ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
UDC/UDK 336.77:336.711]:519.233.5

A COMBINED PRINCIPAL COMPONENT ANALYSIS-REGRESSION ANALYSIS MODEL TO STUDY THE EFFECT ON TECHNICAL EFFICIENCY OF BAD LOANS IN BANK INDUSTRY

Omid Mashhadifarahani, Narges Rezavi, Loghman Hatami-Shirkouhi

Original scientific paper
This paper presents a combined principal component analysis-regression analysis (PCA-RA) model to study the effects of bad loans on the economic performance of banking systems. This model first applies PCA to calculate the overall technical efficiency of banks and then uses regression analysis to find the impact of bad loans on technical efficiency. Bad loans or Non-performing loans (NPLs) include: past due loans; bankrupt and quasi- bankrupt assets, and doubtful assets. The technical efficiency of banks is measured with reference to three input indicators: number of branches, deposits, and costs; and three output indicators: income, profit, and loans. Then technical efficiency is regressed on two explanatory variables of loans and bad loans. Results of a case study in governmental banks in Iran show that bad loans have negative impact on the technical efficiency of banks.

Keywords: *bad loans, banking industry, non-performing loans, principal component analysis, regression analysis, technical efficiency*

Primjena modela sastavljenog od analize glavne sastavnice i regresijske analize u proučavanju utjecaja loših kredita na tehničku učinkovitost u bankarskoj industriji

Izvorni znanstveni članak
U radu je predstavljen model s kombinacijom analize glavne sastavnice i regresijske analize (PCA-RA) za proučavanje djelovanja loših kredita na ekonomsku učinkovitost bankovnih sustava. U modelu se najprije primjenjuje PCA za izračunavanje ukupne tehničke učinkovitosti banaka i zatim koristi regresijska analiza kako bi se ustanovilo djelovanje loših kredita na tehničku učinkovitost. Loši krediti ili Neučinkoviti krediti uključuju: kredite kojima je prošao rok vraćanja, potraživanja kod bankrota ili kvazi-bankrota, i sumnjiva potraživanja. Tehnička učinkovitost banaka mjeri se u odnosu na tri ulazna indikatora: dohodak, profit i krediti. Tada dolazi do regresije tehničke učinkovitosti na dvije objašnjavajuće varijable kredita i loših kredita. Rezultati analize slučaja u državnim bankama u Iranu pokazuju da loši krediti negativno djeluju na tehničku učinkovitost banaka.

Ključne riječi: *analiza glavne sastavnice, bankarska industrija, loši krediti, neučinkoviti krediti, regresijska analiza, tehnička učinkovitost*

1 Introduction

The standard methods applied in banking are the intermediation and production approaches. Under the intermediation approach, banks use purchased funds together with physical inputs to produce various assets (measured by their value). The production approach assumes that banks use only physical inputs such as labour and capital to produce deposits and various assets (measured by the number of deposit and loan accounts at a bank, or by the number of transactions for each product). We adopt a combined approach of the intermediation approach and production approach to assess the technical efficiency of banks.

Berger and Humphrey [1] discussed several approaches of modelling the bank production process: the production approach, user-cost approach, value added approach and dual approach. Berger and Humphrey [1] suggested the intermediation approach is best suited for evaluating bank efficiency, whereas the production approach is appropriate for evaluating the efficiency of bank branches. Koutsomanoli-Filippaki et al. [2] employed the directional technology distance function approach to decompose profit efficiency into its technical and allocative components.

Traditionally, multivariate techniques are extensively used for assessing the performance of banking systems. Data envelopment analysis is one of the popular tools in this respect [3-6]. Frontier techniques such as stochastic frontier analysis (SFA) are also used for bank efficiency assessment. Perera and Skully [7] investigated the consistency of parametric stochastic frontier analysis (SFA) and nonparametric data envelopment analysis (DEA) estimates for bank efficiency. Furthermore, Principal components analysis (PCA) and numerical taxonomy (NT) are used and applied to verify and validate DEA findings in Azadeh et al. [8]. PCA is one of the multivariate analysis techniques usually used for correlation analysis, data reduction and also efficiency assessment [9, 10]. Zhongsheng and Dong [11] analysed the operating efficiency of Chinese commercial banks during 1999 to 2003 based on PCA.

Formally, the bad loan is defined to be a debt instrument (loan) whose contractual interest and principal payments are difficult to collect. The effects of bad loans have also been addressed in the literature. Using DEA, Chen [12] evidenced that bank loss on bad loans is one of the major reasons for lower cost efficiency of privately owned banks than publicly owned banks in Taiwan. Matthews and Zhang [13] considered and incorporated non-performing loans as a bad output for calculating Malmquist total factor productivity growth index of nationwide commercial banks of China. Barros et al. [14] analysed technical efficiency of the Japanese banks based on the Non-radial directional model with undesirable output that takes into consideration not only desirable outputs but also an undesirable output that is represented by non-performing loans (NPLs).

To the best knowledge of the authors there is no study that used PCA for technical efficiency of banks with bad outputs (i.e. Non-operating or bad loans) and their effect on the overall technical efficiency. This paper presents a combined PCA-Regression analysis for efficiency assessment and analysis of the effect of bad loans on the technical efficiency of banks.

The following inputs/outputs variables are defined in this study. The inputs include three indicators: Number of Branches, Deposits, and Costs; and The outputs include three indicators: Income, Profit, and Loans. PCA indicators are defined as the ratio of output variables to input variables hence we have 9 PCA indicators. The notion of dividing outputs by inputs reflects an indication of efficiency from both production and intermediation points of view. PCA will find weights for outputs and inputs and use them to calculate a total efficiency score of the banks namely technical efficiency. Then technical efficiency is regressed on two explanatory variables of loans and bad loans. Finding the relationship between technical efficiency and bad loans will help analysing the effect of bad loans on the technical efficiency of banks under study.

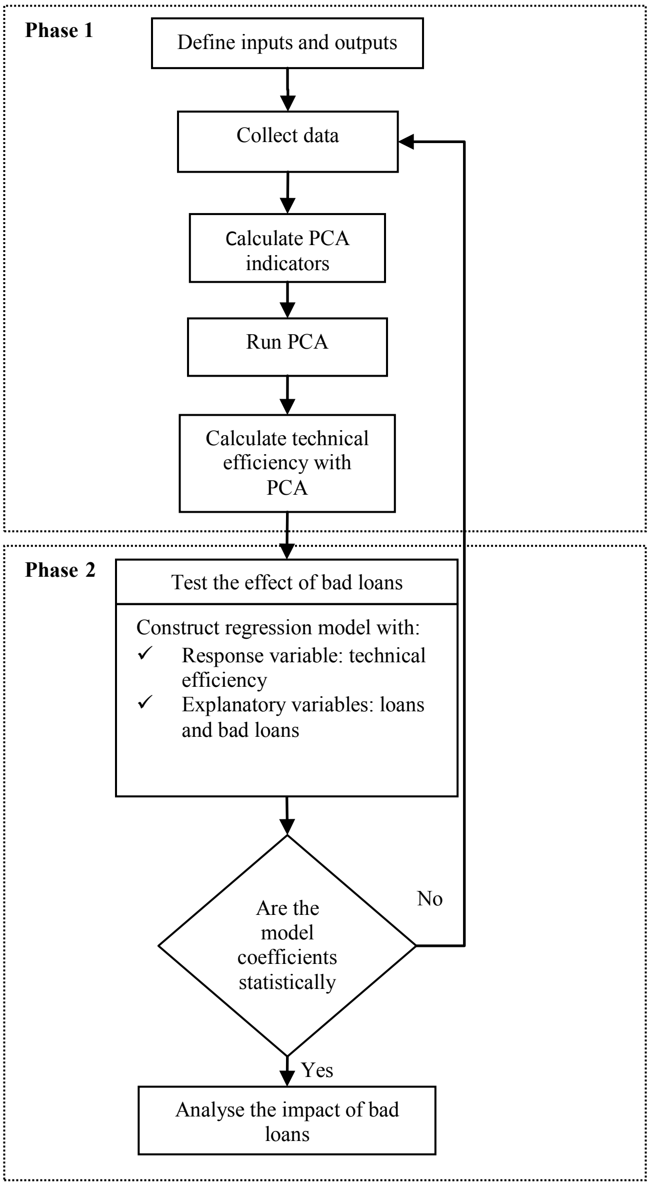

Figure 1 The working algorithm of PCA–regression

This paper is organized as follows. In Section 2, a step-by-step working algorithm of the study is presented. Section 3 introduces the case study; data and variables for technical efficiency of banks. In Section 4, the results of PCA for technical efficiency are presented. Also, the PCA procedure of this study is presented in this section. Section 5 analyses the impact of bad loans on technical efficiency. Section 6 presents the main findings and conclusion.

2 The working algorithm of the study

In this section the proposed working algorithm of the paper is presented. At its early stage, in the first phase, this algorithm proposes to collect data and pre-process them using PCA. This pre-processing involves the use of PCA and Eigen structure of data to find a weight for each of PCA variables to be used for integrating all PCA indicators and calculate an overall efficiency score for each bank in a year. The second phase involves the use of regression analysis as a statistical analysis tool to test the significance of the relationship between bad loans and technical efficiency of banks. The construct of regression model is such that the model uses technical efficiency as the response variable and two variables loans and bad loans as explanatory variables. Standard statistical tools such as MINITAB could be used to estimate the regression coefficient and test the statistical significance of coefficients. This working algorithm is presented in Fig. 1.

3 The case study: data and variables

As a case study, to show the applicability and usefulness of the working algorithm of this paper, we have considered the performance indicator of seven governmental banks in Iran. The data are related to seven performance indicators of these banks in the time period 2006 ÷ 2010. Tab. 1 presents some descriptive statistics of the collected data.

Table 1 Statistical characteristics of the applied variables (Billions Iranian Rials)

Variable	Obs. Mean	Std. dev.	Min	Max
Inputs				
Number of Branches	2 027	825	973	3 277
Deposits	336 299	187 665	53 438	789 314
Costs	13 849	8 007	2 244	36 111
Output				
Income	15 656	8 844	2 360	36 544
Profit	1 806	1 867	112	7 776
Loans	185 027	97 436	36 056	432 850
Non-performing loans	35 495	24 070	4 543	89 228

4 Applied PCA

PCA is a multivariate technique commonly used for data reduction and performance assessment. When it is used for performance assessment of decision-making units (DMUs), first it is applied to find a weight for the variables under study and then ranking of DMUs. Generally, the DMUs utilize a variety of resources as inputs to produce several outputs.

Basically, PCA finds weights according to the Eigen value and Eigen vectors of a specific matrix describing the covariance or correlation between variables. This would eventually lead to scoring and rankings of DMUs. A precise description of procedure of ranking with PCA can be found in Zhu [15] and Premachandra [16], Slišković et al. [17].

In this section, the PCA procedure is illustrated step by step. There are three input variables and three output variables. Dividing each output by all three inputs will yield 9 output/input ratios which are considered as PCA indicators. In the case study we considered the performance data of seven governmental banks in 5 years from 2006 to 2010. So we have 35 DMUs for PCA. Hereafter we call each DMU as bank-year. PCA is performed by identifying Eigen structure of the covariance or singular value decomposition of the original data. Here, the former approach will be used.

Suppose $X = (x_1, x_2)_{35\times 9}$ is a 35×9 matrix composed by x_{ij}'s defined as the value of the jth variable for the i^{th} bank-year and therefore $x_m = (x_{1m}, \ldots, x_{35m})^T$ (m = 1,2,..., 9). Furthermore, suppose $\hat{X} = (\hat{x}_1 \ldots \hat{x}_9)_{35\times 9}$ is the standardized matrix of X and therefore $\hat{x}_m = (\hat{x}_{1m}, \ldots, \hat{x}_{km})^T$. PCA is performed to identify new independent variables or principal components (defined as Y_j for j = 1, 2,..., 9), which are respectively different linear combination of $\hat{x}_1, . \hat{x}_9$. As mentioned, this is achieved by identifying Eigen structure of the covariance of the original data. The principal components are defined by a 35×9 matrix $Y = (y_1, y_9)_{35\times 9}$ composed by y_{ij}'s are shown by:

$$y_m = \sum_{j=1}^{9} l_{mj} \hat{x}_j, \; m = 1, \ldots, 9, \tag{1}$$

where, j_{mj} is the coefficient of the m^{th} variable for the j^{th} principal component. The j_{mj}'s are estimated such that the conditions of Eqs. 6 and 7 are met. Y_1 accounts for the maximum variance in the data, y_2 accounts for the maximum variance that has not been accounted by y_1, and so on.

$$l_{m1}^2 + l_{m2}^2 = 1, \; m = 1, \ldots, 9 \tag{2}$$

$$l_{m1} \cdot l_{n1} + l_{m2} \cdot l_{n2} = 0, \; \text{for all } m \neq n, n = 1, \ldots, 9. \tag{3}$$

For obtaining the l_{ij}'s and consequently vectors ($y_{1j}, \ldots, y_{35j}$) (j = 1,..., 9) and PCA scores the following steps are performed:

Step 1: Calculate the sample mean vector $\overline{x}$ and covariance matrix S (Eqs. 4, 5 and 6):

$$\overline{x} = (\overline{x}_1, \ldots, \overline{x}_9)_{1\times 9}, \tag{4}$$

in which,

$$\overline{x}_j = \frac{1}{35} \sum_{i=1}^{35} x_{ij} \; \text{ for } \; j = 1, \ldots, 9, \tag{5}$$

$$S = (s_{jq})_{2\times 2} = \frac{1}{35} (X - \overline{x})^T (X - \overline{x}) \; \text{ for } \; q = 1, \ldots, 9, \tag{6}$$

Step 2: Calculate the sample correlation matrix.

$R = \left(C_1 \big/ \sqrt{s_{jj}} \right) \cdot S \cdot C_1 \big/ \sqrt{s_{jj}}$ where $C_1 \big/ \sqrt{s_{jj}}$ is a 9×9 diagonal matrix whose j^{th} diagonal element is a $1 \big/ \sqrt{s_{jj}}$ for j = 1,..., 9.

Step 3: Solve the following equation:
$|R - \lambda \cdot I_9| = 0$ where I_9 is a 9×9 identity matrix. We obtain the ordered 9 characteristic roots (eigenvalues) $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_9$ with $\sum_{j=1}^{9} \lambda_j = 9$ and the related 9 characteristic vectors (eigenvectors) ($l_{m1}, \ldots, l_{m35}$) (m = 1,..., 9).

Those characteristic vectors compose the principal components Y_i. The components in eigenvectors are respectively the coefficients in each corresponding Y_i (Eq. 7):

$$Y_m = \sum_{j=1}^{9} l_{mj} \overline{x}_{ij} \; \text{ for } \; m = 1, \ldots, 9 \; \text{ and } \; i = 1, \ldots, 35. \tag{7}$$

Step 4: Calculate the weights w_j of the principal components and PCA scores (Z_i of each country-year i = 1,…, 35). Furthermore, the Z vector ($Z_1, \ldots, Z_{35}$) where Z_i shows the score of each bank-year, is given by Eqs. 8 and 9:

$$w_j = \frac{\lambda_j}{\sum_{j=1}^{9} \lambda_j} = \frac{\lambda_j}{9}, \; j = 1, \ldots, 9, \tag{8}$$

$$z_i = \sum_{j=1}^{9} w_j Y_j, \; i = 1, \ldots, 35. \tag{9}$$

The PCA efficiency scores and ranking for each bank-year are presented in Tab. 2.

The results of full ranking in Tab. 2 show that Refah K. Bank in years 2006 and 2007 has the highest technical efficiency and is the best performing bank. This bank also performed very well in years 2008 and 2009 due to its ranks of 4 and 5 between all 35 DMUs (bank-year). In the table of ranking, Refah K. Bank as the best is followed by Bank Saderat Iran. This bank has got ranks 3, 7, 8, and 10 in years 2006, 2007, 2008, and 2009, respectively. Moreover, Fig. 2 shows the time trend of technical efficiency of banks. This figure shows that technical efficiency in the governmental banks has decreased during 2006 ÷ 2010.

5 Regression analysis – the impact of bad loans

In this section, to test the significance of the relationship between bad loans and technical efficiency, a linear regression model is constructed. The model uses technical efficiency as the response variable and two variables loans and bad loans as explanatory variables. Data are presented in the last three columns of Tab. 2. The MINITAB software is used to estimate the regression coefficient and test the statistical significance of coefficients. Results are presented in Tab. 3.

Table 2 PCA technical efficiency scores and loans data

Bank	Year	Technical efficiency	Rank	Loans	Bad loans
Tejarat Bank	2006	0,948948	6	94 130	26 308
	2007	0,903886	13	126 900	27 630
	2008	0,844313	19	164 152	41 372
	2009	0,795983	24	193 449	47 972
	2010	0,720045	29	233 823	61 139
Bank Sepah	2006	0,906995	11	125 235	8 144
	2007	0,849931	18	153 765	21 104
	2008	0,850215	17	136 969	39 133
	2009	0,870839	15	132 509	34 415
	2010	0,855857	16	152 585	29 024
Bank Saderat Iran	2006	0,966067	3	146 190	26 798
	2007	0,93663	7	179 367	33 368
	2008	0,931413	8	163 804	42 665
	2009	0,907565	10	189 709	48 709
	2010	0,881984	14	216 589	47 591
Bank Maskan	2006	0,816716	22	107 186	4 609
	2007	0,787861	25	130 218	4 543
	2008	0,742201	27	158 000	6 915
	2009	0,656714	32	202 887	10 004
	2010	0,5	35	287 555	12 786
Bank Mellat	2006	0,829893	20	166 556	48 286
	2007	0,724589	28	230 572	62 014
	2008	0,705376	31	248 749	52 435
	2009	0,64225	33	293 013	48 032
	2010	0,540514	34	380 122	38 012
Bank Melli Iran	2006	0,909956	9	203 120	43 204
	2007	0,821794	21	292 319	76 402
	2008	0,798909	23	319 977	80 372
	2009	0,763613	26	367 419	81 117
	2010	0,710925	30	432 850	89 228
Refah K. Bank	2006	1	1	36 056	6 648
	2007	0,977078	2	41 586	10 107
	2008	0,966064	4	46 209	10 596
	2009	0,95812	5	48 738	11 197
	2010	0,904357	12	73 636	10 452

Table 3 The results of regression model estimation

The regression equation is
Technical efficiency = 0,993 + 0,000002 loans – 0,000004 bad loans

Predictor	Coef	SE Coef	T	P-value
Constant	0,99279	0,02315	42,88	0,000
loans	0,00000160	0,00000017	9,25	0,000
bad loans	–0,00000364	0,00000070	–5,21	0,000

$R-Sq = 74{,}6\ \%$

As seen in Tab. 3, the regression coefficients of the model are significant due to a very small value of P-values which are the risk of rejecting the null hypothesis. The null hypothesis is that regression coefficients are not significant. Because these risks are very low, we reject the null hypothesis and conclude that the coefficients are statistically significant. Furthermore, as the final finding, the negative sign of coefficient of bad loans (–0,00000364) suggests that this variable has a negative impact on the technical efficiency of banks.

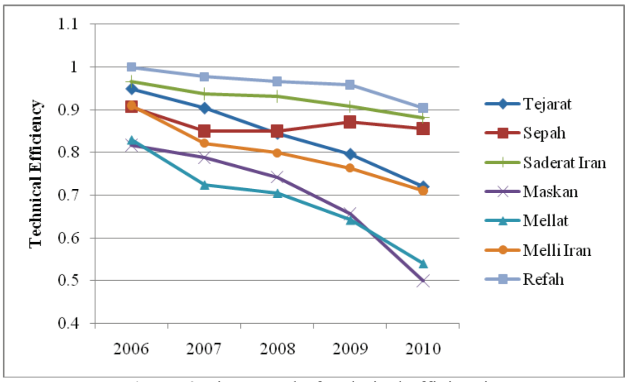

Figure 2 Time trend of technical efficiencies

6 Conclusion

A combined principal component analysis-regression analysis model to study the effects of bad loans on the economic performance of banking systems was presented in this paper. An overall working algorithm was presented which is capable of calculating the overall cost efficiency of banks and analysis of the impact of bad loans on technical efficiency. The applicability of the algorithm is observed via a case study. Bad loans or Non-performing loans (NPLs) include past due loans; bankrupt and quasi-bankrupt assets, and doubtful assets are suspected to have a negative effect on technical efficiency in banking industry. The technical efficiency of banks was measured with reference to three input indicators: Number of Branches, Deposits, and Costs; and three output indicators: Income, Profit, and Loans. Then technical efficiency was regressed on two explanatory variables of loans and bad loans. Results of the case study in governmental banks in Iran show that bad loans have a negative impact on the technical efficiency of banks. For future research, other forecasting and performance assessment methods in certain and fuzzy environment to assess the impact of bad loans on technical efficiency of banks can be applied. These methods are adaptive network-based fuzzy inference system (ANFIS) [18, 19], artificial neural networks (ANN), etc. The results of these methods can be also compared with the results of current study.

7 References

[1] Berger, A. N.; Humphrey, D. B. Efficiency of financial institutions: international survey and directions for future research. // European Journal of Operational Research. 98, 2(1997), pp. 175-212.

[2] Koutsomanoli-Filippaki, A.; Margaritis, D.; Staikouras, C. Profit efficiency in the European Union banking industry: A directional technology distance function approach. // Journal of Productivity Analysis. 37, 3(2012), pp. 277-293.

[3] Azadeh, A.; Ghaderi, S. F.; Mirjalili, M.; Moghaddam, M. Integration of analytic hierarchy process and data envelopment analysis for assessment and optimization of

personnel productivity in a large industrial bank. // Expert Systems with Applications. 38 5(2011), pp. 5212-5225.

[4] Bhandari, A. K. Total factor productivity growth and its decomposition: The Indian banking sector during liberalisation. // Economic and Political Weekly. 47, 1(2012), pp. 68-76.

[5] Elyasiani, E.; Wang, Y. Bank holding company diversification and production efficiency. // Applied Financial Economics. 22, 17(2012), pp. 1409-1428.

[6] Sufian, F.; Habibullah, M. S. Developments in the efficiency of the Malaysian banking sector: The impacts of financial disruptions and exchange rate regimes. // Progress in Development Studies. 12, 1(2012), pp. 19-46.

[7] Perera, S.; Skully, M. On the cross-methodological validation of bank efficiency assessments. // Studies in Economics and Finance. 29, 1(2012), pp. 26-42.

[8] Azadeh, A.; Ghaderi, S. F.; Mirjalili, M.; Moghaddam, M. A DEA approach for ranking and optimisation of technical and management efficiency of a large bank based on financial indicators. // International Journal of Operational Research. 9, 2(2010), pp. 160-187.

[9] Tat, Y. R. Levels of satisfaction among Asian and Western travelers. // International Journal of Quality Reliability Management. 17, 2(2000), pp. 116-131.

[10] Rossi, F.; Thomas, A. A. Analysis of the beverage data Using Cluster Analysis Rotated Principal Components Analysis and LOESS curves. // Food Quality and Preference. 12, 5-7(2001), pp. 437-445.

[11] Zhongsheng, L.; Dong, L. Scale expanding and efficiency improvement of commercial banks of China. // Proceedings - International Symposium on Information Processing, ISIP 2008 and International Pacific Workshop on Web Mining and Web-Based Application, WMWA, 2008. pp. 705-709.

[12] Chen, T. Y. A study of cost efficiency and privatisation in Taiwan's banks: The impact of the Asian financial crisis. // Service Industries Journal. 24, 5(2004), pp. 137-151.

[13] Matthews, K.; Zhang, N. X. Bank productivity in China 1997-2007: Measurement and convergence. // China Economic Review. 21, 4(2010), pp. 617-628.

[14] Barros, C. P.; Managi, S.; Matousek, R. The technical efficiency of the Japanese banks: Non-radial directional performance measurement with undesirable output. // Omega. 40, 1(2012), pp. 1-8.

[15] Zhu, J. Data Envelopment Analysis vs. Principal Component Analysis: An illustrative study of economic performance of Chinese cities. // European Journal of Operational Research. 111, (1998), pp. 50-61.

[16] Premachandra, I. M. A note on DEA versus Principal Component Analysis, An Improvement to Joe Zhu Approach. // European Journal of Operational Research. 132, 3(2001), pp. 553-560.

[17] Slišković, D.; Grbić, R.; Hocenski, Ž. Multivariate statistical process monitoring. // Tehnicki vjesnik-Technical Gazette, 19, 1(2012), pp. 33-41.

[18] Hajialiakbari, F.; Gholami, M. H.; Roshandel, J.; Hatami-Shirkouhi, L. Assessment of the effect on technical efficiency of bad loans in banking industry: a principal component analysis and neuro-fuzzy system. // Neural Computing and Applications, (2013). pp. 1-8. 10.1007/s00521-013-1413-z.

[19] Nazari-Shirkouhi, S.; Keramati, A.; Rezaie, K. Improvement of customers' satisfaction with new product design using an adaptive neuro-fuzzy inference systems approach. // Neural Computing and Applications, (2013), pp. 1-11. 10.1007/s00521-013-1431-x.

Authors' address

Omid Mashhadifarahani
Department of Management
Shahre Qods Branch,
Islamic Azad University
Tehran, Iran

Narges Rezavi
Department of Industrial Engineering
Khatam Institute of Higher Education
Tehran, Iran

Loghman Hatami-Shirkouhi
Corresponding Author
Roudbar Branch
Islamic Azad University
Roudbar, Iran
E-mail: loghmanhatami@yahoo.com
Tel.: (+98)132 6224135