Journal of Money and Economy
Vol. 10, No.2 Spring 2015

Modeling of Banks Bankruptcy in Iran (Multivariate Statistical Analysis)

Ahmadian, Azam[*]
and Gorji, Mahsa[**]

Received: 6/29/2015 Approved: 6/7/2016

Abstract

In this paper we construct a modeling for detection of banks which are experiencing serious problems. Sample and variable set of the study contains 30 banks of Iran during 2006-2014 and their financial ratios. Well known multivariate statistical technique (principal component analysis) was used to explore the basic ifnancial characteristics of the banks, and discriminant Logit and Probit models were estimated based on these characteristics. Results suggest that the model can be used as an analytical decision support tool in both on-site and off-site bank monitoring system to detect the banks which are experiencing serious problems.

Keywords: *Bank failure, Principal component analysis, Logit, Probit*

JEL Classifications: *C49, G21, G33*

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

---

[*] Ph.D. in Economics, Researcher, Monetary and Banking Research Institute.
Email: azam_ahmadian@yahoo.com
[**] MA in Finance

1. Introduction

The two last decades are marked by notable banking and financial crises by their extent as well as their exorbitant financial costs. In fact, many developing countries witnessed serious disturbances in their banking systems.

The study of bank failure is important for two reasons: First, an understanding of the factors related to a bank failure enables regulatory authorities to manage and supervise banks more efficiently. Second, the ability to differentiate between sound banks and troubled ones will reduce the expected cost of bank failure. In other words, if examiners can detect problems early enough, regulatory actions can be taken either to prevent a bank from failing or minimize the costs to the public and taxpayers (Thomson, 1991).

Banks in Iran are faced with various problems such as Non-performing Loan (NPL) to loan ratio> 5%[1], Capital adequacy <8%, low liquidity and interest rate lower than inflation. These problems have led some banks to face risks. Financial ratios such as CAMELS[2] ratios can provide meaningful quantitative information about the changes in internal conditions of banks. However, these issues do not exist and there are no regulations to identify failed banks. In this paper, financial ratios are used to determine important factors which can significantly explain the changes in internal conditions of the banks and NPL to loan ratio> 5% to identify banks at risk. Since this paper is the assessment of banks' performance effects on probability of banking failure, we do not use macroeconomic variables in our model.

This article combines three parametric models (Discriminant, Logit and Probit) with another parametric approach which is Principal Component Analysis (PCA). PCA helps us to explore and understand the underlying patterns of relationship between the financial ratios. By applying PCA to the financial data - the important financial factors-which can significantly explain the changes in financial conditions of banks were explicitly explored, and the five financial factor components (liquidity, earning, sensitivity, capital adequacy and asset quality) were determined. Factor scores were estimated for each bank with respect to the five factors determined and these scores were

١. Supervision of Central Bank of Islamic Republic of Iran and CAMELs rating project in Monetary and Banking Research Academy (Ahmadian, Azam), determined threshold of NPL to loan at 5%. We use this reference for this purpose.

٢. Capital adequacy, asset quality, management, earning, liquidity and sensitivity used as independent variables in estimating discriminant Logit and Probit models.

The rest of this article is organized as follows: Section 2 includes literature review. Section 3 presents the methodology; the sample and variable selection, PCA and the estimation of the early warning models. Finally, Section 4 concludes the article and discusses some future research perspectives.

2. Literature Review

Previous bank studies which employed multivariate statistical analysis include discriminant model (Sinkey, 1975), Logit models [Rose and Kolari, 1985; Pantolone and Platt, (1987)], and Probit model (Cole and Gunther, 1998). Financial ratios were directly used as independent variables to estimate the models in these studies.

Recently, some new studies used Multi-criteria Decision Aid (MCDA) which is originally an operational research approach for assessment of risk of financial failure. Slowinski and Zopounidis (1995) presented a new approach for evaluation of financial failure risk of firms based on the rough set theory. Zopounidis and Doumpos (1999) used UTilit_es Additives Discriminates (UTADIS) method and Mousseau et al. (2000) used ELECTRE TRI method which is a multiple criteria sorting method, i.e., a method that assigns alternatives (firms) to pre-defined categories.

Some other new studies tend to combine the non-parametric approaches with the discriminant or Logit analysis for bank failure prediction; Tam and Kiang (1992) introduced neural network approach to perform discriminant analysis as a promising method of evaluating bank conditions. Jo and Han (1996) suggested an integrated model approach for bankruptcy prediction; the discriminant analysis and two artificial intelligence models, neural network and case-based forecasting, and concluded that the integrated models produced higher prediction accuracy than individual models. Alam et al. (2000) stated that fuzzy clustering algorithm and self-organizing neural networks approaches provide valuable information to identify potentially failing banks. Kolari et al. (2002) used both parametric Logit analysis and the nonparametric trait approach to develop computer based early warning systems to identify large bank failures, and conclude that system provides valuable information about the future viability of large banks. Lam and Moy (2002), combined several discriminant methods, and performed simulation analysis to enhance the accuracy of classification results for classification of problems in discriminant analysis.

The objective of the paper of Canbas et al. (2005) was to propose a methodological framework for constructing the integrated early warning system (IEWS) that can be used as a decision support tool in bank examination and supervision process for detection of banks which were experiencing serious problems. Sample and variable set of the study contained 40 privately owned Turkish commercial banks (21 banks failed during the period 1997–2003) and their financial ratios. Well known multivariate statistical technique (Principal Component Analysis), was used to explore the basic financial characteristics of banks, and discriminant, Logit and Probit models were estimated based on these characteristics to construct IEWS. Also, importance of early warning systems in bank examination was evaluated with respect to cost of failure. Results of the study show that, if IEWS was effectively employed in bank supervision, it can be possible to avoid bank restructuring costs at a significant amount of rate in the long run.

Zaghdoudi (2013) has tried to develop a predictive model of Tunisian banks failures with the contribution of the binary logistic regression method. The specificity of his prediction model is that it takes microeconomic indicators of bank failures into account. The results obtained using his provisional model showed that a bank's ability to repay its debt, the coefficient of banking operations, bank profitability per employee and leverage financial ratio had a negative impact on the probability of failure. The objective of this study was to establish the microeconomic indicators which were able to predict banking defect. The use of collected financial ratios from the Tunisian banks' balance sheets shapes battery of indicators inspired by the CAMEL typology, from which we wanted to select the ratios that have a strong predictive power to construct a prevision model of bank defect from it. The use of a vector of ratios selected in advance by a stepwise regression, like a vector of explanatory variables in our logistic model have provided satisfactory results with expected signs and significations. Likewise, the most pertinent ratios in the explanation of banking defect at the Tunisian banks were the decrease of banking profitability and the ability of banks to repay their debts which appear to be a high odd ratio.

Salam (2012) empirically determined the significant determinants, among credit risk variables, of US bank failure. Applying the Probit model, the paper found that among five credit risk variables, the credit loss provision to net charge off, loan loss allowance to non-current loans, and non-current loans to loans were significant for predicting bank failures. These factors predicted 76.8 to 77.25 percent of total observation correctly. The model predicted 97 out of 121 failures i.e. 80.17 percent correctly. Net charge off to loans and loan loss to non-current loans, though most reliable measures, were not significant predictors for the US bank failures during 2009.

Boyacioglu et al. (2009) aimed to apply various neural network techniques, support vector machines and multivariate statistical methods to the bank failure prediction problem in a Turkish case to present a comprehensive computational comparison of the classification performances of the techniques tested. Twenty financial ratios with six feature groups including capital adequacy, asset quality, management quality, earnings, liquidity and sensitivity to market risk (CAMELS) were selected as predictor variables in the study. Four different data sets with different characteristics are developed using official financial data to improve the prediction performance. Each data set was also divided into training and validation sets. In the category of neural networks, four different architectures namely multi-layer perceptron, competitive learning, self-organizing map and learning vector quantization were employed. The multivariate statistical methods; multivariate discriminant analysis, k-means cluster analysis and logistic regression analysis were tested. Experimental results were evaluated with respect to the correct accuracy performance of techniques. Results showed that multi-layer perceptron and learning vector quantization could be considered as the most successful models in predicting the financial failure of banks.

The purpose of Arabi's paper (2013) was to estimate bank's failure by logistic regression and discriminant analysis. Both the logistic regression and discriminant analysis showed that earning (E) was the most influential measure of bank failure followed by asset quality (A), liquidity (L) and capital adequacy (C). The estimated discriminant function without cross validation obtained the following ratios 0.957, 0.872, 0.764, 1.000, 0.961 for fair, marginal, satisfactory, strong, and unsatisfactory respectively. While using cross validation, it obtained 0.941, 0.872, 0.764, 1.000, and 0.961 respectively. Averages for the first and second method were 0.878 and 0.756 respectively. It is obvious that the estimated function without cross validation was the best for predicting fiscal situation of banks and the most efficient early warning system. A new bank is identified as being of a particular rating dependent upon which discriminant function value is higher.

3. Methodology

3.1. The sample and variable selection

The sample set of the periods 2006-2014, contains financial ratios of 30 Iranian banks. Iranian banks are not explicitly declared bankrupt; this is why nonperforming loan to loan ratio (NPL) is used as an indicator of bankruptcy in this article. The average of the NPL ratio at banking network in the periods 2006-2014 is used as the threshold. Threshold is 5% and if NPL ratio in a bank is more than 5%, the bank is considered insolvent or at risk, otherwise it is assumed a healthy bank.

Initially, the univariate analysis of variance (ANOVA) test was applied to 43 ratios which were determined as the early warning indicators with discriminating ability for healthy and failed banks for one year in advance. Table 1, presents means and standard deviations of the financial ratios for two groups (non- failed and failed), and significance tests for the equality of group means for each ratio. F statistics and their observed significance levels are shown in the last two columns.

Ratios are presented in ascending order, according to the significance level of F statistics in table 1. The significant level is small (<5%) for the first 23 ratios. Hence, the null hypothesis that the means of two groups are equal is rejected at 5% significant level for these ratio. The other ratios displayed in Table 1 were excluded from the analysis. Since they were not able to split the banks into the healthy and failed groups, equality of group means for these ratios cannot be rejected at 5% significant level.

The other test statistics calculated in Table 1 is Wilks'lambda (Λ) which is the ratio of within-group sum of squares to the total sum of squares. Λ takes a value between 0 and 1 ($0 \le \Lambda \le 1$). $\Lambda \cong 1$ means all observed group means are equal. Values close to 0 occur when within- group variability is small compared to the total variability. That is, most of the total variability is attributable to differences between the means of the groups. As can be seen in Table 1, the group means of the first 23 ratios are most different for non-failed and failed banks.

Table 1: Test of Equality of Group Means for the Financial Ratios

Ratio	Symbol	Failed		Non failed		F	Wilks' Lambda	Sign
		Std. deviation	Mean	Std. deviation	Mean			
Capital to total asset	R1	7.917478	10.58593	22.61216	24.29852	18.466	0.83	0
Nonperforming loan to capital	R2	162.9365	174.0836	28.29479	22.41984	15.358	0.854	0
Capital to total liability	R3	11.60874	12.71327	67.69407	50.47874	21.181	0.809	0
Return on asset	R4	0.903939	0.76758	1.948621	2.57476	34.264	0.724	0
Profit before tax to total liability	R5	1.583369	1.6018	6.925829	6.70893	34.04	0.726	0
Liquidity assets to earning assets	R6	7.02862	14.95055	23.27665	28.65879	19.105	0.825	0
Growth of deposits	R7	37.22591	36.92486	215.2205	181.8163	30.785	0.745	0
Long time deposit to total asset	R8	15.68828	28.77775	13.7286	13.80331	13.801	0.867	0
Net interest revenue to earning assets	R9	2.745641	5.13523	5.747669	8.26972	11.514	0.887	0.001
Liquidity assets to total assets	R10	5.799908	11.69355	18.33189	20.24488	11.665	0.885	0.001
Net interest revenue to total asset	R11	9.052506	7.85259	20.64891	18.6239	11.427	0.887	0.001
Bonds to total asset	R12	1.836306	1.35276	19.42358	9.15375	11.907	0.883	0.001
Liabilities sensitive to interest rate / total asset	R13	13.53756	81.73045	22.59768	69.22718	9.234	0.907	0.003
Deposit to branches	R14	334.6661	401.3637	153.4447	161.2215	8.762	0.911	0.004
Liquidity assets to liquidity liability	R15	93.3276	54.10048	210.2864	144.0082	7.591	0.922	0.007
Total cost to total asset	R16	2.808915	9.80874	152.2417	56.22987	7.116	0.927	0.009
Non interest cost to earning assets	R17	2.665229	-5.24893	287.5039	-93.186	7.168	0.926	0.009
Noninterest revenue to total liability	R18	2.231837	-4.49074	418.3276	-128.675	6.754	0.93	0.011
Cost to income	R19	14.21251	86.74839	1478.917	523.314	6.677	0.931	0.011

Ratio	Symbol	Failed		Non failed		F	Wilks' Lambda	Sign
		Std. deviation	Mean	Std. deviation	Mean			
Net revenue to total loan	R20	3.33924	2.78372	515.5482	-150.713	6.793	0.93	0.011
Investment to capital	R21	61.23539	46.31862	13.46617	14.99679	4.618	0.951	0.034
Investment to shareholder equity	R22	61.23539	46.31862	13.46617	14.99679	4.618	0.951	0.034
Earning assets to interest bearing liability	R23	41.79327	129.428	177.9259	176.2998	4.3	0.954	0.041
Liquidity assets to short time liabilities	R24	3.256277	3.89104	6.775415	6.02463	3.816	0.959	0.054
Growth rate of net profit	R25	197.0235	20.00566	5461.812	-1171.83	3.63	0.961	0.06
Risk weighted asset to total asset	R26	175.0637	177.1482	33.9204	99.34252	3.495	0.963	0.065
Short time deposit to total asset	R27	9.22961	18.30028	22.92894	23.88462	2.681	0.971	0.105
Asset of each bank to total asset of banks	R28	4.378566	4.34077	3.770515	2.51704	2.641	0.971	0.108
Liquidity assets to current deposits	R29	161.5694	133.6515	211.123	205.6441	2.536	0.973	0.115
Net liquid assets to deposit	R30	54.26522	-55.2128	79.53034	-78.6437	2.218	0.976	0.14
Off balance sheet to total asset	R31	184.5007	90.73414	36.0139	31.4478	1.827	0.98	0.18
Net profit to capital	R32	12.35255	12.98851	10.68891	17.21238	1.777	0.981	0.186
Cost of fund to profit before tax	R33	2669.968	-670.327	440.6907	-197.882	0.555	0.994	0.458
Due to central bank / total liability	R34	10.96636	7.98287	18.05672	9.11647	0.117	0.999	0.733
Cost of fund to total assets	R35	1.030856	-0.82806	1.354332	-0.90276	0.067	0.999	0.797
Deposit per capita	R36	27.07669	33.09987	44.01464	31.20087	0.054	0.999	0.816
Growth of capital	R37	187.3151	58.32925	66.63217	53.55083	0.011	1	0.916
Loan to deposit	R38	76.52236	107.7497	70.62692	107.919	0	1	0.993

Source: Author's calculations.

In the following sections, Principal Component Analysis (PCA) was applied to 23 early warning ratios and the important factors for explaining changes in financial conditions of bank were determined. Factor scores were calculated for each bank, and these scores were used as independent variables in estimating parsimonious early warning models (discriminant, Logit and Probit).

3.2. Principal component analysis

The main objective of the principal component analysis (PCA) is to determine the important dimensions (characters) which can explain the changes in financial conditions of banks. PCA explores underlying patterns of relationship between the financial ratios; they must be correlated with each other for the PCA to be appropriate. Therefore, before proceeding to PCA, appropriateness of financial data to the PCA was evaluated. The evaluation was performed by Bartlett's test of sphericity. Bartlett's test can be used to test the null hypothesis in correlation matrix. In other words, all of the diagonal elements of the correlation matrix are equal to 1 and the rest of the elements are equal to 0 and no correlations exist between the ratios.

Table2: Results of KMO and Bartlett's test of sphericity

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.633
Bartlett's Test of Sphericity	Approx. Chi-Square	1135.748
	Df.	28
	Sig.	.000

Source: Research findings

Table A in appendix presents the correlation matrix of the ratios. Here, it can be seen that most of the ratios show correlation to each other. Table 2 presents the results of Bartlett's test of sphericity. The value of the chi-square test statistic for sphericity is large and observed significance level is small enough (<1% significant level), so the null hypothesis can be rejected.

In PCA, five common factors needed to represent the financial data, percentages of total variances explained by each factor were estimated (eigenvalues). Table 3, presents the estimated factors and their eigenvalues. In

PCA, financial ratios are expressed in a standardized form, with a mean of 0 and the standard deviation of 1. 23 financial ratios were used in the study; then each ratio's standardized variance is 1 and total variance is 23. Only those factors that account for variances greater than 1 (eigenvalue>1) were included in the model. Factors with variance less than one are not better than a single ratio, since each ratio has a variance of 1. Hence, the first 5 factors (F_1 : Liquidity, F_2 : Earning, F_3 : Sensitivity, F_4 : Capital adequacy and F_5 : Management) were included in the model. The estimated five- common factor model explains 81.32% of the total changes of financial conditions for the Iranian banks.

Table 3: Eigenvalues of the Factors

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	%of Variance	Cumulative %
1	7.924	34.453	34.453	7.924	34.453	34.453
2	4.826	20.983	55.436	4.826	20.983	55.436
3	2.599	11.302	66.738	2.599	11.302	66.738
4	1.811	7.875	74.613	1.811	7.875	74.613
5	1.544	6.712	81.325	1.544	6.712	81.325
6	.986	4.288	85.613			
7	.771	3.353	88.966			
8	.664	2.889	91.855			
9	.538	2.337	94.193			
10	.343	1.490	95.683			
11	.300	1.305	96.988			
12	.274	1.190	98.178			
13	.139	.605	98.782			
14	.116	.503	99.286			
15	.075	.324	99.610			
16	.061	.265	99.875			
17	.014	.060	99.934			
18	.010	.044	99.978			
19	.004	.019	99.997			
20	.000	.002	99.999			
21	.000	.001	100.000			
22	7.169E-5	.000	100.000			
23	2.322E-16	1.010E-15	100.000			

Source: Research findings.

Note: Extraction Method is Principal Component Analysis.

Factor (F_1) is the most important dimension in explaining changes of financial conditions of banks. It explains 34.4% of the total variance of the financial ratios. This result shows that liquidity of banks is the most important factor that affects banking failure. If banks do not have enough liquid assets to face withdrawal of deposits, they fail quickly.

Factor F_2 is earning which explains 20.9% of the total variance. As chronically unprofitable financial institutions risk insolvency, it is important to follow indicators of profitability. Declining trends in those indicators may signal problems regarding the profitability of financial institutions. On the other hand, unusually high profitability may be a sign of excessive risk- taking.

Factor F_3 is sensitivity which explains 11.3% of the total variance. Banks are increasingly involved in diversified operations, all of which involve one or more aspects of market risk. A high share of investments in volatile assets may signal a high vulnerability to fluctuations in the price of those assets. In general, the most relevant components of market risk are interest rate and foreign exchange risk, which tend to have significant impacts on financial institutions, assets and liabilities.

Factor F_4 is capital adequacy which ultimately determines the robustness of financial institutions to shocks facing their balance sheets. Thus, it is useful to track capital adequacy ratios that take into account the most important financial risks—foreign exchange, credit, and interest rate risks—as well risks involved in off-balance sheet operations, such as derivative positions. This factor explains 7.8% of the total variance. A declining trend in this ratio may signal increased risk exposure and possible capital adequacy problems. It is possible to estimate vulnerability based on average sector-wide capital adequacy ratios, but these may be misleading under some circumstances. In addition to adequacy, it may also be useful to monitor indicators of capital quality.

Factor F_5 is management that explains 6.7% of the total variance. Sound management is a key to financial institutions' performance. Indicators of the quality of management, however, are primarily applicable to individual institutions, and cannot be easily aggregated across the sector.

The other objective of the PCA is to calculate factor scores for each of banks according to the five factors determined. All ifnancial ratio are standardized, with a mean of 0 ($\pi \cong 0$) and the standard deviation of 1($\varpi \cong 1$) according to Eq. (1) in PCA. r is ratio and B is bank;

$$Z_{br} \cong \frac{R_{br} \; 0 \; \pi_r}{\varpi_r} \qquad r=1,2,\ldots,23 \qquad B=1,2,\ldots,30 \qquad (1)$$

Estimated factors can be expressed as a function of the observed original variables (ratios). In order to estimate the j th factor score (F_{bj}) for bank B, Equation (2) is used below:

$$F_{bj} \cong \sum_{r \cong 1}^{23} w_{rj} z_{br} \qquad j=1, 2,..,5 \qquad (2)$$

where, w_{rj} is the factor score coefficient, for the j th factor and r th ratio and z_{br} is the standardized value of the r th ratio for bank B. Table 4, presents the factor score coefficient matrix (w_{rj}) estimated by PCA.

To enhance the interpretability of the ifnancial factors, the varimax factor rotation method was used in PCA. This method minimizes the number of variables that have high weights on a factor. Table 5, presents the factor loadings. Here, variables with large loadings for the same factors are grouped and small factor loadings (<0.5) are omitted. Estimated factor represents a speciifc nifancial chracteristic ofthe bnaks under conisderati

After determination of the basic financial factors for the banks, early warning models (discriminant, Logit and Probit) were estimated according to these factors. The basic assumptions of the estimation of early warning models are based on banks possibility to be splited into two groups: The healthy and the failed groups. Thus, banks can be represented by a dummy dependent variable y_i such that, $y_i \cong 0$ if the ith bank is healthy, $y_i \cong 1$ if the ith bank is unhealthy (failed).

3.3. The discriminant model

In the discriminant analysis it is considered that any bank B is characterized by a vector of elements which are measurements of five independent variables (factors). For two populations (failed and healthy banks) it is assumed that the independent variables are distributed within each group according to multivariate normal distribution with different means but equal dispersion matrices.

Table 4: Factor Score Coefficients Matrix (w_{rj})

Ratios	Component				
	F_1	F_2	F_3	F_4	F_5
R15	.112	.011	-.027	-.039	.025
R2	-.045	-.007	.108	.001	-.112
R7	.112	.052	.022	-.006	-.064
R26	.080	.041	-.035	.266	.183
R6	.110	.062	.016	.063	-.038
R12	.108	.036	-.004	.101	-.078
R30	.006	.008	.003	.255	.437
R1	-.048	.008	.106	.390	-.171
R25	.067	-.006	.064	.013	.393
R8	.105	.056	-.011	.096	-.099
R31	.077	.054	.262	-.055	.071
R10	.101	.087	.011	.017	-.047
R29	-.087	.000	.038	.276	.172
R3	-.028	.012	.076	.225	-.348
R28	.062	.033	.007	.243	-.199
R17	.049	-.189	.025	.047	-.024
R11	-.052	.188	-.020	-.021	.016
R5	-.053	.186	-.021	-.022	.021
R19	.049	-.190	.020	.025	-.026
R14	-.013	.005	.372	-.076	.021
R4	-.044	.193	-.010	-.023	.023
R35	-.013	.005	.372	-.076	.021
R24	.101	.058	-.024	-.185	-.066

Source: Research findings. Method: Principal Component Analysis.

Table 5: Factor Loadings

Ratios	Component				
	F_1	F_2	F_3	F_4	F_5
R6	.921				
R7	.912				
R8	.899				
R10	.894				
R12	.885				
R15	.809				
R24	.807				
R26	.645				.562
R29	-.621				
R28	.614				
R4		.996			
R19		-.996			
R17		-.995			
R11		.994			
R5		.990			
R14			.974		
R35			.974		
R31	.626		.705		
R1				.840	
R3				.701	
R30					.818
R25					.563

Source: Research findings

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

The objective of this method is to obtain the linear combination of the independent variables that maximizes the variances between the populations relative to within-group variance. The linear combination of the factors scores provide a D-score for each bank, according to the estimated canonical discriminant model below:

$$D_B \cong 1.306F_1 \;.\; 0.060F_2 \; 0 \; 0.231F_3 \;.\; 0.286F_4 \; 0 \; 0.421F_5 \qquad (3)$$

In Eq 3, D_B is the D-score for bank B and F_1, F_2, F_3, F_4, F_5 represent liquidity, earning, sensitivity, capital adequacy and asset quality.

Table 6: The Statistics of the Estimated Discriminant Model

Eigenvalue	Canonical correlation	Wilks' Lambda
0.806	0.484	0.766

In order to evaluate effectiveness of the estimated discriminant model, the model statistics were calculated in table 6. An effective discriminant model is one that has much between-group variability of D-scores when compared to within- groups variability of D-scores. Coefficients of the discriminant model are chosen so that the ratio of the between-groups to within-groups sum of squares of D-scores is as large as possible. Any other linear combination of the predictor variables will have smaller ratios. The eigenvalue statistic presented in Table 6, is the ratio of the between groups to within-groups sum of squares of D-scores. Large eigenvalue (0.806) shows that the estimated discriminant model has high discriminating ability. Canonical correlation (0.484) is the measure of degree of association between D-scores and the group variable that is coded 0 for non-failed banks and 1 for failed banks. As stated previously, small value of Wilks' lambda (0.566) means that the most of the total variability is attributable to differences between the means of the D-score of the groups.

Based on its D-score and the calculated cut-off score (C) in Eq. (4), a bank is classified in either the failed or the healthy group. The optimum cut-off score is calculated approximately equal to 0.83, as the un-weighted average of the D-scores of the failed and the healthy bank groups: where C cut-off score:

$$C \cong \frac{Dscore_{healthy} \cdot Dscor_{faild}}{2} \cong 0.83 \tag{4}$$

The classification is made by the following procedure: If D-score > C, the bank is classified in theihealthy group, if D-score < C, the bank is classified in the failed group.

3.4. Logit and Probit models

The Logit analysis is based on a cumulative logistic function; providing the probability of a bank belonging to one of the prescribed classes, given the financial characteristics of the bank. In the Logit method the probability of bank going to failure (P_{l_b}) is calculated using the cumulative logistic function:

$$P_{l_b} \cong \frac{1}{1 .\ e^{Z_{l_k}}} \tag{5}$$

where

$$Z_{l_b} \cong \varepsilon_1 F_{1b} \ .\ \varepsilon_2 F_{2b} \ .\ \varepsilon_3 F_{3b} \ .\ \varepsilon_4 F_{4b} \ .\ \varepsilon_5 F_{5b} \tag{6}$$

Based on that probability a bank is classified as failed or non-failed, using a cut-off probability, attempts to minimize the type I (failed banks classified as healthy) and type II (healthy banks classified as failed) errors. Maximization of the log-likelihood function provides the following Z_{l_b} equation in the Logit analysis:

$$Z_{l_b} \cong 2.019F_{1b} \ .\ 4.9F_{2b} \ 0\, 2.2F_{3b} \ .\ 1.7F_{4b} \ 0\, 5.6F_{5b} \tag{7}$$

Table 7: Test Statistics for the Estimated Logit and Probit Models

	Coefficients	Std. error	Z-statistic	Sig.
Logit				
F1	2.01	0.8	2.5	0.000
F2	4.9	2.2	2.2	0.000
F3	-2.2	1.02	-2.1	0.000
F4	1.7	0.45	3.7	0.000
F5	-5.6	2.4	-2.3	0.000
Probit				
F1	3.1	1.4	2.21	0.001
F2	1.9	0.8	2.37	0.005
F3	-1.28	0.56	-2.28	0.000
F4	2.39	1.5	1.59	0.003
F5	2.32	1.2	1.93	0.000

Source: Research findings

In the Probit method the probability (P_{P_b}) of a bank going to failure is given by cumulative standard normal distribution function:

$$P_{P_b} \cong \int_{0\clubsuit}^{Z_{P_b}} \frac{1}{\sqrt{2\sigma}} e^{\frac{0z}{2}} dz . \tag{8}$$

As in the Logit analysis, maximization of the log-likelihood function provides the following (Z_{p_b}) equation in the Probit analysis:

$$Z_{p_b} \cong 3.1F_{1b} \,.\, 1.9F_{2b} \; 0\, 1.28F_{3b} \,.\, 2.39F_{4b} \; 0\, 2.32F_{5b} \tag{9}$$

Table 7 presents the calculated test statistics for the estimated coefficients of the Logit and Probit models. All of the coefficients of the Logit and Probit models are statistically significant according to the observed significance level of z-statistics corresponding to the standard errors of the coefficients. A bank is classified in the failed or healthy group according to the estimated Logit and Probit models, based on a cut-off probability of 0.50 ($P_c \cong 0.50$) and calculated failure probabilities (P_{L_b} and P_{P_b}). The classifications were made by the following procedure:

If failure probability < P_c , the bank is classified in the healthy group; if failure probability $\propto P_c$, the bank is classified in the failed group.

We use Hosmer and Lemeshow Test and Pearson Goodness-of-Fit Test for examination of goodness of fit of Logit and Probit models. Table 8 presents the calculated Goodness-of-Fit test statistics. The significant level is small (<5%) for the two models. This table gives the overall test for the model that includes the predictors. The chi-square value of 20.961 and 46.560 with a p-value of less than 0. 05 tells us that our model as a whole fits significantly.

The -2*log likelihood (103.217) in the table 9 can be used in comparison to nested model. This table also gives two measures of pseudo R-square. We see that Nagelkerke's R^2 is 0.427 which indicates that the model is good. Cox

& Snell's R^2 are the n^{th} root. Thus, we can interpret this as 26% probability of the event passing the exam which is explained by the logistic model.

Table 8: Goodness-of-Fit Test for the Estimated Logit and Probit Models

Test	Chi-square	Def.	Sig.
Logit- Hosmer and Lemeshow Test	20.961	8	.007
Probit- Pearson Goodness-of-Fit Test	46.560	8	.0001

Source: Research findings

Table 9: Other Goodness-of-Fit Test for the Estimated Logit Models

Step	Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	103.217[a]	.269	.427

Source: Research findings

4. Concluding Remarks

Bank failures threaten the economic system as a whole. Therefore, predicting bank financial failures is crucial to prevent and lessen the incoming negative effects on the economic system. This is originally a classification problem to categorize banks as healthy or non-healthy.

The objective of this paper is to construct an appropriate model to identify healthy and non-healthy banks in Iran. Banking system of Iran has no regulations and characteristics that can distinguish between failure and healthy banks. Despite these limitations, we have used non-performing loan (Npl) to loan ratio>5% as an index to identify the failure banks.

For this purpose we have used 30 banks of Iran and their financial ratios during 2006-2014. Well known multivariate statistical technique (principal component analysis) was used to explore the basic ifnancial characteristics of the banks, and discriminant, Logit and Probit models were estimated based on these characteristics to construct the model.

Results of the study show that PCA is a useful tool to explicitly explore the ifinancial characteristics of the banking system and compare the banks with respect to these characteristics, and thus, determine differences in the financial structures of the banks.

In the discriminant analysis for two populations (failed and healthy banks), it is assumed that the independent variables are distributed within each group according to multivariate normal distribution with different means but equal dispersion matrices. Based on D-score and the calculated cut-off score, if D-score > C, the bank is classified as healthy and otherwise, the bank is classified in the failed group. Logit and Probit analysis provides the probability of a bank belonging to one of the prescribed classes.

Finally, PCA, discriminant analysis, Logit and Probit models could be used as alternative or supplementary decision support tools to the CAMELS rating system in bank examination process. These models help to supervise and identify non - healthy (unhealthy) banks and the probability of failure. It is important to find out one bank has failed. This paper proposes a model design to identify the time of failure.

References

Alam, P., L.K. Booth, and T. Thordason, (2000). “The Use of Fuzzy Clustering Algorithm and Self-organizing Neural Networks for Identifying Potentially Failing Banks: An Experimental Study”. *Expert Systems with Applications* 18, 185–199.

Arabi, Khalafalla Ahmed Mohamed, (2013). “Predicting Bank’s Failure: The Case of Banking Sector in Sudan for the Period (2002-2009)”. *Journal of business Studies Quarterly*, Vol, 4. No, 3.

Canbas, Serpil; Cabuk, Altan; Bilgin Kilic, Suleyman, (2005). “Prediction of Commercial Bank Failure via Multivariate Statistical Analysis of Financial Structures: The Turkish Case”’ *European Journal of operational Research* 166 (2005), 528-546.

Cole, R.A., J.W., Gunther, (1998). "Predicting Bank Failures: A Comparison of On-and Off-site Monitoring Systems". *Journal of Financial Services Research* 13 (2), 103–117.

Jo, H., I. Han, (1996). "Integration of Case-based Forecasting, Neural Network, and Discriminant Analysis for Bankruptcy Prediction". *Expert Systems with Applications* 11, 415–422.

Kolari, J., D. Glennon, H. Shin, and, M. Caputo, (2002). "Predicting Large US Commercial Bank Failures". *Journal of Economics & Business* 54, 361–387.

Lam, K.F., and, J. W. Moy, (2002). "Combining Discriminant Methods in Solving Classification Problems in Two-group Discriminant Analysis". *European Journal of Operational Research* 138, 294–301.

Mousseau, V., R. Slowinski, and P., Zielniewicz, (2000). "A User Oriented Implementation of the ELECTRE-TRI Method Integrating Preference Elicitation Support". *Computers & Operational Research* 27, 257–777.

Pantolone, C., M.B., Platt, (1987). "Predicting Commercial Bank Failure since Deregulation". *New England Economic Review*, 37–47.

Rose, P.S., J.W., Kolari, (1985). "Early Warning Systems as a Monitoring Device for Bank Condition". *Quarterly Journal of Business and Economics* 24 (winter), 43–60.

Samad, Abdus, (2012). "Credit Risk Determinants of Failure: Evidence from US Bank Failure". *International Business Research*; Vol. 5, No. 9.

Tam, K.Y., M. Y. Kiang, (1992). "Managerial Applications of Neural Networks: The Case of Bank Failure Predictions". *Management Science* 38 (7).

Thomson, J.B., (1991). "Predicting Bank Failure in the 1980s". *Economic Review* (Federal Reserve Bank of Cleveland) 27.

Zaghdoudi, Taha, (2013). "Bank Failure Prediction with Logistic Regression". *International Journal of Economics and Financial Issues,* Vol. 3, No. 2, PP. 537-543.

Zopounidis, C., M. Doumpos, (1999). "A Multicriteria Decision Aid Methodology for Sorting Decision Problems: The Case of Financial Distress". *Computational Economics* 14 (3), 197-218.

Appendix

Table A: Correlation Matrix of the Ratios

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17	R18	R19	R20	R21	R22	R23
R1	1.0	-0.4	0.9	0.0	-0.2	0.4	-0.4	-0.7	0.4	0.4	-0.4	0.4	0.8	0.7	0.6	0.0	-0.2	0.7	-0.2	0.1	0.5	-0.4	0.5
R2	-0.4	1.0	-0.2	-0.1	0.1	0.0	0.0	0.3	-0.2	-0.1	0.0	0.0	-0.2	-0.1	-0.1	-0.1	0.1	-0.1	0.0	-0.2	-0.2	0.0	-0.1
R3	0.9	-0.2	1.0	0.0	-0.1	0.5	-0.5	-0.7	0.3	0.3	-0.5	0.5	0.8	0.7	0.7	-0.1	-0.1	0.8	-0.2	0.1	0.5	-0.5	0.7
R4	0.0	-0.1	0.0	1.0	0.0	-0.1	0.1	0.0	0.1	-0.3	0.1	-0.1	0.0	0.1	0.2	0.1	0.0	-0.1	0.4	0.2	0.0	0.1	-0.1
R5	-0.2	0.1	-0.1	0.0	1.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	-0.1	0.0	-0.1	0.2	1.0	-0.1	0.0	0.0	0.3	0.0	0.0
R6	0.4	0.0	0.5	-0.1	0.0	1.0	-1.0	-0.3	0.0	0.2	-1.0	1.0	0.3	0.2	0.1	-0.2	0.0	0.1	-0.1	0.1	0.0	-1.0	-0.2
R7	-0.4	0.0	-0.5	0.1	0.0	-1.0	1.0	0.3	0.0	-0.2	1.0	-1.0	-0.3	-0.2	-0.1	0.2	0.0	-0.1	0.1	0.0	-0.1	1.0	0.1
R8	-0.7	0.3	-0.7	0.0	0.0	-0.3	0.3	1.0	0.0	-0.2	0.3	-0.3	-0.5	-0.4	-0.4	-0.1	0.0	-0.7	0.4	0.1	-0.5	0.3	-0.4
R9	0.4	-0.2	0.3	0.1	0.0	0.0	0.0	0.0	1.0	0.4	0.0	0.0	0.7	0.5	0.4	0.4	0.0	0.1	0.3	0.2	0.3	0.0	0.3
R10	0.4	-0.1	0.3	-0.3	-0.1	0.2	-0.2	-0.2	0.4	1.0	-0.2	0.2	0.4	0.2	0.1	0.4	-0.1	0.2	-0.2	-0.1	0.5	-0.2	0.1
R11	-0.4	0.0	-0.5	0.1	0.0	-1.0	1.0	0.3	0.0	-0.2	1.0	-1.0	-0.3	-0.2	-0.1	0.2	0.0	-0.1	0.1	0.0	-0.1	1.0	0.1
R12	0.4	0.0	0.5	-0.1	0.0	1.0	-1.0	-0.3	0.0	0.2	-1.0	1.0	0.3	0.2	0.1	-0.2	0.0	0.1	-0.1	0.0	0.0	-1.0	-0.2
R13	0.8	-0.2	0.8	0.0	-0.1	0.3	-0.3	-0.5	0.7	0.4	-0.3	0.3	1.0	0.8	0.8	0.1	-0.1	0.7	0.1	0.2	0.6	-0.2	0.7
R14	0.7	-0.1	0.7	0.1	0.0	0.2	-0.2	-0.4	0.5	0.2	-0.2	0.2	0.8	1.0	1.0	0.1	0.0	0.6	0.0	0.2	0.4	-0.2	0.7
R15	0.6	-0.1	0.7	0.2	-0.1	0.1	-0.1	-0.4	0.4	0.1	-0.1	0.1	0.8	1.0	1.0	0.0	-0.1	0.7	0.0	0.3	0.4	0.0	0.8
R16	0.0	-0.1	-0.1	0.1	0.2	-0.2	0.2	-0.1	0.4	0.4	0.2	-0.2	0.1	0.1	0.0	1.0	0.2	-0.2	0.0	-0.1	0.3	0.2	-0.1
R17	-0.2	0.1	-0.1	0.0	1.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	-0.1	0.0	-0.1	0.2	1.0	-0.1	0.0	0.0	0.3	0.0	0.0
R18	0.7	-0.1	0.8	-0.1	-0.1	0.1	-0.1	-0.7	0.1	0.2	-0.1	0.1	0.7	0.6	0.7	-0.2	-0.1	1.0	-0.4	0.0	0.5	-0.1	0.8
R19	-0.2	0.0	-0.2	0.4	0.0	-0.1	0.1	0.4	0.3	-0.2	0.1	-0.1	0.1	0.0	0.0	0.0	0.0	-0.4	1.0	0.3	-0.2	0.1	-0.1
R20	0.1	-0.2	0.1	0.2	0.0	0.1	0.0	0.1	0.2	-0.1	0.0	0.0	0.2	0.2	0.3	-0.1	0.0	0.0	0.3	1.0	0.0	0.0	0.1
R21	0.5	-0.2	0.5	0.0	0.3	0.0	-0.1	-0.5	0.3	0.5	-0.1	0.0	0.6	0.4	0.4	0.3	0.3	0.5	-0.2	0.0	1.0	0.0	0.5
R22	-0.4	0.0	-0.5	0.1	0.0	-1.0	1.0	0.3	0.0	-0.2	1.0	-1.0	-0.2	-0.2	0.0	0.2	0.0	-0.1	0.1	0.0	0.0	1.0	0.2
R23	0.5	-0.1	0.7	-0.1	0.0	-0.2	0.1	-0.4	0.3	0.1	0.1	-0.2	0.7	0.7	0.8	-0.1	0.0	0.8	-0.1	0.1	0.5	0.2	1.0