Rough set theory and discriminant analysis to classify financial data

Nursel Selver Ruzgar, Bahadtin Ruzgar and Fahri Unsal

***Abstract*—** The mutual funds play a huge role in today's financial markets. They range from a low risk to very high risk depending on the portfolio of financial instruments included and in what proportions. Typically, funds that are heavy in bonds are low risk while funds heavy in stocks are high risk. Mutual mixed-asset funds, due to their portfolio mix, are influenced by market movements in either direction. That is one reason why this paper focusses on daily price movements and the resulting positive, negative and neutral returns. The statistical discriminant analysis and the mathematical rough set theory approaches were used in the analysis of Turkish mutual fund industry. 14 mixed asset funds were analyzed for daily price and return movements using 2,267 data points from January 2, 2004 to December 31, 2012. Discriminant analysis approach provided a moderate classification in terms of positive or negative returns. The common structures provide by the rough set approach provided generally better results. The resulting models were tested with actual data and the results were very promising.

***Keywords*—** Mutual funds, Discriminant Analysis, Rough Set Theory, Rosetta GUI, Investment Analysis.

I. Introduction

In today's dynamic financial markets, millions of people prefer to buy mutual funds either because they don't have the time to investigate the financial instruments in detail or they don't have the skills to make sound investment decisions. A mutual fund is an investment pool in which thousands of investors put money. The mutual fund managers are professionals who make the daily decisions on what mix of stocks, bonds, and other financial instruments to purchase using the money collected in the pool. They are in constant touch with the companies and sectors they invest in order to make the best possible choices. The investors in a mutual fund own shares of all the stocks and bonds owned by the fund. One share of a mutual fund could potentially give an investor an ownership interest in the stocks and bonds of dozens of different corporations. This kind of diversification reduces risk and potentially the cost of participation for ordinary people who may not have the time and skills to be actively involved in the stock market.

N. S. Ruzgar is with the Department of Mathematics, Ryerson University, Toronto, Canada. (corresponding author phone: 1-416-979 5000/ ext. 3173; e-mail: nruzgar@ryerson.ca).

B. Ruzgar is with Banking and Insurance School (retired from), Marmara University, Istanbul, Turkey. (e-mail: bahadtinruzgar@gmail.com).

F. Unsal is with Marketing Department at Ithaca College, Ithaca, NY, USA. (e-mail: unsal@ithaca.edu).

Depending on the mix, some mutual funds emphasize higher risk and potentially higher returns while others will emphasize lower risk and potentially lower returns. Given the number of financial institutions that can create mutual funds and the number of potential mixes, there can be hundreds of mutual funds on the market. Even a given financial institution might be marketing tens of mutual funds that it has created.

The main objective of this study was to investigate the structure of price movements in mutual funds in Turkey. Daily mutual mixed-asset fund prices were collected from the Turkish market for the period from January 2, 2004 to December 31, 2012, yielding 2,267 (9 years) business day data points [1]. There were 324 mutual funds traded in Turkey. Of these, 116 were defined as A-type funds and the remaining 209 were defined as B-type funds. The A-type funds included: variable funds, index funds, share funds, contributory funds, mixed funds, private funds, private sector funds, and foreign securities equity funds. The B-type funds included: gold and security funds, variable funds, index fund, fund baskets, private funds, mixed fund, liquid fund, government bonds and foreign securities equity funds.

For the purposes of this study, 14 mixed funds were chosen for further analysis. These were all in A-type category and their contents covered from low risk to high risk investment instruments. Actually there were 16 such funds. However, those funds with less than 500 observations (business days) were excluded from the analysis, thus ending up with 14 funds. These funds were selected because they are easily affected from positive or negative market movements. It is believed that many of these mutual funds were very similar in nature, but confused the consumers in the fund selection process. The differences in most cases were artificially created for marketing promotions to different market segments. An attempt will be made in this study to classify the above-mentioned mixed funds using two different approaches, rough set theory and discriminant analysis to see how they compare. It is hoped that the findings will be useful to other researchers, mutual fund managers, and the ultimate consumers.

II. Literature Review

Mutual funds provide opportunities for small investors to participate in the capital market without assuming a very high degree of risk [2]. The relationship between risk and return usually determines the performance of a mutual fund. Studies on mutual funds are abundant for the US market. However, the number of studies conducted in the European countries and emerging countries are quite limited [3]. Some of these studies focus on the level of investor knowledge and information sources. It is generally argued that the individual investors (consumers) have limited information about the funds available and hence their decisions are easily influenced by bank personnel or fund managers. They might also be influenced by advertising online or traditional media [4]. Another study has investigated the relationship between investors' financial knowledge and mutual fund advertising. It concluded that mutual fund ads with financial disclosures are more likely to generate higher levels of recall and positive thoughts regarding advertised information [5].

A number of researchers have studied the structure of mutual funds using techniques such as cluster analysis, factor analysis, time series, regression analysis, rough sets, fuzzy logic, and discriminant analysis. For example, one study using factor and cluster analysis looked into investor security in the mutual fund market [6]. Strieter and Singh identified specific characteristics important in establishing and maintaining mutually beneficial relationships between endowment and pension fund managers and the providers of investment management services [7]. Jonas studied the consumer investment patterns in socially reasonable investments, and profiled mutual funds by cluster and discriminant analysis [8]. This study determined that there is a consumer market segment that is primarily concerned about social responsibility over financial return. Another study that utilized discriminant analysis was conducted in Bosnia and Herzegovina for stock selection purposes [9]. Another study investigated mutual funds using cluster and recursive factor analysis [10].

Among the international studies, one can list the following. The mutual funds market in India, for example, has been studied using cluster analysis [11]. Similarly, the Greek mutual fund market has been investigated using a multi-criteria methodology [12]. Another study using factor and cluster analysis looked into the risk factor in classifying mutual funds as in [13]. A 2012 study evaluated the risk, return and performance measures of selected stocks traded in Belgrade Stock Exchange in Serbia using regression analysis [14]. A study conducted in the Euro area using Bloomberg data for the 2002-2008 period, researchers in Italy have classified stocks into cluster using a three-stage pure statistical analysis with a great deal of success [15]. These researchers proposed fuzzy optimization model to determine the optimal portfolio. During the last two decades, there have been many applications of rough set theory as well in business, medical, and engineering fields.

Classification of the data used in models constitutes the very important part of the statistical analysis. The outcome is usually to assign set of alternatives into predefined homogenous classes. Most of the existing classification methodologies are based on absolute comparisons among the alternatives and some reference profiles (cut-off points) that discriminate the classes. As one might expect, many financial analysis decisions involve the classification of a set of observations into one or several groups using techniques available to researchers from the operations research field. Classification techniques based on traditional statistical approaches include discriminant function analysis and logistic regression [16].

Discriminant analysis is one of the most popular statistical methods to estimate the relationship between a dependent variable and independent variables. The main purpose is to predict group membership based on a linear combination of the interval variables. The procedure begins with a set of observations where both group membership and the values of the interval variables are known. The end result of the procedure is a model that allows prediction of group membership when only the interval variables are known. A second purpose of discriminant analysis is to evaluate data set, create prediction models. The dependent variable is categorical variable which is the combination of two or more categories whereas independent variables are normally distributed interval variables. Discriminant analysis has been widely used for market segmentation studies and in assessing the health of businesses. Several comprehensive review articles on the use of alternative technologies for prediction of business failures have appeared and have included the work. Different approaches and techniques have been used for forecasting the likelihood of failures and it is quite common to use financial ratios [17]. Another application has been made to 24 non-life insurance companies in Turkey for the 2004-2006 period [16]. Overall, the goal of the analysis is to find discriminant function(s) which can differentiate between groups. More technical details of discriminant analysis can easily be found in literature.

As was mentioned above, Rough Set Theory was also used in this study. It is a technique that can reduce dimensionality by using the information contained within the data set and preserving the meaning of the features that are clearly desirable. Rough Set approach can be used as such a tool to discover data dependencies and reduce the number of attributes contained in a data set by purely structural methods [18, 19].

The Rough Set approach accepts both quantitative and qualitative variables by deriving a number of decision rules (deterministic and non-deterministic sorting rules) or if… then rules. First, a range of minimal subsets of independent attributes is constructed. A subset of attributes is called a minimal subset if this subset has the same sorting quality as the whole set of attributes. Then, the core of attributes is defined as the intersection of all minimal subsets. Next, a reduced decision table is constructed, in which the redundant attributes are eliminated. Finally, on the basis of this decision table, the set of sorting rules, the sorting algorithm, is derived and firms are classified by matching their description to the set of sorting rules [20]. In Rough Set Theory, attribute reduction is a key research problem and many useful algorithms have been proposed [21-23]. A reduct (rough subspace) is a subset of all attributes, which not only excludes irrelevant and redundant attributes, but also keeps the granular structure of the original data [24].

As implied above, Rough Set approach is especially helpful in dealing with vagueness and uncertainty in decision situations. The basic idea is to take objects, attributes, and decision values, and create rules for upper, lower, and boundary approximations of the set. With these rules, a new object can easily be classified into one of the regions. Worldwide, there has been a rapid growth in interest in Rough Set approach and its applications in recent decade. Evidence of this can be found in the increasing number of high-quality articles on rough sets and related topics that have been published in a variety of international journals, symposia, workshops, and international conferences in recent years. Rough Set Theory applications have been applied to a variety of business and medical problems, such as predicting business failure, stock market analysis, marketing, medical sciences [18, 25-28].

In summary, grouping analysis is used to divide the units to homogenous groups by using some measures depending on similarities or differences and to define some definite prototypes. While the discriminant analysis classifies the examined area in a category, it requires that the independent variable meets the normality assumption. Rough Set Theory, on the other hand, helps us to determine which independent variables are important for the classification using group properties. Rough Set approach also classifies dependent variable as positive or negative by using the categorization of independent variables. Moreover, it allows the independent variables to be continuous, discrete and categorical variables. Thus, Rough Set approach may be more flexible. This paper utilized both discriminant analysis and rough set approach to classify the insurance companies in Turkey [16].

III. Methodology

Daily mutual mixed-asset fund price were collected from the Turkish market for the period from January 2, 2004 to December 31, 2012, yielding 2267 (9 years) business day using the government sources [1]. Data was collected for 324 mutual funds available in Turkey. Of these, 16 were mutual mixed-asset funds.

However, those funds with less than 500 observations (business days) were excluded from the analysis. As a result, 14 mixed-asset funds were used for further analysis. Mixed-asset funds were selected because they are easily affected from positive and negative market movements. It is believed that many of these mutual funds were very similar in nature, but confused the consumers in the fund selection process. The differences in most cases were artificially created for marketing promotions to different market segments.
This study has three objectives as outlined below:

1. Using discriminant analysis, determine the separation of prices as negative or positive change or neutral on a daily basis for each fund. Then, test the validity of discriminant functions established for each one of the 14 funds using 150-day real data.
2. Using Rough Set Theory, establish common investment ratios for negative or positive change or neutral. Then, for each fund, test the common structures established by the Rough Set approach to see the success rate.
3. Establish the advantages and disadvantages of using discriminant analysis and/or Rough Set approach for the 14 mixed-asset funds that are each owned by a different financial institution, investigated in this study.

IV. Application and Findings

It is generally argued that the individual investors have limited information about the potential performance and risk level of the funds available. Even though there are many financial instruments, the data obtained from market for 14 mutual mix-asset funds were limited to 6 financial instruments: shares, bonds, inverse repo, stock market, foreign investment, and others [1]. In this study, to classify mutual mixed asset funds two different methods were applied discriminant analysis and Rough Set Theory. For both methods, the dependent variable (price) was categorized as a positive return, negative return or neutral. These categories were defined as follows: if the price increased when compared with the previous day's price, it was coded as 1, reflecting a positive change (PC); if the price decreased, it was coded as -1, reflecting a negative change (NC); and if there were no change in price, it was coded as 0, reflecting neutrality (N). In the case of independent variables, since the structures of two methods are different, the real investment ratios were used for discriminant analysis whereas the interval investment ratios were used for Rough Set Theory.

For discriminant analysis, first, Kolmogorov- Smirnov normality test was applied to independent variables for each fund to test whether the independent variables satisfy the normality requirement. The results indicated that they did not satisfy the normality test, due to a light positive skewness. Therefore, to reduce the non-normality, the square root transformation was applied to the independent variables and it was seen that square root transformation worked well and the square root of independent variables satisfied the normality condition.

Table 1: Classification of mixed-asset funds by discriminant analysis

			Predicted group membership							Predicted group membership			
	Original	Code	-1	1	Total	CP*		Original	Code	-1	1	Total	CP*
F1	Count	-1	464	475	939	54.9	F9	Count	-1	462	505	967	51.8
		1	499	720	1254				1	592	707	1299	
	%	-1	49.4	50.6	100.0			%	-1	47.8	52.2	100.0	
		1	40.9	59.1	100.0				1	45.6	54.4	100.0	
F2	Count	-1	553	464	1017	51.9	F10	Count	-1	592	443	1035	51.5
		1	626	622	1248				1	655	575	1230	
	%	-1	54.4	45.6	100.0			%	-1	57.2	42.8	100.0	
		1	50.2	49.8	100.0				1	53.3	46.7	100.0	
F3	Count	-1	411	581	992	55.8	F11	Count	-1	495	503	998	54.0
		1	420	851	1271				1	540	726	1264	
	%	-1	41.4	58.6	100.0			%	-1	49.6	50.4	100.0	
		1	33.0	67.0	100.0				1	42.7	57.4	100.0	
F6	Count	-1	555	450	1005	53.2	F13	Count	-1	522	507	1029	53.3
		1	610	649	1259				1	552	685	1237	
	%	-1	55.2	44.8	100.0			%	-1	50.7	49.3	100.0	
		1	48.5	51.5	100.0				1	44.6	55.4	100.0	

*CP: Percentage of original grouped cases correctly classified

Secondly, discriminant analysis was applied to the dependent and the transformed data, the canonical discriminant function (with unstandardized coefficients) and Fisher's linear discriminant function were found. Then, Box's M was evaluated for each fund and found that M was significant for each funds ($p<0.05$). When Wilk's lambda test was applied, the significance level of chi-squared statistics were found greater than 0.10 for 6 of 14 mutual mixed asset funds, therefore these 6 funds were discarded from the evaluation (F3, F4, F7, F8, F12 and F14) and remaining 8 funds were classified by discriminant analysis as illustrated in Table 1. It was seen that the classification percentages of original grouped cases ranged between 51.5% and 55.8%. These overall percentages show that these 8 of 14 mutual mixed-asset funds, each owned by a different financial institution, have similar tendencies.

Market data only gave the information on daily price and the investment ratios of financial instruments. To obtain more information about the financial instruments for each fund, prospectuses of all 14 funds were examined individually. Investment ratio intervals of financial instruments are shown in Table 2. The F1 fund, for example, has the following limits. The fund manager can invest 20-80 percent of the deposits in shares [20, 80], 20-80 percent in government bonds [20, 80], 0-60 percent in reverse repo [0, 60], and 0-60 in foreign government and private sector bonds [0, 60]. This particular fund does not allow any investments in the stock market and "others" categories. Other funds can be read in a similar fashion.

Table 2: Investment limit intervals of financial instruments for 14 mixed-asset mutual funds

Funds	Shares	Government Bonds	Inverse Repo	Stock Market	Foreign Government and Private Sector Bonds	Others*
F1	[20,80]	[20,80]	[0,60]	[0,0]	[0,60]	[0,0]
F2	[20,80]	[20,75]	[20,75]	[0,0]	[0,60]	[0,0]
F3	[35,65]	[0,30]	[0,30]	[0,20]	[0,20]	[0,0]
F4	[20,60]	[20,75]	[0,55]	[0,0]	[0,0]	[0,0]
F5	[10,90]	[0,90]	[0,90]	[0,0]	[0,0]	[0,0]
F6	[25,100]	[0,49]	[0,49]	[0,20]	[0,49]	[0,40]
F7	[0,80]	[0,80]	[0,80]	[0,20]	[0,20]	[0,50]
F8	[25,80]	[0,55]	[20,75]	[0,0]	[0,0]	[0,50]
F9	[20,70]	[20,70]	[0,60]	[0,20]	[0,0]	[0,0]
F10	[20,80]	[20,75]	[0,75]	[0,20]	[0,0]	[0,5]
F11	[20,80]	[20,80]	[0,60]	[0,20]	[0,60]	[0,25]
F12	[25,95]	[5,75]	[0,75]	[0,0]	[0,0]	[0,0]
F13	[20,80]	[20,80]	[0,80]	[0,25]	[0,25]	[0,5]
F14	[25,80]	[20,75]	[0,55]	[0,20]	[0,49]	[0,0]

*Accession Partnership Documents, Cash, Gold and Precious Metals, Real estate investment trust, Private sector Bonds, Securities, Options, Foreign Shares, and Repo

With daily market data for 2267 business days, there are 2267 rows of data. Given the 6 financial instruments and prices, there are 7 columns. Thus, a 2267x7 information table was created for each one of 14 funds. The analysis was conducted using these tables as the source.

In order to apply Rough Set approach on this data or in classification of mutual funds with Rough Set approach, the first step of the analysis involves recoding the investment limits of financial instruments into categorical variables. This categorization was done by dividing each investment range into 5 equal length sub- intervals with the codes a, b, c, d and e. For example, the interval [20, 80] of shares for F1 is divided into 5 subintervals as a (very low) = [20, 32), b (low) = [32, 44), c (medium) = [44, 56), d (high) = [56, 68) and e (very high) = [68, 80]. The recoding was done by dividing the original domain into sub intervals since such analysis is very useful in drawing general conclusion from the 14 mutual mix-asset funds in terms of dependencies, reducts, and decision rules [25]. Actually, the original domain can be divided into different number of subintervals using different approaches (such as the use of medians and geometric means), but in literature, generally quartiles were used for this purpose. This recoding is a requirement of the Rough Set Theory, but the Rosetta software does not do that automatically. So the user recodes the domain manually. Thus, in this study, the information tables were created by recoding the original data into five subintervals for the actual ratios. Since the investment ranges are different for different funds, range of categorical variables has different ranges for different funds.

The second step of the analysis involves converting daily prices to a categorical variable. To do so, consecutive daily prices were compared. If the price increased when compared with the previous day's price, it was coded as positive change (PC); if the price decreased, it was coded as negative change (NC); and if there is no change in price, it was coded as neutral (N). As in discriminant analysis, this coding is necessary, because this helps the researchers to determine patterns for the price changes by examining daily investment ratios and the price. In other words, it is expected that we find patterns from 2267 business day data for the price chances whether negative, positive or no change when investing on 6 financial instruments in any of the interval very low, low, medium, high or very high.

In the third step, the 2267x7 data matrices formed by 6 coded variables with the price for each 14 mutual funds are examined individually for each day by using the ROSETTA GUI Version 1.4.41 software [29]. For this purpose, genetic algorithms are used for reduces. The process on ROSETTA GUI works as in the following order: 1- select "Reduce"; 2- select "Genetic Algorithm"; 3- select "All objects" for "Object Related" in "Discernibility"; 4- select "Modulo Decision" in "Table interpretation"; 5- select "Okay". As a result of this process, two tables are obtained. While the first one has three columns "reduct", "support" and "length", the second one has nine columns; rules, LHS Support, RHS Support, RHS Accuracy, LHS Coverage, RHS Coverage, RHS Stability, LHS Length and RHS Length. The first table shows the logic and structure of rules. The second table shows all the detail of the rules. Analysing the data with ROSETTA, the following feasible situations have appeared for 14 mutual funds: F1: 23-150; F2: 27-104; F3: 3-8; F4: 12-69; F5: 6-22; F6: 26-120; F7: 12-64; F8: 10-46; F9: 7-31; F10: 10-32; F11: 20-42; F12: 6-16; F13: 17-42; and F14: 5-28. The first number represents the number of feasible solutions for the combination of reduct, support and length; and the second number represents the number of feasible solutions with rough sets.

Examination all feasible situations will take time and may cause ignoring of main necessary subjects, thus, basic filtering is used. The basing filtering process on ROSETTA GUI works in the following order: 1- select "Filter"; 2- select "Basic Filtering"; 3- select "Rule Filter"; 4- select "Remove rules with RHS support" in "Criteria"; 5- select " RHS coverage" in "Remove Rules"; 6- select "Interpret the composed set of criteria as a “Disjunction (OR)” “ in “Connective”; and 7-select “Reduct”. In basic filtering, the intervals [0, 50] and [0, 0.10] are assigned to RHS (right hand side) support and RHS coverage, respectively. By eliminating alternative situations that appear other than these ranges, a new table is established with Rough Set approach values consist of 79 rules. Since there is limited space, a table with 79 rows cannot be shown on paper. For this reason, the system was run also with the new remove restriction interval [0, 0.17] on LHS (left hand side) coverage and a new table was obtained with 18 rules. These rules were met by only 11 funds, and they are illustrated in Table 3.

Table 3: Important structures of financial instruments for 11 mutual mix-asset funds

Funds	Share	Bonds	Inv Repo	Stock_Mar	For_Inv	Others	Decision	LHS Support	RHS Support	RHS Accuracy	LHS Coverage	RHS Coverage	LHS Length	RHS Length
F2	b		d	a			PC, NC	635	363, 272	0.571654, 0.428346	0.280106	0.290064, 0.267453	3	2
F3						c	PC, NC	675	363, 312	0.537778, 0.462222	0.45	0.445963, 0.454148	1	2
	c						PC, NC	255	140, 115	0.549019, 0.45098	0.17	0.171429, 0.167394	1	2
F4	b	a		a			PC, NC	494	293, 201	0.593117, 0.406883	0.217909	0.233652, 0.198617	3	2
		a	d	a			PC, NC	574	338, 236	0.58885, 0.41115	0.253198	0.269537, 0.233202	3	2
	a	a	e	a			PC, NC	398	215, 183	0.540201, 0.459799	0.175562	0.171451, 0.18083	4	2
F5	d	b	a				PC, NC	956	559, 397	0.584728, 0.415272	0.421703	0.438238, 0.400202	3	2
		c	a				PC, NC	518	269, 249	0.519305, 0.480695	0.228496	0.211644, 0.251008	2	2
F7	d	a					PC, NC	275	149, 126	0.541818, 0.458182	0.194209	0.187186, 0.204878	2	2
F9			b				PC, NC	983	576, 407	0.585961, 0.414039	0.433613	0.443418, 0.420889	1	2
F10	c	b		a			PC, NC	441	231, 210	0.52381, 0.47619	0.19453	0.187805, 0.202899	3	2
F11	b	d				a	PC, NC	573	339, 234	0.591623, 0.408377	0.252757	0.267405, 0.234469	3	2
F12		c					PC, NC	1258	695, 563	0.552464, 0.447536	0.554918	0.557338, 0.552502	1	2
	b		a				PC, NC	399	233, 166	0.58396, 0.41604	0.176004	0.186848, 0.162905	2	2
F13	b	c		a		a	PC, NC	481	287, 194	0.596674, 0.403326	0.212175	0.231205, 0.188533	4	2
		b	a	a		a	PC, NC	696	368, 328	0.528736, 0.471264	0.307014	0.297494, 0.318756	4	2
F14		b	b				PC, NC	467	256, 211	0.54818, 0.45182	0.205999	0.204473, 0.208911	2	2
	b	c					PC, NC	530	307, 223	0.579246, 0.420755	0.233789	0.24361, 0.220792	2	2

Table 3 shows which portion of financial instruments is invested for each mutual mix-asset fund to obtain whether negative, positive or no change in price with given limits. The first 7 columns show the type of financial instrument invested, and in which interval the investment was made for 11 mutual funds. Decision column shows the price change according to the investments while LHS support column shows how many days these price changes were seen out of 2267 days with this investment. RHS support column shows how many days positive, negative or no changes occurred with this investment while RHS Accuracy column shows the percentages of these price changes out of total occurrence. LHS Coverage column represents the ratio of number of occurrence days by 2267, while RHS Coverage column shows the ratios of number of positive, negative or no change occurrences by the corresponding total occurrences in 2267 days with any investment. In addition, LHS Length represents the number of different investments while RHS Length represents how many decision variables occurred.

Another objective of this study was to test the validity of the obtained linear discriminant functions. These tests were conducted for the remaining 8 mutual mixed asset funds with 150-day real data (from January 2nd, 2013 to August 1st). This was done and predicted group membership scores were found. These predicted group membership scores were compared with the original membership scores and the results are shown in Table 4. The linear discriminant functions were obtained for the remaining 8 mixed-asset funds. These functions were then tested with 150 days of real data for prediction accuracy. The results showed that two of the predictions are worse than the actual percentages shown in Table 1. Five of the predictions are about the same as the actual values, and one prediction is better that the actual. The overall results indicate that, although not ideal, discriminant analysis and classify mixed-asset funds in an effective way.

Table 4: Discriminant analysis test results

			Predicted group membership							Predicted group membership			
	Original	Code	-1	1	Total	CP*		Original	Code	-1	1	Total	CP*
F1	Count	-1	41	30	71	59.7	F9	Count	-1	36	30	66	51.3
		1	30	48	78				1	43	41	84	
	%	-1	57.7	42.3	100.0			%	-1	54.5	45.5	100.0	
		1	38.5	61.5	100.0				1	51.2	48.8	100.0	
F2	Count	-1	30	37	67	52.7	F10	Count	-1	41	26	67	45.0
		1	34	48	82				1	56	26	82	
	%	-1	44.8	55.2	100.0			%	-1	61.2	38.8	100.0	
		1	41.5	58.5	100.0				1	68.3	31.7	100.0	
F5	Count	-1	52	28	80	54.0	F11	Count	-1	30	36	66	48.3
		1	41	29	70				1	41	42	83	
	%	-1	65.0	35.0	100.0			%	-1	45.5	54.5	100.0	
		1	58.6	41.4	100.0				1	49.4	50.6	100.0	
F6	Count	-1	38	24	62	54.0	F13	Count	-1	25	40	65	52.7
		1	45	43	88				1	31	54	85	
	%	-1	61.3	38.7	100.0			%	-1	38.5	61.5	100.0	
		1	51.1	48.9	100.0				1	36.5	63.5	100.0	

*CP: Percentage of original grouped cases correctly classified

The other objective of the study was to test the validity of the common structures of rough set results with the 150-day real data. Findings were tested with 150-day (7 months) real data from January 2nd, 2013 to August 1st, 2013 to see validity of the patterns and how these patterns can be used in the future. To test the findings, the number of days in 150-day data was searched to see whether the same pattern would be observed as in Table 3. Then, the rough set was applied to those numbers manually. The results are shown in the last five columns of Table 5. When the corresponding five columns in Table 3 and Table 5 were examined with respect to, “LHS Support - LHS Support tested”, “RHS Support - RHS Support tested”, “RHS Accuracy - RHS Accuracy tested”, “LHS Coverage - LHS Coverage tested” and “RHS Coverage - RHS Coverage tested” it was seen that 16 out of 18 Rough Set approach results matched with each other with 88.89 percent accuracy. Two of 18 Rough Set approach results did not match with each other and they were shown in shaded rows in Table 5. Finally, an examination of the first 7 columns of Table 5 indicates that the funds issued by different companies show similarities in terms of content.

As was noted earlier, only 8 of the 14 mixed-asset mutual funds could be analyzed using discriminant analysis (the remaining 6 funds could not be analyzed because they did not meet the normality conditions). Likewise, when the data was analyzed using the Rough Set approach only 11 funds met the conditions for inclusion. There was a common set of funds that were included in both results. These funds were F2, F5, F9, F10, F11, and F13. These observations indicate that there are similarities between discriminant analysis and Rough Set approach in classifying mutual mixed-asset funds.

Table 5: Rough set theory test results

Funds	Share	Bonds	Reverse Repo	Stock_Mar	For_Ins	Others	Decision	LHS Support (Tested)	RHS Support (Tested)	RHS Accuracy (Tested)	LHS Coverage (Tested)	RHS Coverage (Tested)
F2	b		d	a			PC, NC	33	21, 12	0.636364, 0.363636	0.22	0.253012, 0.179104
F3						c	PC, NC	84	51, 33	0.607143, 0.392857	0.56	0.621951, 0.4985294
	c						PC, NC	31	12, 19	0.387097, 0.612903	0.21	0.146341, 0.279412
F4	b	a		a			PC, NC	97	57, 40	0.587629, 0.412371	0.65	0.695122, 0.588235
		a	d	a			PC, NC	56	32, 24	0.571429, 0.428571	0.37	0.390244, 0.352941
	a	a	e	a			PC, NC	25	14, 11	0.56000, 0.44000	0.17	0.170732, 0.161765
F5	d	b	a				PC, NC	28	16, 12	0.571429, 0.428571	0.19	0.228571, 0.150000
		c	a				PC, NC	35	10, 25	0.285714, 0.714286	0.23	0.142857, 0.312500
F7	d	a					PC, NC	24	13, 11	0.541667, 0.458333	0.16	0.144444, 0.186441
F9			b				PC, NC	129	71, 58	0.550388, 0.449612	0.86	0.845224, 0.878788
F10	c	b		a			PC, NC	32	20, 12	0.6250, 0.3750	0.21	0.243902, 0.179104
F11	b	d				a	PC, NC	56	26, 20	0.535088, 0.464912	0.37	0.313253, 0.303030
F12		c					PC, NC	114	61, 53	0.524823, 0.475177	0.76	0.762500, 0.757143
	b		a				PC, NC	29	16, 13	0.551724, 0.448276	0.19	0.200000, 0.185714
F13	b	c		a		a	PC, NC	35	20, 15	0.571429, 0.428571	0.23	0.235294, 0.230769
		b	a	a		a	PC, NC	54	41, 13	0.759259, 0.240741	0.36	0.482353, 0.200000
F14		b	b				PC, NC	38	22, 16	0.578947, 0.421053	0.25	0.261905, 0.242424
	b	c					PC, NC	51	30, 21	0.588235, 0.411765	0.23	0.357143, 0.318182

In general, discriminant analysis provided a moderate amount level of classification with respect to daily positive and negative price changes. Similar results were obtained when actual data was applied to the linear discriminant functions. On the other hand, the analysis conducted with Rough Set approach indicated that common structures could be obtained for positive and negative returns. This was also valid with the test data. It should be noted that the two approaches are quite different in terms of their methodology and solution procedures. Discriminant analysis categorises daily price movements as positive or negative. The Rough Set approach determines positive or negative return structures. Thus, one might argue that the rough set theory approach provides more specific information for the fund managers.

V. Conclusion

The growing literature indicates that a variety of statistical and mathematical models are applied to business problems. These applications are especially very common in the finance field due to evaluate data. Between these two techniques, discriminant analysis usually gets wider acceptance. One of the objectives of the study was to use the same data set for discriminant analysis and Rough Set approach to find out whether the results similar or not.

In this study, the price movements in 14 Turkish mutual mix-asset funds for a 9-year period (2267 business days) were investigated using discriminant analysis and rough sets approaches. The resulting patterns were then tested using additional data for 150 business days. It was noted the two models matched pretty well with a moderate to high level of accuracy. With Rough Set Theory, the accuracy rate was 89 percent while the corresponding value for discriminant analysis was 63 percent.

These findings should be very valuable for fund managers. They should read the daily negative and positive price movements very carefully before they make investment decisions. They should specifically study the classification of investment instruments with Rough Set approach before they make their daily purchases of financial instruments.

Although the patterns determine the positive, negative changes or neutral on the fund prices, it seems that they are not very strong indicators since the differences between the percentages of positive, negative changes or neutral are very small for all investigated funds using the two approaches.

This could be because of other factors that were not considered in this study. It is hypothesized that by including changes in inflation rates and exchange rate fluctuations will improve the results. This is an area that we will focus on in further studies.

One might argue that the investment instruments for the mixed-asset funds are affected from market conditions more than the all other funds. Thus, it is strongly believed that a studying low risk funds with discriminant analysis and Rough Set approach will provide better results. Our immediate research agenda will include a study of low risk and high risk funds using the same methodologies used in this study. This will enable us to view similarities and differences between the two types of funds and provide investment strategies for the fund managers. Our longer-term agenda will include replication of this current study with the introduction of additional environmental factors such as inflation and exchange rates. We also plan to add sensitivity analysis for both studies.

In conclusion, it can be said that the Rough Set approach provided very promising results. The predictions were better than discriminant analysis predictions. Thus, when time is limited to analyze data using both methods, the rough set approach can be useful to fund managers in their daily purchase decision. It is noted that mixed-asset funds are affected by a number of economic factors. It is expected that the findings will be a lot more useful when one uses the same methodology with low-risk funds in future studies.

References

[1] spk.gov.tr, Last access: September 9, 2013.

[2] V. Kumar, and P. Poddar, "Performance evaluation of HDFC mutual funds," *Sumedha J. of Management*, vol. 3, no. 2, Apr.-June 2014

[3] T. P. Korkeamaki, T. I.. Smythe, "Effects of market segmentation and bank concentration on mutual fund expenses and returns: Evidence from Finland," *J. of European Fin. Management*, vol. 10, no. 3, pp. 413-438, 2004.

[4] J. Kozup, E. Howlett and M. Pagano, "The effects of summary information on consumer perceptions of mutual fund characteristics," *J. of Consumer Affairs*, vol. 42, no. 1, pp. 37-59, 2008.

[5] T. D. Lee, T. W. Yun, and E. Haley, "The interplay between advertising disclosures and financial knowledge in mutual fund investment decisions," *J. of Consumer Affairs*, vol. 43, no. 2, pp. 260-287, 2012.

[6] C. M. F. Monterio, J. O. Soares and H. M. P. Teixeire, "A multivariate analysis of the relative impact of performance and relationship attributes on customer satisfaction," in *Proc. 12th WSEAS Int. Conf. on Mathematical and Computational Methods in Science and Engineering (MACMESE'10)*, Wisconsin, USA, 2010, pp. 265-271,

[7] G. Sidana, and D. Acharya, "Classifying mutual funds in India: Some results from clustering," *Indian J. of Econ. and Bus.*, 2007.

[8] N. Jonas, "Segmenting socially responsible mutual fund investors: The influence of financial return and social responsibility," *The Int. J. of Bank Marketing*, vol. 27, no.1, pp. 5-31, 2009.

[9] O. Jasmina, S. Remetić-Horvath and B. Buyukdemir, "Stock selection based on discriminant analysis: Case of capital market of Bosnia and Herzegovina," *J. of Econ. & Social Studies (JECOSS)*, vol. 4, no. 2, pp: 5-30, 2014.

[10] N. S. Ruzgar, and F. Unsal, "Comparison of recursive factor analysis and cluster analysis: A marketing application," *Int. J. of Econ. and Stat.*, vol. 1, .no. 4, pp. 183-190, 2013.

[11] N. L. Kumar and R. Devi, "Cluster analysis of mutual funds," *Int. J. of Multidisciplinary Research,* vol. 1, no. 5, pp. 24-47, Sep. 2010.

[12] K. Pendaraki, C. Zopounidis and M. Doumpos, "On the construction of mutual fund portfolios: A multi-criteria methodology and an application to the Greek market of equity mutual funds," *European J. of Oper. Res.,* vol. 163, no. 2, pp. 462-481, 2005.

[13] D. Archarya, and G. Sidana, "Classifying mutual funds in India: Some results from clustering," *Indian J.of Econ. and Bus.*, vol. 6, no. 1, pp. 71-79, 2007.

[14] M. Momcilovic, J. Njegic, and S. Jovin, "Risk, return and stock performance measures*,"* in Proc. *1st WSEAS Int. Conf. on Finance, Accounting and Auditing (FAA '12*), Zlin, Czech Republic, 2012, pp. 257-261.

[15] F. Lisi and E. Otranto, "Clustering mutual funds by return and risk levels," *CRENOS, Centre for North South Econ. Res., University of Cagliari and Sassari, Sardinia*, 2008. Available: http://crenos.unica.it/crenos/

[16] N. S. Ruzgar, "Classification of insurance companies in Turkey," *The Business Review, Cambridge*, vol. 12, no. 2, pp. 88-95, 2009.

[17] B. K. Wong, V. S. Lai and J. A. Lam, "Bibliography of neural network business applications research: 1994–1998," *Computers and Oper. Res.*, vol. 27, pp. 1045–1076, 2000.

[18] N. S. Ruzgar, F. Unsal and B. Ruzgar, "Predicting business failures using the rough set theory approach: The case of the Turkish banks," *Int. J. of Mathematical Models and Methods in Applied Sciences*, vol. 2, no. 1, pp. 57-64, 2008,

[19] Z. Pawlak, *Rough Sets: Theoretical aspects of reasoning about data*. Boston: Kluwer, 1991.

[20] S. Balcaen and H. Oghe, "Alternative methodologies in studies on business failure: Do they produce better results than the classic statistical methods?" vol. 16, 2004. Available: Vlerick Leuven Gent Management School Working Paper Series.

[21] D. Q. Miao, Y. Zhao and Y. Y. Yao, "Relative reducts in consistent and inconsistent decision tables of the Pawlak rough set model," *Inf. Science,* vol. 179, no. 24, pp. 4140–4150, 2009.

[22] K. Thangavel and A. Pethalakshmi, "Dimensionality reduction based on rough set theory: a review'" *Appl Soft Comput,* vol. 9, no. 1, pp. 1–12, 2009.

[23] J. Zhou, D.Q. Miao and W. Pedrycz, "Analysis of alternative objective functions for attribute eduction in complete decision tables," *Soft Comput,* vol. 15, no. 8, pp. 1601–1616, 2011.

[24] C. Gao, W. Pedrycz and D. Miao, "Rough subspace-based clustering ensemble for categorical data," *Soft Comput*, Springer-Verlag Berlin Heidelberg, vol. 17, pp. 1643–1658, 2013.

[25] M. Beynon, B. Curry and P. Morgan, "Knowledge discovery in marketing: An approach through rough set theory," *European J. of Marketing*, vol. 35, no. 7/8, pp. 915-935, 2001.

[26] B. Ruzgar and N. S. Ruzgar, "Rough sets and logistic regression analysis for loan payment," *Int. J. of Mathematical Models and Methods in Applied Sciences*, vol. 2, n. 1., 2008, Available: http://www.naun.org/journals/m3as/

[27] L. Shen and H.T. Loh, "Applying rough sets to market timing decisions," *Decision Support Systems*, vol. 37, pp. 583-597, 2004.

[28] N. Senthilkumaran and R. Rajesh, "Study on rough set theory for medical image segmentation," *Int. J. of Recent Trends in Eng.*, vol. 2, no. 2, November 2009.

[29] A. Øhrni Rosetta, *Technical Reference Manual*, 1999, Avaliable: http://www.idi.ntnu.no/_aleks/rosetta.