Ekonomik ve Sosyal Araştırmalar Dergisi, Bahar 2007, Cilt:3, Yıl:3, Sayı:1, 3:66-86

THE USE OF SOME MULTIVARIATE METHODS IN ANALYSING STOCK PERFORMANCE: AN APPLICATION TO ISE 100 INDEX

İlhan EGE
Nevşehir University
Faculty of Economics and Administrative Sciences, Nevşehir
iege@nevsehir.edu.tr, ilhanege2005@hotmail.com

Ali BAYRAKDAROĞLU
Nevşehir University
Faculty of Economics and Administrative Sciences, Nevşehir
bayraktar@nevsehir.edu.tr, ali.bayrakdaroglu@hotmail.com

HİSSE SENEDİ PERFORMANSLARININ ANALİZİNDE BAZI ÇOK DEĞİŞKENLİ ANALİZ YÖNTEMLERİNİN KULLANIMI: İMKB 100 ENDEKSİNDE BİR UYGULAMA

Özet

Bu çalışmanın temel amacı İMKB 100 endeksinde yer alan hisse senetlerinin performansını lojistik regresyon yöntemi kullanarak ölçmektir. Bu çalışma özellikle üç soruya cevap aramaktadır: (1) İMKB 100 endeksinde yer alan hisse senetlerinin getirilerini finansal oranlar ile açıklayabilir miyiz? (2) Hisse senetlerinin seçiminde finansal oranlar en kullanışlı seçim kriteri midir? (3) Hisse senedi getilerini lojistic regresyon yöntemi ile analiz edebilir miyiz? Çalışma İMKB 100 endeksindeki imalat işletmerinin 2005 yılındaki dönemsel verilerini kapsamaktadır. Analiz sonucuna gore finansal performanslarına gore İMKB 100 şirketlerinin sınıflandırılması %80 oranından daha yüksek oranda bulunmuştur. Bu çalışmanın sonucuna gore, kontrol değişkenleri gibi diğer değişkenlerde hisse senedi getirilerini efektif olarak açıklamaktadır.

Anahtar Kavramlar: Faktör Analizi, Lojistik Regresyon, Hisse Senedi Performansı, Finansal Oranlar, İMKB 100.

Abstract

The main objective of this paper is measuring the stocks performances of firms in ISE 100 index by using logistic regression method. Specifically, this paper answers three questions: (1) Can we explain yields of stocks in ISE 100 with financial ratios? (2) Are financial ratios usable selection criteria of stocks? (3) Can we analyze stocks' yields with logistic regression model? The study includes only industrial firms over the period of 2005 in ISE 100 index. According to the analysis result, classification success of ISE-100 companies according to their financial performance is found out to be much more than 80%. According to the result of the study, it is possible to state that as well as the control variables, other variable are also effective to explain stock share earnings when related period is taken in to consideration and result are generalized.

Keywords: Factor Analysis, Logistic Regression, Stock Performance, Financial Ratios, ISE 100.

1. Introduction

Return has two components: Yield and capital gain. Yield measures relate cash flows to a price for the security, such as the purchase price or the current market price (Jones, 1999). The main object of this paper is measuring the stocks performances of firms in ISE 100 index by using logistic regression method. Specifically, this paper answers three questions: (1) Can we explain yields of stocks in ISE 100 with financial ratios? (2) Are financial ratios usable selection criteria of stocks? (3) Can we analyze stocks' yields in future period with this model?

In this context, we analyze only industrial firms over the period of 2005 in ISE 100 index. We collect data from periodicals in the web page of ISE. We use two data: Financial tables for using calculation of financial ratios, and stocks price for using calculation of stocks' yields. Service sector firms and financial companies are omitted because they have different characteristic. We analyze 35 firms which have healthy data.

First, in this study we calculated the financial ratios of firms over the period of 2005 for twelve months with using balance sheet and income statement data. In finance literature some studies use quarter period instead of 12 months period (e. g., see Yalçıner et al., 2005; Atan and Catalbas, 2004; Akkum and Vuran, 2005; Kandır, 2005; Karamustafa and Kucukkale, 2002). These studies used quarterly period because they searched effects of changes of financial ratios to stock prices. However, the main focus of our paper is not on the volatility of returns, but rather on the relationship between financial ratios and stock returns over a year. Therefore, we employed monthly data. Because of this we use twelve month period.

Dependent variable is stocks yields and this variable is categorical dependent variables because of logistic regression method. Also independent variables are liquidity, activity, profitability and financial structure ratios which show the firms performance. These 18 ratios are given in the Table-1.

Table 1. Variables of This Study

NO	Variables	Explanation
	Liquidity Ratios	
1	CR	Current Ratio
2	LR	Liquidity Ratio
3	CAR	Cash Ratio
	Activity Ratios	
4	RT	Receivables Turnover
5	ST	Stock Turnover
6	CAT	Current Asset Turnover
7	TAT	Total Asset Turnover
8	ET	Equity Turnover
	Profitability Ratios	
9	ROA	Return on Assets
10	GMM	Gross Merchandise Margin
11	NPM	Net Profit Margin
12	OM	Operating Margin
13	EBTE	Earnings Before Tax/ Equity
	Financial Structure Ratios	
14	ETA	Equity /Total Assets
15	TDTA	Total Debts / Total Assets
16	STDTD	Short Term Debts / Total Debts
17	CCTL	Continuous Capital/Total Liabilities
18	LTDTD	Long Term Debts / Total Debts

We use close price of stocks. Year yield between month t and month t-12 is calculated the following formulas (Jorion, 2001; Zivot, 2002):

$$R_{t(12)} = \frac{P_t - P_{t-12}}{P_{t-12}}$$

In this formula, $R_{t(12)}$ is 12 months (year) yield in month t, P_t is price of security in month t and P_{t-12} is price of security in the twelfth month.

Method of this study includes selection of variables in the model, composing the set of data, determining the techniques for using in the model and analyzing the right categorization success of these techniques.

In the logistic regression model there are two ways in the selection of variables. One is binary logistic regression. In this method all of the variables, which are probabilities (p) less than 0, 25, can be used for the multiple model (Erdogan, 2002).

We use firstly factor analysis method because of elimination of multicollinearity problem in the logistic regression model. In this method we group the financial ratios, determine the explanatory variables according to weighted general scores and lastly forecast yield success. In this context 18 financial ratios eliminate and better explanatory factor variables are defined.

In this study we use binary logistic regression method because independent variables do not have normal distribution and dependent variable is categorical variable (Daştan, 2003). Generally, studies that use the financial ratios, financial ratios have positive skewness instead of normal distribution (Aktaş, 1997). We prefer logistic regression method because this method eliminates these limitations (Tatlıdil and Ozel, 2005). In other words we use Logit model because dependent variable is dummy variable.

2. Analysis of Models

2.1. Method

2.1.1. Factor Analysis

Factor analysis is a statistical technique widely used in psychology and the social sciences (Kline, 1994: 1). Its common objective is to represent a set of variables in terms of a smaller number of hypothetical variables (Kim and Mueller, 1978: 9). In the social sciences factor analysis is usually applied to correlations between variables (Kline, 1994: 3). It enables the social scientist to study behavioural phenomena of great complexity and diversity and to mold his findings into scientific theories (Rummel, 1970: 3).

Factor analytic methods can help scientists to define their variables more precisely and decide what variables they should study and relate to each other in the attempt to develop their science to a higher level. These methods can also help these scientists to gain a better understanding of the complex and poorly defined interrelationships among large numbers of imprecisely measured variables (Comrey and Lee, 1992: 1).

In exploratory factor analysis the aim is to explore the field, to discover the main constructs or dimensions. It was for this purpose that factor analysis was originally developed by Spearmen (1904), in the area of human abilities (Kline, 1994: 7). It attempts to reduce a set of, say ten variables, into two or three underlying “factors”. Confirmatory factor analysis, on the other hand, posits that there are, say, two underlying factors for a set of ten variables and then seeks to determine whether this hypothesis does hold (Kim and Mueller, 1978: 5). In this method, which developed by Joreskog (1973), based upon previous studies or on relevant theory, factor loadings for the variables are hypothesized (Kline, 1994: 10).

Factor analysis will involve all of the following major steps (Comrey and Lee, 1992: 5):

- selecting the variables;
- computing the matrix of correlations among the variables;
- extracting the unrotated factors;
- rotating the factors; and
- interpreting the rotated factors.

In 1904, Charles Spearman involved a single general factor in his paper "General Intelligence, Objectively Determined and Measured" was published in the American Journal of Psychology. He developed the Two-Factor Theory. Karl Pearson (1901), Cyril Burt (1966), Godfrey H. Thomson, J. C. Maxwell Garnett (1919) and Karl Holzinger (1936) studied factor analysis. Garnett used the concept of multiple-factor analysis (Harman, 1976: 3).

Factor analysis has been applied to cross-national data by Cattell (1949) and Berry (1960), world regional patterns by Russett (1966), value orientations of culture groups by Kluckhohn and Strodtbeck (1961), urban crime by Schuessler (1962), classification of groups by Borgatta and Cottrell (1955), classification of primitive tribes by Schuessler and Driver (1956), effectiveness of complex organizations by Godfrey et al. (1958), dimensions of community systems by Jonasssen and Peres (1960), social change data by Gibb (1956), political attitudes by Eysenck (1954), issues of a profession by Somit and Tanenhaus (1963), Senate roll call votes by Harris (1948), UN roll call votes by Alker (1964), and judicial voting behaviour by Schubert (1962) (Rummel, 1970: 13).

In economics evaluating the performance of systems are studied by Burch (1972), investment decisions under uncertainty is studied by Farrar (1962); latent structure of security price changes are studied by King (1964); economic equation systems employing principal components are studied by Kloek and Mennes (1960) (Harman, 1976: 7).

Factor analysis programs (for example SAS, SPSS) used by Cody and Smith (1997), Delwiche and Slaughter (1998), Der and Everitt

(2002), Hatcher (1994), Gardner (2001), Green, Salkind and Akey (2000) (Pett, Lackey and Sullivan, 2003: XV).

Madan (2005) derived a factor analysis from a one period equilibrium model. Dumas (1989), Wang (1996), Chan and Kogan (2002), Lengwiler (2005) and Lengwiler, Malamud, and Trubowitz (2005) analyzed the infuence of heterogeneous preferences on asset prices. Perturbative methods have also been used by Chan and Kogan (2002), Kogan and Uppal (2001), Leippold, Trojani, and Vanini (2006), Lengwiler, Malamud, and Trubowitz (2005) (Malamud and Trubowitz, 2006: 1).

2.1.2. Logistic Regression

Logistic regression allows one to form a multivariate regression relation between a dependent variable and several independent variables (Lee, Ryu and Kim, 2007: 329). Logistic regression, which is a multivariate analysis model, is useful for predicting the presence or absence of a characteristic or outcome based on values of a set of predictor variables (Lee, 2004: 226). The predictor values from the analysis can be interpreted as probabilities (0 or 1 outcome) or membership in the target groups (categorical dependent variables). It should be noted that probability of a 0 or 1 outcome is a nonlinear function of the logit (Nepal, 2003: 316).

It is designed to estimate the parameters of a multivariate explanatory model in situations where the dependent variable is dichotomous, and the independent variables are continuous or categorical. This technique yields coefficients for each independent variable based on a sample of data (Huang, Chai and Peng, 2007: 57). The parameters of the logistic regression model are commonly estimated by maximum

Likelihood (Pardo, Pardo and Pardo, 2005: 93).

The advantage of logistic regression is that, through the addition of an appropriate link function to the usual linear regression model, the variables may be either continuous or discrete, or any combination of both types, and they do not necessarily have normal distributions (Lee, 2004: 226).

Logistic regression models (LRM)with two or more explanatory variables are widely used in practice (Haines and Others, 2007: 91).

LRM is a regression model designed to cope with binary or dichotomous response variables (Balasch, Romero and Ferrer, 2004: 347).

Logistic regression is similar to linear regression the main equation (Cleophas, Zwinderman and Cleophas, 2006: 190):

$Y = a + b_1x_1 + b_2x_2 + \ldots b_nx_n$

where

y = the dependent variable,

x_1 = the independent variable,

x_2 = a second independent variable,

x_n = last independent variable and,

a, b_1, b_2, b_n = coefficients

Logistic regression and logistic discrimination are treated as special cases and generalized linear models are also discussed in introductory books such as Fienberg (1980) and Everitt (1977) and advanced books such as Bishop, Fienberg and Holland (1975), Haberman (1974) and Santner and Duffy (1989) (Christensen, 1997).

There has however been only sporadic interest in optimal designs for such models, with the papers of Sitter and Torsney (1995), Atkinson and Haines (1996), Jia and Myers (2001), Torsney and Gunduz (2001) and Atkinson (2006) and the thesis of Kupchak (2000) providing valuable insights into the underlying problems (Haines and Others, 2007: 91). Some other studies are Theil (1971), Berkson (1944, 1953), Cox and Snell (1989), Hosmer and Lemeshow (1989), Andersen (1990), Agresti (1990), Amemiya (1985) Agresti (1996), andRyan (1997) (Pardo, Pardo and Pardo, 2005: 92; Arias-Nicolás, 2007).

2.2. Factor Analysis Model and Evaluation of Its Results

First correlations between variables were determined and correlation matrix was calculated using 18 financial rates defined above in order to constitute appropriate factors. It is required that correlation values between variables is supposed to be high so that variables can determine any dimension.

Validity of factor models was tested via Bartlett test and Kiaser–Meyer–Olkin (KMO) test. If there exists any relations between variables in the main block and if the correlation matrix is the unit matrix or not are analysed by using sphericity test. Observed coefficient rate and partial correlation coefficient rate were compared. Calculated KMO and Bartlett tests results are shown in Table 2.

Table 2. KMO and Bartlett Test Results

Tests		Relevant Period
KMO Measure of Sampling Adequacy		,540
Bartlett's Test of Sphericity	Approx. Chi-Square	235,299
	df	21
	Sig.	,000

As understood from the above two tests, H_0 hypothesis (H_0 : R = 1) is refused, H_1 hypothesis is accepted. Therefore, because the correlation matrix (R) is not equal to unit matrix (I), there is a relation between variables and 54.0% KMO value supports that (KMO=0.540). Bartlett sphericity test is also found to be meaningful at 0.05% level (P=0.000). It is possible to state that factor analysis can be applied to the variables.

In this research main components method was used as factor derivation model. Obtained communality value is evaluated by means of this method and it is determined that these calculated cooperative variances were not low. In other words, there existed no low variance in the analysis.

Initial eigenvalues statistics was used to determine the factor number. Factors more than 1 were accepted to be meaningful factors in initial eigenvalues statistics. In Table 3 factor numbers derived from initial eigenvalues statistics and explained variance rate are shown.

As it is shown in Table 3, 5 factors explaining 82.80% of total variance were obtained. In this research, as well as initial eigenvalues statistics, scree test was also employed to determine factor numbers. Factor number as indicated by the spot where slope began to disappear was determined in factor analysis linear diagram. According to this, it is determined that linear diagram slope began to disappear after 5^{th} factor (Appendix 2).

Table 3. Total Variance Explained

	Initial Eigenvalues		
Factors	**Total**	**% of Variance**	**Cumulative %**
1	6,014	33,409	33,409
2	3,914	21,745	55,154
3	2,622	14,568	69,722
4	1,342	7,453	77,175
5	1,013	5,627	82,802
6	,911	5,059	87,861
7	,790	4,389	92,250
8	,561	3,114	95,364
9	,350	1,944	97,308
10	,212	1,179	98,487
11	,113	,630	99,117
12	5,062E-02	,281	99,398
13	3,792E-02	,211	99,609
14	3,614E-02	,201	99,810
15	1,875E-02	,104	99,914
16	8,892E-03	4,940E-02	99,963
17	6,603E-03	3,668E-02	100,000
18	3,474E-17	1,930E-16	100,000

"Varimax Method" an axis rotation method was used in order to obtain conceptual meaningfulness and to simplify the evaluation of factor loads. Therefore, reversed factor matrix was calculated and Varimax Method results were used in the evaluations. This matrix is the final result of factor analysis. Reversed factor matrix results are shown in Table 4.

It is seen in Table 4 that under which factors the variables are grouped. Factors were determined through grouping high load ones. Table 4 shows that variables having higher than 0.5% coefficient are found to be variables which have closed relations with the factors. Each factor can be formed relations with the variables collected below them. So factors weren't named because it is not evaluated in terms of conceptual meaningfulness. In binary logistic regression analysis used below factor analysis and dimension reduced data set were used because chosen factor's explanation rates were enough and reversed factor values were taken into consideration. In above mentioned data set these variables take place: CR, LR, TAT, CAT, NPM, GMM, STDTD, LTDTD and ST.

Table 4. Rotated Component Matrix*

Variables (Rations)	Component (Factors) 1	2	3	4	5
CR	,922				
LR	,912				
CAR	,828				
TDTA	-,827				
ETA	,800				
TAT		,947			
CAT		,936			
ET		,807			
GMM		-,710			
NPM			,826		
OM			,807		
EBTE			,734		
ROA			,717		
STDTD				,797	
LTDTD				-,797	
RT				-,661	
CCTL				-,658	
ST					,855

*Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

The second part of the research is composed of binary logistic regression analysis using variables of above determined common basic factors. How this test is conducted and its result is as follows.

2.3. Binary Logistic Regression Model Analysis and Evaluation of Its Results

Binary logistic regression analysis was applied in order to test if there is a difference of financial performance between the companies which have positive stock yield and negative stock yield in the mentioned period and transaction in the ISO 100. More clearly binary logistic regression method was used to determine financial ratios effective in explaining stock yields.

Binary logistic regression used in the analysis has two categories. Therefore, dependent variable (stock yields) has only two different values: 0 and 1. During the analysis, 1 was used for the ISO 100 companies which had positive incomes in that period (2005) and 0 was used for others because changes in stock holder's values were taken into consideration as criteria of financial performance. If categorical dependent variable in binary logistic regression model was symbolized as Y, it was defined as follows:

$$Y = \begin{cases} 1, \text{ Shareholder value up} \\ 0, \text{ Shareholder value down} \end{cases}$$

Financial rates were used as independent variable in the regression model to analyze yield achievement of stock shares. It is known that selection of variables is one of the most important parts of analysis studies having multiple variables. Determination of variables which affect dependent variables most, including all the necessary variables into the model and removing unnecessary variables from the model affect the validity of the model positively and to a great extent. Therefore, as a result of the above factor analysis, related independent variables are determined and model formation was employed to select variables to represent each factor. CR, TAT, NPM, STDTD and ST were used as independent variables in the regression analysis.

In this study, a = 0.10 was accepted in logistic regression analysis and cutting pointing appointing groups is determined to be 0.5. Common use in the literature was taken into consideration in determining these values (Erdogan, 2002).

Forward Stepwise-Wald procedure is used in logistic regression analysis. Firs, the model in which the constant takes place was formed, then each step a variable was added to the model through beginning with the variable which contributed more the model. Finally, meaningful variable were kept in the model. According to this, in a model which consists of only the constant term in the first step, only the units of a group can be grouped correctly.

Firstly we add the constant to the model and in the third iteration of this phase; -2logL statistics is downturn less than 0.001. Because of this iteration is finished and in the third iteration of first phase coefficient of constant is 0.916 and -2LogL statistics is 41.879 (Table 5).

Table 5. 1. Section Iteration History

Iteration		**-2Log Likelihood**	**Coefficients Constant**
Step 0	**1**	41,904	,857
	2	41,879	,916
	3	41,879	,916

For the model to be forecasted, rates were examined one by one and statistically meaningful 4 models were estimated. In other words, in logit model, the best one explaining the yield achievement was obtained among 20 variables. Among these, only the model which had the highest rate to explain income achievement was tested in terms of coefficient and its general meaningfulness. The results are summarized below.

Model:

$\mathbf{Y_i = \beta_0 + \beta_2 + \beta_2}$

$\mathbf{Y_i = -2.145+1.343(TAT) +38.025(NPM)-0.019(ST)}$

Y_i = Stock Yield (Positive; 1, Negative; 0)

TAT: Total Assets Turnover

NPM: Net Profit Margin

ST: Stock Turnover

Standart error of ST = 0,013,

Standart error of TAT = 0,859,

Standart error of NPM = 19,560.

Comparison of the observed value and estimated value in logistic regression depends on log likelihood. Admission requirement for a good model is that the observed results must form high likelihood. Log likelihood value has values between 0-1. This rate shows estimating likelihood of the dependent variable by independent variables. Logarithm of the numbers less than 1 is between 0 and $-\infty$. LogL statistics is estimated by maximum likelihood algorithm. Because -2LogL statistics is approximately similar Chi-Square distribution, -2LogL statistics is similar to sum of squares of error term in logistic regression analysis, too. If odds ratio equals to 1, -2LogL statistics will equal to 0. This means that -2 LogL statistics is low. If the model has exact adaptation likelihoods 1 and -2LogL statictics is 0. In this study, Model Chi-Squares statistics was used to test its general meaningfulness. This statistics which test guessed logit model in general, tests if all logit coefficients, except for the fixed term are 0, or not. It was calculated via determining the difference -2 LogL statistics of a model without an independent variable and -2LogL statistics of a model with independent variables. According to this, in the logit model consisting of the constant -2LogL likelihood value is 41,879. -2LogL value of the model consisting of the constant and related independent variables is 27,430. Model Chi-Squares value, 14,499 is the difference between these two-2LogL likelihood values. P value of this test result meaningfulness rate is 0,002. This value is 5% meaningful and the hypothesis which states the coefficients are equal to "0" is refused. That is, these hypothesis,

$H_0=\beta_1= \beta_2= \beta_3=0$; model is meaningless is general

$H_1= \beta_1\neq \beta_2=\neq\beta_3\neq0$; model is meaningful in general were tested and it was decided that estimated logit model was meaningful in general (App-3).

In forecast equation, for each company units are classified by calculating financial data Z_1 values. P_1 values obtained as result of transforming Z_1 values determine whether company will be in the successful class or unsuccessful class. If P1 value is higher than 0, 50 stock shares (Y_i=1) are placed in successful class, if it is lower than 0, 50 (Y_i=0) they are placed in unsuccessful class.

Table 6. Classification Table of Model

Observed		Predicted		
		Group		Perc. % Correct
		Unsuccessful	Successful	
Group	Unsuccessful (Neg. Return)	6	2	60,0
	Successful (Poz. Return	4	23	92,0
	Total	10	25	82,9

Correct classification rate, in the dissociation received by the help of P_i values obtained by using antilog of Z_i values of the model is 82, 9 %. Table-6 shows that companies providing positive stock share earnings, in short successful companies have 92,0% correct classification rate, while companies having negative stock share income, in short unsuccessful companies have 60,0% correct classification rate. Moreover, the table shows that 2 companies defined to be unsuccessful at he beginning and 6 companies defined to be successful at he beginning are classified incorrectly. As opposed to this 23 companies defined to be successful are shown to have correct classification.

Table-7 shows binary logistic regression result of variables in the model. β value shown in the table was used determine whether the dependent variable 0 or 1. β_2 value is ST ratios' coefficient and is fond -0,019. Raised of this coefficient decreases logarithm of diversity ratio of firms which including in high yield group, because this ratio is negative.

Table 7. Variables in the Equation

Variables	β	Std. Error	Wald	df	(P)	Exp (β)
TAT	1,343	,859	2,444	1	0,018	3,829
NPM	38,025	19,560	3,779	1	0,053	3,265
ST	-,019	,013	2,239	1	0,035	,981
Constant	-2,145	1,501	2,042	1	0,070	,117

Wald statistics is used to test the meaningfulness of logistic regression coefficient and meaningfulness rates are shown by P value in the model. Therefore, fixed term and NPM independent variable is 0, 10 meaningful, TAT and ST variables are 0, 05 meaningful, which shows that it is meaningful statistically. In Table 7, value of β_1 which is coefficient of ratio ST is 1,343 and positive and value of β_2 which is coefficient of ratio NKM is 38,025 and positive. Log of difference ratio of firms coming into high yield firms raises respectively 1,343 and 38,025 unit because of rising these ST and NKM ratios. However ratio of OM decreases this log of difference ratio of firms coming into high yield firms and probility of firms coming into better firm group.

When the Exp (β) value which is the contribution of these variables to the model is evaluated, it is found that TAT variable has the most contributions. Next ones are NPM and ST rate variables. It is determined that Cox and Snell statistics for the model was 33, 8% and relation rate between the dependent and independent variables is 48, 5% Nagelkerke statistics (App-4)

3. Conclusion and Evaluation

In this study, the relation between financial rates and stock share earnings in mentioned period was analysed and binary logistic regression analysis was conducted by using financial rates. Companies having low and high earnings are determined by finding financial rate which explain earning with logistic regression analysis.

Factor analysis was used in determining financial rate to be added to the model. The reason to use factor analysis in the study was that binary logistic regression was too sensible to multicollinearity. After grouping the variables by factor analysis is, studying with the variables which could represent the group correctly provided easier analysis.

NPM, ST and TAT rate were added respectively to the model which has 71,4% correct classification rate of only constant term and stock share earning and this rate increased to 82,9%. At this stage, the relation between control variables were explained to be 48, 5% according to Nagelkerke statistics.

According to the analysis result, classification success of ISE-100 companies according to their financial performance is found out to be much more than 80%. When evaluated fro the investors point of view, it is concluded that it is possible to determine the companies having positive earning, to forecast stock shares having earning by examining financial rates.

According to the result of the study, it is possible to state that as well as the control variables, other variable are also effective to explain stock share earnings when related period is taken in to consideration and result are generalized. It should be kept in mind that stock share prices may be changed by any kind of information about the company. Moreover, it should also be kept in mind that these prices may chance according to general economics conjuncture, investors or investors groups expectations and always possible speculative investments.

In general, it is possible to state that financial rates of companies can direct their future earnings under normal conditions. However it is extremely difficult that any change in the performance of a company is noticed at he same time by the markets under conditions in which no activity is experienced.

In forecasting earning success, instead of adding financial rate to logit model one by one, factor analysis result obtained by using variables from liquidity rate group, activity rates group, profitableness rate group, capital structure rates group are used in the analysis of the relation between financial rates and stock share earnings. Consequently, binary logistic regression method, as an investment criterion, helps the investor to form an opinion about the stock shares to be invested. However, it should be kept in mind that the model using this method may cause incorrect results when used alone.

In further studies, other variables which can affect earning rate of stock shares can be added to the model as well as financial rate as an independent variable. In this study, data for 12 months was taken into consideration and at the end of 12th. Month, stock share prices were compared with the previous year, and the success was determined. In further studies data for each 3 month can be used and different success can be defined differently. Moreover, studies to determine the most prominent financial rate in selecting stock shares and other studies to find out the factors peculiar to the company and general economic conjuncture will be able to contribute to the literature.

Acknowledgements

The authors express their sincere thanks to participants of The 11th IFAC Intl. Symposium Computational Economics and Financial and Industrial Systems (CEFIS) 2007, October 09-11, Istanbul, Turkey for their valuable contributions, useful comments and remarks on this research.

References

Akkum, T., Vuran, B. (2005) "Türk Sermaye Piyasasındaki Hisse Senedi Getirilerini Etkileyen Makroekonomik Faktörlerin Arbitraj Fiyatlama Modeli ile Analizi", *İktisat İşletme ve Finans Dergisi,* Yıl 20, Ağustos. 28-45

Aktaş, R (1995) Mali Başarısızlık (Işletme Riski) Tahmin Modelleri, T. İ Bankası Kültür Yayınları.

Albayrak, A. S. (2006) Uygulamalı Çok Değişkenli İstatistik Teknikleri. Ankara: Asil Yayınları,.

Arias-Nicolás, J. P. , Pérez, C. J. , Martín, J. (2007) "A Logistic Regression-Based Pairwise Comparison Method to Aggregate Preferences", Group Decision and Negotiation, Springer Netherlands.

Atan, M., Çatalbaş, E. (2004) "Çok Değişkenli İstatistiksel Analiz Yöntemleri ile Türk Bankacılık Sektöründe Çok Boyutlu Mali Başarısızlık Tahmin Modelleri Oluşturulması", *IV. İstatistik Günleri Sempozyumu,* Dokuz Eylül Üniversitesi, İzmir-Kuşadası, 19-22 Mayıs.

Balasch, S. , Romero, R. , Ferrer, A. (2004) “A Logistic Regression Model Applied to Evaluate the Influence of Operating Time of AgI Ground Acetonic Generators on the Size and Hardness of Hail”, Natural Hazards, Vol. 32, No 3, 345-355.

Benli, Y. K. (2005) “Sektörel Farklılıkların Oranlar Üzerine Etkisi: Ampirik Bir Çalışma”, Gazi *Üniversitesi Endüstriyel Sanatlar Eğitim Fakültesi Dergisi*, Sayı 16, 14-30.

Büyüköztürk, Ş (2003) Veri Analizi El Kitabı, Geliştirilmiş 3. Baskı, Pegem Yayıncılık.

Christensen, R. (1997) Log-Linear Models and Logistic Regression. Netherlands: Springer.

Cleophas, T. J. , Zwinderman, A. H. , Cleophas, T. F. (2006) Post-Hoc Analyses in Clinical Trials, A Case for Logistic Regression Analysis. Netherlands: Springer.

Comrey, A. L., Lee, H. B. (1992) A First Course in Factor Analysis. Lawrence Erlbaum Associates.

Daşdan S. (2003) Lojistik Regresyon Tekniği ile Hisse Senedi Getirilerinin Analizi, Yüksek Lisans Tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü.

Erdoğan, B. E. (2002) Bankaların Mali Performanslarının Lojistik Regresyon ile Analizi ve İleriye Yönelik Tahmin ile Bir Uygulama, Doktora Tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü.

Gujarati, N. D. (2006) Temel Ekonometri, Çev. Ümit Şenesen, Gülay Günlük Şenesen, İstanbul: Literatür Yayıncılık.

Haines, L. M. and Others (2007) “D-optimal Designs for Logistic Regression in Two Variables, mODa 8-Advances in Model-Oriented Designed and Anaysis”, Physica-Verlag HD, 91-98.

Harman, H. H. (1976) Modern Factor Analysis. University of Chicago Press.

Hosmer, D.W, Lemeshow, Jr. S. (1995) Applied Logistic Regression. John Wiley&Sons.

Huang, Q. , Cai, Y. , Peng, J. (2007) Modeling the Spatial Pattern of Farmland Using GIS and Multiple Logistic Regression: A Case Study of Maotiao River Basin, Guizhou Province, China, Environmental Modeling and Assessment, Vol. 12, No. 1, 55–61.

Jones, P. C. (2000) Investment Analysis and Management, Seventh Edition. John Wiley & Sons, Inc.

Jorion, P. (2001) Value at Risk: The New Benchmark For Managing Financial Risk. New York: Mcgraw-Hill, 2nd Edition.

Kalaycı, Ş. (2006) SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri. Ankara: Asil Yayın Dağıtım Ltd. Şti.

Kandır, S. Y. (2006) "Tüketici Güveni ve Hisse Senedi Getirileri İlişkisi: İMKB Mali Sektör Şirketleri Üzerinde Bir Uygulama", *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, Cilt 15, Sayı 2, 217–230.

Karamustafa, O., Küçükkale, Y. (2002) "Hisse Senedi Getirileri ve Makroekonomik Değişkenlerin Koentegrasyon ve Nedensellik İlişkileri", *6. Türkiye Finans Eğitimi Sempozyumu*, Aralık, Isparta.

Kılıçkaplan, S. (1997) "Avrupa Birliği Karşısında Türkiye'nin Rekabet Gücünün Ölçülmesinde Açıklanmış Karşılaştırmalı Üstünlük İndeksine Alternatif Bir Yaklaşım Olarak Temel Bileşenler Analizi", *III. Ulusal Ekonometri ve İstatistik Sempozyumu*, 29-30 Mayıs, Uludağ Üniversitesi, Bursa.

Kim, J. , Mueller, C. W. (1978) Introduction to Factor Analysis: What It Is and How to Do It. SAGE University Paper.

Kline, P. (1994) An Easy Guide to Factor Analysis. Routledge.

Lee, S. (2004) "Application of Likelihood Ratio and Logistic Regression Models to Landslide Susceptibility Mapping Using GIS", Environmental Management, Vol. 34, No. 2, 223-232.

Lee, S. , Ryu J. , Kim, L. (2007) "Landslide Susceptibility Analysis and Its Verification Using Likelihood Ratio, Logistic Regression, and Artificial Neural Network Models: Case Study of Youngin, Korea", Landslides, Vol. 4, 327–338.

Malamud S., Trubowitz, E. (2006) Rational Factor Analysis, Working Paper.

Nakip, M. (2003), Pazarlama Araştırmaları: Teknikler ve SPSS Destekli Uygulamalar,.Ankara: Seçkin Yayınları.

Nepal, S. K. (2003) "Trail Impacts in Sagarmatha (Mt. Everest) National Park, Nepal: A Logistic Regression Analysis", Environmental Management, Vol. 32, No. 3, 312-321.

Pardo, J. A. , Pardo, L. , Pardo, M. C. (2005) "Minimum Ø-divergence Estimator in Logistic Regression Models", Statistical Papers, No 47, 91-108.

Pett, M. A, Lackey, N. R., Sullivan, J. J. (2003) Making Sense of Factor Analysis: The Use of Factor Analysis for Instrument Development in Health Care Research. SAGE.

Rummel, R. J. (1970) Applied Factor Analysis. Northwestern University Press.

Tatlıdil, H., Özel, M. (2005) "Firma Derecelendirme Çalışmaları Konusunda Çok Değişkenli İstatistiksel Analize Dayalı Karar Destek Sistemlerinin Kullanımı", *Bankacılar Dergisi*, Türkiye Bankalar Birliği Yayını, Sayı 54, Eylül.

Yalçıner, K. and others (2004) "İMKB 30 Endeksinde Etkinlik Analizi ile Hisse Senedi Seçimi", *I. Uluslararası Manas Üniversitesi Ekonomi Konferansı*, 23-24 Eylül, Manas Üniversitesi, Bişkek/ Kırgızistan.

Zivot, E. Introduction to Computational Finance and Financial Econometris: Chapter 1 Asset Return Calculation, http://faculty.washington.edu, 26.03.2007

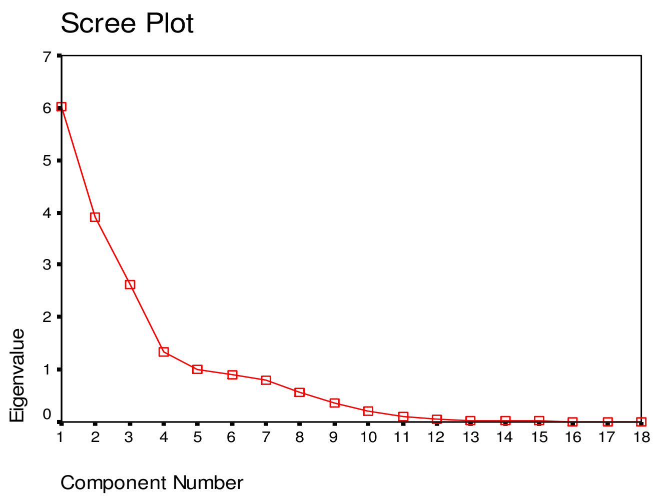
Appendix-1

Communalities

		Initial	Extraction
1	CR	1,000	,917
2	LR	1,000	,917
3	CAR	1,000	,842
4	RT	1,000	,918
5	ST	1,000	,948
6	CAT	1,000	,886
7	TAT	1,000	,842
8	ET	1,000	,814
9	ROA	1,000	,885
10	GMM	1,000	,814
11	NPM	1,000	,561
12	OM	1,000	,748
13	EBTE	1,000	,854
14	ETA	1,000	,819
15	TDTA	1,000	,641
16	STDTD	1,000	,904
17	CCTL	1,000	,768
18	LTDTD	1,000	,828

Extraction Method: Principal Component Analysis

Appendix -2: Result of Scree Test



Appendix -3: Result of Omnibus Tests

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	6,840	1	,009
	Block	6,840	1	,009
	Model	6,840	1	,009
Step 2	Step	2,928	1	,087
	Block	9,768	2	,008
	Model	9,768	2	,008
Step 3	Step	4,680	1	,031
	Block	14,449	3	,002
	Model	14,449	3	,002

Appendix-4: Cox & Snell R^2 and Nagelkerke R^2 Values of Model

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	35,039	,178	,254
2	32,111	,244	,349
3	27,430	,338	,485