ELSEVIER

European Journal of Operational Research 114 (1999) 263–280

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

Business failure prediction using rough sets

A.I. Dimitras [a], R. Slowinski [b], R. Susmaga [b], C. Zopounidis [a,*]

[a] Decision Support Systems Laboratory, Technical University of Crete, University Campus, 73100 Chania, Greece
[b] Institute of Computing Science, Poznan University of Technology, Piotrowo 3a, 60-965 Poznan, Poland

Received 22 April 1997; accepted 3 February 1998

Abstract

A large number of methods like discriminant analysis, logit analysis, recursive partitioning algorithm, etc., have been used in the past for the prediction of business failure. Although some of these methods lead to models with a satisfactory ability to discriminate between healthy and bankrupt firms, they suffer from some limitations, often due to the unrealistic assumption of statistical hypotheses or due to a confusing language of communication with the decision makers. This is why we have undertaken a research aiming at weakening these limitations. In this paper, the rough set approach is used to provide a set of rules able to discriminate between healthy and failing firms in order to predict business failure. Financial characteristics of a large sample of 80 Greek firms are used to derive a set of rules and to evaluate its prediction ability. The results are very encouraging, compared with those of discriminant and logit analyses, and prove the usefulness of the proposed method for business failure prediction. The rough set approach discovers relevant subsets of financial characteristics and represents in these terms all important relationships between the image of a firm and its risk of failure. The method analyses only facts hidden in the input data and communicates with the decision maker in the natural language of rules derived from his/her experience.

Keywords: Business failure prediction; Rough set theory; Discriminant analysis; Decision rules; Classification

1. Introduction

Business failure prediction is a scientific field which many academic and professional people have been working for, at least, the three last decades. Also, financial organizations, such as banks, credit institutions, clients, etc., need these predictions for firms in which they have an interest (of any kind).

The first approach to predict business failure started with the use of empirical methods proposed by large banks in USA (the "five C" method, the "LAPP" method, and the "credit-men" method, see Zopounidis (1995)). Then, the financial ratios methodology was developed for the business failure prediction problem. These ratios have been long considered as objective indicators of firm failure (insolvency risk, see Beaver, 1966;

* Corresponding author. E-mail: kostas@hra.ergasya.tuc.gr

0377-2217/99/$ – see front matter
PII: S 0 3 7 7 - 2 2 1 7 (9 8) 0 0 2 5 5 - 0

Courtis, 1978; Altman, 1993). The approach of the financial ratios (also called univariate statistical approach), gave rise to the methods for business failure prediction based on the multivariate statistical analysis. In 1968 already, Altman proposed to use the discriminant analysis (a discriminant function with five financial ratios has been assessed) for predicting the business failure risk. Subsequently, the use of this method has continued to spread out to the point that today we can speak of discriminant models of predicting business failure. At the same time, however, the generalization of the discriminant analysis resulted in several critical studies (see Eisenbeis, 1977; Ohlson, 1980; Dimitras et al., 1996).

Since the work of Altman (1968), several studies proposing other methods have been used to overcome some disadvantages of the discriminant analysis and to provide higher prediction accuracy. Among these studies, we can cite the study of Ohlson (1980) using logit analysis and the study of Zmijewski (1984) using probit analysis. Frydman et al. (1985) first employed the recursive partitioning algorithm while Gupta et al. (1990) used mathematical programming methods for the business failure prediction problem. Other methods used were survival analysis by Luoma and Laitinen (1991), expert systems by Messier and Hansen (1988), neural networks by Altman et al. (1994) and multi-factor model by Vermeulen et al. (1998). Moreover, several methods were developed based on multicriteria decision aid methodology (MCDA). They classify firms into categories according to their business failure risk (see Zollinger, 1982; Zopounidis, 1987; Siskos et al., 1994; Andenmatten, 1995; Dimitras et al., 1995; Zopounidis et al., 1995). The use of multicriteria decision aid methods in the prediction of business failure circumvents many of the problems that exist when using discriminant analysis (see Eisenbeis, 1977 and Dimitras et al., 1996 among others).

Most of the methods mentioned above have already been investigated in the course of comparative studies related in several review articles (see Scott, 1981; Zavgren, 1983; Altman, 1984; Jones, 1987; Keasey and Watson, 1991 among others). Dimitras et al. (1996) and Zopounidis (1995) gave a complete review of methods used for the prediction of business failure and of new trends in this area.

Recently, new methods of predicting business failure have been developed. Due to the advancement of computer and information science, they offer to the managers of financial institutions significant aid in the evaluation and selection of viable firms for financing.

This paper presents a new method called the rough set approach for the analysis and prediction of business failure. The concept of a rough set, introduced by Pawlak (1982), proved to be an effective tool for the analysis of information tables (financial information tables) describing a set of objects (firms) by a set of multi-valued attributes (financial ratios). Rough set approach has already been used for the analysis and explanation of financing decisions in a Greek industrial development bank called ETEVA (Slowinski and Zopounidis, 1995). The aim of the present study is to test the ability of the rough set approach in predicting business failure, and to compare it with two other methods: discriminant analysis and logit analysis. This application of the rough set approach has been made with the help of the credit manager of a large Greek commercial bank.

Section 2 presents the basic concepts of the rough set theory. The application of the rough set approach on a sample of Greek firms is presented in Section 3 while the comparison of its results with those obtained by the discriminant analysis and logit analysis are presented in Sections 4 and 5 respectively. In the concluding remarks the merits of the proposed method are discussed and possible new trends in the field of business failure prediction are given.

2. Basic concepts of the rough set theory

2.1. Introductory remarks

Rough set theory was introduced by Pawlak (1982). It has attracted the attention of many researchers and practitioners all over the world, who contributed to its development and applications during the last decade (see e.g. Pawlak, 1991;

Slowinski, 1992; Ziarko, 1994; Lin and Wildberger, 1995; Pawlak et al., 1995; Pawlak and Slowinski, 1994; Slowinski, 1995). Just to name a few possible uses of rough set theory, this theory may be used to describe dependencies between attributes, to evaluate significance of attributes, and to deal with inconsistent data. As an approach to handling imperfect data (uncertainty and vagueness), it complements other theories that deal with data uncertainty, such as probability theory, evidence theory, fuzzy set theory, etc.

The rough set philosophy is founded on the assumption that with every object of the universe of discourse we associate some information (data, knowledge). Objects characterized by the same information are indiscernible in view of the available information about them. The indiscernibility relation generated in this way is the mathematical basis for the rough set theory.

Any set of all indiscernible objects is called an elementary set, and forms a basic granule of knowledge about the universe. Any set of objects being a union of some elementary sets is referred to as crisp (precise) – otherwise the set is rough (imprecise, vague). Consequently, each rough set has boundary-line cases, i.e. objects which cannot be classified with certainty as members of the set or of its complement.

Therefore, a rough set can be represented by a pair of crisp sets, called the lower and the upper approximation. The lower approximation consists of all objects which certainly belong to the set and the upper approximation contains objects which possibly belong to the set.

2.2. Information table and indiscernibility relation

For algorithmic reasons, knowledge about objects will be represented in the form of an information table. The rows of the table are labelled by *objects*, whereas columns are labelled by *attributes* and entries of the table are *attribute values*. In general, we will use the notion of attribute instead of criterion, because the former is more general than the latter; the domain (scale) of a criterion has to be ordered according to decreasing or increasing preference while the domain of the attribute does not have to be ordered. We will use the notion of criterion only when the preferential ordering of the attribute domain will be important in a given context.

Formally, by an *information table* we understand the 4-tuple $S=\langle U, Q, V, f\rangle$, where U is a finite set of objects, Q is a finite set of *attributes*, $V=\bigcup_{q\in Q} V_q$ and V_q is a domain of the attribute q, and $f: U\times Q\rightarrow V$ is a total function such that $f(x, q)\in V_q$ for every $q\in Q$, $x\in U$, called an *information function* (Pawlak, 1991).

Let $S=\langle U, Q, V, f\rangle$ be an information table and let $P\subseteq Q$ and $x, y\in U$. We say that x and y are indiscernible by the set of attributes P in S iff $f(x, q)=f(y, q)$ for every $q\in P$. Thus every $P\subseteq Q$ generates a binary relation on U, called *P-indiscernibility relation*, denoted by I_P. Obviously, I_P is an equivalence relation for any P. Equivalence classes of the relation I_P are called *P-elementary sets* in S and $I_P(x)$ denotes the P-elementary set containing the object $x\in U$.

2.3. Approximation of sets

Let $P\subseteq Q$ and $Y\subseteq U$. The *P-lower approximation* of Y, denoted by $\underline{P}Y$, and the *P-upper approximation* of Y, denoted by $\overline{P}Y$, are defined as

$$\underline{P}Y=\{x\in Y: I_P(x)\subseteq Y\}, \tag{1}$$

$$\overline{P}Y=\bigcup_{x\in Y} I_P(x). \tag{2}$$

The *P-boundary* (doubtful region) of set Y, denoted by $\mathrm{BN}_P(Y)$, is defined as

$$\mathrm{BN}_P(Y)=\overline{P}Y-\underline{P}Y. \tag{3}$$

Set $\underline{P}Y$ is the set of all elements of U, which can be certainly classified as elements of Y, employing the set of attributes P. Set $\overline{P}Y$ is the set of elements of U, which can be possibly classified as elements of Y using the set of attributes P. The set $\mathrm{BN}_P(Y)$ is the set of elements, which cannot be certainly classified to Y using the set of attributes P.

With every set $Y\subseteq U$ we can associate the *accuracy of approximation* of the set Y by P, or in short, accuracy of Y, defined as

$$\alpha_P(Y) = \frac{\mathrm{card}(\underline{P}Y)}{\mathrm{card}(\overline{P}Y)}, \tag{4}$$

where card( ) means cardinality of a set.

Let S be an information table, P a subset of attributes from Q, and let $\mathcal{Y} = \{Y_1, Y_2, \ldots, Y_n\}$ be a classification, or partition, of U. The origin of this classification is independent from attributes contained in P. Subsets Y_i, $i = 1, \ldots, n$, are classes of classification $\mathcal{Y}$. By P-lower (P-upper) approximation of $\mathcal{Y}$ in S we mean sets $\underline{P}\mathcal{Y} = \{\underline{P}Y_1, \underline{P}Y_2, \ldots, \underline{P}Y_n\}$ and $\overline{P}\mathcal{Y} = \{\overline{P}Y_1, \overline{P}Y_2, \ldots, \overline{P}Y_n\}$, respectively. The coefficient,

$$\gamma_P(\mathcal{Y}) = \frac{\sum_{i=1}^{n} \mathrm{card}(\overline{P}Y_i)}{\mathrm{card}(U)} \tag{5}$$

is called the *quality of approximation of classification $\mathcal{Y}$* by the set of attributes P, or in short, *quality of classification*. It expresses the ratio of all P-correctly classified objects to all objects in the system.

2.4. Reduction and dependency of attributes

We assume that the set of attributes $R \subseteq Q$ depends on the set of attributes $P \subseteq Q$ in S (denotation $P \to R$) iff $I_P \subseteq I_R$. Discovering dependencies between attributes is of primary importance in the rough set approach to information table analysis.

Another important issue is that of attribute reduction, which is performed in such a way that the reduced set of attributes P, $P \subseteq Q$, provides the same quality of classification $\gamma_P(\mathcal{Y})$ (cf. formula (5)) as the original set of attributes Q. The minimal subset $R \subseteq P \subseteq Q$ such that $\gamma_P(\mathcal{Y}) = \gamma_R(\mathcal{Y})$ is called *$\mathcal{Y}$-reduct* of P (or, simply, *reduct*, if there is no ambiguity in the understanding of $\mathcal{Y}$) and is denoted by $\mathrm{RED}_{\mathcal{Y}}(P)$. Let us notice that an information table may have more than one $\mathcal{Y}$-reduct. Intersection of all $\mathcal{Y}$-reducts is called the $\mathcal{Y}$-core of P, i.e. $\mathrm{CORE}_{\mathcal{Y}}(P) = \cap \mathrm{RED}_{\mathcal{Y}}(P)$. The core is a collection of the most relevant attributes in the table.

2.5. Decision rules

An information table can be seen as *decision table* assuming that $Q = C \cup D$ and $C \cap D = \varnothing$, where set C contains so called *condition attributes*, and D contains *decision attributes*.

From the decision table $S = \langle U, C \cup D, V, f \rangle$, defined as in Section 2.2, a set of *decision rules* can be derived. Let us assume that D is a singleton, i.e. $D = \{d\}$, which does not decrease the generality of further considerations. The d-elementary sets in S are denoted by Y_j ($j = 1, \ldots, n$) and called *decision classes*. Describing decision classes in terms of condition attributes from C, one gets lower and upper approximations, $\underline{C}Y_j$ and $\overline{C}Y_j$, respectively, as well as the boundary $\mathrm{BN}_C(Y_j) = \overline{C}Y_j - \underline{C}Y_j$, $j = 1, \ldots, n$, according to formulae (1)–(3).

A *decision rule* can be expressed as a logical statement:

IF *conjunction of elementary conditions*
THEN *disjunction of elementary decisions*

The elementary condition formulae over subset $A \subseteq C$ and domain V_{a_i} of attribute $a_i \in A$ are defined as: $a_i = v_i$, where $v_i \in V_{a_i}$. By cond_A we denote a conjunction of elementary condition formulae, i.e. $(a_1 = v_1) \wedge \ldots \wedge (a_r = v_r)$ for all $a_i \in A$, and by $[\mathrm{cond}_A]$ we mean the set of all objects satisfying conjunction cond_A. Obviously, if object $x \in [\mathrm{cond}_A]$ then $[\mathrm{cond}_A] = I_A(x)$.

Similarly, we define elementary decision formula $d = v_j$, where $v_j \in V_d$. By dec_D we denote a disjunction of elementary decision formulae, i.e. $(d = v_1) \vee \cdots \vee (d = v_s)$, where $1 \leqslant s \leqslant n$. If $s > 1$, then the elementary decisions indicate decision classes represented by objects belonging to the boundary of a decision class. By $[\mathrm{dec}_D]$ we understand a set of objects belonging either to C-lower approximation of decision class Y_j, if $s = 1$, or, otherwise, to C-boundary of decision class Y_j. Precisely,

$$[\mathrm{dec}_D] = \begin{cases} \underline{C}Y_j, & \text{if } s = 1, \\ BN_c(Y_j), & \text{otherwise.} \end{cases} \tag{6}$$

The decision rule "IF cond_A THEN dec_D" is *consistent* iff $[\mathrm{cond}_A] \subseteq [\mathrm{dec}_D]$. If $s = 1$, i.e. dec_D consists of one elementary decision only, the decision rule is *exact*, otherwise it is *approximate*. Approximate rules are consequences of an approximate description of decision classes in terms of blocks of objects (granules) indiscernible by condition attributes. It means that using the available knowledge, one is unable to decide whether some objects (from the boundary region) belong to a given decision class or not.

Each decision rule r is characterised by the *strength* of its suggestion, which means the number of objects satisfying the condition part of the rule (we say, *covered* by the rule) and belonging to the suggested decision class. In the case of approximate rules, the strength is calculated for each possible decision class separately. Stronger rules are usually more general, i.e. their condition parts are shorter and less specialised.

Procedures for generating decision rules from a decision table operate on inductive learning principles. The objects are considered as examples of decisions. In order to induce decision rules describing a set of objects $[\mathrm{dec}_D]$, the examples belonging to $[\mathrm{dec}_D]$ are called *positive* and all the others *negative*. A decision rule is *discriminant* if it is consistent, i.e. it distinguishes positive examples from negative ones, and it is minimal, i.e. removing any elementary condition from cond_A (producing $\mathrm{cond}_{A'}$) would result in $[\mathrm{cond}_{A'}] \nsubseteq [\mathrm{dec}_D]$ (violation of rule consistency).

It may be also interesting to look for *partly discriminant* rules (Mienko et al., 1996b). These are rules, which could cover a limited number of negative examples besides positive ones. They are characterised by a coefficient called *level of discrimination* specifying to what extent the rule is consistent, i.e. what is the ratio of positive examples to all examples covered by the rule.

Procedures for induction of decision rules from decision tables were presented by Grzymala-Busse (1992), Skowron (1993), Stefanowski and Vanderpooten (1994), Mienko et al. (1996b), and by Ziarko et al. (1993).

The existing induction algorithms use one of the following strategies:

(a) generation of a minimal set of rules covering all objects from a decision table,
(b) generation of an exhaustive set of rules consisting of all possible rules for a decision table,
(c) generation of a set of 'strong' decision rules, even partly discriminant, covering relatively many objects each but not necessarily all objects from the decision table.

2.6. Decision support using decision rules

Decision rules derived from a decision table can be used for recommendations concerning new objects. Specifically, the classification of a new object can be supported by matching its description to one of the decision rules. The matching may lead to one of four situations (cf. Slowinski and Stefanowski, 1994):

(a) the new object matches one exact rule,
(b) the new object matches more than one exact rule indicating, however, the same decision class,
(c) the new object matches one approximate rule or several rules indicating different decision classes,
(d) the new object does not match any of the rules.

In cases (a) and (b), the recommendation is univocal. In case of ambiguous matching (c), the user is informed about the total strength of all matching rules with respect to suggested decision classes.

In case of no rules matching the new object (d), one can help the user by presenting him/her a set of the rules "nearest" to the description of the new object. Slowinski (1993), proposed a distance measure based on a *valued closeness relation*, VCR, manifesting several advantageous properties. It involves indifference, strict difference and veto thresholds on particular attributes, used in concordance and discordance tests. The goals of these tests are to:

(i) characterize a group of attributes considered to be in concordance with the affirmation "*object x is close to rule y*", and assess the relative importance of this group,
(ii) characterize among the attributes, which are not in concordance with the above affirmation, the ones whose opposition is strong enough to reduce the credibility of the closeness, which would result from taking into account just the concordance, and to calculate the possible reduction that would thereby result.

The same tests have been used for construction of the outranking relation by Roy (1985).

2.7. Calculation of the valued closeness relation (VCR)

Applying the *VCR* prevents a major difference on one attribute from being compensated by a number of minor differences on other attributes.

Since a decision rule may have less elementary conditions than the description of an object to be classified, the closeness of the rule to the object is computed for attributes represented in the rule only. It is understood that there is no difference between the object and the rule on other attributes.

Let a given new object x be described by values $a_1(x), a_2(x), \dots, a_m(x)$ corresponding to attributes represented in the condition part of the rule r ($m \leqslant \text{card}(C)$). Let us assume, as is the case in our application, that all attributes are quantitative ones. Object x will be compared to each rule r in order to assess the credibility of the closeness relation x(VCR)r. The calculation of the credibility $\delta(x, r)$ of the closeness is based on the common sense: the formula for determining the value of $\delta(x, r)$ over the interval [0,1] is constructed so as to respect certain qualitative principles, and, in particular, excludes the possibility of undesired compensation. Credibility $\delta(x, r) = 1$ if the assertion x(VCR)r is well-founded; $\delta(x, r) = 0$ if there is no argument for closeness of x and r. In order to carry out the calculations, the user must express explicitly and numerically:

(i) the relative importance (weight) k_l that the user wishes to confer on attribute a_l in the calculation of concordance,

(ii) the minimum value of discordance, which gives attribute a_l the power to take all credibility away from the affirmation of the closeness, even if opposed to all the other attributes in concordance with the affirmation; it is denoted by $v_l[a_l(x)]$ and called the *veto threshold* of attribute a_l.

The comprehensive concordance index is defined as

$$C(x,r) = \sum_{l=1}^{m} k_l c_l(x,r) \Bigg/ \sum_{l=1}^{m} k_l, \tag{7}$$

where $c_l(x, r)$ is a partial concordance index for attribute a_l. Calculation of $c_l(x, r)$ involves two thresholds: $0 \leqslant q_l[a_l(x)] \leqslant p_l[a_l(x)]$, called *indifference* and *strict difference thresholds*, respectively.

The weights and the thresholds defined above come from the domain knowledge of the user (expert, decision maker). The definition of concordance index $c_l(x, r)$ and discordance index $d_l(x, r)$ is given graphically in Fig. 1.

The degree of credibility $\delta(x, r)$ of the closeness relation x(VCR)r is obtained from the comprehensive concordance index weakened by discordance indices (up to the point of its annulment):

$$\delta(x,r) = C(x,r)\prod_{l\in L}\frac{1-d_l(x,r)}{1-C(x,r)},$$
$$L = \{l\colon d_l(x,r) > C(x,r)\}. \tag{8}$$

The rules with the greatest values of $\delta(x, r)$ are presented to the user together with information

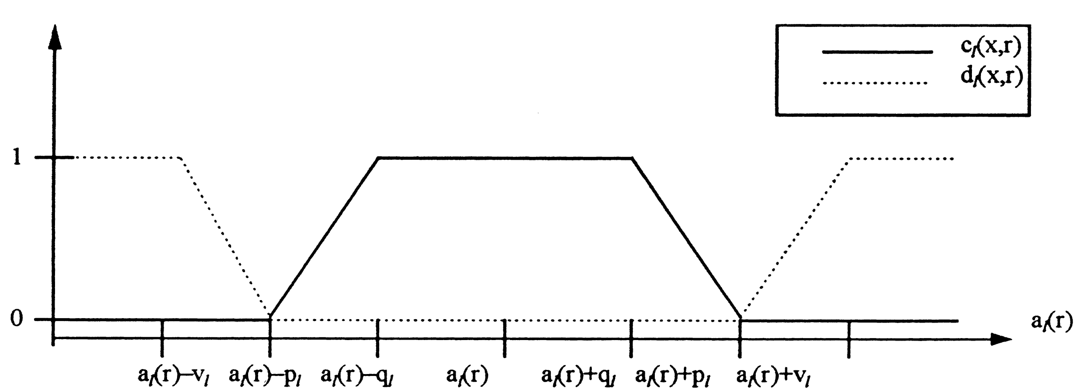

Fig. 1. Concordance and discordance indices for object x and rule r, with respect to attribute a_l.

about the strength of the decision classes being suggested.

3. Application of the rough set approach

3.1. The data

A large number of firms which failed in Greece in the years 1986–1990 were collected. From this large set, 40 firms from 13 industries meeting the criteria of (a) having been in business for more than five years and (b) data availability were selected. The number of firms in each industry is presented in Table 1. The financial statements of these firms were collected for a period of five years, starting from year −5 (five years before bankruptcy) and ending with the year −1 (one year prior to the year of bankruptcy, the last year that the firm had been in business). Obviously, the actual year of bankruptcy (year 0) is not the same for all the firms, as they did not all fail in the same year.

The 40 failed firms were matched one by one to 40 "healthy" firms, i.e. firms that did not fail for bankruptcy. The healthy firms were chosen among those of the same industry and having also similar total assets and number of employees for the year −1 to the corresponding failed firm. Similarly to the failed firms, the financial statements for the healthy ones were also collected for five years. This way of composing the sample of firms was also used by several researchers in the past, e.g. Beaver (1966), Altman (1968), Zavgren (1985), among others. Its aim is to minimise the effect of such factors as industry or size that in some cases can be very important.

Except from the above learning sample a second testing sample was collected using a similar approach. The testing sample consisted of 19 firms that went bankrupt in the period from 1991 to the mid of 1993, as well as 19 healthy firms. The number of the bankrupt firms in each industry is presented in Table 2.

Using the financial statements of the firms (i.e. balance sheets and income statements), 28 financial ratios were calculated. These are presented in Table 3.

The small number of ratios is due to missing data and no availability of the sales volume reported by Greek firms, mainly for taxation reasons. Unfortunately, qualitative characteristics of the firms providing satisfactory reliability were not available, as is the case in many countries. More details about the design of the samples, their characteristics and summary statistics are given in Dimitras (1995).

For the application of the rough set approach, the credit manager of a large Greek bank was employed to act as a decision maker (DM). The DM played an important role in:

- the choice of the attributes (financial ratios) entering the information table,

Table 1
Number of bankrupt firms in each industry in the learning sample

Industry	Number of firms
Food	2
Textile	12
Wear and footwear	3
Wood	1
Paper	1
Impression-publications	2
Plastics	2
Chemical	2
Nonmetallic minerals	6
Metallurgical	2
Metallic objects	3
Transport vehicles	1
Various	3
Total	40

Table 2
Number of bankrupt firms in each industry in the testing sample

Industry	Number of firms
Food	1
Textile	7
Wear and footwear	2
Wood	1
Plastics	1
Chemical	2
Nonmetallic minerals	1
Metallic objects	3
Transport vehicles	1
Total	19

Table 3
Initial set of financial ratios

Net Income / Gross Profit
Gross Profit / Total Assets
Net Income / Total Assets
Net Income / Net Worth
Net Income / (Long Term Debt + Current Liabilities)
Inventories / Total Assets
Inventories / Current Assets
Current Liabilities / (Long Term Debt + Current Liabilities)
Net Fixed Assets / Total Assets
Current Assets / Current Liabilities
Quick Assets / Current Liabilities
Working Capital / Total Assets
Working Capital / Current Assets
(Long Term Debt + Current Liabilities) / Net Worth
(Long Term Debt + Current Liabilities) / Net Fixed Assets
Net Worth / (Net Worth + Long Term Debt)
Net income / Working Capital
Current Liabilities / Inventories
Current Liabilities / Net Worth
Net Worth / Net Fixed Assets
Inventories / Working Capital
(Long Term Debt + Current Liabilities) / Working Capital
Net Worth / Total Assets
Current Liabilities / Total Assets
Quick Assets / Total Assets
Working Capital / Net Worth
Current Assets / Total Assets

- the discretization of the continuous attributes, by setting norms dividing the original domains of the attributes into sub-intervals,
- the selection of a satisfactory reduct of attributes from among all reducts calculated for the learning sample, and
- the test of decision rules on the testing sample.

Firstly, from the set of the 28 available financial ratios, 12 were selected by the DM to enter the information table (see Table 4). This choice was justified by:

(i) the fact that the selected ratios represent well all three categories proposed by Courtis (1978), i.e. (a) profitability, (b) managerial performance and (c) solvency ratios,

(ii) a primary analysis of the characteristics of the two groups of firms (bankrupt versus healthy firms) presented to the DM, and

(iii) the preferences of the DM, his knowledge, his experience about the Greek economy and Greek industrial firms, as well as the commercial policy of the bank he is working for.

Table 4
Attributes (financial ratios) considered in the information table

Attribute	
a_1	Net Income / Gross Profit
a_2	Gross Profit / Total Assets
a_3	Net Income / Total Assets
a_4	Net Income / Net Worth
a_5	Current Assets / Current Liabilities
a_6	Quick Assets / Current Liabilities
a_7	(Long Term Debt + Current Liabilities) / Total Assets
a_8	Net Worth / (Net Worth + Long Term Debt)
a_9	Net Worth / Net Fixed Assets
a_{10}	Inventories / Working Capital
a_{11}	Current Liabilities / Total Assets
a_{12}	Working Capital / Net Worth

Then, the DM was asked to discretize the continuous financial ratios providing norms according to his knowledge and the primary analysis of the groups. The discretization is performed because the precision of financial ratios is rather doubtful and, moreover, it prevents drawing general conclusions from data in terms of dependencies, reducts and decision rules.

The sub-intervals proposed for discretization are presented in Table 5. Firms for which the values of these financial ratios are in the same sub-intervals are supposed to have very similar characteristics and behaviour. The codes that are used to refer to each sub-interval do not represent any preferential ordering. The rough set theory does not take into account any order, since it is based on the indiscernibility relation, while furthermore for the same reason the selection of the codes does not affect the obtained results.

The coded information table prepared for further analysis consisted of 80 firms (objects) belonging to the learning sample; they were described by 12 coded attributes (financial ratios), using data from one year before bankruptcy (year –1), and by the binary assignment to a decision class (healthy or bankrupt, coded by 1 and 0, respectively).

3.2. Presentation of the rough set results

The rough set analysis of the coded information table has been performed using the systems

Table 5
Sub-intervals (norms) and their codes defined for the 12 attributes

Attribute	Interval / code 1	2	3	4	5
a_1	$(-\infty, 0.00]$	$(0.00, 0.25]$	$(0.25, 1.00]$	$(1.00, +\infty)$	
a_2	$(-\infty, 0.00]$	$(0.00, 0.25]$	$(0.25, 0.50]$	$(0.50, +\infty)$	
a_3	$(-\infty, -0.05]$	$(-0.05, 0.05]$	$(0.05, 0.20]$	$(0.20, +\infty)$	
a_4	$(-\infty, 0.00]$	$(0.00, 0.25]$	$(0.25, +\infty)$		
a_5	$[0.00, 0.70]$	$(0.70, 1.00]$	$(1.00, 1.50]$	$(1.50, 2.00]$	$(2.00, +\infty)$
a_6	$[0.00, 0.50]$	$(0.50, 0.80]$	$(0.80, 1.00]$	$(1.00, 1.20]$	$(1.20, +\infty)$
a_7	$[0.00, 0.667]$	$(0.667, 0.80]$	$(0.80, 1.00]$	$(1.00, +\infty)$	
a_8	$(-\infty, 0.00]$	$(0.00, 0.50]$	$(0.50, 1.00]$		
a_9	$(-\infty, 0.00]$	$(0.00, 0.30]$	$(0.30, +\infty)$		
a_{10}	$(-\infty, 0.00]$	$(0.00, 0.50]$	$(0.50, 0.75]$	$(0.75, 1.00]$	$(1.00, +\infty)$
a_{11}	$[0.00, 0.25]$	$(0.25, 0.50]$	$(0.50, 0.75]$	$(0.75, 1.00]$	$(1.00, +\infty)$
a_{12}	$(-\infty, 0.00]$	$(0.00, 0.50]$	$(0.50, +\infty)$		

RoughDAS and ProFIT (Mienko et al., 1996a). It produced the following results.

(1) The whole set of attributes C provided perfect approximation of the decision classes as well as the quality of classification. It means that $\alpha_c(Y^1) = 1$ and $\alpha_c(Y^0) = 1$, where Y^1 and Y^0 are the decision classes and, therefore, $\gamma_c(\{Y^1, Y^0\}) = 1$ (quality of approximation of objects' classification is equal to one); cf. formulae (4) and (5).

(2) The core of attributes was empty (cf. Section 2.4). This indicates that no single attribute is absolutely necessary for perfect approximation of the decision classes. Non-empty core would indicate that there are attributes in the system which are indispensable from the discriminating point of view, because removal of any of the attributes contained in the core leads immediately to the decrease of the quality of approximation. On the other hand, a non-empty core helps in determining the most important attributes as far as the approximation of classes is concerned.

(3) 54 reducts were obtained for the coded information table (cf. Section 2.4). They contain 5–7 attributes, which is considerably smaller than 12 – the total number of attributes. This result gives the idea of reduction a strong support because each of the reducts contains fewer attributes, but, on the other hand, ensures the same value of quality of approximation as the whole set of attributes Q (equal to 1.0). The attribute with the highest frequency of occurrence in reducts is a_{11} (47 reducts) and with the lowest frequency: a_3 (15 reducts). The reducts are presented in Table 6.

Table 6
Reducts of the coded information table

#	Reduct	#	Reduct
1	$\{a_4, a_8, a_{10}, a_{11}, a_{12}\}$	28	$\{a_1, a_5, a_7, a_8, a_{11}, a_{12}\}$
2	$\{a_1, a_7, a_9, a_{10}, a_{11}, a_{12}\}$	29	$\{a_1, a_2, a_6, a_7, a_{12}\}$
3	$\{a_1, a_3, a_8, a_{10}, a_{11}, a_{12}\}$	30	$\{a_1, a_2, a_5, a_7, a_{11}\}$
4	$\{a_1, a_7, a_8, a_{10}, a_{11}, a_{12}\}$	31	$\{a_1, a_4, a_5, a_9, a_{11}, a_{12}\}$
5	$\{a_4, a_7, a_9, a_{10}, a_{11}\}$	32	$\{a_1, a_4, a_5, a_8, a_{11}, a_{12}\}$
6	$\{a_4, a_5, a_9, a_{10}, a_{11}, a_{12}\}$	33	$\{a_1, a_3, a_5, a_7, a_{11}\}$
7	$\{a_1, a_3, a_6, a_{11}, a_{12}\}$	34	$\{a_2, a_4, a_6, a_7, a_9, a_{12}\}$
8	$\{a_1, a_3, a_5, a_{11}, a_{12}\}$	35	$\{a_2, a_4, a_5, a_6, a_9, a_{12}\}$
9	$\{a_3, a_4, a_6, a_9, a_{11}, a_{12}\}$	36	$\{a_1, a_2, a_3, a_5, a_6, a_{12}\}$
10	$\{a_2, a_4, a_6, a_9, a_{11}\}$	37	$\{a_2, a_3, a_4, a_6, a_9, a_{10}, a_{12}\}$
11	$\{a_1, a_4, a_6, a_9, a_{11}, a_{12}\}$	38	$\{a_1, a_2, a_4, a_6, a_{10}, a_{12}\}$
12	$\{a_1, a_2, a_7, a_{10}, a_{11}\}$	39	$\{a_1, a_2, a_4, a_5, a_6, a_{12}\}$
13	$\{a_1, a_2, a_4, a_6, a_{11}, a_{12}\}$	40	$\{a_1, a_6, a_7, a_9, a_{10}, a_{11}\}$
14	$\{a_1, a_2, a_4, a_5, a_{11}\}$	41	$\{a_1, a_6, a_7, a_8, a_{10}, a_{11}\}$
15	$\{a_1, a_3, a_7, a_{10}, a_{11}\}$	42	$\{a_1, a_5, a_6, a_7, a_9, a_{11}\}$
16	$\{a_4, a_5, a_7, a_9, a_{11}\}$	43	$\{a_1, a_5, a_6, a_7, a_8, a_{11}\}$
17	$\{a_3, a_4, a_5, a_9, a_{11}, a_{12}\}$	44	$\{a_1, a_2, a_3, a_8, a_{10}, a_{11}\}$
18	$\{a_2, a_4, a_5, a_9, a_{11}\}$	45	$\{a_1, a_2, a_3, a_6, a_{10}, a_{11}\}$
19	$\{a_4, a_6, a_8, a_{11}, a_{12}\}$	46	$\{a_1, a_2, a_3, a_5, a_{11}\}$
20	$\{a_4, a_5, a_7, a_8, a_{11}\}$	47	$\{a_1, a_2, a_6, a_7, a_9, a_{11}\}$
21	$\{a_3, a_4, a_5, a_8, a_{11}, a_{12}\}$	48	$\{a_1, a_2, a_3, a_6, a_9, a_{11}\}$
22	$\{a_2, a_4, a_5, a_8, a_{11}\}$	49	$\{a_2, a_4, a_8, a_{10}, a_{11}\}$
23	$\{a_4, a_6, a_7, a_9, a_{11}, a_{12}\}$	50	$\{a_4, a_7, a_8, a_{10}, a_{11}\}$
24	$\{a_4, a_5, a_6, a_9, a_{11}, a_{12}\}$	51	$\{a_1, a_2, a_4, a_6, a_{10}, a_{11}\}$
25	$\{a_1, a_6, a_7, a_9, a_{11}, a_{12}\}$	52	$\{a_2, a_4, a_6, a_8, a_{11}\}$
26	$\{a_1, a_5, a_7, a_9, a_{11}, a_{12}\}$	53	$\{a_1, a_2, a_6, a_7, a_8, a_{11}\}$
27	$\{a_1, a_6, a_7, a_8, a_{11}, a_{12}\}$	54	$\{a_1, a_2, a_3, a_6, a_8, a_{11}\}$

Table 7
The minimal set of decision rules

	Elementary conditions					Decision	
Rule #	a_4	a_5	a_7	a_9	a_{11}	d	Strength
#1	1		3			0	15
#2		1				0	12
#3			4			0	12
#4			2		4	0	4
#5	1	3				0	10
#6	2		3			0	3
#7		2	2			0	3
#8				2		0	9
#9		2	1			0	1
#10	2		1			1	18
#11	3			3		1	18
#12	2	3			3	1	5
#13		4	2			1	1

(4) The reducts were presented to the DM who was asked to select the one that best fits his/her preferences. This selection was made taking into account two criteria:

(i) the reduct should contain as small a number of attributes as possible,

(ii) the reduct should not miss the attributes judged by the DM as the most significant for evaluation of the firms.

(5) The reduct selected was the #16, which includes: a_4 (profitability ratio), a_5, a_7, a_9 (solvency ratios) and a_{11} (managerial performance ratio).

(6) The remaining attributes were then eliminated from the coded information table and a set of decision rules has been derived from the reduced table. Since the quality of classification was equal to 1.0 (this is guaranteed by reducing the decision table using one of the generated reducts, which always ensures the quality of the approximation to be equal to the original value), the boundaries of decision classes were empty and thus all decision rules were exact (one elementary decision per rule); cf. Section 2.5.

(7) The decision rules have been derived from the reduced coded information table according to all three possible strategies listed at the end of Section 2.5.

(i) the minimal set of rules is composed of 13 rules and it is presented in Table 7,

(ii) the exhaustive set of rules, composed of 45 rules, was discarded because of low readability and relatively poor classification accuracy in the reclassification tests,

(iii) the set of 'strong' and discriminant rules supported by at least eight objects each, is presented in Table 8; assuming the minimum level of discrimination equal to 90% and the minimum strength equal to 8 we obtained 10 rules presented in Table 9.

Let us observe that the shorter the rule, the stronger it is. In all three sets of decision rules the number of rules describing the bankrupt firms ($d = 0$) is greater than the number of rules describing the healthy firms ($d = 1$) – 6 or 9 vs. 4 or 5. This means that it is harder to generalize the description of the bankrupt firms than the healthy ones.

Before testing the decision rules, the DM was asked to provide information necessary for the definition of the VCR. The thresholds and the weights corresponding to the five attributes are presented in Table 10. According to the DM's preferences, the most important attributes are a_5 and a_{11}, followed by a_7, a_9 and a_4.

3.3. Reclassification tests using the three sets of decision rules and VCR

The three sets of decision rules, shown in Tables 7–9, were tested first on the firms from the learning sample, and then applied to data from the

Table 8
The set of 'strong' decision rules (strength >= 8)

	Elementary conditions					Decision	
Rule #	a_4	a_5	a_7	a_9	a_{11}	d	Strength
#1		1				0	12
#2			4			0	12
#3				1		0	12
#4					5	0	9
#5				2		0	9
#6	1		3			0	15
#7	1	2				0	11
#8	1	3				0	10
#9	1				4	0	12
#10		5				1	12
#11	2		1			1	18
#12			1		2	1	13
#13	3	3				1	9
#14	3			3		1	18

Table 9
The set of 'strong' and partly discriminant decision rules (strength >= 8, level of discrimination >= 90%)

	Elementary conditions					Decision		Decision	
Rule #	a_4	a_5	a_7	a_9	a_{11}	d	Strength	d	Strength
#1		1						0	12
#2			4					0	12
#3				1				0	12
#4					5			0	9
#5				2				0	9
#6	1					1	1	0	33
#7	2				4	1	11	0	1
#8	3					1	18	0	1
#9			1			1	24	0	1
#10		5				1	12		

Table 10
Parameters of the VCR

Parameter	Attribute				
	a_4	a_5	a_7	a_9	a_{11}
q_l	0	0	0	0	0
p_l	1	3	1	1	3
v_l	2	4	3	2	4
k_l	1.3	1.5	1.4	1.3	1.5

years −2, −3, −4 and −5, i.e. 2, 3, 4 and 5 years before the state described in the information table.

Next, the three sets of decision rules were used for the classification of firms from the testing sample in years −1, −2 and −3.

All reclassification tests were run twice: first without and then with the VCR technique for classification of objects which did not match any rule during the first run. This gave a clear insight into the efficiency of the VCR technique. On the average, 60% of objects not classified by exactly matching rules were classified correctly by the VCR-nearest rules, while 40% were classified incorrectly. This is a better result than random classification of these objects.

Another general conclusion is that the two weak rules #9 and #13 (each supported by only one object) were almost never used in the reclassification test. This means that the corresponding

Table 11
Classification accuracy for the minimal set of rules

Classification accuracy	Learning sample					Testing sample		
	Year − 1	Year − 2	Year − 3	Year − 4	Year − 5	Year − 1	Year − 2	Year − 3
Bankrupt firms	100%	85.0%	82.1%	82.5%	71.1%	84.2%	57.9%	42.1%
Healthy firms	100%	87.5%	71.8%	75.0%	78.9%	57.9%	52.6%	68.4%
Total	100%	86.3%	76.9%	78.8%	75.0%	71.1%	55.3%	55.3%

Table 12
Classification accuracy for the set of 'strong' rules

Classification accuracy	Learning sample					Testing sample		
	Year − 1	Year − 2	Year − 3	Year − 4	Year − 5	Year − 1	Year − 2	Year − 3
Bankrupt firms	97.5%	85.0%	79.5%	72.5%	65.8%	73.7%	47.4%	36.8%
Healthy firms	97.5%	85.0%	87.2%	80.0%	81.6%	57.9%	68.4%	68.4%
Total	97.5%	85.0%	83.3%	76.3%	73.7%	65.8%	57.9%	52.6%

supporting objects no. 25 and 73 are marginal in this case.

The classification accuracies in percent of correctly classified objects by the three sets of rules for the learning sample in 5 years prior to the reference year and for the testing sample in 3 years prior to the reference year are summarised in Tables 11–13.

Going back from year −1, the classification accuracy is usually decreasing faster for bankrupt firms than for healthy ones. This was expected because several years prior to the bankruptcy, characteristic problems for bankrupt firms could not yet be present, while the healthy firms should have relatively stable and good performance during all five years.

The classification accuracy for the testing sample is not so good as for the learning sample. In general, however, the results obtained are quite satisfactory. It can be observed that while the set of 'strong' and partly discriminant rules performs slightly better on the testing sample, the minimal set of rules performs slightly better on the learning sample. Indeed, the differences are rather small.

4. Comparison of the rough set approach with discriminant analysis

Although the philosophy of the rough set approach and discriminant analysis is very different, their comparison on a common set of data is well founded because the two methodologies can be applied to the business failure prediction problem. The choice of discriminant analysis has been made since this method was the first and the most

Table 13
Classification accuracy for the set of 'strong', partly discriminating rules

Classification accuracy	Learning sample					Testing sample		
	Year − 1	Year − 2	Year − 3	Year − 4	Year − 5	Year − 1	Year − 2	Year − 3
Bankrupt firms	95.0%	85.0%	82.1%	75.0%	71.1%	94.7%	78.9%	42.1%
Healthy firms	90.0%	85.0%	79.5%	72.5%	78.9%	57.9%	42.1%	57.9%
Total	92.5%	85.0%	80.8%	73.8%	75.0%	76.3%	60.5%	50.0%

widely used for classification and prediction in financial management problems (see Altman et al., 1981; Altman, 1993; Dimitras et al., 1995). Discriminant analysis is a parametric method and must respect several statistical assumptions such as normality of distributions of the variables, equality of variance–covariance matrices of the variables for the two classes of firms and specification of a priori probabilities and/or costs of misclassification. It is hard indeed to meet all these assumptions in practice; this is why their validity has been questioned by several authors (see Eisenbeis, 1977; Ohlson, 1980; Zavgren, 1985). Unlike discriminant analysis, the rough set approach does not need any assumptions about data prior to the analysis (see Krusinska et al. (1992) and Stefanowski (1992) for a critical discussion of the two methods).

In order to compare the two methods, the same set of data has been used in its original (non-discretised) form. Two discriminant functions were derived. One for the reduced information table, i.e. with attributes from the same reduct as in the rough set analysis and another for the complete information table, i.e. with all 12 attributes. The discriminant functions' coefficients are shown in Tables 14 and 15.

Tables 16 and 17 provide classification accuracy results in the reclassification test with both discriminant functions applied to the learning and the testing samples, respectively.

It can be observed that better results are obtained for the discriminant function derived from the complete information table, however, these results are less satisfactory than those obtained using the rough set approach. More specifically, the comparison of discriminant analysis with the rough set approach either in case of 5 attributes (Table 11) or 12 attributes (Tables 12 and 13) reveals the superiority of rough sets over discriminant analysis, except for year −3 of the testing sample. For example, the advantage of the minimal set of decision rules over the discriminant function derived from the reduced information table ranges from +15% to −2.6% with the average of +6.1% per case.

Table 14
Discriminant function's coefficients for the attributes of the reduced information table

Attribute	a_4	a_5	a_7	a_9	a_{11}	Constant
Coefficient	0.3061	0.8633	−2.0721	0.0804	1.1754	−0.3297

Table 15
Discriminant function's coefficients for the attributes of the complete information table

Attribute	a_1	a_2	a_3	a_4	a_5	a_6	
Coefficient	0.0093	1.9154	2.4196	0.1245	1.2882	−0.9008	
Attribute	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	Constant
Coefficient	−0.7149	0.0004	0.0342	−0.0168	0.6294	0.0022	−1.1510

Table 16
Classification accuracy for the discriminant function defined on the attributes of the reduced information table

Classification accuracy	Learning sample					Testing sample		
	Year − 1	Year − 2	Year − 3	Year − 4	Year − 5	Year − 1	Year − 2	Year − 3
Bankrupt firms	92.5%	75.0%	67.5%	65.0%	65.0%	47.4%	42.1%	42.1%
Healthy firms	77.5%	82.5%	75.0%	80.0%	72.5%	73.7%	68.4%	73.7%
Total	85.0%	78.8%	71.3%	72.5%	68.8%	60.5%	55.3%	57.9%

Table 17
Classification accuracy for the discriminant function defined on the attributes of the complete information table

Classification accuracy	Learning sample					Testing sample		
	Year − 1	Year − 2	Year − 3	Year − 4	Year − 5	Year − 1	Year − 2	Year − 3
Bankrupt firms	87.5%	75.0%	67.5%	55.0%	55.0%	63.2%	42.1%	36.8%
Healthy firms	92.5%	87.5%	87.5%	85.0%	80.0%	68.4%	73.7%	73.7%
Total	90.0%	81.3%	77.5%	70.0%	67.5%	65.8%	57.9%	55.3%

Moreover, from the user's point of view, the decision rules express the dependencies between financial characteristics of a firm and its status in a much more convenient way than the discriminant function. The decision rules speak the natural language of decision examples given by the expert, while the discriminant function gives only a global view.

The above comparative results give evidence of the ability of the rough set approach to respond efficiently to the business failure prediction problem, being a reliable alternative to discriminant analysis.

5. Comparison of the rough set approach with logit analysis

The problems on the application of discriminant analysis in real world financial classification problems led many financial researchers to the exploitation of the capabilities of other multivariate statistical methods to overcome these limitations. Among others, conditional probability models, including logit and probit analysis have been used as alternatives to discriminant analysis. Both logit and probit models provide the probability of occurrence of an outcome described by a dichotomous (or polytomous) dependent variable using coefficients of the independent variables (Zavgren, 1985; Vranas, 1992). The main difference between logit and probit analysis is that the former uses the cumulative logistic probability function, while the latter is based on the cumulative standard normal distribution function. A significant advantage of these models over discriminant analysis is that they do not require the independent variables to be multivariate normal (Keasey and Watson, 1991). Furthermore, they provide much more useful information to the financial/credit analysts in bankruptcy prediction, since except for the classification of the firms into bankrupt and non-bankrupt ones, they also provide the probability of failure of a firm.

Having in mind these comparative advantages of conditional probability models in bankruptcy prediction over discriminant analysis, it was also decided to compare the rough set approach to logit analysis. Logit analysis was preferred to probit analysis, since its form is quite similar to the cumulative normal function, while it is computationally more tractable (probit analysis requires non-linear estimation, cf. Altman et al. (1981) and Gloubos and Grammatikos (1988)).

To apply logit analysis in this case study, as for discriminant analysis, the original (non-discretised) set of data has been used. Following the methodology used in the application of rough sets and discriminant analysis, two logit models were developed using the maximum likelihood estimation procedure. These two models correspond to the reduced and the complete information table, respectively.

Tables 18 and 19 present the parameters of the two logit models based on the attributes of the reduced and the complete information tables, while Tables 20 and 21 present the classification accuracy results in the reclassification test with both models applied to the learning and the testing samples, respectively. At this point it should be pointed out that the negative coefficients of some ratios in the developed logit model indicate that these ratios are negatively correlated with the probability of failure (they decrease the bankruptcy risk), while the variables with positive

Table 18
Logit model's coefficients for the attributes of the reduced information table

Attribute	a_4	a_5	a_7	a_9	a_{11}	Constant
Coefficient	−0.0870	−1.4837	7.9405	−0.4841	−2.4980	−2.2071

Table 19
Logit model's coefficients for the attributes of the complete information table

Attribute	a_1	a_2	a_3	a_4	a_5	a_6	
Coefficient	−0.2755	−3.1684	−16.1094	−0.0242	−4.6816	3.8213	
Attribute	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	Constant
Coefficient	8.6756	0.0278	−0.3657	0.2321	−4.0971	0.0069	0.1456

Table 20
Classification accuracy for the logit model defined on the attributes of the reduced information table

Classification accuracy	Learning sample					Testing sample		
	Year − 1	Year − 2	Year − 3	Year − 4	Year − 5	Year − 1	Year − 2	Year − 3
Bankrupt firms	87.50%	75.00%	70.00%	65.00%	62.50%	63.16%	31.58%	36.84%
Healthy firms	80.00%	80.00%	80.00%	80.00%	75.00%	57.89%	84.21%	84.21%
Total	83.75%	77.50%	75.00%	72.50%	68.75%	60.53%	57.89%	60.53%

Table 21
Classification accuracy for the logit model defined on the attributes of the complete information table

Classification accuracy	Learning sample					Testing sample		
	Year − 1	Year − 2	Year − 3	Year − 4	Year − 5	Year − 1	Year − 2	Year − 3
Bankrupt firms	92.50%	77.50%	77.50%	65.00%	67.50%	63.16%	42.11%	36.84%
Healthy firms	87.50%	87.50%	80.00%	85.00%	82.50%	57.89%	89.47%	73.68%
Total	90.00%	82.50%	78.75%	75.00%	75.00%	60.53%	65.79%	55.26%

coefficients have a positive effect on the probability of failure (they increase the bankruptcy risk).

According to these results, it is clear that the logit model developed on the attributes of the complete information table provides higher classification accuracy than the model which is based on attributes of the reduced information table. Compared to discriminant analysis, the two logit models are superior in all years, except year −1 for the testing sample. On the contrary, the comparison with the rough set approach: (a) in case of 5 attributes reveals the dominance of the rough set over logit analysis, except for years −2 and −3 of the testing sample, and (b) in case of 12 attributes reveals the dominance of the rough set over logit analysis, except for years −5 and −4 of the learning sample in Tables 12 and 13, respectively, and years −2 and −3 of the testing sample.

Moreover, one should also consider some significant drawbacks in the application of logit analysis in financial decision making, involving mainly the interpretation of the models parameters, especially in the case of multi-group classification problems (Altman et al., 1981) which are very common in finance (Zopounidis, 1987).

6. Concluding remarks

In this study, the rough set method is proposed as an operational decision tool for the prediction of business failure in Greece. This method, especially conceived for multi-attribute classification problems, suits well the problem of business failure prediction. The prediction model has the form of decision rules. The decision rules take into account preferences of the DM (in our case, the credit manager of a Greek commercial bank), who takes part in the construction process. Moreover, the derived decision rules reveal the most relevant attributes which should be considered by the credit manager in order to evaluate the risk of failure of a firm. In the present study, the rough set analysis has underlined the importance of the financial profitability, liquidity, debt capacity and working capital ratios. It is important to mention that the rules were derived from a particular data set and as such they represent a generalised description of the experience of one particular bank. Following this, the rules cannot be applied uncritically to other banks. If such a need arises, however, a new data set may be created and the rough set method can be used to analyse this data set and generate appropriate rules, that will represent the experience of any bank being considered.

Concerning the classification of firms considered by the credit manager of the Greek commercial bank, the rough set method produced very satisfactory results. The results were generally better than those obtained by the classical discriminant analysis and logit analysis, although the superiority over logit analysis was not so distinct as that over discriminant analysis. This result is very important because rough set approach becomes, for the future, a strong alternative tool for the analysis of financial management problems.

Finally, compared to other existing methods, rough set approach offers the following advantages:

- It discovers important facts hidden in data and expresses them in the natural language of decision rules.
- It accepts both quantitative and qualitative attributes and specifies their relevance for approximation of the classification.
- It can contribute to the minimization of the time and cost of the decision making process (rough set approach is an information processing system in real time).
- It offers transparency of classification decisions, allowing for their argumentation.
- It takes into account background knowledge of the decision maker.
- It can be incorporated into an integrated DSS for the evaluation of corporate performance and viability (see Siskos et al., 1994; Zopounidis et al., 1992, 1996a, b).

References

Altman, E.I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The Journal of Finance 23, 589–609.

Altman, E.I., Avery, R., Eisenbeis, R., Stinkey, J., 1981. Application of classification techniques in business, banking and finance. Contemporary Studies in Economic and Financial Analysis, vol. 3. JAI Press, Greenwich.

Altman, E.I., 1984. The success of business failure prediction models: An international survey. Journal of Banking and Finance 8 (2), 171–198.

Altman, E.I., 1993. Corporate Financial Distress and Bankruptcy. John Wiley, New York.

Altman, E.I., Marco, G., Varetto, F., 1994. Corporate distress diagnosis: Comparisons using discriminant analysis and neural networks (the Italian experience). Journal of Banking and Finance 18, 505–529.

Andenmatten, A., 1995. Evaluation de risque de défaillance des émetteurs d́obligations: Une approche par l'aide multicritère à la décision. Presses Polytechniques et Universitaires Romandes, Lausanne.

Beaver, W.H., 1966. Financial ratios as predictors of failure. Empirical Research in Accounting: Selected Studies. Journal of Accounting Research, Supplement to vol. 4, pp. 71–111.

Courtis, J.K., 1978. Modelling a financial ratios categoric framework. Journal of Business Finance and Accounting 5 (4), 371–386.

Dimitras, A.I., 1995. Multicriteria methods for business failure prediction (in Greek). Ph.D. Dissertation, Technical University of Crete, Chania, Greece.

Dimitras, A.I., Zanakis, S.H., Zopounidis, C., 1996. A survey of business failures with an emphasis on prediction methods and industrial applications. European Journal of Operational Research 90, 487–513.

Dimitras, A.I., Zopounidis, C., Hurson, Ch., 1995. A multicriteria decision aid method for the assessment of business failure risk. Foundations of Computing and Decision Sciences 20 (2), 99–112.

Eisenbeis, R.A., 1977. Pitfalls in the application of discriminant analysis in business and economics. The Journal of Finance 32, 875–900.

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.

Gloubos, G., Grammatikos, T., 1988. The success of bankruptcy prediction models in Greece. Studies in Banking and Finance 7, 37–46.

Grzymala-Busse, J.W., 1992. LERS – a system for learning from examples based on rough sets. In: Slowinski, R. (Ed.), Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory. Kluwer Academic Publishers, Dordrecht, pp. 3–18.

Gupta, Y.P., Rao, R.P., Bagghi, P.K., 1990. Linear goal programming as an alternative to multivariate discriminant analysis: a note. Journal of Business Finance and Accounting 17 (4), 593–598.

Jones, F.L., 1987. Current techniques in bankruptcy prediction. Journal of Accounting Literature 6, 131–164.

Keasey, K., Watson, R., 1991. Financial distress prediction models: a review of their usefulness. British Journal of Management 2, 89–102.

Krusinska, E., Slowinski, R., Stefanowski, J., 1992. Discriminant versus rough set approach to vague data analysis. Applied Stochastic Models and Data Analysis 8, 43–56.

Lin, T.Y., Wildberger, A. (Eds.), 1995. Soft Computing: Rough Sets, Fuzzy Logic, Neural Networks, Uncertainty Management, Knowledge Discovery. Simulation Councils, San Diego, CA.

Luoma, M., Laitinen, E.K., 1991. Survival analysis as a tool for company failure prediction. Omega 19 (6), 673–678.

Messier, W.F., Hansen, J.V., 1988. Including rules for expert system development: an example using default and bankruptcy data. Management Science 34 (12), 1403–1415.

Mienko, R., Slowinski, R., Stefanowski, J., Susmaga, R., 1996a. Rough family – software implementation of rough set based data analysis and rule discovery techniques. In: Tsumoto, S., Kobayashi, S., Yokomori, T., Tanaka, H., Nakamura, A. (Eds.), Proceedings of the 4th International Workshop on Rough Sets, Fuzzy Sets and Machine Discovery, Tokyo, pp. 437–440.

Mienko, R., Stefanowski, J., Toumi, K., Vanderpooten, D., 1996b. Discovery-oriented induction of decision rules, Cahier du LAMSADE no. 141, Université de Paris Dauphine, Paris, 1996.

Ohlson, J.A., 1980. Financial ratios and the probabilistic prediction of bankruptcy. Journal of Accounting Research (Spring), 109–131.

Pawlak, Z., 1982. Rough sets. International Journal of Information and Computer Sciences 11, 341–356.

Pawlak, Z., 1991. Rough Sets. Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht.

Pawlak, Z., Grzymala-Busse, J.W., Slowinski, R., Ziarko, W., 1995. Rough sets. Communications of the ACM 38 (11), 89–95.

Pawlak, Z., Slowinski, R., 1994. Rough set approach to multi-attribute decision analysis. European Journal of Operational Research 72, 443–459.

Roy, B., 1985. Méthodologie multicritère d'aide à la décision. Economica, Paris.

Scott, J., 1981. The probability of bankruptcy: a comparison of empirical predictions and theoretical models. Journal of Banking and Finance 5, 317–344.

Siskos, Y., Zopounidis, C., Pouliezos, A., 1994. An integrated DSS for financing firms by an industrial development bank in Greece. Decision Support Systems 12, 151–168.

Skowron, A., 1993. Boolean reasoning for decision rules generation. In: Komorowski, J., Ras, Z.W. (Eds.), Methodologies for Intelligent Systems, Lecture Notes in Artificial Intelligence, vol. 689. Springer, Berlin, pp. 295–305.

Slowinski, R. (Ed.), 1992. Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory. Kluwer Academic Publishers, Dordrecht.

Slowinski, R., 1993. Rough set learning of preferential attitude in multi-criteria decision making. In: Komorowski, J., Ras, Z.W. (Eds.), Methodologies for Intelligent Systems. Lecture Notes in Artificial Intelligence, vol. 689. Springer, Berlin, pp. 642–651.

Slowinski, R., 1995. Rough set approach to decision analysis. AI Expert Magazine 10 (3), 18–25.

Slowinski, R., Stefanowski, J., 1994. Rough classification with valued closeness relation. In: Diday, E. et al. (Eds.), New Approaches in Classification and Data Analysis. Springer, Berlin, pp. 482–488.

Slowinski, R., Zopounidis, C., 1995. Application of the rough set approach to evaluation of bankruptcy risk. International Journal of Intelligent Systems in Accounting, Finance and Management 4, 27–41.

Stefanowski, J., 1992. Rough sets theory and discriminant methods as tools for analysis of information systems. A comparative study. Foundations of Computing and Decision Sciences 17 (2), 81–98.

Stefanowski, J., Vanderpooten, D., 1994. A general two-stage approach to inducing rules from examples. In: Ziarko, W. (Ed.), Rough Sets, Fuzzy Sets and Knowledge Discovery. Springer, London, pp. 317–325.

Vermeulen, E.M., Spronk, J., Van der Wijst, N., 1998. The application of the multi-factor model in the analysis of corporate failure. In: Zopounidis, C. (Ed.), Operational Tools in the Management of Financial Risks. Kluwer Academic Publishers, Dordrecht, pp. 59–73.

Vranas, A.S., 1992. 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.

Zavgren, C.V., 1983. The prediction of corporate failure: the state of the art. Journal of Financial Literature 2, 1–37.

Zavgren, C.V., 1985. Assessing the vulnerability to failure of American industrial firms. A logistic analysis. Journal of Business Finance and Accounting 12 (1), 19–45.

Ziarko, W. (Ed.), 1994. Rough Sets, Fuzzy Sets and Knowledge Discovery. Springer, London.

Ziarko, W., Golan, D., Edwards, D., 1993. An application of DATALOGIC/R knowledge discovery tool to identify strong predictive rules in stock market data. In: Proc. AAAI Workshop on Knowledge Discovery in Databases, Washington DC, pp. 89–101.

Zmijewski, M.E., 1984. Methodological issues related to the estimation of financial distress prediction models. Studies on Current Econometric Issues in Accounting Research, pp. 59–82.

Zollinger, M., 1982. L'analyse multicritère et le risque de crédit aux entreprises. Revue Franyçaise de Gestion, pp. 56–66.

Zopounidis, C., 1987. A multicriteria decision making methodology for the evaluation of the risk of failure and an application. Foundations of Control Engineering 12 (1), 45–67.

Zopounidis, C., Pouliezos, A., Yannacopoulos, D., 1992. Designing a DSS for the assessment of company performance and viability. Computer Science in Economics and Management 5, 41–56.

Zopounidis, C., 1995. Evaluation du risque de défaillance de l'enterprise: Méthodes et cas d'application. Economica, Paris.

Zopounidis, C., Dimitras, A.I., Le Rudulier, L., 1995. A multicriteria approach for the analysis and prediction of business failure in Greece. Cahier du LAMSADE, no. 132, Université de Paris Dauphine.

Zopounidis, C., Matsatsinis, N.F., Doumpos, M., 1996a. Developing a multicriteria knowledge-based decision support system for the assessment of corporate performance and viability: The FINEVA system. Fuzzy Economic Review 1 (2), 35–53.

Zopounidis, C., Doumpos, M., Matsatsinis, N.F., 1996b. Application of the FINEVA multicriteria knowledge-based decision support systems to the assessment of corporate failure risk. Foundations of Computing and Decision Sciences 21 (4), 233–251.