The International Journal of Banking and Finance, Volume 10 (Number 1), 2013: Pages 34-56

MODELING CREDIT RISK: AN APPLICATION OF THE ROUGH SET METHODOLOGY

Reyes Samaniego Medina and Maria Jose Vazquez Cueto
Pablo de Olavide University, Spain and Seville University, Spain

Abstract

The Basel Accords encourages credit entities to implement their own models for measuring financial risk. In this paper, we focus on the use of internal ratings-based (IRB) models for the assessment of credit risk and, specifically, on one component that models the probability of default (PD). The traditional methods used for modeling credit risk, such as discriminant analysis and logit and probit models, start with several statistical restrictions. The rough set methodology avoids these limitations and as such is an alternative to the classic statistical methods. We apply the rough set methodology to a database of 106 companies that are applicants for credit. We obtain ratios that can best discriminate between financially sound and bankrupt companies, along with a series of decision rules that will help detect operations that are potentially in default. Finally, we compare the results obtained using the rough set methodology to those obtained using classic discriminant analysis and logit models. We conclude that the rough set methodology presents better risk classification results.

***Keywords*:** Rating, Credit risk, Basel Accords, Rough sets

JEL Classification: G21, G32

1. Introduction

The Basel Accords opened the way for and encouraged credit entities to implement their own models for measuring financial risks. In this paper, we focus on the use of internal ratings-based (IRB) models for the assessment of credit risk and, specifically, on one of the approaches to model the probability of default (PD).

The traditional methods used for modeling credit risk, such as discriminant analysis and logit and probit models, start with several statistical restrictions. The rough set methodology avoids these limitations and is presented as an alternative to the classic statistical methods.

The objective of our study is to apply the rough set methodology to a database composed of 106 companies that are debtors of the same financial entity to obtain the ratios that best discriminate between healthy and bankrupt companies. A second objective is to find a series of decision rules that will help detect potentially failing credit operations as a first step in modeling the probability of default. Finally, we compare the results obtained using the rough set methodology to those obtained using classic discriminant analysis and logit models. We conclude that the rough set methodology presents good risk classification results.

This paper is structured as follows. Section 2 reviews the most significant empirical studies. Section 3 introduces the theory of rough sets. In section 4, we continue with a description of the sample of companies used for the empirical study. The empirical application is described in section 5 wherein we first use the rough set methodology to determine the variables that may explain the default, and then compare the results obtained using this methodology to those obtained using classic discriminant analysis and logit models. Finally, in section 6, we draw a series of conclusions.

2. Literature Review

The models for predicting business failure and estimating the probability of default (PD) required by the Basel Accords have been the subject of several studies, conducted not only by academics but also by the financial sector itself. All of the theoretical effort has been focused on the modeling of the stochastic process associated with insolvency and on determining the variables that must be included in these models. Among these traditional models, we can distinguish between univariate and multivariate models. Univariate models examine the behavior of each variable separately to explain any insolvency. One of the classic studies using this method was conducted by Beaver (1966), who found a number of financial ratios that could discriminate between healthy and bankrupt companies during the 5-year period prior to the occurrence of the actual default. Other notable studies were those conducted by Courtis (1978) and Altman (1993).

Unlike the univariate models, the multivariate models combine the information provided by a set of variables. The study that pioneered this method was performed by Altman in 1968, in

which he proposed a discriminant analysis that combined the information provided by 25 financial ratios. A wide variety of studies have been based on discriminant analysis, including those by Dambolena (1980) and Laitinen (1991). In Spain, Cabedo et al. (2004) presented an adaptation of the discriminant model to calculate the probability of default in companies. The model was applied to a portfolio of hypothetical borrowers from the same sector to calculate the regulatory capital according to the foundational IRB method of the Basel Accord. The importance of this technique is demonstrated in the bibliographic review by Dimitras (1996). After analyzing 158 articles on the prediction of business insolvency during the period from 1932 to 1994, Dimitras concluded that discriminant analysis is the model that is most frequently used to resolve this type of problem.

Other authors have opted for logit and probit analysis. Ohlson (1980) was the first to apply this type of technique to predict company insolvency. Wilson (1997) developed the Credit Portfolio View model for McKinsey, establishing a discrete process with multiple periods. With this methodology, the probability of default is obtained through logit functions of indices of macroeconomic variables that, in some ways, represent the functioning of the economy (Zmijewski, 1984). Dimitras (1996) found that the logit model is the second most frequently used model for resolving the problem of company bankruptcy.

Fernández (2005), in an attempt to combine univariate and multivariate analysis, conducted an empirical study in which he used a prior univariate analysis to select those ratios with greater discriminant power within each of the categories of ratios established from among the 23 ratios initially considered.[1] Subsequently, he performed a logit and probit multivariate analysis to obtain scores for each company; these scores enabled a rating system to be established and default probabilities to be assigned. Trucharte et al. (2002) developed a system for *rating* borrowers by estimating a logistic regression model that utilizes economic and financial information. The scores obtained are used to establish homogeneous categories in which the various borrowers are classified or rated and the probability of default that can be assigned to each category.

More recently, Altman and Sabato (2007) developed a distress predictor model specifically for the SME sector and analyzed its effectiveness against that of a generic corporate model. They

[1] These categories were liquidity, leverage, activity, debt cover and productivity.

used a logit regression model technique on panel data from more than 2,000 US firms over the period from 1994 to 2002.

In parallel with these studies, other methods have been explored to overcome the restrictive hypotheses that models of statistical inference impose on the variables. These hypotheses usually do not conform to reality and distort the results obtained. For these reasons, Eisenbeis (1977), Ohlson (1980) and Zavgren (1983) questioned the validity of the traditional models. In particular, techniques originating from the field of artificial intelligence have begun to be used; programs have been produced that are capable of generating knowledge from empirical data and then utilizing that knowledge to make inferences based on new data. Within this approach, we can distinguish techniques that seek knowledge by identifying patterns in the data. Among these are various classes of neural networks and other techniques that infer decision rules from the base data. The rough set methodology belongs to this last group of techniques. Authors such as Dimitras et al. (1998) and Daubie et al. (2002) have applied this technique to the classification of commercial loans. Other authors, such as Ahn et al. (2000), have combined the rough set methodology with neural networks to predict company failure. In Spain, various studies can similarly be found that apply the rough set methodology for the prognosis of company insolvency. Segovia et al. (2003) applied this technique to the prediction of insolvency in insurance companies, and Rodríguez et al. (2005) utilized it for the same purpose in a sample of small- and medium-sized enterprises (SMEs).

3. Theoretical Framework of the Rough Set Methodology

Rough set theory, first proposed by Z. Pawlak in 1982, is considered as an appropriate tool for handling cases in which there is considerable vagueness and imprecision. More specifically, the method is effective at working with problems of multidimensional classification (Pawlak et al., 1994). The basic idea rests on the indiscernibility relation that describes elements that are indistinguishable from each other. Its principal objective is to find basic decision rules that enable the acquisition of new knowledge. Its key concepts are *discernibility, approximation, reducts* and *decision rules.*

The point of departure for the method is the existence of an information/decision table in which each element is characterized by a set of variables (attributes) and a decision variable that classifies the element into one of two or more categories. *Indiscernibility* exists when two elements are characterized by the same properties for all variables, but the categories in which they are classified do not coincide. This is the basis of the rough set theory. In such a case, for

each class of decision or category X and for each subset B that contains variables, two sets are constructed; these sets are called the set of the *lower approximation* of the decision class and the set of the *upper approximation* of the decision class, respectively. The set of the lower approximation of decision class X with respect to the variables B, which is called $\underline{B}X$, is given by the group of all elements that, being characterized by B, belong to class X with complete certainty. The set of the upper approximation of decision class X, which is called $\overline{B}X$, is given by the group of elements that, based on the information B that we possess, may belong to class X, but about which we cannot be sure. The elements that are different between the two sets form the "doubtful" elements, which are those elements that, using only the information contained in B, are not known with complete certainty to belong to class X. When these different elements exist (i.e., when the difference is not zero), it is said that class X is a rough set with respect to the subset of variables B. This set can be characterized numerically by the quotient between the cardinal of the set of lower approximation and the cardinal of the set of upper approximation. This quotient is known as the "accuracy of approximation". If various decision classes exist, the sum of the cardinals of the lower approximations divided by the total of all elements is known as the "quality of approximation of the classification, by means of set B", and this is the percentage of the elements that have been correctly classified.

Another important aspect of this technique is the reduction of the initial table of data, which eliminates the redundant information. This process is performed using the *reducts*. A reduct is a minimum set of variables that conserve the same capacity for classifying the elements as the full table of information. A reduct is thus an essential part of the information and constitutes the most concise way of differentiating between the decision classes.[2]

The final stage of the rough set analysis is the creation of *decision rules*, rules that allow us to determine whether a given element belongs to particular decision classes. These rules represent knowledge and are generated by combining the reducts with the values of the data analyzed. A decision rule is a logical statement of the following type: "IF particular conditions are met, THEN the element belongs to a particular decision class". These rules allow us to classify new elements easily.[3]

---

[2] The reducts were obtained based on the equivalence classes that defined the indiscernibility relation on the set of observations.

[3] For more detail on the formal mathematical aspects of the methodology, Komorowski et al. (1999) may be consulted.

4. Data and Variables

A. Selection of the Sample.

We adopted the following approach for both selecting the sample and choosing the independent variables. Following Altman (1968), we have paired a number of healthy and bankrupt companies of similar size and sector, thus taking a sample in which the bankrupt companies represent 50% of the total. When selecting the sample, the data under consideration should be obtained for the same period of time for healthy and bankrupt companies alike. However, the companies in bankruptcy or suspension of payments tend to delay the presentation of their accounting data in the time period prior to their declarations of insolvency. To overcome this inconvenience, we have collected the accounting data from the last full year prior to the bankruptcy from the most recent data available.

In the development of our model, we employed a database, provided by a Spanish savings bank that contains information on companies that requested and obtained a loan from the bank. These companies were divided into two groups: healthy and bankrupt. In particular, the sample of bankrupt companies used for the analysis only included those companies with loans whose unpaid debt, whether of interest or principal, amounted to a percentage of more than 10% of the full risk accepted. The computation date was December 31, 2003.

The group of healthy companies, i.e., those that did not default in the time horizon considered, was selected using the individual pairing technique, controlled by those characteristics that could affect the relationships between financial ratios and failure. Each company in the failed group was matched with a healthy company of the same industry and same approximate size. In relation to the sector, the pairing was achieved at a level of four digits of the C.N.A.E. (National Classification of Economic Activities) of 1993. The criterion adopted for pairing by size was the total assets.

As a homogenizing factor for all of the companies, we controlled for the conditions that the total amount of the customer's operations with the savings bank, i.e., their live risk, should exceed 60,120 euros and that the companies should all be public limited companies (PLCs), which facilitates access to their accounting statements.

In total, the sample contained 106 Spanish companies, 53 failed and 53 healthy, with a very diverse spread of economic activities.[4]

B. Selection of the Independent Variables for the Models

The independent variables chosen for the construction of the models were selected from the financial statements, principally from the balance sheet and the profit and loss account, of the companies that comprise the sample. These accounting statements were extracted from the SABI database, developed by Informa, S.A., which includes more than 95% of the companies that present their accounts in the Mercantile Register in Spain. Given that most of the companies that went bankrupt presented their financial statements neither in the preceding year nor in the two years prior to the date of default, we obtained the latest data available corresponding to the year prior to the company bankruptcy, as explained above. Thus, the year t-1 corresponds to that of the latest available accounts.

Table 1: Ratios considered in the analysis

This table summarizes the variables that are potentially explanatory for company bankruptcy. These, in general, include ratios of liquidity, indebtedness, structure, rotation, generation of resources and profitability.

Liquidity Ratios	
Degree to which the company's assets that can be liquidated, in the short term, are sufficient to meet the payments required for the short-term debts contracted.	R1=Current assets / current liabilities R2=(Quick + available assets) / current liabilities R3=Available assets / current liabilities R4=(Quick + available assets - current liabilities) / (Operating costs + Personnel costs + Variation provisions + Other operating costs)
Indebtedness Ratios	

[4] We have excluded property development and property sales companies from the analysis, as these companies have characteristics that are very peculiar and different from other companies. In the assessment of the loan application made by this type of company, the decisive factor for granting the loan is the viability of the specific project for which the loan is sought. This information is not reflected in corresponding accounting statements.

Relationship between the different components of the liabilities, in the short and long term, and the company's own funds and between the cost of the debt and the liabilities or the profits and funds generated.	R5=Long-Term Debt / Net Worth R6=Net Worth / Total Liabilities R7=Long-Term Debt / (Long-Term Debt + Current Liabilities) R8=Financial Costs / (Long Term Debt + Current Liabilities) R9=Financial Costs / (Gross Profits + Provision for Amortization) R10=Financing costs / Gross profits R11=Long-Term Debt / Total Liabilities
Structural Ratios	
Proportionality between the balance sheet items of assets and liabilities and in the composition of these items.	R12=(Current Assets - Current Liabilities) / Total Assets. R13=Current Assets / Total Assets.
Rotation Ratio	
Measure of the dynamism of the business activity in relation to the structure of the company.	R14=(Current Assets – Current Liabilities) / (Net Turnover + Other Income from Operations)
Resource Generation Ratios	
Relationship of the self-financing capacity of the company to various accounting magnitudes.	R15=(Net Profit/Loss for period + Amortization Provision) / (Net Turnover + Other Income from Operations)

	R16=(Net Profit/Loss for period + Amortization Provision) / Current Liabilities R17=(Net Profit/Loss for period + Amortization Provision) / (Long-Term Debt + Current Liabilities) R18=(Net Profit/Loss for period + Amortization Provision) / Total Liabilities R19=(Gross Profits + Amortization Provision) / Current Liabilities
Profitability Ratios	
Comparison of the profit obtained at various levels with the resources invested	R20=(Operating Profit/Loss + Financial Income + Profits from financial investments + Exchange rate gains) / Total Assets R21=Profit/Loss from ordinary activities / Total Liabilities R22=Pre-tax Profits / Net Worth R23=Pre-tax Profits / Total Liabilities R24=Profit/Loss for the period / Net Worth R25=Gross Profits / Total Assets

Source: Trujillo *et al*., (2004)

The accounting information derived from the sample selected was subjected to a meticulous study with the aim of detecting and resolving possible anomalies or significant incidents that could distort the final analysis. Those atypical companies with clear and insuperable anomalies in their accounts were excluded from the sample. For example, those companies that presented profits despite being in a situation of default were eliminated.

Twenty-five ratios were selected by choosing a broad set of variables that are potentially explanatory for company bankruptcy based on the frequency and efficacy with which the ratios have been used in other predictive models of company insolvency or in the analysis of banking risks.

The variables used include ratios of liquidity, indebtedness, structure, rotation, generation of resources and profitability. The specific ratios considered in the analysis are given in Table 1.

5. Empirical Application

A. Application of the Rough Set Mrthodology

For the empirical application, the values of the 25 economic/financial ratios shown in Table 1 were calculated for each of the 53 bankrupt companies for the financial year before entering into default, and a similar procedure was adopted for each matched healthy company. This process produced a table of information containing 106x25 data items. An additional column that indicates whether the company in question is in a situation of bankruptcy or health is included in the table of information. Thus, we have assigned the value 0 to bankrupt companies and the value 1 to the matched healthy companies. We thus obtain an information-decision table with 106x26 data items.

From these data, we determine which ratio or ratios of the 25 variables serve to explain a company's state of default as the first step in calculating the probabilities of default.

First, Napierian logarithms of the values of the ratios were computed to avoid problems with the normality of the variables when applying the discriminant analysis. Then, given the nature of the variables considered, we proceeded to discretize the values. This is not an essential requirement for the application of the technique, but it facilitates the interpretation of the results; it is a more consistent way to identify bankrupt or healthy companies when the values of the variables considered fall within the same range but do not coincide exactly. For this, we utilized the codification given in Table 2.[5]

The next step is to determine the accuracy provided by the explanatory variables using the ROSE software. The quality of the approximation is 1.[6]

[5] We discretized the variables by grouping them into four ranges based on the number of observations belonging to each range. For this, we utilized the ROSE software, provided by the Institute of Computing Science of the Poznan University of Technology, and we thank the Institute for making this software available to us.

[6] The quality of the approximation is expressed by the ratio between the number of companies classified correctly and the total number of companies that comprise the sample.

Table 2: Codification ranges of the variables

Napierian logarithms of the values of the ratios are first computed. Then, the values are discretized according to the ranges identified in this table.

Variables	***CODIFIED VALUE***			
	0	***1***	***2***	***3***
R1	(-inf, 0.00434058)	(0.00434058, 0.00437793)	(0.00437793, 0.00467188)	(0.00467188, +inf)
R2	(-inf, 0.00131066)	(0.00131066, 0.00215777)	(0.00215777, 0.00653424)	(0.00653424, +inf)
R3	(-inf, 3.01154e-005)	(3.01154e-005, 4.19928e-005)	(4.19928e-005, 0.000818944)	(0.000818944, +inf)
R4	(-inf, -0.00126886)	(-0.00126886, -0.000469693)	(-0.000469693, -0.000412292)	(-0.000412292, +inf)
R5	(-inf, -0.000651859)	(-0.000651859, 0.000583821)	(0.000583821, 0.00176257)	(0.00176257, +inf)
R6	(-inf, 1.38138e-005)	(1.38138e-005, 0.00072438)	(0.00072438, 0.00077608)	(0.00077608, +inf)
R7	(-inf, 1.46802e-005)	(1.46802e-005, 6.811e-005)	(6.811e-005, 0.000260608)	(0.000260608, +inf)
R8	(-inf, 4.45802e-005)	(4.45802e-005, 6.1332e-005)	(6.1332e-005, 0.000253627)	(0.000253627, +inf)
R9	(-inf, -0.00032564)	(-0.00032564, 0.00129406)	(0.00129406, 0.00412803)	(0.00412803, +inf)
R10	(-inf, -0.00032564)	(-0.00032564, 0.00170486)	(0.00170486, 0.00355543)	(0.00355543, +inf)
R11	(-inf, 0.00308829)	(0.00308829, 0.00355161)	(0.00355161, 0.00430769)	(0.00430769, +inf)
R12	(-inf, -0.00169722)	(-0.00169722, 3.00724e-005)	(3.00724e-005, 0.000163319)	(0.000163319, +inf)
R13	(-inf, 0.00131974)	(0.00131974, 0.00255195)	(0.00255195, 0.00265304)	(0.00265304, +inf)
R14	(-inf, -0.000937993)	(-0.000937993, 0.000271131)	(0.000271131, 0.00183654)	(0.00183654, +inf)
R15	(-inf, 9.4422e-	(9.44223e-006,	(7.94461e-005,	(0.000744896,

	006)	7.94461e-005)	0.000744896)	+inf)
R16	(-inf, 3.03552e-005)	(3.03552e-005, 0.000232274)	(0.000232274, 0.000290301)	(0.000290301, +inf)
R17	(-inf, 6.49981e-006)	(6.49981e-006, 0.000152804)	(0.000152804, 0.000620383)	(0.000620383, +inf)
R18	(-inf, 6.4581e-006)	(6.45818e-006, 1.90195e-005)	(1.90195e-005, 0.000153841)	(0.000153841, +inf)
R19	(-inf, 5.4507e-005)	(5.45079e-005, 5.56532e-005)	(5.56532e-005, 0.000473715)	(0.000473715, +inf)
R20	(-inf, -0.000115164)	(-0.000115164, 5.14362e-005)	(5.14362e-005, 0.00121164)	(0.00121164, +inf)
R21	(-inf, -0.000148011)	(-0.000148011, 4.68013e-005)	(4.68013e-005, 0.00120903)	(0.00120903, +inf)
R22	(-inf, 1.16457e-005)	(1.16457e-005, 0.000203684)	(0.000203684, 0.00417805)	(0.00417805, +inf)
R23	(-inf, 7.0378e-007)	(7.0378e-007, 1.22549e-005)	(1.22549e-005, 0.000117764)	(0.000117764, +inf)
R24	(-inf, 1.10402e-005)	(1.10402e-005, 0.000426896)	(0.000426896, 0.00416212)	(0.00416212, +inf)
R25	(-inf, 1.2036e-005)	(1.2036e-005, 2.53061e-005)	(2.53061e-005, 0.000161758)	(0.000161758, +inf)

Inf. = Infinite.

Next, we construct the reducts. Because there are correlations between the explanatory variables, the program produces many reducts. Specifically, the program produced 18,241 reducts with and between 6 to 12 variables. Importantly, there are no core elements; there is no variable that is essential for the classification or is more relevant than any other. The frequency of appearance of each variable is shown in Table 3.

Table 3: Frequency of appearance of each variable in the reducts
The program produced 18,241 reducts with and between 6 to 12 variables

Variable	**R1**	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**
Frequency	23.78%	30.55%	42.46%	23.26%	26.31%	29.58%	24.89%
Variable	**R8**	**R9**	**R10**	**R11**	**R12**	**R13**	**R14**

Frequency	44.01%	32.34%	48.32%	33.71%	23.26%	38.31%	28.20%
Variable	**R15**	**R16**	**R17**	**R18**	**R19**	**R20**	**R21**
Frequency	31.96%	31.96%	24.75%	25,60%	24.33%	28.39%	30.72%
Variable	**R22**	**R23**	**R24**	**R25**			
Frequency	39.72%	37.99%	34.67%	29.52%			

We selected 3 of all possible reducts. The selection criteria are the following: first, the reduct should contain the smallest possible number of variables; second, the variables should present a high frequency of appearance in the different reducts; and finally, they should be formed by the smallest number of ratios for each category considered. The ratios belonging to each of the reducts are shown in Table 4.

Table 4: Selected reducts
The ratios belonging to each of the reducts

Reducts	Variables
1	{R3,R10,R13,R17,R22,R25}
2	{R3,R10,R13,R14,R17,R20,R25}
3	{R3,R11,R13,R14,R15,R23}.

The first 2 reducts have been chosen because they include the ratios R3, R10 and R13, the quotients that have high percentages of appearance and contain a ratio in each of the categories studied. The third reduct was chosen because it presents ratios from all of the categories. We limited our search to those reducts formed from a maximum of 7 ratios (one more than the number of categories considered).

With each of the 3 reducts, decision rules were generated using the lem2 procedure;[7] these are shown in Table 5. The following points can be noted regarding the items in the table. We can observe that the number of rules varies from one reduct to another; we find that 24, 28 and 22 rules are required to classify 100% of the observations correctly. The first reduct is the one that

[7] Chan *et al*., (1994).

Table 5: Selected reducts and corresponding decision rules
In all sub-tables below the decision rules are generated using the lem2 procedure.

Table 5.1: Reduct 1: Ratios R3, R10, R13, R17, R22, and R25

Rules of Classification	Correct Classification
(D1 = 0): Healthy	
1. (R3 = 2) & (R13 = 3) & (R17 = 1) & (R25 = 3)	9.43%
2. (R3 = 0) & (R10 = 3)	26.42%
3. (R10 = 2) & (R17 = 2) & (R22 = 1) & (R25 = 3)	3.77%
4. (R17 = 0)	33.96%
5. (R10 = 3) & (R17 = 2) & (R22 = 2)	5.66%
6. (R3 = 2) & (R13 = 1) & (R25 = 2)	11.32%
7. (R3 = 0) & (R17 = 1)	13.21%
8. (R10 = 1) & (R13 = 3) & (R22 = 3)	3.77%
9. (R3 = 0) & (R17 = 3)	5.66%
10. (R3 = 2) & (R13 = 0) & (R25 = 2)	5.66%
11. (R22 = 0)	37.74%
(D1 = 1): Bankrupt	
12. (R10 = 1) & (R22 = 2) & (R25 = 3)	49.06%
13. (R10 = 2) & (R17 = 2) & (R22 = 2) & (R25 = 3)	16.98%
14. (R3 = 2) & (R17 = 3) & (R22 = 2)	9.43%
15. (R13 = 2)	9.43%
16. (R3 = 3) & (R17 = 3)	24.53%
17. (R3 = 2) & (R13 = 3) & (R25 = 2)	5.66%
18. (R25 = 1)	3.77%
19. (R3 = 2) & (R13 = 1) & (R25 = 3)	15.09%
20. (R3 = 0) & (R17 = 2) & (R25 = 2)	1.89%
21. (R3 = 2) & (R17 = 2) & (R22 = 3)	3.77%
22. (R3 = 1) & (R25 = 3)	7.55%
23. (R3 = 2) & (R10 = 3) & (R13 = 3) & (R17 = 2) & (R22 = 1)	1.89%
24. (R3 = 3) & (R25 = 2)	3.77%

requires the fewest rules to identify the bankrupt companies. We can see also that the first reduct is the one that has the rules with the greatest power of classification. Thus, the rules 11 [(R22 = 0) => (D1 = 0)] and 12 [(R10 = 1) & (R22 = 2) & (R25 = 3) => (D1 = 1)] of the first reduct classify, respectively, 37.74% of the bankrupt companies and 49.06% of the healthy ones. No other individual rule of the second or third reducts reaches such high percentages.

The above 2 rules tell us, first, that if the value of the Napierian logarithm of the profitability ratio R22 (Pre-tax profits / Net Worth) is less than 1.16457e-005, then the company must be classified as bankrupt. Second, if the value of the Napierian logarithm of the ratio of indebtedness R10 (Financial Costs /Gross Profits) is between 0.00032564 and 0.00170486, that of the profitability ratio R22 (Pre-tax profits / Net Worth) is between 0.000203684 and 0.00417805 and that of the ratio R25 (Gross Profits / Total Assets) is greater than 0.000161758, the company must be classified as healthy.

Table 5.2: Reduct 2: Ratios R3, R10, R13, R14, R17, R20, and R25

Rules of Classification	Correct Classification
(D1 = 0): Healthy	
1. (R13 = 3) & (R14 = 1) & (R17 = 1) & (R25 = 3)	11.32%
2. (R10 = 3) & (R25 = 2)	22.64%
3. (R10 = 3) & (R14 = 2)	9.43%
4. (R10 = 0)	16.98%
5. (R3 = 2) & (R14 = 0) & (R17 = 2)	5.66%
6. (R10 = 2) & (R14 = 3)	1.89%
7. (R10 = 3) & (R13 = 0)	15.09%
8. (R3 = 0) & (R13 = 3)	26.42%
9. (R3 = 3) & (R13 = 0) & (R17 = 2)	1.89%
10. (R20 = 0)	32.08%
11. (R3 = 0) & (R17 = 3)	5.66%
12. (R14 = 1) & (R20 = 1) & (R25 = 2)	9.43%
13. (R13 = 0) & (R14 = 0)	16.98%
14. (R10 = 2) & (R13 = 3) & (R20 = 1)	3.77%

15. (R20 = 3)	1.89%
16. (R10 = 3) & (R20 = 2)	1.89%
(D1 = 1): Bankrupt	
17. (R10 = 1) & (R17 = 3) & (R20 = 2)	28.30%
18. (R10 = 2) & (R14 = 1) & (R17 = 2) & (R20 = 2)	15.09%
19. (R3 = 2) & (R10 = 1) & (R20 = 2)	20.01%
20. (R3 = 2) & (R13 = 1) & (R25 = 3)	15.09%
21. (R13 = 2)	9.43%
22. (R10 = 1) & (R14 = 2)	32.08%
23. (R3 = 1) & (R25 = 3)	7.55%
24. (R3 = 0) & (R13 = 1) & (R17 = 2)	3.77%
25. (R3 = 2) & (R13 = 3) & (R25 = 2)	5.66%
26. (R3 = 3) & (R13 = 1)	15.09%
27. (R14 = 2) & (R17 = 3)	26.42%
28. (R3 = 2) & (R10 = 3) & (R13 = 3) & (R14 = 1) & (R17 = 2) & (R20 = 1)	1.89%

Table 5.3: Reduct 3: Ratios R3, R11, R13, R14, R15, and R23

Rules of Classification	Correct Classification
(D1 = 0): Healthy	
1. (R3 = 0) & (R11 = 2)	26.42%
2. (R11 = 3)	26.42%
3. (R15 = 1) & (R23 = 3)	3.77%
4. (R3 = 2) & (R13 = 1) & (R14 = 0)	3.77%
5. (R23 = 0)	47.17%
6. (R13 = 0) & (R23 = 1)	3.77%
7. (R11 = 2) & (R13 = 3) & (R15 = 2) & (R23 = 2)	5.66%
8. (R15 = 3) & (R23 = 2)	7.55%
9. (R14 = 0) & (R15 = 3)	5.66%

10. (R3 = 2) & (R11 = 0) & (R14 = 3)	7.55%
11. (R13 = 1) & (R23 = 1)	1.89%
12. (R11 = 0) & (R13 = 1) & (R14 = 1) & (R23 = 2)	1.89%
(D1 = 1): Bankrupt	
13. (R13 = 1) & (R15 = 2) & (R23 = 3)	20.75%
14. (R3 = 2) & (R15 = 2) & (R23 = 3)	30.19%
15. (R3 = 2) & (R11 = 2) & (R15 = 1) & (R23 = 2)	7.55%
16. (R11 = 0) & (R14 = 2)	30.19%
17. (R3 = 3) & (R15 = 2)	30.19%
18. (R11 = 1) & (R23 = 2)	11.32%
19. (R13 = 2)	9.43%
17. (R3 = 3) & (R15 = 2)	30.19%
18. (R11 = 1) & (R23 = 2)	11.32%
19. (R13 = 2)	9.43%
20. (R3 = 1) & (R14 = 3)	1.89%
21. (R3 = 1) & (R15 = 2)	5.66%
22. (R3 = 2) & (R11 = 2) & (R13 = 3) & (R23 = 1)	1.89%

To verity these classification results, we performed a cross validation with 10 passes. We present the results for each reduct in Table 6.

The first and third reducts are clearly more robust than the second, with respect to percentages of correct classification. The third reduct, in addition to meeting the previously imposed requirements, is the one that presents fewest type I and II errors (13.42% and 11.18%, respectively, for the validation sample).

Given the above results, we are thus able to confirm that the rough set methodology leads to good results in the classification of healthy and bankrupt companies and indicates which variables are the most relevant of those considered. Thus, by applying these rules to new credit operations, a bank would be able to detect possible defaults.

Table 6: Percentages of correct classification

This table includes the results of a cross validation with 10 passes.

Reduct	Correct classification: bankrupt	Correct classification: healthy	Correct classification: total
R3,R10,R13,R17,R22,R25	83.98%	87.90%	86.00%
R3,R10,R13,R14,R17,R20,R25	66.67%	84.83%	76.27%
R3,R11,R13,R14,R15,R23	86.58%	88.82%	88.82%

B. Comparison with the Discriminant Analysis

As mentioned above, discriminant analysis was the first technique that was widely used. It still remains the most frequently utilized technique for measuring company insolvency. Therefore, we have compared the results of the rough set methodology with those produced by the discriminant analysis.

We used the same set of data to perform this comparison[8]. First, we utilized Box's M-statistic to test the equality of the variance-covariance matrices between the two groups. From the results, we can accept this hypothesis at a 5% level of significance. The discriminant function (obtained using the *ascending stepwise* method, with Snedecor's F criterion for entry between 0.06 and 0.09) and the statistics associated with the model are shown in Table 7.

Table 7: Discriminant functions and statistics of the models

This table reports information on the discriminant analysis. The discriminant function is obtained using the ascending stepwise method, with Snedecor's F criterion for entry between 0.06 and 0.09.

Discriminant function (discriminant canonical function with standardized coefficients)	Wilks' lambda	P-Value
Z = 0.554 R4 + 0.769 R6 – 0.337 R9	0.691	<0.0000

Base on the avove table, we can deduce that, according to the discriminant analysis, the liquidity and indebtedness ratios are the most relevant items for determining the possible

[8]The Napierian logarithms of the data were used, and their values were typified.

insolvency of a company. The liquidity ratio positively affects a company's probability of being classified as healthy. With respect to the ratios of indebtedness, their influence will depend on their definition. The ratio R6 (Net Worth / Total liabilities) has a positive influence, while the ratio R9 (Financial Costs / (Gross Profits + provision for amortization)) has a negative influence on the probability of the company in question being considered healthy.

Table 8 gives the results of the classification with the discriminant function applied to the original sample and validated using the cross procedure.

Table 8: Percentages of correct classification base on discriminant analysis

This table reports the results of the classification with the discriminant function. The original sample is used. The results are validated using the cross procedure.

	Correct classification: bankrupt	**Correct classification: healthy**	**Correct classification: total**
Original sample	71.7%	83.0%	77.4%
Validation sample	71.7%	83.0%	77.4%

From tables 6 and 8 together, it can be deduced that the rough set methodology is a more useful tool than discriminant analysis for the classification of the defaulting companies in the database considered for our empirical application. With the correct choice of reducts, only 6 of the 25 variables considered in the analysis are required to correctly classify 100% of the original observations. Furthermore, more validation samples are correctly classified, giving type I and II errors of 13.42% and 11.18%, respectively, compared to the errors of 28.3% and 17%, respectively, that are given by applying discriminant analysis and utilizing the *forward* method for the twenty-five variables.

5.3. Comparison with the Logit Analysis

Because logit analysis is a widely used technique for categorizing two groups, we have compared the results obtained using the rough set methodology with the results obtained using the logit model. The practical benefits of the logit methodology are that it does not require the restrictive assumptions of the discriminant analysis.

In the logit model, we applied a statistical forward stepwise selection procedure on the selected variables. The logit model function resulting from our sample of companies is shown in Table 9. The Wald test for each of the predictors is statistically significant. Additionally, the logit-likelihood test is statistically significant.

Table 9: Logit model results

This table includes the details of applying the logit model. A statistical forward stepwise selection procedure is employed.

	Coefficient (β)	Standard error	Wald	gl	Sig.	Exp(β)
R4	1.271	.602	4.462	1	.035	3.566
R6	4.671	1.851	6.370	1	.012	106.765
R8	-9.568	5.268	3.299	1	.069	.000
R21	12.630	5.589	5.106	1	.024	305642.138
Constant	.002	.492	.000	1	.996	1.002
Z = 1.271 R4 + 4.671R6 – 9.568R8 + 12.630R21 + 0.002						

Table 10 shows the results of the designation of healthy and failed companies. Based on tables 6 and 10, it can be deduced that the rough set methodology is more useful than the logit model over the sampl of companies considered.

Table 10: The result of the classification between healthy and failed companies (logit model).

	Forecast group in which the company falls		Total
	Failed	**Healthy**	
Failed	81,6%	18,4%	100%
Healthy	26,5%	73,5%	100%
	Overall correct percentage		77,6%

6. Conclusion

In this study we present an alternative methodology to the classic discriminant analysis and logit model for determining the variable(s) that serve to explain the failure of a company to meet its debt repayments as the first step in determining the probability of default (PD).

Basing our arguments on a sample of sound and bankrupt companies and on a set of 25 financial ratios that are potential explanatory factors for the defaults occurring in the sample, we have shown that the rough set methodology can be a valid alternative to discriminant analysis and to the logit model when there is a need to classify objects into two different classes. In addition to obtaining acceptable percentages of correct classification using rough sets, there is no need to assume any type of prior statistical behavior of the variables involved in the classification, unlike discriminant analysis, which requires normality in the distributions and equality in the variance-covariance matrices. Furthermore, the variables are included as they are presented, with no need for any transformation. Among the more significant advantages of this methodology is that it eliminates redundant information, and it expresses the dependencies between the variables considered and the results of the classification through decision rules for which the language is closer to that normally utilized by the experts.

Authors Information: Reyes Samaniego Medina is the corresponding author and is with Pablo de Olavide University, Department of Financial Economics and Accounting, Carretera de Utrera, Km. 1 (41013) Seville, Spain. Tel.: +34 954 34 98 45. E-mail address: rsammed@upo.es. M. José Vázquez Cueto, Departamento de Economía Aplicada III, Universidad de Sevilla (41018) Sevilla, Spain (España).

References

Altman, E.I. (1968). Financial ratios, discriminate analysis and the prediction of corporate bankruptcy. *Journal of Finance.*September 589-609.

Altman, E.I. (1993). Corporate financial distress and bankruptcy. John Wiley. New York.

Altman, E.I., Sabato, G. (2007). Modelling credit risk for smes: evidence from the U.S. market. *Abacus*, Vol. 43, 3, 332-357

Ahn, B.S., Cho, S.S., Kim, C.Y. (2000).The integrated methodology of rough set theory and artificial neural network for business failure prediction. *Expert Systems with Applications,* 18.

Basle Committee on Banking Supervision (2001).The internal ratings-based approach. *Consultive Document. Supporting Document to the New Basle Capital Accord, January.*

Basel Committee On Banking Supervision (1999). Credit risk modelling: current practices and applications. April.

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research* 4, 71-127

Cabedo,J., Reverte, J.A., Tirado, J.M. (2004). Riesgo de crédito y recursos propios mínimos en entidades financieras. *Revista Europea de Dirección y Economía de la Empresa*, 13, 2.

Chan. C., Grzymala-Busse, J. (1994).On the two local inductive algorithms: PRISM and LEM2. *Foundations of Computing and Decision Sciences*, 19, 185–204.

Comité De Supervisión Bancaria De Basilea (2004). Convergencia internacional de medidas y normas de capital. June.

Courtis, J.K. (1978). Modelling financial ratios: categories framework. *Journal of Business, Finance and Accounting*, 5 (4).

Dambolena, I.G., Khoury, S.J. (1980). Ratio stability and corporate failure. *The Journal of Finance*, 35.

Daubie, M.; Leveck, P.; Meskens, N. (2002). A comparison of the rough sets and recursive partitioning induction approaches: an application to commercial loans. *International Transactions in Operational Research* 9, 681-694.

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.

Dimitras, A.I., Slowinski, R., Susmaga, R, Zopounidis, C. (1998). Business failure prediction using rough sets. *European Journal of Operational Research,* 114.

Eisenbeis, R.A. (1977). Pitfalls in the application of discriminant analysis in business and economics. *Journal of Finance*, 32, 875-900.

Fernandez, J.E. (2005). Corporate credit risk modeling: quantitative rating system and probability of default estimation. SSRN: http://ssrn.com/abstract=722941

Komorowski, Z. (1999). Rough sets: a tutoriain rough-fuzzy hybridization - a new trend in decision makin. *S.K. Pal and A. Skowron, Eds.*, 3-98, Springer-Verlag Singapore Pte Ltd.

Laitinen, E.K. (1991). Financial ratios and different failure processes. *Journal of Business Finance and Accounting*,18/5.

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

Pawlak Z. (1982). Rough sets. Int. J. *Computer and Information Sci.*, 11,341-356

Pawlak,Z.; Slowinski,R.(1994). Decision analysis using rough sets. *International Transactions in Operational Research* 1, 107-114.

Rodriguez, M., Díaz, F. (2005). La teoría de los rough sets y la predicción del fracaso empresarial. Diseño de un modelo para pymes. *Revista de la Asociación Española de Contabilidad y Administración de Empresas, 74*.36-39

Segovia, M.J., Gil, J.A., Vilar, L., Heras, A.J. (2003). La metodología rough set frente al análisis discriminante en la predicción de insolvencia en empresas aseguradoras. *Anales del Instituto de Actuarios Españoles*, 9.

Trujillo, A., Martin, J.L. (2004). El rating y la fijación de precios en préstamos comerciales: aplicación mediante un modelo logit. *Revista Europea de Dirección y Economía de la Empresa*, 13. Nº. 2.

Trucharte, C., Marcelo, A. (2002). Un sistema de clasificación (rating) de acreditados. *Estabilidad Financiera*, 2.

Wilson, T. (1997). Portfolio credit risk. *Risk Magazine*, Sept., 111-117.

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

Zmijewski, M.E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Studies on Current Econometric Issues in Accounting Research.*