Comparison of Recursive Factor Analysis and Cluster Analysis: A Marketing Application

N. S. Ruzgar, F. Unsal

Abstract—Daily price movements for 264 Turkish mutual funds were investigated using two different approaches, recursive factor analysis and cluster analysis. Data for 1952 business days starting from January 1, 2004 until September 30, 2011 were used. With the recursive factor analysis, during stage 1, 2 factors were extracted and during the stage 2, 5 sub-factors were extracted. With the hierarchical cluster analysis using Pearson distance, however, the same data were grouped into 7 significant clusters. When comparing the results of both analyses, we found that the recursive factor analysis provided more meaningful grouping.

Keywords—Hierarchical cluster analysis, mutual funds, recursive factor analysis, trends of funds.

I. Introduction

Mutual funds are pooled funds that are divided into units of equal value and sold to investing public. These collective investment vehicles are professionally managed by a bank or other types of financial institutions. In today's dynamic environment many people prefer mutual funds because they feel that they can get higher return from these investments as opposed directly getting involved in the stock market. These funds are a collection of financial instruments including stocks, bonds, and other types. 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. Do the consumers really understand the differences between these funds? Are the financial institutions creating multiple funds for market segmentation purposes? Could hundreds of funds be grouped into a few classes? These are some of the questions investigated in this study.

A very large number of studies have been conducted for the US mutual funds. However, the number of studies conducted in the European countries and emerging countries are quite limited [1]. 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 who would like to sell a particular type of mutual fund and/or advertising online or traditional media [2], [3]. A more recent study has investigated the relationship between investors' financial knowledge and mutual fund advertising. The researchers concluded that mutual fund ads with financial disclosures are more likely to generate higher levels of recall and positive thoughts regarding advertised information [4]. This creates more favorable attitudes toward the mutual fund, and they are more likely to purchase it.

The main objective of this study is to investigate whether one could use recursive factor analysis technique to classify hundreds of mutual funds marketed in Turkey into a few groups. If this can be done, the consumers could make easier purchase decisions. After selecting a "group" that perhaps satisfies their investment criteria, they can then perhaps compare financial institutions that provide these funds, the cost, and other factors. A secondary objective was to apply cluster analysis to the same data to find out which method will group the funds more meaningfully.

II. Literature Review

Factor analysis and Cluster analysis will be the two main analytical techniques used in this study. A review of literature indicates these techniques are used in a number of fields for a variety of applications ranging from marketing, to finance, insurance, regional planning, and engineering. Multivariate statistical techniques, including factor analysis and multiple regression analysis, were used, for example, in a study that modeled the satisfaction of frequent customers of a group of hotels in Portugal [5]. Reference [6] shows how to analyze the marketing mix of the life insurance industry in India using factor analysis. An application of cluster analysis was conducted to test consumer behavior in a food market in the Czech Republic [7]. In another marketing study, researchers investigated the factors that determine customer satisfaction in durable products using a factor analysis approach [8].

Factor and cluster analysis has also been used in regional studies. For example, researchers in Croatia have made an attempt to create country cluster in the European Union (EU 27) after the more recent expansion [9]. In another regional study, Greece has been clustered in to 13 regions for strategic employment planning, as in [10]. A similar approach was followed by researchers in Czech Republic to assess competitiveness of regions using data from Eurostat [11].

A number of analytical techniques have been used by researchers in investigating financial market data. These techniques have included factor analysis, cluster analysis, time series, regression analysis, rough sets, fuzzy logic, and

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

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

discriminant analysis. For example, one study using factor and cluster analysis looked into investor security in the mutual fund market [12]. The style factors are identified as in [13]. A number of other studies have used factor analysis models to classify hedge funds [14]-[17]. Other studies, such as indexes of CTA returns have been used rather than the returns of individual managers, [13], [18]. Reference [19] identified specific characteristics important in establishing and maintaining mutually beneficial relationships between endowment and pension fund managers and the providers of investment management services. Reference [20] shows how IFAs perceive they add value to the decision making of consumers when purchasing pension products in order to understand how they compete and the nature of the strategic groups within this channel by using cluster analysis. Another study investigated private pension funds using cluster and factor analysis [21].

The mutual funds market in India, for example, has been studied using cluster analysis [22]. Similarly, the Greek mutual fund market has been investigated using a multi-criteria methodology [23]. Another study using factor and cluster analysis looked into the risk factor in classifying mutual funds as in [24]. Reference [25] shows how to measure the performance of closed–ended mutual funds and classified them into several categories for the guidance of the investors. A 2012 study evaluated the risk, return and performance measures of selected stocks traded in Belgrade Stock Exchange in Serbia using regression analysis [26]. A study conducted in the Euro area using Bloomberg data for the 2002-2008 period, researchers in Italy have classified sticks into cluster using a three-stage pure statistical analysis with a great deal of success [27]. Cluster analysis was utilized to categorize the huge amount of equity mutual funds into several groups based on four evaluation indices, namely, rates of return, standard deviation, turnover rate, and Treynor index, in order to aid investors in making the investment decision [28]. These researchers proposed fuzzy optimization model to determine the optimal portfolio.

As reported above, a great deal of research is conducted to test the performance of the market, to classify financial instruments, and to investigate how different statistical techniques can be used for the analysis. Few studies look at the consumer side. Do the investors really understand the technical nature and expected performance of the mutual funds that they purchase? Do they always try to maximize their returns? How do they, in fact, decide which mutual funds to purchase? One study has focused on these points [29]. The findings indicate that investors consider many nonperformance related variables. Using cluster analysis, the researchers have classified the investors into several groups. Only a small segment of the market was found to be quite knowledgeable about the funds that they purchased. Most investors, however, were found to be rather naive, having little knowledge of the investment strategies or financial details of their investments.

As was mentioned earlier, this particular study will utilize recursive factor analysis and cluster analysis to classify mutual funds in Turkey. The reader can refer to [30]-[39] among others for a description of these methods used in this paper.

III. THE THEORICAL MODEL

A. Recursive Factor Analysis

Let the data input matrix ***X*** consist of p observations and n variables. The correlation matrix of **X**, **R**=Corr(**X**), will be an n*x*n symmetrical matrix. The eigenvalues of this symmetrical matrix, λ_i, are the roots of the characteristic equation det(λ**I** - **R**)=0. It is clear that expanding and simplifying the n*x*n determinant det(λ**I** - **R**) yields a polynomial of degree n in which the coefficient of λ^n is 1; that is, det(λ**I** - **R**) is of the form

$$\det(\lambda \mathbf{I} - \mathbf{R}) = \lambda^n + c_1\lambda^{n-1} + c_2\lambda^{n-2} + \cdots + c_n$$

where $c_1, c_2, \ldots, c_n$ are arbitrary constants and

$$(\lambda \mathrm{I} - \mathrm{R}) = \begin{bmatrix} \lambda & 0 & \cdots & 0 \\ 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda \end{bmatrix}_{n \times n} - \begin{bmatrix} 1 & r_{12} & \cdots & r_{1n} \\ r_{12} & 1 & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1n} & r_{2n} & \cdots & 1 \end{bmatrix}_{n \times n} = \begin{bmatrix} \lambda-1 & -r_{12} & \cdots & -r_{1n} \\ -r_{12} & \lambda-1 & \cdots & -r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -r_{1n} & -r_{2n} & \cdots & \lambda-1 \end{bmatrix}_{n \times n}$$

This is called characteristic polynomial of **R** [40]. The polynomial equation has n roots ranging from λ_1 to λ_n. Each one of these roots is responsible in explaining a certain percentage of the total variance. Thus, the value of each λ_i explaining the percentage of the total variance can be denoted as:

$v_i = \left(\frac{\lambda_i}{\lambda_1 + \lambda_2 + \cdots + \lambda_n} \right) \times 100$. It is generally accepted in factor analysis literature that $\lambda_i \geq 1$ values explain a larger proportion of the variance while $0 < \lambda_i < 1$ values explain a very insignificant proportion of the total variance. In factor analysis, the specification of q factors in explaining **X** input matrix should be based on the potential impact on total variance. Therefore, q roots that satisfy $\lambda_i \geq 1$ for the matrix, **X,** is kept in the analysis while the roots (n-q of them) based on $0 < \lambda_i < 1$ are left out for further analysis. Thus, the **X** matrix with n variables is grouped into q factors. The factor loads needed in specifying the factors are computed from $\ell = \sqrt{\lambda}\, e$ relationship. Here the eigenvalues are shown as ($\lambda_1, \lambda_2, \ldots, \lambda_q$) and **e** eigenvectors are shown as ($\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_q}$). For example, the factor loads for factor 1 are computed as follows:

$$\ell_{11} = \sqrt{\lambda_1}\, e_{11}, \ell_{12} = \sqrt{\lambda_1}\, e_{12}, \ell_{13} = \sqrt{\lambda_1}\, e_{13}, \cdots, \ell_{1q} = \sqrt{\lambda_1}\, e_{1q}$$

Similarly, the factor loads for factor 2 can be computed as follows:

$$\ell_{21} = \sqrt{\lambda_2}\, e_{21}, \ell_{22} = \sqrt{\lambda_2}\, e_{22}, \ell_{23} = \sqrt{\lambda_2}\, e_{23}, \cdots, \ell_{2q} = \sqrt{\lambda_2}\, e_{2q}$$

If we continue with this line of argument, the factor loads for factor q can be determined as follows:

$$\ell_{q1} = \sqrt{\lambda_q}\, e_{q1}, \ell_{q2} = \sqrt{\lambda_q}\, e_{q2}, \ell_{q3} = \sqrt{\lambda_q}\, e_{q3}, \cdots, \ell_{qq} = \sqrt{\lambda_q}\, e_{qq}.$$

It should be noted, however, that q factor groups created from n variables may not be always meaningful due to the nature of factor analysis. In other words, the groups created after analysis may not always confirm the expectations. This is expected due to the characteristics of eigenvalues. For example, if we consider the **X** input matrix first and obtain the $\mathbf{R}_{nxn}$ correlation matrix, and then remove some of the variables or extract sub-matrices (removing rows and columns for the same variables), and finally obtain a new correlation matrix, the resulting eigenvalues will be different and hence we will end up with different factor groupings. In other words, the $\mathbf{R}_k$ sub-correlation matrix and the eigenvalues, λ_j, of the resulting ($\lambda\mathbf{I} - \mathbf{R}_k$) matrix will be different than the original eigenvalues. As a result of this, the eigenvectors, $\mathbf{e}_j$, specified from λ_j eigenvalues will be different. Likewise, the factor loads computed from $\ell = \sqrt{\lambda}\, e$ will be different as well. The newly computed factor loads will result in different sets of factor classifications. Thus, we can conduct factor analysis in a recursive manner and come to an ideal solution after several stages of analysis. For example, we can start with factor analysis of the original data obtain the factor groupings, and then take each factor specified and conduct a second factor analysis for each group determined in stage 1. During this second stage, one can drop certain variables from further analysis. This process can be repeated until the most meaningful groups are formed.

B. Cluster Analysis

Cluster analysis groups data objects into clusters in such a way that the objects belonging to the same cluster are similar, while those belonging to different ones are dissimilar [41]. In other words, the objective of cluster analysis is to assign observations to groups so that observations within each group are similar to one another with respect to variables or attributes of interest, and the groups themselves stand apart from one another [39]. The groups or clusters should be as homogeneous as possible and the differences among the various groups as large as possible. Conducting cluster analysis can be divided into two fundamental steps, choice of a proximity measure and choice of group-building algorithm [42]. The proximity between data points is measured by a distance or similarity matrix. For the proximity, Euclidean distance, Square Euclidean distance, Manhattan Distance, Pearson correlation distance measures etc. can be used. Meanwhile, there are essentially two types of clustering methods: hierarchical algorithms and nonhierarchical algorithms. The main difference between the two clustering techniques is that in hierarchical clustering once groups are found and elements are assigned to the groups, this assignment cannot be changed. In nonhierarchical techniques, on the other hand, the assignment of objects into groups may change during the algorithm application.

The hierarchical algorithms can be divided into agglomerative and splitting procedures. The first type of hierarchical clustering starts from the finest partition possible (each observation forms a cluster) and groups them. The second type starts with the coarsest partition possible: one cluster contains all of the observations. It proceeds by splitting the single cluster up into smaller sized clusters [38], [39]. For agglomerative hierarchical clustering procedures, one can use single linkage-nearest neighbor method, average linkage method, complete linkage method, McQuitty linkage method, Centroid linkage method, Ward linkage method, etc. In clustering analysis, to evaluate the distances between continuous variables in data matrix generally Euclidean distance or Square Euclidean distance methods are suggested. However, to measure the relationship between two or more variables one can focus on a correlation analysis with Pearson correlation distance method. The most widely-used type of a correlation coefficient is Pearson correlation coefficient [37].

In the case of more data points, a visualization of the implication of clusters is desirable. A graphical representation of the sequence of clustering is called a *dendrogram*. It displays the observations, the sequence of clusters and the distances between the clusters. The vertical axis displays the indices of the points, whereas the horizontal axis gives the distance between the clusters. Large distances indicate the clustering of heterogeneous groups. Thus, if we choose to "cut the tree" at a desired level, the branches describe the corresponding clusters. If the distances between the groups are desired to be minimum, "between group linkage method" is suggested, while the distances between groups are desired to be maximum and the relationship in the within the group is high, "within group linkage method "is suggested [36], [38], [39], [41].

IV. METHODOLOGY

In this study, daily mutual fund price data were collected from the Turkish market for the January 1, 2004 to September 30, 2011 a period, yielding 1952 data points [43]. At the present time, there are 299 funds available. However, those funds with less than 500 observations (business days) were excluded from the analysis. As a result, 264 funds were used for further analysis. Of these 264, 106 were defined as A-type funds and the remaining 158 were defined as B-type funds. The A-type funds included: variable fund (43), index fund (20), share fund (19), contributory fund (2), mixed fund (15), private fund (5), private sector fund (1) and foreign securities equity fund (1). The B-type funds included: gold and security fund (7), variable fund (54), index fund (2), fund basked (1), private fund (2), mixed fund (1), liquate fund (46), government bonds (41) and foreign securities equity fund (4).

It was 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. Thus, the main objective of this study is to find out whether the 264 mutual funds could be grouped into groups based on their similarities using both factor analysis technique and cluster analysis. A secondary objective was to determine the ideal number of groups. The third objective of this study is to find out which method will group the funds more sensitively. A final objective of the study was to determine whether the funds having the same trends according to daily fund prices are included in the same factor groups.

The first aspect of this study was to apply what we coined "Recursive Factor Analysis." To the best of our knowledge, this is the first time this procedure has been applied in data analysis. In summary, this procedure involves applying factor analysis to the data first, examine the results, and if the resulting factor groupings did not meet the expectations given the information about the market and each mutual fund, continue with the second stage. Here each of the factors obtained from the first step are subjected to another factor analysis. In other words, if the original factor analysis yielded three factors A, B, and C with less than desirable groupings, factor analysis was conducted for each one of the above factors. This recursive (stepwise) procedure was repeated until ideal groups were formed. The following paragraphs summarize these steps in more detail for the Turkish data we analyzed.

- The evaluation of the 264 mutual funds was performed in three stages. First, principal component analysis was utilized. This method seeks values of the loading that bring the estimate of the total communality as close as possible to the total of observed variances [44]. Then, Varimax rotation method, which seeks the rotated loading that maximize the variance of the squared loading for each factor was used. In this study, Varimax method was applied to 264 mutual funds using their prices for the 1952 business days under investigation. During the evaluation using factor analysis, the weight of a fund in any factor loads higher than 0.60 was included into that factor as suggested in the literature. Two factors were created at this stage.
- In the second stage, factor analysis was reapplied to factor 1 and factor 2 where dense accumulation occurred. Factor 1 was first divided into 3 sub-factors and factors 1.1, 1.2, and 1.3 were established. Then, factor 2 was also divided into 2 sub-factors and grouped as factor 2.1, and factor 2.2. The structure of this procedure is provided as a tree diagram in Fig. 1.
- In the third stage, trends of all funds were examined to determine if the funds having similar trends were included in the same factor groups. It was assumed that the different mutual funds were affected from the economical events in the same manner for the 1952 business days of the study period. Thus, no corrections were made for economic fluctuations during the study period.

The second aspect of this study was the creation of groups that using the hierarchical cluster analysis. We used Pearson correlation distance method for the proximity, and the average linkage method for the hierarchical clustering. The reason for the choice Pearson correlation distance method was to compare the findings of factor analysis and cluster analysis easily since the correlation matrix is also used for the factor analysis. A compromise method is average linkage, under which the distance between two clusters is the average of the distances of all pairs of observations, one observation in the pair taken from the first cluster and the other from the second cluster.

V. Application

When conducting factor analysis, in deciding the number of factors to be formed, usually cumulative explained variance percentages of initial eigenvalues are considered. If “q” eigenvalues explains 80-90 percent of total variance, q factors are determined. Alternatively, observation of the scree plots may be used for this determination. One of the approaches to determine the number of factors is suggested by Kaiser and here it is suggested that one should give preference to eigenvalues above zero. The other approach is to determine a factor numbers based on eigenvalues above one [38].

An observation of the scree plot indicated that the ideal number factors during stage 1 should be equal to 2. Besides, a large proportion of the total variance was explained by these two factors. If we do not consider the factors whose factor loads are below 0.60, analysis will stop here necessary calculations after 2 factors are not made. According to Kaiser-Meyer-Olkin Measure of Sampling Adequacy, the sufficient explanatory level found 0.97 for the first factor analysis and 0.98 and 0.98 for the recursive factor analyses, respectively. These levels show that data are suitable for principal components analysis, because the KMO values exceed the heuristic of 0.70, indicating that the correlations are adequate for factor analysis. Our initial factor analysis indicates that 2 factors explain 94 percent of the total variance. It was also observed that 152 funds placed under factor 1 explained 87.5 percent of the total variance, 88 funds under factor 2 explained 6.5 percent of the total variance. Similarly, three sub-factors extracted from factor 1 explain 97 percent of the total variance and two sub-factors extracted from factor 2 explain 87 percent of the total variance.

The initial findings from the factor analysis are summarized in a cross-tabulation format in Table I. The “A type funds” seen as a column variable are heavily weighed in stock funds with heavier risk factor. In contrast, the “B type funds” are heavily weighed in government bonds and similar instruments with lower risk.

An examination of Table I indicates that out of 264 funds investigated, 152 were classified under factor 1 (58 percent), 88 were classified under factor 2 (33 percent), and 24 were classified under neither one (9 percent). After a closer review of the contents of factor 1, one can easily conclude that it mainly consists of B-type funds. It includes 130 B-type funds and only 22 A-type funds. A similar review factor 2 indicates that it attracts mostly A-type funds (77 of them) versus a small number of B-type funds (only 11 of them). It was also noted 10 out of 11 of these B-type funds were somewhat similar to A-type funds.

Table I. Crosstabulation of Factors and Fund Types after the First Stage of the Factor Analysis

MUTUAL FUNDS			FUNDS		Total
			A TYPE FUNDS	B TYPE FUNDS	
FACTORS	Factor 1	Count	22	130	**152**
		% within FACTORS	14%	86%	100%
		% within FUNDS	21%	82%	58%
	Factor 2	Count	77	11	**88**
		% within FACTORS	88%	12%	100%
		% within FUNDS	73%	7%	33%
	Excluded Factors	Count	7	17	**24**
		% within FACTORS	29%	71%	100%
		% within FUNDS	6%	11%	9%
Total		Count	**106**	**158**	**264**
		% within FACTORS	40 %	60%	100%
		% within FUNDS	100%	100%	100%

In the second stage of the analysis, each one of the above factors was investigated further through factor analysis to see whether other sub-factors could be extracted from them. When factor analysis was applied to the original Factor 1, three sub-factors were extracted from it. The results are summarized in Table II. When factor analysis was applied for the original factor 2, two sub-factors were extracted. The results are summarized in Table III.

An examination of Table II indicates that Factor 1.1 includes 41 liquid funds and 16 variable funds out of a total 77 funds. In other words, 74 percent of the funds in Factor 1 are either liquid or variable funds ((41+16)/77). Further examination shows that 96 percent of the liquid funds (41/43) and 100 percent of the gold funds (7/7) are included in Factor 1.1. There are 51 funds in Factor 1.2. Of these, 94 percent are either variable funds or bonds (48/51). Further examination shows that 42 percent of the variable funds (20/47) and 74 percent of the bonds (28/38) are included in Factor 1.2.

Table II. Sub-Factors Extracted from Factor 1

FACTOR 1 TYPE FUNDS (22-A;130-B)			FUNDS					Total
			Variables Funds	Liquid Funds	Bonds	Gold Funds	Others	
FACTORS	Factor 1.1	Count	3-A;13-B	41	10	7	3	**77**
		% within FACTORS	4-A;17-B%	53%	13%	9%	4%	100%
		% within FUNDS	6-A;28-B%	96%	26%	100%	18%	50.7%
	Factor 1.2	Count	2-A;18-B	1	28	0	2	**51**
		% within FACTORS	4-A;35-B %	2%	55%	0%	4%	100%
		% within FUNDS	4-A;38-B %	2%	74%	0%	12%	33.6%
	Factor 1.3	Count	4-A;4-B	0	0	0	11-A;1-B	**20**
		% within FACTORS	20-A;20-B %	0%	0%	0%	55-A;5-B%	100%
		% within FUNDS	8.5-A;8.5-B%	0%	0%	0%	64-A;6-B%	13.1%
	Excluded Factors	Count	2-A;1-B	1	0	0	0	4
		% within FACTORS	50-A;25-B%	25%	0%	0%	0%	100%
		% within FUNDS	4-A;2-B%	2%	0%	0%	0%	2.6%
Total		Count	**(11-A;36-B) 47**	**43**	**38**	**7**	**(11-A;6-B) 17**	**152**
		% within FACTORS	7-A;24-B%	28%	25%	5%	11%	100%
		% within FUNDS	100%	100%	100%	100%	100%	100%

There are 20 funds in Factor 1.3 and of these 75 percent are A-type funds (15/20). It was shown earlier that a total of 22 A-type funds were included in Factor 1 (Table I). Of these 22, 15 have been included in Factor 1.3. In other words, 68 percent of the A-type funds have been included in Factor 1.3 (15/22). One might recall that 86 percent of the funds in Factor 1 were B-type and the remaining 14 were A-type. When the second stage of factor analysis was conducted, it was noted these A-type funds are forming their own group under Factor 1.3. This is one of the unique advantages of recursive factor analysis suggested in this paper. The sub-groups can provide more uniform groups as seen here. There were 4 funds that could not be classified under any of the sub-groups. Three of these were variable funds and one liquid fund.

Table III. Sub-Factors Extracted from Factor 2

FACTOR 2 TYPE FUNDS (77-A;11-B)			FUNDS					Total
			Variables Funds	Index Funds	Share Funds	Mixed Funds	Others	
FACTORS	Factor 2.1	Count	20-A;5-B	9	11	4	4	**53**
		% within FACTORS	38-A;9-B%	17%	21%	7.5%	7.5%	100%
		% within FUNDS	50-A;12.5-B%	47%	69%	57%	66%	60%
	Factor 2.2	Count	7-A;3-B	10	5	3	1-A;1-B	30
		% within FACTORS	24-A;10-B%	33%	17%	10%	3-A;3-B%	100%
		% within FUNDS	17.5-A;7.5-B%	53%	31%	43%	17-A;17-B%	34%
	Excluded Factors	Count	3-A;2-B	0	0	0	0	**5⁺**
		% within FACTORS	60-A;40-B%	0%	0%	0%	0%	100%
		% within FUNDS	7.5-A;5-B%	0%	0%	0%	0%	6%
Total		Count	**(30-A;10-B) 40**	**19**	**16**	**7**	**(5-A;1-B) 6**	**88**
		% within FACTORS	34-A;11-B%	22%	18%	8%	6-A;1-B%	100%
		% within FUNDS	100%	100%	100%	100%	100%	100%

A review of Table III shows that out of 53 funds in Factor 2.1, 85 percent (45/53) consist of variable, index, and share funds. Further examination from the fund point of view, one can see that 62.5 percent of the variable funds (25/40), 69 percent of the share funds (11/16), and 57 percent of the mixed funds (4/7) are grouped under Factor 2.1. Out of 30 funds classified under Factor 2.1, 67 percent are variable and index funds (20/30). One can also see that 53 percent of the index funds (10/19) and 43 percent of the mixed funds (3/7) are grouped under Factor 2.1. It can also be noted all 5 funds that could not be classified in neither of the factors are variable funds.

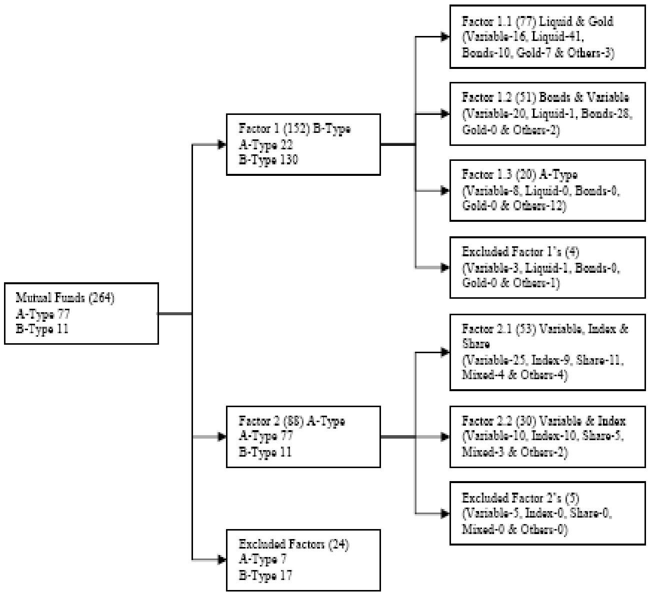

Fig. 1 Two stages of factor analysis with factor classifications

The above discussion is summarized in a tree diagram format in Fig. 1. It shows that the initial factor analysis with 264 mutual funds, created Factor 1 with 152 funds and Factor 2 with 88 funds. At this stage, 24 funds were excluded from the classification. During the second stage of the analysis, factor analysis was applied to both of these factors to obtain sub-factors within each group. Thus, three sub-factors were obtained under Factor 1 and 2 sub-factors under Factor 2.

As can be seen in the above diagram, certain funds were not classified under none of the factors. For example one can observe variable funds in Factor 1.1, Factor 1.2, Factor 1.3, Factor 2.1 and Factor 2.2. Nevertheless, there were other variable funds that could not be grouped under none of the above sub-factors.

Similarly, according to the results of hierarchical cluster analysis to find out the relationship among the mutual funds, the funds are grouped into maximum 146 clusters. Since most of the groups included one observation and the rest include 2 or three observations, it was decided that the number of clusters should be 7, even if it was not perfect. The *dendrogram* obtained from hierarchical cluster analysis was re-graphed by keeping the distance measures same. The resulting *dendrogram* is illustrated in Fig. 2 below.

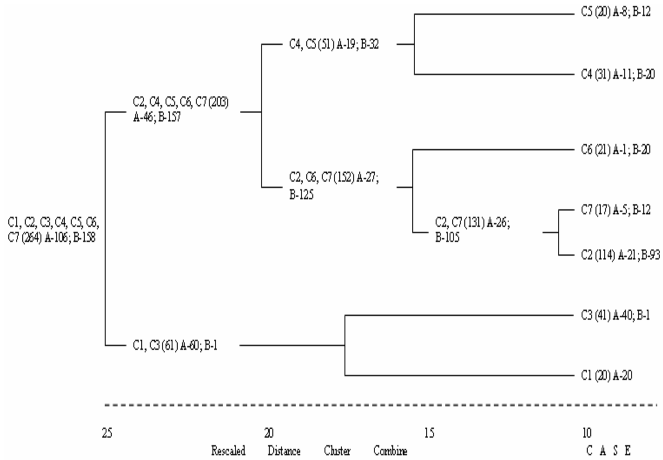

Fig. 2 Hierarchical cluster analysis dendrogram [using average linkage (within group)]

As shown in Fig. 2 above, when the funds are divided into two clusters, 57% (60/106) of A-type funds and approximately 0% (1/158) are in C1 and C3. It can also be observed that 43% (46/106) of A-type funds and approximately 100% (157/158) are grouped in C2, C4, C5, C6 and C7. This means that most of the A-type funds are grouped in C1 and C3 and almost all of B-type funds are grouped in C2, C4, C5, C6 and C7. Detailed contents of clusters are summarized in Table IV.

Table IV. The Distribution of Funds by Clusters

Clusters	A-Type (106)						B-Type (158)						
	Variable (43)	Index (20)	Share (19)	Mixed (15)	Others (9)	Sum (A)	Variable (54)	Liquid (46)	Bonds (41)	Gold (7)	Others (10)	Sum (B)	Sum (A+B)
C1	7	10	2	-	1	20	-	-	-	-	-	0	20
C3	18	2	8	9	3	40	1	-	-	-	-	1	41
C2	4	3	5	6	3	21	29	31	24	4	5	93	114
C7	3	1	1	-	-	5	4	3	4	-	1	12	17
C6	-	-	-	-	1	1	8	4	7	-	1	20	21
C4	7	2	1	-	1	11	7	6	5	-	2	20	31
C5	4	2	2	-	-	8	5	2	1	3	1	12	20
C2, C7	7	4	6	6	3	26	33	34	28	4	6	105	131
C2, C6, C7	7	4	6	6	4	27	41	38	35	4	7	125	152
C4, C5	11	4	3	-	1	19	12	8	6	3	3	32	51
C1, C3	25	12	10	9	4	60	1	-	-	-	-	1	61
C2, C4, C5, C6, C7	18	8	9	6	5	46	53	46	41	7	10	157	203
C1, C2, C3, C4, C5, C6, C7	43	20	19	15	9	106	54	46	41	7	10	158	264

When comparing the results from the two analytical tools, one can see that cluster C1 with variable and index funds is similar to Factor 2.2; cluster C3 with variable, share and mixed funds is similar to Factor 2.1; cluster C2 with variable liquid and bonds funds is similar to Factor 1.1 and Factor 1.2. While cluster C6 contains only B-type funds except one A-type fund, C7, C4, and C5 contain both A-type funds (26 percent) and B-type funds (74 percent). Although clustering into 7 groups seemed reasonable, the mutual funds are not well grouped as expected. Perhaps one needs to examine the structure of each fund for better understanding.

The reasons for the exclusion of certain funds during analysis were investigated by analyzing the rules and regulation of each fund issuing organizations in terms of how they define the lower and upper limits of financial instrument in a particular mutual fund. It was noted that each institution defined content limits of a mutual fund differently and this might have caused the exclusion certain types of funds in the factor and cluster analysis process. Several financial institutions (Isbank, Garanti Bankasi, Vakifbank, Yapi Kredi Bankasi, ING Bank ve Global Menkul Degerler) with large mutual fund volume were investigated for rules and regulations in defining the content of a fund and the findings are summarized in Table V and Table VI below. When reviewing these tables, one should note that in each cell the lower and upper limits of the content are provided. For example, for a "Variable Fund," the stock content of Isbank ranges from 0 to 20 percent while the stock content ranges from 0 to 100 percent for Garanti Bank. The freedom of the intuitions to define their funds with liberal ranges might have caused different funds to bring in similar returns.

Table V. The Content Distribution of B-Type Funds Issued by Major Banks
(Minimum and Maximum Percentage Limits Provided in Each Cell)

B Type Funds (n=158)	Variable (n=54)				Liquid (n=46)				Bonds (n=41)				Gold (n=7)			
	Isbank	Garbank	Vakbank	Yapikredibank	Isbank	Garbank	Vakbank	Yapikredibank	Isbank	Garbank	Vakbank	Yapikredibank	Isbank	Garbank	Vakbank	Yapikredibank
Share	0 20	0 100	0 100	0 100					0 5	0 49	0 49	0 49	0 49	0 49	0 49	0 49
Government Bonds	0 100	0 100	0 100	0 100	0 100	0 100	0 100	0 100	51 100	51 100	51 100	51 100	0 49	0 49	0 49	0 49
Private Sector Bonds	0 100	0 100	0 100	0 100	0 100	0 100	0 100	0 100	0 100	0 49	0 100	0 49	0 49	0 49	0 49	0 49
Repo	0 100	0 100	0 100	0 100	0 100	0 100	0 100	0 100	0 49	0 49	0 49	0 49	0 49	0 49	0 49	0 49
Foreign Government and Private Sector Instruments	0 25	0 100	0 100	0 100	0 10			0 100	0 25	0 49	0 49	0 49	0 49	0 49	0 49	0 49
Accession Partnership Documents	0 20	0 10			0 20				0 20	0 10				0 10		
Cash	0 20			0 20	0 20			0 20	0 20			0 20				0 20
Gold and precious metals	0 5	0 100	0 100	0 100					0 5	0 49	0 49	0 49	51 100	51 100	51 100	51 100
Real estate investment trust		0 10	0 40	0 100			0 40		0 10	0 10	0 40	0 49	0 10	0 10	0 49	0 49
Stock Market		0 20				0 20				0 20			0 10	0 20		0 20
Securities		0 100		0 100				0 100		0 49		0 49		0 49		0 49

Table VI. The Content Distribution of A-Type Funds Issued by Major Banks
(Minimum and Maximum Percentage Limits Provided in Each Cell)

A Type Funds (n=106)	Variable (n=43)				Share (n=19)				Index (n=20)				Mixed (n=15)			
	Isbank	Garbank	Vakbank	Globalmenk	Isbank	Garbank	ING bank	YapiKredibank	Isbank	Garbank	Vakbank	Yapikredibank	Isbank	Garbank	Globalmenk	Yapikredibank
Share	25 90	25 100	50 100	25 95	51 100	51 100	51 100	51 100	80 100	80 100	50 100	80 100	20 80	20 80	25 80	20 80
Government bonds	0 75	0 75	0 49	0 75	0 49	0 49	0 49	0 49	0 20	0 20	0 49	0 20	20 80	20 80	0 55	20 75
Private bonds	0 75	0 75	0 49	0 50	0 49	0 49	0 49	0 49	0 20	0 20	0 49	0 20	0 60	0 60	0 55	20 75
Repo	0 75	0 75	0 49	0 75	0 49	0 49	0 49	0 49	0 20	0 20	0 49	0 20	0 60	0 60	20 75	0 75
Foreign investment	0 10	0 75	0 49		0 49	0 49	0 40	0 25		0 20	0 25	0 20	0 60	0 60		
Accession Partnership Documents	0 20	0 10			0 20	0 10				0 10			0 20	0 10	0 10	
Cash	0 20				0 20			0 20	0 5			0 20	0 20		0 5	0 20
Gold and precious metals	0 5	0 75	0 49		0 5	0 49		0 49		0 20	0 49	0 20	0 10	0 60		
Real estate investment trust	0 10	0 10	0 49	0 10	0 10	0 10		0 49	0 10	0 10	0 49	0 20	0 10	0 10	0 10	0 55
Stock market						0 20				0 20			0 20	0 20		
Securities								0 49		0 20		0 20				

After grouping the data by recursive factor analysis, a question may arise. Is the recursive factor analysis determining if the relationships between the variables in the hypothesized model resemble the relationship between the variables in the observed data set? More formally, does the analysis determine the extent to which the proposed covariance matches the observed covariance? To answer this question, we applied to the confirmatory factor analysis and discovered that the proposed covariance matched the observed covariance.

An attempt was made to show the daily fund prices for all 264 funds to obtain a visual map of the data points to see whether there were any visible patterns or not. The graphs were scattered all over the page and no patterns could be observed.

Then the graphs were drawn for the funds that were grouped in Factor 1 (Fig. 3) and Factor 2 (Fig. 4). Again the reader should note that these graphs were introduced only to show the general trends in a visual format rather than exact graphs. An inspection of the graphs presented in Figure 3 and Figure 4 showed distinct patterns. It was clear that the funds grouped in Factor 1 showed similar patterns. Similarly, the funds grouped under Factor 2 showed similar patterns. Yet the two sets of graphs were much different from each other. It looked as if we could get the graph with 264 funds if we were to superimpose Fig. 3 and Fig. 4.

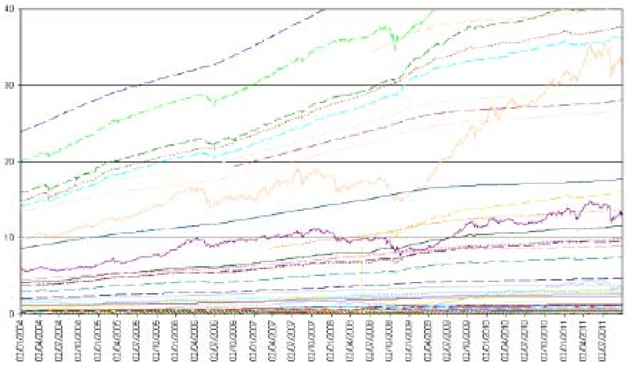

Fig. 3 Factor 1 Investment Funds (n=152)

The same graphical exercise was conducted for all of the 5 sub-factors. The examination of these five charts also indicated there were similarities of trends within a particular sub-factor while there were wide differences between charts of different sub-factors.

Fig. 4 Factor 2 Investment Funds (n=88)

VI. CONCLUSIONS AND FURTHER STUDIES

In this study, two methods, the recursive factor analysis and hierarchical cluster analysis, were applied to the daily prices of 264 mutual funds with 1952 data points, to classify them into ideal number of groups. When conducting factor analysis creating groups that are similar in nature, the researcher also needs to investigate whether these groups are meaningful or not for the particular industry studied. If the first factor analysis applied to the data does not provide enough factors and/or if the factors formed are meaningful in explaining the real word, the research should not just stop. One can apply further factor analysis to the factors created in the first stage. This kind of recursive process might yield much better results.

The initial factor analysis provided 2 factors. Attempts to create the third or fourth factors from the original data were unsuccessful. As a result, there was a need to apply a second of factor analysis to the factors identified in the first stage. This process provided three sub-factors in Factor 1 and two sub-factors in Factor 2. The researcher then tried further factor analysis to each one of the five sub-factors with no success. Thus, they ended the study and investigated the properties of the five sub-factors determined in stage 2.

When the hierarchical cluster analysis was applied to 1952 data points, the funds were grouped into maximum 146 groups. Since some of the groups have one data point and some of the others have two or three data points, it was decided that the ideal number of groups by cluster analysis was 7. When the results of recursive factor analysis and cluster analysis are compared, it was seen that some of the factors from recursive factor analysis coincided to the groups from cluster analysis, such as C1 and Factor 2.2, C3 and Factor 2.1, C2 and Factor 1.1 and 1.2. However, some groups and factors did not match each other. Although both analyses factored or grouped the data into seven groups, we discovered that the data extracted by recursive factor analysis was more meaningful.

The main conclusion of this study is that recursive factor analysis where one can apply further factor analysis into factors created in an earlier stage is a useful analytical tool that can help more meaningful groups.

REFERENCES

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

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

[3] M. A. Jones, V. P. Lesseig, T. I. Smythe, V.,Taylor, “Mutual fund advertising: Should investors take notice?”, *J. Financial Services Marketing*, vol. 12, no. 3, pp. 242-254, 2007.

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

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

[6] M. C. Garg, A. Verma,, "An empirical analysis of marketing mix in the life insurance industry in India", *IUP J. Management Research*, vol. IX, no. 2, 2010, pp. 7-21.

[7] P. Turcinek, J. Stastny, A. Motycka, "Usage of cluster analysis in consumer behaviour research", *Proc. 12th WSEAS Int. Conf. on Applied Informatics and Communications (AIC '12),* Istanbul, Turkey, 2010, pg. 172-177.

[8] R, A. Krishnan, K. Hari, "Factors determining customer satisfaction in consumer durable white goods: factor analysis approach", *Asia Pacific J. Res. in Business Management*, vol. 2, no. 6, Jun2011, pp. 32-46.

[9] K, N. Zivadinovic, K. Dumicic, A. C. Casni, "Multivariate analysis of structural economic indicators for Croatia and EU 27", *Proc. 2nd WSEAS Int. Conf. on Multivariate Analysis and its Application in Science and Engineering (MAASE 2009),* Istanbul, Turkey, 2009, pg. 134-139.

[10] P. Michaelides, G. Economakis, D. Logos, "Strategic regional planning and regional employment in Greece: A clustering analysis approach", *Proc. IASME/WSEAS Int. Conf. on Energy, Environment, Ecosystems and Sustainable Development*, Athens, 2005.

[11] P. Petr, J. Krupka, R. Provaznikova, "Statistical approach to analysis of the regions", *Proc. 10th WSEAS Int. Con. on Applied Computer Science (ACS'10)*, Wisconsin, USA, 2010, pg. 280-285.

[12] G. Sidana, D. Acharya, "Classifying mutual funds in India: Some results from clustering", *Indian J. Economics and Business*, 2007.

[13] W. K. H. Fung, D. A. Hsieh, "Asset-based style factors for hedge funds", *Financial Analyst J.*, vol. 58, no. 1, 2002, pp. 16-27.

[14] Agarwal, V., Naik, N, "Generalised style analysis of hedge funds", *J. Portfolio Management*, vol. 1, no. 1, 2000, pp. 93-109.

[15] S. J. Brown, W. N. Goetzmann, "Hedge funds with style", *J. Portfolio Management*, vol. 29, no. 2, 2003, pp. 101-112.

[16] N. Das, "Hedge fund classification using k-means clustering method", *9th Int. Conf. on Computing in Economics and Finance*, Seattle, July 11-13, 2003.

[17] R. Gibson, S. Gyger, "The style consistency of hedge funds", *European Financial Management*, vol. 13, no. 2, 2006, pp. 287-308.

[18] E. Noma, A. Alibair, W. T. Long, "*Time Frames, Research Quality and Strategy: The Differentiating Factors for CTAs? Directory of Funds of Hedge Funds"*, Barclay Hedge, 2008.

[19] J. C. Strieter, S. Singh, "The determinants of acquisition of outside investment management service providers in public and corporative pension plans and endowments", *Int. J. Bank Marketing,* vol. 23, no. 2/3, 2005, pp. 218.

[20] O. Gough, "Independent financial advisers – Why they remain the strongest distribution route for pensions", *Service Industries J.*, vol. 25, no. 5, 2005, pp. 709–720.

[21] B. Ruzgar, F. Unsal, "Evaluation of private pension funds with factor and cluster analysis", *Business Review*, Cambridge, vol.13, no. 1, 2009, pp. 120-128.

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

[23] K. Pendaraki, C. Zopounidis, 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. Operational Research,* vol. 163, no. 2, 2005, pp. 462-481.

[24] D. Archarya, G. Sidana, "Classifying mutual funds in India: Some results from Clustering", *Indian J. Economics and Business*, vol. 6, no. 1, 2007, pp. 71-79.

[25] K., S. Iqbal, A. Zaheer, S., S. M. Amir, "Performance evaluation of close-ended mutual funds by investment objectives in Pakistan's economy", *Proc. Int. Conf. on Applied Computer Science (ACS'2010)*, Malta, 2010, pp. 96-107.

[26] M. Momcilovic, J. Njegic, S. Jovin, "Risk, return and stock performance measures*", Proc. 1st WSEAS Int. Conf. on Finance, Accounting and Auditing (FAA '12),* Zlin, Czech Republic, Sep. 20-22, 2012, pg 257-261.

[27] F. Lisi, E. Otranto, "Clustering mutual funds by return and risk levels", *CRENOS*, Centre for North South Economic Research, University of Cagliari and Sassari, Sardinia , 2008, available in http://EconPapers.repec.org/RePEc:cns:cnscwp:200813

[28] L-H. Chen, L. Huang, "Portfolio optimization of equity mutual funds with fuzzy return rates and risks", *Expert Systems with Applications,* vol. 36, 2008, pp. 3720–3727.

[29] N. Capon, G. J. Fitzsimons, R. A. Prince, "An individual level analysis of the mutual fund investment decision", *J. Financial Services Research*, vol. 10, 1996, pp. 59-82.

[30] A. R. Johnson, W. W. Dean, "*Applied Multivariate Statistical Analysis"*, Pearson/Prentice Hall, 2007.

[31] C. T. Ragsdale, "*Spreadsheet Modeling and Decision Analysis",* 3rd ed., South-Western College

[32] T. J. McClave, P. G. Benson, T. A. Sincich, "*First Course in Business Statistics*", 8th ed, NY: Prentice Hall, 2001.

[33] A. C. Mertler, R. A. Vannatta, "*Advanced and Multivariate Statistical Methods"* Pyrczak Publishing, 2005.

[34] F. J. Hair, E. A. Rolph, R. L. Tahtam, W.C. Black, "*Multivariate Data Analysis*", 5th ed., NY: Prentice Hall. Publishing, 2001

[35] D. Iacobucci, "*Marketing Models: Multivariate Statistics and Marketing Analytics"*, South- Western- Cengage Learning, 2013.

[36] K. Ozdamar, "*Statistical Data Analysis by Software"*, Kaan Publishing, 2004.

[37] R. Ho, "*Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS"* Taylor & Francis Group, 2006.

[38] W.K. Härdle, L. Simar, "*Applied Multivariate Statistical Analysis"*, Springer-Verlag Berlin Heidelberg, 2012, pg. 331-349.

[39] L. Meyers, G. Gamst, A. Guarino, "*Applied Multivariate Research Design and Interpretation"*, Sage publications, 2006.

[40] H. Anton, R. Busby, "*Contemporary Linear Algebra"*, John Wiley & Sons, Inc, 2003.

[41] M. Kasparova, J. Krupka, J. Chylkova, "Comparison of cluster analysis algorithms for heavy metals contamination in atmospheric precipitation" *WSEAS Trans. on Information Science and Applications*, vol. 7, no. 3, March 2010.

[42] S. E. Brian, "*Multivariable Modelling and Multivariate Analysis for the Behavioral Sciences",* CRC Press, Taylor & Francis Group, 2009, pg. 239-258.

[43] *Capital Markets Board of Turkey*, www.spk.gov.tr

[44] M. Massa, W. N. Goetzman, G. Rouwenhorst, "*Behavioral factors in mutual funds flows",* Yale School of Manegement, 1999.

N. S. Ruzgar became a Member (M) of **NAUN** in 2007. She was born in Pinarhisar, Turkey in 1962. She received the Ms degree in Mathematics from Istanbul Technical University, Istanbul, Turkey in 1989 and Ph. D. degree in operation research from Istanbul University, Istanbul, Turkey, in 1998.

She worked more than 20 years at Computer and Electronic Education Department of Technical Education Faculty, Marmara University, Istanbul Turkey. Currently, she is working in Department of Mathematics, Ryerson University, Toronto, Canada. She has four books, an author of more than 30 papers in refereed journals and more than 50 papers in conference proceedings. Her research interests are system simulation, applied statistics, modeling, and distance education.

Ass. 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.

F. Unsal was born in Turkey. He received his Bs and Ms from the American University of Beirut. He completed his Ph. D. at Cornell University, Ithaca, New York in 1979. His primary field was Marketing and he minored in Economic Theory and Quantitative Analysis.

He is currently a Professor of Marketing/International Business at Ithaca College, Ithaca, New York. His research interests include technology, online education, and applications of quantitative tools in business.