Rough sets and logistic regression analysis for loan payment

Bahadtin Ruzgar, Nursel Selver Ruzgar

Abstract**—**Risk classification is an important part of the financial processes. In small business loans, there is always a risk for nonpayment or non-refunding of loans though very detailed examinations are made about the company. In this study, behaviors that increase the risk in loans or causing non-refunding are tried to be determined by using the rough Set (RS) approach and logistic regression (LR). For this purpose, 121 regularly refunded and 121 irregularly refunded loans, drawn from a bank in 2006 year, were randomly selected and their attributes were examined in 2007. Examination is made in three sections for qualitative variables, for quantitative variables and for both qualitative and quantitative variables. As a result, RS model is applicable to a wide range of practical problems pertaining to loan payment prediction, but LR does not classify refund or non-refund of loan payment as good as RS, so LR can not be used for prediction. Moreover, the results show that the RS model is a promising alternative to the conventional methods for financial prediction. In fact, RS gives the attributes that affect the results negatively or positively with their measures which are used for predictions.

Keywords**—**Classification, loan payment, logistic regression, rough sets.

I. Introduction

The financial system plays a crucial role in economic development as responsible for the allocation of resources over time and among different alternatives of investment by pricing the postposition of consumption and pricing the risk [1]. Decompositions of balances on financial systems affect markets badly. Therefore, these may force sound banks to go to bankrupt, large or small business to go to failure, or loans payments to go to failure. In this paper, we propose an approach to predict small business loan payment failure based on genetic algorithm techniques of RS and LR [2].

Business failure prediction is a scientific field which many academic and professional people have been working for, at least, the three last decades. The high individual and social costs encountered in corporate bankruptcies make this decision problem very important to parties such as auditors, management, government policy makers, and investors. Also, financial organizations, such as banks, credit institutions, clients, etc., need these predictions for firms in which they have an interest (of any kind) [3]. Many financial decisions involve the classification of a set of observations (firms, stocks) into one or several groups of categories, what leads many researches to apply operational research methods to management problems. Methods such as discriminant analysis (DA), logit analysis, RS, recursive partitioning algorithm, and several others have been used in the past for the prediction of business failure. Although some of these methods led to models with a satisfactory ability to discriminate between healthy and potentially risky (candidates for bankruptcy) businesses, they suffer from some limitations, often due to the unrealistic assumption of statistical hypotheses or due to a confusing language of communication with the decision makers [4]. There exists an extensive literature devoted to the study of classification problems in the financial field, i.e. the work by Chen and Yeh [5] dealing with financial management or the works by Dimitras et al. [6] and Tam [7] in business failure [2]. Over the last 35 years, academic researchers from all over the world have dedicated their time to the search for the best corporate failure prediction model that can classify companies according to their financial health or failure risk [6], [8], [9]. Most papers deal with insurance audits, purchase intentions, purchase channel studies, methodologies for investigating customer purchasing intentions, and customer satisfaction [10]. Many research papers have quantified the problem in order to simplify the parameters, such as social parameters, and use statistical tools to analyze data. This approach, however, is only good for crisp types of data sets and certain data values. If the value of data is continuous or uncertain we must apply fuzzy theory [11]. In this paper, two methods, RS approach and LR are used to provide a set of rules able to discriminate between healthy and failing firms in order to predict business loan payments [3].

The Rough Set Theory (RST) has been successfully applied in many real-life problems in medicine, pharmacology, engineering, banking, financial and market analysis and others. The RS methodology has found many real-life applications [12]. RST overlaps with many other theories, such as fuzzy sets, evidence theory, and statistics. Nevertheless, it can be viewed in its own right as an independent, complementary, and no competing discipline [13]. The most important problems that can be solved by RST are: finding description of sets of objects in terms of attribute values, checking dependencies (full or partial) between attributes, reducing attributes, analyzing the significance of attributes, and generating decision rules [14], [15]. The application of the RS approach in

Manuscript received December 14, 2007: Revised version received March 8, 2008.

N. S. Ruzgar, is with Vocational School of Social Science, Marmara University, Istanbul, Turkey, (phone: 90-212-517 20 16; fax: 90-216-517 20 12; e-mail: nruzgar@marmara.edu.tr).

B. Ruzgar is with the Actuaries Department, Banking and Insurance School, Istanbul, Marmara University, Turkey, (phone: 90-216-414 99 89; fax: 90-216-347 50 86; e-mail: bruzgar@marmara.edu.tr).

business failure prediction was investigated by Slowinski et al. [16] and Dimitras et al. [3]. In the past a large number of methods have been proposed to predict business failure; however, the special characteristics of the insurance sector have made most of them unfeasible. Up to date, just a few of them have been applied to this sector. Most approaches applied to insurance companies are statistical methods such as DA or logit analysis [17], [18] and use financial ratios as explicative variables.

While the DA classifies the examined area in a category, it requires the independent variable data necessary for that area to remain under the normality assumption. The LR analysis classifies the examined area in two categories (it may be also multiple) and it doesn't require the independent variable data necessary for that area to remain under the normality assumption [19]. Besides, it allows the independent variables to be continuous, discrete and categorical variables. LR is one of the sophisticated statistical methods used by the banking industry to select credit rating variables. Extending the method to insurance risk classification seems natural. But Insurance risks are usually classified in a larger number of classes than good and bad, as is usually the case in credit rating. Ruzgar used LR analysis to classify life and non-life insurance companies [19] while Williamson used binary LR to calculate the retirement age of people depending on age, sex, economical and social statuses [20]. While Braniv used logistic analysis to examine and estimate the classification of the financial and non-financial values of their final bankruptcy resolutions [21], Temple with multi LR analysis examined 2001 data of national health researches of the Australian households passing rate to private health insurance [22]. Also, the other studies examined and classified the disability risk and disability risk insurances according to workability limits, non-workability situations and the need to get health aid of people who retired for reason of a physical disability [23].

The rest of the paper is structured as follows: Section 2 gives determinants of loan problems and Section 3 introduces main concepts of RST and LR briefly and explains methodologies. Section 4 describes the empirical models and the main results we obtained. Finally, Section 5 highlights the main conclusions that can be outlined from the analysis.

II. THE DETERMINANTS OF LOAN PROBLEMS

Loaning is a significant part of commercial transactions. The loan transactions are important for the loan drawing companies as well as for the loan issuing companies. While the small business companies try to solve their problems by drawing the credit they need, the financial institutions issuing the loan try to secure back payment of the loan (try to decrease the risk to minimum). In small business loans, there is always a risk for nonpayment or non-refunding of loans though very detailed examinations are made about the company. The important think here is to minimize the risk and not cause bad loans. There are many reasons that increase non-refunding risk of small business loans including financial straits faced by the company, low profit volumes, insufficient assets, sales not meeting the expected increase, changing of market, fluctuations in economical indicators, changing of competition conditions in related activity field. The loan issuing financial institutions make detailed examinations about companies they issue loan to secure refunding of loan. Though there are various differences between the financial institutions, they generally examine the company that demands small business loan under 7 headings. These include information about the company owners, information about guarantors, information about managers, information about structure of company, information about assets of company, information about balance sheet of company for last two years (this is also valid for the guarantors if any) and information about utilization of loan. In this study, the information about the company that demands small business loan is considered in 7 headings which consist 99 articles totally. In the first part, information from 21 articles consist of information about the owner and partners (if any) of company is required. In the second part, information from 14 articles consist of information about the guarantor (if any, about the second and third guarantors) is required. In the third part, the information about the management and activities of the company is required in 5 articles. In the forth part, information about nature and address of the company is required in 5 articles. In the fifth part, general information about the company is questioned in 21 articles. In the sixth part, information about assets, bank investment instruments, financial registers and balance sheet of the company are collected in 9 articles and required together with the related documents. In the seventh part, 24 articles question information about the loan demand and loan utilization. Though the loan issuing financial institutions examine in detail the company that demands loan and try to secure refunding of the loan (by creating a mortgage on assets), the risk on nonpayment or non-refunding of commercial loan is high. However, whatever the conditions of a loan are, it has vital importance for the loan issuing and drawing companies and it is obligatory.

III. THE METHODOLOGY

A. Main Concepts of RST

RST has attracted worldwide attention of many researchers and practitioners, who have contributed essentially to its development and applications. Despite this, RST may be considered as an independent discipline in its own right [24]. The RST proposed by Pawlak provides an effective tool for extracting knowledge from data tables. To represent and reason about the extracted knowledge, a decision logic (DL) was proposed in Pawlak [12], [25]. From a philosophical point of view, RST is a new approach to vagueness and uncertainty, and from a practical point of view, it is a new method of data analysis [13]. Therefore, RS theory is related in some aspects to other tools that deal with uncertainty. However, RS approach is somewhat different to statistical probability, which deals with random events in nature or Fuzzy Set Theory (FST), which deals with objects that may not belong only to one category but may belong to more than one category by differing degrees [1]. RST can deal with inexact, uncertain, and vague datasets [26]. Both FST and RST are used with the indiscernibility relation and perceptible knowledge. The major difference between them is that RST does not need a membership function; thus, it can avoid pre-assumption and one-sided information analysis. In RST, the data is grouped into classes called *elementary sets*. Feature/attribute selection is crucial in data processing that consists of relevant (or maybe irrelevant) object patterns, but it may be redundant in data pattern recognition [26]. The RS approach works by discovering dependencies between attributes in an information table, and reducing the set of attributes by removing those that are not essential to characterize knowledge. A *reduct* is defined as the minimal subset of attributes which provides the same quality of classification as the set of all attributes [2]. If the information table has more than one reduct, the intersection of all of them is called the core and is the collection of the most relevant attributes in the table [17].

The RS method is useful for exploring data patterns because of its ability to search through a multi-dimensional data space and determine the relative importance of each attribute with respect to its output [26]. An information table which contains condition and decision attributes is referred as a decision table. A decision table specifies what decisions (actions) should be undertaken when some conditions are satisfied. So a reduced information table may provide decision rules of the form "*if* conditions *then* decisions". These rules can be *deterministic* when the rules describe the decisions to be made when some conditions are satisfied and *non-deterministic* when the decisions are not uniquely determined by the conditions so they can lead to several possible decisions if their conditions are satisfied. The number of objects that satisfy the condition part of the rule is called the *strength* of the rule and is a useful concept to assign objects to the strongest decision class when rules are non-deterministic [17]. The rules derived from a decision table do not usually need to be interpreted by an expert as they are easily understandable by the user or decision maker. The most important result in this approach is the generation of decision rules because they can be used to assign new objects to a decision class by matching the condition part of one of the decision rule to the description of the object. So rules can be used for decision support [17].

The RST has the following important advantages: it provides efficient algorithms for finding hidden patterns in data; it finds reduced sets of data (data reduction); it evaluates significance of data; it generates minimal sets of decision rules from data; it is easy to understand; it offers straightforward interpretation of results; it can be used in both qualitative and quantitative data analysis; and it identifies relationships that would not be found using statistical methods [13].

B. Main Concept of LR

In cases when categorical results such as successful-unsuccessful, ill-not ill, good-fair-bad are obtained especially as a result of evaluation of data, the logistic regression is a rather suitable statistical method. The logistic regression establishes very useful functional relation with independent variables (it may be cross sectional, continuous and categorical) in case the dependent variable is a categorical variable depended on two situations (it may be more). In this way, it gives opportunity for the categorical classification by using the regression analysis structure [19].

When the statistics related to the logistic regression are examined it is found that Cox-Snell R^2 (CS-R^2) and Nagelkerke R^2 (Nag R^2) values that show the degree of relation between the dependent and independent variables in the logistic regression models are higher and -2LogL (-2log likelihood=-2LL) statistic is lower. When the model exactly represents the data, the likelihood is 1 and -2LL statistics is zero. For this reason, the lower -2LL statistic always shows a better model [18, 19]. When the statistics related to testing of meaningfulness of model are examined, Chi-square (χ^2) statistic, -2LL statistic and Block Chi-square (Bχ^2) statistic should be considered. The χ^2 statistic tests the logistic regression model in general. The χ^2 statistic firstly shows the fault only when there is a constant term in the model and then it determines whether all the logistic coefficients except the constant term are equal to zero. The χ^2 statistic conforms to χ^2 distribution with degree of freedom equals to difference between the parameter number of examined model and parameters of model with constant term [27]. In logistic regression, -2LL statistic shows the fault of model when an independent variable is added to model. For this reason, it is the measure of unexplained variance in a dependent variable and the non-meaningful statistic is a desirable situation. The Bχ^2 statistic shows the changing in the χ^2 statistic when a block variable is added to model [28]. The mostly used R^2 statistics of regression analysis for the logistic regression are CS-R^2 statistic and Nag R^2 statistic [29], [30]. CS-R^2 statistic may have a value higher than zero. Its value below 1 strengthens the interpretation of statistic. Nag R^2 statistic was developed to ensure CS-R^2 statistic to have values between 0 and 1. The statistic closer to 1 means the relation is high.

Other common statistics in logistic regression are odds ratios (EXP(B)) and Hosmer-Lemeshow goodness-of-fit statistic [31]. The odds ratio (EXP(B)), for a given independent variable represents the factor by which the odds (event) change for a one-unit change in the independent variable. If B has positive value, odds ratio will increase, if B has negative value, the odds ratio will decrease, and if B is zero odds ratio will not change. An odds ratio of 1 indicates that the condition or event under study is equally likely in both groups. An odds ratio greater than 1 indicates that the condition or event is more likely in the first group. And an odds ratio less than 1 indicates that the condition or event is less likely in the first group. As the odds of the first group approaches zero, the odds ratio approaches zero. As the odds of the second group approaches zero, the odds ratio approaches positive infinity [31]. Moreover, Hosmer-Lemeshow goodness-of-fit statistic is more robust than the traditional goodness-of-fit statistic used in logistic regression, particularly for models with continuous covariates and studies with small sample sizes. It is based on grouping cases into deciles of risk and comparing the observed probability with the expected probability within each decile. Goodness-of-fit statistics help you to determine whether the model adequately describes the data. The Hosmer-Lemshow statistic evaluates the goodness-of-fit by creating 10 ordered groups of subjects and then compares the number actually in the each group (observed) to the number predicted by the logistic regression model (predicted). Thus, the test statistic is a chi-square statistic with a desirable outcome of non-significance, indicating that the model prediction does not significantly differ from the observed. The Hosmer-Lemeshow statistic indicates a poor fit if the significance value is less than 0.05 [32].

In addition to all above statistics, A Wald test is used to test the statistical significance of each coefficient (β) in the model. A Wald test calculates a *Z* statistic. This z value is then squared, yielding a Wald statistic with a chi-square distribution [33]. However, several authors have identified problems with the use of the Wald statistic. Menard [33] warns that for large coefficients, standard error is inflated, lowering the Wald statistic (chi-square) value. Agresti [29] states that the likelihood-ratio test is more reliable for small sample sizes than the Wald test. There are a few other things to note about the output. The first is that although we have only one predictor variable, the test for the odds ratio does not match with the overall test of the model. This is because the test of the coefficient is a Wald chi-square test, while the test of the overall model is a likelihood ratio chi-square test. While these two types of chi-square tests are asymptotically equivalent, in small samples they can differ [29]. Also, we have the unfortunate situation in which the results of the two tests give different conclusions. This does not happen very often. In a situation like this, it is difficult to know what to conclude. One might consider the power, or one might decide if an odds ratio of this magnitude is important from a clinical or practical standpoint [30].

IV. EMPIRICAL RESULTS

As we have previously mentioned, RS approach is especially well suited to classification problems. One of these problems is a multiattribute classification problem which consists of assignment of each object, described by values of attributes, to a predefined class or category [10].

The information given by the companies that demand business loan can be collected under two groups. In the first group, there is linguistic information about the company that demand loan and its guarantors. In this group, 81 variables can be defined. In the second group, there is information about the financial structure of the company that demands loan. In this group, 18 variables can be defined. Some of the business loans drawn from a bank in 2006 year were randomly selected, 121 regularly refunded and 121 irregularly refunded loans were examined in 2007 and 242*99 information table was created. However, not all the questioning articles consist of 99 attributes contain information suitable to be used as variable. Since 45 of 99 attributes have lack of information, only 54 attributes are taken into consideration. 54 attributes that can be defined as variable are divided into two sections.

Table I. Definition of attributes

Variable: Definition
A1: The activity period of company is more than 2 years
A2: The sector experience of the dominant partner is 4 years or more
A3: Place of company belongs to company or partners
A4: The company operates at the current address for more than 2 years
A5: The dominant partner of company is older than 25 years
A6: The place of dominant partner belongs to him
A7: There are immovable assets owned by the dominant partner and company
A8: There is administrative or legal prosecution against to dominant partner
A9: A protested cheque or bond of company or dominant partner has been put to execution process within last 2 years
A10: An execution process has been filed against to company or dominant partner within last 2 years for their loans
A11: Guarantor of partner
A12: 3rd person security
A13: There have been problems in related to loans they drawn previously
A14: There is mortgage on assets of company
A15: There are pledged vehicles of company
A16: The name and surname of the 1st guarantor exist
A17: The name of the first guarantor and his partner company exist
A18: The name and surname of the second guarantor exist
A19: The name of the second guarantor and his partner company exist
A20: The guarantors have assets
A21: There is administrative or legal prosecution against to guarantors
A22: The protested cheque, bond of guarantors, company or dominant partner has been put to execution process within last 2 years
A23: An execution process has been filed against to, guarantors, company or dominant partner within last 2 years for their loans

Table II. Initial set of attributes

Variable	0	1	2
A1		Yes (241)	No (1)
A2	Unknown (5)	Yes (0)	No (237)
A3	Unknown (179)	Yes (0)	No (63)
A4	Unknown (10)	Yes (232)	No (0)
A5	Unknown (1)	Yes (241)	No (0)
A6	Unknown (93)	Yes (149)	No (0)
A7	Unknown (88)	Yes (154)	No (0)
A8		Yes (1)	No (241)
A9		Yes (1)	No (241)
A10		Yes (0)	No (242)
A11		Yes (126)	No (116)
A12		Yes (131)	No (111)
A13		Yes (8)	No (234)
A14		Yes (2)	No (240)
A15		Yes (5)	No (237)
A16	Non (15)	Okay (227)	
A17	Non (158)	Okay (84)	
A18	Non (183)	Okay (59)	
A19	Non (217)	Okay (25)	
A20		Yes (217)	No (25)
A21		Yes (0)	No (242)
A22		Yes (1)	No (241)
A23		Yes (0)	No (242)

36 attributes consist of linguistic elements in the first section are selected in order to be used in examination. 13 of them are put out of examination as they require many data and remaining 23 attributes are given in Table II with their recoded values, 0, 1, and 2. The definitions of 23 attributes are shown in Table I. Moreover, Table II shows information about the attribute and related frequency in parenthesis.

In the second group, 18 attributes containing the balance sheet information for last 2 years of company (guarantors) are examined (information that contain 18 attributes for the 1st guarantor and 18 attributes for the 2nd guarantor are put out of examination). However, 6 attributes are put out of examination as they do not contain sufficient information about the balance sheet of company and 12 attributes are used. There is not missing balance sheet information consists 12 attributes about the companies that use loans. The definitions of these continuous variables are given in Table III.

Table III. Definitions of attributes

Variable: Definition
B1: 2005 year net sales
B2: 2004 year net sales
B3: 2005 year activity profit
B4: 2004 year activity profit
B5: 2005 year revolving assets
B6: 2004 year revolving assets total
B7: 2005 year short-term loans total
B8: 2004 year short-term loans total
B9: 2005 year short-term bank loans
B10: 2004 year short-term bank loans
B11: 2005 year equity capitals total
B12: 2004 year equity capitals total

This paper examines the effects of loan payments refunds and nonrefunds in three sections by using both RS approach and LR analysis. In the first section, 23 attributes consisting linguistic elements (qualitative variables), in the second section, 12 attributes consisting continuous variables (quantitative variables) and in the third section, all 35 (=23+12) attributes (both qualitative and quantitative variables) are examined.

By means of definition of attributes in Table I and Table III, 36 variables- total 35 condition variables and one decision variable showing whether the companies have refunded the loans- form the study. As a result of this data analysis, 242*36 information table is prepared in order to be used for RS approach. The first results obtained from RS analysis of the actual information table were; the approximation of the decision classes and their quality of classification were equal to one.

Class	# objects	L-App.	U-App.	Acc- App.	App- Qua.
1 (Regular Payment)	121	121	121	1.0	1.0
0 (Irregular Payment)	121	121	121	1.0	1.0

These results obtained show that the firms are very well discriminated among them. (Consequently, the boundary region is empty for the two decision classes) 242*36 information table consists of altitude values prepared for the study is evaluated with the software ROSETTA [34]. The lacking cells in this information table are completed with “Complete→Mean/Mode Fill” and “1” code value for 121 companies that regularly pay their loans and “0”code value for companies that irregularly pay their loans are used for the decision variable. Evaluation of the information table is made in 3 sections.

In the first section, the effectiveness of 23 condition variables (A1, A2, ..., A23) consist of information about the company and guarantors in decision variable and its success in classification are examined. For this purpose, genetic algorithms are used for reduces (Reduce→Genetic Algorithm →Object Related→Okay) and as a result of this process, 722 feasible situations have appeared. As to examine all feasible situations will take time and may cause ignoring of main necessary subjects, basic filtering (reduct) is used. In basic filtering, de RHS accuracy [0, 0.75], RHS (right hand side) coverage [0, 0.075] and LHS (left hand side) length [8, 1000] ranges are used for reduction. By eliminating alternative situations that appear other than these ranges, Table 4 is established with RS approach values consist of 11 rules. As a result of the basic filtering, it is found that A1, A2, A5, A8, A9, A10, A13, A21, A22 and A23 variables are insignificant. The most marked behavior in Table 5 is changing of altitude value of A6, A11, A17, A18 and A20 variables and changing of altitude value of decision variable. When the definitions of variables effecting the decision are examined in general, it is determined that the loan drawing companies are lack of necessary guarantees. In other words, in Table IV, the basic reason for non-refunding of loans is the loans without guarantee given by guarantors.

Table IV. The 11 rules algorithm

Conditions																	
A6	A7	A11	A12	A14	A15	A16	A17	A18	A19	A20	Decisions	Strength	RHS Accuracy	LHS Coverage	RHS Coverage	RHS Stability	LHS Length
		1	2				1		1		1	10	1.0	0.041	0.083	1.0	5
		1	2				1	1			1	11	1.0	0.045	0.091	1.0	5
	1	1					1	1		1	1	10	1.0	0.041	0.083	1.0	6
1	1	1					1	1		1	1	10	1.0	0.041	0.083	1.0	7
	1	1					1	1		1	1	10	1.0	0.041	0.083	1.0	7
1		2				1	0	0		2	0	12	1.0	0.050	0.099	1.0	7
1		2	1				0	0		2	0	11	1.0	0.045	0.091	1.0	7
1			1		2		0	0		2	0	13	1.0	0.054	0.107	1.0	7
1		2		2			0	0		2	0	14	1.0	0.058	0.116	1.0	7
	1	2		2			0	0		2	0	12	1.0	0.050	0.099	1.0	7
	1	2				1	0	0		2	0	10	1.0	0.041	0.083	1.0	7

When LR is applied for the 23 attributes that defined in Table I and Table II, -2LL, Cox & Snell R Square and Nagelkerke R Square are found 309.810, 0.101 and 0.134, respectively. According to these values, high value of -2LL statistics shows that model does not represent the data totally. Moreover, low values of Cox & Snell R Square and Nagelkerke R Square statistics show that explanation power of independent variable of dependent variable is very low. The recommended test for overall fit of a logistic regression model is the Hosmer and Lemeshow test, also called the Chi-square test. It is considered more robust than the traditional chi-square test, particularly if continuous covariates are in the model or sample size is small. A finding of non-significance corresponds to the researcher concluding the model adequately fits the data. When Hosmer -Lemeshow test is applied, Chi-square is found 5.342 with the degree of freedom 8 and significance value 0.720. In fact, the Hosmer-Lemeshow statistic indicates a poor fit if the significance value is less than 0.05, but, here, the model adequately fits the data because of p=0.720.

Table V. LR classification for 23 attributes.

	Observed		Predicted		
			result		Percentage Correct
			0	1	0
Step 1	result	0	79	42	65.3
		1	47	74	61.2
	Overall Percentage				63.2

*The cut value is 0.500

Table V shows overall percentage of LR classification (63.2%) with correct classification numbers. However, Table 6 shows coefficients of variables, standard errors of these coefficients, Wald statistics, significance levels, odds ratio, Exp(B), and 95% of confidence intervals for Exp(B) in different columns. The odds ratio for a given independent variable represents the factor by which the odds (event) change for a one-unit change in the independent variable. If B has positive value, odds ratio will increase, if B has negative value, the odds ratio will decrease, and if B is zero odds ratio will not change.

Table VI. LR equation for quantitative variables.

	B	S.E.	Wald	df	Sig.	Exp(B)	95.0 % C.I.for EXP(B) Lower	Upper
Step 1*a1(1)	0.377	1.772	0.045	1	0.831	1.458	0.045	47.013
a2(1)	-0.570	0.863	0.436	1	0.509	0.566	0.104	3.068
a3(1)	-0.053	0.324	0.026	1	0.871	0.949	0.503	1.790
a4(1)	0.166	0.788	0.045	1	0.833	1.181	0.252	5.528
a5(1)	-0.781	1.898	0.169	1	0.681	0.458	0.011	18.914
a6(1)	0.045	0.350	0.017	1	0.897	1.046	0.527	2.077
a7(1)	-0.052	0.356	0.021	1	0.884	0.949	0.472	1.908
a8(1)	-1.019	1.282	0.632	1	0.427	0.361	0.029	4.452
a9(1)	-0.539	1.043	0.267	1	0.605	0.583	0.075	4.510
a11(1)	0.652	0.351	3.453	1	0.063	1.919	0.965	3.814
a12(1)	-0.505	0.371	1.854	1	0.173	0.603	0.292	1.249
a13(1)	0.268	0.817	0.108	1	0.743	1.308	0.263	6.489
a14(1)	-0.680	1.281	0.282	1	0.595	0.506	0.041	6.236
a15(1)	-0.692	1.265	0.299	1	0.584	0.501	0.042	5.973
a16(1)	-0.290	0.610	0.226	1	0.634	0.748	0.226	2.472
a17(1)	-0.559	0.351	2.543	1	0.111	0.572	0.288	1.137
a18(1)	-0.601	0.396	2.304	1	0.129	0.548	0.252	1.192
a19(1)	0.671	0.583	1.324	1	0.250	1.956	0.624	6.131
a20(1)	-0.315	0.306	1.055	1	0.304	0.730	0.401	1.331
a22(1)	-1.359	1.223	1.235	1	0.266	0.257	0.023	2.823
Constant	0.036	1.839	0.000	1	0.984	1.037		

*Variable(s) entered on step 1: a1, a2, a3, a4, a5, a6, a7, a8, a9, a11, a12, a13, a14, a15, a16, a17, a18, a19, a20, a22.

In Table VI, it is seen that odds ratio for a1, a4, a11, a13 and a19 will increase, odds ratio for a2, a3, a5, a7, a8, a9, a12, a14, a15, a16, a17, a18, a20 and a22 will decrease. Since all results of the variables a10, a21 and a23 are same, they are removed from the evaluation. With 0.05 significance levels, the significance tests of LR coefficients are given with Wald statistics related to Chi-square distribution. According to Wald statistics in Table VI, all LR coefficients are not significant, but for the variables a1, a4, a6, a11, a13 and a19, it can be said that they are effective and preferable to other variables for non-refunding or refunding of loan payments because their odds ratios are greater than 1. Thus, LR model of qualitative variables (23 attributes) is not a good model in categorizing the loans whether refunded or no refunded.

In the second section, 12 continuous variables are examined with RS and LR respectively. The first analysis we have made was to recode the ratios (continuous variables) into qualitative terms (low, medium, high and very high) with corresponding numeric values such us 1, 2, 3 and 4. This recoding has been made dividing the original domain into subintervals. This recoding is not imposed by the RS theory but it is very useful in order to draw general conclusions from the ratios in terms of dependencies, reducts and decision rules. 12 attributes that contain balance sheet information about the companies are classified as (0, 0.25]-low, (0.25, 0.50]-medium, (0.50, 0.75]-high and (0.75, ∞)-very high. These classified attributes are given 1, 2, 3 and 4 recoded values and categorical variables are appointed. In Table III, definitions of 12 attributes are presented and in Table VII, list of subintervals (quartiles) are given.

Table VII. List of subintervals (quartiles)

Variable	1st (Low-1)	2nd (Medium-2)	3rd (High-3)	4th (Very high-4)
B1	(0,54800]	(54800,173683]	(173683,453649]	(453649,∞)
B2	(0,29041]	(29041,109989]	(109989,298118]	(298118,∞)
B3	(-∞,3199]	(3199,5999]	(5999,14833]	(14833,∞)
B4	(-∞,1311]	(1311,4299]	(4299,12269]	(12269,∞)
B5	(0,12670]	(12670,40329]	(40329,163472]	(163472,∞)
B6	(0,6309]	(6309,15321]	(15321,109280]	(109280,∞)
B7	(0,1800]	(1800,8680]	(8680,97120]	(97120,∞)
B8	(0,107]	(107,252]	(252,88904]	(88904,∞)
B9	(0,8710]	(8710,22690]	(22690,49418]	(49418,∞)
B10	(0,2420]	(2420,12560]	(12560,42981]	(42981,∞)
B11	(0,5620]	(5620,21081]	(21081,71391]	(71391,∞)
B12	(0,2050]	(2050,7974]	(7974,36595]	(36595,∞)

The information table consists of balance sheet information of the company that demands loan and valued with Table VII are evaluated by means of the software ROSETTA [34]. As a result of this reduct made with genetic algorithms, 3844 feasible situations have appeared (Reduce→Genetic Algorithm→Object Related →Okay). However, it is not possible to examine all these feasible situations. For this reason, basic filtering is applied and RHS accuracy [0, 0.75], RHS coverage [0, 0.075] and LHS length [6, 1000] ranges are used fort he basic filtering. Feasible situations other than these ranges are eliminated and Table VIII that consists of 22 rules is found. Table VIII shows the values found for RS approach.

When Table VIII is examined, it is found that B1, B2, B3 and B4 attributes are effective in classification of the decision. The basic reason of non-refunding of the loan is the low sales and profit of company within last two years and less economical earning.

When LR is applied for the 12 attributes that defined in Table 3, -2LL, Cox & Snell R Square and Nagelkerke R Square are found 291.204, 0.167 and 0.2230, respectively. According to these values, high value of -2LL statistics shows that model does not represent the data totally. Moreover, low values of Cox & Snell R Square and Nagelkerke R Square statistics show that explanation power of independent variable of dependent variable is very low. When Hosmer- Lemeshow test is applied, Chi-square is found 18.273 with the degree of freedom 8 and significance value 0.019. Goodness-of-fit statistics help you to determine whether the model adequately describes the data and it is known that he Hosmer-Lemeshow statistic indicates a poor fit if the significance value is less than 0.05, so, this model does not adequately fits the data because of p=0.019.

Table VIII. The 22 rules algorithm

Rule #	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	Decisions	Strength	RHS Accuracy	LHS Coverage	RHS Coverage	RHS Stability	LHS Length
	Conditions																		
1				2	3		2						1	11	1.0	0.045	0.091	1.0	3
2		2									3	2	1	10	1.0	0.041	0.083	1.0	3
3				2		2					3		1	11	1.0	0.045	0.091	1.0	3
4							2				3	2	1	10	1.0	0.041	0.083	1.0	3
5				2			2				3		1	10	1.0	0.041	0.083	1.0	3
6				2				2			3		1	10	1.0	0.041	0.083	1.0	3
7				2	3			2					1	10	1.0	0.041	0.083	1.0	3
8				2	3	2							1	10	1.0	0.041	0.083	1.0	3
9		2			3							2	1	11	1.0	0.045	0.091	1.0	3
10				2	3							2	1	12	1.0	0.050	0.099	1.0	3
11		2				2					3		1	12	1.0	0.050	0.099	1.0	3
12					3		2					2	1	11	1.0	0.045	0.091	1.0	3
13				2							3	2	1	10	1.0	0.041	0.083	1.0	3
14					3	2	2						1	11	1.0	0.045	0.091	1.0	3
15					3	2			3		3		1	12	1.0	0.050	0.099	1.0	4
16	3				3		2		3				1	10	1.0	0.041	0.083	1.0	4
17						2		2	3		3	2	1	13	1.0	0.054	0.107	1.0	5
18						2			3		3	2	1	13	1.0	0.054	0.107	1.0	4
19		1	1	1								2	0	11	1.0	0.045	0.091	1.0	4
20	1		1	1								2	0	12	1.0	0.050	0.099	1.0	4
21		1	2								2		0	10	1.0	0.041	0.083	1.0	3
22											1		0	22	1.0	0.091	0.182	1.0	1

Table IX. LR classification for 12 attributes.

Observed			Predicted		
			result		Percentage Correct
			0	1	0
Step 1	result	0	81	40	66.9
		1	44	77	63.6
	Overall Percentage				65.3

*The cut value is 0.500

Table IX shows overall percentage of LR classification (65.3%) with correct classification numbers. However, Table X shows coefficients of variables, standard errors of these coefficients, Wald statistics, significance levels, odds ratio, Exp(B), and 95% of confidence intervals for Exp(B) in different columns.

In Table X, it is seen that odds ratio for b1, b4, b5, b8, b10, b11 and b12 will increase, odds ratio for b2, b3, b6, b7 and b9 will decrease. According to Wald statistics with the significance level 0.05 in Table 10, all LR coefficients, except b3, b4 and b11 which are shaded, are not significant, but the for variables b1, b4, b5, b8, b10, b11 and b12, it can be said that they are effective and preferable to other variables for non-refunding or refunding of loans because their odds ratios are greater than 1. Thus, LR model of continuous variables of balance sheet (12 attributes) is not a good model in categorizing the loan payments whether refunded or no refunded.

Table X. LR equation for quantitative variables.

	B	S.E.	Wald	df	Sig.	Exp(B)	95 % C.I. for EXP(B) Lower	Upper
Step 1* b1	0.316	0.292	1.178	1	0.278	1.372	0.775	2.430
b2	-0.107	0.260	0.169	1	0.681	0.898	0.539	1.496
b3	-0.385	0.177	4.737	1	0.030	0.680	0.481	0.962
b4	0.287	0.167	2.955	1	0.056	1.332	0.961	1.846
b5	0.525	0.347	2.283	1	0.131	1.690	0.856	3.340
b6	-0.157	0.371	0.179	1	0.672	0.855	0.413	1.768
b7	-0.407	0.364	1.255	1	0.263	0.665	0.326	1.357
b8	0.366	0.364	1.013	1	0.314	1.442	0.707	2.941
b9	-0.830	0.593	1.957	1	0.162	0.436	0.136	1.395
b10	0.120	0.636	0.036	1	0.850	1.127	0.324	3.918
b11	0.603	0.262	5.303	1	0.021	1.828	1.094	3.054
b12	0.092	0.236	0.153	1	0.695	1.097	0.691	1.742
Constant	-0.771	1.460	0.279	1	0.597	0.462		

*Variable(s) entered on step 1: b1, b2, b3, b4, b5, b6, b7, b8, b9, b10, b11, b12.

In the third section, what the dominant situation for loans will be when all the information is examined together is tried to be determined. Will the situation change when the basic results in Table IV and Table VII are examined together? With participation of all variables, will the basic variables to be focused by 242*36 information table change? When the table is evaluated with the genetic algorithms of the software ROSETTA to get replies of all these questions, 13799 feasible situations appear [34]. (Reduce→Genetic Algorithm →Object Related→Okay). As it is not possible to examine all 13799 situations, the basic filtering is applied and RHS accuracy [0, 0.75], RHS coverage [0, 0.10] and LHS length [5, 1000] ranges are used for the filtering. Feasible situations other than these ranges are eliminated, and Table XI that consists of 32 rules is found. A1, A4, A8, A9, A10, A13, A21, A22, A23, B3 and B7 variables are redundant, and therefore, they could be eliminated.

Table XI. The 32 rules algorithm

Rule #	A2	A3	A5	A6	A7	A11	A12	A14	A15	A16	A17	A18	A19	A20	B1	B2	B4	B5	B6	B8	B9	B10	B11	B12	Decision	Strength
	Conditions																									
1													0						2				3	2	1	13
2			1																2				3	2	1	13
3																			2		3		3	2	1	13
4		0										1												4	1	19
5										1	0			2										4	1	15
6					0		2													4		4			1	14
7	2	0										1												4	1	19
8		0	1									1											4		1	18
9		0										1					4		4						1	14
10		0										1											4		1	18
11												1					4			4				4	1	13
12		0										1						4						4	1	13
13		0										1			4								4		1	14
14		0										1				4							4		1	14
15												1					4		4					4	1	14
16		0			1							1												4	1	15
17		0							2			1											4		1	18
18		0								1		1												4	1	19
19		0								1		1											4		1	18
20		0				1						1											4		1	15
21		0						2				1											4		1	18
22		0										1							4				4		1	15
23		0								1				2										4	1	15
24										1				2		4								4	1	13
25		0							2			1												4	1	19
26		0		1								1												4	1	16
27		0												1		1								2	0	14
28		0														1	1							2	0	16
29					0												1		2					2	0	14
30					0												1					3		2	0	14
31					0												1			2				2	0	13
32																							1		0	22

When Table XI is examined, it is determined that there a clear differentiation in the balance sheet information. Very high balance sheet information causes refunding of loan and very low and medium balance sheet information causes non-refunding of commercial loan. Lack of information about the company that demands loan results with the behavior is causing the differentiation in Ai variables section.

When LR is applied for the 35 attributes, 23 qualitative and 12 quantitative variables, -2LL, Cox & Snell R Square and Nagelkerke R Square are found 272.255, 0.230 and 0.307, respectively. According to these values, high value of -2LL statistics shows that model does not represent the data totally. Moreover, low values of Cox & Snell R Square and Nagelkerke R Square statistics show that explanation power of independent variable of dependent variable is very low. When Hosmer -Lemeshow Test is applied, Chi-square is found 6.337 with the degree of freedom 8 and significance value 0.610. The Hosmer-Lemeshow statistic indicates a poor fit if the significance value is less than 0.05. Thus, the model adequately fits the data (p=0.610).

Table XII. LR classification for 12 attributes.

Observed			Predicted		
			Result		Percentage Correct
			0	1	0
Step 1	Result	0	81	40	66.9
		1	37	84	69.4
	Overall Percentage				68.2

*The cut value is 0.500

Table XII shows overall percentage of LR classification (68.2%) with correct classification numbers. However, Table 13 shows coefficients of variables, standard errors of these coefficients, Wald statistics, significance levels, odds ratio, Exp(B), and 95% of confidence intervals for Exp(B) in different columns.

In Table XIII, it is seen that odds ratio for a1, a7, a11, a13, a19, b1, b4, b5, b8, b11 and b12 will increase, odds ratio for a2, a3, a4, a5, a6, a8, a9, a12, a14, a15, a16, a17, a18, a20, a22, b2, b3, b6, b7, b9 and b10 will decrease. Since all results of the variables a10, a21 and a23 are same, they are removed from the evaluation. According to Wald statistics with the significance level 0.05 in Table 13, all LR coefficients, except a17, b3 and b11 which are shaded, are not significant, but the for variables a1, a7, a11, a13, a19, b1, b4, b5, b8, b11 and b12, it can be said that they are effective and preferable to other variables for non-refunding or refunding of loan payments because their odds ratios are greater than 1. Thus, LR model of qualitative and quantitative variables (35 attributes) is not a good model in categorizing the loans whether refunded or no refunded.

Table XIII. LR equation for qualitative and quantitative variables.

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B) Lower	Upper
Step 1* a1(1)	0.252	1.843	0.019	1	0.891	1.286	0.035	47.625
a2(1)	-0.475	0.967	0.242	1	0.623	0.622	0.093	4.135
a3(1)	-0.058	0.365	0.025	1	0.874	0.944	0.461	1.931
a4(1)	-0.104	0.827	0.016	1	0.900	0.901	0.178	4.554
a5(1)	-1.003	2.060	0.237	1	0.626	0.367	0.006	20.806
a6(1)	-0.107	0.396	0.072	1	0.788	0.899	0.413	1.955
a7(1)	0.267	0.412	0.418	1	0.518	1.306	0.582	2.929
a8(1)	-0.928	1.560	0.354	1	0.552	0.395	0.019	8.410
a9(1)	-0.355	1.189	0.089	1	0.765	0.701	0.068	7.203
a11(1)	0.389	0.398	0.955	1	0.328	1.476	0.676	3.221
a12(1)	-0.084	0.422	0.039	1	0.842	0.920	0.402	2.104
a13(1)	0.918	0.995	0.851	1	0.356	2.503	0.356	17.589
a14(1)	-0.766	1.397	0.301	1	0.583	0.465	0.030	7.180
a15(1)	-1.054	1.444	0.533	1	0.465	0.348	0.021	5.903
a16(1)	-0.575	0.710	0.656	1	0.418	0.562	0.140	2.264
a17(1)	-0.674	0.387	3.035	1	0.052	0.510	0.239	1.088
a18(1)	-0.589	0.446	1.743	1	0.187	0.555	0.231	1.331
a19(1)	0.777	0.651	1.424	1	0.233	2.174	0.607	7.783
a20(1)	-0.676	0.348	3.786	1	0.052	0.508	0.257	1.005
a22(1)	-1.837	1.292	2.021	1	0.155	0.159	0.013	2.005
b1	0.339	0.308	1.209	1	0.272	1.404	0.767	2.570
b2	-0.123	0.273	0.202	1	0.653	0.884	0.518	1.510
b3	-0.340	0.193	3.093	1	0.049	0.712	0.487	1.040
b4	0.216	0.182	1.415	1	0.234	1.241	0.869	1.773
b5	0.410	0.372	1.218	1	0.270	1.507	0.727	3.124
b6	-0.166	0.402	0.170	1	0.680	0.847	0.385	1.863
b7	-0.317	0.396	0.643	1	0.422	0.728	0.335	1.581
b8	0.338	0.402	0.708	1	0.400	1.402	0.638	3.080
b9	-0.874	0.654	1.787	1	0.181	0.417	0.116	1.503
b10	-0.067	0.702	0.009	1	0.923	0.935	0.236	3.698
b11	0.752	0.293	6.600	1	0.010	2.121	1.195	3.765
b12	0.192	0.254	0.570	1	0.450	1.212	0.736	1.994
Constant	-0.287	2.446	0.014	1	0.907	0.751		

*Variable(s) entered on step 1: a1, a2, a3, a4, a5, a6, a7, a8, a9, a11, a12, a13, a14, a15, a16, a17, a18, a19, a20, a22, b1, b2, b3, b4, b5, b6, b7, b8, b9, b10, b11, b12.

V. CONCLUSION

The business loans in financial transactions are very risky though they are quite important for both loan issuing and drawing companies. In this study, by using the RS approach, behaviors that increase the risk in loan payments or causing non-refunding and by using LR, classification of refunded and no refunded loan payments are tried to be determined. RS theory provides a very good differentiation in classification of loan payments and introduces the basic characteristics in three different examinations with qualitative, with quantitative and with both qualitative and quantitative variables as class differentiations. It is impossible for a company that can not get sufficient profit and has bad sales to refund the loan. For this reason, with RS approach, it is seen that the companies with high balance sheet information pay their loans and this differentiation can be clearly introduced by using the RS approach. However, LR analysis results for three models show that all LR models are not good in classification of refunded and no refunded loans according to significance values of LR coefficients with the significance level 0.05. When odds ratios of variables are examined, it is seen that some of variables can be the reason or one of the effects for refund or nonrefundable of loan payments, because their odds ratios are greater than 1.
Consequently, in classification of refundability of loans, the RS approach is very successful but LR is not. This success of RS approach clearly introduces to the loan issuing financial institutions to which information of loan demanding company they can focus on.

REFERENCES

[1] A. Sanchis, M.J. Segovia, J.A. Gil, A. Heras, J.L. Vilar, "Rough sets and the role of the monetary policy in financial stability and the prediction of insolvency in insurance sector," *European J of Oper Res*, vol. 181, no:3, 16 September 2007, pp. 1554-1573.
[2] S. S. Sancho, J. L. F. Villacañas, M. J. S. Vargas, C. B. Calzon, "Genetic programming for the prediction of insolvency in non-life insurance companies," *Computers & Oper Res*, vol. 32, no: 4, April 2005, pp. 749-765.
[3] A. I., Dimitras, R. Slowinski, R. Susmaga, C. Zopounidis, "Business failure prediction using rough sets," *European J of Oper Res*, vol. 114, no:2, 16 April 1999, pp. 263-280.
[4] R. G. West, "A factor-analytic approach to bank condition," *J of Banking and Finance*, vol. 9, pp. 254–266, 1985.
[5] S. H. Chen, C. H. Yeh, "Using genetic programming to model volatility in financial time series," in *Proc. of the 2nd Annu Conf. on Genetic Programming*, 1997, pp. 58–63.
[6] A. I. Dimitras, S. H. Zanakis, and C. Zopounidis, "A survey of business failures with an emphasis on prediction methods and industrial applications," *European J of Oper Res*, vol. 90, no:3, 1996, pp. 487–513.
[7] K. Y., "Tam Neural network models and the prediction of bankruptcy," *Omega* , vol. 19, no: 5, 1991, pp. 429–445.
[8] S. Balcaen, H. Ooghe, "Alternative methodologies in studies on business failure: Do they produce better results than the classic statistical methods?" *Vlerick Leuven Gent Management School Working Paper Series*, *no:* 16, 2004.
[9] E. I. Altman, "The success of business failure prediction models: An international survey," *J of Banking and Finance*, vol. 8, no: 2, 1984, pp. 171-198.
[10] T. Hennig-Thurau, A. Klee, "The impact of customer satisfaction and relationship quality on customer retention: A critical reassessment and model development," *Psychology & Marketing*, vol. 14, no: 8, 1997, pp. 737–764.
[11] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no: 3, 1965, pp. 338–353.
[12] Z. Pawlak, *Rough sets, theoretical aspects of reasoning about data*. Dordrecht/Boston/London: Kluwer Academic Publishers, 1991.
[13] Z. Pawlak, "A primer on rough sets: A new approach to drawing conclusions from data," *Cardozo Law Review*, vol. 22, 2002, pp.1407–1415.
[14] F. E. H. Tay, L. Shen, "Economic and financial prediction using rough sets model," *European J of Oper Res*, vol. 141, 2002, pp. 641–659
[15] Z. Pawlak, "Rough sets, in: T.Y. Lin, N. Cercone, (Eds.), Rough sets and data mining," Dordrecht: Kluwer Academic Publisher, 1997, pp. 3–8.
[16] R Slowinski, C. Zopounidis, A. I. Dimitras, R. Susmaga, Rough set predictor of business failure, in: R. A. Ribeiro, H. J. Zimmermann, R. R Yager, J. Kacprzyk, (Eds.), Soft Computing in Financial Engineering, Physica-Verlag, Wurzburg, 1999, pp. 402–424.
[17] M. J. Segovia, J. A. Gil, A. Heras, J. L. Vilar, A. Sanchis, Using rough sets to predict insolvency of Spanish nonlife insurance companies, Documentos de trabajo de la facultad de ciencias económicas y empresariales, Avaliable: http://econpapers.repec.org/paper/ucmdoctra/
[18] J. M. Ambrose, A. M. Carroll, "Using best's ratings in life insurer insolvency prediction," *The J of Risk and Insurance*, vol. 61, no: 2, 1994, pp. 317-327.
[19] B. Ruzgar, N. S. Ruzgar, "Classification of insurance sector with logistic regression", in *Proc. 9th WSEAS Int. Con. On Automatic Control, Modeling & Simulation*, 2007, pp. 52-57.
[20] J. B. Williamson, T. K. McNamara. (2001, May). Why some workers remain in the labor force beyond the typical retirement age? Avaliable: http://www.bc.edu/centers/crr/papers/ wp_ 2001-09.pdf. 6. 2006.
[21] R. Bran v, A. Agarwal, R. Leach. (2006, June). Predicting bankruptcy resolution. Avaliable: http:// bear.cba.ufl.edu/ agarwal/ vita/PredictingBRResolutionJBFA.pdf. 6. 2006.
[22] J. Temple. (2006). Explaining the private health insurance coverage for older Australians. Avaliable: http://www.elecpress. monash.edu.au/pnp/cart/download/free.php ?paper=178.17. 2006.
[23] A. Chandra, A. A. Samwick. (2006). Disability risk and the value of disability insurance. Avaliable: http://www.nber.org /papers/w11605.
[24] Z. Pawlak, A. Skowron, "Rough sets: Some extensions," *Information Sciences*, vol. 177, 2007, pp. 28–40.
[25] T.-F. Fan, D.-R. Liu, G.-H. Tzeng, "Rough set-based logics for multicriteria decision analysis," *European J of Oper. Res.*, vol. 182, 2007, pp. 340–355.
[26] J.Yu. Shyng, F. K. Wang, G. H. Tzeng, K. S. Wu, "Rough set theory in analyzing the attributes of combination values for the insurance market," *Expert systems with applications*, vol. 32, no: 1, January 2007, pp. 56-64.
[27] S. Weisberg, *Applied linear regression*, CA: Wiley Pub., 2005, pp. 255-265.
[28] A. S. Albayrak, *Uygulamalı cok degiskenli istatistik teknikleri*, Ankara: Asil Yayin, 2006, pp. 439-462.
[29] A. Agresti, *Categorical data analysis*, Florida: Wiley Pub., 2002, pp. 165-257.
[30] S. F. Hair, R. E. Anderson, R. L. Tahtam, W. C. Black, *Mutivariate data analysis*, NJ: Prentice-Hall, 1998, pp. 276-281, 314-321.
[31] H. Tatl dil, *Uygulamalı cok degiskenli istatistiksel analiz*, Ankara, 1992, pp. 225-232.
[32] K. Ozdamar, *Paket programlar ile istatistiksel veri analizi*, Eskisehir: Kaan Kitabevi, 2004, pp. 601-608.
[33] S. Menard, *Applied logistic regression analysis*, USA: Sage Pub., 1995, pp. 37-42.
[34] Rosetta Software, Avaliable: http://rosetta.lcb.uu.se/general/

B. Ruzgar was born in Musulca, Edirne, Turkey in 1962. He received the MS degree in Mathematics from Marmara University, Istanbul, Turkey in 1986 and Ph. D. degree in Quantitative Methods from Mamara University, Istanbul, Turkey, in 1992.

He worked as a Mathematician, Research Assistant at Department of Management, Marmara University, Istanbul Turkey. Currently, he works as an Assistant Professor at Baking and Insurance School, Marmara University, Istanbul, Turkey. He has five books, an author of more than 15 papers in refereed journals and more than 50 papers in conference proceedings. His research interests are fuzzy logic, applied statistics, quantitative methods.

Assis. Prof. Dr. Ruzgar is a Member of Mathematics Association of Turkey, Member of Operation Research Society of Turkey, Member of Informatics Association of Turkey and Member of Association of Econometrics.

N. S. Ruzgar was born in Pinarhisar, Kirklareli, Turkey in 1962. She received the MS degree in Mathematics from Istanbul Technical University, İstanbul, Turkey in 1989 and Ph. D. degree in Quantitative Methods from Istanbul University, Istanbul, Turkey, in 1998.

She worked as a Mathematician, Lecturer, and Assistant Professor at Computer and Electronic Education Department of Technical Education Faculty, Marmara University, Istanbul Turkey. Currently, she works as an Associte Professor at Vocational School of Social Sciences, Marmara University, Istanbul, Turkey. She has four books, an author of more than 15 papers in refereed journals and more than 60 papers in conference proceedings. Her research interests are system simulation, applied statistics, quantitative methods, and distance education.

Assoc. Prof. Dr. Ruzgar is a Member of Mathematics Association of Turkey, Member of Operation Research Society of Turkey, Member of Informatics Association of Turkey and Member of Association of Econometrics.