

An implementation of Soft Set Theory in the Variables Selection Process for Corporate Failure Prediction Models. Evidence from NASDAQ Listed Firms

Apostolos G. Christopoulos^a, Ioannis G. Dokas^b, Iraklis Kollias^c, John Leventides^d

Abstract

The foremost aim of this paper is to propose a reliable methodology regarding the selection process of financial ratios as input variables in the construction of corporate failure prediction models. In this paper soft set theory is introduced. In the first stage, emphasis is given on the state of liquidity as a measure for the classification of a group of NASDAQ listed firms in two a priori groups (failed and non-failed) using four liquidity criteria as follows: current ratio < 1 , current liabilities to total liabilities $> 70\%$, Equity to Liabilities ≤ 0 and Total Debt to Total Assets $> 70\%$. In the second stage, a parameter reduction algorithm is applied in order to determine, from a group of ratios, those which provide significant predictive power and optimize the classification accuracy of the model. A tabular representation of a soft set is constructed in order to select the input variables in the model based on the importance degree of each financial ratio. The findings show that the primary assumptions relevant to the definition of failure based on the soft set theory approach are confirmed, though the majority of the significant ratios in the applied sample of listed firms are related to the analysis of profitability.

JEL classification numbers: C2, C4, G33

Keywords - Failure prediction, definition of failure, liquidity and profitability, financial ratios, soft set theory, Logistic regression.

b Department of Economics, Democritus University of Thrace, National Technical University of Athens, Faculty of Applied Mathematics and Physical Sciences. Postgraduate program "Mathematical Modeling in modern Technologies and Financial Engineering" (intokas@econ.duth.gr).

a,c,d Faculty of Economics, National and Kapodistrian University of Athens, National Technical University of Athens, Faculty of Applied Mathematics and Physical Sciences. Postgraduate program "Mathematical Modeling in modern Technologies and Financial Engineering" (axristop@econ.uoa.gr, hkollias@econ.uoa.gr, ylevent@econ.uoa.gr)

1 Introduction

In the world of business, the ability to predict and avoid firms' bankruptcy plays a significant role in the decision making process. A significant issue, in the study of the corporate failure prediction models, is the definition of failure. For many researchers like Ohlson (1980), Zmijewski (1984), Dirickx and Van Landeghem (1994), Charitou et al. (2004), Hauser and Booth (2011), the legal status of bankruptcy constitutes the main definition of failure. Alternative approaches though provide the stage of financial distress as a "step" before bankruptcy. The main characteristics of financial distress are insolvency, long term low or even negative profitability, low financial flexibility relevant to the investment financing and dividend policy. Although financial distress does not necessarily lead to failure, empirical studies show that the majority of distressed firms cannot avoid bankruptcy.

During the last five decades numerous studies have focused on investigating the causes of firm failure in order to determine the likelihood of bankruptcy. Beaver (1966) and Altman (1968), the pioneers in this field, have established two different methodological approaches, univariate and multiple discriminant analysis, in order to predict and avoid bankruptcy. According to their studies, the estimation of a firm's financial status is based on the results of ratio analysis. Ratio analysis offers findings relevant to the performance of the firm in the fields of liquidity, profitability and management quality. According to the traditional literature, these fields constitute the major factors of the firm's viability. However, the distributional properties of financial ratios create several limitations with respect to the use of this tool in the construction of reliable predictive models. These limitations characterized mainly the econometric methods used like multiple discriminant analysis. The major problem concerned the fact that the ratios were not normally distributed.

However, looking back at the evolution of research in this field, it is clear that emphasis was given on the development of the computational methods, while financial ratios remain the basic type of explanatory variables in the majority of corporate failure prediction models. The above mentioned problems of financial ratios were dealt with various techniques. McLeay and Omar (2000) indicate a transformation approach which is based on the ratio's form and can lead to an improvement in the classification accuracy of prediction models.

The introduction of the non-parametric models provided a reliable solution in the use of financial ratios in the construction process of the predictive models. These models were based on advanced mathematical tools like artificial neural networks (Iturriaga & Sanz, 2015, Christopoulos et al. 2013, Youn & Gu, 2010, Jardin, 2010, Tang & Chi, 2005), data envelopment analysis (Cielen et al., 2004, Premachandra et al., 2011), recursive partitioning algorithm (Frydman et al., 1985), multicriteria methods (Doumpos & Zopounidis, 1999, Zopounidis & Doumpos, 2001), rough set theory (Dimitras et al. 1999), DEA and rough sets (Yeh, et al. 2010, Beynon & Peel, 2001), and soft set theory (Molodtsov, 1999, Xu, et al. 2014, Kong, et al. 2015).

A common characteristic between the various types of predictive models is the use of the same group of financial ratios and the diversification concerns the type of technique which is based on the model structure. These approaches are characterized by the limitation of the static theoretical framework, while the availability of accounting information can contribute to the creation of a strong theoretical framework of failure causes, giving alternative definitions according to the overall financial status of firms.

The scope of this study is to propose a new methodology on the financial ratios' selection process for the construction of a failure prediction model. The primary discrimination of our sample in two groups is based on the second pillar of the Courtis model (1978) and, specifically, on financial distress on the basis of the liquidity status. Our approach focuses mainly on the examination of liquidity as we know that firms which are active in the capital markets concentrate their strategy on the profitability performance. In this sense, we assume that in difficult periods, problems will mainly arise in the performance of liquidity which comes second in the planned strategies of the public firms.

The contribution of this study concerns primarily the proposal of a reliable methodology at the stage of the variables selection process. In essence, this approach is based on soft set theory as an advanced method for improving the classification accuracy of a sample of firms in two groups: distressed and non-distressed.

The methodology of this study is structured as follows: In the first stage, the theoretical framework of failure is defined. Specifically, giving emphasis on the state of liquidity a group of NASDAQ listed firms in the sectors of chemicals and metals is classified in two a priori groups (distressed and non-distressed) based on their liquidity status. The fundamental assumption relevant to the liquidity status of distressed firms includes four criteria: current ratio < 1 , current liabilities to total liabilities $> 70\%$, Equity to Liabilities ≤ 0 , and Total Debt to Total Assets $> 70\%$.

In the second part of this stage, the measures which estimate the financial status of firms are also established. In the second stage, a parameter reduction algorithm is applied in order to select those financial ratios which provide significant predictive power and optimize the classification accuracy of the sample in the two a priori groups. For this purpose the logistic regression (LR) is fitted on the training data set in order to model the probability of a firm being in normal (non-failed) status. The tabular representation of the soft set (F, A) , where A is the set of financial ratios and F is an appropriate mapping, is obtained based on the results of the LR model. This representation is subsequently used in order to calculate the parameter importance degree of each financial ratio based on which the input variables in the model under construction are selected.

The remainder of this paper is structured as follows: Section 2 presents a brief literature review, while in Section 3 the discussion focuses on methodology. Section 4 introduces the model specification and presents the results. Section 5 discusses the conclusions and provides recommendations for future work.

2 Literature Review

In the fields of accounting and finance several studies focused on the ability to predict and avoid firms' bankruptcy. The surveys of Beaver (1966), Beaver et al (2005), and Altman (1968, 1977, 2006) constituted the starting point of this effort and their theoretical framework was based on decoding the accounting figures using financial ratios. The performance of a firm in the capital market, the solvency status and the negative equity (Wilcox, 1971), constituted classical approaches in the definition of failure. However, numerous researchers like Deakin (1972) and Almamy, et al (2016) define failure according to the legal bankruptcy, while Vranas (1992) included also the case of the takeover from the creditors. Liang et al (2016) defined failure according to the business

regulations of the Stock Exchange, while according to Doumpos and Zopounidis (1999) the definition of failure includes the characteristics of inefficient liquidity status and negative asset value. In essence, all these approaches present a point of convergence in the basis of the Courtis model (1978). In summary, the pillars of liquidity, management quality and profitability create a reliable framework relevant to the causes of firms' failure.

Soft set theory, first proposed in a paper by Molodtsov (1999), is an intelligent technique based on a generalization of fuzzy set theory in order to deal with uncertainty in a parametric manner. In particular, a soft set is a collection of subsets of a set, X . This collection is determined via mapping parameters to corresponding arbitrary subsets of X . Two examples of soft sets are provided in Appendix A. A soft set generalizes a fuzzy set by assigning a set rather than a number to every element of an underlying set. Soft set theory is free from the problem of setting up the membership function which is present in fuzzy set theory. The application of fuzzy sets in financial classification is relatively limited. See, Scherger, et al (2014), Vigier and Terceño (2008). Unlike statistical methods used for the purpose of financial classification and bankruptcy prediction, soft set theory is free from the several problems related to the application of such statistical methods. A nice overview of these methods and their problems is Balcaen and Ooghe (2004). Maji et. al (2003) defined the notions of equality of two soft sets, subset and super set of a soft set, complement of a soft set, null soft set, and absolute soft set giving examples. They also defined soft binary operations like AND, OR and the operations of union and intersection.

In addition, they verified DeMorgan's laws in soft set theory. The work of Maji et. al (2003) was further improved by Chen et. al (2005). Yang et. al (2007) investigated the idea of fuzzy soft sets and provided some immediate outcomes. Their work was extended by Kharal and Ahmad (2011) who introduced the notion of mapping on the classes of fuzzy soft sets which constitutes a pivotal notion regarding the advanced development of any new area of mathematical sciences. Ali et. al (2009) provided some new notions such as the restricted intersection, the restricted union, the restricted difference, and the extended intersection of two soft sets along with a new notion regarding the complement of a soft set. Majumdar and Samanta (2010, 2011) studied the notion of similarity between two fuzzy soft sets as well as the application of similarity transformation between fuzzy soft sets.

3 Methodology

In this section the discussion focuses on the stage of methodology. In the first part the theoretical status of liquidity and the group of financial ratios, which at a theoretical level are used in viability analysis, are presented. In the second part of this section we provide the theoretical framework of soft set theory and the process which is followed in order to develop the proposed model.

3.1 Theoretical Status of Liquidity

The term of liquidity refers to the ability of a firm to create the essential cash inflows in order to pay current liabilities. In order to separate our sample in two groups (distressed and non-distressed firms), on the basis of their liquidity status, four assumptions are introduced. The first assumption is that the Current Ratio < 1 which means that current liabilities are financing a percentage of the fixed asset whereas the ratio

Equity to Liabilities indicates the guaranteeing of creditors by the firm's assets. In this paper, the sample is classified in distressed and non-distressed firms based on their liquidity ratios compared to the critical values. Such values have been set based on a semi structured interview with a group of twenty experts out of which seven were financial and insurance experts, and the rest were experts from the banking sector. The results of the semi structured interview were discussed further with academics who agreed with these conclusions. This interview can be used as a pilot survey for evaluation and the accuracy of financial failure prediction model variables. According to the results of this research, the average price for the current liabilities to total liabilities ratio was 0,7 with standard deviation 0,12 while the average critical value for the total debt to total assets ratio was 0,71 with a standard deviation 0,15.

The applied approach to investigate the financial status is based on the three pillars of Courtis model. For that the following ratios are introduced:

The **Current Ratio** is the main liquidity ratio which offers estimation for the potential of working capital to cover current liabilities. However, the reliability of this ratio depends on the activity ratios like inventories turnover ratio, receivables turnover ratio, and short term liabilities turnover ratio. The **Acid Test Ratio** belongs to the group of basic liquidity ratios but is more reliable than the current ratio because it combines the defense assets with the short term liabilities. The **Cash Ratio** presents the relationship between cash items and cash equivalents with current liabilities. Its information concerns the ability of liquid assets to cover current liabilities. The **Defensive Interval Ratio** constitutes a measure relevant to the period for which the defense assets cover the daily operating cost. The **Current Liabilities to Total liabilities** ratio constitutes a significant tool in the investigation process of the firms' liquidity status which shows the structure of liabilities. An increase of current liabilities in relation to the long term liabilities offers significant findings concerning solvency risk, especially in the positive phase of the business cycle. The **Equity to Liabilities** ratio constitutes a pledge measure to the firm's creditors and especially to the short term creditors. The informative value of this ratio is significant for the creditors because it enhances the reliability of their credit risk analysis and contributes to the configuration of their credit policy. If the creditors ensure their receivables, then they adopt a large credit policy, to their clients. This fact means longer credit lines and lower financial cost for the clients. The **total Debt to Total Asset ratio**, also known as the debt ratio, provides information for the level of the external financing. When this ratio takes on high values, the failure probability for liquidity reasons is higher, especially in the case that a main part of liabilities includes short term liabilities. The **Operating Expenses to Average Current Assets ratio** shows the ability of a firm to pay its operating expenses from the liquidation of working capital. The **Depreciation to Current Liabilities** is a liquidity measure because the depreciations create cash inflows. **ITR (Inventories Turnover Ratio)** is a figure which belongs to the family of the activity ratios. It plays a supporting role in the illustration process of the basic liquidity ratios. The use of this ratio offers findings relevant to the degree of the inventories effective management. Therefore ITR constitutes a measure which is used in order to evaluate the management performance. The **Receivables Turnover Ratio (RTR)** ratio evaluates the quality of receivables portfolio. In addition it constitutes a measure to evaluate the efficiency of receivables management. The **CLTR (Current Liabilities Turnover Ratio)** is an informative measure relevant to the ability of firms to manage effectively their current liabilities. A low value of this ratio can be a signal of ineffective management of

the current liabilities in the case which the essential cash outflows to the creditors are delayed. However a low value of this ratio may be justified by the existence of longer credit lines by the creditors. The **Fixed Asset Turnover Ratio** offers significant estimation relevant to the ability of firms to manage effectively the fixed assets. The **Net Working Capital to Current Asset**. **Net Working Capital to non –Current Liabilities**. This ratio offers estimations relevant to the ability of firm's working capital to cover a part of long term liabilities. It's a reliable estimated measure for the long term liquidity of a firm.

Equity to Total Asset. It is a leverage ratio and offers information relevant to the structure of capital. **Net Change in Cash to Current liabilities**. This ratio is encountered in the relative literature. It offers information about the liquidity status of firms. **Gross Profit Margin**, constitutes a crucial profitability ratio. It offers information relevant to the ability of firm to minimize the Cost of Goods Sold (COGS) relevant to the sales. **Net Profit Margin**. This ratio constitutes a fundamental profitability measure and offers findings relevant to the degree of the effective management of the operation cost. **Return on Asset, and Return on Equity** constitute crucial profitability ratios and offer information relevant to the performance of the total and Shareholders' capital. A combined analysis of these ratios offers significant findings relevant to the contribution of the external funding in the firms income. **Financial Leverage** show the participation of debt in the asset financing, while the ratio **Financial Expenses to Operating Expenses** is a measure for the size of financial cost as percentage of the total operating cost. The size of financial cost affects the illustration of the financial leverage results. **Operating Expenses/Revenues**. This ratio offer information relevant to the size of operating cost as percentages of revenues. **Net Income/Weighted Average Common Shares Outstanding** shows the relation between earnings after taxes and Average Common Shares Outstanding and finally the ratio **Depreciation/ Total Cost** is a measure of earnings quality.

Table 1: Financial Ratios

Primary Group of Financial Ratios		
1	Current Ratio	CR
2	Acid Test Ratio	ATR
3	Cash Ratio	CSR
4	Defensive Interval Ratio	DIR
5	Current Liabilities to Total Liabilities	CL/TL
6	Equity to Liabilities	E/L
7	Total Debt to Total Asset	TD/TA
8	Operating Expenses to Avg Current Assets	OE/CA
9	Depreciation to Current liabilities	D / CL
10	Inventories Turnover Ratio	ITR
11	Receivables Turnover Ratio	RTR
12	Current Liabilities Turnover Ratio	CLTR
13	Fixed Asset Turnover Ratio	FATR

14	Net Working Capital to Current Asset	NWC/CA
15	Net Working Capital to Non - Current Liabilities	NWC/NCL
16	Equity to Total Asset	E/TA
17	Net Change in Cash to Current Liabilities	NCC/CL
18	Gross Profit Margin	GPM
19	Net Profit Margin	NPM
20	Return on Asset	ROA
21	Return on Equity	ROE
22	Financial Leverage	FL
23	Financial Expanses to Operating Expanses	FE/OE
24	Operating Expanses/Revenues	OE/R
25	Net Income /Weighted Avg Common Shares Outstanding	NI /WCSO
26	Depreciation /Total Cost	D /TC

3.2 Preliminaries in Soft Set Theory

Originated by Molodtsov (1999), soft set theory deals with complicated problems in economics, engineering, environment, medical science etc., that involve various types of uncertainties typical for those problems. There are many theories, *viz.*, theory of probability, theory of fuzzy sets (Zadeh, 1965), and theory of interval mathematics (Atanassov, 1994) which can be considered as mathematical tools dealing with uncertainties. However, all these theories have their inherent difficulties as pointed out in Molodtsov (1999). The reason for these difficulties is, possibly, the inadequacy of the parameterization tool of the theories.

In classical mathematics, we construct a model of an object and define the notion of an exact solution. Usually this solution is associated with a high degree of complexity so that the idea of an approximate solution is introduced. On the other hand, in soft set theory the initial description of the object has an approximate nature which is free from any restrictions thus making this theory particularly appealing and easily applicable in practice. Any parameterization we prefer such as words and sentences, real numbers, functions, and so on can be used in the description of the object. Potential fields of application of soft set theory include game theory, operations research, Riemann integration, Perron integration, probability theory, and measurement theory.

For a set X we denote by $P(X)$ the power set of X , that is the set of all subsets of X and introduce the following

Definition 3.2.1: Let X be an initial universe set and A a set of parameters. A pair (F, A) , where F is a map from A to $P(X)$, is called a soft set over X .

In other words, a soft set is a parameterized family of subsets of the set X . Two trivial examples of soft sets are the null soft set and the total or absolute soft set defined respectively as follows:

(a) The null soft set $\Phi = (F, A) : F(a) = \emptyset$ for all $a \in A$

(b) The total soft set $\Psi = (F, A) : F(a) = X$ for all $a \in A$

Molodtsov (1999) considered several examples of soft sets, for illustration purposes, one of which is presented in Appendix A together with an added similar example. Appendix B contains some basic definitions regarding operations with soft sets.

3.3. Parameter reduction in soft sets

Parameter reduction in soft sets is discussed by various authors. Maji et al. (2002) considered the initial level reduction of soft sets with the help of rough set approach. A new notion of parameter reduction in soft sets was introduced by Chen et al. (2005) who pointed out the errors in the previous approach. Kong et al. (2008) introduced the normal parameter reduction and its algorithm in soft sets which was later simplified by Ma et al. (2011). Kong et al. (2015) examine normal parameter reduction in soft sets based on the particle swarm algorithm. This algorithm has been successfully applied to solve combinatorial optimization problems.

One of the most popular algorithms for parameter reduction in soft sets is that proposed by Kong et al. (2008). We introduce a modification of this algorithm whereby the relevant parameters are sorted according to their importance degree and the optimal normal parameter reduction is obtained based on some threshold value. Let $X = \{h_1, h_2, \dots, h_n\}$ be a universe set, $A = \{a_1, a_2, \dots, a_m\}$ be a set of parameters and (F, A) be a soft set with tabular representation $\{h_{ij}\}$. Define $f_E(h_i) = \sum_{j:a_j \in E} h_{ij}$ for

$E \subset A$. Clearly, $f_A(h_i) = \sum_{j=1}^m h_{ij}$. Now consider a partition

$C_A = \{\{h_1, h_2, \dots, h_i\}_{f_1}, \{h_{i+1}, h_{i+2}, \dots, h_j\}_{f_2}, \dots, \{h_k, h_{k+1}, \dots, h_n\}_{f_s}\}$ of the universe set X according to all possible values of f_A , where for a subclass $\{h_v, h_{v+1}, \dots, h_{v+w}\}_{f_i}$ it holds that $f_A(h_v) = f_A(h_{v+1}) = \dots = f_A(h_{v+w}) = f_i$ and $f_1 \geq f_2 \geq \dots \geq f_s$, s being the number of subclasses. Hence, objects with the same value of $f_A(\cdot)$ are partitioned into a same subclass.

Consider a soft set (F, A) , where $A = \{a_1, a_2, \dots, a_m\}$ is the set of parameters over the initial universe set $X = \{h_1, h_2, \dots, h_n\}$, together with a partition $C_A = \{\{h_1, h_2, \dots, h_i\}_{f_1}, \{h_{i+1}, h_{i+2}, \dots, h_j\}_{f_2}, \dots, \{h_k, h_{k+1}, \dots, h_n\}_{f_s}\}$ of objects in X written in simpler notation as $C_A = \{A_{f_1}, A_{f_2}, \dots, A_{f_s}\}$. If the parameter a_i is deleted from the set A then the partition is changed and can be denoted as $C_{A-a_i} = \{\{h_1^*, h_2^*, \dots, h_i^*\}_{f_1^*}, \{h_{i+1}^*, h_{i+2}^*, \dots, h_j^*\}_{f_2^*}, \dots, \{h_k^*, h_{k+1}^*, \dots, h_n^*\}_{f_s^*}\}$ or $C_{A-a_i} = \{\overline{A-a_{if_1^*}}, \overline{A-a_{1f_2^*}}, \dots, \overline{A-a_{1f_s^*}}\}$ in simpler notation. We can now define an importance degree of a_i as

$$r_{a_i} = \frac{1}{|X|} \sum_{k=1}^s \alpha_{k,a_i}$$

where $|\cdot|$ denotes the cardinality of a set and

$$\alpha_{k,a_i} = \begin{cases} |A_{f_k} - \overline{A - a_{if_z^*}}|, & \text{if } \exists z': f_{z'} = f_k, 1 \leq z' \leq s', 1 \leq k \leq s \\ |A_{f_k}|, & \text{otherwise} \end{cases}$$

The proposed algorithm of normal parameter reduction consists of the following sequence of steps:

1. Input the soft set (F, A) ;
2. Compute the parameter importance degree $r_{a_i}, 1 \leq i \leq m$;
3. Sort the parameters according to the magnitude of their importance degrees;
4. Set some threshold value below which a parameter is discarded as not being helpful in BFP;
5. Determine the maximal subset $E \subset A$ whose elements fall short of the threshold in step 4;
6. Compute $A - E$ as the optimal normal parameter reduction.

3.4 Obtaining a tabular representation for a soft set

In order to assess the relationship between financial ratios and the status of firms, we apply the logistic regression model (LR) (Harrell, 2013) due to its simplicity and easy interpretability. The status of each firm $j, 1 \leq j \leq n$, is represented by the binary variable Y_j defined as

$$Y_j := \begin{cases} 0, & \text{if firm } j \text{ is in normal status} \\ 1, & \text{otherwise} \end{cases}$$

The LR model uses the cdf of the logistic distribution in order to model the probability $P_j \equiv P(Y_j = 1 | \mathbf{x}_j)$, where \mathbf{x}_j is the vector of covariates. Given that the resulting model is nonlinear in the parameters, and cannot therefore be estimated by OLS, the model is transformed by using the odds ratio $\frac{P_j}{1 - P_j}$ given by the function $\exp(\mathbf{x}_j^T \boldsymbol{\beta})$ which

implies that $\log \frac{P_j}{1 - P_j} = \mathbf{x}_j^T \boldsymbol{\beta}$ or $L_j = \mathbf{x}_j^T \boldsymbol{\beta}$, where $L_j = \log \frac{P_j}{1 - P_j}$ is the Logit variable

and \log denotes the natural logarithm of the variable. Coefficient β_i is the marginal effect of x_i on the log of the odds ratio $\frac{P_j}{1-P_j}$. The coefficients vector $\boldsymbol{\beta}$ is estimated by maximum likelihood (ML) based on the training data set $\{Y_j, \mathbf{x}_j\}$, $1 \leq j \leq n$, where $\mathbf{x}_j = (x_{1j}, x_{2j}, \dots, x_{mj})$ is the vector of covariates. Hence for Y_j , $\exp\left(\hat{\beta}_i x_{ij}\right)$ represents the contribution of the i -th variable of the j -th firm to its odd ratio $\frac{P_j}{1-P_j}$, where $\hat{\beta}_i$ is the maximum likelihood estimate of β_i , $1 \leq i \leq m$. If $\exp\left(\hat{\beta}_i x_{ij}\right)$ is large (greater than, or equal to, a critical value c to be determined), the i -th variable is helpful in predicting that the j -th firm is in failure status. If $\exp\left(\hat{\beta}_i x_{ij}\right) < c$ then the i -th variable is useful in predicting that the j -th firm is in normal status. Then, we use the ground truth about Y_j to verify the prediction. If the prediction from the covariate x_{ij} , $1 \leq i \leq m$, $1 \leq j \leq n$, is that the j -th firm is in normal status and $Y_j = 0$, we are sure that x_{ij} is helpful in BFP. If the prediction from x_{ij} is that the j -th firm is in failure status and $Y_j = 1$, then we also confirm that x_{ij} is useful for BFP. Otherwise, if the prediction does not coincide with the true status of the firm, the covariate x_{ij} is not useful for BFP. In summary, the following algorithmic procedure is applied:

1. LR is fitted on the training data set $\{Y_j, \mathbf{x}_j\}$ to obtain the MLE of the coefficients vector $\boldsymbol{\beta}$
2. The set of financial ratios and the set of firms are inputted
3. The tabular representation of the soft set is constructed where the entries are given by the formula

$$\chi_{ij} = \begin{cases} 1, & \text{if } \exp\left(\hat{\beta}_i x_{ij}\right) \geq c \text{ and } Y_j = 1; \\ 1, & \text{if } \exp\left(\hat{\beta}_i x_{ij}\right) < c \text{ and } Y_j = 0; \\ 0, & \text{otherwise} \end{cases}$$

with $1 \leq i \leq m$ and $1 \leq j \leq n$. The choice of c is based on optimizing the content of financial information sufficient for BFP according to step 5 of the parameter reduction

algorithm proposed in section 3.2.2. Together with the value of c , we obtain the soft set tabular representation and select the set of financial ratios using the proposed algorithm.

4 Model Specification and Presentation of Results

The Logit model estimated via maximum likelihood (ML) includes as control variables the twenty six ratios analyzed in section 3.0. The estimation results are provided in table 2.

Table 2: Binary Logit Estimation Results

Variable	Coefficient	Std. Error	z-Statistic	Prob.
CR	3.176.439	1.524.948	2.082.982	0.0373
ATR	-2.384.148	2.532.989	-0.941239	0.3466
CSR	0.090487	1.532.069	0.059062	0.9529
DIR	0.004989	0.002868	1.739.405	0.0820
CLTL	4.437.831	4.696.686	0.944886	0.3447
EL	0.371637	1.449.549	0.256381	0.7977
TDTA	1.372.461	4.191.294	3.274.553	0.0011
OECA	5.930.996	3.638.293	1.630.159	0.1031
DCL	-1.104.930	6.716.614	-1.645.070	0.1000
ITR	0.246471	0.193488	1.273.832	0.2027
RTR	0.129573	0.039932	3.244.823	0.0012
CLTR	0.583454	0.379391	1.537.870	0.1241
FATR	-1.009.585	0.657064	-1.536.508	0.1244
NWCCA	-2.017.791	5.841.922	-3.453.985	0.0006
NWCNCL	3.815.618	0.997663	3.824.557	0.0001
ETA	-3.209.138	6.442.365	-4.981.304	0.0000
NCCCL	0.090120	0.862611	0.104474	0.9168
GPM	5.377.901	4.619.034	1.164.291	0.2443
NPM	-1.993.990	1.232.840	-1.617.396	0.1058
ROA	-1.075.724	9.279.062	-1.159.303	0.2463
ROE	-0.443943	0.742458	-0.597937	0.5499
FL	0.008846	0.022354	0.395726	0.6923
FEOE	0.495017	1.638.328	0.302148	0.7625
OER	-8.659.072	5.141.928	-1.684.013	0.0922
NIWCSO	-0.046576	0.174822	-0.266420	0.7899
DTC	-1.043.300	7.022.447	-1.485.664	0.1374

In total, eight out of the twenty six coefficients turned out to be statistically significant at the 10% level of statistical significance. The prediction ability of the estimated Logit model is summarized in table 3.

Table 3: Prediction Ability of the Estimated Binary Logit Model

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)≤C	364	6	370	369	58	427
P(Dep=1)>C	5	52	57	0	0	0
Total	369	58	427	369	58	427
Correct	364	52	416	369	0	369
% Correct	98.64	89.66	97.42	100	0	86.42
% Incorrect	1.36	10.34	2.58	0	100	13.58
Total Gain*	-1.36	89.66	11.01			
Percent Gain**	NA	89.66	81.03			

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
E(# of Dep=0)	359.77	8.88	368.65	318.88	50.12	369
E(# of Dep=1)	9.23	49.12	58.35	50.12	7.88	58
Total	369	58	427	369	58	427
Correct	359.77	49.12	408.89	318.88	7.88	326.76
% Correct	97.5	84.7	95.76	86.42	13.58	76.52
% Incorrect	2.5	15.3	4.24	13.58	86.42	23.48
Total Gain*	11.08	71.11	19.24			
Percent Gain**	81.59	82.29	81.94			

*Change in "% Correct" from default (constant probability) specification

**Percent of incorrect (default) prediction corrected by equation

From the total number of 369 firms in normal status the model correctly predicts 364 of them or 98.64%. Similarly, from the total number of 58 firms in failure status 52 of them or 89.66% is correctly predicted by the estimated binary logit model. In total, the estimated model correctly predicts 97.42% of the firms as to their real status. The calculation of the financial ratios importance degrees, with the methodology described in section 3.2.2, gives the following results for each of the twenty six ratios.

Table 4: Importance Degrees of Financial Ratios

Financial Ratios			Importance Degree
	Ratios		
1	Current Ratio	CR	14.75%
2	Acid Test Ratio	ATR	85.25%
3	Cash Ratio	CSR	14.75%
4	Defensive Interval Ratio	DIR	14.75%
5	Current Liabilities to Total Liabilities	CL/TL	14.75%
6	Equity to Liabilities	E/L	13.82%
7	Total Debt to Total Asset	TD/TA	14.75%
8	Operating Expenses to Avg Current Assets	OE/CA	15.21%
9	Depreciation to Current liabilities	D / CL	84.79%
10	Inventories Turnover Ratio	ITR	14.52%
11	Receivables Turnover Ratio	RTR	14.75%
12	Current Liabilities Turnover Ratio	CLTR	14.74%
13	Fixed Asset Turnover Ratio	FATR	85.25%
14	Net Working Capital to Current Asset	NWC/CA	87.33%
15	Net Working Capital to Non - Current Liabilities	NWC/NCL	11.52%
16	Equity to Total Asset	E/TA	86.18%
17	Net Change in Cash to Current Liabilities	NCC/CL	50.00%
18	Gross Profit Margin	GPM	14.75%
19	Net Profit Margin	NPM	78.57%
20	Return on Asset	ROA	78.34%
21	Return on Equity	ROE	79.26%
22	Financial Leverage	FL	14.29%
23	Financial Expenses to Operating Expenses	FE/OE	11.75%
24	Operating Expenses/Revenues	OE/R	84.79%
25	Net Income /Weighted Avg Common Shares Outstanding	NI /WCSO	77.42%
26	Depreciation /Total Cost	D /TC	84.56%

Based on the value of the above financial ratio importance degrees we come to reject the following set of ratios: CR (14.75%), CSR(14.75%), DIR(14.75%), CLTL(14.75%), EL(13.82%), TD/TA(14.75%), OECA(15.21%), NCC/CL (50,00%), ITR(14.52%), RTR(14.75%), CLTR(14.74%), NWCNCL(11.52%), GPM(14.75%), FL(14.29%), and FEOE(11.75%) as not helpful in BFP.

According to the findings of Table 4, the implementation of soft set theory provides a smaller group of ratios which can be used as predictive variables. The 11 significant ratios cover the three pillars of the Curtis model confirming the reliability of the primary assumptions of this study in order to discriminate the total sample in two a priori groups: failed and non failed. The ratios D/CL and D/TC can be connected with the liquidity

status, since the value of sales includes depreciation as a part of the operating cost. The firm's liquidity level is enhanced by the amount of depreciation which remains in its cash items after the liquidation of its receivables. The findings of Table 4 in regard of the quality management, with FATR, NWC/CA and OE/R contribute in the evaluation of the firms' management of fixed and current assets and their performance in the cost management. In the case of profitability the significant ratios are NPM, ROA and ROE, while the ratio E/TA offers an estimation of the debt level. In essence, according to the results of soft set theory, the majority of the profitability ratios constitute a set of significant variables. This finding can be justified by the kind of the firms' sample. This study is applied in NYSE listed firms and their corporate policy which gives emphasis in the field of profitability and in their performance in the capital market (NI/WCSO).

5 Conclusions and Future Work

The purpose of this study is to provide an alternative selection process for variables which are usually used in the construction of corporate failure prediction models. According to the findings of this study, the final group of significant firms covers the basic pillars of the Courtis model. Noteworthy is the fact that the final output of soft set theory, based on the calculated importance degrees of each financial ratio, includes the majority of the profitability ratios which confirms the strategy of the listed firms to give emphasis in profitability and the performance of their stock in the stock market. Comparing the results of the simple LR model and soft set theory it seems that the first method provides eight ratios as strong predictive variables. However these variables are unable to express the theoretical framework of the Courtis model because this group of ratios does not include ratios from the fields of profitability and quality of management. Therefore, soft set theory reflects more satisfactorily the strategy of firms proving to be superior in this respect in comparison to a simple logistic regression model.

Soft set theory is a novel mathematical theory which has only recently found applications in various areas of research. Maji et al (2003) provided a first practical application of soft sets in decision making problems. As a newly emerging area of interdisciplinary research, soft set theory has attracted the attention of many researchers from around the world. A testament to this fact, is the growing number of high-quality works in soft sets that have published over the past few years. A nice overview of the emerging trends in soft set theory and related topics is Feng et al (2015). The primary advantage of soft set theory has to do with the ordering of factors in terms of their importance degree. This allows extracting only the essential information of the phenomenon under investigation. A second advantage linked with the above is that this classification need not be based on statistical hypotheses or even quantitative data. Furthermore, because at its core lies a binary representation of some set this method is robust to statistical errors or small deviations in data. Concluding, in the present study soft set theory is implemented as a tool for selecting financial ratios helpful for business failure prediction which in fact proved to be in line with financial theory. Due to its flexibility, the method can be extended to variables selection in a variety of business failure prediction models such as bankruptcy prediction, default prediction, credit ratings and a number of other topics related to possible future work.

References

- Ali, M. I., Feng, F., Liu, X., Min, W. K., & Shabir, M. (2009). "On some new operations in soft set theory". *Computers & Mathematics with Applications*, 57(9), 1547-1553.
- Almamy, J., Aston, J., Ngwa, L N (2016) "An evaluation of Altman's Z-score using cash flow ratio to predict corporate failure amid the recent financial crisis: Evidence from the UK". *Journal of Corporate Finance*, 36, 278–285.
- Altman E.I., (1977). "ZETA analysis A new model to identify bankruptcy risk of corporations", *Journal of Banking & Finance*, 1(1), 2–54.
- Altman, E. (1968). "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy". *The Journal of Finance*, 23 (4), 589 - 609.
- Altman, E., Hotchkiss, E. (2006). "*Corporate Financial Distress and Bankruptcy: Predict and Avoid Bankruptcy, Analyze and Invest in Distressed Debt*", 3rd Ed. Wiley Finance.
- Atanassov, K.T. (1994). "Operators over Interval Valued Intuitionistic Fuzzy Sets". *Fuzzy Sets and Systems*, 64(2), 159-174.
- Balcaen, S., & Ooghe, H. (2004). "35 years of studies on business failure: an overview of the classical statistical methodologies and their related problems". *Journal of Vlerick Leuven Gent Working Paper*, 15, 6.
- Beaver, W. (1966). "Financial ratios as predictors of failure". *Journal of Accounting Research*, 71-111.
- Beaver, W. H., McNichols, M. F., & Rhie, J.-W. (2005). "Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy", *Review of Accounting Studies*, 10, 93-122.
- Beynon, M J., Peel, MJ (2001) "Variable precision rough set theory and data discretisation: an application to corporate failure prediction" *Omega*, 29 561–576.
- Charitou, A., Evi Neophytou, E., Charalambous, C (2004) "Predicting corporate failure: empirical evidence for the UK". *European Accounting Review*, 13(3), 465-497.
- Chen, D., Tsang, E. C. C., Yeung, D. S., & Wang, X. (2005). "The parameterization reduction of soft sets and its applications". *Computers & Mathematics with Applications*, 49(5), 757-763.
- Christopoulos, A G., Dokas, I G., Dimitrios H. Mantzaris, D H(2013) "The estimation of corporate liquidity management using artificial neural networks". *International Journal of Financial Engineering and Risk Management*, 1(2), 193 -210.
- Cielen, A., Peeters, L., & Vanhoof, K. (2004). "Bankruptcy prediction using a data envelopment analysis". *European Journal of Operational Research*, 154(2), 526-532.
- Courtis, J K (1978) "Modeling a financial ratios categoric framework". *Journal of Business Finance and Accounting*, 5(4), 371 – 386.
- Deakin, E. (1972) "A Discriminant Analysis of Predictors of Business Failure". *Journal of Accounting Research*, 10, 167-179.
- Dimitras, A I., Slowinski, R. Susmaga C. Zopounidis, C (1999) "Business failure prediction using rough sets". *European Journal of Operational Research*, 114 263-280
- Dirickx Y., Van Landeghem G (1994) "Statistical failure prevision problems". *Tijdschrift voor Economie en Management*, 39(4), 429 - 462.
- Doumpos, M., Zopounidis, C (1999) "A multinational discrimination method for the prediction of financial distress: the case of Greece" *Multinational Finance Journal*,

- 3(2), 71–101.
- Feng, F., Cagman, N., Leoreanu-Fotea, V., & Akram, M. (2015). Emerging Trends in Soft Set Theory and Related Topics. *The Scientific World Journal*, 2015.
- Frydman, H., Altman, E. I., & KAO, D. L. (1985). “Introducing recursive partitioning for financial classification: the case of financial distress”. *The Journal of Finance*, 40(1), 269-291.
- Harrell, F. E. (2013). *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. Springer Science & Business Media.
- Hauser, R P., Booth, D(2011) “Predicting Bankruptcy with Robust Logistic Regression”. *Journal of Data Science* 9,565-584.
- Iturriaga, F J L., Sanz, I P (2015) “Bankruptcy visualization and prediction using neural networks: A study of U.S. commercial banks”. *Expert Systems with Applications*, 42, 2857–2869
- Jardin, P (2010) “Predicting bankruptcy using neural networks and other classification methods: The influence of variable selection techniques on model accuracy”. *Neurocomputing*, 73, 2047–2060.
- Kharal, A., & Ahmad, B. (2011). “Mappings on soft classes”. *New Mathematics and Natural Computation*, 7(03), 471-481.
- Kong, Z., Gao, L., Wang, L., & Li, S. (2008). “The Normal Parameter Reduction of Soft Sets and its Algorithm”. *Computers & Mathematics with Applications*, 56(12), 3029-3037.
- Kong, Z., Jia, W., Zhang, G., & Wang, L. (2015). “Normal Parameter Reduction in Soft Set Based on Particle Swarm Optimization Algorithm”. *Applied Mathematical Modelling*, 39(16), 4808 – 4820.
- Lee, K., Booth, D., & Alam, P. (2005). “A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms”. *Expert Systems with Applications*, 29(1), 1-16.
- Liang, D., Lu, C C., Tsai, C F., Shih, G A (2016) “Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study”. *European Journal of Operational Research*, 252, 561–572.
- Ma, X., Sulaiman, N., Qin, H., Herawan, T., & Zain, J.M. (2011). “A new Efficient Normal Parameter Reduction Algorithm of Soft Sets”. *Computers & Mathematics with Applications*, 62(2), 588-598.
- Maji, P. K., Biswas, R., & Roy, A. (2003). “Soft set theory”. *Computers & Mathematics with Applications*, 45(4), 555-562.
- Majumdar, P., & Samanta, S. K. (2010). “On soft mappings”. *Computers & Mathematics with Applications*, 60(9), 2666-2672.
- Majumdar, P., & Samanta, S. K. (2011). “On similarity measures of fuzzy soft sets”. *Int. J. Adv. Soft. Comput. Appl*, 2, 1-8.
- Mcleay, S. and Omar, A. (2000). “The sensitivity of prediction models to the non-normality of bounded and unbounded financial ratios”. *The British Accounting Review*, 32 (2), 213-230.
- Molodtsov, D. (1999). “Soft set theory—first results”. *Computers & Mathematics with Applications*, 37(4), 19-31.
- Ohlson, J A (1980) “Financial Ratios and the probabilistic prediction of bankruptcy Management”. *Journal of Accounting Research*, 18 (1), 109-131.
- Premachandra, I. M., Chen, Y., & Watson, J. (2011). “DEA as a tool for predicting corporate failure and success: A case of bankruptcy assessment”. *Omega*, 39(6),

620-626.

- Scherger, V., Vigier, H. P., & Barberà-Mariné, M. G. (2014). "Finding business failure reasons through a fuzzy model of diagnosis". *Fuzzy economic review*, 19(1), 45.
- Tang, T C., Chi, L C (2005) "Neural networks analysis in business failure prediction of Chinese importers: A between-countries approach". *Expert Systems with Applications*, 29, 244 – 255.
- Vigier, H. P., & Terceño, A. (2008). "A model for the prediction of "diseases" of firms by means of fuzzy relations". *Fuzzy Sets and Systems*, 159(17), 2299-2316.
- Vranas, A S (1991) "The significance of financial characteristics in predicting business failure: an analysis in the Greek context". *Foundations of Computing and Decision Sciences*, 17(4), 257 – 275.
- Wilcox, J. N(1971) "A simple theory of financial ratios as predictors of failure". *Journal of Accounting Research*, 389 – 395.
- Xu, W., Xiao, Z., Dang, Z., Yang, D., Yang, X. (2014). "Financial Ratio Selection for Business Failure Prediction using Soft Set Theory". *Knowledge-Based Systems*, 63, 59 -67.
- Yang, X., Yu, D., Yang, J., & Wu, C. (2007). *Generalization of soft set theory: from crisp to fuzzy case*. In *Fuzzy Information and Engineering* (pp. 345-354). Springer Berlin Heidelberg.
- Yeh, C C, Chi, D J, Hsu, M F (2010) "A hybrid approach of DEA, rough set and support vector machines for business failure prediction" *Expert Systems with Applications*, 37, 1535-1541.
- Youn, H., Gu , Z (2010) "Predicting Korean lodging firm failures: An artificial neural network model along with a logistic regression model" *International Journal of Hospitality Management* 29 (2010) 120–127.
- Zadeh, L.A. (1965). "Fuzzy Sets". *Information and Control*, 8(3), 338-353.
- Zmijewski, M E (1984) "Methodological Issues Related to the Estimation of Financial Distress Prediction Models". *Journal of Accounting Research*, 22, 59 -82.
- Zopounidis, C., Doumpos M(2001)"A preference disaggregation decision support system for financial classification problems" *European Journal of Operational Research*, 130(2), 402–413.

APPENDIX A

Two examples of soft sets are provided. Example 1 is drawn by Molodtsov (1999).

Example 1: A soft set is used to describe the attractiveness of the houses which Mr. Y is going to buy. In this case X is the set of houses under consideration and A the set of parameters, where each parameter is a word or sentence:

$$A = \{ \text{expensive; beautiful; wooden; cheap; in the green surroundings; modern; in good repair; in bad repair} \}$$

In this case, to define a soft set means to point out expensive houses, beautiful houses, and so on. Suppose that there are six houses in the initial universe set:

$$X = \{h_1, h_2, h_3, h_4, h_5, h_6\}$$

and express the parameter set A as:

$$A = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8\},$$

where the element a_1 stands for the parameter "expensive", a_2 stands for the parameter "beautiful", a_3 stands for the parameter "wooden", and so on until the last element a_8 which stands for the parameter "in bad repair". The soft set (F, A) is a parameterized family $\{F(a_i)\}$, $1 \leq i \leq 8$, of subsets of the initial set X . For example suppose that

$$F(a_1) = \{h_1, h_3\}, F(a_2) = \{h_1, h_2, h_3, h_4\}, F(a_3) = \emptyset, F(a_4) = \{h_5\}$$

$$F(a_5) = \{h_1, h_2, h_3, h_4, h_6\}, F(a_6) = \{h_1, h_2, h_3, h_6\}, F(a_7) = \{h_3, h_6\}, F(a_8) = \emptyset$$

The sets $F(a_i)$ may be arbitrary. Some of them may be empty or some may have a non-empty intersection as is clear with this example. Below we provide one more example of a soft set.

Example 2: A soft set is used to describe the type of risk faced by banks whose stock Mr. Z is going to buy. In this case X is the set of banks under consideration and A the set of parameters, where each parameter is a word or sentence:

$$A = \{ \text{high credit risk; low credit risk; high operational risk; low operational risk; high market risk; low market risk; high systemic risk; low systemic risk} \}$$

In this case, to define a soft set means to point out banks with a high credit risk, banks with a low credit risk and so on. Suppose that there are eight banks in the initial universe set:

$$X = \{b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8\}$$

and express the parameter set A as:

$$A = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8\}$$

where the element a_1 stands for the parameter "high credit risk", a_2 stands for the parameter "low credit risk", a_3 stands for the parameter "high operational risk", and so on until the last element a_8 which stands for the parameter "low systemic risk". The soft set (F, A) is a parameterized family $\{F(a_i)\}$, $1 \leq i \leq 8$, of subsets of the initial set X . For example suppose that

$$F(a_1) = \{b_1, b_2\}, \quad F(a_2) = \{b_3, b_4, b_5\}, \quad F(a_3) = \{b_1, b_6\}, \quad F(a_4) = \{b_2, b_4, b_5\}, \\ F(a_5) = \emptyset, \quad F(a_6) = X, \quad F(a_7) = \{b_1\}, \quad F(a_8) = \{b_3, b_4, b_5, b_6\}$$

Again as is clear with this example the sets $F(a_i)$ may be arbitrary. Some of them may be empty or some may have a non-empty intersection.

APPENDIX B

Some basic definitions regarding operations with soft sets are provided. If we denote by $SS(X, A)$ the family of all soft sets (F, A) over the universe set X , then we may introduce the following definitions:

Definition 2.2.1: Given $(F, A), (G, A) \in SS(X, A)$ we say that the pair (F, A) is a *soft subset* of (G, A) if $F(p) \subseteq G(p)$ for every $p \in A$.

Definition 2.2.2: Let $(F, A), (G, A) \in SS(X, A)$. The *soft union* of these soft sets is the soft set $(H, A) \in SS(X, A)$ where the map $H : A \rightarrow P(X)$ is defined as $H(p) := F(p) \cup G(p)$, for every $p \in A$

Definition 2.2.3: Let $(F, A), (G, A) \in SS(X, A)$. The *soft intersection* of these soft sets is the soft set $(H, A) \in SS(X, A)$, where the map $H : A \rightarrow P(X)$ is defined as $H(p) := F(p) \cap G(p)$, for every $p \in A$

Definition 2.2.4: Let $(F, A) \in SS(X, A)$. The *soft complement* of (F, A) is the soft set (H, A) , where the map $H : A \rightarrow P(X)$ is defined as $H(p) := X \setminus F(p)$, for every $p \in A$

Definition 2.2.5: The soft set $(F, A) \in SS(X, A)$, where $F(p) = \emptyset$, for every $p \in A$ is called the *A-null soft set* of $S(X, A)$