since the introduction. This Z-score model is based on discriminant analysis and distinguishes between bankrupt and non-bankrupt firms.

MDA is a statistical technique used to classify an observation into one of the, a priori determined, groupings dependent on the observation's individual characteristics. This MDA model is primarily used for predicting qualitative dependent variables, so very suitable for the bankruptcy prediction and bond rating prediction problems[1]. The MDA technique has the advantage to take several characteristics into account at the same time, as well as their inter-relational behavior. Another strength of MDA is the ability to reduce the "a priori" groupings into a small dimension. MDA reduces the dimension in the bankruptcy prediction problem to the simplest form, where the a priori groupings are bankrupt and nonbankrupt. Because of this, the MDA model used

[1] Bankruptcy prediction; bankrupt, nonbankrupt. Bond rating; the ordinal scaled rating classes

for bankruptcy prediction can be treated as "simple" discriminant analysis (DA). For our problem, the bond rating prediction, the MDA model will make use of more than 2 groupings.

3.1.1 Technical background: Multiple discriminant analysis

Before exaggerating about the MDA application for the bond rating prediction problem, we need to have a clear description of discriminant analysis (DA) on its own. Discriminant analysis, like analysis of variance, is an analysis of dependence method which actually is a variant of canonical correlation (Lattin et al. (2003)). However in the case of discriminant analysis the dependent variables are categorical, which divide the set of observations into mutually exclusive and collectively exhaustive groups. Simple discriminant analysis, which has only two groups, only needs a single dichotomous dependent variable to indicate group membership. For MDA we need $n-1$ dichotomous variables to indicate group membership across n groups. Discriminant analysis make use of information about independent variables, and turn that information into the clearest possible separation between or among groups. The Fisher approach, a well known discriminant analysis approach, finds the linear combination of independent variables that produces the best discriminant score. If we look at the scatter plot 3.1 [2] and keep the two plots of figure 3.2 in mind, we see that figure 3.4 represents the best linear combination to discriminate between the two groups.

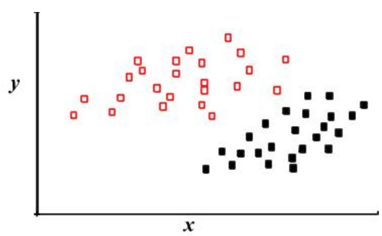

Figure 3.1: Scatter plot showing two groups.

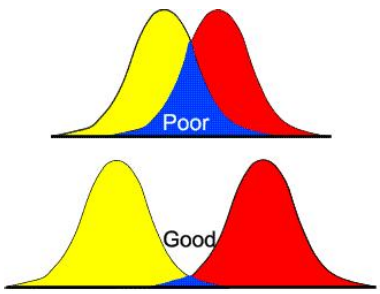

Figure 3.2: Visualization of two possible linear combinations of two independent variables.

[2]Source of information:
http://www.doe-mbi.ucla.edu/~parag/multivar/dawords.htm

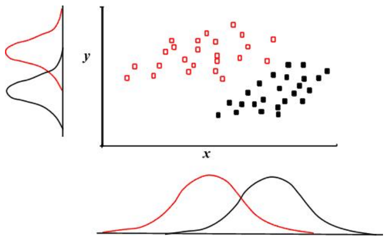

Figure 3.3: Scatter plot of two groups using x or y to discriminate between both groups.

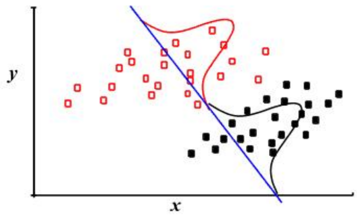

Figure 3.4: The best linear combination of x and y to discriminate between both groups.

When we look at the technical background and when $\mathbf{k}$ denotes the linear combination, then the discriminant scores are given by;

$$\mathbf{t} = \mathbf{X}\mathbf{k} \tag{3.1}$$

The linear combination of $\mathbf{k}$ has to maximize the ratio of the between-group sum of squares to the within-group sum of squares of the discriminant scores $\mathbf{t}$. This ratio is proportional to:

$$\frac{\mathbf{k}'\mathbf{d}\mathbf{d}'\mathbf{k}}{\mathbf{k}'\mathbf{C}_W\mathbf{k}} \tag{3.2}$$

where $\mathbf{d} = (\bar{\mathbf{x}}_{(2)} - \bar{\mathbf{x}}_{(1)})$ is a vector describing the difference between the means of the two groups, and $\mathbf{C}_W$ is the pooled within-group covariance matrix of $\mathbf{X}$. So the smaller the within-group variation the larger the objective function. In fact, equation 3.2 is maximized by choosing $\mathbf{k}$ as follows[3]:

$$\mathbf{k} \propto \mathbf{C}_W^{-1}\mathbf{d} \tag{3.3}$$

Formula 3.1 calculates the discriminant function scores for all the observations in the data set.

A cutoff score, t_c, can be used to categorize observations. All observations with discriminant function scores $t > t_c$ are assigned to one group, the others are assigned to the other group. The following formula will do to calculate the cutoff score for a two group discriminant analysis

$$t_c = \frac{(\bar{t}_{(1)} + \bar{t}_{(2)})}{2}, \tag{3.4}$$

where $\bar{t}_{(1)} = \bar{\mathbf{x}}'_{(1)}\mathbf{k}$ and $\bar{t}_{(2)} = \bar{\mathbf{x}}'_{(2)}\mathbf{k}$ are the discriminant function scores of the two group centroids. This formula is only appropriate when the two groups are of equal size. When the groups sizes differ the following formula for the cutoff score is

$$t_c = \frac{(n_1\bar{t}_{(1)} + n_2\bar{t}_{(2)})}{n_1 + n_2} \tag{3.5}$$

This equation minimizes the expected probability of misclassification.

[3]See Lattin et al. (2003), page 436 for the technical derivation of this formula

At this stage the observations can be classified into one of the two possible groups. With these classified observations we can be formulate an accuracy matrix(see table 3.1, Altman (1993)). Where H stands for correct classifications, M_1 for a Type I error and M_2 represents a Type II error. To see how accurate the model is, you have to sum the H diagonal and divide this number by the total number of firms in the data set. This will give the accuracy percentage.

	Predicted Group Membership	
Actual Group Membership	Group 1	Group 2
Group 1	H	M_1
Group 2	M_2	H

Table 3.1: The accuracy matrix used to determine the accuracy of a discriminant analysis model.

This accuracy, or hit rate, is a simple and intuitive approach to measure the goodness of fit for the classification model. Though, the question is; "is this accuracy percentage reasonable?" For this reason the accuracy should be compared to some sort of benchmark. To test the accuracy of the model the frequently used benchmark method, the proportional chance criterion, is implemented. This method creates the following accuracy matrix based on the relative frequency with which each group appears in the data.

	Expected number classified by chance		
Actual Group Membership	Group 1	Group 2	All
Group 1	np^2	$np(1-p)$	np
Group 2	$np(1-p)$	$n(1-p)^2$	$n(1-p)$
All	np	$n(1-p)$	n

Table 3.2: The accuracy matrix based on the relative frequency with which each group appears in the data.

Where p is the chance that an observation will be assigned to 'Group 1' and $1-p$ is the chance the observation will be assigned to the 'Group 2'. The expected correctly classified observations which follow from the matrix 3.2 is

$$h_{expected} = np^2 + n(1-p)^2 \tag{3.6}$$

With this expected hits we can calculate expected hit ratio

$$hr_{expected} = \frac{h_{expected}}{n}, \tag{3.7}$$

which obviously return 50 percent for $hr_{expected}$ when the group sizes are equal. With this $h_{expected}$ and $hr_{expected}$ we can do a t-test. This t-test calculates if the chance of coincidence of a certain accuracy will be negligible. In fact the accuracy is compared with the expected accuracy, based on the probability of classifying an observation in the right way. Before we can accomplish this *t*-test we need to calculate the last missing variable. The standard deviation of the number of expected hits from the proportional chance criterion. This standard deviation is given by

$$s_{CPRO} = \sqrt{nhr_{expected}(1 - hr_{expected})} \tag{3.8}$$

We can now execute the *t*-test, which is given by

$$t = \frac{h_{actual} - h_{expected}}{s_{CPRO}} \tag{3.9}$$

The objective in for MDA does not differ from the objective in DA. The only difference is the fact that the number of groups exceeds two with MDA, which results in more than one dependent variable[4]. Letting **W** denote the within-group sum of squares matrix and **A** denote the across-group sum of squares matrix, then the objective for this discriminant problem is to find **k** which maximizes the following formula

$$\lambda = \frac{\mathbf{k}'\mathbf{A}\mathbf{k}}{\mathbf{k}'\mathbf{W}\mathbf{k}}, \tag{3.10}$$

When we take the derivative and solve the first order condition for **k** we get

$$\mathbf{W}^{-1}\mathbf{A}\mathbf{k} = \lambda\mathbf{k}, \tag{3.11}$$

where λ is the eigenvalue, which brings along the familiar structure of an eigenvalue problem. A major difference between MDA and DA is the fact that there may be more than two discriminant functions in the solution to the problem with MDA.

Testing for accuracy and significance of a MDA model differs from DA as well, although it shows similarities. Table 3.2 will be transformed into table 3.3, which has to obey to the following restriction;

$$\sum_{i=1}^{c} p_i = 1, \qquad \text{where } c = \text{number of classes} \tag{3.12}$$

Where p_i is the chance that an observation will be assigned to 'Group i'. The proportional chance criterion, which follow from table 3.3 is

$$hr_{expected} = \frac{n\sum_{i=1}^{c} p_i^2}{n} = \sum_{i=1}^{c} p_i^2 \tag{3.13}$$

The execution of a t-test works the same for MDA as for DA, therefore functions 3.8 and 3.9 suffice to do this test.

[4]For three groups we need two indicator variables. For instance, Y_1 and Y_2 which indicate two groups and if both of them are false $\Rightarrow$ third group

	Expected number classified by chance				
Actual Group Membership	Group 1	Group 2	...	Group i	All
Group 1	np_1^2	np_1p_2	...	np_1p_i	np_1
Group 2	np_2p_1	np_2^2	...	np_2p_i	np_2
...	...	...	...	...	...
Group i	np_ip_1	np_ip_2	...	np_i^2	np_i
All	np_1	np_2	...	np_i	n

Table 3.3: The MDA accuracy matrix based on the relative frequency with which each group appears in the data.

3.1.2 Multiple discriminant analysis application for bond rating prediction

MDA is a frequently used technique to construct BRC models. Section 3.1 presented some of the most important reasons why MDA is as successful as it is. A MDA model which is described in many literature studies was introduced by Belkaoui (1983). He developed a well accepted, accurate bond rating prediction model based on MDA. For this reason, we use his model as the concept for our benchmark model within this report. He introduced his model as a response to several criticisms on the existing bond rating prediction models. Belkaoui quoted the following: "A bond rating is primarily a judgement of the investment quality of a long-term obligation of a firm." The variables of a reliable ratings model must be related to the firm, the market and the indentures of the company.

Belkaoui's model is based on MDA, because of the fact that these models can cope with the ordinal behavior of bond ratings really well. The MDA model was constructed and tested with four randomly selected samples of industrial bonds with a bond rating of at least B by Standard and Poor's. Nine independent variables, showed in figure 3.4, were selected in order to discriminate between the six classes. The model showed

#	Independent variable	#	Independent variable
X_1	Total assets	X_6	Fixed charge coverage ratio
X_2	Total debt	X_7	Five year cash flow as percentage of five year growth needs
X_3	Long-term debt/total invested capital		
X_4	Short-term debt/total invested capital	X_8	Stock price/common equity per share
X_5	Current assets/current liabilities	X_9	Subordination

Table 3.4: Nine independent variables for the MDA model introduced by Belkaoui. For further details on these financial input variables, see Belkaoui (1983) Chapter 4.

a relatively high prediction accuracy, and classified 72.93 percent in one of the correct bond rating classes. The model showed a prediction accuracy of 67.8 percent on another, independent test set. The prediction models seen up until this point did not exceed 65.9 percent, therefore we can denominate this study as an improvement on the existing bond rating prediction models. On top of that, most of the misclassified bonds were rated in the adjacent rating scale to their true ratings. Belkaoui proposed his model as a good alternative next to the ratings given by the official rating companies, especially if the rating agencies do not rate a company of interest for an investor or manager. Whether we can say if this model is the best alternative, stays a subjective opinion. What we can say is that his MDA model is a model with proper characteristics, which showed relatively good results.

3.2 Probabilistic Fuzzy Systems

Probabilistic Fuzzy Systems (PFSs) will be described in this section, a technique which will act as a classifier for bond rating prediction in this report. It is useful to understand the fundamental idea behind fuzzy systems, before we enter the field of these PFSs.

PFSs are an expansion to fuzzy systems, an AI technique initiated by Zadeh (1965). This technique was introduced to cope with the vagueness and impreciseness of the real world[5]. Human beings deal with linguistic uncertainty every day, while mathematical and other AI techniques are not able to manage these linguistic uncertainties. "Tall people are smart" is, for example, a sentence which is difficult to translate into a mathematical or expert model. The difficulty lies in the fact that both the adjectives are "fuzzy". For example, a certain someone can be seen as tall by one while another person would described this certain someone as little. So far both mathematical and AI techniques were not able to cope with these fuzzy, human language, descriptions and therefore always assigned every example to one certain set by using crisp boundaries, defined as a classical set. The translation of the given statement, or a comparable statement, into an expert system causes loss of semantic value. In many cases these semantic values are extremely valuable, for instance in the case were an expert system has to make fuzzy decisions. Examples of these kind of expert systems are automatic pilots in the metro, and diagnostic programs to help a physician making his advisory decisions. Traditional logic is not able to cope with these fuzzy, intuitive decisions, because they are not able to assign one example to more than one set. This deficiency motivated Zadeh (1965) to initiate fuzzy sets in his seminal paper. These fuzzy sets differ from the "classical sets" in the way that they do not have crisp boundaries, in order that a certain example can be assigned to more than one set. This paper has conducted a major break-through in the development of expert systems ever since its release. The technical details on these fuzzy sets/systems as well as of the PFSs, which in fact are of major importance for this report, will be sketched in the following paragraphs.

3.2.1 Technical background: Probabilistic Fuzzy Systems

Fuzzy sets

Fuzzy sets differ from classical sets, which are used in traditional logic. Jang et al. (1997) quoted the following example for a classical set: "A classical set A of real numbers greater than 6 can be expressed as $A = \{x|x > 6\}$, where there is a clear, unambiguous boundary 6 such that if x is greater than this number, then x belongs to the set A; otherwise x does not belong to the set." These traditional sets are not able to assign an example to more than one set, as described in the previous paragraph. Fuzzy systems are able to assign one example to more than one set, by using fuzzy sets which make use of membership functions. These membership functions bring about smooth

[5]As from here this vagueness and impreciseness of the real world will be defined as linguistic uncertainty

transitions, which indicate that one example can belong for 0.8 to set A and 0.2 to set B. Therefore these fuzzy sets express the degree to which an example belongs to a set. This membership value falls in a domain between 0 and 1, whenever using the characteristic function of a fuzzy set. The membership value obtains a value of 1 if it represents absolute truth and a value of 0 if it represents absolute falseness. A fuzzy set A in X is expressed as

$$A = \{x, \mu_A(x) | x \in X\}, \tag{3.14}$$

if X is the universe of discourse a collection of objects indicated by x, where $\mu_A(x)$ is the membership function for the fuzzy set A. Fuzzy and classical sets differ in the fact that the membership value in 3.14 is permitted to have a value between 0 and 1, where the characteristic value in a traditional set is restricted to 0 or 1. The construction of a fuzzy set is dependent on two items, namely on the identification of a proper universe of discourse and on the definition of an appropriate membership function. Membership values are determined by membership functions (MFs), which are outlined by experts or by a fuzzy clustering algorithm. These MFs can be defined in many different ways. Figures 3.5, 3.6, 3.7 and 3.8[6] illustrate four classes of parameterized MFs[7].

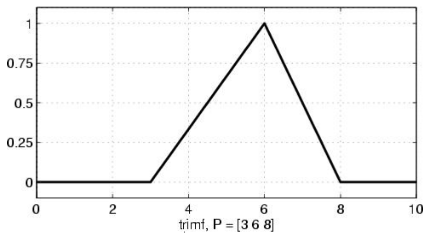

Figure 3.5: Triangular MF

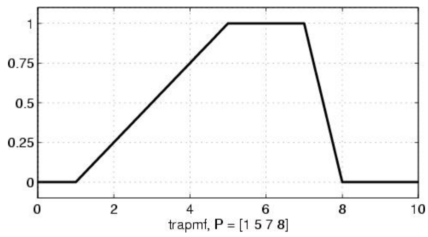

Figure 3.6: Trapezoidal MF

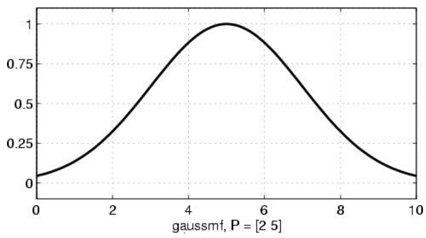

Figure 3.7: Gaussian MF

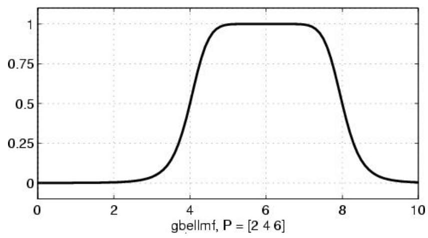

Figure 3.8: Generalized Bell MF

Fuzzy systems

Fuzzy systems are depicted by a set of linguistic statements, which are usually expressed in "if-then" rules, like the following statement; "If Peter is tall then Peter is

[6]These illustrations are derived from the MATLAB web-page: www.mathworks.com

[7]Standard functions for these memberships and other membership functions can be found in Jang et al. (1997)

smart”. Generally these statements can be described as

If antecedent ***then*** consequent.

A fuzzy rule can also have more than one antecedents, for example; “***If*** Peter is tall and ***if*** Peter is wearing classes ***then*** Peter is smart”. One fuzzy rule can have as many antecedents as required.

The concept of the previous described fuzzy rules are combined in a fuzzy inference system (FIS)[8]. A FIS, or simply fuzzy system is a popular framework which can be used for many artificial intelligence applications, like experts systems, classification models, decision analysis and many more. Every FIS comprises of three components;

Rule base, which includes the fuzzy rules used for the particular FIS

Database, which points out the MF used in the fuzzy rules in the particular FIS

Reasoning mechanism, which acts as the component that tries to deduct a reasonable solution from the proposed fuzzy rules

FISs can handle fuzzy or crisp inputs and produces fuzzy or crisp outputs, this differs per FIS and will be illustrated in the following paragraphs. If the FIS produces fuzzy outputs a *defuzzification* method needs to be introduced to extract a crisp value which best represents the fuzzy set created by the FIS. Figure 3.9[9] is an example of a FIS with crisp inputs and a fuzzy output, which is defuzzified in the last step. The FIS, with all three components, is visualized in the quadrilateral in this figure. According to Jang et al. (1997), FISs can be seen as systems which implement nonlinear mappings from its input space to its output space. This mapping is performed by the fuzzy if-then rules in the FIS, where the antecedent defines the fuzzy region in the input space, while the consequent takes care of the output side, which can be fuzzy or crisp. The following paragraph will expound two of the most heavily used FISs, followed by a the explanation on probabilistic fuzzy systems PFSs, the technique examined in this report.

Mamdani fuzzy inference system [10] The main difference between the two FISs discussed in this summary can be retrieved in the consequents of their fuzzy rules. The consequents in a Mamdani FIS are always fuzzy, which implies defuzzification at the end of the FIS. Figure 3.10[11] shows the process of a Mamdani FIS, with all the five steps to be taken by each example. Step two and three show that the consequents are fuzzy indeed. The last step, step five, is the defuzzification step. As described before this defuzzification step is nothing more than retrieving a crisp value out of a fuzzy set, and can be executed in different ways. Most of these defuzzification methods are based on experimental results, because they are not easily subjected to mathematical analysis. This implies the fact that it could be less time consuming and maybe even less biasing to use another FIS, which does not need to make use these defuzzification methods.

[8]This paragraph is based on Jang et al. (1997) Chapter 4

[9]Source of information:
http://www.owlnet.rice.edu/~davids/vlsi/blockdiagram1.gif

[10]Mamdani and Assilian (1975)

[11]Source of information: www.mathworks.com

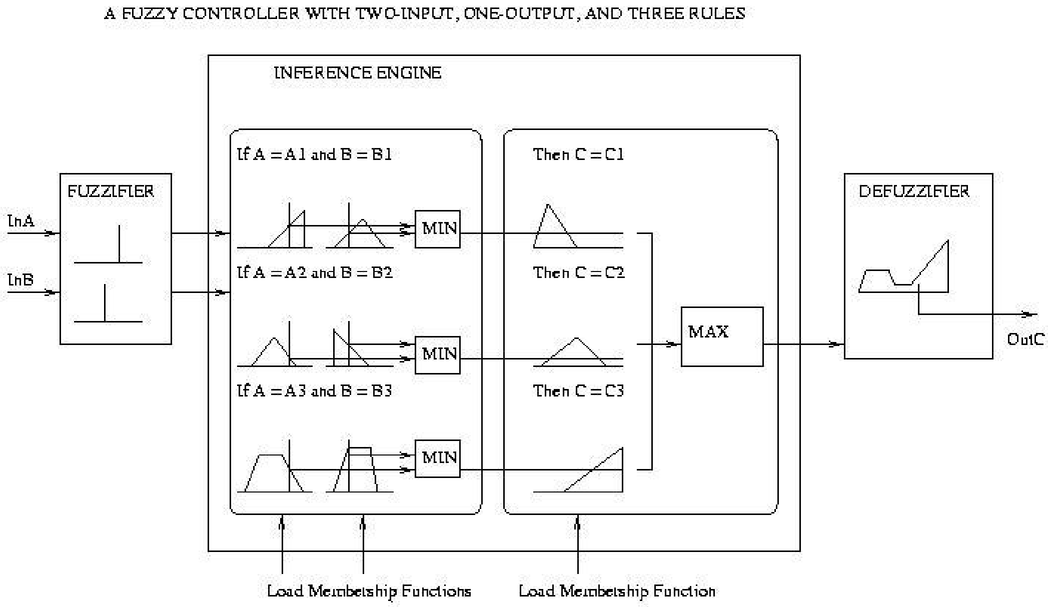

Figure 3.9: Block diagram for a FIS

Takagi-Sugeno fuzzy models [12] Takagi-Sugeno FISs differ from the aforementioned FISs, mainly because their consequents are crisp[13]. This implies that these FISs do not use the defuzzification methods, as described in the previous paragraphs. The Takagi-Sugeno FIS is called a first-order model if the consequent is a first-order polynomial, if the consequent is a constant it is called a zero-order model. In the last case the consequent can be seen as a fuzzy singleton. Figure A.2[14], in Appendix A, shows a zero-order Takagi-Sugeno fuzzy model. The overall output of a Takagi-Sugeno fuzzy model is obtained by a weighted average or weighted sum procedure, which are mathematically expressed as follows;

$$z = \frac{\sum_{i=1}^{n} w_i z_i}{\sum_{i=1}^{n} w_i} \qquad \text{and} \tag{3.15}$$

$$z = \sum_{i=1}^{n} w_i z_i, \tag{3.16}$$

where z_i is the consequent for fuzzy rule i, which can be a constant or a first-order polynomial, and w_i is the firing strength of each rule.

[12] Takagi and Sugeno (1985)

[13] A fuzzy rule with one antecedent in a first-order Takagi-Sugeno fuzzy model: "if x is A then $y = f(x)$"

[14] Source of information: www.mathworks.com

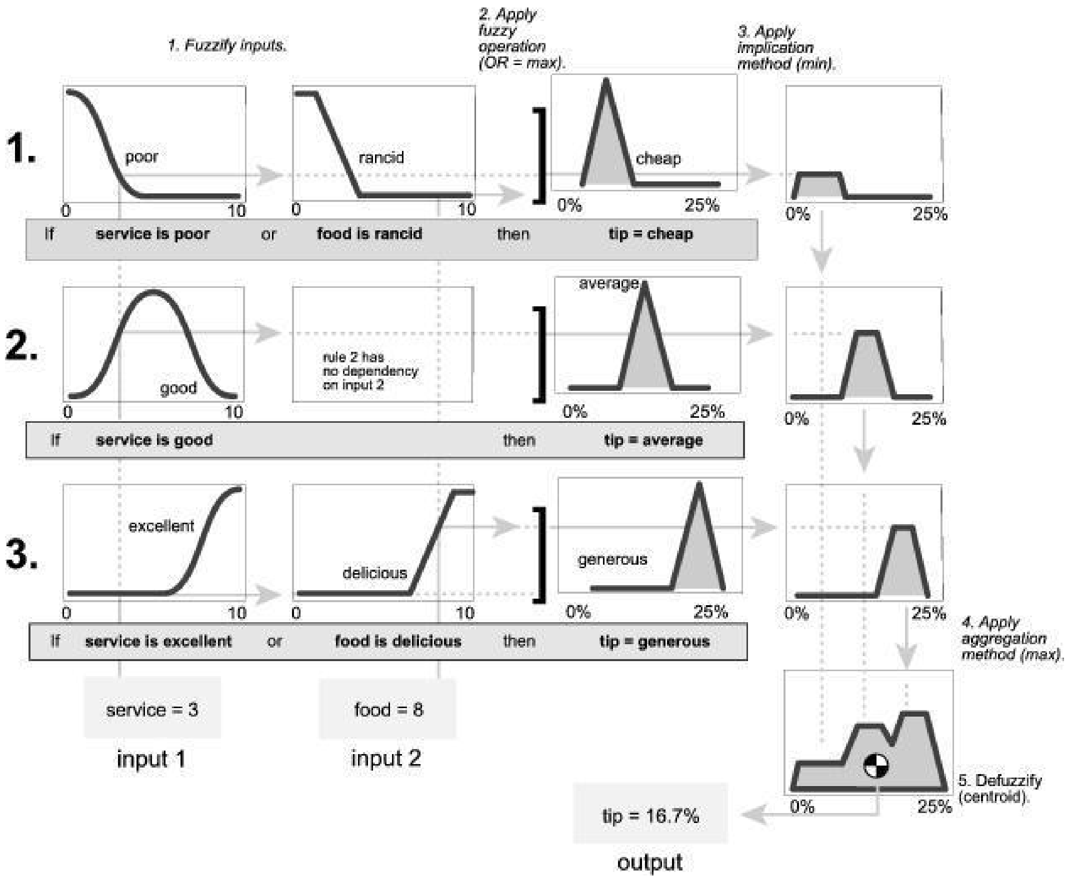

Figure 3.10: Three-rule Mamdani FIS deriving an overall output from two crisp inputs service and food

Probabilistic fuzzy systems

So far the basics on fuzzy systems have been discussed, and we have showed that fuzzy systems are powerful by the fact that they are able to cope with linguistic uncertainty. Different fuzzy systems, like the Mamdani and the Takagi-Sugeno fuzzy inference systems, are exemplified in the previous subsections. These FISs are proficient to return a single output value, after have been feeded with a single input value, even though some FIS make use of fuzzy consequents.

In some cases it is valuable to get a grip on the degree or probability to which an input value belongs to each of the possible output value. In these cases an expansion to the FISs seen before can offer an effective solution. On top of that many researches, outside the fuzzy community, criticized the fuzzy systems approach[15]. According to Kaymak and van den Berg (2003a), most of these researchers commented on the fact that they think that function approximation by fuzzy systems are heuristic driven and do not have any relation to the probabilistic nature of uncertainty.

Probabilistic fuzzy systems adapt to the aforementioned problems, in consequence

[15] This paragraph is based on the following articles; Kaymak and van den Berg (2003a), Kaymak and van den Berg (2003b), Waltman et al. (2005) and Waltman (2005)

they can be applied to problems with both linguistical and probabilistic uncertainty. Therefore PFSs, in comparison with ordinary fuzzy systems, do not only return a single output value, but return a probability distribution over all possible output values. In this report we concentrate PFSs appropriate for classifier problems, which are based on the principles of the Takagi-Sugeno FISs.

Consider the bond rating classification (BRC) problem, a problem with more than one output class. First of all we are interested in a system which is interpretable, it has to be easy to read, even for non-experts. The linguistic uncertainty aspect of fuzzy systems go along with this restriction. Secondly the classifier needs to cope with the probabilistic uncertainty, which is intercepted by the probabilistic concepts of the, subsequently proposed, PFSs. Both of these two concepts are of relevance to the BRC problem. The indicators on the financial health condition of the examined companies ought to be processed in the model, and have to be easy to read. For example, "if debt-to-equity is low and profit-to-assets is high then bond rating is X". On the other hand, the relation between the input variables, the financial ratios, and the output value, the bond rating class, can be stochastic. Above all, every data set possesses incorrect values, as a result of which the correctness of the single output of the classifier cannot be guaranteed. Because of these reasons, it may be required to receive more than one output value. Instead, it would be better to receive as many outputs as, in advance stipulated, classes with a probability to each of these classes. If we look at the example for the BRC problem again, we could say; "if debt-to-equity is low and profit-to-assets is high then bond rating is X_1 with probability y and X_2 with probability $1-y$", if there are only two possible output classes. In short, PFSs combine two types of uncertainty, namely linguistic uncertainty and probabilistic uncertainty. The mathematical expression of these PFSs will be described in the following paragraphs.

In this paragraph the mathematical background to PFSs will be highlighted[16], the technique which will have central role in this report. Assume that we try to determine the class $y \in \{C_1, ..., C_n\}$ for each data point $\mathbf{x} = (x_1, ..., x_n) \in X$. The following general form of the PFS with its rules is suitable to perform this task,

$$\begin{aligned} \text{If } \mathbf{x} \text{ is } A_j \text{ then } \underline{y} &= C_1 \text{ with probability } p_{j,1} \text{ and} \\ &\cdots\cdots\cdots\cdots\cdots \\ \underline{y} &= C_c \text{ with probability } p_{j,c}, \end{aligned} \tag{3.17}$$

where $A_j(j = 1, ..., a)$ are defined by fuzzy sets, which are defined in the d-dimensional input space X, $\underline{y}$ is the stochastic consequent variable equal to one of the values y. The probability parameters must obey to the following to restriction

$$p_{j,k} \geq 0 \qquad \text{for } j = 1, \ldots, a \text{ and } k = 1, \ldots, c \tag{3.18}$$

and to

$$\sum_{k=1}^{c} p_{j,k} = 1 \qquad \text{for } j = 1, \ldots, a. \tag{3.19}$$

Let $\mu_{A_j}(\mathbf{x})$ define the membership function of the fuzzy set A_j. The conditional probability distribution of $\underline{y}$ given by $\mathbf{x}$, is provided by the PFS with the rules following from

[16] This paragraph is based on Waltman et al. (2005)

3.17. The estimate $\hat{p}(C_k|\mathbf{x})$ of the conditional probability is obtained by the following equation

$$\hat{p}(C_k|\mathbf{x}) = \sum_{j=1}^{a} \bar{\mu}_{A_j}(\mathbf{x})\, p_{j,k}. \tag{3.20}$$

where $\bar{\mu}_{A_j}(\mathbf{x})$ defines the normalized MF given by

$$\bar{\mu}_{A_j}(\mathbf{x}) = \frac{\mu_{A_j}(\mathbf{x})}{\sum_{j'=1}^{a} \mu_{A_{j'}}(\mathbf{x})}. \tag{3.21}$$

The normalized membership function 3.21 determines the degree of fulfillment of the probabilistic fuzzy rules, given $\mathbf{x}$. In order to classify each data point $\mathbf{x}$, we use the estimated conditional probability function $\hat{p}(\underline{y}|\mathbf{x})$. Rule 3.22 minimizes the probability of misclassification.

$$\hat{y} = \underset{y \in \{C_1,\ldots,C_c\}}{\operatorname{argmax}} \hat{p}(y|\mathrm{x}). \tag{3.22}$$

Sequential parameter estimation of PFSs consist of two phases. The antecedent membership functions are determined during the first phase, the probability parameters of the rule consequents are defined under phase two. In this report, we have chosen to apply simultaneous parameter estimation proposed by Waltman et al. (2005). Before exemplifying this methodology, we will have a look at the sequential parameter estimation methodology.

In the traditional approach, the parameters are divided into two disjoint sets, which are estimated separately assuming in phase two the fact that the parameters in the first set are constant. As expected, the antecedent parameters form set one and set two consists of the probability parameters for the rule consequents. Generally, the antecedent parameters are estimated with an unsupervised learning technique or defined by expert knowledge. On the other hand, the probability parameters for the rule consequents are estimated by some sort of statistical formula.

In this report we use Gaussian membership functions for each dimension of the input space X. The sum of these Gaussian membership functions can mathematically be expressed as

$$\mu_{A_j}(\mathbf{x}) = \exp\left(-\sum_{l=1}^{d} \frac{(x_l - c_{jl})^2}{\sigma_{jl}^2} \right) \tag{3.23}$$

The center and the width of the MF in each dimension of the input space need to be estimated, which are expressed as a vector $\mathbf{c}_j = \{c_{j1}, \ldots, c_{jd}\}$ and by the vector $\mathbf{\Sigma}_j = \{\sigma_{j1}, \ldots, \sigma_{jd}\}$. Next to the estimation of vectors $\mathbf{c}_j$ and $\mathbf{\Sigma}_j$, we have to calculate the probability parameters p_{jk}, satisfying to restrictions 3.18 and 3.19.

The following items show how the parameters are estimated by the aforementioned 'sequential' method.

Antecedent parameters As mentioned before, antecedent parameters estimation by the traditional methodology is accomplished by unsupervised learning or by expert knowledge. Estimating vectors $\mathbf{c}_j$ and $\mathbf{\Sigma}_j$ is the target of this first phase. An applicable unsupervised learning technique is fuzzy c-means clustering. This technique is able to estimate vector $\mathbf{c}_j$, by handing in the normalized data points, see function 3.24, and a predefined number of cluster centers.

$$\bar{x}_{il} = \frac{x_{il}-\mu_l}{\sigma_l}, \qquad \begin{aligned} &\text{where } l = && \text{an input variable and} \\ &\mathbf{x}_i(i = 1,\ldots,n) = && \text{is the vector of data points} \end{aligned} \tag{3.24}$$

The widths of the MFs, $\mathbf{\Sigma}_j$, are left for estimation. Nearest neighbor heuristic is used in this paper, for the estimation of these vectors. In this report, function 3.25 is utilized for the $\mathbf{\Sigma}_j$ vectors estimation.

$$\sigma_{jl} = \min_{j' \neq j} \parallel \mathbf{c}_j - \mathbf{c}_{j'} \parallel, \qquad \text{for } l = 1,\ldots,d \tag{3.25}$$

where $\parallel \mathbf{c}_j - \mathbf{c}_{j'} \parallel$ is the Euclidian distance between $\mathbf{c}_j$ and $\mathbf{c}_{j'}$.

Probability parameters So far, vectors $\mathbf{c}_j$ and $\mathbf{\Sigma}_j$ are determined, which leaves the estimation of the probability parameters for the rule consequents. Kaymak and van den Berg (2003a) opted [17] to set the p_{jk} equal to the conditional probabilities which results in

$$p_{jk} = \frac{\sum_{i=1}^{n} \bar{\mu}_{A_j}(\mathbf{x}_i)\chi_{C_k}(y_i)}{\sum_{i=1}^{n} \bar{\mu}_{A_j}} \tag{3.26}$$

Function $\chi C_k(y)$ equals 1 if $y = C_k$ and equals 0 $y \neq C_k$. In this way, both sets of parameters are defined for the 'traditional' estimation method.

The 'sequential' parameter estimation, as described, cannot be seen as the optimal methodology, due to the fact that this technique does not estimate both sets of parameters simultaneously. On top of that, this method uses unsupervised learning for the antecedent parameters, as a results of which it does not take class labels into account. This may effect the performance of the PFS negatively. Moreover, function 3.26 does not maximize the probability of observing the data set available for parameter estimation, because it does not implement maximum likelihood (ML) estimates.

This paragraph describes the ML method for parameter estimation in a PFS, introduced by Waltman (2005) and implemented in this report. This method does implement ML and estimates both antecedent parameters, $\mathbf{c}_j$ and $\mathbf{\Sigma}_j$, and the probability parameters simultaneously. If the examples in a data set are independent of each other, the likelihood of a data set is given by

$$L = \prod_{i=1}^{n} \hat{p}(y_i|\mathbf{x}_i) \tag{3.27}$$

Minimization of the minus log-likelihood is equivalent to the maximization of function 3.27, expressed as

$$E = -\sum_{i=1}^{n} ln\hat{p}(y_i|\mathbf{x}_i) \tag{3.28}$$

Determining the $\mathbf{c}_j$, $\mathbf{\Sigma}_j$ and p_{jk} that maximizes function 3.27 or minimizes function 3.28 is a constraint optimization problem, as p_{jk} is restricted to the constraints 3.18 and 3.19. This problem can be transformed into an unconstrained optimization problem by

[17] In Kaymak and van den Berg (2003a) and Kaymak and van den Berg (2003b)

using auxiliary variables $u_{j,k}(j = 1, \ldots, a$ and $k = 1, \ldots, c)$. The following function describes the relation between the auxiliary variables and the probability parameters

$$p_{j,k} = \frac{e^{u_{j,k}}}{\sum_{k'=1}^{c} e^{u_{j,k'}}} \tag{3.29}$$

In this report we do not optimize the auxiliary variables $u_{j,c}(j = 1, \ldots, a)$, which are given a fixed value of 0, resulting in an unconstrained minimization problem which determines $\mathbf{c}_j$, $\boldsymbol{\Sigma}_j$ and p_{jk}.

A gradient descent algorithm is used for the minimization of the error function 3.28, for the initial values of the antecedent parameters we use fuzzy c-means clustering.

In order to measure the performance of the PFSs, two error functions can be implemented. The first error function is expressed as

$$E_1 = \frac{n_{errors}}{n} \tag{3.30}$$

This is a common used error function, where n_{errors} is the number of test examples which are misclassified by the PFS and n is the total number of test examples. The second error function in fact is a normalized function of 3.28 and is given by

$$E_2 = -\frac{1}{n} \sum_{i=1}^{n} ln\hat{p}(y_i|\mathbf{x}_i) \tag{3.31}$$

Error function 3.31 is more appropriate for performance measurement of PFSs[18], because of the following two reasons:

1. It is of interest to know the degree of confidence which the PFS attaches to every classification. This error function is able to evaluate the accuracy of the confidence measure.

2. The relation between the input variables and the output classes may be stochastic, which can be caused by lack of relevant information at the input side. Which indicates that is impossible to construct a PFS model that always produces correct answers.

3.2.2 Strengths and weaknesses of Probabilistic Fuzzy Systems

The majority of the strengths of the fuzzy systems (FSs) and PFSs have been illustrated in the previous sections. The two most important strengths of PFSs, as comprehensively described in the previous section, are the capability to cope with linguistic uncertainty as well as with probabilistic uncertainty. The ability to cope with linguistic uncertainty was seen before in ordinary FSs, and showed relatively large successes. Secondly, a PFS does not return a single output value, but it returns a probability distribution over all possible output values.

Of course, (P)FSs provoke some objections in the non-fuzzy professional community. According to Kaymak and van den Berg (2003a), researchers decry the fact that

[18] For further explanation see Waltman (2005) Chapter 5

they think FSs are heuristic driven and not related to the probabilistic nature of uncertainty. PFSs responses to these criticisms, as PFSs take the probabilistic nature of uncertainty into account.

The PFS method proposed by Waltman et al. (2005), which is applied in this report, does not utilize grid partitioning of the input space. This results in a system that searches for an classifier which optimizes the prediction accuracy. A possible loss in interpretability is a disadvantage of this type of system, as the system defines the membership functions itself. This type of optimizing could cause overlap between two or more membership functions for the same variable, through which these membership functions become inseparable.

Another point of attention is the input space dimensionality. The number of input variables should be small, if the goal of a (P)FS application is to construct an easily interpretable prediction system. Adding an extra input variable can improve the prediction accuracy and can at the same time reduce the interpretability power of the system.

3.2.3 Probabilistic Fuzzy Systems application for Bond Rating

So far, no one examined an application of PFS for the BRC problem, most likely because the PFS technique is relatively new. Due to the novelty of PFS, this application of PFS for BRC will be totally new, and as a result, all the more interesting. On the other hand we cannot compare our model and results with an existing PFS application, that's why we choose to benchmark the constructed PFS model with an application of the common used MDA technique.

In Chapter 4 the construction of a PFS model, for BRC, will be clarified. Besides that all encountered problems will be described in this Chapter. First of all the data set, the financial statements and the the financial ratios to be used will be clarified carefully. This is crucial for the construction of this prediction model. Secondly, the construction of both models, the MDA and the PFS model, will get attention. In this section the complete preparation of both these models will be done for this economical classification problem.

Chapter 4

Experimental Setup

The experimental setup of this report will be outlined in this Chapter. This Chapter starts with the financial setup in section 4.1. This section consists of three subsections, in which the data set and financial ratios, derived from this data set, will be explained. The architecture of the models will be sketched in section 4.2.

4.1 Financial Setup

This section pays attention to the data collection and to the derivation of the financial ratios. Before we go deeper into these financial ratios, the balance sheet and in the profits and losses accounts (P&L), which are used for data representation, will be illustrated.

4.1.1 Data set

The data set used within this report consists of a selection of 161 insurance companies, located in the United States of America, Europe and South Africa. All of these companies are rated by Standard and Poor's with a rating of at least B. The dataset, and the underlying components for each of the, in subsection 4.1.3 described, financial ratios were downloaded from the Thomson One Banker web-site. The associated bond ratings have been found on the official web-site of Standard and Poor's (S & P's).

For a start 304 insurance companies located in one of the aforementioned continents were selected through Thomson One Banker. After gathering these companies, the accompanying ratings had to be searched for on the S & P's web-site. Unfortunately only 174 of the selected 304 companies were rated by S & P's, as a result of which the data set decreased in size. Since in some cases the ratings were below the domain of 'AAA' to 'B' or because of the fact that selected insurance companies simply had to many missing values in the financial information, only 161 of the selected insurance companies could be used for this report. Except for some companies, most of them were provided with all the financial information needed. In order to realize a complete and reliable data set, we searched for the missing financial information in the annual reports of the companies. In most cases this was sufficient, in some cases we searched for broker reports or we thought up values for the missing values ourselves.

The motivation for this data set is based on a number of reasons and restrictions. First of all the companies had to be rated by the same rating companies. Furthermore, the companies in the data set had to belong to the same industry group. Besides that, the financial information on the companies had to be public, otherwise it just would have been to hard to get all the relevant financial information about these companies. Without these restrictions this research could not produce reliable results.

In this data set, the problem is to classify each example into one of the top-six rating classes defined by Standard and Poor's, where it is presumed that, for example, companies rated by AA+, AA and AA− are all rated by rating class AA. The used ratings all are organization ratings. Unfortunately, although expected in advance, the ratings in the data set are not homogeneously distributed. The largest classes are A and BBB, only a few companies are rated by classes AAA and B, respectively 5 and 6. Histogram 4.1 illustrates the top 6 rating classes, which are absorbed into the insurance companies data set. Histogram B.1, in Appendix B, visualizes the complete density of the used bond ratings, histogram B.2 visualizes the countries of origin of the companies in the data set. For a complete overview of the in the data set absorbed insurance companies, see tables B.1, B.2 and B.3 in Appendix B.

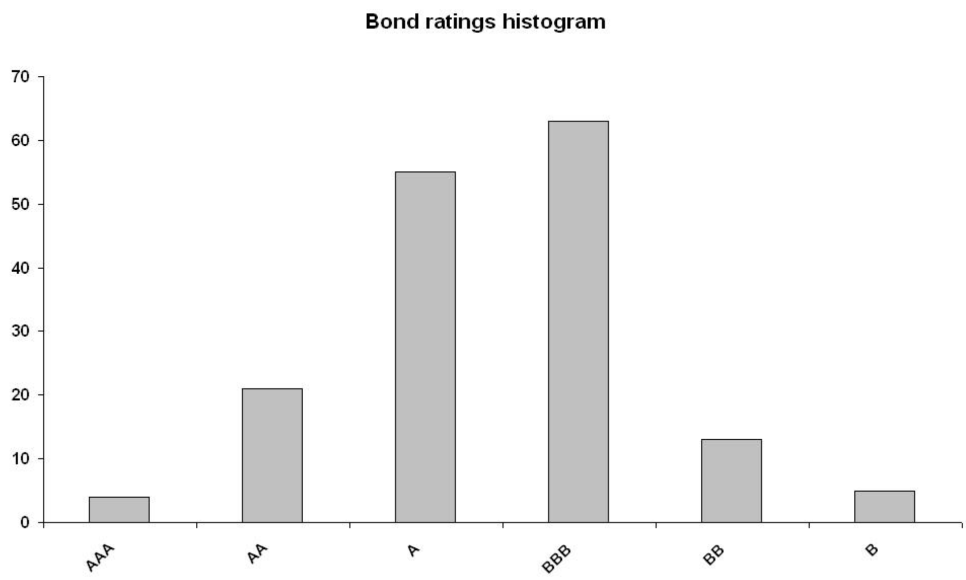

Figure 4.1: Bond rating histogram of the insurance companies data set.

4.1.2 Financial Statement Analysis

The financial statement of a firm encapsulates the balance sheet and the profits and losses account[1], which must obey to strict accounting regulations. These accounting regulations vary enormously among different countries, which matters in the scope of

[1] See tables C.1 and C.2 in Appendix C for a representation of a standardized balance sheet and profits & losses account

this report because the companies absorbed into the data set originate from different countries. Besides these variations, financial statements of different industry sectors can diverge as well. Especially the financial statement of industrials and of financial companies differ, among other things like the fact that industrials have a lot of inventory on their balance where financial companies have not. The financial information on the insurance companies used in this report is downloaded from Thomson One Banker, an information data-bank that normalizes the information they provide, which indicates the fact that our information is ready for use.

4.1.3 Financial variables

The financial performance of a company can be measured by looking at several financial ratios. According to Beaver (1966), "a financial ratio is a quotient of two numbers, where both numbers consist of financial statement items." These financial ratios can be allocated to five groups, including liquidity, profitability, leverage, solvency and activity. Although financial ratios provide a fast and easy way to compare different companies, some caution has to be taken whenever using them. As described before, financial statements of companies can differ enormously, not only because of industry differences but also because of different notations in the financial statements.

For the scope of this report, we did an extensive literature study on financial variables and ratios, which all have explanation power in BRC models. In this way we ground the choice for the selected financial variables and ratios used in this report. Table 4.1 shows eight studies on BRC models which followed from this literature study, which all utilized financial variables and ratios as inputs. Except for the study done by Belkaoui (1983), the focus of the displayed studies was based on AI models for BRC problems. The left column represents 20 of the modal variables and ratios, which actually could be composed by means of the available information at the data providers[2]. By virtue of this literature study, and the information found at the data providers, we managed to create a data set containing 14 financial variables and ratios. These variables and ratios all are respectable present in the data set, in other words, these underlying values of these financial ratios are relatively easy provided by Thomson One Banker. Table 4.2 reports the selected variables and ratios, on the base of which our BRC models will be constructed and tested. Subsection 4.1.4 will illustrate how the classification and regression trees (CART) algorithm is applied, in order to point out the input variables which are of most importance in determining the dependent variable to be explained.

[2] Thomson One Banker and Standard & Poor's

Variables	Study							
	Belkaoui (1983)	Tan (2000)	Shin and Han (2001)	Dutta and Shekhar (1988)	Kim et al. (1993)	Chaveesuk et al. (1997)	Kaplan and Urwitz (1979)	Huang et al. (2004)
Total assets	●	●	●		●	●	●	●
Total debt	●				●	●		●
LT Debt / total invested capital	●				●	●		●
ST Debt / total invested capital	●				●	●		
EBIT / interest expenses		●						●
EBIT / total debt		●						
Debt ratio		●		●				●
Debt-to-equity ratio		●						
Net gearing		●						
Return on equity		●						●
Return on total assets		●						●
Operating income / sales		●						
Net profit margin		●		●		●		●
Stockholders' equity			●					
Sales			●					
Total liabilities to assets			●			●		
Liabilities / (cash + assets)				●				
Sales / Net worth				●				
Fixed charge coverage ratio						●		
LT Debt / total assets							●	

Table 4.1: Total selection of financial variables and financial ratios examined in order to select the used financial ratios

#	Independent variable	#	Independent variable
X_1	Total assets	X_8	Operating income / sales
X_2	Total debt	X_9	Net profit margin
X_3	Long-term debt /total invested capital	X_{10}	Sales
X_4	Debt ratio	X_{11}	Total liabilities to assets
X_5	Debt-to-equity ratio	X_{12}	Liabilities / (cash + assets)
X_6	Net gearing	X_{13}	Sales / net worth
X_7	Return on equity	X_{14}	Long term debt / total assets

Table 4.2: Fourteen selected financial input variables and ratios. (Total invested capital = long-term debt + preferred stock + common equity; Debt ratio = long-term debt / (long-term debt + equity + minority interest); Net gearing = (total liabilities - cash) / equity); Net worth = total assets - total liabilities

4.1.4 Variable selection: classification and regression trees

The previous subsection ventilated about the accomplished literature study on variable[3] selection. Table 4.1 showed a number of input variables which are significant for BRC models. From these tables, we selected fourteen variables, as illustrated in table 4.2. As mentioned in Chapter 3, the interpretability power of a PFS decreases by increasing the number of input variables. For that reason, it is important to see whether variable reduction is possible and sensible as well. Whether variable selection is sensible indeed, will be answered in Chapter 5.

There are many ways to reduce the amount of input variables. Some of these ways are based on statistical models others on AI techniques. Trying all possible combinations of input variables is another proficient way to select input variables which are of importance in explaining the dependent variable. Unfortunately this method is time consuming. For example if we look at our BRC problem, where we now have 14 input variables, but we may want to reduce this amount to 5 input variables. What happens if we try to select 5 variables if we apply this last method is the following

$$\frac{14!}{5! * 9!}, \tag{4.1}$$

which results in 2002 possible combinations. It is to expensive to run the model for 2002 times, in order to point out which combination produces the highest accuracy. For that reason we have selected an alternative AI technique which is commonly used for variable reduction.

Classification and regression trees (CART) algorithm introduced by Breiman et al. (1984) will point out our input variables, which are of importance in explaining the dependent variable. Decision trees constructed by the CART algorithm are represented by a set of questions which split a data set into smaller and smaller parts. As a result decision trees partition a data set into mutually exclusive regions. When decision trees are used for classification problems, these trees are called classification trees. If it concerns a regression problem then they are often called regression trees. CART is a technique which combines both of them. The CART algorithm only asks yes or no

[3]From now on financial input variables and ratios will be denominated as input variables.

questions like: ”Are the *total assets* $> x$?” The algorithm searches for all possible input variables in order to find the best split.

CART can function directly as a model, or can be used for structure identification for other techniques. Identify input variables, which are of most importance in explaining the dependent variable, is an important strength of CART[4]. For this reason CART is used within this report.

The objective of variable reduction is to secure the input space dimensionality at a certain level. Figures E.1, E.2 and E.3 show the structure of the pruned and unpruned trees for the insurance companies data set. By means of these trees we have selected two sets of input variables which could produce accurate BRC models, namely set $\{X_1, X_2, X_3, X_8, X_{10}, X_{11}, X_{14}\}$ and set $\{X_1, X_2, X_3, X_8, X_{14}\}$. Chapter 5 will expel whether these reduced input sets show high prediction accuracies on the test sets.

4.2 Architecture of the models

This section encapsulates a comprehensive explanation on the construction procedure and on the accuracy testing methods used for both models. The construction and testing procedures of the models are accomplished in MATLAB 6.5, mainly because MATLAB is a convenient tool to construct the models used for this report. Subsection 4.2.1 starts with the architecture of a MDA model, which is partly based on the research done by Belkaoui (1983). Subsection 4.2.2 follows with an overview of the architecture of the PFS model constructed within this report.

4.2.1 Multiple discriminant analysis

As mentioned before MDA will form a benchmark model within this report. To construct this MDA model based on the variables discussed in subsection 4.1.3 the following action plan is useful.

1. **Find the discriminant coefficients** They are determined by utilizing equations 3.10 and 3.11 as stated in subsection 3.1.1. Before identification of the discriminant coefficients we need to calculate the within-group sum of squares matrix and the across-group sum of squares matrix.

2. **Determine the accuracy** In this step the overall accuracy, or hit ratio, of the model will be be calculated. This is the fraction between the total number of correctly classified observations and the total number of observations.

3. **Create an actual accuracy matrix** Figure 3.3 shows the accuracy matrix based on the relative frequency with which each group appears in the data. The actual accuracy matrix replaces the values with the predicted values. This actual accuracy matrix illustrates the classification of the observations perfectly.

4. **Test the accuracy** The accuracy, or hit ratio, is a simple and intuitive approach to measure the goodness of fit for the classification model. Though, the question is; “is this accuracy percentage reasonable?” For this reason the accuracy should

[4]For further technical details on the CART algorithm see Breiman et al. (1984), or Sprengers (2005) which describes an application of the CART algorithm for bankruptcy prediction

be compared to some sort of benchmark. To test the accuracy of the model, we implement the frequently used benchmark method, or in other words the proportional chance criterion. This method creates an expected accuracy matrix 3.3 based on the relative frequency with which each group appears in the data. It is possible to test the accuracy, with this expected accuracy matrix. Functions 3.13, 3.8 and 3.9 lead to the *t*-test which is able to assess the accuracy for significance.

This aforementioned action plan will be accomplished according to the rules of leave-one-out cross-validation, a train and test technique which squeezes a data set optimally and produces reliable test result[5].

The results for the MDA prediction model will be described extensively in Chapter 5. Besides that the encountered problems will be tagged, and the possible solutions for these problems will be expounded.

4.2.2 Probabilistic fuzzy systems

The aforementioned procedure will partly be applicable for the PFS model. Nevertheless there are some differences, like the way of constructing the system, which differs from the way of constructing the MDA model. The data set, and of course the underlying input variables and output classes, will be exactly the same for both applications, in order to compare both accuracies in a proper way.

Maximum likelihood (ML) is utilized for parameter estimation, as discussed in subsection 3.2.1 and introduced by Waltman (2005). In that way, both the antecedent parameters and the probability parameters $p_{j,k}$ for the rule consequents are estimated simultaneously. Furthermore, in this report, both the antecedent parameters and the probability parameters are estimated by the ML methodology.

With the classifications of the individual observations we can determine the accuracies and accuracy matrices. Besides that the accuracy will be tested twice, we firstly test the constructed PFS by implementing functions 3.8 and 3.9, which is the same way of testing the MDA model. Secondly, the PFS model will be exposed to function 3.31, as described and clarified in subsection 3.2.1. This function gives a more valuable prediction accuracy than the normal error/accuracy functions used for MDA. Because this function is not applicable to MDA models, unfortunately we cannot compare these results with our through MDA constructed model.

The realization of the PFS model and the test results will be described in the next Chapter. The problems encountered will be mentioned and solved whenever possible.

[5] Section D.1 in Appendix D gives a moderate overview of the different training versus testing techniques.

Chapter 5

Results and analysis

The results of the experiments, as described in the previous chapters, are reported and discussed in this chapter. First of all the intention of the experiments will be depicted in section 5.1. Section 5.2 describes the procedure and results for the multiple discriminant analysis. The composition of section 5.3 is for the greater part the same as for section 5.2, although this section goes more into details about the choices which have been made concerning variable and rule selections. Finally, this chapter will be concluded with a short comparison and summary of the models constructed by both techniques.

5.1 Setup of the models

A general scheme to produce and test our BRC models was handed over in the previous chapter. The exact setup of the models will be sketched in this paragraph.

The results of the different models are reviewed in the following sections. Before the models can produce these results, they have to be trained and tested by a data set. The data set prepared for the scope of this report is discussed in Chapter 4. This data set consists of 161 insurance companies, which are all rated by S & P's with a company rating of at least 'B'. Fourteen proper independent variables were selected, based on quantitative financial information. Variables which seem to be of most importance in explaining the dependent variable were selected by classification and regression trees, in order to reduce the input space complexity. This resulted in three different data sets. Table 5.1 illustrates the independent variables used in each of the data sets.(From now one respectively denoted by: set 1, set 2 and set 3) We will expose the models to these three different data sets, in order to check wether these variable reductions have positive either negative influence on the results.

Model training and testing is achieved by leave-one-out cross-validation, as described in appendix D. This means that every iteration the models are trained by $N-1$ data points and tested by 1 data point. Leave-one-out cross-validation is particularly appropriate given the relatively small size of our data set.

Different sets				
#	*Independent variable*	*set 1*	*set 2*	*set 3*
X_1	Total assets	•	•	•
X_2	Total debt	•	•	•
X_3	LT Debt / total invested capital	•	•	•
X_4	Debt ratio	•		
X_5	Debt-to-equity ratio	•		
X_6	Net gearing	•		
X_7	Return on equity	•		
X_8	Operating income / sales	•	•	•
X_9	Net profit margin	•		
X_{10}	Sales	•	•	
X_{11}	Total liabilities to assets	•	•	
X_{12}	Liabilities / (cash + assets)	•		
X_{13}	Sales / Net worth	•		
X_{14}	LT Debt / total assets	•	•	•

Table 5.1: Independent variables selected for the three sets of data.

5.2 Multiple discriminant analysis

As mentioned before, an application of MDA for bond rating classification will function as benchmark model within this report. Chapter 3 and 4 sketched the technical details and introduced the 'action plan' in order to construct and test the model.

The test results and analysis of the three different data sets will be reported successively in this section. Unfortunately we cannot implement both error measures 3.30 and 3.31 described in section 3.2.1, as MDA assigns examples to a single class with probability 1. In order to measure the performance of the constructed MDA models, we use a prediction accuracy measure instead of an error measure. This prediction accuracy, or 'actual hit ratio' is expressed as

$$hr_{actual} = 1 - E_1 = \frac{(n - n_{errors})}{n} \tag{5.1}$$

Because we cannot use error measure 2 as expressed in function 3.31, we are only allowed to compare the results of the actual hit ratio of the MDA model with the actual hit ratio of the PFS model. On the other hand, we can observe the accuracy matrix. This matrix provides an insight in the predicting capabilities of the model, and offers an alternative to compare models with each other.

The actual accuracy will be tested on significance by conducting a t-test of the difference between the classification performance of the MDA model versus the proportional chance criterion. We need to calculate this proportional chance criterion, before we can conduct this t-test. This proportional chance criterion is calculated with formula 3.13, or by dividing the sum of the diagonal of matrix B.3 with N.

Results

The MDA model constructed by set 2 showed the highest prediction accuracy. This MDA model produced an accuracy of 54.04 %. For each model we conducted a t-test of the difference between the actual classification performance and the proportional chance criterion. This t-test showed that the results, prediction accuracies, are significant for all three models at a significance level of 0.01.

Accuracy results for set 1, 2 and 3	
	accuracy
14 independent variables	0.5280
7 independent variables	0.5404
5 independent variables	0.5155

Table 5.2: Test results for set 1, 2 and 3 based on leave-one-out cross-validation.

Accuracy matrices

The accuracy matrices produced by our MDA models are presented below. The 'red' diagonals show the examples which are correctly classified. We can see that the models tend to classify most of the examples towards rating classes 'A' and 'BBB'. We are not amazed by this observation, given the fact that these rating classes are excessively present in the data sets.

	Predicted rating class						
Actual rating class	AAA	AA	A	BBB	BB	B	All
AAA	0	1	2	1	0	0	4
AA	0	5	10	4	2	0	21
A	1	4	31	18	1	0	55
BBB	2	1	10	48	2	0	63
BB	0	0	4	9	0	0	13
B	0	0	1	3	0	1	5
All	3	11	58	83	5	1	161

Table 5.3: Accuracy matrix of the MDA model based on 14 independent variables (Set 1)

Predicted rating class							
Actual rating class	AAA	AA	A	BBB	BB	B	All
AAA	0	1	1	2	0	0	4
AA	1	3	9	7	0	1	21
A	1	3	31	20	0	0	55
BBB	1	1	9	51	0	1	63
BB	0	0	4	9	0	0	13
B	0	0	1	2	0	2	5
All	3	8	55	91	0	4	161

Table 5.4: Accuracy matrix of the MDA model based on 7 independent variables (Set 2)

Predicted rating class							
Actual rating class	AAA	AA	A	BBB	BB	B	All
AAA	0	0	2	2	0	0	4
AA	0	4	11	5	0	1	21
A	0	5	24	26	0	0	55
BBB	0	1	8	53	0	1	63
BB	0	0	3	8	0	2	13
B	0	0	1	2	0	2	5
All	0	10	49	96	0	6	161

Table 5.5: Accuracy matrix of the MDA model based on 5 independent variables (Set 3)

5.3 Probabilistic fuzzy systems

The test results of the probabilistic fuzzy approaches to BRC are presented in this section. As mentioned before, the testing procedure of the constructed PFS applications is largely the same as for MDA. However, what differs is the fact that we can variate the number of rules for our PFSs. In this way we can create several PFSs with a different number of rules, for each of the three sets. The PFS models constructed for this report are based on 2 to 14 rules. On top of that we can calculate error measure 2, in order to evaluate the accuracy of the confidence measure.

We choose to start training and testing the models with a ten-fold cross-validation, considering that PFS model training and testing is computationally expensive. In this way, we are able to get a feeling for the prediction accuracies that we can expect after running the models with leave-one-out cross-validation.

The following subsection starts with the presentation of the results for the ten-fold cross-validation. A number of PFSs with high prediction accuracies will be selected to be exposed to leave-one-out cross-validation. The results of these leave-one-out cross-validations will be illustrated subsequently.

Results

It is hard to conclude from the following table which of the model structures works most accurate. The 'red' colored figures point to the best results, the 'green' figures to the worst. From these colored figures we could interpret that the PFS model based on 4 rules for data set 2 performs best, and generally speaking the PFSs with 6 rules or less perform better than the others. For that reason we have chosen to perform several leave-one-out cross-validations for the PFS models with 1 to 6 rules. We did not perform these tests for set 1, because of the poor results produced by set 1, and because of the fact that these tests are computationally expensive. Table 5.7 illustrates the results of these cross-validation tests. This table indicates that the PFS model with

Accuracy results for set 1, 2 and 3 from a ten-fold cross-validation						
	set 1		*set 2*		*set 3*	
# rules	accuracy	E_2	accuracy	E_2	accuracy	E_2
2	0.4915	1.4449	0.5103	1.3651	0.5228	1.3338
3	0.4603	1.5134	0.5099	1.3630	0.5165	1.3344
4	0.4978	1.4922	0.5349	1.3597	0.5349	1.3809
5	0.4294	1.6315	0.5158	1.3666	0.5158	1.3666
6	0.5037	1.6274	0.5029	1.4379	0.4974	1.4312
7	0.4478	1.6440	0.5029	1.3991	0.4912	1.4232
8	0.4794	1.5064	0.4724	1.5249	0.4665	1.4585
9	0.4603	1.5949	0.4974	1.5445	0.4853	1.4843
10	0.4232	1.6691	0.5037	1.4704	0.5048	1.4259
11	0.4735	1.6961	0.5033	1.5832	0.5169	1.4931
12	0.4232	1.6772	0.4978	1.602	0.5162	1.4854
13	0.4728	1.6641	0.5162	1.5682	0.4974	1.5249
14	0.4469	1.6804	0.4787	1.7162	0.4735	1.4965

Table 5.6: Results of the PFS models from a ten-fold cross-validation (This t-test showed that the prediction accuracies, are significant for all systems at a significance level of 0.01.)

the highest prediction accuracy equalizes the result of the MDA model discussed in section 5.2. This means that this PFS model, based on 6 rules and trained and tested by set 3, correctly classified 54.04% of the 161 insurance companies in the data set.

Accuracy results for set 2 and 3				
	set 2		*set 3*	
# of rules	accuracy	E_2	accuracy	E_2
2	0.5217	1.3623	0.5031	1.3330
3	0.5031	1.4071	0.5155	1.3492
4	0.5217	1.4430	0.5155	1.4069
5	0.4845	1.4580	0.5031	1.3587
6	0.5217	1.3543	0.5404	1.3890

Table 5.7: Results of the PFS models from a leave-one-out cross-validation

Accuracy matrix

The following table is the representation of the accuracy matrix of the PFS model based on 6 rules which is trained and tested by set 3. This table shows again the symptom seen with the accuracy matrices of the MDA models, namely the fact that the models tend to classify the examples in rating classes 'A' and 'BBB'. This tendency toward these classes is even more obvious for this PFS model than for the MDA models described before.

	Predicted rating class						
Actual rating class	AAA	AA	A	BBB	BB	B	All
AAA	0	0	4	0	0	0	4
AA	0	6	11	4	0	0	21
A	0	9	33	13	0	0	55
BBB	0	2	13	48	0	0	63
BB	0	0	3	10	0	0	13
B	0	1	2	2	0	0	5
All	0	18	66	77	0	0	161

Table 5.8: Accuracy matrix of the PFS model with 6 rules, based on 5 independent variables (Set 3)

Interpretability

Next to the prediction accuracies, we are interested in the interpretability of the constructed PFS models. As mentioned before, (P)FSs are able to produce interpretable prediction models. Though, in the first place we have chosen to maximize the prediction accuracy of our PFS models, by implementing maximum likelihood without grid partitioning. PFS model training without grid partitioning could bring about negative effects on the interpretability of the model. In order to get an insight into the PFS model described in the previous subsection, we have created plots of the membership functions produced by this PFS model. Figure 5.1 reflects the membership functions of variable X_1, for all 6 rules. We can conclude that the membership functions for rule 2, 4 and 5 are almost the same, what implicates that we can give one label to these membership functions. Other distinctive labels can be given to rules 3 and 6. On the other hand, the membership function of variable X_1 in rule 1 is a typical example of the fact that we constructed the PFS without grid partitioning. This phenomenon is inherent to this way of constructing a PFS, since the PFS model is simply searching for the membership functions that show the optimal prediction capabilities. Membership functions for variables X_2, X_3, X_8 and X_{14} are displayed in Appendix F, that also shows the representations of the 6 rules with the probabilities distributions for the consequents.

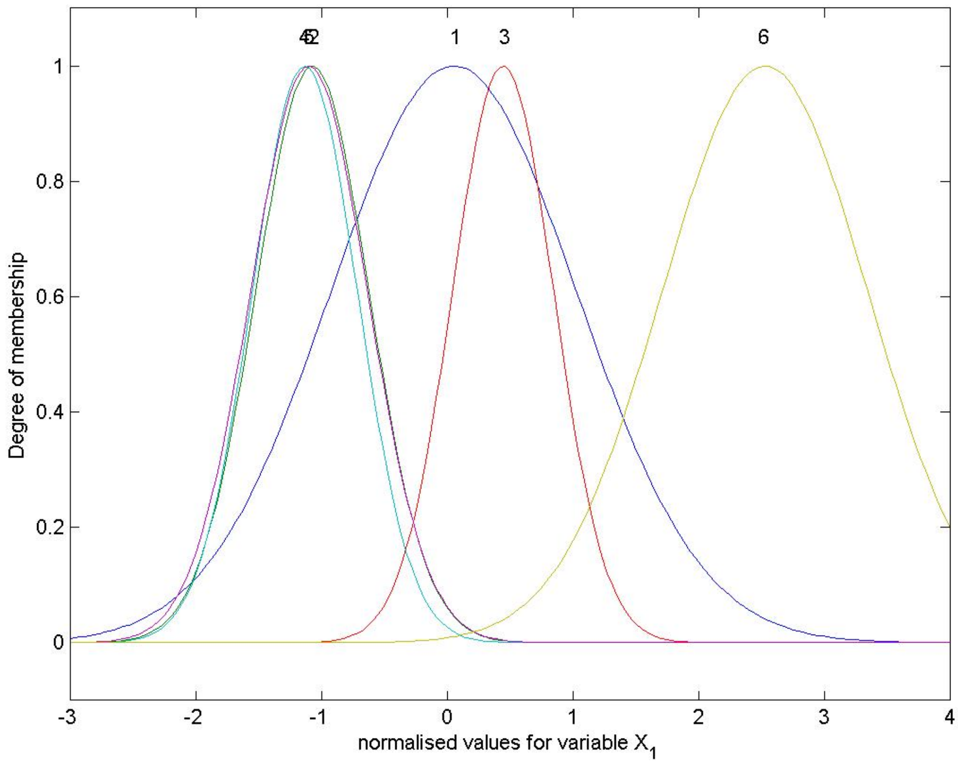

Figure 5.1: Membership functions of variable X_1 for rules $1 - 6$

By means of rule 5 and 6, see figures F.14 and F.16, we can argue that our PFS is interpretable. The probabilities for the consequents in both rules show opposite distributions just like the membership functions, which show opposite symptoms. The linguistic rules for rule 5 and 6 can be written down as follows:

Rule 5 ***If*** X_1(total assets) is low **and** X_2(total debt) is reasonably low **and** X_3(long-term debt / total invested capital) is medium **and** X_8(operating income/debt) reasonably high **and** X_{14}(long-term debt/total assets) is reasonably high ***then*** the rating class is

AAA *with probability* 0.0011396 and

AA *with probability* 0.0043962 and

A *with probability* 0.011836 and

BBB *with probability* 0.62509 and

BB *with probability* 0.35516 and

B *with probability* 0.0023709

Rule 6 ***If*** X_1(total assets) is high **and** X_2(total debt) is reasonably high **and** X_3(long-term debt / total invested capital) is medium **and** X_8(operating income/debt) low **and** X_{14}(long-term debt/total assets) is low ***then*** the rating class is

AAA *with probability* 0.0045722 and

AA *with probability* 0.56374 and

A *with probability* 0.34335 and

BBB *with probability* 0.016738 and

BB *with probability* 0.0048282 and

B *with probability* 0.066769

5.4 Discussion

The following table shows the difference in prediction accuracies for both models on the different data sets. The prediction accuracies equalize each other, as mentioned before. Although, there is a difference in the performance of the two classifiers. The PFS model was able to classify 54.04% correct by means of data set 3, which consisted of 5 independent variables instead of 7. Looking at tables 5.6 and 5.7 we could say that this is a coincidence.

Accuracy results for set 2 and 3				
	set 2		*set 3*	
Method	accuracy	E_2	accuracy	E_2
MDA	0.5404		0.5155	
PFS, 6 rules	0.5217	1.3543	0.5404	1.3890

Table 5.9: Results of the MDA and PFS models from a leave-one-out cross-validation

Furthermore we can look at accuracy matrices 5.4 and 5.8. If we compare them we see that MDA classifies the examples correct in 4 of the rating classes, where PFS did not surpass 3 rating classes in which it classified correct. Nevertheless, if we look at figure 5.2 which is a representation of the misclassified examples, PFS scores slightly better. The *x*-axis represents the number of ratings a misclassified example is away from the actual rating given by S & P's.

Compared to other studies seen before we can conclude that our MDA model performs considerable good. Huang et al. (2004) listed five studies on BRC which utilized MDA as the benchmark technique, which all tried to classify the examples into 5 or 6 rating classes. The prediction accuracy of these studies range between 36.20% and 62.00%.

Huang et al. (2004) also listed several studies on AI techniques for BRC. These studies also tried to classify the examples into 5 or 6 rating classes, and are for that reason comparable. On the other hand, the data sets used differ completely as well as the AI techniques used. This implies that we can only use these results as an indication of the possible prediction accuracies. Of these techniques, rule-based systems performed worst with only 31.03% prediction accuracy, where most of the neural networks models showed prediction accuracies between 55.17% and 72.50% for BRC

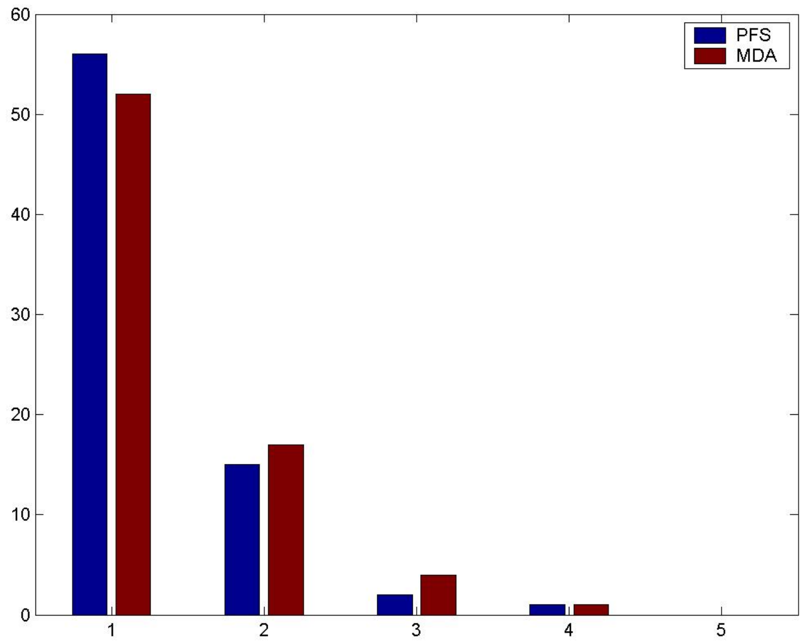

Figure 5.2: Misclassification histogram for the PFS model made by set 3 and the MDA model made by set 2.

with 5 or 6 rating classes. Most of the other mentioned AI techniques performed similar, except for a combination of case-based reasoning (CBR) with genetic algorithm constructed by Shin and Han (2001). This combination was able to produce 75.50% on a data set, where normal CBR showed 62.00% and ID3 showed $53 - 54\%$ on the same data set.

Unfortunately our PFS did not contribute as much as we hoped to the interpretability of a BRC model, due to the fact that we have chosen to maximize its prediction capabilities by not utilizing grid partitioning. Nevertheless, this probabilistic fuzzy system proved to be competitive classifier for our BRC problem.

Chapter 6

Conclusions and Future Research

6.1 Conclusions

In this report we applied the newly introduced AI technique "probabilistic fuzzy systems" for bond rating classification. The goal set was to create an accurate BRC model which was able to predict bond rating classes by virtue of financial quantitative data with a probabilistic fuzzy approach. The data set that had been used for this report consisted of 161 insurance companies from the United States of America, Europe and South Africa.

The PFS model was benchmarked with an application of multiple discriminant analysis. MDA is a statistical technique which is currently most commonly used for the bond rating classification problem ever since the publication of a study conducted by Belkaoui (1983). Nowadays this technique often functions as a benchmark method for other techniques. Huang et al. (2004) reported six studies that implemented MDA for bond rating classification, which showed prediction accuracies between 36.20% and 62.00%.

The objective of this report was to construct an accurate prediction model for BRC based on PFS. On top of that we strived for a model which was better interpretable than most of the other AI techniques described in previous studies.

Both the MDA and the PFS model showed 54.04 % prediction accuracy with leave-one-out cross-validation. Compared with studies conducted in the past, we can argue that the MDA model performed average. The PFS model was more difficult to compare, because this is the first study conducted on PFSs for the BRC problem. The results of the studies on AI techniques for BRC conducted previously can only give an indication of performances. Each technique and each data set differs, which results in the fact that our PFS was difficult to compare with one of the studies conducted previously.

As regards to the interpretability of the constructed PFS model, we have to mention that our PFS did not bring forward an easy interpretable model. This is most probably caused by the input space partitioning, which did not utilize grid partitioning of the input space. Optimizing the prediction accuracy was the main goal for this report, which has as a consequence that our PFS determined the partitioning of the input space itself. This resulted in membership functions for the different input variables which are hard to keep apart. On the other hand we can say the PFS model scored relatively good if we take the relatively small number of input variables into account.

6.2 Future Research

Several problems and points of attention came across during the writing of this report. Most of these issues are related to the interpretability of the PFS models or with the improvement of the prediction accuracies. The following topics suggest some areas for future research to address the aforementioned points of attention.

1. Reduction of the input space dimensionality is frequently seen as a point of interest for (P)FSs. In order to classify an example into the correct rating class, the PFS model needs to be fed with several independent variables. These variables variate from quantitative financial variables to qualitative managerial measurements. Due to this, usually many input variables are selected that are of importance in explaining the dependent variable. The actual amount of input variables can exceed the desired amount of variables, concerning the input space dimensionality. Reduction of this amount could be accomplished by one of the following suggestions:

    - *Hierarchical fuzzy systems*. A technique which counteracts the phenomena that the number of rules increase exponentially with the number of variables. Actually this technique constructs chaining rules, or multi-stage fuzzy systems (Torra (2002)). In this way the system is able group input variables into fewer input variables, with the result that the final rules are easier to interpret.
    - *Combining grid partitioning with genetic algorithms.* In the first place genetic algorithms could identify the most important input variables, which can be used by the (P)FSs. Subsequently, the grid partitionings share of the combination leads to membership functions for each of the variables that are easily separable.
    - *Expert knowledge*. An intense cooperation with several specialists on bonds could offer insight in the leading variables in explaining the dependent variable.

2. Unravel the bond rating models used by the rating companies like Standard & Poor's and Moody's. So far the rating procedures implemented by the official rating companies are surrounded with mystery. Perhaps an extensive research on these rating procedures can take away this mystery for the greater part. This could imply it will become easier to create a competitive bond rating classifier.

3. Joining membership functions. The membership functions produced by our PFS models showed overlap or similarity in many cases, as a consequence of the fact that we did not implement grid partitioning. According to Babuska et al. (1996), rule base simplification with similarity measures could contribute to the interpretability of (P)FS models which produced rules which showed similarity. In this way the membership functions that show high overlap are combined into one membership function. This can cause negative influence on the prediction

accuracy, but on the other hand this probably results into more effective linguistic descriptions.

As this reports shows, in the area of probabilistic fuzzy systems for bond rating classification several topics are still left unexplored which offer great opportunities for future research.

Bibliography

E.I. Altman. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 23(4):569–608, September 1968.

E.I. Altman. *Corporate Financial Distress and Bankruptcy: A Complete Guide to Predicting & Avoiding Distress and Profiting from Bankruptcy, Second Edition*. John Wiley & Sons, 1993.

J.S. Ang and K.A. Patel. Bond rating methods: Comparison and validation. *Journal of Finance*, 30(2):11, May 1975.

T.A. Atkinson and E.T. Simpson. Trends in corporate bond quality. Technical report, National Bureau of Economic Research, 1967.

R. Babuska, M. Setnes, U. Kaymak, and H.R. van Nauta Lemke. Rule base simplification with similarity measures. *Fuzzy systems*, 3, 1996.

W. Beaver. Financial Ratios as Predictors of Failures. In Empirical Research in Accounting, selected studies supplement to the. *Journal of Accounting research*, 1(4): 71–127, 1966.

A. Belkaoui. *Industrial Bonds and the Rating Process*, volume 1. Greemwood Press, 1983. ISBN 0-89930-046-4.

R.A. Brealey and S.C. Myers. *Principles of Corporate Finance*. McGraw-Hill, 2003. ISBN 0-07-246766-5.

L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone. *Classification and regression trees*. Chapman & Hall, Inc., 1984. ISBN 0-412-04841-8.

Ravipim Chaveesuk, Chat Srivaree-ratana, and Alice E. Smith. Alternative neural network approaches to corporate bond rating. *Journal of Engineering Valuation and Cost Analysis*, 1997.

S. Dutta and S Shekhar. Bond rating: A nonconservative approach of neural networks. *Proceedings of the IEEE International Conference on Neural Networks*, pages 443–450, July 1988.

F.J. Fabozzi and S.V. Mann. *The Handbook of Fixed Income Securities*. McGraw-Hill, 2005. ISBN 0-07-144099-2.

L. Fisher. Determinants of risk premium on corporate bonds. *Journal of Political Economy*, pages 217–237, 1959.

J.A. Gentry, D.T. Whitford, and P. Newbold. Predicting industrial bond ratings with a probit model and funds flow components. *The Financial Review*, 23(3), 1988.

R.C. Grinold and R.N. Kahn. *Active Portfolio Management: A quantitative approach for providing superior returns and controlling risk*, volume 2. McGraw-Hill, 1999. ISBN 0-07-024882-6.

G. Harold. *Bond rating and investment Guide*. New York: Ronald Press, 1938.

W.B. Hickman. *Corporate Bonds, Quality and Investment Performance*. Princeton University Press, 1958.

J.O. Horrigan. The determination of long-term credit standing with financial ratios. *Journal of Accounting Research*, Supplement to volume 4:44–62, 1966.

Z. Huang, H. Chen, C.-J. Hsu, W.-H. Chen, and S. Wu. Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision Support Systems*, 37:543–558, 2004.

Investopedia. Investopedia online financial dictionary. `http://www.investopedia.com/terms/b/bond.asp`, 2005.

J.-S.R. Jang, C.-T. Sun, and E. Mizutani. *Neuro-Fuzzy and Soft Computing*. Prentice Hall, 1997.

R.S. Kaplan and G. Urwitz. Statistical models of bond ratings: A methodological inquiry. *The Journal of Business*, 52(2):231–261, 1979.

U. Kaymak and J. van den Berg. On probabilistic connections of fuzzy systems. *Proceedings of the 15th Belgian-Dutch Conference on Artificial Intelligence*, pages 187–194, 2003a.

U. Kaymak and J. van den Berg. A fuzzy additive reasoning scheme for probabilistic mamdani fuzzy systems. *Proceedings of the 12th IEEE International Conference on Fuzzy Systems*, pages 331–336, 2003b.

J.W. Kim, W.H. Rolan, and R.T Redmond. Expert systems for bond rating: A comparative analysis of statistical, rule-based and neural network systems. *Expert Systems*, 10(3):167–171, 1993.

Y.S. Kwon, I.G. Han, and K.C. Lee. Ordinal pairwise partitioning (opp) approach to neural networks training in bond rating. *Intelligent Systems in Accounting, Finance and Management*, 6:23–40, 1997.

J. Lattin, J. Douglas Carroll, and P.E. Green. *Analyzing Multivariate Data*. Thompson, Duxbury, 2003. ISBN 0-534-34974-9.

J.J. Maher and T.K. Sen. Predicting bond ratings using neural networks: A comparison with logistic regression. *International Journal of Intelligent Systems in Accounting, Finance & Management*, 6:59–72, 1997.

Website: MainstayInvestments. Mainstay investments financial glossary. http://www.nylim.com/mainstayfunds/0,2058,20_1002860,00.html, 2005.

E.H. Mamdani and S. Assilian. An experiment in linguistic synthesis with fuzzy logic controller. *International journal of Man-Machine Studies*, 7(1):1–13, 1975.

J. Moody and J. Utan. Architecture selection strategies for neural networks application to corporate bond rating. *Neural Networks in the Capital Markets, Wiley, Chichester,*, pages 277–300, 1995.

Moody's. Historical default rates of corporate bond issuers, 1920-1999. Technical report, Moody's Investor Services, 2000.

G.E. Pinches and K.A. Mingo. A multivariate analysis of industrial bond ratings. *The Journal of Finance*, 28(1):1–18, 1973a.

G.E. Pinches and K.A. Mingo. A multivariate analysis of industrial bond ratings. *Journal of Finance*, pages 1–18, 1973b.

G.E. Pinches and K.A. Mingo. The role of subordination and industrial bonds. *Journal of Finance*, pages 201–206, 1975.

T.F. Pogue and R.M. Soldovsky. What is in a bond rating? *Journal of Financial and Quantitative Analysis*, pages 201–228, 1969.

J.R. Quinlan. Introduction of decision trees. *Machine Learning*, 1(1):81–106, 1986.

D. Schutzer. *Artificial Intelligence: An application-oriented approach*, volume 1. Van Nostrad Reinhold Company Inc., 1987. ISBN 0-442-28034-3.

M. Shaw and J. Gentry. Inductive learning for risk classification. *IEEE Expert*, pages 47–53, 1990.

K.S. Shin and I. Han. A case-based approach using inductive indexing for corporate bond rating. *Decision Support System*, 32:41–54, 2001.

J.C. Singleton and A.J. Surkan. Neural networks for bond rating improved by multiple hidden layers. *Proceedings IEEE International Conference on Neural Networks*, pages 163–168, 1990.

M.A. Sprengers. Bankruptcy Prediction using Classification and Regression Trees, August 2005. Bachelor's thesis, Erasmus University Rotterdam.

Standard and Poor's. Corporate rating criteria 2005. Technical report, Standard and Poor's, 2005.

T. Takagi and M. Sugeno. Fuzzy indentification of systems and its applications to modeling and control. *Ieee Transactions on Systems, Man, and Cybernatics*, 15: 116–132, 1985.

R. Tan. Credit rating prediction using self-organising maps. Master's thesis, Erasmus University Rotterdam, July 2000.

Vicenç Torra. A review of the construction of hierarchical fuzzy systems. *International journal of intelligent systems*, 17:531–543, 2002.

P.C. van Aalst et al. *Financiering en Belegingen deel 2*, volume 7. Rhobeta Consultants, 1997. ISBN 90-73282-35-7.

J.P. van der Ent. De verklaring van bond rating op de kapitaalmarkt. Master's thesis, Erasmus University Rotterdam, August 1992.

J.P. van der Ent and D.J.C. Smant. *Credit rating and kapitaalmarkt*. Wolters-Noordhoff, 1994.

L. Waltman. Probabilistic fuzzy systems. Master's thesis, Erasmus University Rotterdam, February 2005.

L. Waltman, U. Kaymak, and J. van den Berg. Maximum likelihood parameter estimation in probabilistic fuzzy classifiers. *Proceedings of the 14th IEEE International Conference on Fuzzy Systems*, pages 1098–1103, 2005.

R.R. West. An alternative approach to predicting corporate bond ratings. *Journal of Accounting Research*, pages 118–127, 1970.

L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

Appendix A

Appendix A: Figures, charts and tables

A.1 The rating process

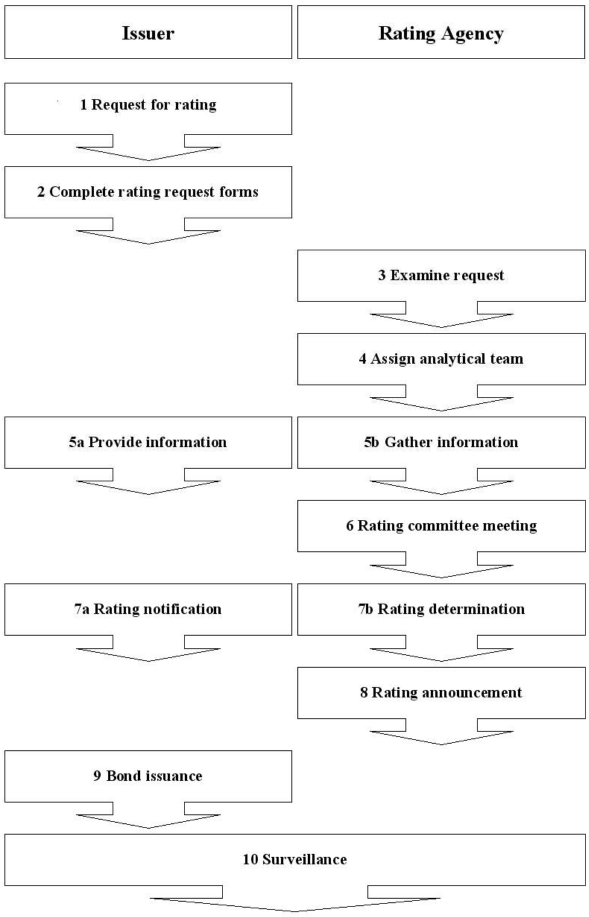

Figure A.1: Flow chart of an on request rating process.

A.2 General fuzzy inference system model

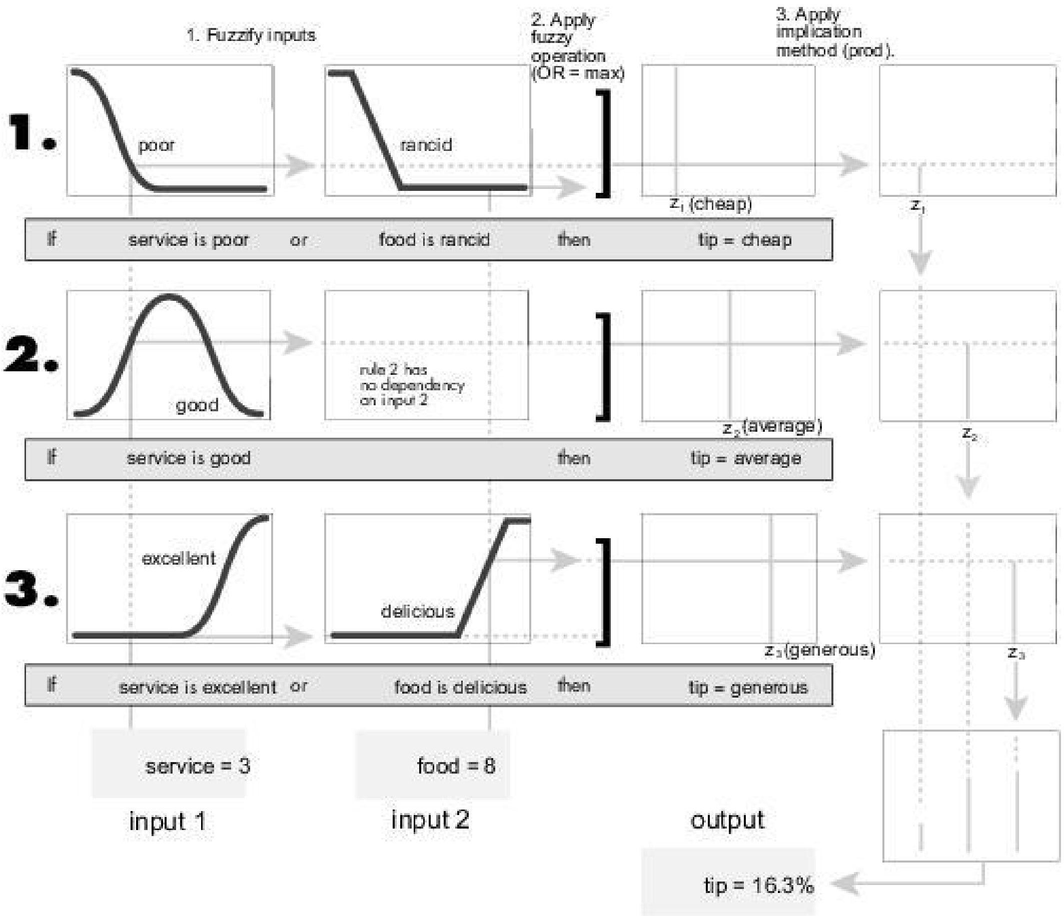

Figure A.2: Three-rule Takagi-Sugeno FIS deriving an overall output from two crisp inputs service and food

Appendix B

Appendix B: Data set and accompanying histograms

B.1 List of insurance companies

List of insurance companies				
#	**Company:**	**Country:**	**Actual rating**	**Assigned rating**
1	21st Century Insurance Group	USA	BBB+	**BBB**
2	Ace Limited	USA	BBB+	**BBB**
3	Aegon NV	Netherlands	A+	**A**
4	Aflac Inc	USA	A	**A**
5	AGF-Assurance Generale De France SA	France	A	**A**
6	Alleanza	Italy	AA	**AA**
7	Allianz AG	Germany	AA-	**AA**
8	Allianz Lebensvicherung AG	Germany	AA-	**AA**
9	ALM Brand A/S	Danmark	BBBpi	**BBB**
10	AMB Generali Holdings AG	Germany	AA	**AA**
11	Ambac Financial Group Inc	USA	AA	**AA**
12	American Equity Investment Life Holding	USA	BB+	**BB**
13	American Financial Group Inc	USA	BBB	**BBB**
14	American International Group Inc	USA	AA	**AA**
15	American National Insurance Company	USA	AA	**AA**
16	American Physicians Capital Inc	USA	BBB-	**BBB**
17	American Safety Insurance Holdings Limit	USA	Bpi	**B**
18	Amerus Group Company	USA	BBB+	**BBB**
19	Amlin PLC	Great Britain	BBB+	**BBB**
20	AON Corp.	USA	BBB+	**BBB**
21	Arch Capital Group Limited	USA	BBB	**BBB**
22	Argonaut Group Inc	USA	BBB-	**BBB**
23	Assurant Inc	USA	BBB+	**BBB**

Table B.1: Complete list of insurance companies and their ratings

List of insurance companies, continued				
#	Company:	Country:	Actual rating	Assigned rating
24	Aviva PLC	Great Britain	A+	**A**
25	AXA	France	A	**A**
26	AXA Colonia Lebensversicherung	Germany	AA-	**AA**
27	AXA Portugal	Portugal	BBpi	**BB**
28	Baloise	Switzerland	BBBpi	**BBB**
29	Berkshire Hathaway Inc	USA	AAA	**AAA**
30	Bristol West Holdings Inc	USA	BB+	**BB**
31	Cattolica Assicurazioni	Italy	Api	**A**
32	Ceres Group Inc	USA	AAA	**AAA**
33	Chubb Corp.	USA	A	**A**
34	Cincinnati Finance	USA	A	**A**
35	CNA Financial Corp.	USA	BBB-	**BBB**
36	Codan A/S	Danmark	A-	**A**
37	Commerce Group Inc	USA	BBB	**BBB**
38	Conseco Inc	USA	BB-	**BB**
39	Converium Holding	USA	BBB+	**BBB**
40	Corp. Mapfre CIA Internacional	Spain	AA-	**AA**
41	Covanta Holding Corp.	USA	B+	**B**
42	DBV-Winterthur Holding	Germany	BBBpi	**BBB**
43	Delphi Financial Group Inc	USA	BBB	**BBB**
44	Deutsche Aerztversicherung	Germany	BBBpi	**BBB**
45	Domestic & General Group PLC	Great Britain	Api	**A**
46	Donegal Group Inc	USA	BBBpi	**BBB**
47	E-L Financial Corp. Limited	Canada	A	**A**
48	EMC Insurance	USA	BBBpi	**BBB**
49	Ergo Versicherung AG	Germany	A-	**A**
50	Erie Family Life Insurance Company	USA	Api	**A**
51	Erie Indemnity Company	USA	Api	**A**
52	Euler Hermes	France	AA-	**AA**
53	Everest RE Group Limited	USA	A-	**A**
54	Fairfax Financial Holdings Limited	Canada	BB	**BB**
55	FBD Holdings PLC	Ireland	BBBpi	**BBB**
56	FBL Financial Group Inc	USA	BBB	**BBB**
57	Fidelity National Financial Inc	USA	BBB-	**BBB**
58	Fidelity National Title Group	USA	BBB-	**BBB**
59	First American Corp.	USA	BBB+	**BBB**
60	Fondiaria-SAI RNC	Italy	BBB	**BBB**
61	Fpic Insurances Group	USA	BBpi	**BB**
62	Friends Provident PLC	Great Britain	A-	**A**
63	Generali	Italy	AA	**AA**
64	Genworth Financial Inc	USA	A	**A**
65	Gerling Konzern Allgemeine	Germany	A-	**A**

List of insurance companies, continued				
#	**Company:**	**Country:**	**Actual rating**	**Assigned rating**
66	Great American Financial Resources Inc	USA	BBB-	**BBB**
67	Great West Life Company Inc	Canada	AA	**AA**
68	Hannover Rueckversicherung AG	Germany	AA-	**AA**
69	Hanover Insurance Group Inc	USA	BB+	**BB**
70	Harleysville Group Inc	USA	BBB-	**BBB**
71	Hartford Financial Services Group	USA	A-	**A**
72	HCC Insurance Holdings	USA	A	**A**
73	Helvetia Pattria Holding	Germany	BBBpi	**BBB**
74	Hilb Rogal & Hobbs Company	USA	BB	**BB**
75	Hiscox PLC	Great Britain	A-	**A**
76	Horace Mann Corp.	USA	BBB	**BBB**
77	Industrial Alliance Insurance & Financia	Canada	A+	**A**
78	Infinity Property & Casualty Corp.	USA	BBB	**BBB**
79	ING Canada Inc	Canada	A+	**A**
80	IPC Holdings Limited	USA	BBB+	**BBB**
81	Irish Life & Permanent PLC	Ireland	A+	**A**
82	Kansas City Life Insurance Company	USA	A+	**A**
83	Kentucky Investors Inc	USA	AA-	**AA**
84	Kingsway Financial Services Inc	Canada	BBB-	**BBB**
85	Koelnische Rueckversicherung AG	Germany	AAA	**AAA**
86	Landamerica Financial Group	USA	BBB-	**BBB**
87	Legal & General Group PLC	Great Britain	AA-	**AA**
88	Lincoln National Corp.	USA	A-	**A**
89	Lindsey Morden Group Inc	Canada	B	**B**
90	Loews Corp.	USA	A	**A**
91	Manulife Financial Corp.	Canada	AA-	**AA**
92	Markel Corp.	USA	BBB-	**BBB**
93	Marsh & McLennan Companies	USA	BBB	**BBB**
94	MBIA Inc	USA	AA	**AA**
95	Mercury General Corp.	USA	A	**A**
96	Metlife Inc	USA	A	**A**
97	Milano Assicurazioni	Italy	BBB	**BBB**
98	Mutual & Federal Insurance Company Limit	South Africa	Api	**A**
99	National Western Life Insurance Company	USA	A	**A**
100	Nationwide Financial Services	USA	A-	**A**
101	Navigators Group Inc	USA	BBB	**BBB**
102	Nuernberger Beteiligung AG	Germany	BBB+	**BBB**
103	Odyssey RE Holdings Corp.	USA	BBB-	**BBB**
104	Ohio Casualty Corp.	USA	BB+	**BB**
105	Partnerre Limited	USA	A	**A**

List of insurance companies, continued				
#	**Company:**	**Country:**	**Actual rating**	**Assigned rating**
106	Phoenix Companies Inc	USA	BBB	**BBB**
107	Pohjola Group PLC	Finland	A+	**A**
108	Presidential Life Corp.	USA	B-	**B**
109	Principal Financial Group Inc	USA	A	**A**
110	Proassurance Corp.	USA	BBB-	**BBB**
111	Progressive Corp. Ohio	USA	A+	**A**
112	Protective Life Corp.	USA	A	**A**
113	Prudential Financial Inc	USA	A-	**A**
114	Prudential PLC	Great Britain	AA-	**AA**
115	Pxre Corp.	USA	BBB-	**BBB**
116	RAS RNC	Italy	AA-	**AA**
117	Renaissancere Holdings Ltd	USA	A-	**A**
118	Rheinland Holding AG	Germany	BBpi	**BB**
119	RLI Corp.	USA	BBB+	**BBB**
120	Royal & Sun Alliance Insurance Group PLC	Great Britain	BBB	**BBB**
121	Safeco Corp.	USA	BBB+	**BBB**
122	Safety Insurance Group Inc	USA	Api	**A**
123	Saint James's Place Capital PLC	Great Britain	A-	**A**
124	Saint Paul Travellers Companies Inc	USA	BBB+	**BBB**
125	Sampo PLC	Finland	A	**A**
126	Santam Limited	South Africa	Api	**A**
127	Scor SA	France	A-	**A**
128	Scottish RE Group Limited	USA	BBB-	**BBB**
129	Scpie Holdings Inc	USA	Bpi	**B**
130	Selective Insurance Group Inc	USA	BBB	**BBB**
131	Skandia Forsakrings AB	Sweden	A	**A**
132	South Africa Eagle Insurance Company Lim	South Africa	BBBpi	**BBB**
133	Stancorp Financial Group Inc	USA	BBB+	**BBB**
134	State Auto Financial Corp.	USA	BBB	**BBB**
135	Stewart Information Services Corp.	USA	A-	**A**
136	Storebrand ASA	Norway	BBB+	**BBB**
137	Sun Life Financial Inc	Canada	AA-	**AA**
138	Swiss Life Holding	Switzerland	BBB	**BBB**
139	Swiss Reinsurance Company	Switzerland	AA	**AA**
140	Topdanmark A/S	Danmark	BBBpi	**BBB**
141	Toro	Italy	A-	**A**
142	Tower Group Inc	USA	BBB-	**BBB**
143	Transatlanitc Holdings Inc	USA	A-	**A**
144	Uici	USA	BBB-	**BBB**
145	Unipol	Italy	A-	**A**

List of insurance companies, continued				
#	**Company:**	**Country:**	**Actual rating**	**Assigned rating**
147	United America Indemnity Limited	USA	BBBpi	**BBB**
148	United Fire & Casualty Company	USA	A	**A**
149	Unitrin Inc	USA	BBB+	**BBB**
150	Unumprovident Corp.	USA	BB+	**BB**
151	Usi Holdings Corp.	USA	BB-	**BB**
152	Vaudoise	Switzerland	A-	**A**
153	Vittoria Assicurazioni	Italy	BBBpi	**BBB**
154	Wesco Financial Corp.	USA	AAA	**AAA**
155	White Mountains Insurance Group	USA	BBB	**BBB**
156	Wiener Staedt VZ AG	Austria	A+	**A**
157	Willis Group Holdings Limited	Great Britain	BBB-	**BBB**
158	Wuerttembergische Lebensvicherung Regist	Germany	A-	**A**
159	Yadkin Valley Corp.	USA	A	**A**
160	Zenith National Insurance Corp.	USA	BB+	**BB**
161	Zurich Financial Services	Switzerland	BBB	**BBB**

B.2 Descriptive analysis

Variable	Statistics per variable				
	mean	**median**	**stdev**	**minimum**	**maximum**
Total assets	62480,87	11722	150854,2	22,63822	1302894
Total debt	5123,431	367,0746	29860,62	0	356591,1
LT Debt / total invested capital	0,219181	0,206221	0,172965	0	0,894892
Debt ratio	0,215025	0,209164	0,167695	0	0,840287
Debt-to-equity ratio	0,413986	0,264484	0,786433	0	8,514038
Net gearing	12,78848	5,931134	25,77467	-0,33377	221,1164
Return on equity	0,125694	0,123532	0,088294	-0,51773	0,388267
Operating income / sales	0,129147	0,109124	0,106448	-0,07001	0,841035
Net profit margin	0,080463	0,073507	0,076221	-0,1878	0,588057
Sales	10188,29	3040,8	20222,07	0,237248	124621,7
Total liabilities to assets	0,821379	0,857221	0,142125	0,176778	0,995504
Liabilities / (cash + assets)	0,796847	0,825274	0,153576	0,121786	0,994133
Sales / Net worth	2,507691	1,791328	3,379935	0,12359	31,88826
LT Debt / total assets	0,046796	0,028524	0,069501	0	0,5975

Table B.2: Summary statistics per variable. (Year 2004)

Predicted rating class							
Actual rating class	AAA	AA	A	BBB	BB	B	All
AAA	0.10	0.52	1.37	1.57	0.32	0.12	4
AA	0.52	2.74	7.17	8.22	1.70	0.65	21
A	1.37	7.17	18.79	21.52	4.44	1.71	55
BBB	1.57	8.22	21.52	24.65	5.09	1.96	63
BB	0.32	1.70	4.44	5.09	1.05	0.40	13
B	0.12	0.65	1.71	1.96	0.40	0.16	5
All	4	21	55	63	13	5	161

Table B.3: Accuracy matrix based on the relative frequency of each rating class within the insurance data set

B.3 Histograms

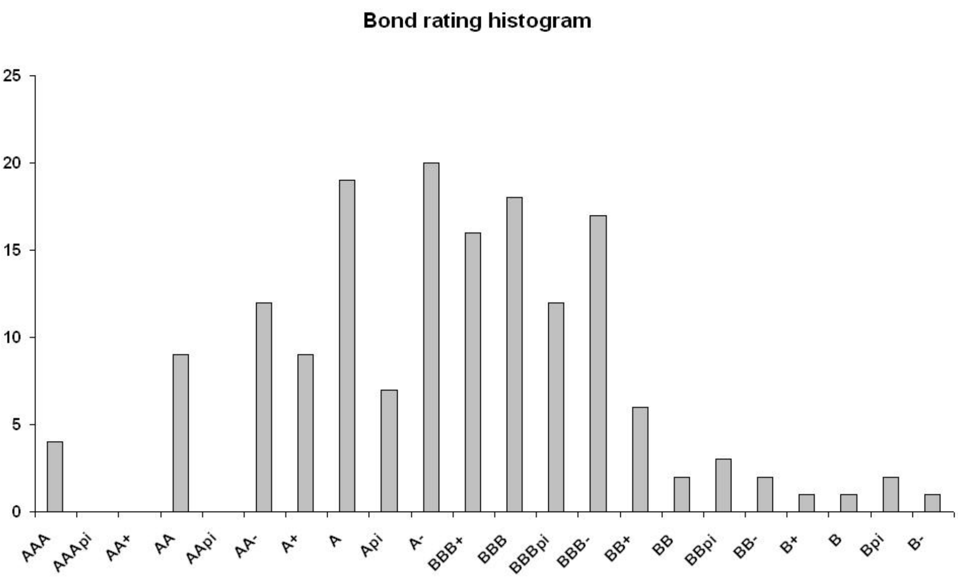

Figure B.1: Extensive bond rating histogram of the insurance companies data set.

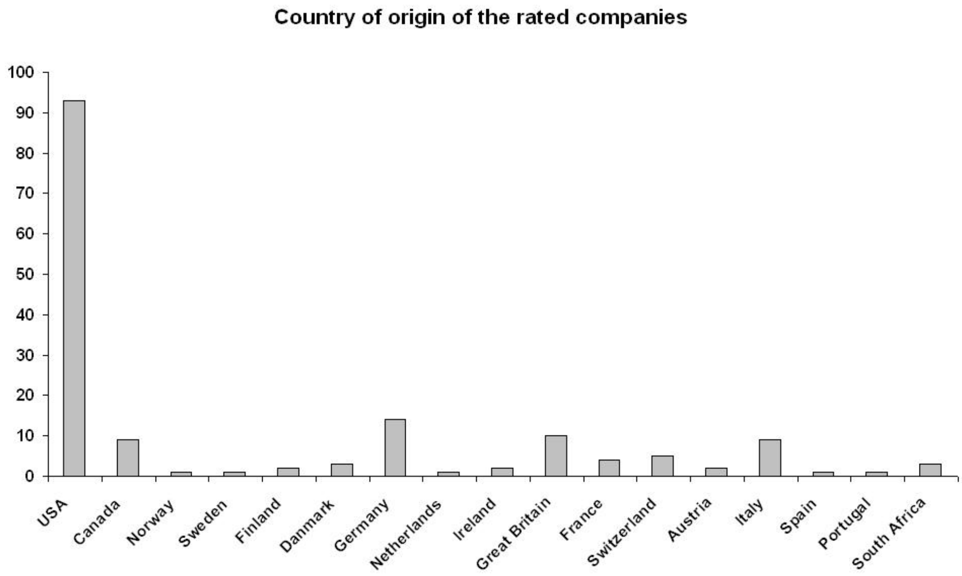

Figure B.2: Histogram with the countries of origin of the companies in the data set

Appendix C

Appendix C: Balance sheet and profits & losses accounts

Consolidated balance sheet

Assets/Period	Liabilities/Period
Tangible fixed assets	Issued and paid up capital
Financial fixed assets	Reserves
Other fixed assets	Loss
Total Fixed assets	**Shareholders' equity**
Inventories	
Receivables	Provisions
Cash and cash equivalents	Long-term debt
Total Current assets	**Long-term liabilities**
	Short-term bank debt
Intangible assets	Other short-term debt
	Accounts payable
	Current liabilities
Other assets	**Other liablilities**
Total assets	**Total Liabilities**

Period	Period
Highest share price	Number of outstanding shares
Lowest share price	
Average share price	

Period	Nominal value share
Number of employees	

Table C.1: Standardized balance sheet.

Consolidated profits & losses account

Period
Net sales (excluding turnover tax)
Cost of goods sold
Gross margin
Wages and salaries
Operating costs
Operating income (EBITDA)
Depreciation
Income from equity participations
Interest expenses
Result on ordinary activities
Extraordinary profits and losses: provisions
Other extraordinary profits and losses
Result before tax
Income tax expense
Result after tax
Minority interest
Net result

Table C.2: Standardized profits & losses account.

Appendix D

Appendix D: Training & testing of a data set

D.1 Cross validation methods

This Appendix gives a moderate overview of the most common used training versus testing methodologies.

Holdout method The idea behind this technique is extremely simple. It simply splits a data set into two individual data set, a training and a test data set.[1] So it leaves the test set out of the training the models. In this way the test data set can function as an independent data set to determine the accuracy of the model. The advantages of this method is its simpleness and low computational costs. The main disadvantage is the possibility of high variance in the evaluation. Mainly, because this evaluation totally depends on the way how the data set is separated.

K-fold cross validation This technique is an improvement over the holdout method described before. This technique splits the data set into *k* subsets, so the holdout method can be repeated *k* times. Every iteration one subset is used as the test data set, while $k-1$ subsets form the training data set. Then the average error of the *k* tests is computed. Because of this technique some of the disadvantages of the holdout method are banned. For instance the fact that this technique uses several different test data sets, it reduces the change to score high variances. As a disadvantage there is the problem of high computational cost, because it has to train and test the tree *k* times. The number which they usually assign to *k* is 10. A variant to this *k*-fold cross validation is to randomly divide the data set into a training and test data set *k* different times.

Leave-one-out cross validation This is the same as *k*-fold cross validation, only now it assigns the size of the total data set to *k*. This means that every iteration the tree is trained by $N-1$ data points and tested by 1 data point. In this way the data set is optimally utilized.

[1] Where the test set usually is smaller than the training set.

It depends on the data set and on the situation which of the aforementioned techniques to elect as the best technique. But as you can see the last two techniques make more use of a data set than the holdout method and obtain smaller variances in the evaluation.

Appendix E

Appendix E: Classification and regression trees

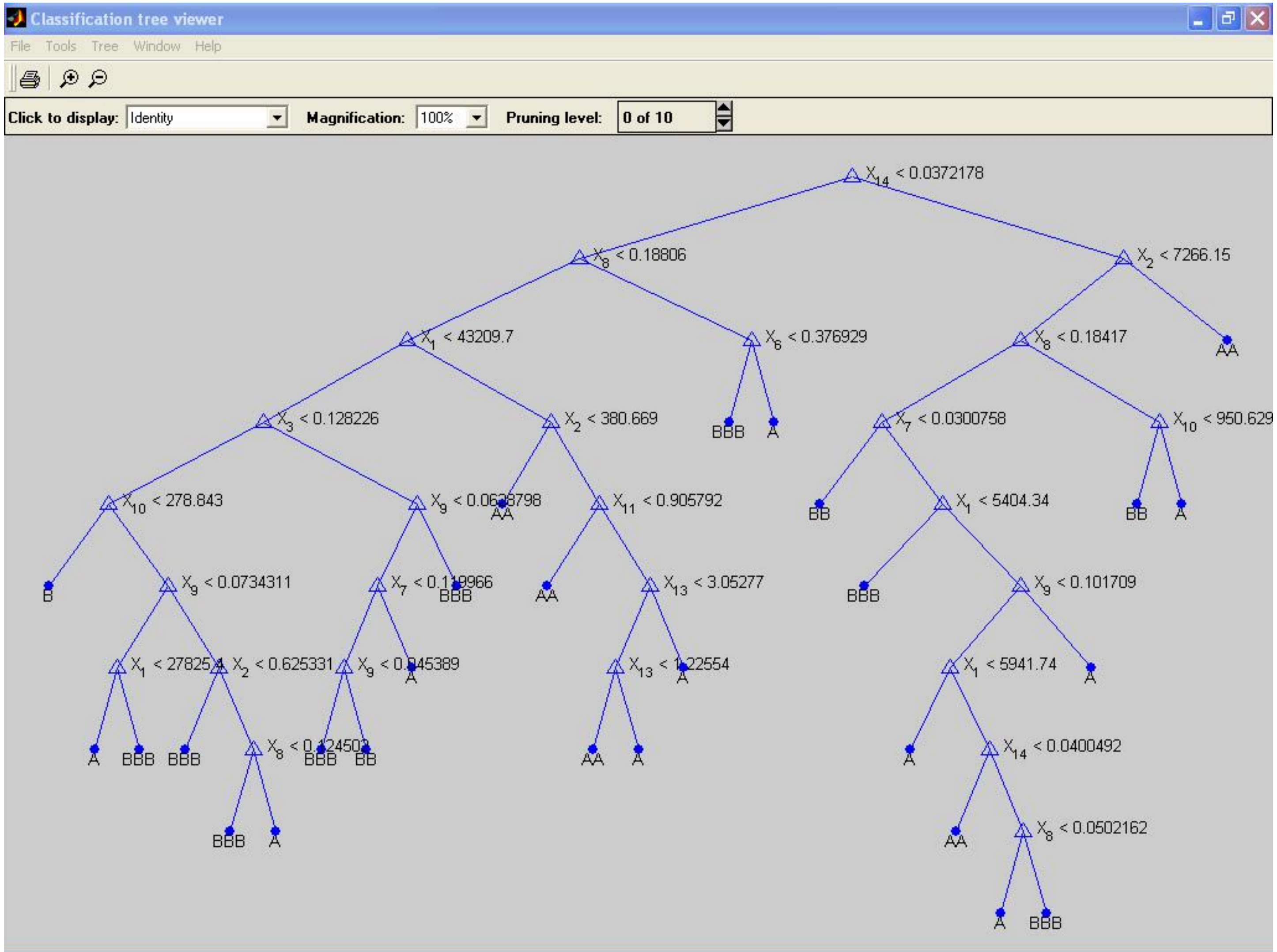

Figure E.1: Unpruned tree built by the CART algorithm of the insurance companies data set

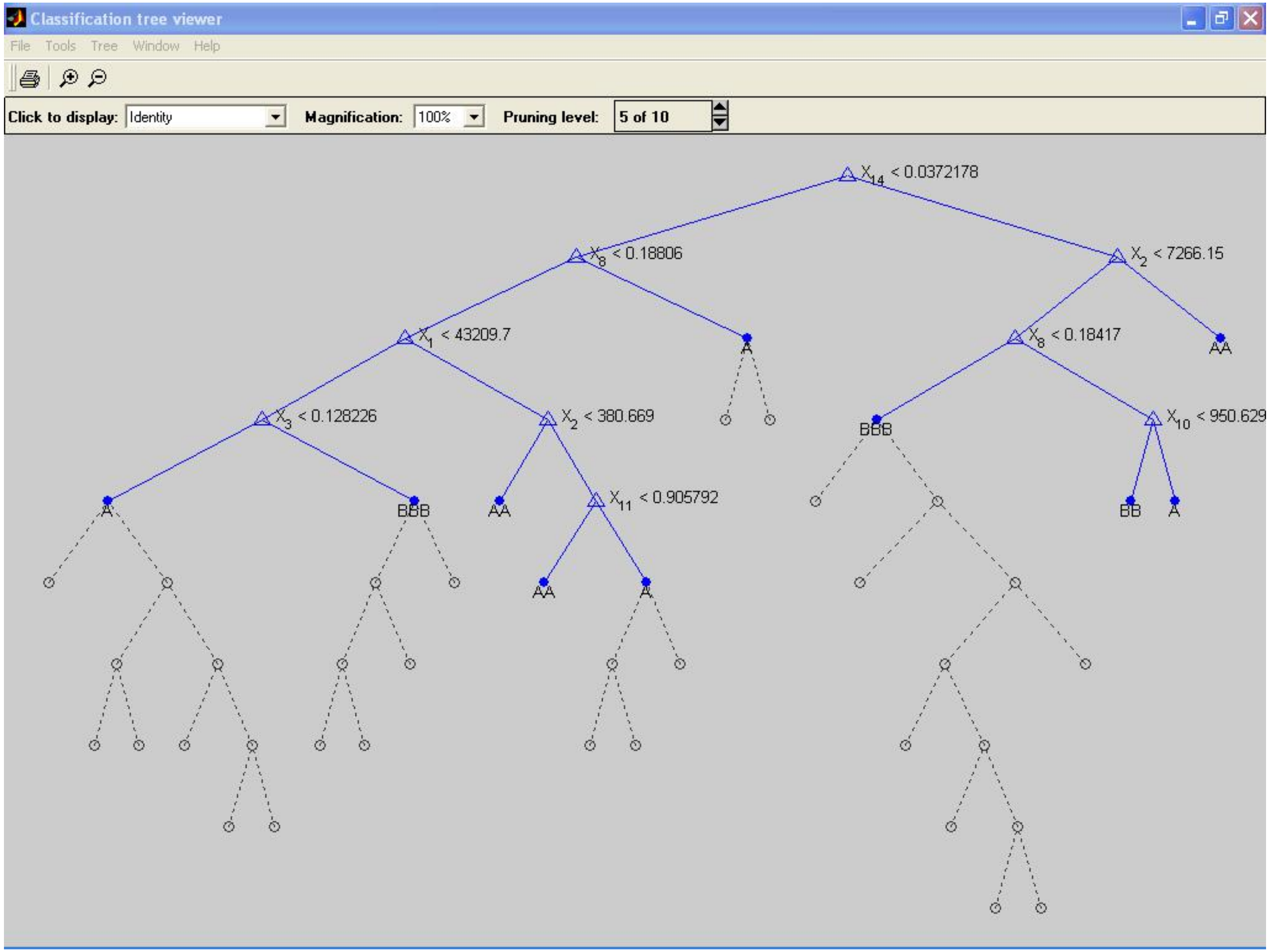

Figure E.2: Pruned tree built by the CART algorithm of the insurance companies data set

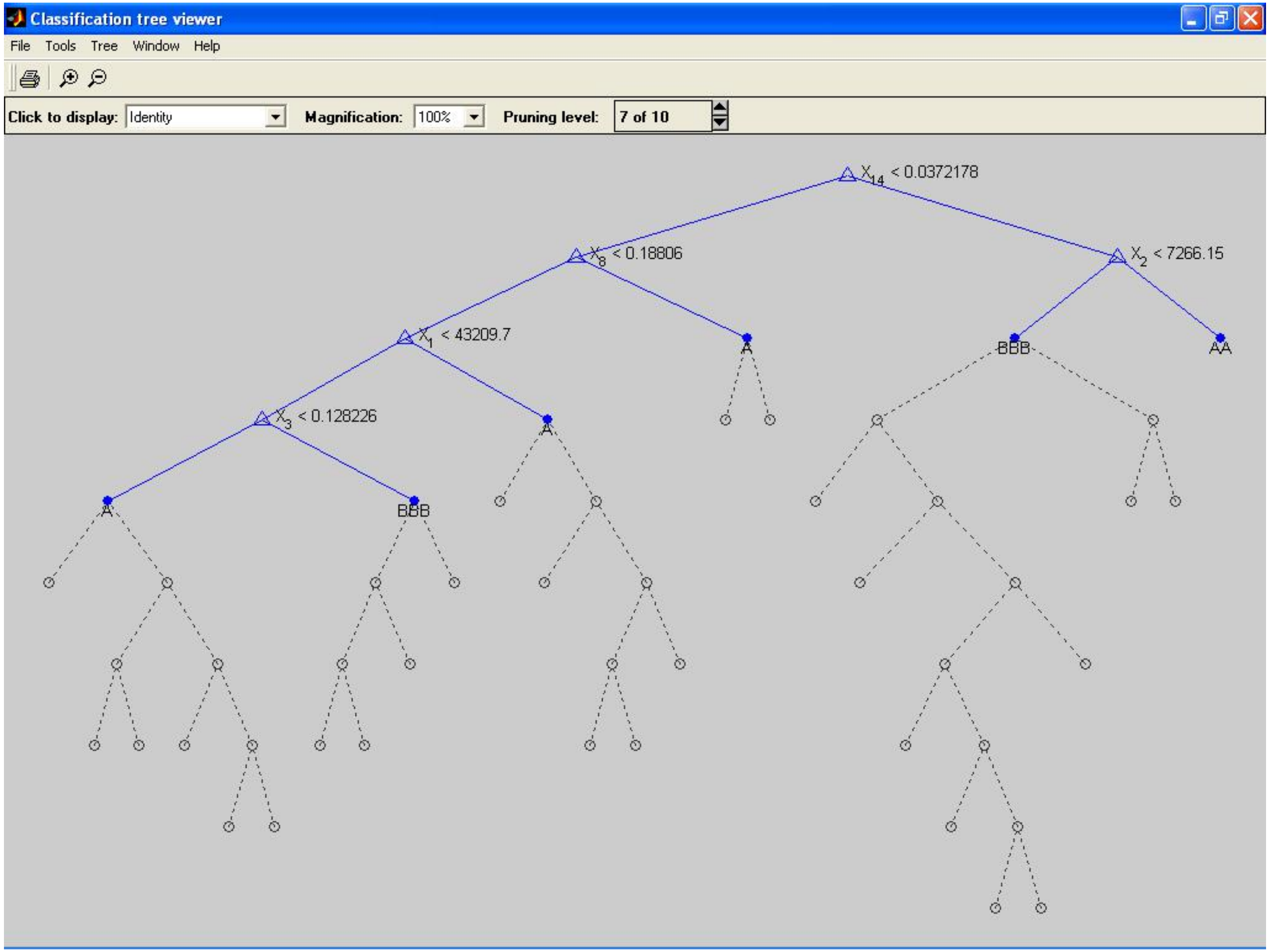

Figure E.3: Pruned tree built by the CART algorithm of the insurance companies data set

Appendix F

Appendix F: Probabilistic fuzzy systems

The membership functions illustrated in this appendix are based on the PFS model constructed with 6 rules and trained and tested by set 3. (See Chapter 5 for explanation on this set)

F.1 Membership functions

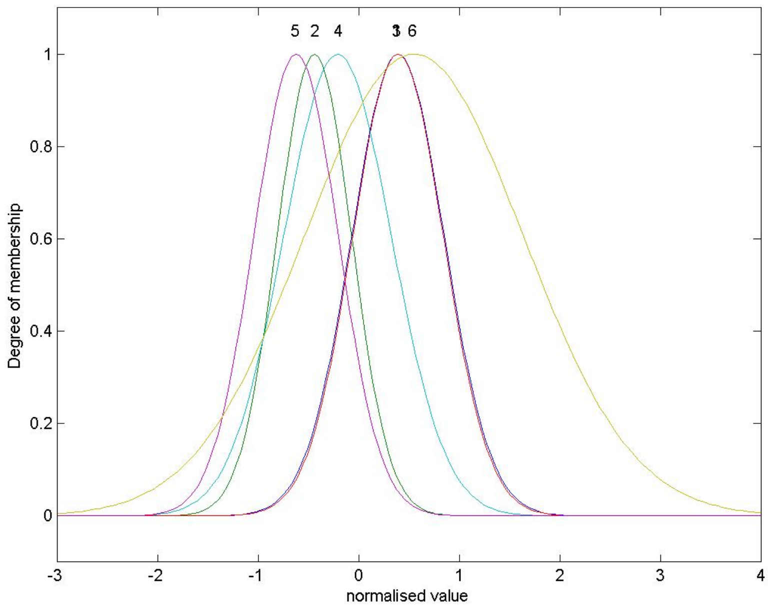

Figure F.1: Membership functions of variable X_2 for rules $1-6$

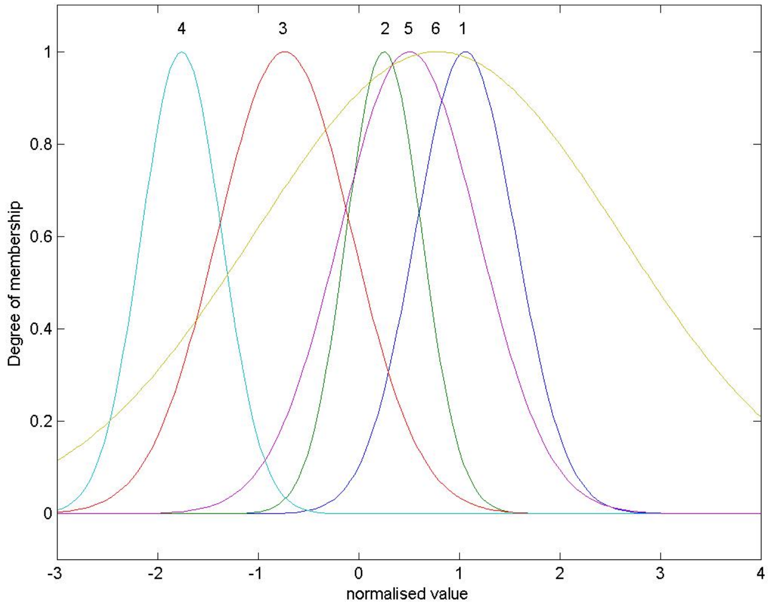

Figure F.2: Membership functions of variable X_3 for rules $1-6$

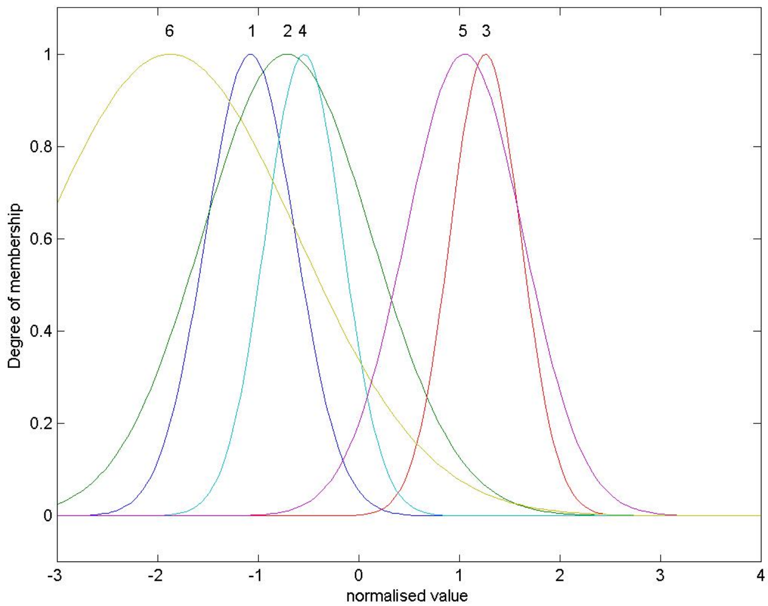

Figure F.3: Membership functions of variable X_8 for rules $1-6$

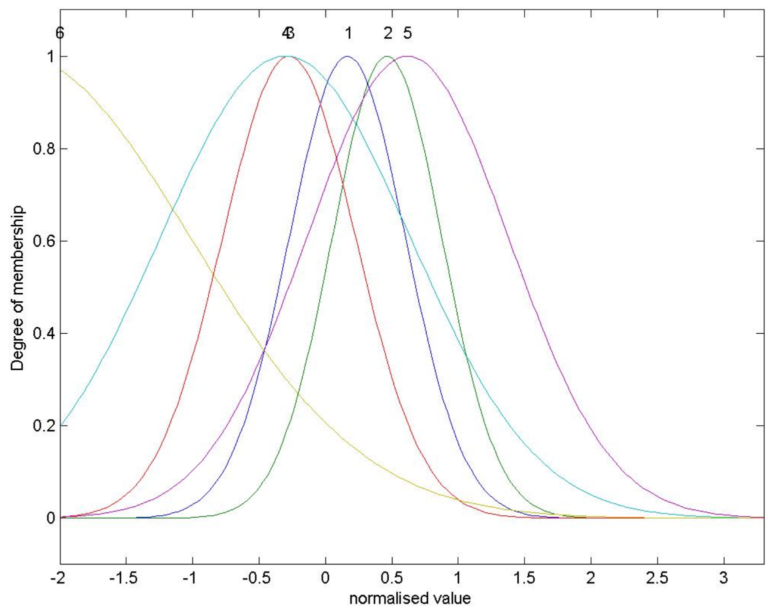

Figure F.4: Membership functions of variable X_{14} for rules $1 - 6$

F.2 Probabilistic fuzzy rules

The figures below show the 6 rules defined by our 'most accurate' PFS model, as described in Chapter 5. The rule-figures show the membership functions of the individual independent variables above each other. The bar-charts show the probabilities belonging to these rules, where the bar on the left side points at rating class 'AAA' and so on.

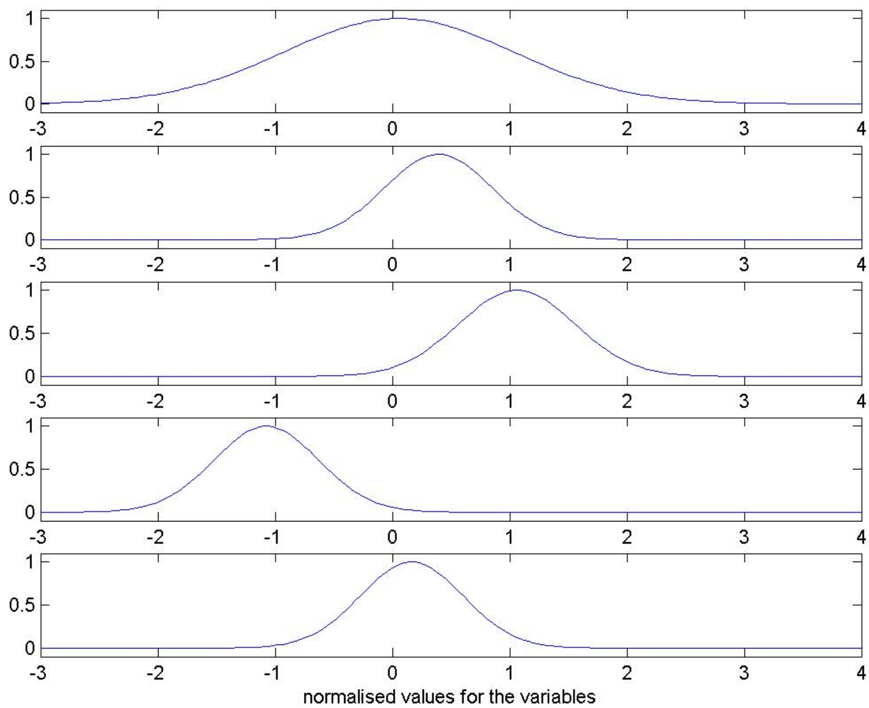

Figure F.5: Representation of rule 1

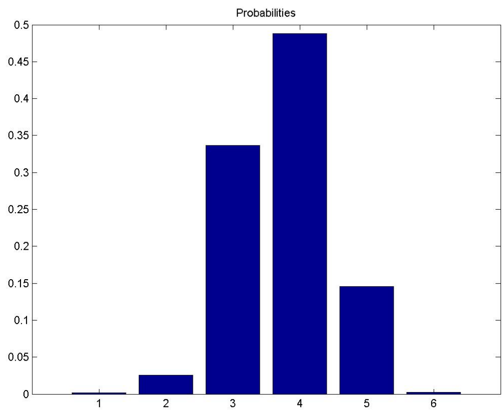

Figure F.6: Probabilities belonging to rule 1

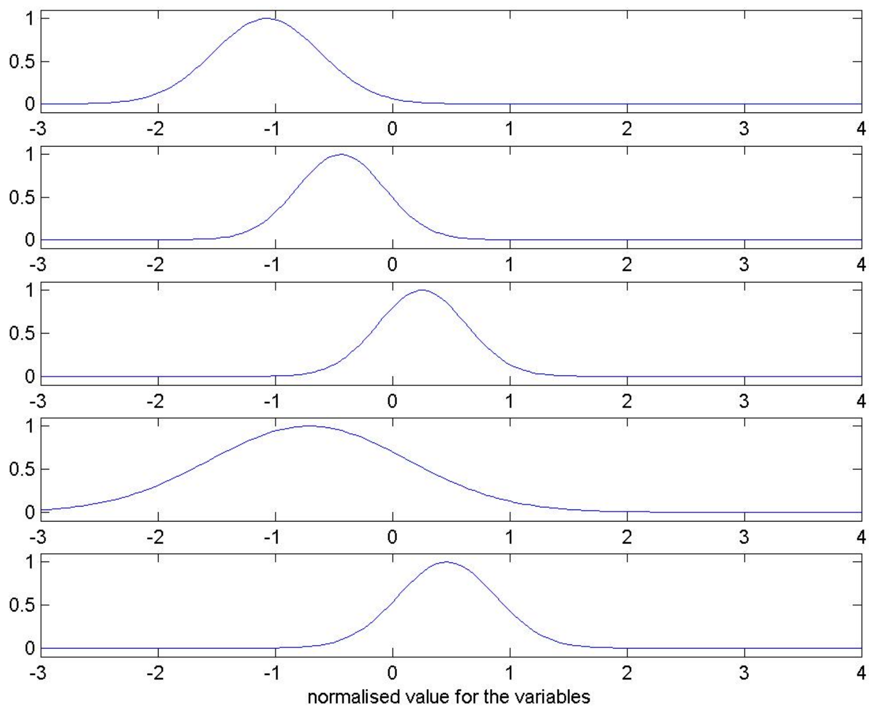

Figure F.7: Representation of rule 2

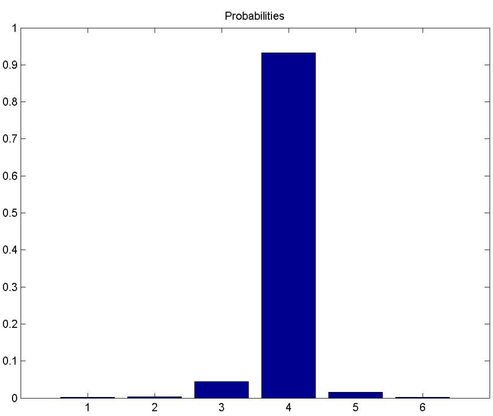

Figure F.8: Probabilities belonging to rule 2

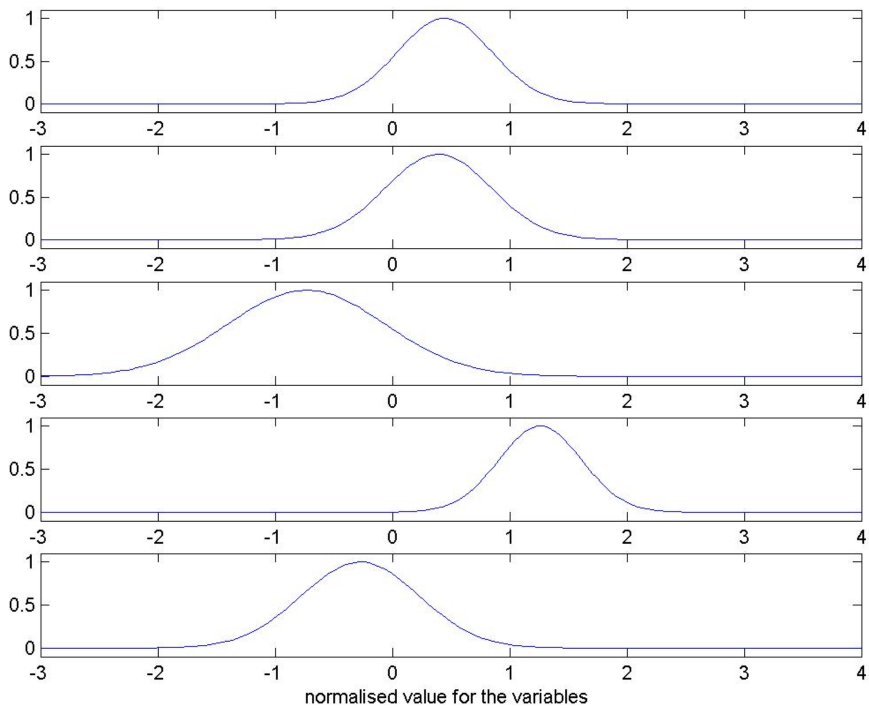

Figure F.9: Representation of rule 3

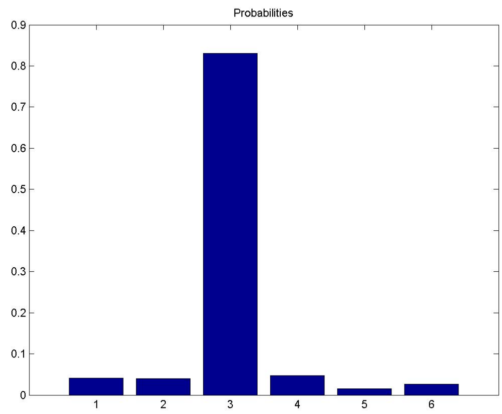

Figure F.10: Probabilities belonging to rule 3

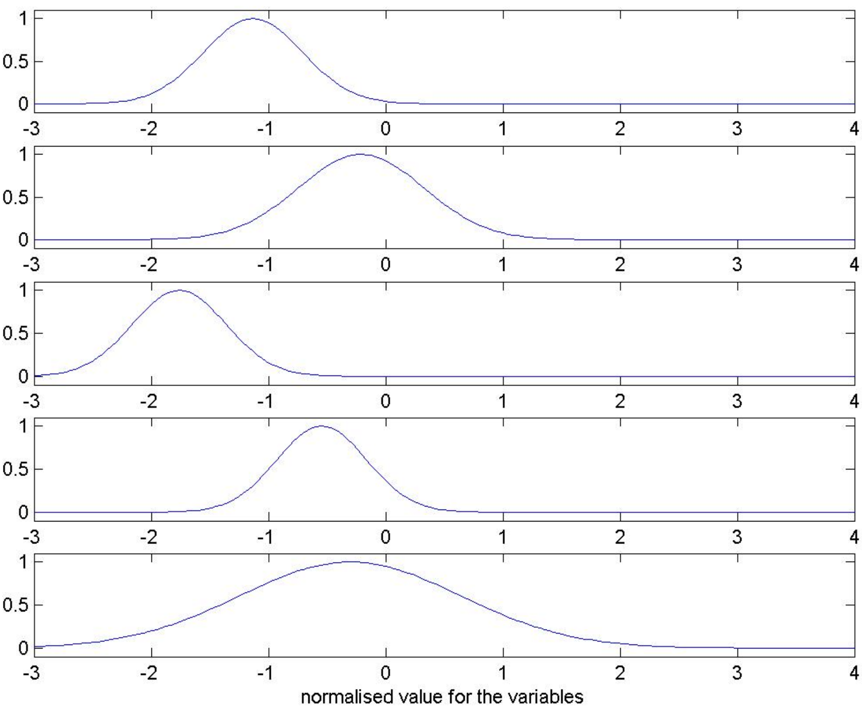

Figure F.11: Representation of rule 4

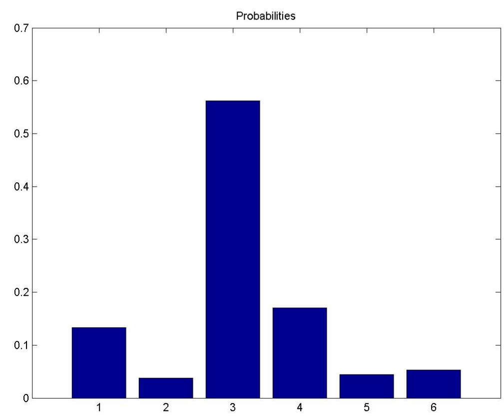

Figure F.12: Probabilities belonging to rule 4

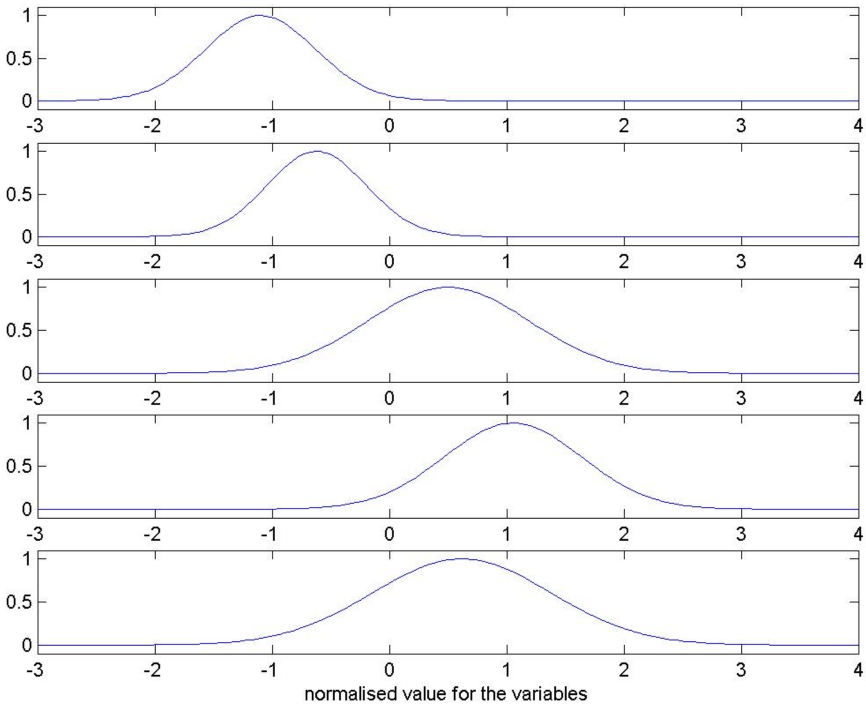

Figure F.13: Representation of rule 5

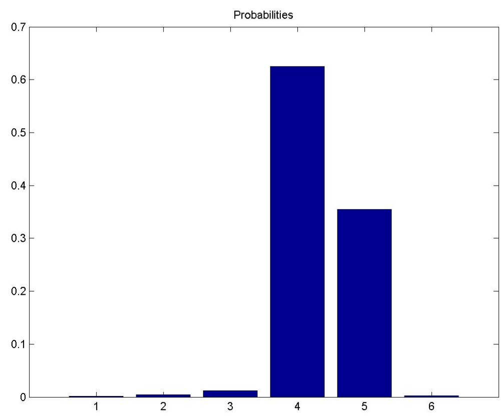

Figure F.14: Probabilities belonging to rule 5

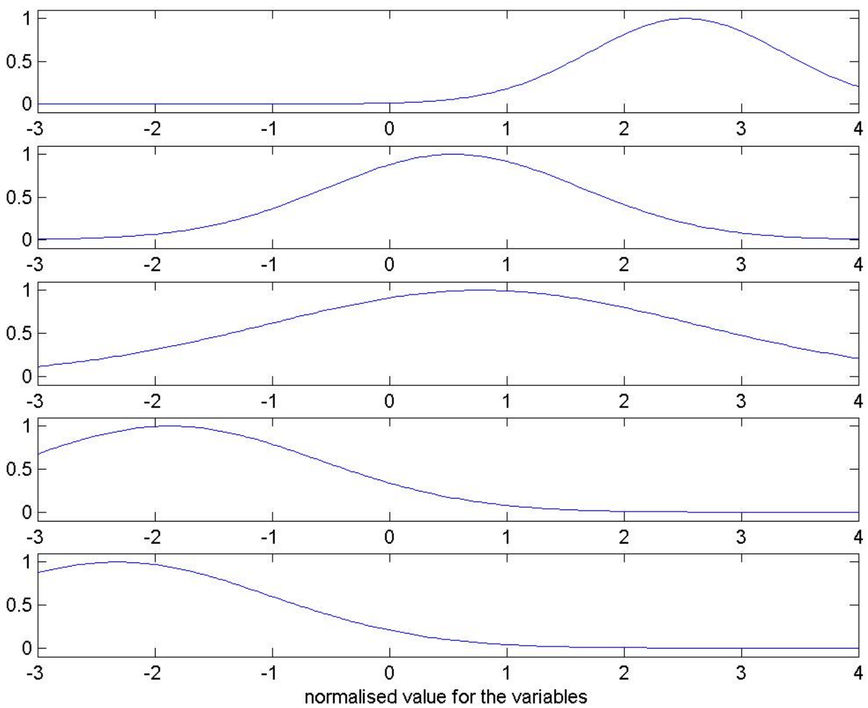

Figure F.15: Representation of rule 6

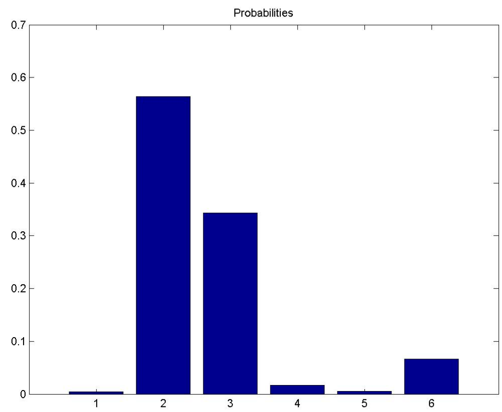

Figure F.16: Probabilities belonging to rule 6