

**MIDDLE EAST TECHNICAL UNIVERSITY**

**INSTITUTE OF APPLIED MATHEMATICS**

**FINANCIAL MATHEMATICS**

**IAM 589 TERM PROJECT**

**COMPARISON OF MARS, CMARS AND CART  
IN PREDICTING DEFAULT PROBABILITIES FOR  
EMERGING MARKETS**

**ERKAN BÜYÜKBEBECİ**

**1144831**

**e114483@metu.edu.tr**

**Supervisor: Prof. Dr. Gerhard Wilhelm Weber**

**AUGUST 2009**

## ABSTRACT

### COMPARISON OF MARS, CMARS AND CART IN PREDICTING DEFAULT PROBABILITIES FOR EMERGING MARKETS

Erkan BÜYÜKBEBECİ

M.Sc., Department of Financial Mathematics

Supervisor: Prof. Dr. Gerhard Wilhelm Weber

The concept of data mining has become popular as a business information management tool where it is expected to reveal knowledge structures that can guide decisions in conditions of limited certainty [43]. Data mining is commonly used in a wide range of financial fields such as credit scoring, loan management and financial forecasting, which are very important in business decision making. The main objective of this term project is to evaluate the performance of commonly used data mining techniques – Classification and Regression Trees (CART) and Multivariate Adaptive Regression Splines (MARS) in predicting default probabilities for emerging markets. Moreover, MARS algorithm is modified by constructing a penalized residual sum of squares (PRSS) as a Tikhonov regularization problem and this problem is solved by conic quadratic programming (CQP) which provides us an alternative modeling technique for MARS, named CMARS. Here “C” represents not only the word **conic** but also **convex** and **continuous** [53]. To point out which method is more effective, the financial information of 45 emerging markets in the period from 1980 to 2005 is analyzed and the three methods are compared by the use of statistical performance criteria. Ratings on the results of best method is done by using receiver operating characteristic curve. The results of the analyze show that CMARS outperform traditional MARS and CART in terms of classification accuracy rate and hence provide an efficient alternative in predicting default probabilities.

Keywords: Default Probability, Classification and Regression Tree, Multivariate Adaptive Regression Splines, CMARS, Optimization, Accuracy Ratio

# TABLE OF CONTENTS

## CHAPTER

<b>1 INTRODUCTION.....</b>	<b>4</b>
<b>2 HISTORY OF METHODS USED IN CREDIT SCORING AND PREDICTING FINANCIAL CRISIS .....</b>	<b>6</b>
<b>3 METHODS.....</b>	<b>12</b>
3.1 Multivariate Adaptive Regression Splines .....	12
3.1.1 The Approach.....	13
3.1.2 Advantages and Disadvantages of MARS.....	18
3.2 CMARS.....	19
3.2.1 Tikhonov Regularization and the Construction of the Conic Quadratic Programming.....	20
3.2.2 MOSEK.....	22
3.2.3 Advantages and Disadvantages of CMARS .....	22
3.3 Classification and Regression Tree (CART).....	23
3.3.1 Criterion of Diversity.....	24
3.3.2 Goodness of Split Criterion.....	24
3.3.3 The Class Assignment Rule to the Terminal Nodes and Resubstitution Estimates.....	27
3.3.4 Selecting the Correct Complexity of a Tree.....	28
3.3.5 Advantages and Disadvantages of CART.....	29
3.4 Theoretical Relationships between MARS and CART.....	30
<b>4 APPLICATION.....</b>	<b>32</b>
4.1 Data and Methodology .....	32
4.2 Application Results .....	33
4.2.1 Classification and Regression Trees.....	33
4.2.2 Multivariate Adaptive Regression Splines.....	33
4.2.3 CMARS.....	34
4.2.4 Results Between Different Scoring Models.....	35
4.2.5 Type I and Type II Errors of the Constructed Models.....	35

4.2.6 The Receiver Operating Characteristic Curves of the Constructed Models...	36
4.2.7 Variable Selection Results in Multivariate Adaptive Regression Splines .....	38
<b>5 CONCLUSION.....</b>	<b>40</b>
<b>REFERENCES.....</b>	<b>42</b>

# CHAPTER 1

## INTRODUCTION

Credit scoring is the set of decision models and their underlying techniques that aid lenders in the granting of consumer credits. These techniques decide who will get a credit, how much credit they should get, and what operational strategies will enhance the profitability of borrowers to the lenders [50].

Over the past 25 years, credit scoring has become a vital issue which takes place in issuing credit cards, auto loans, some other kinds of consumer loans and especially in mortgage origination which is new for Turkish Real Estate Market. Credit scoring has lots of benefits in terms of loan evaluation, diminishing the risks associated with inappropriate credit decisions of financial institutions and decreasing the load of probable economic crisis on government and firms. Also, in the context of the Basel II Accord, there is an increasing emphasis on quantitative techniques for estimation of credit risk [21].

Since credit scoring is an important task for the credit industry, there has been an increase in the development of credit scoring methods during the past few years. We can classify credit scoring methods into two groups which are traditional statistical methods and data mining. These methods have been developed to successfully handle the credit scoring tasks. In the traditional group we have discriminant analysis, logistic and probit regression. In the other side, we have the Data Mining tools such as Classification and Regression Tree (CART) and Multivariate Adaptive Regression Splines (MARS) which are the effective alternatives of traditional methods [32].

Some researchers have criticized statistical credit scoring techniques because of their model assumptions. In contrast, Data Mining approach is becoming a common alternative for making credit scoring models due to its associated memory characteristic, generalization capability, and outstanding credit scoring capability. For this reason, Classification and Regression Tree (CART) and Multivariate Adaptive

Regression Splines (MARS) have become very popular in credit scoring models and classification tasks [40].

In parallel with the credit scoring models, in the last decade, many empirical studies have concentrated their attention in developing models able to timely signal the occurrence of a financial crisis, the so-called *early warning system* (EWS). Assessing and predicting these crisis is of great importance for governments and decision makers. In this study we consider the debt crisis and intend to capture the the signals of debt crisis before crisis exists. In this project, we assess the performance and the effectiveness of predicting default probabilities using the financial information of 45 emerging markets in the period from 1980 to 2005 by using CART, MARS and CMARS. The data of dependent variable: "whether a country is in debt crisis or not" in the period 1980-2004 is obtained from Fioramanti, Marco's paper [13]. Moreover, the debt crises in 2005 is supplied according to his definition: "A country is defined to be in debt crises if it is classified as being in default by Standard and Poor's or if it receives a large nonconcessional IMF loan defined as access in excess of 100 percent of quota." [13].

The organisation of this term project is as follows. Chapter 2 includes a brief overview of the literature in credit scoring and predicting debt crisis. Chapter 3 contains the fundamentals of CART, MARS and CMARS which we have applied in our work. Moreover, the advantages and disadvantages of each method is given in this chapter. In Chapter 4, we have compared and discussed the success of each model in the estimation of debt defaults. Finally, the conclusion and related future work conclude this term project.

## CHAPTER 2

### HISTORY OF METHODS USED IN CREDIT SCORING AND PREDICTING FINANCIAL CRISIS

Default is the disability of satisfying obligations. When a default occurs, it loads very big costs. Therefore, the prediction of default has been a great concern of scientists since 1930s. First studies were usually focused on bankruptcy prediction and credit scoring of individuals [23]. The research in the area of credit scoring started in the 1930's. This period can be called as a primitive age of credit scoring because this includes very basic applications. In this part, research was based only on a ratio analysis. In those years, scientists compared ratios of default and non-default companies and tried to develop an idea of companies financial performances. As it can be guessed, these types of methods had no predictive power and so they were not very suitable [24].

After that date, many different works and methods have entered the literature of credit scoring. According to the types of methods, we can split the credit scoring methods into two groups [40, 28].

#### **Parametric credit scoring methods:**

- Linear regression model,
- Probit and Logit models,
- Discrimination analysis-based models,

#### **Non-parametric credit scoring models:**

- Neural Networks,
- Classification and Regression Trees (CART),
- Multivariate Adaptive Regression Splines (MARS).

In 1936, Fisher [14] first proposed discriminant analysis as a classification technique. Up to date, it has been reported as the most commonly used technique in modeling classification and the credit scoring problems [33, 51]. Since discriminant analysis is

mainly used for identifying the linear relationship between variables, the method has some weaknesses upon classification. Also discriminant analysis requires normally distributed data but the credit data are generally not normal and categorized. It follows from this fact that new methods are developed for credit scoring.

To overcome the normality requirements of discriminant analysis, researchers use an extension of the linear discriminant analysis which is known as logit model. The logistic regression model does not necessarily require the assumptions of some other regression models, like assuming that the variables are normally distributed in discriminant analysis [35]. Usually, both models (probit and logit) are estimated using the maximum likelihood method and these methods can deal with categorized data. As a result of this situation, several studies found that logit model outperforms discriminant analysis [55], and more recently, logit analysis became one of the main approaches of classification in credit scoring practices of banks. Nevertheless, logit models are very sensitive to high correlation among explanatory variables like discriminant analysis. Another disadvantage of this method is the sensitivity of missing values (all observations with missing values have to be deleted) [52].

Examples of the application of logistic regression can be given as follows. In 1995, Lawrence and Arshadi [31] used the logit model for the analysis of the management of problem loans and of determinants of resolution choices using a series of borrower and bank variables. In the area of mortgage lending, Campbell and Dietrich [3] utilized a logit model to show that the age of a mortgage, the loan-to-value ratio, interest rates, and unemployment rates are significant in explaining mortgage prepayments, delinquencies and defaults. In 1989, Gardner and Mills [19] recognizing that delinquent borrowers do not necessarily end up in default, employ a logit regression model to estimate the probability of default for currently delinquent loans. They recommend that bankers use this method to identify the severity of the problem and thereby formulate an appropriate response to the delinquency.

In addition to discriminant analysis and logistic regression, there has been a development of neural networks approach in credit scoring. Neural networks provide an alternative to discriminant analysis and logistic regression, particularly in situations where the dependent and independent variables exhibit complex nonlinear relationships.

Even though neural networks have been reported to have better credit scoring capability than discriminant analysis and logistic regression [8, 25, 39, 54], they are also being criticized for their long training process in designing the optimal network's topology and inability to identify the relative importance of potential input variables, as a result of which they are limited its applicability in handling credit scoring problems [38].

Because of the above-mentioned drawbacks of discriminant analysis, logistic regression, and neural networks, two commonly discussed data mining techniques, classification and regression tree (CART) and multivariate adaptive regression splines (MARS) attract the researchers attention to these methods in credit scoring. The rationale to use CART and MARS in credit scoring is fourfold. Firstly, unlike discriminant analysis and logistic regression, both CART and MARS approaches exhibit the capability of modeling complex relationship between variables without strong model assumptions. Besides, unlike neural networks, they both are able to identify "important" independent variables through the built tree and basis functions when many potential variables are considered. Thirdly, CART and MARS do not need a long training process and hence can save lots of modeling time when the data set is huge. Finally, one strong advantage of CART and MARS over other classification techniques is that the resulting classification model can be easily interpreted. It not only points out which variables are important in classifying objects/observations, but also indicates that a particular object/observation belongs to a specific class when the built rules are satisfied [32].

Examples of the applications of MARS and CART in credit scoring can be given as follows. The first to use the CART method in the credit scoring area were Frydman, Altman and Kao [18] in 1985 who found it to outperform discriminant analysis. Relevant to retail lending is the study by Devaney [9] in 1994, who used logit and CART methods to choose which financial ratios are the best predictors of households default.

If we want to look at the recent studies; in 2006, Lee, Chiu, Chou and Lu [32] performed credit scoring task worked on one bank credit data set in order to evaluate the feasibility and effectiveness of using CART and MARS in building credit scoring models. They found that CART and MARS both have better average correct

classification rate in comparison with discriminant analysis, logistic regression, neural networks. Besides, CART and MARS not only have better credit scoring accuracies, but also lower Type II errors associated with high misclassification costs and therefore have better overall credit scoring capabilities. The research findings provide efficient alternatives in conducting credit scoring tasks.

In 2002, Kolyshkina and Brookes [29] use CART to evaluate insurance risks in workers compensation and hospital costs. In the first case they find that CART performs better than logistic regression and in the second case they use MARS (Multivariate Adaptive Regression Splines) a modification of the CART methodology designed to improve performance where the response is continuous rather than binary or categorical.

In 2006, Chen, Song and Ji [4] compared MARS and CART in credit scoring using a database of personal credit records in a city of China. The data are composed of 3216 sets of records, 868 of which are bad credit ones, and all of the data is classified into 14 characteristic variables, including registration of population, age, gender, marriage status, fixed phone number, cell phone number, occupation, post, position, technical title, education background, credit card information and loan information. They found that the classification accuracy rate of MARS model is a bit higher than that of CART model. But the difference is small, and MARS model is more robust than CART model.

In 2006, Satchidananda and Simha [41] used the data from two banks in India pertaining to the agricultural production loans given to farmers in and around Honavar, a backward block in Karnataka, India. In this paper, they evaluated and contrasted decision tree classifiers with logistic regression classifiers for credit scoring. The evaluation was done by looking at the performance in terms of classification accuracy and the complexity of the trained classifiers. It was found that the decision tree classifiers had a good performance and outperformed the logistic regression on all the parameters.

The increase in debt crisis in emerging markets after 1980s also aroused the interests of scientists in the area of sovereign defaults [23]. Financial crises occurred in emerging countries in the last decade of 20th century have revived theoretical and empirical

interest in the topic in order to understand their causes and consequences as well as to develop statistic and econometric models that can timely signal their occurrence [13].

In the last decade, many empirical studies have concentrated their attention in developing models able to timely signal the occurrence of a financial crisis, the so-called *early warning system* (EWS). Using statistical and econometric techniques these models are applied to predict the likelihood of financial crises using a wide number of indicators related to internal and external factors, as well as social and political condition [13].

According to the type of approach, models can be classified between parametric and non-parametric. In 1996, Frankel and Rose [16] and in 1998, Kaminsky, Lizondo and Reinhart [27] are the seminal papers in the two classes of approach applied to currency crises prediction. Using a probit, Frankel and Rose estimate the probability of currency crises for more than 100 developing countries from 1971 to 1992, finding that crises occur when GDP (gross domestic product) and the ratio of FDI (foreign direct investment) to external debt are low, while growth of domestic credit and foreign interest rates are high. Kaminsky et al. proposed the non-parametric “signal approach” which involves monitoring the evolution of a number of economic indicators that show a behaviour which is different in tranquil period and prior to a crisis. When an indicator exceeds a particular threshold, this is interpreted as a signal that a crisis could occur in the following 24 months. Using a dataset of 23 countries from 1970 to 1995 and a wide set of indicators, they show that international reserves, real exchange rate, domestic credit, credit to public sector, and domestic inflation are very useful in signalling a crisis [13].

In 2006, Fioramanti [13] shows that Artificial Neural Networks (ANN) with only two layers can outperform a traditional Early Warning System in predicting a sovereign debt crisis, if one chooses the right number of hidden unit, training epochs and an efficient training algorithm. This is possible because if the relation that links debt crisis indicator and explanatory variables is highly nonlinear, the flexibility of ANNs should give, theoretically, results that are at least as good as those of the parametric traditional methods [13].

The purpose of this study is to explore the performance of the estimation of debt defaults using two commonly discussed data mining techniques, classification and regression tree (CART) and multivariate adaptive regression splines (MARS), without confronting with the drawbacks of discriminant analysis, logistic regression, and neural networks.

# CHAPTER 3

## METHODS

### 3.1 Multivariate Adaptive Regression Splines

*Multivariate Adaptive Regression Splines (MARS)*, an implementation of techniques developed in 1991 by Friedman [17], is an adaptive regression procedure for solving high dimensional problems where there are many explanatory variables. Multivariate adaptive regression spline (MARS) denotes a tool from statistics, important in classification and regression, with applicability in many areas of finance, science and technology. It is very useful in high dimensional problems and shows a great promise for fitting nonlinear multivariate functions [47]. The first word multivariate expresses that MARS is able to deal with multidimensional data, examine individual features and possible interactions among them. The second word adaptive simply means selective. MARS automatically deletes certain number of predictors if they do not contribute enough to the performance of the final model. The next word regression refers to the normally used statistical term, which is often represented as a general prediction function (linear case):

$$y = \beta_0 + \sum_{j=1}^p \beta_j x_j, \quad (3.1.1)$$

where  $y$  is the target value,  $\beta_0$  is the constant term,  $\beta_j$  are the coefficients and  $x_j$  are the predictor values. The last word “splines” indicates a wide class of piecewise defined functions that are used in applications requiring data interpolation and/or smoothing. In order to develop a spline, the original space is divided into a conventional number of regions [53].

MARS is a nonparametric regression procedure that makes no assumption about the underlying functional relationship between the dependent and independent variables. The MARS technique has become particularly popular in the area of data mining because it does not assume or impose any particular type or class of relationship

(e.g., linear, logistic, etc.) between the predictor variables and the dependent (outcome) variable of interest. Instead, useful models (i.e., models that yield accurate predictions) can be derived even in situations where the relationship between the predictors and the dependent variables is non-monotone and difficult to approximate with parametric models [5].

The MARS procedure can be thought of as a generalisation of stepwise regression. MARS not only uses a stepwise procedure to introduce and delete explanatory variables, but very importantly also considers transformations and interactions between the variables. This relationship between the variables is totally data driven. The MARS algorithm works by partitioning each of the explanatory variables into regions, with each region having its own regression equation. MARS has been shown to be competitive with neural network models [10] especially if the data contain low order interactions [20]. Also a special advantage of MARS lies in its ability to estimate the contributions of the basis functions so that both the additive and the interactive effects of the predictors are allowed to determine the response variable [48].

The MARS method generates a model in a two-stage process: forward and backward. In the first stage, MARS constructs an extra large number of basis functions (BFs), which deliberately overfit the data. These BFs represent distinct intervals of every predictor divided by knots, and in an intensive search, every possible knot location is tested. The MARS model is actually, in each dimension, a linear summation of certain BFs, and interactions among them if needed. Then, some of the BFs are removed as they contribute least to the overall performance. Therefore, the forward construction initially includes many incorrect terms. In the backward pruning step, these erroneous terms are eventually excluded. Thus, the backward step reduces the complexity of the model without degrading the fit to the data. By allowing arbitrary shapes of BFs and their interactions, MARS has the capacity of reliably tracking very complex data structures that often hide in high dimensions [53, 11].

### **3.1.1 The Approach**

MARS is a nonparametric that essentially builds flexible models by fitting piecewise linear regressions; that is, the nonlinearity of a model is approximated through the use

of separate linear regression slopes in distinct intervals of the independent variable space. Therefore, the slope of the regression line is allowed to change from one interval to the other as the two ‘knot’ points are crossed. The variables to be used and the end points of the intervals for each variable are found through a fast but intensive search procedure. In addition to searching variables one by one, MARS also searches for interactions between variables, allowing any degree of interaction to be considered as long as the built model can better fit the data [34].

MARS can be expressed in an expanded form of the piecewise linear basis function,  $(x - t)_+$  and  $(t - x)_+$  with a knotting value at  $t$ . The following two functions are truncated where  $x \in \mathbb{R}$  [5] :

$$(x - t)_+ = \begin{cases} x - t, & \text{if } x > t, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$(t - x)_+ = \begin{cases} t - x, & \text{if } x < t, \\ 0, & \text{otherwise.} \end{cases} \tag{3.1.2}$$

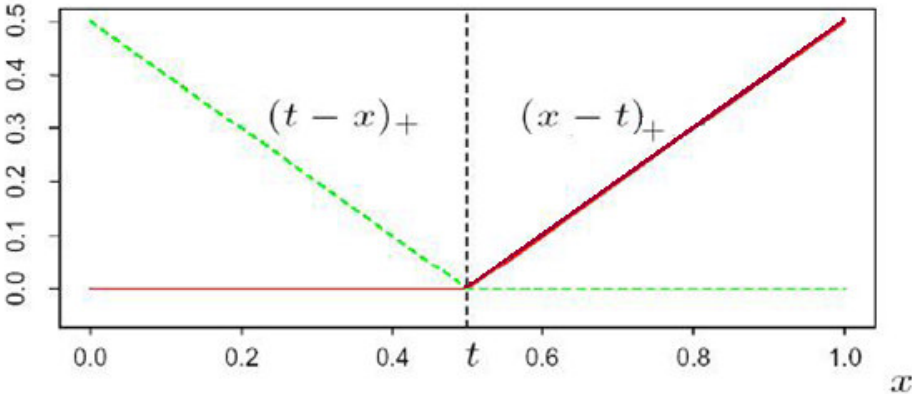


Figure 3.1: The BFs  $(x - t)_+$  and  $(t - x)_+$  used by MARS [22].

In Figure 3.1, each function is piecewise linear whose knot is on the point of value  $t$ . The equation  $(.)_+$  means that only the positive parts are used, if not it is given a zero value. These two functions are named as a reflected pair. Here, as to every input  $x_{ij}$ , each input  $x_j$  with knot constitutes a reflected pair. So the collection of the BFs is [4]:

$$C := \left\{ (x_j - t)_+, (t - x_j)_+ \mid t \in \{x_{1j}, x_{2j}, \dots, x_{Nj}\}, j \in \{1, 2, \dots, p\} \right\}. \tag{3.1.3}$$

If each input value is not equal to any other one, there will be  $2Np$  BFs totally. Although every BF is only related to single  $x_j$ , we can still consider it a function over the whole input space  $\mathbb{R}^p$  [4].

The usual method for generalizing spline fitting in higher dimensions is to employ BF that are the tensor products of univariate spline functions. Therefore, multivariate splines BFs take the following form:

$$B_m(\mathbf{x}) = \prod_{k=1}^{K_m} [s_{km} \cdot (x_{v(km)} - t_{km})]_+, \quad (3.1.4)$$

where  $K_m$  is the total number of truncated linear functions in the  $m$ th BF,  $x_{v(km)}$  is the input variable corresponding to the  $k$ th truncated linear function in the  $m$ th basis function,  $t_{km}$  is the corresponding knot value and  $s_{km} \in \{\pm 1\}$  [53].

The way to construct model is analogous to forward stepwise linear regression, but it allows the use of functions and their products from the set  $C$ , not the initial value. The form of the model is as follows:

$$f(\mathbf{x}) = \beta_0 + \sum_{m=1}^M \beta_m B_m(\mathbf{x}) + \epsilon, \quad (3.1.5)$$

where each  $B_m(\mathbf{x})$  is a function from the set  $C$ , or the product of two or more functions from  $C$  [4].

Given  $B_m$  a choice, coefficient  $\beta_m$  could be estimated by minimizing the sum of squares of residual differences, that is, by standard linear regression. However, the important issue is how to structure the function  $B_m(\mathbf{x})$ . Firstly, there is only a constant function  $B_0(\mathbf{x}) = 1$ , and the other functions in set  $C$  are candidate functions [4].

At each phase, we consider the product of the function  $B_m$  in model collection  $\mathcal{M}$  and the function of its reflective pair in  $C$  a new basis function pair. Then putting the following formula as an item into the model  $\mathcal{M}$ , we get:

$$\hat{\beta}_{M+1} B_l(\mathbf{x})(x_j - t)_+ + \hat{\beta}_{M+2} B_l(\mathbf{x})(t - x_j)_+, \quad B_l \in \mathcal{M}. \quad (3.1.6)$$

This procedure minimizes training errors. Here,  $\hat{\beta}_{M+1}$  and  $\hat{\beta}_{M+2}$  are the coefficients which can be estimated by least squares along with other  $M + 1$  coefficients. The better product can be put into the model. This process will restart until the number of items in the collection  $\mathcal{M}$  reaches the given maximum number [4].

For example, the following BFs are possible candidates [30]:

- $x_j, j = 1, 2, \dots, p,$
- $(x_j - t_k)_+,$  if  $x_j$  is already in the model,
- $x_l x_j,$  if  $x_l$  and  $x_j$  are already in the model,
- $(x_j - t_k)_+ x_l,$  if  $x_l x_j$  and  $(x_j - t_k)_+$  are already basis functions,
- $(x_j - t_k)_+ (x_l - t_h)_+,$  if  $(x_j - t_k)_+ x_l$  and  $(x_l - t_h)_+ x_j$  are already in the model.

This process can also be figured below as follows. The figures in the left hand side are the basis functions of the current model and right hand side are the alternative basis functions when constructing the model. Firstly in the left column, we have initially a constant function  $B_0(\mathbf{x}) = 1$ . At the end of this process, the ‘‘total model’’ is obtained [4].

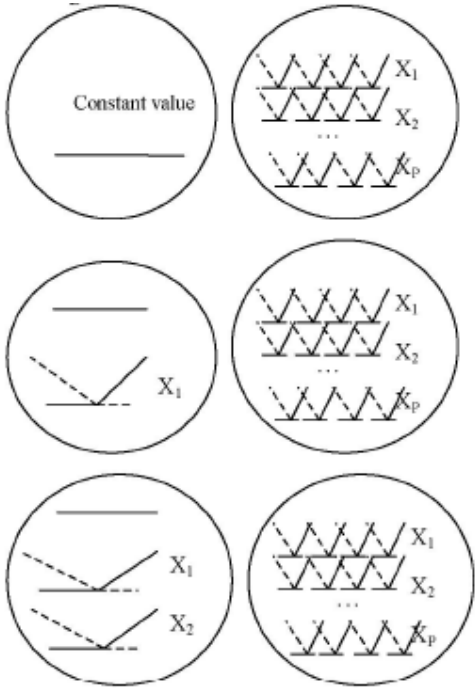


Figure 3.2: MARS forward model construction process [22].

The total model normally overfits the data, in order to avoid from this we start the backward elimination process. In each step we delete the term which causes the smallest increase in the residual squared error. We continue until we attain the best model  $\hat{f}_\alpha$  with the optimal value  $\alpha$ .

In order to estimate the optimal value of  $\alpha$  reducing computational costs, *generalized cross-validation* is used. This formulation, also known as lack of fit criterion, is defined as [45]:

$$GCV_{Friedman} = \frac{\sum_{i=1}^N (y_i - \hat{f}_\alpha(x_i))^2}{\left(1 - \frac{M(\alpha)}{N}\right)^2}, \quad (3.1.7)$$

where  $N$  is the number of data samples, and  $M(\alpha)$  is the number of valid parameters being fit.

The numerator of  $GCV_{Friedman}$  is the known common residual sum of squares, which is penalized by the denominator. This denominator accounts for the increasing variance in the case of increasing the model complexity [53].

If there are  $r$  linearly independent functions in the model,  $M(\alpha) = r + cK$  where  $K$  represents the number of knots which are selected in the forward process and, the quantity  $c$  shows a cost for each BF optimization and generally,  $c = 3$  [22]. If the model is limited to addition model, a penalty of  $c = 2$  is used. A smaller  $M(\alpha)$  generates a larger model with more BFs, a larger  $M(\alpha)$  creates a smaller model with less BFs. Using lack of fit criteria, we choose the best model according to the backward sequence that minimizes  $GCV_{Friedman}$  [11, 22].

The MARS models in this study are fitted using MARS (Version 2, Salford Systems, San Diego, Calif., USA). MARS allows the user to set control parameters to explore different models and find the “best” model. The MARS package developed by Salford Systems is available at [5].

### 3.1.2 Advantages and Disadvantages of MARS

#### Advantages:

- Useful tool for simplifying high-dimensional problems where there are many explanatory variables.
- MARS is capable of identifying a relatively small number of predictor variables which are complex transformations of initial variables.
- MARS uses piecewise linear functions which produces continuous models and provides a more effective way to model nonlinearities.
- MARS is not computationally intensive and is straightforward to implement in order to look for suitable interactions between independent variables, which makes it in particular preferable whenever there is a large number of interacting variables [42].
- Both the additive and the interactive effects of the predictors are allowed to determine the response variable [42].
- MARS can handle complex (nonlinear) relationships and interactions providing an interpretable model.
- MARS identifies interactions and also produces graphs that help visualize and understand interactions.
- MARS has automated capabilities for handling missing data, a common feature of large databases.

#### Disadvantages:

- The MARS methodology has a risk of overfitting because of very exhaustive search that is conducted to identify nonlinearities and interactions.

- The dataset has to be large enough to make the use of MARS possible and worthwhile.

### 3.2 CMARS

The MARS algorithm for estimating the model function consists of two algorithms, these are the *forward* and the *backward stepwise algorithm* that mentioned in Section 3.1 in detail. In CMARS, instead of using the backward stepwise algorithm, we can construct a penalized residual sum of squares for MARS as a *Tikhonov regularization* problem which is also known as *ridge regression*. We treat this problem using *continuous optimization* techniques which we consider to become an important complementary technology and model based alternative to the concept of the backward stepwise algorithm. In particular, we apply the framework of *conic quadratic programming* [48].

Let use below formulation for the set of BFs:

$$S := \left\{ (X_j - \tau)_+, (\tau - X_j)_+ \mid \tau \in \{x_{1,j}, x_{2,j}, \dots, x_{N,j}\}, j \in \{1, 2, \dots, p\} \right\}. \quad (3.2.8)$$

If all of the input values are distinct, there are  $2Np$  BFs altogether. Thus we can represent  $f(\mathbf{X})$  by a linear combination which is successively built up by the set  $S$  and with the intercept  $\theta_0$  such that (3.2.8) takes the form

$$Y = \theta_0 + \sum_{m=1}^M \theta_m \psi_m(\mathbf{X}) + \varepsilon, \quad (3.2.9)$$

where  $M$  is the number of BFs suitable for the data [53]. Here,  $\psi_m$  ( $m = 1, 2, \dots, M$ ) represents a BF from  $S$  or products of two or more such functions,  $\psi_m$  is taken from a set of  $M$  linearly independent basis elements, and  $\theta_m$  is the unknown coefficient for the  $m$ th BF ( $m = 1, 2, \dots, M$ ) for the constant 1,  $m$  equals to zero [48].

Provided the observations represented by the data  $(x_i, y_i)$  ( $i = 1, 2, \dots, N$ ), the form of the  $m$ th basis function is as follows [48]:

$$\psi_m(\mathbf{x}) := \prod_{j=1}^{K_m} \left[ s_{K_m^j} \cdot (x_{K_m^j} - \tau_{K_m^j}) \right]_+, \quad (3.2.10)$$

where  $K_m$  is the number of truncated linear functions multiplied in the  $m$ th basis function,  $x_{K_m^j}$  is the input variable corresponding to the  $j$ th truncated linear function in the  $m$ th basis function,  $\tau_{K_m^j}$  is the knot corresponding to the variable  $x_{K_m^j}$ , and  $s_{K_m^j}$  is the selected sign +1 or -1.

In the previous section, in order to estimate the optimal value of  $\alpha$ , which shows the complexity of our estimation, we use *generalized cross-validation*. This formulation, also known as lack of fit criterion, which is also mentioned in the previous section, is defined for CMARS as follows:

$$GCV := \frac{1}{N} \frac{\sum_{i=1}^N (y_i - \hat{f}_\alpha(x_i))^2}{(1 - M(\alpha)/N)^2}, \quad (3.2.11)$$

where  $M(\alpha) := u + dK$ ,  $\alpha$  depending on  $(u, d, K)$  [6]. Here,  $N$  is the number of sample observations,  $u$  is the number of linearly independent BFs,  $K$  is the number of knots selected in the forward process and  $d$  is the cost of BF optimization.

### 3.2.1 Tikhonov Regularization and the Construction of the Conic Quadratic Programming

From now on, instead of running the backward stepwise algorithm, we use the penalized residual sum of squares with (PRSS) with  $M_{max}$  BFs having been accumulated in the *forward* stepwise algorithm in order to control the the lack of fit from the point of the complexity of the estimation [53]. For the MARS model, PRSS has the following form:

$$PRSS := \sum_{i=1}^N (y_i - f(\bar{x}_i))^2 + \sum_{m=1}^{M_{max}} \lambda_m \sum_{\substack{|\alpha|=1 \\ \alpha=(\alpha_1, \alpha_2)}}^2 \sum_{\substack{r < s \\ r, s \in V(m)}} \int \theta_m^2 [D_{r,s}^\alpha \psi_m(\mathbf{t}^m)]^2 d\mathbf{t}^m, \quad (3.2.12)$$

where  $V(m) = \{K_j^m | j = 1, 2, \dots, K_m\}$  is the variable set associated with the  $m$ th basis function  $\psi_m$ ,  $\mathbf{t}^m = (t_{m_1}, t_{m_2}, \dots, t_{m_{K_m}})^T$  represents the vector of variables which contribute to the  $m$ th basis function  $\psi_m$ . Moreover,

$$D_{r,s}^\alpha \psi_m(\mathbf{t}^m) = \frac{\partial^\alpha \psi_m}{\partial \alpha_1 t_r^m \partial \alpha_2 t_s^m}(\mathbf{t}^m),$$

for  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$ ,  $|\boldsymbol{\alpha}| = \alpha_1 + \alpha_2$ , where  $\alpha_1, \alpha_2 \in \{0, 1\}$ . Our optimization problem bases on the tradeoff between both accuracy, i.e., a small sum of error squares, and not too high a complexity. This tradeoff is established through the penalty parameters  $\lambda_m$  [53].

It follows from [53, 47, 48] that PRSS can be defined as follows:

$$PRSS = \|\mathbf{y} - \psi(\bar{\mathbf{d}})\boldsymbol{\theta}\|_2^2 + \sum_{m=1}^{M_{max}} \lambda_m \sum_{i=1}^{(N+1)^{K_m}} L_{im}^2 \theta_m^2, \quad (3.2.13)$$

where  $\psi(\bar{\mathbf{d}}) := (\psi(\bar{\mathbf{d}}_1), \dots, \psi(\bar{\mathbf{d}}_N))^T$  is an  $(N \times (M_{max} + 1))$ -matrix,  $\|\cdot\|_2$  denotes the Euclidean norm and the numbers  $L_{im}$  are defined by their roots

$$L_{im} := \left[ \left( \sum_{\substack{|\boldsymbol{\alpha}|=1 \\ \boldsymbol{\alpha}=(\alpha_1, \alpha_2)}}^2 \sum_{\substack{r < s \\ r, s \in V(m)}} \int [D_{r,s}^{\boldsymbol{\alpha}} \psi_m(\hat{\mathbf{x}}_i^m)]^2 \right) \Delta \hat{\mathbf{x}}_i^m \right]^{1/2}. \quad (3.2.14)$$

For each derivative in (3.2.13) instead of using distinct penalty parameters we can use only one penalty parameter say  $\lambda = \lambda_m (= \varphi^2)$  and write PRSS as follows [6]:

$$PRSS = \|\mathbf{y} - \psi(\bar{\mathbf{d}})\boldsymbol{\theta}\|_2^2 + \lambda \|\mathbf{L}\boldsymbol{\theta}\|_2^2, \quad (3.2.15)$$

where  $\mathbf{L}$  is a diagonal  $(M_{max} + 1) \times (M_{max} + 1)$ -matrix. Then our PRSS problem looks as a classical *Tikhonov regularization problem* with  $\varphi > 0$ , i.e.,  $\lambda = \varphi^2$  for some  $\varphi \in \mathbb{R}$  [46].

We can tackle the Tikhonov regularization problem with *conic quadratic programming (CQP)* which is a continuous optimization technique [37]. We can formulate PRSS as a CQP problem based on an appropriate choice of a bound  $\bar{M}$  as follows [48]:

$$\begin{aligned} \min_{\mathbf{t}, \boldsymbol{\theta}} \quad & t, \\ \|\psi(\bar{\mathbf{d}})\boldsymbol{\theta} - \mathbf{y}\|_2 & \leq t, \\ \|\mathbf{L}\boldsymbol{\theta}\|_2 & \leq \sqrt{\bar{M}}. \end{aligned} \quad (3.2.16)$$

### 3.2.2 MOSEK

To solve “well-structured” convex problems like conic quadratic problem, *interior point methods* can be applied [37]. In order to solve the CQP problem, we use MOSEK optimization tool [43] and transform the problem in (3.2.15) into the MOSEK format.

The MOSEK, which is a MATLAB add-on, is an optimization tool for solving large-scale mathematical optimization problems [43]. For its excellent speed and stability, the MOSEK interior-point optimizer is well known [53]. The interior-point optimizer is an implementation of the homogeneous and self-dual algorithm and it computes the interior point solution which is an arbitrary optimal solution [53]. The values  $\sqrt{M}$  in our optimization problem are determined by a *model-free* (train and error) method. When we access the  $\sqrt{M}$  values in our CMARS code, CMARS provides us several solutions [53].

The related study on generalized additive models, spline regression and optimization methods are prepared and implemented by the Institute of Applied Mathematics [49] successfully under numerical-statistical criteria [23]. Code development and use of our contribution on MARS is a Project at METU, Ankara. We invite the reader to find a lot of closer information and examples by visiting the two links <http://144.122.137.55/gweber/> and <http://www.iam.metu.edu.tr/mscthesis/fatmayerlikayathesis.pdf>.

### 3.2.3 Advantages and Disadvantages of CMARS

#### Advantages:

- CMARS provides its solutions by using CQP. In this respect, it has the advantage of speed and complexity [37].
- CMARS offers alternative models for decision makers.
- Permits the use of powerful interior point methods [48].

### **Disadvantages of CMARS:**

- Unlike MARS, CMARS does not select a best model.
- Computational time is long with respect to MARS. CMARS does not represent the results as a model form. This takes time when compared with MARS software [53].

### **3.3 Classification and Regression Tree (CART)**

*Classification and Regression Tree (CART)*, a statistical procedure introduced by Breiman, Frydman, Olshen and Stone [2], is primarily used as a classification tool, where the objective is to classify an object into two or more populations. We know that the objective of credit scoring models is to assign credit applicants to either a “good credit” group that is likely to repay financial obligation or a “bad credit” group who has high possibility of defaulting on the financial obligation. Therefore, credit scoring problems are basically in the scope of the more general and widely discussed classification problems [26].

CART is a tool for analyzing both continuous and categorical variables. For categorized dependent variables, the result of the CART approach is a classification tree and for continuous dependent variables, the result of the CART approach is a regression tree. In both cases CART produces a binary classification tree and in credit scoring since we have a categorical dependent variable we focus on classification trees.

Basic idea of CART is constructing a tree that will separate the data in the "best" way by finding binary splits on variables; finding *the best splitting variable* and *the best splitting point* at each stage with minimizing diversity.

CART algorithm has a four-step classification procedure which are as follows [24]:

***Step 1:*** *a criterion of diversity,*

***Step 2:*** *goodness of split criterion,*

***Step 3:*** *the class assignment rule* to the terminal nodes and resubstitution estimates,

***Step 4:*** *selecting the correct complexity* of a tree.

### 3.3.1 Criterion of Diversity

The CART binary tree consists of a root node, internal nodes and leaf (terminal) nodes. Each root and internal node is a parent node with two daughter nodes. Each node, say  $t$ , is described by the subset of the original learning sample that it contains. For all but the leaf nodes, this subset is divided into two groups, going to daughter nodes  $t_l$  and  $t_r$ . The split at each node is described by a rule that depends on one selected feature. Let this feature be  $X$ , and assume, first, that the  $X$  is categorical. The splitting rules are of the form  $X \in S$  or  $X \notin S$ , where  $S$  is some nonempty subset of  $X$ 's possible categories. When  $X$  is qualitative with  $L$  categories, CART tries all possible distinct binary splits,  $2^{L-1}$  in number. At each node of the tree the program searches through the features one by one, determines the best split for each  $X$ , and then the best  $X$  to split on at that node. In other words, the feature  $X$  is selected among all possible ones, and  $S$  is selected among all possible splits, with a view towards minimizing the *diversity* of the resulting subsamples forwarded to the two daughter nodes. Diversity of a subsample, roughly speaking, is a measure of its heterogeneity [12].

The CART algorithm is a recursive procedure; starting at the root node, and then at every internal node, it selects a single feature, and a threshold value  $s$  to split the group of individuals at the node into two groups to be placed at two new daughter nodes. CART grows the largest tree possible, called a maximal tree, that is the tree whose leaves (terminal nodes) cannot be split any further. A node may not be split any further either because it contains only cases that belong to a single class, or because no reduction in total diversity can be obtained by further splitting [12].

### 3.3.2 Goodness of Split Criterion

The goodness of split criterion is applied to the each split point at a node in order to select best split point for each variables and then, for node. The goodness of split criterion is an index based on impurity functions [24]. A splitting rule is derived from a diversity function called impurity function. We can define the impurity function and the belonging impurity measure as follows [2]:

**An impurity function** is a function  $\phi$  defined on the set of  $J$  classes with prior probability vector  $\delta = (\delta_1, \delta_2, \dots, \delta_j)$  satisfying  $\delta_j \geq 0$  ( $j = 1, 2, \dots, J$ ) and  $\sum_{j=1}^J \delta_j = 1$  with the following properties:

- (i)  $\phi$  has a maximum only at the point  $\delta = (\frac{1}{J}, \dots, \frac{1}{J})$ ,
- (ii)  $\phi$  achieves its minimum only at the points  $\delta = (1, 0, \dots, 0)$ ,  $\delta = (0, 1, \dots, 0)$ ,  $\dots$ ,  
 $\delta = (0, 0, \dots, 1)$ ,
- (iii)  $\phi$  is a symmetric function of  $\delta_1, \delta_2, \dots, \delta_j$ .

Given an impurity function  $\phi$ , define **the impurity measure**  $i(t)$  of any node  $t$  as

$$i(t) := \phi(p(1|t), p(2|t), \dots, p(J|t)).$$

CART offers several splitting methods. The best known rules for binary recursive partitioning are **Gini**, **Twoing** and **Entropy** [2].

Before starting these splitting methods, it is better we define some probabilities related to search material given from the studies [24, 2, 18].

Let  $N$  be the total learning sample size, and let  $N_{ND}$  be the number of non-default firms in our learning sample and  $N_D$  be the number of default firms.

Suppose CART has  $t = 1, 2, \dots, T$  nodes. In node  $t$ , there are  $N(t)$  observations. Let  $N_{ND}(t)$  represent the number of non-default firms in node  $t$  and similarly,  $N_D(t)$  represents the number of default firms.

Let  $p(ND, t)$  be the probability of a firm is non-default firm and falls into node  $t$ . Similarly,  $p(D, t)$  be the probability of a firm is default firm and falls into node  $t$ . So we can write these probabilities as

$$p(ND, t) = \delta_{ND} \frac{N_{ND}(t)}{N_{ND}}, \quad (3.3.17)$$

$$p(D, t) = \delta_D \frac{N_D(t)}{N_D}, \quad (3.3.18)$$

where  $\delta_{ND}, \delta_D$  stands for the prior class probabilities.

Also the probability that an observation falls into node  $t$  is

$$p(t) = \sum_{j=ND}^D p(j, t). \quad (3.3.19)$$

Following Bayes rule, the probability of a firm in the node  $t$  to be a non-default firm is denoted by the following equality

$$p(ND|t) = \frac{p(ND, t)}{p(t)}. \quad (3.3.20)$$

Similarly the probability that a firm in the node  $t$  to be a default firm is

$$p(D|t) = \frac{p(D, t)}{p(t)} \quad (3.3.21)$$

and

$$\sum_{j=ND}^D p(j|t) = 1. \quad (3.3.22)$$

### **Gini Index**

*Gini index* is a quadratic impurity measure. Given the the left(l) and right(r) node probabilities  $p(t_l)$  and  $p(t_r)$  and impurities  $i(t_l)$  and  $i(t_r)$  the change in the impurity at node  $t$  is [2, 7]:

$$\Delta i(t, s) := i(t) - p(t_l)i(t_l) - p(t_r)i(t_r), \quad (3.3.23)$$

where

$$i(j) := p(ND|j)p(D|j). \quad (3.3.24)$$

Given the way how Gini index is computed it becomes obvious that this impurity measure can be quite effective. First, it is relatively cheap in terms of computation speed and second Gini index is relatively robust to the effect of outliers; a few outliers can not drastically change the values of  $p(j|t)$  [1].

## Entropy Index

Like Gini index, the *entropy index* criterion is a nonlinear measure of impurity.

According to [7], the rule of entropy index is to

$$\max \Delta i(t, s) := i(t) - p(t_l)i(t_l) - p(t_r)i(t_r),$$

where

$$i(j) := -p(ND|j)\ln(p(ND|j)) - p(D|j)\ln(p(D|j)). \quad (3.3.25)$$

## Twoing Index

The *twoing function*, with daughter nodes  $t_l$  and  $t_r$ , and where the probabilities  $p(t_l)$  and  $p(t_r)$  are the proportions of cases going to nodes  $t_l$  and  $t_r$ , respectively, is

$$\Delta i(t, s) := \frac{p(t_l)p(t_r)}{4} \left( \sum_{j=D}^{ND} |p(j|t_l) - p(j|t_r)| \right)^2. \quad (3.3.26)$$

We remark that the Entropy and the Gini index diversity functions refer to the diversity of cases at a given node. Therefore as a tool for splitting cases at a node, a change in diversity from that of the parent node, to the sum of diversity at the daughter nodes is required. The twoing function, on the other hand, measures a class-prevalence distance between the daughter nodes, anticipating that the diversity within the daughter nodes will decline when the split achieves a higher degree of difference in the prevalence of the different classes in the two daughter nodes. Thus, to achieve the highest reduction in diversity, one chooses the split  $s$  that maximizes the twoing function [12].

### 3.3.3 The Class Assignment Rule to the Terminal Nodes and Resubstitution

#### Estimates

After the goodness of split step, we obtain best splits for each node and get the maximal tree say  $T_{max}$ . Let  $\tilde{T}$  denote the set of terminal nodes.

**A class assignment rule** assigns a class  $j \in \{1, 2, \dots, J\}$  to every terminal node  $t \in \tilde{T}$ . The class assigned to node  $t \in \tilde{T}$  is denoted by  $j(t)$  [2].

The class assignment rule requires the minimization of cost of misclassification of a node one of the classes. In CART analysis, the observed expected cost of misclassification of each assignment is called the resubstitution risk [18]. Let we denote the resubstitution risk of assigning a terminal node  $t \in \tilde{T}$  to  $j^{th}$  by  $R_j(t)$ .

Suppose that the misclassification cost of classifying a default firm as non-default be  $c(ND|D) \geq 0$  and the misclassification cost of classifying a non-default firm as default be  $c(D|ND) \geq 0$ . Then the resubstitution risk of classifying the observation which falls into the terminal node  $t$  as a nondefault firm can be obtained as follows [24]:

$$R_{ND}(t) = c(ND|D)p(D, t) \quad (3.3.27)$$

$$= c(ND|D)p(t|D)\delta_D \quad (3.3.28)$$

$$= c(ND|D)\delta_D \frac{N_D(t)}{N_D}. \quad (3.3.29)$$

In the same manner,

$$R_D(t) = c(D|ND)\delta_{ND} \frac{N_{ND}(t)}{N_{ND}}. \quad (3.3.30)$$

It follows that our class assignment rule  $j(t)$  will be

$$j(t) = \begin{cases} ND, & \text{if } R_D(t) \geq R_{ND}(t), \\ D, & \text{otherwise.} \end{cases} \quad (3.3.31)$$

This means, if the resubstitution risk of assigning node  $t$  to the class of non-default firms is greater or equal to that of default firms assign to node  $t$  to the class of default firms. Otherwise, we assign node  $t$  to the class of non-default firms [24].

### 3.3.4 Selecting the Correct Complexity of a Tree

The idea of selecting the correct complexity of a tree is that the maximal tree is hard to interpret and usually overfits the data although it will be able to make better predictions with small resubstitution risks. On the other hand, small trees have a limited predicting abilities without getting a penalty for its size. Hence, we have to select a tree by optimizing the correct complexity with smaller resubstitution risk [24].

The resubstitution risk of of any tree being a subtree of the  $T_{max}$  is

$$R(\tilde{T}) = \sum_{t \in \tilde{T}} R(t), \quad (3.3.32)$$

where  $\tilde{T}$  is a set of terminal nodes. For any subtree  $T \leq T_{max}$ , define the number of terminal nodes  $|\tilde{T}|$  as a measure of its complexity. Then the complexity measure that is used to optimize the size of the decision tree is

$$R_\alpha(T) := R(T) + \alpha|\tilde{T}|, \quad (3.3.33)$$

where  $\alpha \geq 0$  is a complexity parameter and  $\alpha|\tilde{T}|$  is cost component [1]. For a given  $\alpha$ , our decision rule is to find the optimal tree which minimizes the complexity measure  $R_\alpha(T)$ . The more complex is the tree (the higher is the number of terminal nodes) – the lower is  $R(T)$  but, at the same time, the higher is the penalty  $\alpha|\tilde{T}|$ .

Although  $K$  can have infinite number of values, the number of subtrees of  $T_{max}$  resulting in minimization of  $R_\alpha(T)$  is finite. Hence, pruning of  $T_{max}$  leads to creation of subtrees sequence  $T_1, T_2, T_3, \dots$  with a decreasing number of terminal nodes. Since the sequence is finite, if  $T(\alpha)$  is an optimal subtree for some arbitrary  $\alpha$ , then it will remain optimal until complexity parameter is not changed to some  $\alpha'$  when  $T(\alpha')$  becomes a new optimal subtree until complexity parameter value is  $\alpha''$  and so on [1].

The CART based classification model in this study is fitted using MATLAB 7 Statistics Toolbox. Gini index is used for the splitting criteria, because it is relatively cheap in terms of computation speed and is relatively robust to the effect of outliers.

### 3.3.5 Advantages and Disadvantages of CART

#### Advantages:

- CART is a nonparametric method that makes no distributional assumptions of any kind for dependent and independent variables.
- CART is well suited to mixed data types that is the explanatory variables can be a mixture of categorical, continuous and interval [24].

- CART deals effectively with large data sets and the issues of higher dimensionality.
- CART is computationally fast and the interpretation of results summarized in a tree is more easily understood than other model constructions like mathematical expressions and nonlinear equations.
- CART can handle multicollinearities and outliers that effect parametric procedures [42,44].
- CART has the ability to detect and reveal variable interactions in the data set.
- It handles missing data well [44].

**Disadvantages:**

- Because of the hierarchical structure of the process, the variance of the estimation increases causing instability of the trees [42].
- The relative importance of variables are unknown.
- CART has a difficulty in modelling an additive structure [42].
- It is a discrete scoring system.

**3.4 Theoretical Relationships between MARS and CART**

In fact, MARS could be considered as the generalization of stepwise linear regression, or as a modification of CART to improve its ability in regression process. We can infer that there is a close relationship between MARS and CART. It is reasonable to generate the regression tree by weakest link pruning that provides a sequence of limited subtree in CART [4].

Although they look much different from each other, MARS and CART are strongly analogous to each other in the following aspects [4]:

- (1) The forward process of MARS is the same as the tree growth algorithm of CART.
- (2) MARS and CART apply similar ways to deal with quantitative and qualitative "mixed" forecasters.

MARS makes up and improves some deficiencies of CART, especially, in the following ways [4]:

- (1) MARS can effectively improve the gliding attribute of forecasting scope.
- (2) MARS does not apply the tree structure, so that it can overcome the difficulties that CART suffers from in the course of capturing the addition structure.
- (3) MARS implements the piecewise linear basis function, which can strengthen the partial operational skills and reduce computing times.
- (4) MARS produces model function but CART does not. Thus, we can observe that which variables are used. Interactions of variables and the contribution of them to the model function can be easily understood.

# CHAPTER 4

## APPLICATION

### 4.1 Data and Methodology

The advantages and disadvantages of CART, MARS and CMARS were mentioned in previous chapter. However, we did not consider which one is the best in scoring. In this section, the validation of methods mentioned in Chapter 3 will be presented and applied.

The data used in this part were collected from the period of 1980-2005. The data set includes the financial information of 45 emerging markets including 1045 observations. Among of them, 410 are defaulters. The data of dependent variable: “whether a country is in debt crisis or not” in the period 1980-2004 is obtained from Fioramanti, Marco’s paper: Predicting Sovereign Debt Crises Using Artificial Neural Networks: A Comparative Approach [13]. The data set also includes 13 explanatory variables:

$X_1$ : Bank liquid reserves to bank assets ratio,

$X_2$ : Changes in net reserves / GDP ( Gross Domestic Product),

$X_3$ : Current account balance (% of GDP),

$X_4$ : Exports of goods and services (% of GDP),

$X_5$ : External debt total / Total Reserves,

$X_6$ : Long-term debt / GDP,

$X_7$ : GDP growth (annual %),

$X_8$ : Liquid liabilities as % of GDP,

$X_9$ : Total debt service (% of exports of goods services and income),

$X_{10}$ : Short-term debt (% of exports of goods services and income),

$X_{11}$ : Trade (% of GDP),

$X_{12}$ : Use of IMF credit / GDP,

$X_{13}$ : Inflation consumer prices (annual %).

In the application part, the data set is divided into two parts:

- Training sample: 777 observations - 1980-1999,
- Validation sample: 268 observations - 2000-2005.

The first sample was used in the estimation of the model parameters and the second sample was used in the validation of the model.

## 4.2 Application Results

### 4.2.1 Classification and Regression Trees

The prediction results of CART in the validation sample is shown in Table 4.1. From the results revealed in Table 4.1, we can observe that the average correct classification rate is 81.34% with 6 default market misclassified as non-default market and 44 non-default market misclassified as default market.

Table 4.1: Prediction results of classification and regression tree.

Validation Sample		Actual	
		<i>D</i> (Default)	<i>ND</i> (Non-default)
CART	<i>D</i> (Default)	45 (88.24%)	44 (20.28%)
	<i>ND</i> (Non-default)	6 (11.76%)	173 (79.72%)

Average correct classification rate: 81.34%.

### 4.2.2 Multivariate Adaptive Regression Splines

The prediction results of MARS in the validation sample is shown in Table 4.2. We can realize that the average correct classification rate is 85.45% with 6 default market misclassified as non-default market and 33 non-default market misclassified as default market.

Table 4.2: Prediction results of multivariate adaptive regression splines.

Validation Sample		Actual	
		<i>D</i> (Default)	<i>ND</i> (Non-default)
<b>MARS</b>	<i>D</i> (Default)	45 (88.24%)	33 (15.21%)
	<i>ND</i> (Non-default)	6 (11.76%)	184 (84.79%)

Average correct classification rate: 85.45%.

### 4.2.3 CMARS

The prediction results of CMARS in the validation sample is shown in Table 4.3. We can notice that the average correct classification rate is 86.94% with 7 default market misclassified as non-default market and 28 non-default market misclassified as default market.

Table 4.3: Prediction results of CMARS.

Validation Sample		Actual	
		<i>D</i> (Default)	<i>ND</i> (Non-default)
<b>CMARS</b>	<i>D</i> (Default)	44 (86.27%)	28 (12.90%)
	<i>ND</i> (Non-default)	7 (13.73%)	189 (87.10%)

Average correct classification rate: 86.94%.

In order to evaluate the classification capabilities of the three built debt crisis scoring models, the scoring results of the training and the validation samples are summarized in Table 4.4.

#### 4.2.4 Results Between Different Scoring Models

Table 4.4: Classification and prediction results of the three built models.

	Training Sample			Validation Sample		
	{D-D}	{ND-ND}	Average correct classification rate	{D-D}	{ND-ND}	Average correct classification rate
<b>CART</b>	94.43%	97.13%	95.88%	88.24%	79.72%	81.34%
<b>MARS</b>	84.68%	91.87%	88.55%	88.24%	84.79%	85.45%
<b>CMARS</b>	81.89%	92.58%	87.64%	86.27%	87.10%	86.94%

From Table 4.4, we can infer that, in the training sample, CART scoring model have better classification capability in terms of the average correct classification rate. However, in the validation sample estimates MARS and CMARS approaches give more accurate results. Besides, as the classification accuracy of the validation samples is only slightly lower than those of the corresponding training samples, we can conclude that MARS and CMARS discover the main structure of the data and consequently can be successfully applied in the validation sample after the model building procedure.

#### 4.2.5 Type I and Type II Errors of the Constructed Models

It is well known that, in order to justify the prediction performance of the designed models, the misclassification probability and misclassification costs have to be taken into account in order to obtain a model with the smallest expected misclassification costs. Hence, special attention also needs to be paid to misclassification costs in order to evaluate the prediction aptitude of the built models [32, 26].

Table 4.5: Type I and Type II error definitions.

		Actual	
		Default	Non-default
<b>Model Classification</b>	Default	1-Type II	Type I
	Nondefault	Type II	1-Type I

Type I and Type II errors of the three models need to be compared in order to ratify the overall prediction capability. Table 4.5 shows the error definitions. Type I and Type II errors of the three built models are summarized in Table 4.6.

Table 4.6: Type I and Type II error estimates of the three model.

	<b>Types of Errors</b>			
	<b>Training Sample</b>		<b>Validation Sample</b>	
	<b>Type I</b>	<b>Type II</b>	<b>Type I</b>	<b>Type II</b>
<b>CART</b>	0.0287	0.0557	0.2028	0.1176
<b>MARS</b>	0.0813	0.1532	0.1521	0.1176
<b>CMARS</b>	0.0742	0.1811	0.129	0.1373

According to Table 4.6, CART, MARS and CMARS have close Type II errors in comparison with each other. All three models have small Type II errors which reduce the possible risks associated with Type II errors. In terms of Type II error, MARS and CART detects default records slightly better than CMARS. However, in terms of Type I error, we can conclude that MARS and, especially, CMARS has a better estimate compared with CART. CMARS clearly detects non-default records better than MARS and CART.

#### 4.2.6 The Receiver Operating Characteristic Curves of the Constructed Models

To validate the models we also use the *Receiver Operating Characteristic Curves (ROC)* for each method. The ROC curve is a curve of True Positive Rate with respect to False Positive Rate.

$$\text{True Positive Rate}(C) = P(\text{Score of a default country is } \leq C),$$

and

$$\text{False Positive Rate}(C) = P(\text{Score of a nondefault country } \leq C).$$

where  $C$  is any arbitrary cut-off point score [23].

Figures 4.1, 4.2 and 4.3 show the plotted ROC curves for training and validation samples for our analysis, respectively.

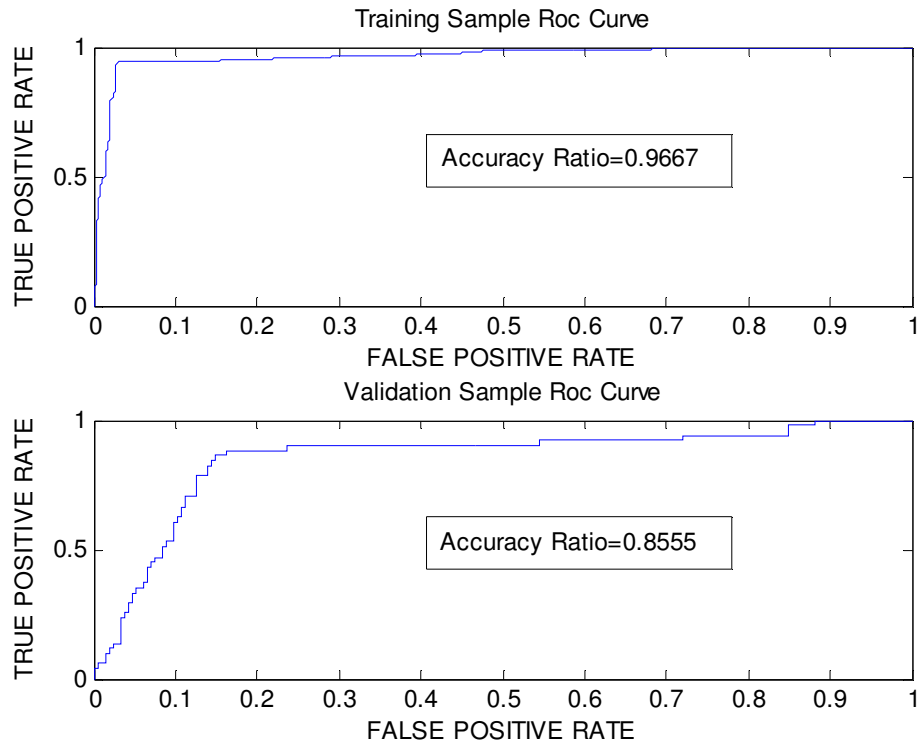


Figure 4.1: Receiver operating characteristic curves for CART.

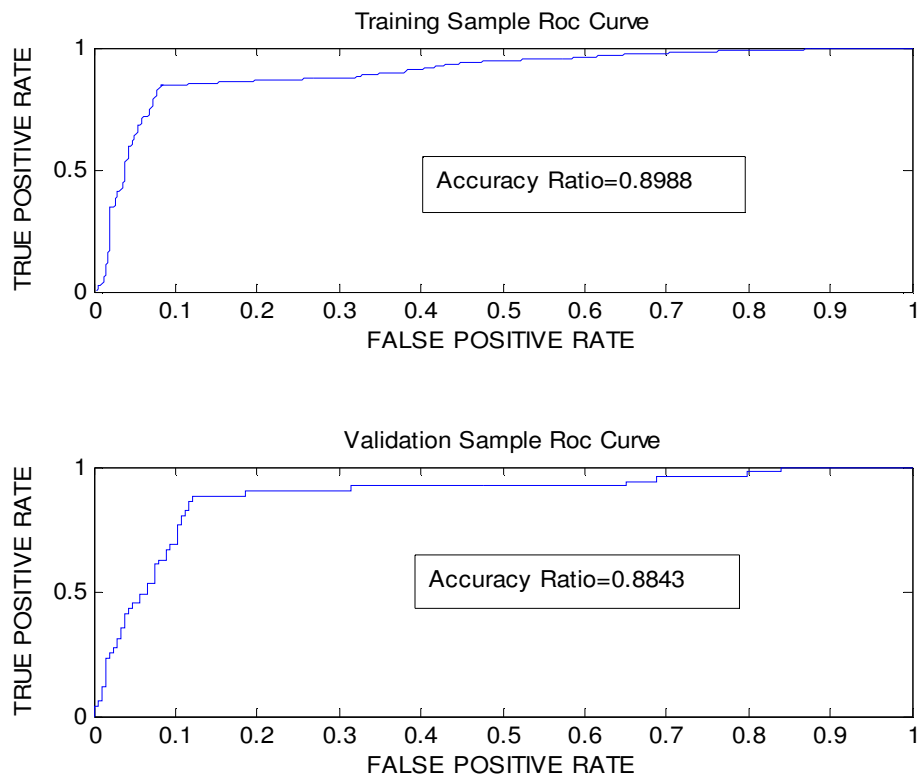


Figure 4.2: Receiver operating characteristic curves for MARS.

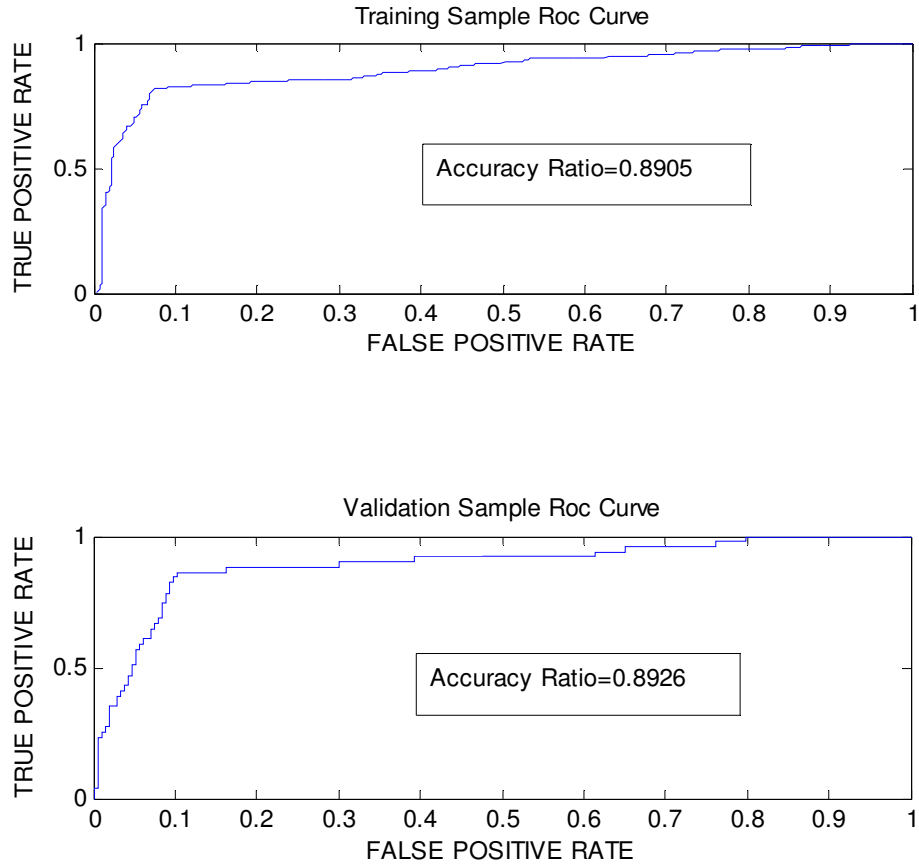


Figure 4.3: Receiver operating characteristic curves for CMARS.

From these figures, we could not be able to observe any big differences in the validation and training sample results for MARS and CMARS. On the other hand, CART does not sustain its discrimination success in training sample and become the worst model in the validation sample results. In conclusion, the results show that MARS and, especially, CMARS provides better accuracy results in validations. It follows that MARS and CMARS are robust enough for practical application.

#### 4.2.7 Variable Selection Results in Multivariate Adaptive Regression Splines

One of the objectives of this research is to investigate which variables are likely to be of value in predicting debt defaults. To do this we have to look at the generalized cross validation (GCV) procedure. This procedure can be used to determine which variables to keep in the model, as they produce the best fit, and which to eliminate [15].

The variable selection results using MARS are summarized in Table 4.7. MARS indicates 6 explanatory variables are really important classifiers. It is inferred from the table that the most important classifiers in the estimation of debt defaults are *the use of IMF credit / GDP, inflation consumer prices (annual %), liquid liabilities as % of GDP, trade (% of GDP), long-term debt / GDP and exports of goods and services (% of GDP)*.

Table 4.7: Variable selection results using MARS.

Variable	Importance	- GCV
$X_{12}$ : Use of IMF credit / GDP	100	0.191
$X_{13}$ : Inflation consumer prices (annual %)	76.595	0.161
$X_8$ : Liquid liabilities as % of GDP	62.925	0.147
$X_{11}$ : Trade (% of GDP)	60.410	0.145
$X_6$ : Long-term debt / GDP	53.142	0.139
$X_4$ : Exports of goods and services (% of GDP)	49.001	0.136

# CHAPTER 5

## CONCLUSION

The increase in debt crisis in emerging markets after 1980s, aroused the interests of scientists in the area of sovereign defaults. In this study, we also consider the debt crisis, and we aim to capture the signals of debt crisis before crisis exists. In this work, we give an overview about the theoretical aspects of CART and MARS and their applications in estimating debt defaults and credit scoring.

This study on classification also provides a new contribution to MARS algorithm. The MARS algorithm is modified by constructing a penalized residual sum of squares (PRSS) as a Tikhonov regularization problem. This problem is solved by using continuous optimization, especially, conic quadratic programming (CQP). We employ the program packages of MOSEK, which is an optimization tool for solving large-scale mathematical optimization problems, Salford MARS and special codes written in Matlab, and provide an alternative modeling technique for MARS, named CMARS [53].

In the application part, we analyze the financial information of 45 emerging markets including 1045 observations collected from the period of 1980-2005. Firstly, we verify the ability of CART, MARS and CMARS to adapt data and secondly we check the prediction accuracies of the of all methods mentioned in this work. We also intend to determine the important financial ratios in predicting debt defaults.

According to the results of application, CART approach gives more accurate results in training sample estimates, however, in testing sample estimates MARS and CMARS give more accurate results. MARS and CMARS discover the main structure of the data better and consequently can be successfully applied in the validation sample after the model building procedure.

The results also indicate that there is not a significant difference between CMARS and MARS solutions. However, CMARS is more robust than MARS and performance

measures of CMARS show higher stability. CMARS includes an improvement on the second part of MARS algorithm making the predictive accuracy of MARS better. CMARS seriously reduces the chance of making Type I error compared with other methods. The results of the analyze also show that *the use of IMF credit / GDP, inflation consumer prices (annual %), trade (% of GDP), long-term debt / GDP, liquid liabilities as % of GDP and exports of goods and services (% of GDP)* are really important classifiers.

By this project, we intended to bring in a new point of view to the theory, methods and applications of mathematical data mining for displaying its success in financial theory by the use of modern continuous optimization. The results of the analyze show that CMARS can be used as much as MARS in the estimation of debt defaults and future research challenges in credit scoring.

As a future work, MARS and CMARS can be compared with other modeling techniques such as artificial neural networks, generalized additive models, logistic regression and robust regression for analyzing continuous data sets and different distributions of variables. Moreover, under the normality assumption, MARS and CMARS can be compared with linear regression and discriminant analysis. As we said, CMARS includes an improvement on the second part of MARS algorithm. By using clustering techniques, a similar improvement can also be made for the first part of the algorithm. The first part of MARS algorithm, e.g., forward stepwise algorithm, determines knot points among the data points for obtaining basis functions. Increasing in the number of data points results in a one-to-one manner in an increase in the number of knot points. Therefore it gives rise to complexity. For this reason we can try to determine suitable knot points for the data set *by using clustering theory* [53]. Also we should not forget to improve CMARS in order to make it a user friendly data mining tool and reduce computational time.

## REFERENCES

- [1] Andriyashin, A., *Financial Applications of Classification and Regression Trees*, 2005, Master Thesis, Center of Applied Statistics and Economics Humboldt University, Berlin.
- [2] Breiman, L., Frydman, H., Olshen, R.A., and Stone, C.J., *Classification and Regression Trees*, Chapman and Hall, New York, London, 1984.
- [3] Campbell, T.S., and Dietrich, J.K., *The Determinants of Default on Insured Conventional Residential Mortgage Loans*, *Journal of Finance*, 38 (1983) 1569-1581.
- [4] Chen, L., Song, J., and Ji F., *MARS-based Research of Personal Credit Scoring: Verification of Chinese Data*, Management Science and Engineering, International Conference on; Lille, France (2006) 1498-1501.
- [5] Copyright StatSoft, Inc., *Multivariate Adaptive Regression Splines*, <http://www.statsoft.com/textbook/stmars.html> (accessed 15 June 2009).
- [6] Craven, P., and Wahba, G., *Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation*, *Numerische Mathematik*, 31 (1979) 377-403.
- [7] Crook, J.N., Edelman, D.B., and Thomas L.C., *Credit Scoring and Its Applications*, SIAM, 2002.
- [8] Desai, V.S., Crook, J.N., and Overstreet, Jr., G.A., *A comparison of neural networks and linear scoring models in the credit union environment*, *European J. Oper. Res.*, 95 (1996) 24-37.
- [9] Devaney, S., *The Usefulness of Financial Ratios as Predictors of Household Insolvency: Two Perspectives*, *Financial Counseling and Planing*, 5 (1994) 15-24.

- [10] Deveaux, R.D., Psychogios, D.C., and Ungar, L.H., *A comparison of two nonparametric estimation schemes: MARS and neural networks*, Computers in Chemical Engineering, 17, 8 (1993) 819-837.
- [11] Di, W., *Long Term Fixed Mortgage Rate Prediction Using Multivariate Adaptive Regression Splines*, School of Computer Engineering, Nanyang Technological University, 2006.
- [12] Feldman, D., and Gross, S., *Mortgage Default: Classification Trees Analysis*, The Pinhas Sapir Center For Development Tel-Aviv University, Discussion Paper No. 3, 2003.
- [13] Fioramanti, M., *Predicting Sovereign Debt Crisis Using Artificial Neural Networks: A Comparative Approach*, Preliminary version, 2006.
- [14] Fisher, R.A., *The use of multiple measurements in taxonomic problems*, Annals of Eugenics, 7 (1936) 179-188.
- [15] Francis, L., *Is MARS beter than Neural Networks?*, <http://www.salford-systems.com/doc/03wf027.pdf> (accessed 31 July 2009).
- [16] Frankel, J.A. and Rose, A.K., *Currency Crashes in Emerging Markets: An Empirical Treatment*, Journal of International Economics, 41 (1996) 351-366.
- [17] Friedman, J.H., *Multivariate adaptive regression splines*, The Annals of Statistics, 19, 1 (1991) 1-141.
- [18] Frydman, H., Altman, E., and Kao, D.L., *Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress*, The Journal of Finance, 40 (1985) 269-291.
- [19] Gardner, M.J., and Mills, D.L., *Evaluating the Likelihood of Default on Delinquency Loans*, Financial Management, 18 (1989) 55-63.

- [20] Glasson, S., *Censored regression techniques for credit scoring*, The Doctoral thesis of Philosophy, School of Mathematical and Geospatial Sciences Science, Engineering and Technology Portfolio, RMIT University, July 2007.
- [21] Hamerle, A., Liebig, T., and Rösch, D., *Credit Risk Factor Modeling and the Basel II IRB Approach*, Deutsche Bundesbank Discussion Paper Series No:2 Banking and Financial Supervision, 2003.
- [22] Hastie, T., Tibshirani, R., and Friedman, J., *The Elements of Statistical Learning, Data Mining, Inference and Prediction*, Springer, 2001.
- [23] İşcanoğlu, A., Weber, G.W., and Taylan, P., (2007), *Predicting default probabilities with generalized additive models for emerging markets*, lecture at the Graduate Summer School on Recent Advances in Statistics, METU, Ankara, Available via DIALOG, <http://www.statsummer.com>.
- [24] İşcanoğlu, A., *Credit Scoring Methods and Accuracy Ratio*, 2005, Master Thesis, METU, Ankara.
- [25] Jensen, H.L., *Using neural networks for credit scoring*, *Managerial Finance*, 18 (1992) 15-26.
- [26] Johnson, R.A., and Wichern, D.W., *Applied Multivariate Statistical Analysis, (Fifth Edition)*, Prentice-Hall, Upper Saddle River, NJ., 2002.
- [27] Kaminsky, G.L., Lizondo, S. and Reinhart, C.M. (1998), *Leading Indicators of Currency Crises*, IMF Staff Papers, 45, 1 (1998).
- [28] Kiss, F., *Credit scoring process from a knowledge management perspective*, *Periodical polytechnica*, 11, 1 (2003) 95-110.
- [29] Kolyshkina, I., and Brookes, R., *Data Mining Approaches to Modeling Insurance Risk*, Report, Price Waterhouse Coopers, 2002.

- [30] Kriner, M., *Survival Analysis with Multivariate adaptive Regression Splines*, 2007. Dissertation, LMU Munchen: Faculty of Mathematics, Computer Science and Statistics.
- [31] Lawrence, E., and Arshadi, N., *A Multinomial Logit Analysis of Problem Loan Resolution Choices in Banking*, *Journal of Money, Credit and Banking*, 27 (1995) 202-216.
- [32] Lee, T., Chiu, C., Chou, Y., and Lu, C., *Mining the customer credit using classification and regression tree and multivariate adaptive regression splines*, *Computational Statistics & Data Analysis*, 50 (2006) 1113-1130.
- [33] Lee, G., Sung, T.K., and Chang, N., *Dynamics of modeling in data mining: Interpretive approach to bankruptcy prediction*, *Journal of Management Information Systems*, 16 (1999) 63-85.
- [34] Lee, T., and Chen, I., *A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines*, *Expert Systems with Applications*, 28 (2005) 743-752.
- [35] Liu, Y., and Schumann, M., *Data mining feature selection for credit scoring models*, *Journal of the Operational Research Society*, 56 (2005) 1099-1108.
- [36] MOSEK, *A very powerful commercial software for CQP*, <http://www.mosek.com> (accessed 04 August 2009).
- [37] Nemirovski, A., *A lectures on modern convex optimization*, *Israel Institute of Technology* ( 2002), <http://iew3.technion.ac.il/Labs/Opt/opt/LN/Final.pdf> (accessed 31 July 2009).
- [38] Piramuthu, S., *Financial credit-risk evaluation with neural and neurofuzzy systems*, *European J. Oper. Res.*, 112 (1999) 310-321.
- [39] Piramuthu, S., Shaw, M.J., and Gentry, J.A., *A classification approach using multi-layered neural networks*, *Decision Support Systems*, 11, 5 (1994) 509-525.

- [40] Sabzevari, H., Soleymani, M., and Noorbakhsh, E., *A comparison between statistical and data mining methods for credit scoring in case of limited available data*, <http://www.crc.man.ed.ac.uk/conference/archive/2007/papers/sabzevari-et-al.pdf> (accessed 20 July 2009).
- [41] Satchidananda, S., and Simha, J.B., *Comparing decision trees with logistic regression for credit risk analysis*, [http://www.abibasystems.com/credit\\_risk\\_modeling.pdf](http://www.abibasystems.com/credit_risk_modeling.pdf) (accessed 15 July 2009).
- [42] Staicu, A., *Classification Trees and MARS*, <http://www.maths.bris.ac.uk/maams/Mar.16pdf> (accessed 04 August 2009).
- [43] StatSoft, Inc., *Data Mining Techniques*, <http://www.statsoft.com/textbook/stdatmin.html> (accessed 31 July 2009).
- [44] Steinberg, D., *Predicative Modeling and Data Mining for Actuaries*, [http://www.actuaries.asn.au/IAA/upload/public/iInsights\\_0704\\_Steinberg\\_Presentation.pdf](http://www.actuaries.asn.au/IAA/upload/public/iInsights_0704_Steinberg_Presentation.pdf) (accessed 04 August 2009).
- [45] Steinberg, D., Colla, P.L., and Martin, K., *MARS User Guide*, San Diego, CA: Salford Systems, 1999.
- [46] Taylan, P., Weber, G.W., Yerlikaya, F., Batmaz İ. and Köksal, G., *MARS Algoritmasında Tikhonov Düzenlemesi ve Çok Amaçlı Optimizasyon Kullanımı*, <http://www3.iam.metu.edu.tr/iam/images/4/4f/Preprint105.pdf> (accessed 31 July 2009).
- [47] Taylan, P., Weber, G.W., and Yerlikaya, F., *Continuous optimization applied in MARS for modern applications in finance, science and technology*, in the ISI Proceedings of 20th Mini-EURO Conference Continuous Optimization and Knowledge-Based Technologies, Neringa, Lithuania (2008) 317-322.

- [48] Taylan, P., Weber, G.W., and Yerlikaya, F., *A new approach to multivariate adaptive regression spline by using Tikhonov regularization and continuous optimization*, to appear in TOP (the Operational Research journal of SEIO (Spanish Statistics and Operations Research Society), Selected Papers at the Occasion of 20th EURO Mini Conference Continuous Optimization and Knowledge-Based Technologies (Neringa, Lithuania, May 20-23, 2008).
- [49] Taylan, P., Weber, G.-W. and Beck, A., *New approaches to regression by generalized additive models and continuous optimization for modern applications in finance, science and technology*, in the Special issue of Optimization, 2007 56(5–6): 1-24.
- [50] Thomas, L.C., Edelman D.B., and Crook, J.N., *Credit Scoring and Its applications*, Society for Industrial and Applied Mathematics, 2002.
- [51] Thomas, L.C., *A survey of credit and behavioral scoring: Forecasting financial risks of lending to customers*, International Journal of Forecasting, 16 (2000) 149-172.
- [52] Vojtek, M., and Kocenda, E., *Credit Scoring Methods*,  
[http://journal.fsv.cuni.cz/storage/1050\\_s\\_152\\_167.pdf](http://journal.fsv.cuni.cz/storage/1050_s_152_167.pdf) (accessed 15 June 2009).
- [53] Yerlikaya, F., *A New Contribution to Nonlinear Robust Regression and Classification with Mars and Its Applications to Data Mining for Quality Control in Manufacturing*, 2008, Master Thesis, METU, Ankara.
- [54] West, D., *Neural Network Credit Scoring Models*, Computer and Operation Research, 27 (2000) 1131-1152.
- [55] Wiginton, J.C., *A Note on the Comparison of Logit and Discriminant Models of Consumer Credit Behavior*, Journal of Financial and Quantitative Analysis, 15 (1980) 757-770.