PERGAMON

Omega 29 (2001) 561–576

omega
The International Journal of Management Science

www.elsevier.com/locate/dsw

Variable precision rough set theory and data discretisation: an application to corporate failure prediction

Malcolm J. Beynon *, Michael J. Peel

Cardiff Business School, Aberconway Building, Colum Drive, Cardiff, CF10 3EU, UK

Received 18 February 2000; accepted 6 June 2001

Abstract

Since the seminal work of Pawlak (International Journal of Information and Computer Science, 11 (1982) 341–356) rough set theory (RST) has evolved into a rule-based decision-making technique. To date, however, relatively little empirical research has been conducted on the efficacy of the rough set approach in the context of business and finance applications. This paper extends previous research by employing a development of RST, namely the variable precision rough sets (VPRS) model, in an experiment to predict between failed and non-failed UK companies. It also utilizes the FUSINTER discretisation method which negates the influence of an 'expert' opinion. The results of the VPRS analysis are compared to those generated by the classical logit and multivariate discriminant analysis, together with more closely related non-parametric decision tree methods. It is concluded that VPRS is a promising addition to existing methods in that it is a practical tool, which generates explicit probabilistic rules from a given information system, with the rules offering the decision maker informative insights into classification problems. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Data mining; Failure prediction; FUSINTER data discretisation; Rough set theory; Variable precision rough set theory

1. Introduction

Since the nascence of computerisation, together with the evolution of Artificial Intelligence (AI), there has been an explosion in the application of advanced decision-making techniques to solving business problems [1–5]. Following the pioneering study of Altman [6], who used multivariate discriminant analysis (MDA) to differentiate between failed and non-failed US firms, a large body of research has focused on corporate failure prediction (see [7–10] for literature reviews). The prediction of corporate failure continues to be viewed as a matter of considerable interest to both academics and practitioners (including credit and investment analysts), and has obvious importance for the stakeholders (investors, creditors, employees, managers) of a firm. This is evidenced by the recent application of neural networks (NNs), recursive partitioning algorithm (RPA) and case based reasoning to this issue [11–15]. A key advantage of these contemporary methods over their traditional counterparts (such as MDA and logit analysis) is that they do not require pre-specification of a functional form, nor the adoption of restrictive assumptions concerning the distributions of model variables and errors [12,16,10].

More recently, a further non-parametric technique, rough set theory (RST), which has its foundations in mathematical set theory, has been applied to decision problems [17,18]. RST was originated by Pawlak [19] and has been described as 'a new mathematical tool to deal with vagueness and uncertainty. This approach seems to be of fundamental importance to AI and cognitive sciences, especially in the areas of machine learning, knowledge acquisition, decision analysis, knowledge discovery from databases, expert systems, decision support systems, inductive reasoning and pattern recognition ... One of the advantages of RST is that it does

* Corresponding author. Tel./fax: +44-29-2087-5747.
E-mail address: beynonmj@cardiff.ac.uk (M.J. Beynon).

0305-0483/01/$ - see front matter
PII: S0305-0483(01)00045-7

not need preliminary or additional information about data, such as probability distributions in statistics, basic probability assignment in the Dempster Shafer theory of evidence, or grade of membership of the value of possibility in fuzzy set theory' [20, p. 89].

RST incorporates the use of indiscernibility (equivalence) relations to approximate sets of objects by upper and lower set approximations and, as noted by Slowinski and Zopounidis [21, p. 79], 'it is a formal framework for discovering deterministic and non-deterministic rules from a given representation of knowledge ... [it] ... assumes knowledge representation in a decision table form which is a special case of an information system'. Initial RST applications focused on medical diagnosis, drug research and process control [22,23], but more recently it has been extended to cover credit fraud detection, stock market rule-generation, market research, climate change and the development of expert systems for the NASA space centre [24,25,20].

Slowinski and Zopounidis [21] also investigated the use of RST to assess the risk of a Greek bank's clients (firms) in terms of granting finance. Although they did not examine the predictive accuracy of the RST rules, they did conclude [21, p. 39] that (based on financial ratios and other firm-specific variables), RST 'is a useful tool for discovering a preferential attitude of the decision maker in multi-attribute sorting problems'. More recently, Dimitras et al. [26, p. 278] reported that (on the basis of financial ratios) a rough set approach to predicting between failed and non-failed Greek firms 'was generally better than those obtained by classical discriminant and logit models'. A limitation of these studies is that the continuous data used to derive the rough set rules, have been discretised (a requirement of RST) with the aid of a selected 'expert'. Clearly different experts may proffer different views and the operational costs and complexities of using RST (and related techniques) will increase when there is over-reliance on an expert. In this context An et al. [27, p. 647] have stated that 'It has to be emphasised ... that the question of how to optimally discretise the attribute (variable) values, is unsolved, and a subject of on-going research'. This paper therefore employs a new (and more objective) discretisation method, namely the FUSINTER technique. However, the motivation for data discretisation extends beyond the requirements of RST, to include discretising data of an imprecise quality ('noisy' data). The ability to formulate rules from interval data (via discretisation) may also facilitate a more informed understanding of the interaction of the characteristics of objects. In this context, it is of interest to note that, even with regard to traditional statistical estimators (logit/discriminant analysis), it has recently been advocated that continuous variables (financial ratios) should be rank-transformed to improve their distributional properties in a failure prediction setting [28].

A further RST innovation has been the development by Ziarko [29] of a variable precision rough sets (VPRS) model, which incorporates probabilistic decision rules. This is an important extension, since as noted by Kattan and Cooper [30, p. 468], when discussing computer based decision techniques in a corporate failure setting, 'In real world decision making, the patterns of classes often overlap, suggesting that predictor information may be incomplete... This lack of information results in probabilistic decision making, where perfect prediction accuracy is not expected'.

An et al. [27] applied VPRS (which they termed 'Enhanced RST') to generating probabilistic rules to predict the demand for water. Relative to the traditional rough set approach, VPRS has the additional desirable property of allowing for partial classification compared to the complete classification required by RST. More specifically, when an object is classified using RST it is assumed that there is complete certainty that it is a correct classification. In contrast, VPRS facilitates a degree of confidence in classification, invoking a more informed analysis of the data, which is achieved through the use of a *majority inclusion* relation [29].

This paper extends previous work by providing an empirical exposition of VPRS, where we present the results of an experiment which applies VPRS rules to the corporate failure decision. In addition, we mitigate the impact of using the subjective views of an expert (as employed in previous studies) to discretise the data, by utilising the sophisticated FUSINTER discretisation technique which is applied to a selection of attributes (variables) relating to companies' financial and non-financial characteristics. The discretised data, in conjunction with other nominal attributes, are then used in this new VPRS framework to identify rules to classify companies in a failure setting.

To facilitate a comparison of our experimental VPRS results with those of existing techniques, we present the predictive ability of classical statistical methods—logit analysis and MDA—together with two more closely related non-parametric decision-tree methods, RPA and the Elysee method, which utilises ordinal discriminant analysis (see [15,31], for an exposition of these methods). However in the spirit of previous experimental research—and more particularly the previous failure prediction study of Frydman et al. [15, p. 239], who concluded that 'we feel that the attributes of new techniques like RPA can be presented and evaluated in a rigorous framework without the necessity of proving its absolute superiority over existing procedures'—the comparative classification results are not meant to be definitive, but rather to illustrate the potential of VPRS. In this context, research on the criteria to select the most efficacious and parsimonious set of VPRS rules (for predictive purposes) is still in its infancy [27].

The remainder of the paper is organised as follows: The next section gives a brief exposition of the VPRS method and a discussion of the FUSINTER discretisation method. The results of the empirical experiments are then reported, including a discussion of the predictive ability of VPRS relative to other existing parametric and non-parametric methods.

Table 1
Example of a decision table

	Condition attributes (C)						Decision attribute (D)
Objects	c_1	c_2	c_3	c_4	c_5	c_6	d
o_1	1	0	1	1	0	1	L
o_2	1	0	0	0	0	0	L
o_3	0	0	1	0	0	0	L
o_4	1	0	1	1	0	1	H
o_5	0	0	0	0	1	1	H
o_6	1	0	1	1	0	1	H
o_7	0	0	0	0	1	0	H

2. An overview of VPRS

VPRS (as with RST) operates on what may be described as a decision table or *information system*. As is illustrated in Table 1, a set of *objects* $\mathbf{U}$ $(o_1, \ldots, o_7)$ are contained in the rows of the table. The columns denote *condition attributes* $\mathbf{C}$ $(c_1, \ldots, c_6)$ of these objects and a related *decision attribute* $\mathbf{D}$ (d). A value denoting the nature of an attribute to an object is called a *descriptor*. As noted above, a VPRS data requirement is that it must be in discrete or categorical form. Table 1 shows that, with this particular example, the condition attribute descriptors comprise 0's and 1's (for example, denoting yes and no answers), and the decision attribute values are L and H (for example, denoting low and high). The table shows that the objects have been classified into one of these decision values, which are also referred to as *concepts*.

For the condition attributes in this example, all of the objects ($\mathbf{U}$) can be placed in five groups: $\mathbf{X}_1 = \{o_1, o_4, o_6\}$, $\mathbf{X}_2 = \{o_2\}$, $\mathbf{X}_3 = \{o_3\}$, $\mathbf{X}_4 = \{o_5\}$ and $\mathbf{X}_5 = \{o_7\}$. The objects within a group are indiscernible to each other so that, objects o_1, o_4 and o_6 in $\mathbf{X}_1$ have the same descriptor values for each of the condition attributes. These groups of objects are referred to as *equivalence classes* or *conditional classes*, for the specific attributes. The equivalence classes for the decision attribute are: $\mathbf{Y}_\mathrm{L} = \{o_1, o_2, o_3\}$ and $\mathbf{Y}_\mathrm{H} = \{o_4, o_5, o_6, o_7\}$. The abbreviation of the set of equivalence classes for the conditional attributes $\mathbf{C}$, is denoted by $\mathbf{E}(\mathbf{C}) = \{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4, \mathbf{X}_5\}$ and for the decision attribute, it is defined $\mathbf{E}(\mathbf{D}) = \{\mathbf{Y}_\mathrm{L}, \mathbf{Y}_\mathrm{H}\}$.

VPRS measurement is based on ratios of elements contained in various sets. A case in point is the conditional probability of a concept given a particular set of objects (a condition class). For example:

$$\Pr(\mathbf{Y}_\mathrm{L}|\mathbf{X}_1) = \Pr(\{o_1, o_2, o_3\}|\{o_1, o_4, o_6\})$$
$$= \frac{|\{o_1, o_2, o_3\} \cap \{o_1, o_4, o_6\}|}{|\{o_1, o_4, o_6\}|} = 0.333.$$

It follows that this measures the accuracy of the allocation of the conditional class $\mathbf{X}_1$ to the decision class $\mathbf{Y}_\mathrm{L}$ [29]. Hence for a given probability value β, the β-positive region corresponding to a concept is delineated as the set of objects with conditional probabilities of allocation at least equal to β. More formally:

β-positive region of the set $\mathbf{Z} \subseteq \mathbf{U}$: $POS_\mathbf{P}^\beta(\mathbf{Z})$
$$= \bigcup_{\Pr(\mathbf{Z}|\mathbf{X}_\mathrm{i}) \geqslant \beta} \{\mathbf{X}_\mathrm{i} \in \mathbf{E}(\mathbf{P})\} \text{ with } \mathbf{P} \subseteq \mathbf{C}.$$

Following An et al. [27], β is defined to lie between 0.5 and one. Hence for the current example, the condition equivalence class $\mathbf{X}_1 = \{o_1, o_4, o_6\}$ have a majority inclusion (with at least 60% majority needed, i.e. $\beta = 0.6$) in $\mathbf{Y}_\mathrm{H}$, in that most objects (2 out of 3) in $\mathbf{X}_1$ belong in $\mathbf{Y}_\mathrm{H}$. Hence $\mathbf{X}_1$ is in $POS_\mathbf{C}^{0.6}(\mathbf{Y}_\mathrm{H})$. It follows $POS_\mathbf{C}^{0.6}(\mathbf{Y}_\mathrm{H}) = \{o_1, o_4, o_5, o_6, o_7\}$.

Corresponding expressions for the β-boundary and β-negative regions are given by Ziarko [29], as follows:

β-boundary region of the set $\mathbf{Z} \subseteq \mathbf{U}$: $BND_\mathbf{P}^\beta(\mathbf{Z})$
$$= \bigcup_{1-\beta < \Pr(\mathbf{Z}|\mathbf{X}_\mathrm{i}) < \beta} \{\mathbf{X}_\mathrm{i} \in \mathbf{E}(\mathbf{P})\} \text{ with } \mathbf{P} \subseteq \mathbf{C},$$

β-negative region of the set $\mathbf{Z} \subseteq \mathbf{U}$: $NEG_\mathbf{P}^\beta(\mathbf{Z})$
$$= \bigcup_{\Pr(\mathbf{Z}|\mathbf{X}_\mathrm{i}) \leqslant 1-\beta} \{\mathbf{X}_\mathrm{i} \in \mathbf{E}(\mathbf{P})\} \text{ with } \mathbf{P} \subseteq \mathbf{C}.$$

Using $\mathbf{P}$ and $\mathbf{Z}$ from the previous example, with $\beta =$ 0.6, then $BND_\mathbf{C}^{0.6}(\mathbf{Y}_\mathrm{H}) = \emptyset$ (empty set) and $NEG_\mathbf{C}^{0.6}(\mathbf{Y}_\mathrm{H}) = \{o_2, o_3\}$. Similarly for the decision class $\mathbf{Y}_\mathrm{L}$ it follows $POS_\mathbf{C}^{0.6}(\mathbf{Y}_\mathrm{L}) = \{o_2, o_3\}$, $BND_\mathbf{C}^{0.6}(\mathbf{Y}_\mathrm{L}) = \emptyset$ and $NEG_\mathbf{C}^{0.6}(\mathbf{Y}_\mathrm{L}) = \{o_1, o_4, o_5, o_6, o_7\}$.

VPRS applies these concepts by firstly seeking subsets of the attributes, which are capable (via construction of decision rules) of explaining allocations given by the whole set of condition attributes. These subsets of attributes are termed *β-reducts* or *approximate reducts*. Ziarko [29] states that a β-reduct, a subset $\mathbf{P}$ of the set of conditional attributes $\mathbf{C}$ with respect to a set of decision attributes $\mathbf{D}$, must satisfy the following conditions: (i) that the subset $\mathbf{P}$ offers the same quality of classification (subject to the same β value) as the whole set of condition attributes $\mathbf{C}$; and (ii) that no attribute can be eliminated from the subset $\mathbf{P}$ without affecting the quality of the classification (subject to the same β value).

The quality of the classification is defined as the proportion of the objects made up of the union of the β-positive regions of all the decision equivalence classes based on the condition equivalence classes for a subset $\mathbf{P}$ of the condition attributes $\mathbf{C}$. Associated with each conditional class is an upper bound on the β value above which there is no opportunity for majority inclusion and hence not in a β-positive region—in the previous example $\Pr(\mathbf{Y}_\mathrm{L}|\mathbf{X}_1) = 0.333$ and $\Pr(\mathbf{Y}_\mathrm{H}|\mathbf{X}_1) = 0.667$, hence if $\beta = 0.7$ then $\mathbf{X}_1$ is not in the associated β-positive region, since the upper bound on β (for majority inclusion) is 0.667. The lowest of these upper bounds (amongst the condition classes) on the β values is defined $\beta_{\min}$ and relates to the overall level of confidence in

Table 2
Examples of β-reducts ($\beta_{\min}$, β-reduct Quality of classification) for data shown in Table 1

0.571,	0.667,	0.667,	1.00,
$\{c_2\}, 1$	$\{c_1, c_3\}, 1$	$\{c_1, c_4\}, 1$	$\{c_4, c_5\}, 0.571$

Table 3
Decision rules for the β-reduct $\{c_1, c_4\}$

Rule		c_1		c_4		d	S	P
1	If	1	and	0	Then	L	1	1
2	If	0			Then	H	3	0.667
3	If			1	Then	H	3	0.667

classification by the particular β-reduct. The quality of classification associated with all the condition attributes, is dependent on the $\beta_{\min}$ value. Hence for $\beta_{\min} \in (0.500, 0.667]$ [1] the quality of classification is one (with $\mathbf{X}_1$ in $POS_{\mathbf{C}}^{\beta}(\mathbf{Y}_{\mathrm{H}})$), for $\beta_{\min} \in (0.667, 1]$ the quality of classification is 0.571 (i.e., classifying four of the seven objects, with $\mathbf{X}_1$ in $BND_{\mathbf{C}}^{\beta}(\mathbf{Y}_{\mathrm{H}})$). Table 2 provide examples of four such β-reducts based on the condition attributes in Table 1.

For the β-reducts given in Table 2, it can be seen that the $\beta_{\min}$ values range from 0.571 to one. As an illustration, for the β-reduct $\{c_1, c_4\}$ in Table 2, $\beta_{\min} = 0.667$ implying that the objects are classified with at least 66.7% confidence with a quality of classification of unity—that is, all objects are given a classification. Using the information in Table 1, this is further illuminated by an examination of the relevant equivalence classes, that is, $\mathbf{E}(\mathbf{D}) = \{\mathbf{Y}_{\mathrm{L}}, \mathbf{Y}_{\mathrm{H}}\} = \{\{o_1, o_2, o_3\}, \{o_4, o_5, o_6, o_7\}\}$ and $\mathbf{E}(\{c_1, c_4\}) = \{\{o_1, o_4, o_6\}, \{o_2\}, \{o_3, o_5, o_7\}\}$. Since the quality of classification is unity in this case, a valid value of β is obtained when the β-boundary regions for each decision equivalence classes ($\mathbf{Y}_{\mathrm{L}}$ and $\mathbf{Y}_{\mathrm{H}}$) are empty [27]. An inspection of the sets in $\mathbf{E}(\{c_1, c_4\})$ shows that both $\{o_1, o_4, o_6\}$ and $\{o_3, o_5, o_7\}$ have the worst majority inclusion of 66.7%, that is, $\beta_{\min} = 0.66$ (as given in Table 2). As $\{c_1, c_4\}$ is a β-reduct, the quality of its classification is equal to that obtained from the set of condition attributes, that is unity, when $\beta \in (0.500, 0.667]$. With an adequate β-reduct identified (see below), the rule generation aspect of VPRS can commence. Following An et al. [27], the aim is to construct a rule for each (or group of) condition class for a given β-reduct, by establishing for each rule, whether there are condition attributes whose values are irrelevant in determining the value in the concept.

As an example, Table 3 provides the minimal rules associated with the β-reduct $\{c_1, c_4\}$ shown in Table 2. Rule 1 in Table 3 can be used to illustrate the rule generation process. Each equivalence class associated with $\{c_1, c_4\}$ must be examined with reference to the two decision equivalence classes $\mathbf{Y}_{\mathrm{L}}$ and $\mathbf{Y}_{\mathrm{H}}$. Only rule 1 classifies objects into the decision class $\mathbf{Y}_{\mathrm{L}}$, in that $\{o_2\} \in \mathbf{E}(\{c_1, c_4\})$ is definitely classified to $\mathbf{Y}_{\mathrm{L}}$, so $\{o_2\}$ needs to be differentiated from the other members of $\mathbf{E}(\{c_1, c_4\})$. By comparing $\{o_2\}$ to $\{o_1, o_4, o_6\} \in \mathbf{E}(\{c_1, c_4\})$, it can be seen (from Table 1) that the two sets of objects have the same c_1 value (i.e. 1). Hence it must be established that $c_4 = 0$ to differentiate $\{o_2\}$ from $\{o_1, o_4, o_6\}$. Similarly, $\{o_2\}$ is only different from $\{o_3, o_5, o_7\}$, with respect to the c_1 attribute. In consequence, it must be established that $c_1 = 1$ to differentiate $\{o_2\}$ from $\{o_3, o_5, o_7\}$. From this analysis a set of criterion ($c_1 = 1$ and $c_4 = 0$) is obtained, and their conjunction infers $\{o_2\}$ is differentiated from the other members of $\mathbf{E}(\{c_1, c_4\})$ and is classified to $\mathbf{Y}_{\mathrm{L}}$. The S column relates to the strength of each rule—that is how many objects are classified by it. The P column relates to the proportion of those objects correctly classified by the rule. Two alternative formats of rule 1 can be presented:

$$(c_1 = 1) \wedge (c_4 = 0) \overset{1/7 \;\; 100\%}{\longrightarrow} \mathrm{L}:$$

or equivalently

Classifying 1 *object*: *If* $c_1 = 1$ *and* $c_4 = 0$, *then* L *with* 100% *confidence in the classification.*

As discussed above, the VPRS analysis (and RST) requires that the data are presented in categorical form. Hence continuous data (e.g., financial ratios in our case) must first be subjected to a process of discretisation. To obviate the requirement of an expert (as used in previous research) within the discretisation process, this paper utilises a continuous variable discretisation (CVD) approach to data discretisation. [2] Both *local* and *global* methods of supervised CVD have been developed. Localised CVD considers only sub-regions of the decision table, and commonly involves the discretisation of one condition (continuous) attribute at a time. Such techniques include the minimum entropy method [32], together with Chi-Merge [33] and Chi2 [34] which both employ variants of the Chi-square statistic as a measure of goodness of fit. In contrast, global CVD considers groups of condition attributes in the discretisation process. For example, Chmielewski and Grzymala-Busse [35], formulate a global clustering method, adopting a stopping criteria based on RST.

In this paper, a supervised CVD technique is utilised. Since not all of the condition attributes which will be used are continuous (see Table 4), a local discretisation method is employed, namely the FUSINTER method [36], which

[1] That is, for the range 0.5 (not including 0.5) up to and including 0.667.

[2] As with neural network training, CVD may be based on supervised (utilising the decision classification) and unsupervised (considering only the group of continuous variables in question) learning. Unsupervised CVD can be viewed as a simple clustering mechanism—for example, with data discretised on the basis of equal-frequency-intervals or equal-sized-width-intervals.

Table 4
Definition of variables

Attribute	Description
Size	
c_1	Sales (£000)
Profit	
c_2	ROCE: profit before tax/capital employed (%)
c_3	FFTL: funds flow (earnings before interest, tax & depreciation)/total liabilities
Gearing	
c_4	GEAR: (current liabilities + long-term debt)/total assets
c_5	CLTA: current liabilities/total assets
Liquidity	
c_6	CACL: current assets/current liabilities
c_7	QACL: (current assets – stock)/current liabilities
c_8	WCTA: (current assets – current liabilities)/total assets
Other	
c_9	LAG: number of days between account year end and the date the annual report and accounts were filed at company registry.
c_{10}	AGE: number of years company has been operating since incorporation date.
c_{11}	CHAUD: coded 1 if changed auditor in previous three years, 0 otherwise
c_{12}	BIG6: coded 1 if company auditor is a Big6 auditor, 0 otherwise

is a *bottom up* algorithm (merging sub-intervals) whose objective is to partition condition attributes into ordinal groups subject to the optimisation of a specific measure (i.e., quadratic entropy). As discussed above a key advantage of the FUSINTER technique is that it provides an objective method (without recourse to an expert) to discretise the data into the requisite number of intervals. Zighed et al. [36] have noted two further advantages of this method; firstly, that it avoids the problem of thin partitioning (i.e. the presence of intervals with very few objects in them) and secondly, it explicitly takes into account finite sample sizes. [3]

3. Data and variables

All data were obtained from the financial analysis made easy (FAME, [37]) CD-ROM UK corporate database. FAME contains financial and non-financial information (largely derived from corporate annual reports and accounts) for in excess of 200,000 British companies. The database covers all medium-sized and large firms, together with a large sample of smaller firms. The FAME computer software was utilised to generate the population of all actively trading (i.e. non-failed) private companies, together with those private companies which had appointed a receiver (i.e. had failed), whose primary operating activities were in the manufacturing sector (1992, SIC codes 15–36), whose annual account year ends fell in the period 1-4-97 to 31-3-98 (the latest annual period on the database), whose latest turnover was greater than £2.8m, and whose latest total assets was greater than, or equal to, £1.4m.

The rationale for the selection criteria are as follows: (i) previous studies have reported that financial variables (particularly ratios) vary significantly according to industrial classification [38,10]. In this context, Taffler [8] has argued that separate failure prediction models should be developed for manufacturing and non-manufacturing firms. [4] To control for potential confounding structural differences, we therefore focus on manufacturing firms; (ii) a common fiscal year (the latest available) was selected to control for any confounding intertemporal macro influences [10] which might otherwise bias the results; and (iii) under UK Company Acts, smaller firms are defined as those with total assets not exceeding £1.4m and with sales less than £2.8m [37, p. 18]. Such companies are permitted to file modified (truncated) accounts (i.e. often do not disclose full financial data). In addition, it is unclear how smaller firms are selected for inclusion on the FAME database. In consequence, small firms are excluded from the current study.

In total, 45 failed and 10,274 non-failed manufacturing firms met the sampling criteria. Data for the 45 failed firms was collected from the FAME database, together with data for 45 randomly selected non-failed ones. [5] To facilitate out-of-sample (holdout) testing, 30 failed and 30 non-failed firms were randomly selected to form the within-sample rule-construction (training) sample, with the remaining 15 failed and 15 non-failed firms being utilised as a holdout sample. The sample size is smaller than some previous studies (see e.g. [9,10] for reviews) but larger than a recent study by Laitinen and Kankaanpaa [41] whose specific aim was to compare the failure prediction accuracy of various

[3] For a full description of the FUSINTER algorithm see [36]. In this paper the quadratic entropy measure is used together with the default values $\alpha = 0.975$ and $\lambda = 1$.

[4] However, as noted by a referee to this paper, sample selection on this basis clearly limits the generalisation of empirical findings to the industrial sector studied.

[5] Given the computational effort required, the collection of data for 90 companies was thought sufficient to illustrate the properties of VPRS and FUSINTER. In this context, both the VPRS and FUSINTER methods were self-programmed by one of the authors (using the MAPLE software package [39]). However it should be noted that researchers/practitioners may acquire the appropriate computer software to operationalise the techniques (in the case of VPRS, see [40]; and for FUSINTER, see [36]).

statistical techniques. [6] In the context of the current expository study, the sample size was selected with reference to previous RST studies of [21,26]—which employed training samples containing 39 and 80 companies, respectively—to enable a reasonably sized set of rules to be extracted for evaluation. In any event, the VPRS rules are not 'biased' by a relatively small sample (in that the rules are derived from the self-contained information system), but in common with other techniques, larger samples permit more confidence in the inferences drawn from, and the generalisations of, the reported results.

In the analysis which follows, the decision attribute is company status, which is coded zero for a non-failed firm and unity for a failed company. [7] A total of 12 variables (condition attributes) were collected for potential rule generation. The variables were collected with reference to the extant literature [42,12,9,10], with the aim of covering the key attributes which have been found to be associated with corporate failure. Table 4 contains the variable definitions and the corporate attributes they refer to.

The first variable (c_1) is the annual sales figure, which is a key size attribute [10]. Smaller firms have been reported to be characterised by a significantly higher probability of failing, and company size has been found to be negatively associated with financial distress levels [38,10]. The next two variables (c_2, c_3) in Table 4 relate to profit attributes. Not surprisingly, failing firms have consistently been found to be less profitable than non-failing ones. Two standard ratios, which have been employed in previous research [9,10] are used in the current study: the ratio of profit before tax to capital employed (ROCE), and the ratio funds flow to total liabilities (FFTL). The level of financial gearing has also been found to be a key attribute associated with corporate failure (ibid), with higher gearing being positively associated with corporate financial distress [10,26]. Two standard ratios (c_4, c_5) are employed in the current study: the ratio of current liabilities plus long-term debt to total assets (GEAR), and the ratio of current liabilities to total assets (CLTA).

Corporate liquidity (solvency) is another key attribute associated with corporate collapse, with less liquid firms exhibiting a higher probability of collapse [8,43]. The three liquidity variables (c_6–c_8) shown in Table 4 have been found to be significant predictors in previous studies [9,10]. These are the ratio of current assets to current liabilities (CACL, known as the 'current' ratio), the ratio of current assets minus stock to current liabilities (QACL, known as the 'quick' ratio), and the ratio of working capital to total assets (WCTA).

The remaining four variables (c_9–c_{12}) described in Table 4 relate to various non-financial corporate attributes which have been found to be associated with corporate failure. The first (LAG) is the time lag (in days) between the date of the account year end of each company and the date the annual report and accounts were filed (i.e. became public) at the Company Registry (a requirement under UK law). Previous studies have reported that companies with longer reporting lags are more likely to fail [43,10]. This is thought to occur because companies in financial distress may attempt to delay publication of their accounts (in an attempt to remedy the situation before the accounts are published), and/or because the auditing process is more time consuming for financially distressed firms [44,45]. The second variable (AGE) is the time (in years) a company had been operating since its incorporation date. Corporate age has been found to be negatively associated with corporate failure, with younger (start-up) firms being particularly risk-prone [46,10].

The final two variables (c_{11}, c_{12}) are binary in nature and relate to auditor characteristics. The first is whether or not a company had changed its auditor over the preceding three years (CHAUD). Previous studies have reported that failing firms are more likely to be characterised by auditor switches (up to three years before failure), largely in consequence of disputes between auditors and managers over accounting methods, together with disagreements in respect of audit opinions/qualifications [47,48,10]. The second (BIG6) is whether or not the company was audited by a Big6 accounting firm. [8] Recent evidence suggests that, because of the increased litigation threat attached to auditing risky clients (including those which are failing), large (Big6) auditors are less likely to audit clients in financial distress [49,50,38]. Hence, on the basis of current theory and evidence, CHAUD and BIG6 are expected to be positively and negatively associated with corporate failure respectively.

Table 5 reports summary statistics for the variables (attributes) utilised in the rule-generating samples. In terms of the mean and median values, all variables exhibit the anticipated differences in magnitude as between the failed and non-failed firms, with most of the differences being statistically significant at conventional levels on the basis of standard parametric and non-parametric univariate tests [51]. In summary, the statistics in Table 5 suggest that smaller (c_1), less profitable (c_2, c_3) firms, characterised by higher gearing (c_4, c_5) and lower liquidity (c_6, c_7, c_8) levels, which exhibit longer reporting lags (c_9), which are younger (c_{10}), which have experienced an auditor switch (c_{11}), and which are audited by a non-Big6 auditor (c_{12}), are more likely to fail.

[6] The recent Laitinen and Kankaanpaa study [41] used a total sample of only 76 Finnish companies (38 failed and 38 non-failed), with the holdout results based on the Lachenbruck jacknife method and by splitting the estimation sample in half. They concluded [41, p. 84], based on evaluation of various methods (including RPA, NNs, Logit and MDA), that 'no superior method has been found'.

[7] In all cases, the data for the failed firms were published prior to the first announcement of failure (i.e. that a receiver had been appointed).

[8] The Big6 accounting firms comprise Arthur Anderson, Coopers and Lybrand, Ernst and Young, KPMG, Price Waterhouse and Touche Ross. Following the merger between Price Waterhouse and Coopers and Lybrand in July 1998, the Big6 became the Big5.

Table 5
Summary statistics for rule-generating sample

Attribute	Units	Failed Mean	Failed Median	Non-failed Mean	Non-failed Median	All Firms Mean	All Firms Median
c_1 (Sales)	000s	11347.1*	5986.0	24355.6*	7975.0	17851.3	7299.5
c_2 (ROCE)	ratio (%)	−2.676**	−1.991†††	4.462**	5.409†††	0.8930	0.7885
c_3 (FFTL)	ratio	0.0698***	0.0651†††	0.1986***	0.1785†††	0.1341	0.1021
c_4 (GEAR)	ratio	0.8364	0.8053†††	0.7307	0.6406†††	0.7836	0.7428
c_5 (CLTA)	ratio	0.7034***	0.7019†††	0.5275***	0.4921†††	0.6154	0.6216
c_6 (CACL)	ratio	1.012***	0.9319†††	1.434***	1.325†††	1.223	1.153
c_7 (QACL)	ratio	0.6725***	0.5594†††	1.044***	1.033†††	0.8584	0.7319
c_8 (WCTA)	ratio	−0.0397***	−0.0412†††	0.1203***	0.1409†††	0.0309	0.0758
c_9 (LAG)	days	269.1***	295.0††	213.6***	204.5††	241.4	266.5
c_{10} (AGE)	years	20.8*	14.0†††	31.7*	26.5†††	26.3	23.0
c_{11} (CHAUD)[a]	binary	30.0%	N/A	13.3%	N/A	21.7%	N/A
c_{12} (BIG6)[a]	binary	36.7%■■	N/A	66.7%■■	N/A	51.7%	N/A

[a] Indicates a binary variable where the mean denotes the % of the sample with a value of unity.

Note: ***, **, * Indicates means are significantly different on basis of t-test at the 1%, 5% and 10% significant levels, respectively (two-tailed tests). †††, †† Indicates medians are significantly different on the basis of a chi-square medians test at the 1% and 5% significance levels, respectively. ■■ Indicates chi-square test is significant at the 5% level.

Table 6
Condition attributes intervals for FUSINTER discretisation[a]

Attribute	Interval '0'	Interval '1'	Interval '2'	Interval '3'
c_1 (SALES)	[2857.00, 5694.00)	[5694.00, 32683.50]	[32683.50, 167370.00]	
c_2 (ROCE)	[−37.34970, −9.31125)	[−9.31125, −6.30660)	[−6.30660, 1.71560)	[1.71560, 33.84510]
c_3 (FFTL)	[−0.32830, 0.02900)	[0.02900, 0.12095)	[0.12095, 0.21375)	[0.213750, 0.63120]
c_4 (GEAR)	[0.12120, 0.57495)	[0.57495, 0.79300)	[0.79300, 1.09850)	[1.09850, 3.53360]
c_5 (CLTA)	[0.12120, 0.46500)	[0.46500, 0.70260)	[0.70260, 1.48650]	
c_6 (CACL)	[0.49740, 1.16945)	[1.16945, 1.37075)	[1.37075, 4.44650]	
c_7 (QACL)	[0.28470, 0.96165)	[0.96165, 3.36350]		
c_8 (WCTA)	[−0.74700, 0.05720)	[0.05720, 0.11970)	[0.11970, 0.44820]	
c_9 (LAG)	[55.00, 288.00)	[288.00, 421.00]		
c_{10} (AGE)	[2.0, 24.5)	[24.5, 90.0]		

[a] Figures in square brackets indicate closed intervals, including boundary values. Figures in one square and one curved bracket indicate one closed and one open end of the interval excluding the boundary value.

4. Discretisation of the data

The FUSINTER discretisation method described above was used to categorise those condition attributes which are continuous in nature (c_1–c_{10} in Table 4). Appendix A to the paper lists the original values of the condition attributes for all companies in the training set and Appendix B shows the associated discretised values. Table 6 reports the interval ranges constructed for each of the relevant attributes. It shows that the discretisation process has produced between two and four intervals for each of the condition attributes. Since no attribute is discretised to a single interval, all the variables are influential in terms of allocation–determination for particular concepts. For example, the first attribute, c_1 (SALES measured in £000s), has three intervals: interval '0' from 2857 to 5694, interval '1' between 5694 and 32683.5 and interval '2' between 32683.5 and 167370.

The analysis revealed 58 different condition classes (i.e. combinations of discretised and nominal condition attribute values), with only two groups of objects (namely: $\{o_3, o_{27}\}$ and $\{o_{37}, o_{41}\}$) exhibiting the same combination of discretised and nominal condition attribute values (Appendix B). However, the objects in each of the two groups have the same decision value (e.g., objects o_3 and o_{27} both have the decision attribute value 0). This indicates that there is no contradictory information in the system. The associated quality of classification is therefore 1, for any $\beta_{\min}$ value in the range $(0.5, 1]$.

5. VPRS analysis

As discussed above, in performing VPRS analysis, the β-reducts associated with the (post discretisation)

Table 7
Rules from β-reduct $\{c_4, c_7, c_8, c_9\}$

Rule 1: $(0.12120 \leqslant c_4 < 0.57495) \overset{14/16\ 92.8\%}{\longrightarrow} 0$:
Rule 2: $(1.09850 \leqslant c_4 < 3.53360) \wedge (55 \leqslant c_9 < 288) \overset{3/60\ 100\%}{\longrightarrow} 0$:
Rule 3: $(0.7930 \leqslant c_4 \leqslant 1.0985) \wedge (0.96165 \leqslant c_7 < 3.3635) \overset{2/60\ 100\%}{\longrightarrow} 0$:
Rule 4: $(0.0572 \leqslant c_8 \leqslant 0.1197) \wedge (288 \leqslant c_9 \leqslant 421) \overset{2/60\ 100\%}{\longrightarrow} 0$:
Rule 5: $(0.57495 \leqslant c_4 \leqslant 0.7930) \wedge (0.2847 \leqslant c_7 < 0.96165) \wedge (0.1197 \leqslant c_8 < 0.4482) \overset{3/60\ 100\%}{\longrightarrow} 0$:
Rule 6: $(0.96165 \leqslant c_7 < 3.3635) \wedge (0.0572 \leqslant c_8 \leqslant 0.1197) \wedge (55 \leqslant c_9 < 288) \overset{1/60\ 100\%}{\longrightarrow} 0$:
Rule 7: $(0.57495 \leqslant c_4 \leqslant 0.793) \wedge (-0.747 \leqslant c_8 < 0.0572) \wedge (55 \leqslant c_9 < 288) \overset{7/60\ 57.14\%}{\longrightarrow} 0$:
Rule 8: $(0.57495 \leqslant c_4 \leqslant 0.793) \wedge (0.96165 \leqslant c_7 < 3.3635) \wedge (0.1197 \leqslant c_8 < 0.4482) \wedge (55 \leqslant c_9 < 288) \overset{1/60\ 100\%}{\longrightarrow} 0$:
Rule 9: $(0.7930 \leqslant c_4 \leqslant 1.0985) \wedge (0.2847 \leqslant c_7 < 0.96165) \overset{15/60\ 100\%}{\longrightarrow} 1$:
Rule 10: $(-0.747 \leqslant c_8 < 0.0572) \wedge (288 \leqslant c_9 \leqslant 421) \overset{5/60\ 100\%}{\longrightarrow} 1$:
Rule 11: $(0.57495 \leqslant c_4 \leqslant 0.793) \wedge (0.96165 \leqslant c_7 < 3.3635) \wedge (288 \leqslant c_9 \leqslant 421) \overset{2/60\ 100\%}{\longrightarrow} 1$:
Rule 12: $(0.57495 \leqslant c_4 \leqslant 0.793) \wedge (0.2847 \leqslant c_7 < 0.96165) \wedge (0.0572 \leqslant c_8 \leqslant 0.1197) \wedge (55 \leqslant c_9 < 288) \overset{5/60\ 80\%}{\longrightarrow} 1$:

information table shown in Appendix B must be examined. A key aspect of this process is to decide which β-reduct to select. However, research of the appropriate selection criteria (e.g., for forecasting) is still evolving, with the criteria suggested in the RST literature to date, including minimising the number of attributes [18], selection based on the increase in the quality of classification by successive augmentation of the subset of attributes [17,52], and the use of human expertise [26].

Currently there is a dearth of empirical evidence to guide users on the appropriate selection criteria to employ. Added to this, VPRS brings with it a further possible measure effecting β-reduct selection, namely the $\beta_{\min}$ value, where the choice of β-reduct may depend on the acceptable level of its associated $\beta_{\min}$ value. As explained above, the aim of this paper is not to formulate an optimum β-reduct selection criterion for forecasting failure (although clearly this would be a worthy topic to pursue in future research), and hence the set of β-reducts satisfying the criteria of Ziarko [29] and described previously were obtained. [9] A β-reduct, namely $\{c_4, c_7, c_8, c_9\}$, was then selected to illustrate the VPRS analysis. [10] Two parts of the analysis are now considered—classification and prediction. Following the method of rule construction in [27] described previously, Table 7 shows the *minimal* set of rules constructed for the β-reduct $\{c_4, c_7, c_8, c_9\}$.

[9] We would also note that the problem of selection criteria is not limited to VPRS. For example, the recent failure prediction study of Laitinen and Kankaanpaa [41, p. 76] used only three (self-selected) financial ratios, to compare the predictive ability of a number of statistical techniques, in the interests of simplicity.

[10] This β-reduct was selected with reference to including a reasonable number of attributes for expositional purposes [18], which performed with a reasonable degree of accuracy in the training sample [17].

With reference to prior expectations, and the summary statistics in Table 5, in general the rules appear logical and consistent with the relative characteristics of the failed and non-failed firms in the training sample. For example, rule 1, which—consistent with prior expectations—specifies a relatively low interval range for GEAR, correctly classifies 13 non-failed firms, with only one failed firm (o_{42}) being misclassified. Interestingly this firm exhibits the lowest gearing level in the failed sub-sample (Appendix A). In addition, rule 9 in Table 7 correctly classifies all 15 failed firms to which it is applied on the basis of relatively high and low interval values for GEAR and QACL, respectively, which is also consistent with prior expectations. However, Table 7 reveals that the worst performing rule (rule 7) is responsible for misclassifying three of the five misclassified companies in the training set. An inspection of Appendix A shows that the three misclassified failed firms (o_{35}, o_{54} and o_{55}), have relatively low gearing ratios (GEAR)—in fact lower than those of the four non-failed firms (o_{13}, o_{21}, o_{29} and o_{30}) correctly classified by rule 7—whereas the four correctly classified non-failed firms have relatively low values of working capital to total assets (WCTA), which is usually associated with failing firms (Table 5).

Overall, the three failed and four non-failed firms covered by rule 7 have characteristics more usually associated with their counterparts in the non-failed and failed samples, respectively—and in conventional failure prediction models would be termed 'grey area' firms or outliers [10]. Despite this the VPRS rule was still able to correctly classify 4 (57.1%) of these seven firms. Although the within-sample classification performance of the VPRS model on the FUSINTER discretised data is encouraging, a more exacting classification test is to asses its predictive accuracy out-of-sample. The VPRS rules in Table 7, together with the boundary interval values derived from the FUSINTER data discretisation process, were applied to the holdout

Table 8
Use of VPRS rules[a]

Rule (Table 7)	1	2	3	4	5	6	7	8	9	10	11	12
Training set	14 [1] (23.3%)	3 (5.0%)	2 (3.3%)	2 (3.3%)	3 (5.0%)	1 (1.7%)	7 [3] (11.7%)	1 (1.7%)	15 (25.0%)	5 (8.3%)	2 (3.3%)	5 [1] (8.3%)
Holdout sample	8 [1] (26.7%)	1 [1] (3.3%)	3 [1] (10.0%)	0 (0.0%)	3 [2] (10.0%)	0 (0.0%)	3 [1] (10.0%)	0 (0.0%)	8 [2] (26.7%)	3 [1] (10.0%)	0 (0.0%)	1 (3.3%)

[a]Figures in round brackets show percentage of sample classified by the rule; figures in square brackets indicate the number of companies misclassified by the rule.

sample of 15 failed and 15 non-failed firms described above.

This was achieved by establishing whether the condition attributes in the β-reduct exactly matched any rule. For those companies that were not exactly matched to a rule, the rule that most closely matched the companies' relevant condition attributes was then selected. Following Stefanowski [53] and Slowinski and Stefanowski [25], a measure of the distance between each classifying rule and each new object was computed for those attributes which determine each rule. Each given new object x is described by values $c_1(x), c_2(x), \ldots, c_m(x)$ for the condition attributes, with $m \leqslant \text{card}(\mathbf{C})$. It follows that the distance of object x from rule y is measured by

$$D = \frac{1}{m}\left\{\sum_{l=1}^{m}\left[k_l\left(\frac{|c_l(x) - c_l(y)|}{v_{l\max} - v_{l\min}}\right)^p\right]\right\}^{1/p},$$

where $p = 1, 2, \ldots$ is a natural number selected by an analyst, $v_{l\max}, v_{l\min}$ the maximal and minimal value of c_l, respectively, k_l the importance coefficient of criterion c_l, m the number of condition attributes in a rule.

In this case the values $p = 2$, and $k_l = 1$ were used for all l, indicating equal importance (weight) amongst the criteria/attributes. It follows from this analysis that the value of p determines the importance of the nearest rule. A small value of p allows a major difference with respect to a single criterion to be compensated by a number of minor differences with regard to other criteria, whereas high values of p will over-value the larger differences and ignore minor ones. Following [25], a value of $p = 2$ is used, thereby implicitly considering *least squares* fitting for each rule.

Table 8 reports the relative frequency with which the 12 VPRS rules were used to classify companies in the training and holdout samples. The analysis shows that there is reasonable degree of consistency in rule usage in the training and holdout samples, with rules 1 and 9 together used to classify 48.3% and 53.4% of firms in the training and holdout samples, respectively. Table 8 also reveals that three of the 15 non-failed and six of the 15 failed firms in the holdout were misclassified by the VPRS rules, with the rules which were responsible for misclassifying the objects exhibiting no systematic pattern. Overall, the VPRS rules correctly classified 91.7% of objects in their training sample, falling to 70.0% in the holdout sample.

6. Comparative predictive accuracy

As explained above, although the purpose of this paper has not been to provide definitive evidence as to whether VPRS outperforms the forecasting ability of existing techniques, [11] in line with previous research we present the classification accuracy of the VPRS method, relative to other techniques, in order to assess its potential and properties. As well as comparing VPRS with the classical parametric methods of logit analysis and MDA, we also present results for the more closely related non-parametric decision tree techniques of RPA [12] and the Elysee ordinal discriminant method.

The RPA was introduced as a failure prediction method by Frydman et al. [15]. The model is in the form of a binary classification tree, based on pattern recognition amongst all the variables in their original form (i.e. continuous for financial ratios). Fig. 1 shows the binary RPA classification tree for the training sample and variables employed in the current study.

With reference to Fig. 1, each node $(i_1, i_2, \ldots, i_6)$ via an iterative process partitions a set of companies on the basis of a particular condition attribute (c_1–c_{12}). Prior probabilities of the failed (π_f) and non-failed (π_n) firms and associated misclassification costs (c_{fn}, c_{nf}) are required. The RPA classification tree in Fig. 1, is calibrated using $c_{fn} = c_{nf} = 1$ and $\pi_f = \pi_n = 0.5$, to facilitate comparisons with the other methods. For the first node (i_1) in Fig. 1, the split point $c_6 \leqslant 1.16945$ and $c_6 > 1.16945$ partitions all of the 60 objects in the training sample. Values further down each branch

[11] This would, inter alia, require an examination of the predictive accuracy of the various methods with respect to misclassification costs (see e.g., [15]).

[12] As with RPA, ID3 [54] is a similar decision tree algorithm acting on categorical and continuous attribute values, which uses an entropy measure to calculate the necessary optimum split points. In common with RPA, ID3 considers all possible attribute binary partitions (splits). For a discussion of ID3 within a corporate failure and loan default setting see [55–57] and for a comparison of ID3 with other techniques, including MDA, RPA and RST, see [30,58,59].

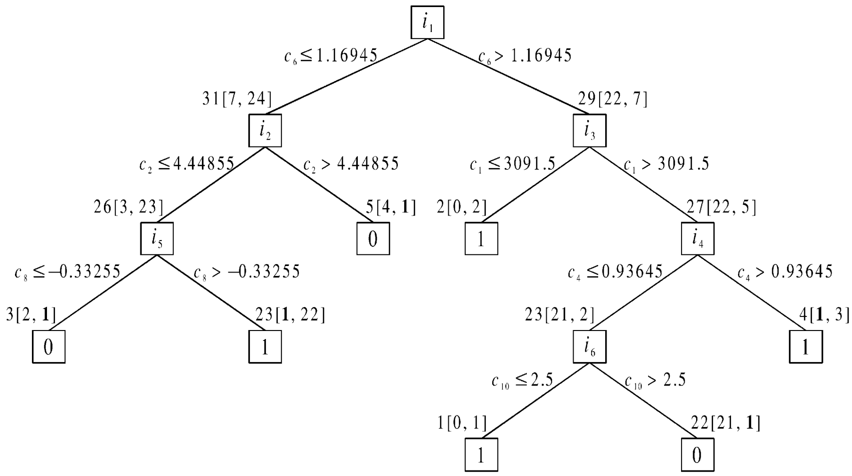

Fig. 1. Binary RPA tree for training set. (Bold figures in square brackets indicate the number of missclassified firms.)

(e.g. from i_1, 31 [7,24] and 29 [22,7]) indicate the split sizes—for example, 31 [7,24] indicates 31 companies satisfy $c_6 \leqslant 1.16945$, of which 7 and 24 have decision class 0 (non-failed) and 1 (failed), respectively. [13]

The end nodes in Fig. 1 have a single value 0 or 1 in the box signifying the objects classified as non-failed or failed, respectively. For example, the bottom left end node (i_5) classifies to 0 (non-failed), where we have 3[2, **1**], signifying that three companies satisfy the successive criteria to arrive at this end node, with 1 firm (identified in bold) misclassified. Although various RPA stopping criteria have been advocated [15] the stopping process employed here is when the number of misclassified companies was less than or equal to that in the VPRS analysis, to facilitate an analysis of the comparative holdout sample forecasting accuracy of the two methods. Hence, Fig. 1 reveals, a total of five firms (shown in bold) were misclassified as per the VPRS method.

The second non-parametric decision tree technique, is the Elysee ordinal discriminant method [60]. Unlike MDA, but in common with VPRS, the Elysee method is designed to discriminate on the basis of ordinal or nominal attributes (see [31]). However, the Elysee method partitions groups of objects based on the distribution of the decision class values within the separated groups, with optimum splitting derived from the classical χ^2 distribution [31]. One of the stopping criteria given in [31], and used here, is simply when an end node comprises all objects from the same decision class. This infers a 100% correct classification rate in training (i.e. overfitting), but is employed here for simplicity, where we focus on the holdout sample predictive accuracy of the non-parametric techniques relative to VPRS.

[13] It is of interest to note that this cut off value also occurred in the discretisation process for the VPRS analysis, and applied to all 60 objects in the training sample.

In the context of the current study, it is also interesting to note that a recent failure prediction study [28] reported that a logit model which employed rank transformed ordinal explanatory variables (ratios) outperformed (in terms of classification accuracy) one which included the cardinal values of the ratios. In consequence, as well as presenting the classification accuracy of logit and MDA models which use the original condition attributes, we also present the results of similar models which incorporate the discretised values (as described above) of the variables.

Table 9 presents the classification accuracy of the various models (labelled 1–13) for both the training and holdout samples. As expected (given the above discussion), it shows that the VPRS, RPA and Elysee methods (1–3, respectively) exhibit high training sample accuracy, with overall correct classification rates of 91.7%, 91.7% and 100%, respectively, but that their accuracy falls significantly out-of-sample, to 70.0%, 63.3% and 46.7%, respectively, and with the Elysee method exhibiting particular poor (worse than random) accuracy. Also noteworthy is the fact that the three methods perform relatively poorly in terms of classifying failed firms, but that the VPRS method outperforms the RPA and Elysee techniques out-of-sample. [14]

[14] As noted by a referee to this paper, the training results, relative to the holdout ones, suggest overfitting may be a problem. In addition, there is a relatively high misclassification rate for the failed firms. As with other non-parametric techniques, and indeed non-linear ones such as NNs, this is probably inevitable, with misclassification rates usually being significantly higher in holdout samples [10]. In any event, the VPRS model was not selected with the aim to minimise classification errors—nor to minimise misclassification costs as between failed and non-failed firms (see [15])—although such issues are clearly worthy of future research.

Table 9
Classification accuracy of methods

Correct classification: training sample		Failed (30)	Non-failed (30)	All (60)
1	VPRS on $\{c_4, c_7, c_8, c_9\}$	26/30 (86.7%)	29/30 (96.7%)	55/60 (91.7%)
2	RPA	28/30 (93.3%)	27/30 (90.0%)	55/60 (91.7%)
3	Elysee method	30/30 (100%)	30/30 (100%)	60/60 (100%)
4	Logit on $\{c_4, c_7, c_8, c_9\}$	23/30 (76.6%)	22/30 (73.3%)	45/60 (75.0%)
5	Logit on $\{c_7, c_9, c_{10}, c_{12}\}$ stepwise	23/30 (76.7%)	24/30 (80.0%)	47/60 (78.3%)
6	Logit on all attributes (c_1–c_{12})	25/30 (83.3%)	23/30 (76.7%)	48/60 (80.0%)
7	Logit on discrete $\{c_4, c_7, c_8, c_9\}$	22/30 (73.3%)	21/30 (70.0%)	43/60 (71.7%)
8	Logit on discrete $\{c_7, c_9, c_{10}, c_{12}\}$ stepwise	26/30 (86.7%)	21/30 (70.0%)	47/60 (78.3%)
9	MDA on $\{c_4, c_7, c_8, c_9\}$	23/30 (76.6%)	22/30 (73.3%)	45/60 (75.0%)
10	MDA on $\{c_7, c_9, c_{10}, c_{12}\}$ stepwise	22/30 (73.3%)	22/30 (73.3%)	44/60 (73.3%)
11	MDA on all attributes (c_1–c_{12})	25/30 (83.3%)	22/30 (73.3%)	47/60 (78.3%)
12	MDA on discrete $\{c_4, c_7, c_8, c_9\}$	22/30 (73.3%)	21/30 (70.0%)	43/60 (71.7%)
13	MDA on discrete $\{c_7, c_9, c_{10}, c_{12}\}$ stepwise	25/30 (83.3%)	22/30 (73.3%)	47/60 (78.3%)
14	RPA on discrete $\{c_1$–$c_{10}\}$ and c_{11}, c_{12}	27/30 (90.0%)	29/30 (96.7%)	56/60 (93.3%)

Correct classification: holdout sample		Failed (15)	Non-failed (15)	All (30)
1	VPRS on $\{c_4, c_7, c_8, c_9\}$	9/15 (60.0%)	12/15 (80.0%)	21/30 (70.0%)
2	RPA	7/15 (46.7%)	12/15 (80.0%)	19/30 (63.3%)
3	Elysee method	5/15 (33.3%)	9/15 (60.0%)	14/30 (46.7%)
4	Logit on $\{c_4, c_7, c_8, c_9\}$	10/15 (66.7%)	10/15 (66.7%)	20/30 (66.7%)
5	Logit on $\{c_7, c_9, c_{10}, c_{12}\}$ stepwise	12/15 (80.0%)	10/15 (66.7%)	22/30 (73.3%)
6	Logit on all attributes (c_1–c_{12})	14/15 (93.3%)	10/15 (66.7%)	24/30 (80.0%)
7	Logit on discrete $\{c_4, c_7, c_8, c_9\}$	11/15 (73.3%)	10/15 (66.7%)	21/30 (70.0%)
8	Logit on discrete $\{c_7, c_9, c_{10}, c_{12}\}$ stepwise	15/15 (100%)	10/15 (66.7%)	25/30 (83.3%)
9	MDA on $\{c_4, c_7, c_8, c_9\}$	10/15 (66.7%)	9/15 (60.0%)	19/30 (63.3%)
10	MDA on $\{c_7, c_9, c_{10}, c_{12}\}$ stepwise	12/15 (80.0%)	9/15 (60.0%)	21/30 (70.0%)
11	MDA on all attributes (c_1–c_{12})	14/15 (93.3%)	8/15 (53.3%)	22/30 (73.3%)
12	MDA on discrete $\{c_4, c_7, c_8, c_9\}$	11/15 (73.3%)	10/15 (66.7%)	21/30 (70.0%)
13	MDA on discrete $\{c_7, c_9, c_{10}, c_{12}\}$ stepwise	15/15 (100%)	11/15 (83.3%)	26/30 (86.7%)
14	RPA on discrete $\{c_1$–$c_{10}\}$ and c_{11}, c_{12}	10/15 (66.7%)	12/15 (80.0%)	22/30 (73.3%)

Models 4 and 9 in Table 9 report the classification results for logit and MDA models which incorporate the same non-discretised variables used in the VPRS analysis (i.e. $\{c_4, c_7, c_8, c_9\}$). They show that the VPRS method outperforms the logit and MDA models in both training sample (overall accuracy = 75% for both models) and the holdout sample (66.7% and 63.3%, respectively). Although not definitive, these results are encouraging, in that, although non-parametric techniques such as VPRS (and indeed non-linear ones such as NNs), are prone to overfitting problems, the VPRS method using FUSINTER discretised variables outperforms the other methods (which use the original variable values) in the holdout sample.

To assess the relative performance of the logit and MDA techniques when not constrained by the use of the VPRS variables, alternative models were estimated. Models 5 and 10 in Table 9 report stepwise selected variables, where the best fitting (significant) set of explanatory variables from all those available (c_1–c_{12}) are included in the respective models (in both cases, the logit and MDA stepwise procedures selected the same four variables, that is, c_7, c_9, c_{10} and c_{12}). The table shows that, relative to models 4 and 9, the within-sample classification of the logit model increases, whereas that of the MDA model declines—but both models exhibit improved holdout sample classification accuracy (to 73.3% and 70.0%, respectively). Hence, in the holdout sample, the MDA classification accuracy now equals that of the VPRS method (70.0%), with the logit accuracy marginally improving on it.

Following [26], logit and MDA classification results are also presented for models which utilise the whole set of explanatory variables [15] (c_1–c_{12}). Table 9 shows that both models (6 and 11) achieve their highest training sample classification accuracy (at 80% and 78.3%, respectively) and that in the holdout sample, the MDA (73.3%) and the logit model (80.0%) now outperform the VPRS method—with the logit results being particularly impressive. We would note, however, that this superior performance results from

[15] We are grateful to an anonymous referee for making this suggestion.

utilising all 12 original explanatory variables, whereas the VPRS results are derived from a subset of four discretised condition attributes.

Table 9 also reports the classification accuracy of logit and MDA models which incorporate the discretised values for the variables used in the VPRS analysis (models 7 and 12) and those selected by the stepwise procedure (models 8 and 13). Models 7 and 12 have identical correct overall classification rates in both the training sample (71.7%) and the holdout sample (70.0%)—the latter being equal to the predictive accuracy of VPRS method which used the same four discretised variables. Indeed, in the holdout sample these models outperform their counterparts (models 4 and 9) which utilised the original variables.

Furthermore, models 8 and 13 (which incorporate discretised values of the stepwise selected variables), have the highest overall classification rates of all the reported models in the holdout sample (83.3% and 86.7%), which, perhaps surprisingly, is higher than their training sample classification accuracy (78.3% in both cases). Prima facie, these results suggest (in relation to the samples analysed), that the FUSINTER data discretisation technique produced variables which did not lead to a reduction in classification accuracy of conventional statistical estimators—a result which is at least partly consistent with the recent failure study [28] discussed above, which focused on rank-transformed financial ratios.

Finally model 14 in Table 9 reports the classification accuracy of the RPA decision tree which incorporates the FUSINTER discretised variables (c_1–c_{10}) along with the original c_{11} and c_{12} variables. Consistent with the preceding analysis, model 14 outperforms model 2 in terms of overall predictive accuracy in the holdout sample (73.3% versus 63.3%), and has a marginally superior performance to the VPRS model in the holdout sample (73.3% versus 70.0%). However, the RPA model utilised 8 discretised values (c_1–c_7, c_9), relative to the subset of four incorporated in the VPRS model.

7. Conclusion

This paper has provided an exposition of variable precision rough sets (VPRS) model which has been developed from rough set theory (RST) as originally formulated by Pawlak [19]. Based on our experimental analysis, the paper has added to the literature in terms of demonstrating the application of VPRS to an important business decision problem, corporate failure prediction. The results of the empirical analysis were encouraging. The use of the FUSINTER method for data discretisation (together with a least square method of nearest rule calculation for classification), mitigated the requirement of the input of a human expert—which may be deemed undesirable by potential users of VPRS, since the use of human expertise may be impractical, relatively costly and may introduce an unacceptable level of subjective bias into the analysis.

Compared to classical statistical methods, and more closely related non-parametric decision tree techniques, the VPRS rules when applied to FUSINTER discretised variables, were found to predict with a reasonable degree of accuracy in training and holdout samples. In addition, it was demonstrated that the logit models, which incorporated the discretised variables, outperformed those based on continuous variables. This may well stem from the fact that the FUSINTER method has the added desirable property of eliminating the influence of outliers. [16]

As is common with experimental research, our results perhaps raise as many questions as they answer. What is clear, however, is that VPRS, building on RST, offers an innovative approach to rule induction, knowledge discovery and management classification problems. Moreover, in a failure prediction setting, VPRS has a number of desirable properties, in terms of information quality and the formulation of explicit rules (which can be assessed by the user for logical interpretation and consistency), and which can be utilised in expert systems—with NASA, for example, currently using applications in this field.

As with decision tree techniques, ceteris paribus, a clear benefit to users of VPRS is the ability to interpret individual rules in a decision-making context (as opposed to interpreting coefficients in conventional statistical models). Hence VPRS generated rules are relatively simple, comprehensible and are directly interpretable with reference to the decision domain. For example, users are not required to possess the technical knowledge and expertise associated with interpreting classical models. These VPRS characteristics are particularly useful to decision makers, [17] who are interested in interpreting the rules (based on factual cases) with direct reference to the outcomes they are familiar with—for example a bank's decision whether or not to grant credit, or in respect of auditors' materiality judgements (see, e.g. [61]).

Although classical statistical methods often rely (particularly in corporate failure studies) on relatively 'crude' methods of variable selection (e.g. stepwise procedures), and the selection of optimal stopping criteria for existing non-parametric decision tree techniques are still evolving, further research in respect of optimal VPRS β-reduct and rule selection criteria is clearly warranted. In addition, the VPRS results presented in this paper are experimental, and do not aim to be definitive in terms of demonstrating the superiority of the new technique over existing methods.

[16] As noted by a referee to this paper, the FUSINTER method provides a 'natural' method to eliminate outliers and may be superior to the rank transformation approach utilised in a recent failure prediction study [28]. Clearly the efficacy of the FUSINTER technique, relative to other methods, in dealing with outliers would be an interesting topic to pursue in future research.

[17] As commented by a referee, these advantages may also be more pronounced where an expert is used to discretise the data, in that the rules derived from such data offer clear and logical interpretation.

Further research is therefore required on these issues, together with an investigation of the asymptotic and sampling properties of VPRS models. Furthermore, the potential loss of information to the user when continuous data are discretised to facilitate RST and VPRS analysis, is an additional issue which is worthy of further attention. It is hoped that the research presented in this paper will stimulate additional work on these important topics.

Appendix A. Variable values for training set

Firm	SALES c_1	ROCE c_2	FFTL c_3	GEAR c_4	CLTA c_5	CACL c_6	QACL c_7	WCTA c_8	LAG c_9	AGE c_{10}	CHAUD c_{11}	BIG6 c_{12}	FAIL d_1
o_1	6762	7.5364	0.1545	0.6233	0.6233	1.5489	0.7356	0.3422	96	74	0	0	0
o_2	16149	−1.0712	0.0271	1.2218	1.2218	0.6236	0.3153	−0.4599	287	29	0	1	0
o_3	8086	15.2024	0.6163	0.3307	0.3307	2.3553	1.7513	0.4482	64	51	0	1	0
o_4	7646	31.2239	0.6312	0.5205	0.4829	1.6397	1.4935	0.3089	286	25	0	0	0
o_5	36067	10.9613	0.354	0.3786	0.3786	1.5852	1.1626	0.2216	301	33	0	1	0
o_6	3205	−1.979	−0.0103	0.3548	0.3548	1.8835	1.2034	0.3134	314	19	0	1	0
o_7	9819	33.8451	0.5326	0.7018	0.7018	1.1864	1.1762	0.1308	249	10	0	1	0
o_8	16737	17.9949	0.4388	0.5236	0.4347	1.1748	0.9795	0.076	55	26	0	1	0
o_9	4168	−6.7091	0.0426	0.786	0.6715	1.2963	0.6288	0.199	298	57	0	0	0
o_{10}	38917	11.1326	0.1898	0.4019	0.1401	1.1728	1.0867	0.0242	241	14	1	1	0
o_{11}	3813	4.408	0.217	0.5398	0.4533	1.1907	0.8894	0.0864	168	27	0	1	0
o_{12}	4049	9.4607	0.2922	0.6316	0.5297	1.3874	0.8809	0.2052	136	18	0	0	0
o_{13}	6912	24.5325	0.3483	0.7866	0.7866	0.942	0.7063	−0.0457	212	39	0	1	0
o_{14}	11362	−37.3497	−0.0814	3.5336	0.6301	1.1944	1.164	0.1225	188	22	0	1	0
o_{15}	24261	9.0815	0.2643	0.5736	0.5072	1.3534	0.444	0.1792	268	25	0	0	0
o_{16}	47679	10.0875	0.2596	0.817	0.3999	1.6964	1.2495	0.2785	271	3	0	1	0
o_{17}	22461	2.637	0.1672	0.1212	0.1212	4.4465	3.3635	0.4176	197	90	0	0	0
o_{18}	24001	−9.2212	0.0051	0.876	0.2761	2.1925	1.0913	0.3293	303	25	0	1	0
o_{19}	4818	−8.8515	−0.0955	0.5081	0.5014	1.6034	1.3363	0.3025	302	28	0	1	0
o_{20}	3518	−2.9022	0.0249	0.7489	0.406	1.2253	0.9207	0.0915	421	36	1	1	0
o_{21}	2857	4.4519	0.1437	0.7228	0.5458	0.9225	0.6929	−0.0423	77	9	0	0	0
o_{22}	9433	6.3359	0.2667	0.4549	0.3983	1.5444	1.2766	0.2169	192	81	0	0	0
o_{23}	103541	14.2607	0.2153	0.7412	0.7412	1.1223	0.7921	0.0906	304	28	1	1	0
o_{24}	3310	6.7887	0.4796	0.5074	0.3426	1.4407	1.3364	0.151	191	51	0	0	0
o_{25}	99520	−7.1057	−0.0101	0.6495	0.1538	1.5268	1.2591	0.081	117	13	1	1	0
o_{26}	6336	4.4826	0.1305	0.7355	0.6198	1.0993	0.4721	0.0616	160	16	0	1	0
o_{27}	5814	8.0062	0.3611	0.4034	0.4034	1.5611	1.2662	0.2263	264	33	0	1	0
o_{28}	3514	−16.8028	−0.1064	1.1831	1.1831	0.7944	0.6491	−0.2432	183	6	0	1	0
o_{29}	7864	1.7552	0.1238	0.7569	0.6965	0.5044	0.4414	−0.3452	117	24	0	0	0
o_{30}	37415	−8.3372	−0.026	0.7883	0.7883	0.7976	0.5631	−0.1596	147	40	0	1	0
o_{31}	3691	4.0961	0.1781	0.7286	0.7034	1.1661	0.5598	0.1169	172	16	1	0	1
o_{32}	6398	−11.5928	−0.0359	0.8613	0.7081	1.063	0.5592	0.0446	265	3	0	1	1
o_{33}	8604	−16.6667	−0.0363	1.0139	1.0139	0.7485	0.5265	−0.255	301	90	0	0	1
o_{34}	17295	5.3741	0.1691	0.8813	0.7585	0.8556	0.5182	−0.1095	309	2	0	1	1
o_{35}	3340	−4.2769	0.0439	0.5763	0.56	0.8433	0.5173	−0.0877	243	17	1	0	1
o_{36}	5574	−0.4819	0.0866	0.8008	0.7772	0.7131	0.5667	−0.223	393	32	0	0	1
o_{37}	3055	−31.2538	−0.2121	1.2343	1.2024	0.734	0.5056	−0.3199	360	2	0	0	1
o_{38}	5105	26.5006	0.4277	0.7005	0.7005	1.3217	1.2668	0.2254	296	2	0	1	1
o_{39}	11528	1.3275	0.066	0.7124	0.6377	0.9967	0.9438	−0.0021	295	40	0	0	1
o_{40}	29300	0.0745	0.0683	0.5977	0.4767	1.1994	0.7513	0.0951	233	25	0	0	1
o_{41}	2958	−9.4013	0.0145	1.4865	1.4865	0.4974	0.2847	−0.747	301	11	0	0	1
o_{42}	2978	8.5486	0.3285	0.3898	0.3883	2.0519	1.8493	0.4084	265	9	0	1	1
o_{43}	47237	−2.5559	0.0735	1.0072	0.8516	0.7437	0.4385	−0.2183	86	8	0	1	1
o_{44}	8125	1.0502	0.1358	0.8098	0.7585	0.6738	0.4596	−0.2474	297	59	0	0	1
o_{45}	10339	4.4452	0.1176	0.7461	0.656	1.1399	0.6296	0.0918	144	19	0	0	1
o_{46}	27599	−5.9041	0.0519	0.7977	0.4824	0.8959	0.7514	−0.0502	295	21	0	0	1
o_{47}	8070	13.9864	0.2122	0.9969	0.2986	2.0674	0.7283	0.3188	303	26	1	1	1
o_{48}	48162	6.1787	0.1181	0.9996	0.7437	1.3007	0.907	0.2236	303	34	0	0	1
o_{49}	6953	0.0648	0.1447	0.952	0.4976	1.041	0.5355	0.0204	362	14	1	1	1
o_{50}	3900	−2.684	0.1451	0.6925	0.6672	0.6086	0.4879	−0.2611	289	3	1	0	1

Firm	SALES c_1	ROCE c_2	FFTL c_3	GEAR c_4	CLTA c_5	CACL c_6	QACL c_7	WCTA c_8	LAG c_9	AGE c_{10}	CHAUD c_{11}	BIG6 c_{12}	FAIL d_1
o_{51}	4718	−10.01	−0.0231	0.9967	0.5726	0.7646	0.4307	−0.1348	301	8	0	0	1
o_{52}	4028	−4.9692	0.0309	0.8171	0.7586	1.0696	0.927	0.0528	301	14	1	0	1
o_{53}	5566	−9.4513	−0.0179	0.7722	0.6596	0.7774	0.4643	−0.1468	293	11	0	1	1
o_{54}	13909	−1.4262	0.0642	0.689	0.5916	0.9457	0.312	−0.0321	179	74	0	0	1
o_{55}	3981	−3.6023	0.0333	0.7144	0.5261	0.816	0.6391	−0.0968	283	32	0	0	1
o_{56}	4999	1.676	0.0565	0.9994	0.7112	1.1862	0.7188	0.1324	207	2	1	1	1
o_{57}	2859	−3.5788	0.0341	0.639	0.5742	1.3541	1.2728	0.2033	301	24	0	0	1
o_{58}	28145	0.5269	0.0856	0.9049	0.8618	0.742	0.4457	−0.2223	153	8	1	1	1
o_{59}	3601	−30.7797	−0.3283	0.7444	0.7444	1.1202	0.3757	0.0895	243	5	1	1	1
o_{60}	8393	−5.4902	0.0619	0.8307	0.732	0.9182	0.8027	−0.0599	301	14	0	0	1

Appendix B. Discretised variable values for training set

Firm	SALES c_1	ROCE c_2	FFTL c_3	GEAR c_4	CLTA c_5	CACL c_6	QACL c_7	WCTA c_8	LAG c_9	AGE c_{10}	CHAUD c_{11}	BIG6 c_{12}	FAIL d_1
o_1	1	3	2	1	1	2	0	2	0	1	0	0	0
o_2	1	2	0	3	2	0	0	0	0	1	0	1	0
o_3	1	3	3	0	0	2	1	2	0	1	0	1	0
o_4	1	3	3	0	1	2	1	2	0	1	0	0	0
o_5	2	3	3	0	0	2	1	2	1	1	0	1	0
o_6	0	2	0	0	0	2	1	2	1	0	0	1	0
o_7	1	3	3	1	1	1	1	2	0	0	0	1	0
o_8	2	3	3	0	0	1	1	1	0	1	0	1	0
o_9	0	1	1	1	1	1	0	2	1	1	0	0	0
o_{10}	2	3	2	0	0	1	1	0	0	0	1	1	0
o_{11}	0	3	3	0	0	1	0	1	0	1	0	1	0
o_{12}	0	3	3	1	1	2	0	2	0	0	0	0	0
o_{13}	1	3	3	1	2	0	0	0	0	1	0	1	0
o_{14}	1	0	0	3	1	1	1	2	0	0	0	1	0
o_{15}	1	3	3	0	1	1	0	2	0	1	0	0	0
o_{16}	2	3	3	2	0	2	1	2	0	0	0	1	0
o_{17}	1	3	2	0	0	2	1	2	0	1	0	0	0
o_{18}	1	1	0	2	0	2	1	2	1	1	0	1	0
o_{19}	0	1	0	0	1	2	1	2	1	1	0	1	0
o_{20}	0	2	0	1	0	1	0	1	1	1	1	1	0
o_{21}	0	3	2	1	1	0	0	0	0	0	0	0	0
o_{22}	1	3	3	0	0	2	1	2	0	1	0	0	0
o_{23}	2	3	3	1	2	0	0	1	1	1	1	1	0
o_{24}	0	3	3	0	0	2	1	2	0	1	0	0	0
o_{25}	2	1	0	1	0	2	1	1	0	0	1	1	0
o_{26}	1	3	2	1	1	0	0	1	0	0	0	1	0
o_{27}	1	3	3	0	0	2	1	2	0	1	0	1	0
o_{28}	0	0	0	3	2	0	0	0	0	0	0	1	0
o_{29}	1	3	2	1	1	0	0	0	0	0	0	0	0
o_{30}	2	1	0	1	2	0	0	0	0	1	0	1	0
o_{31}	0	3	2	1	2	0	0	1	0	0	1	0	1
o_{32}	1	0	0	2	2	0	0	0	0	0	0	1	1
o_{33}	1	0	0	2	2	0	0	0	1	1	0	0	1
o_{34}	1	3	2	2	2	0	0	0	1	0	0	1	1
o_{35}	0	2	1	1	1	0	0	0	0	0	1	0	1
o_{36}	0	2	1	2	2	0	0	0	1	1	0	0	1
o_{37}	0	0	0	3	2	0	0	0	1	0	0	0	1
o_{38}	0	3	3	1	1	1	1	2	1	0	0	1	1

Firm	SALES c_1	ROCE c_2	FFTL c_3	GEAR c_4	CLTA c_5	CACL c_6	QACL c_7	WCTA c_8	LAG c_9	AGE c_{10}	CHAUD c_{11}	BIG6 c_{12}	FAIL d_1
o_{39}	1	2	1	1	1	0	0	0	1	1	0	0	1
o_{40}	1	2	1	1	1	1	0	1	0	1	0	0	1
o_{41}	0	0	0	3	2	0	0	0	1	0	0	0	1
o_{42}	0	3	3	0	0	2	1	2	0	0	0	1	1
o_{43}	2	2	1	2	2	0	0	0	0	0	0	1	1
o_{44}	1	2	2	2	2	0	0	0	1	1	0	0	1
o_{45}	1	3	1	1	1	0	0	1	0	0	0	0	1
o_{46}	1	2	1	2	1	0	0	0	1	0	0	0	1
o_{47}	1	3	2	2	0	2	0	2	1	1	1	1	1
o_{48}	2	3	1	2	2	1	0	2	1	1	0	0	1
o_{49}	1	2	2	2	1	0	0	0	1	0	1	1	1
o_{50}	0	2	2	1	1	0	0	0	1	0	1	0	1
o_{51}	0	0	0	2	1	0	0	0	1	0	0	0	1
o_{52}	0	2	1	2	2	0	0	0	1	0	1	0	1
o_{53}	0	0	0	1	1	0	0	0	1	0	0	1	1
o_{54}	1	2	1	1	1	0	0	0	0	1	0	0	1
o_{55}	0	2	1	1	1	0	0	0	0	1	0	0	1
o_{56}	0	2	1	2	2	1	0	2	0	0	1	1	1
o_{57}	0	2	1	1	1	1	1	2	1	0	0	0	1
o_{58}	1	2	1	2	2	0	0	0	0	0	1	1	1
o_{59}	0	0	0	1	2	0	0	1	0	0	1	1	1
o_{60}	1	2	1	2	2	0	0	0	1	0	0	0	1

References

[1] Keasey K, McGuiness P, Short H. Multilogit approach to predicting corporate failure—further analysis and the issue of signal consistency. OMEGA 1990;18(1):85–94.

[2] Luoma M, Laitinen E. Survival analysis as a tool for company failure prediction. OMEGA 1991;19(6):673–80.

[3] Mar Molinero C, Ezzamel M. Multidimensional scaling, applied to corporate failure. OMEGA 1991;19(4):259–74.

[4] Trippi R, Turban E. Neural networks in finance and investment. London: Probus Publishing, 1993.

[5] Wong B, Bodnovich T, Selvi Y. A bibliography of neural network business applications research: 1988–1994. Expert Systems 1995;12(3):253–61.

[6] Altman E. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. Journal of Finance 1968;4(4):589–609.

[7] Altman E. Corporate financial distress: a complete guide to predicting, avoiding and dealing with bankruptcy. New York: Wiley, 1983.

[8] Taffler R. Empirical models for the monitoring of UK corporations. Journal of Banking and Finance 1984;8: 199–227.

[9] Peel M. The liquidation/merger alternative. Avebury: Aldershot, 1990.

[10] Morris R. Early warning indicators of corporate failure: a critical review of previous research and further empirical evidence. Ashgate: Aldershot, 1997.

[11] Tam K, Kiang M. Managerial applications of neural networks: the case of bank failure prediction. Management Science 1992;38(7):926–47.

[12] Altman E, Marco G, Varetto F. Corporate distress diagnosis: comparisons using linear discriminant analysis and neural networks. Journal of Banking and Finance 1994;18:505–29.

[13] Lee K, Han I, Kwon Y. Hybrid neural network models for bankruptcy prediction. Decision Support Systems 1996;18:63–72.

[14] Bryant S. A case-based reasoning approach to bankruptcy prediction modelling. Intelligent Systems in Accounting, Finance and Management 1997;6:195–214.

[15] Frydman H, Altman E, Kao D. Introducing recursive partitioning for financial classification: the case of financial distress. The Journal of Finance 1985;XL(1):269–91.

[16] Maddala G. Introduction to Econometrics. 2nd ed. London: Prentice-Hall, 1992.

[17] Moraczewski I, Sudnik-Wojcikowska B, Borkowski W. Rough sets in floristic description of the inner-city of Warsaw. FLORA 1996;191:253–60.

[18] Wakulicz-Deja A, Paszek P. Diagnose progressive encephalopathy applying the rough set approach. International Journal of Medical Informatics 1997;46:119–27.

[19] Pawlak Z. Rough sets. International Journal of Information and Computer Sciences 1982;11(5):341–56.

[20] Pawlak Z, Grzymala-Busse J, Slowinski R, Ziarko W. Rough sets. Association for Computing Machinery: Communications of the ACM 1995;38(11):89–96.

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

[22] Pawlak Z, Slowinski K, Slowinski R. Rough classification of patients after highly selective vagotomy for duodenal ulcer. International Journal of Man–Machine Studies 1986;24: 413–33.

[23] Slowinski R. Intelligent decision support: handbook of applications and advances of the rough sets theory. Dordrecht: Kluwer Academic Publishers, 1992.

[24] Grzymala-Busse J. LERS—a system for learning from examples based on rough sets. In: Slowinski R, editor.

Intelligent decision support: handbook of applications and advances of the rough sets theory. Dordrecht: Kluwer Academic Publishers, 1992. p. 3–18.

[25] Slowinski R, Stefanowski J. Rough DAS and rough class software implementations of the rough sets approach. In: Slowinski R, editor. Intelligent decision support: handbook of applications and advances of the rough sets theory. Dordrecht: Kluwer Academic Publishers, 1992. p. 445–56.

[26] Dimitras A, Slowinski R, Susmaga R, Zopounidis C. Business failure using rough sets. European Journal of Operational Research 1999;114:263–80.

[27] An A, Shan N, Chan C, Cercone N, Ziarko W. Discovering rules for water demand prediction: an enhanced rough-set approach. Engineering Applications in Artificial Intelligence 1996;9(6):645–53.

[28] Kane G, Richardson F, Meade N. Rank transformation and the prediction of corporate failure. Contemporary Accounting Research 1998;15(2):145–66.

[29] Ziarko W. A variable precision rough set model. Journal of Computer and System Sciences 1993;46:39–59.

[30] Kattan MW, Cooper RB. The predictive accuracy of computer-based classification decision techniques. A Review and Research Directions OMEGA 1998;26(4):467–82.

[31] Teghem J, Benjelloun M. Some experiments to compare rough sets theory and ordinal statistical methods. In: Slowinski R, editor. Intelligent decision support: handbook of applications and advances of the rough sets theory. Dordrecht: Kluwer Academic Publishers, 1992. p. 267–86.

[32] Fayyad U, Irani K. Multi-interval discretization of continuous-valued attributes for classification learning. In: Proceedings of the 13th International Joint Conference on Artificial Intelligence. Chambery, France: Kaufmann, 1993. p. 1022–7.

[33] Kerber R. ChiMerge: discretization of numeric attributes. In: Proceedings of the Ninth International Conference on Artificial Intelligence (AAAI). Cambridge, MA: AAAI Press/The MITS Press, 1992. p. 123–8.

[34] Liu H, Setiono R. Feature selection via discretization. IEEE Transactions on Knowledge and Data Engineering 1997;9(4):642–5.

[35] Chmielewski M, Grzymala-Busse J. Global discretization of continuous attributes as preprocessing for machine learning. International Journal of Approximate Reasoning 1996;15: 319–31.

[36] Zighed D, Rabaseda S, Rakotomala R. FUSINTER: a method for discretisation of continuous attributes. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 1998;6(3):307–26.

[37] FAME. FAME for windows user guide. London: Bureau Van Dijk Electronic Publishing Limited, 1995.

[38] Jones F, Raghunandan K. Client risk and recent changes in the market for audit services. Journal of Accounting and Public Policy 1998;17:169–81.

[39] MAPLE V Release 4 software package. Waterloo Maple Inc. 57 Erb Street W. Waterloo, Ont. Canada N2L 6C2.

[40] Ziarko W, Shan N. KDD-R: a comprehensive system of knowledge discovery using rough sets. In: Lin TY, editor. The Third International Workshop on Rough Sets and Soft Computing Proceedings. San Jose State University, San Jose, California, 1994. p. 164–73.

[41] Laitinen T, Kankaanpaa M. Comparative analysis of failure prediction methods: the Finnish case. The European Accounting Review 1999;8(7):67–92.

[42] Altman E. Corporate financial distress and bankruptcy. New York: Wiley, 1993.

[43] Storey D, Keasey K, Watson R, Wynarczyk P. The performance of small firms. London: Croom Helm, 1987.

[44] Ohlson J. Financial ratios and the probabilistic prediction of bankruptcy. Journal of Accounting Research 1980;15(1): 3–25.

[45] Keasey J, Watson R. Non-financial symptoms and the prediction of small company failure: a test of Argenti's hypotheses. Journal of Business Finance and Accounting 1987;14(3):335–54.

[46] Hudson J. The age, regional and industrial structure of company liquidations. Journal of Business Finance and Accounting 1987;14(2):200–15.

[47] Chow C, Rice S. Qualified audit opinions and auditor switching. The Accounting Review 1982;April:326–35.

[48] Schwartz K, Menon K. Auditor switches by failing firms. The Accounting Review 1985;14(2):248–60.

[49] Holland K, Light L, Galen M. Big six firms are firing clients. Business Week 1993;3307:76–7.

[50] Simunic D, Stein M. The impact of litigation risk on audit pricing: a review of the economics and the evidence. Auditing: A Journal of Practice and Theory 1996;15(Suppl.): 119–34.

[51] Siegal S, Castellan N. Non-parametric statistics. London: McGraw-Hill, 1988.

[52] Slowinski R, Zopounidis C, Dimitras A. Prediction of company acquisition in Greece by means of the rough set approach. European Journal of Operational Research 1997;100:1–15.

[53] Stefanowski J. Classification support based on the rough set theory. In: Proceedings IIASA Workshop on User-Oriented Methodology and Techniques of Decision Analysis and Support, Serock, 1992.

[54] Quinlan JR. Introduction of decision trees. Machine Learning 1986;1:81–106.

[55] Messier WF, Hansen JV. Inducing rules for expert system development: an example using default and bankruptcy data. Management Science 1988;34(12):1403–15.

[56] Kim H, Koehler GJ. Theory and practice of decision tree induction. OMEGA 1995;23(6):637–52.

[57] Lee KC, Han I, Kwon Y. Hybrid neural network models for bankruptcy prediction. Decision Support Systems 1996;18: 63–72.

[58] Kattan MW, Cooper RB. A simulation of factors affecting machine learning techniques: an examination of partitioning and class proportions. OMEGA 2000;28:501–12.

[59] Tsumoto S, Tanaka H. Induction of probabilistic rules based on rough set theory, algorithmic learning theory ALT93. Proceedings of the 4th International Workshop. Tokyo, Japan: Springer, 1993. p. 410–23.

[60] Ellard C. Programme elysee: presentation and application. Revue Metra 1967;1:503–20.

[61] Steinbart P. The construction of a rule-based expert system as a method for studying materiality judgements. Accounting Review 1987;97–116.