International Journal of Statistika and Mathematika, ISSN: 2277- 2790 E-ISSN: 2249-8605, Volume 8, Issue 2, 2013 pp 60-70

Evaluation of Efficiency and Benchmarking Commercial Banks in India: A Combined PCA and DEA Approach

R. Chandrasekaran[1a], R. Madhana Gopal[2b]

[1]Professor and Head (Retired), [2]Assistant Professor, Department of Statistics

Madras Christian College, Chennai 600 059, Tamil Nadu, INDIA.

Corresponding Addresses:

[a]drrcmcc@gmail.com, [b]madhan.stat@gmail.com

Research Article

Abstract: Data Envelopment Analysis (DEA) is a non-parametric linear programming technique used to determine the relative efficiency of Decision Making Units (DMU's). To evaluate efficiency, multiple input and output variables are used. There are exciting opinions available in the literature relating to the use of various input and output variables in assessing performance of DMU's. Even though there is no limit on the number of variables, the use of excessive number of variables will tend to reduce the discriminatory power within efficient as well as inefficient DMU's. In order to keep the number of input and output variables to a manageable level, it is possible to combine Principal Component Analysis (PCA) with DEA. In this paper, PCA is applied first to all inputs and outputs separately. With the intention to reduce the number of variables in the analysis, the principal components are chosen appropriately. Then PCA scores of selected principal components are treated as input and output variables for DEA and the performance of commercial banks in India are determined.

Keywords: Data envelopment analysis, Principal components analysis, Data reduction, Commercial Indian Banks, Efficiency Analysis

1. Introduction

Data envelopment analysis (DEA) is a non-parametric linear programming technique that evaluates the relative efficiency of decision making units (DMU). DEA is most valuable in multifaceted situations where there are multiple inputs and outputs, which cannot be easily analyzed with other techniques like ratio, indices, regression etc. Among various efficiency measurement tools, such as conventional statistical methods, non-parametric methods, and artificial intelligence methods, developed in the literature, it has been affirmed that DEA can effectively measure the relative efficiencies of multiple decision making units (DMUs) with similar goals and objectives [28]. It is simplest, popular and most efficient method than other frontier analysis such as stochastic frontier analysis (SFA), distribution free approach (DFA) and thick frontier analysis (TFA) etc. There are many literature based on DEA application in various fields, such as, *school and university* ([38], [26], [11], [1], [10]), *Port* ([39], [30], [44], [47], [9], [20], [32], [21], [37]), *hospitals* ([22], [14], [41], [25], [35], [40]), *Agriculture* ([6], [5], [12]). In this paper, efficiency of commercial banks in India is studied, since banks normally serve as a main channel for financial intermediation [17]. Due to technology development now a day's all the business transactions are done through the banks world-wide. Banks play a major role in the payment system of the country that allows financial and real resources to flow freely to their highest-returns uses. The productivity of economy is linked directly to efficiency of banks [34]. A small slowdown of the bank may affect not only its customers or share holders but also the economy of the countries. Therefore evaluation of technical efficiency of the banks and monitoring their financial positions are of critical importance to government bodies, investors, bank managers as the efficiency scores are informative signals of management quality ([8], [48]). Selection of input and output variables to be used in an assessment of comparative performance is the most important stage. In order to examine relative efficiency of a set of units it is necessary to define a production function which captures the key points of the production process [18]. Also appropriate number of input and output variables be used because too many variables tend to shift the units towards the efficiency frontier, resulting in a large number of units with high efficiency scores ([24]; [29]). [19] provides two rules for the selection of sample size; i) $\mathbf{n} \geq \max(\mathbf{r} * \mathbf{m})$ where $\mathbf{n}$ is the sample size, $\mathbf{r}$ is the number of inputs and $\mathbf{m}$ is the number of outputs. This states that sample size should be greater than or equal to product of inputs and outputs; ii) $\mathbf{n} \geq 3(\mathbf{r} + \mathbf{m})$, states that the number of observation in the data should be at least three times the sum of the inputs and outputs. This can be achieved by principal component analysis (PCA), which is able to reduce the data to a few principal components whilst minimizing the loss of information. The idea of integrating PCA with

DEA was first proposed by [46] and [2], [3] independently. Only few research papers are found in the literatures that are based on PCA-DEA, such as, [36], [27], [4], [15]. Applying PCA separately for input and output of DEA variables and then selected principal components (PC's) are used as new inputs and outputs of DEA are proposed by [42]. Similar approach is followed in this study to improve the discriminate power of DEA and to find efficiency of commercial banks in India. Rest of this paper is organized as follows. A brief introduction of techniques used in this study is provided in Sections 2 and 3. Section 4 briefly describes the methodology of the present study. Section 5 deals with source of data and variables selection. Section 6 presents the result and discussion, followed by conclusion in section 7.

2. Principal component analysis (PCA)

The technique of PCA was first described by Person (1901) and is one of the simplest techniques of the multivariate methods. The main objectives of PCA are:

- Identify new meaningful underlying (latent) variables;
- Discover or to reduce the dimensionality of the data set.

PCA involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components. The objective of the analysis is to take k variables X_1, X_2, X_3, …, X_k and find combinations of these to produces indices Z_1, Z_2, Z_3, … , Z_k that are uncorrelated. The lack of correlation is most important and is a useful property because the uncorrelated variables are measuring different dimensions in the data. These uncorrelated variables are ordered based on their variation i.e., $\mathrm{var}(Z_1) \geq \mathrm{var}(Z_2) \geq \mathrm{var}(Z_3) \geq \ldots \geq \mathrm{var}(Z_k)$. Therefore first principal component (PC_1) accounts for as much of the variability in the data as possible, and each succeeding component (PC_2, PC_3, …, PC_k) accounts for as much of the remaining variability as possible (Manly, 1986).

2.1. Steps involved in construction of PCA

Let us consider k variables say X_1, X_2, X_3, …, X_k for the study

- First normalize the data.
- Calculate the correlation matrix C.
  *If data set is not normalized then calculate the covariance matrix C.
- Find the eigen values λ_1 , λ_2 , λ_3,… , λ_k and the corresponding eigen vectors a_1, a_2 , a_3, … , a_k. The coefficients of the i^{th} principal component are the given by a_i while λ_i is its variance.
- Discard any components that only account for a small proportion of the variation in the data.

3. Data Envelopment Analysis (DEA)

DEA was originally introduced by [23] and improved by [16], also referred as frontier analysis, is a non-parametric, special linear programming model for deriving the comparative efficiency of single or multiple-input and single or multiple-output Decision-Making Units (DMU's). DEA does not require any underlying assumption of a functional form relating to inputs and outputs but always assume to have non-negative empirical data values.

3.1 Charnes, Cooper and Rhodes (CCR) Model (CRS):

The name of the model was given against the name of the authors. The measure of efficiency for each DMU is obtained as maximum of a ratio of weighted outputs to weighted inputs.

$Eff = \frac{\sum_K u_k\, y_{kj}}{\sum_i v_i\, x_{ij}}$, where u and v are weights and

x_{ij} = the observed amount of input of the i^{th} type of the j^{th} DMU.

y_{kj} = the observed amount of output of the k^{th} type for the j^{th} DMU.

The efficiency ranges from 0 to 1. The weights of all the DMU's are uniform rather arbitrary. The main idea of the DEA is that for each DMU's to set its own weights. The optimization problem is, maximize the efficiency of DMU subject to the condition that all efficiency of other DMU's remain less than or equal to 1. That is,

$$\text{Max}\,\lambda_0 = \frac{\sum_k u_k\, y_{ko}}{\sum_i v_i\, x_{io}}$$

Where $k = 1, 2, \ldots, S$ and $i = 1,2, \ldots, m$

Subject to

$$\frac{\sum_k u_k\, y_{kj}}{\sum_i v_i\, x_{ij}} \leq 1 \ \ \forall\, j\ , j = 1,2, \ldots, j_o, \ldots, n$$

$$u_k, v_i \geq 0 \ \ k = 1, 2, \ldots, S \ and \ i = 1,2, \ldots, m$$

If $u^* \ and\ v^*$ are optimal then for positive scalar c ($cu^* \ and \ cv^*$) are optimal therefore problem has infinite solutions. Therefore a simple work around is to fix the denominator to a constant value say (1) which can be interpreted as setting a constraint on the weights v_i . This result in

$$Max\, Z_o = \sum_k u_k\, y_{ko}$$

Subject to

$$\sum_i v_i\, x_{i,j_0} = 1$$

$$\sum_k u_k\, y_{k,j} \leq \sum_i v_i\, x_{i,j} \ \ \forall j, j = 1,2, \ldots, j_o, \ldots, n$$

$$u_k, v_i \geq 0 \ \ k = 1, 2, \ldots, S \ and \ i = 1,2, \ldots, m$$

x and y are data , u and v are decision variables.

The dual of the above model is,

$$Min\, P_o = \Theta_o$$

$$\sum_j \lambda_j\, y_{k,j} \geq y_{ko} \qquad k = 1, 2, \ldots, S$$

$$\Theta_o x_{io} \geq \sum_j \lambda_j\, x_{i,j} \qquad i = 1,2, \ldots, m$$

$$\lambda_j \geq 0$$

Both the linear programming problems yield the optimum solution Θ^*, which is the efficient score for the particular DMU_o and repeating them for each DMU_j, j = 1, 2, … , n, the efficient scores for all DMUs are obtained.

3.2 Banker, Charnes and Cooper (BCC) Model (VRS):

Another basic model is BCC, wherein some variants set a lower bound on u_k and v_i to prevent zero weights, that is; $u_k,\ v_i \geq \varepsilon$. The model was based on the dual, and adds a restriction on the λ_j.

$$Min\, P_o = \Theta_o$$
$$\sum_j \lambda_j\, y_{k,j} \geq y_{ko}$$
$$\Theta_o x_{io} \geq \sum_j \lambda_j\, x_{i,j}$$
$$\sum_j \lambda_j = 1$$
$$\lambda_j \geq 0\ k = 1, 2, \dots, S;\ i = 1,2, \dots, m\ and\ j = 1,2, \dots, n$$

This transforms the model from being "constant returns-to-scale" to "variable returns-to-scale". The scores from this model are sometimes called "pure technical efficiency scores" as they eliminate scale-efficiency from the analysis [13].

4. Methodology of Research

In order to keep the number of input and output variables to a manageable level PCA is applied first to all inputs and outputs separately. With the intention to reduce the number of variables in the analysis, the principal components are chosen appropriately. Then PCA scores of selected principal components are treated as input and output variables for DEA and the efficient commercial banks in India are identified. The proposed PCA-DEA approach is illustrated below in figure 4.1.

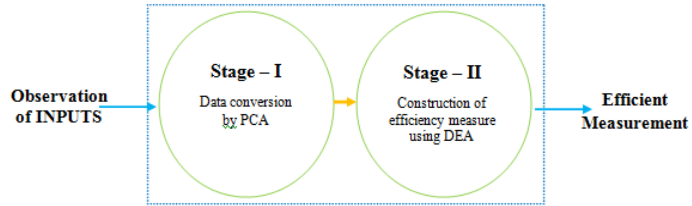

***Figure 4.1*: PCA-DEA Approaches**
Source: ling-Jing Kao et.al (2010)

5. Data and Sources

The present study deals with the secondary data of the year 2012 published in web pages of Reserved Bank of India (RBI) and Indian Banks' Association (IBA). According to [43], in banking theory, there are two approaches for selection of input and output variables for DEA, *viz.,* the production approach and intermediation approach. Production approach is more suitable for the analysis of bank branch efficiency and at the same time between banks, intermediation approach is most suitable. Therefore, in this paper, we have used intermediation approach for identification of inputs and outputs. Variables selection is very crucial in DEA. After carefully studying the efficiency analysis of banks through DEA, the following input and output variables are selected. Banks (DMUs) for this study is determined based on the following criteria i) Banks should be active in the Indian business market for a minimum period of five years (2008 – 2012) ii) Every selected bank should have more than 3 branches and 100 employees and iii) Banks should not be continuously in loss for 2 years. Based on the above conditions 55 commercial banks from public, private and foreign are selected for the analysis. Table 5.1 lists the descriptive Statistics for input and output variables and their codes.

Table 5.1: Descriptive Statistics for Inputs and Outputs of Banks in India

	Variables	Codes	Min	Max	Average	SD
Inputs	Loanable Funds	LOF	348563.50	117065293.00	13149975.00	18149858.00
	Fixed Asset	FIA	969.78	546654.92	102356.80	122018.50
	Number of branches	NOB	5.00	14316.00	1512.20	2192.70
	Number of Employees	NOE	195.00	215481.00	18370.31	30725.65
	Interest Expenses	INE	20108.28	6323036.87	778283.00	1005369.00
	Operating Expenses	OPE	7226.82	2606899.21	244967.80	379509.00
Outputs	Loans	LOA	191518.60	86757889.01	9163343.00	13206250.00
	Other Income	OTI	2236.92	1435144.57	152964.30	231619.50
	Interest Earned	INE	34187.67	10652145.34	1180420.00	1611877.00
	Net Interest Income	NII	14079.39	4329108.47	402136.70	625850.50
	Investment	INV	11330.70	3121976.10	397616.70	506353.20
	Net profit	NEP	2590.13	1170728.86	146547.00	200804.80

Source: Author's Calculation

To ensure the validity of the DEA model specification, an isotonicity test is carried out. An isotonicity test involves the calculation of all inter-correlations between inputs and outputs for identifying whether increasing amounts of inputs lead to greater outputs. Only variables with positive and statistically significant inter-correlations between inputs and outputs meeting the requirements of DEA isotonicity principle are selected to ensure that greater quantities of the selected inputs will not cause decreasing output ([7]; [45], [33]). Table 5.2 and Figure

5.1 show the Pearson correlation coefficients and it is observed that all the correlation coefficient values between input and output variables are positive and significant at the 0.01 level (2 tailed). Therefore, the present input and output variables passed the test and included for analysis.

Table 5.2: Pearson Correlation Coefficients for inputs and outputs of DEA variables

	LOF	FIA	NOB	NOE	IEX	OEX	LAA	OTI	INE	NII	INV	NEP
LOF	1.000											
FIA	0.876	1.000										
NOB	0.958	0.786	1.000									
NOE	0.953	0.755	0.958	1.000								
IEX	0.993	0.887	0.953	0.938	1.000							
OEX	0.963	0.826	0.919	0.974	0.942	1.000						
LAA	0.999	0.861	0.965	0.964	0.990	0.966	1.000					
OTI	0.933	0.872	0.837	0.896	0.911	0.969	0.928	1.000				
INE	0.997	0.870	0.960	0.966	0.993	0.973	0.997	0.937	1.000			
NII	0.973	0.815	0.940	0.981	0.950	0.994	0.977	0.950	0.981	1.000		
INV	0.980	0.909	0.907	0.916	0.976	0.961	0.971	0.964	0.981	0.957	1.000	
NEP	0.937	0.862	0.835	0.880	0.911	0.947	0.928	0.975	0.934	0.943	0.962	1.000

Source: Author's Calculation

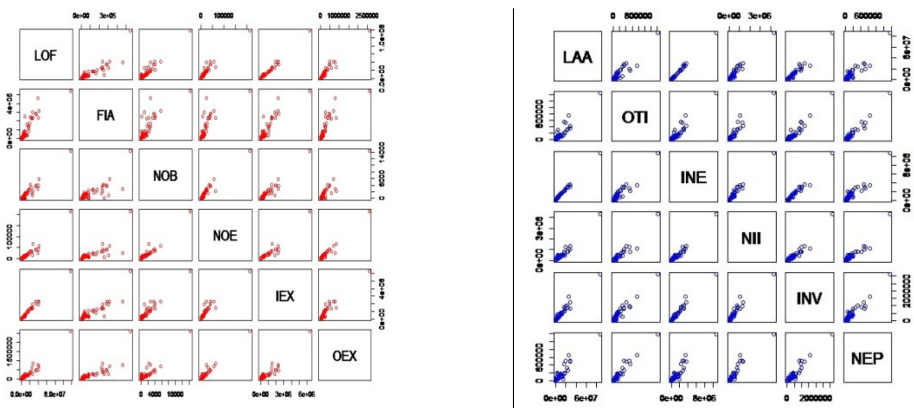

Figure 5.1**:** Scatter plot matrix of Input and Output variables

Figure 5.1 presents the scatter plot matrix of the results from all 55 banks for the Input and output variables. All the scatterplots in the diagram suggest that there is a strong positive correlation between each pair of variables.

6. Results and Discussion

6.1. Results of PCA

As mentioned in Section 1, one of the major drawbacks of DEA is a large number of input and output variables compared to the number of DMUs to be evaluated. The larger the number of inputs and outputs compared to the number of units to be evaluated, the greater the chances that the units will allocate appropriate weights to a single subset of inputs and outputs that will make them appear efficient [31]. To overcome this difficulty, first PCA is applied for input and output variables separately. R-2.15.3 software is used to perform PCA. Table 6.1 shows the percentage of variance of all the six principal components for inputs and output variables. For input data, PCA offered the following results. The first two components account for approximately 98% of the variance and for output variables the first component account for approximately 97% of the variance. Barplot of each component's variance of inputs and outputs shows the domination of first two components in input and first component in output (Figure 6.1). Based on 95% variance, two PC's (PC_{I1} and PC_{I2}) on input and one PC (PC_{O1}) on output explained most of variance in the original data. Therefore, scores of first two PC of input and first PC of output is used to estimate efficiency of banks using DEA.

Table 6.1.1 Percentage Variance and Loadings of Components of Input and Output variables

INPUT						
	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**
SD	2.3595	0.5460	0.2966	0.1814	0.1021	0.0567
% of Var	92.79	4.97	1.47	0.55	0.17	0.05
Cum % of Var	92.79	97.76	99.23	99.78	99.95	100.00
LOF	-0.4214	0.0097	-0.0893	0.4801	-0.3725	-0.6671

FIA	-0.3751	0.8433	0.0700	-0.3516	0.1163	-0.0781
NOB	-0.4094	-0.2807	-0.6252	-0.4960	-0.3061	0.1522
NOE	-0.4098	-0.4214	0.2422	-0.2353	0.6624	-0.3188
IEX	-0.4192	0.0804	-0.2705	0.5856	0.3523	0.5269
OEX	-0.4128	-0.1611	0.6815	-0.0399	-0.4372	0.3826
OUTPUT						
SD	2.4074	0.3581	0.2022	0.1601	0.0946	0.0306
% of Var	96.59	2.14	0.68	0.43	0.15	0.02
Cum % of Var	96.59	98.73	99.41	99.84	99.99	100.00
LAA	-0.4087	0.4644	-0.0823	0.1410	-0.5880	0.4948
OTI	-0.4053	-0.5246	0.1907	-0.6345	-0.3461	-0.0412
INE	-0.4108	0.4006	-0.1386	-0.0248	-0.0615	-0.8045
NII	-0.4092	0.2175	0.7213	0.0207	0.4905	0.1549
INV	-0.4110	-0.0218	-0.6448	-0.2241	0.5372	0.2759
NEP	-0.4044	-0.5485	-0.0417	0.7254	-0.0386	-0.0788

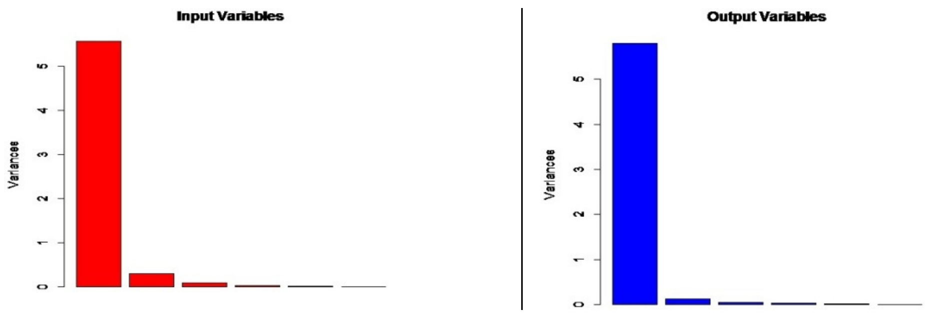

Figure 6.1.1 Barplot of Variance Explained by Components for Input and Output variables

6.2 Results of DEA and PCA-DEA

The present study used an input oriented DEA models with variable returns to scale. The efficiency scores of DEA and PCA-DEA for 55 commercial sector banks in India are summarized in Table 6.2.1 Model–A consists of original inputs and outputs, Partial PCA-DEA (ie., Original inputs and PC score of output, and PC scores of inputs and original output) are used in Model–B and Model-C. Complete PCA-DEA (ie., PC scores of input and output) are considered in Model-D.

Table 6.2.1 Efficiency Scores of Different model

Bank Codes	Model -A	Model –B (6 original input + 1PC output)	Model –C (2PC input + 6 original output)	Model -D
	DEA	**PARTIAL PCA-DEA**		**Complete PCA-DEA**
B01	1.0000	0.0047	1.0000	1.0000
B02	1.0000	0.0761	0.9646	1.0000
B03	1.0000	0.0518	0.9622	0.9836
B04	0.9730	0.0941	0.8541	0.9885
B05	0.9680	0.0614	0.9539	0.9929
B06	0.9850	0.0725	0.9493	0.9949
B07	1.0000	0.0301	0.9267	0.9731
B08	1.0000	0.0446	1.0000	1.0000
B09	1.0000	0.0149	0.9440	0.9537
B10	1.0000	0.0163	0.8557	0.9583
B11	0.9540	0.0544	0.9325	1.0000
B12	1.0000	0.0172	0.8290	0.9406
B13	0.9740	0.0211	0.8582	1.0000
B14	1.0000	0.0419	1.0000	0.9821
B15	0.9950	0.0644	0.9284	0.9868
B16	1.0000	0.0300	0.6543	0.9113
B17	1.0000	0.0392	0.8208	0.9399
B18	0.9650	0.0265	0.8724	0.9705
B19	1.0000	0.0357	0.8441	0.9463
B20	0.9690	0.0802	0.8481	0.9874

B21	1.0000	0.0126	0.9567	0.9835
B22	0.9960	0.0291	0.9271	0.9897
B23	1.0000	0.0357	0.9816	1.0000
B24	1.0000	0.0218	0.8381	0.9527
B25	0.9710	0.0530	0.9015	0.9865
B26	0.9640	0.0612	0.9237	0.9904
B27	0.9110	0.3119	0.9004	1.0000
B28	1.0000	0.2767	0.8962	0.9913
B29	0.8910	0.3048	0.8912	0.9989
B30	1.0000	0.0870	0.9183	0.9841
B31	0.9420	0.1279	0.8798	0.9845
B32	1.0000	0.1094	0.8877	0.9742
B33	0.9630	0.1494	0.9118	0.9944
B34	0.9750	0.1533	0.8934	0.9854
B35	0.9760	0.2888	0.8814	0.9967
B36	1.0000	1.0000	0.8887	1.0000
B37	1.0000	0.8328	0.8846	0.9976
B38	0.9840	0.1238	0.8837	0.9895
B39	1.0000	0.2747	0.8981	0.9894
B40	1.0000	0.0264	0.8228	0.8828
B41	0.9770	0.6511	0.8704	0.9995
B42	1.0000	0.0232	1.0000	1.0000
B43	1.0000	0.0178	0.6262	0.8126
B44	0.9800	0.1241	0.8545	0.9654
B45	1.0000	0.1210	0.8977	0.9647
B46	1.0000	0.1133	0.9061	0.9624
B47	1.0000	1.0000	0.8840	0.9737
B48	1.0000	1.0000	0.8836	0.9839
B49	1.0000	1.0000	0.8671	0.9932
B50	1.0000	0.1298	0.8323	0.9089
B51	1.0000	0.4167	0.8869	0.9761
B52	1.0000	0.6238	0.8853	0.9639
B53	1.0000	0.1514	0.8245	0.9044
B54	1.0000	0.4472	0.8888	0.9746
B55	1.0000	0.1067	0.6569	0.9086
No. Efficiency banks	35	4	4	9
% of Efficiency banks	64	7	7	16
Mean	0.9875	0.2015	0.8860	0.9722
SD	0.0226	0.2830	0.0750	0.0357
MAD	0.0165	0.2039	0.0486	0.0248
MD	≈ 0.0000	≈0.0000	≈0.0000	≈0.0000

≈ □□□□□□□□□□□□□□, □□□ – □□□□ □□□□□□□□ □□□□□□□□□□, □□ – □□□□ □□□□□□□□□ □□□ □□ – □□□□□□□□ □□□□□□□□□□

Source: Author's Calculation

From Table 6.2.1, the average measure of efficiency score for Model–A is estimated at 97% (0.9695), for Model– B, Model-C and Model-D efficiency scores are estimated at 20% (0.2015), 88% (0.8860) and 97% (0.9722) respectively. Sixty four percentages of banks are efficient in Model-A, seven percentages of banks is efficient in Model–B and Model-C respectively and in Model–D, sixteen percentages of banks are efficient. No banks are efficient in all the models. On seeing the standard deviation of all models, Model-B is high with 0.2830, next comes Model–C with 0.0750, Model–D with 0.357 and Model-A has the least 0.226. To estimate the performances of models the mean absolute deviation (MAD) and mean deviations (MD) are calculated. MD is to know the models that estimate over or under. If MD value is negative then on the average model overestimates. MAD results suggest that Model–B (0.2039) is the highest when compared to other models and on considering MD all models yield the same result zero ($MD_A \approx MA_B \approx MD_C \approx MD_D \approx 0$) suggesting that no model overestimates. ANOVA is performed to know whether any significant difference exists between the models. Since estimated efficiency scores of DEA models fails in normality test an equivalent non-parametric test Kruskal-Wallis Test is applied. From Table 6.2.2, Chi–square value is 141.60 and P-value is ≈ 0 which is < 0.05, indicating that models are not similar.

Table 6.2.2 Kruskal-Wallis test on efficiency scores of DEA models

Test Statistics	
	Models
Chi–Square	141.60
df	3
p- value	171E-30 (≈ 0)

Source: Author's Calculation

Figure 6.2.1 shows the distribution of efficiency score of four models. The distribution of efficient scores is very high in Model-B. Model–A and Model-B are distributed in same manner but the efficient banks differ hugely. Number of efficient banks is same in Model–B and Model-C.

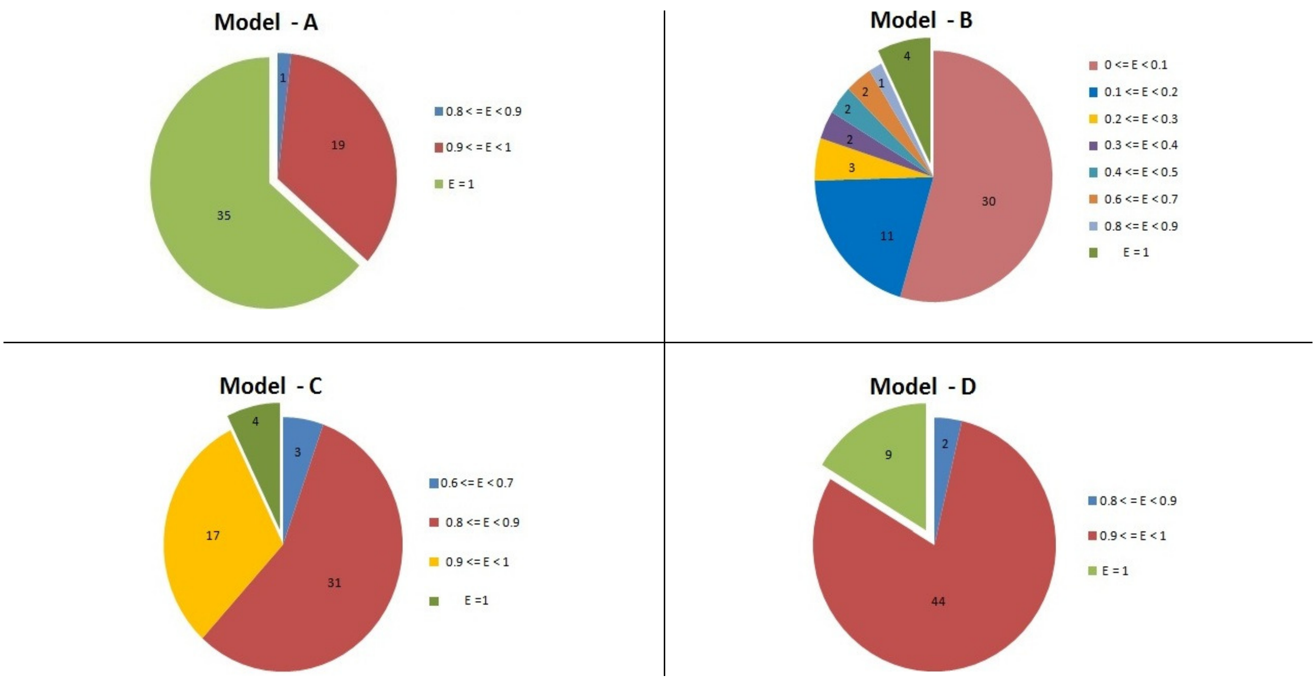

Figure 6.2.1 Range of Efficiency Scores of each Model

6.2.1 Comparison of models

Based on the Table 6.2.1 and Figure 6.21, when compared to other models,

- The average efficiency of Model–B is less.
- Least number of efficient banks.
- The highest standard deviation and mean absolute deviation, and
- The distribution of efficient score spread widely.

This indicates that the Model–B has more ability to distinguish between the performances of banks. Therefore benchmarking and raking of banks are done based on Model–B.

Table 6.2.3 Benchmarking and Ranking of banks based on Model–B

Bank Code	Reference Set			weights			Peer Count	Rank
B01	B49	B48	B36	0.100	0.268	0.632	0	55
B02	B49	B36	B48	0.023	0.752	0.224	0	29
B03	B49	B36	B48	0.056	0.657	0.287	0	36
B04	B49	B48	B36	0.085	0.216	0.699	0	26
B05	B49	B36	B48	0.027	0.644	0.329	0	32
B06	B49	B36	B48	0.043	0.608	0.349	0	30
B07	B49	B36	B48	0.026	0.722	0.252	0	42
B08	B49	B36	B48	0.011	0.726	0.264	0	37
B09	B49	B36	B48	0.007	0.556	0.437	0	53
B10	B49	B36	B48	0.022	0.617	0.362	0	52
B11	B49	B36	B48	0.058	0.835	0.107	0	34

B12	B49	B36	B48	0.021	0.586	0.394	0	51
B13	B49	B36	B48	0.021	0.838	0.141	0	49
B14	B36	B48		0.571	0.429		0	38
B15	B49	B36	B48	0.002	0.704	0.294	0	31
B16	B49	B36	B48	0.005	0.246	0.749	0	43
B17	B49	B36	B48	0.043	0.745	0.213	0	39
B18	B49	B36	B48	0.035	0.672	0.293	0	45
B19	B49	B36	B48	0.029	0.608	0.363	0	40
B20	B49	B36	B48	0.067	0.67	0.263	0	28
B21	B49	B36	B48	0.051	0.686	0.263	0	54
B22	B49	B36	B48	0.029	0.752	0.219	0	44
B23	B36	B48		0.818	0.182		0	41
B24	B49	B36	B48	0.046	0.688	0.266	0	48
B25	B36	B48		0.814	0.186		0	35
B26	B36	B48		0.778	0.222		0	33
B27	B49	B36		0.000	1.000		0	10
B28	B49	B36	B48	0.014	0.804	0.182	0	13
B29	B49	B36		0.142	0.858		0	11
B30	B49	B36	B48	0.042	0.796	0.162	0	27
B31	B49	B36	B48	0.242	0.645	0.113	0	19
B32	B49	B36	B48	0.046	0.566	0.387	0	24
B33	B49	B48	B36	0.04	0.209	0.751	0	17
B34	B49	B48	B36	0.032	0.271	0.697	0	15
B35	B49	B36	B48	0.042	0.861	0.097	0	12
B36	B36			1.000			50	01
B37	B36	B48		0.824	0.176		0	05
B38	B49	B36	B48	0.011	0.81	0.18	0	21
B39	B49	B48	B36	0.065	0.178	0.756	0	14
B40	B49	B36	B48	0.295	0.383	0.322	0	46
B41	B49	B36		0.306	0.694		0	06
B42	B49	B36	B48	0.308	0.453	0.239	0	47
B43	B49	B36	B48	0.228	0.338	0.433	0	50
B44	B49	B36	B48	0.325	0.438	0.236	0	20
B45	B49	B48	B36	0.310	0.306	0.385	0	22
B46	B49	B48	B36	0.066	0.577	0.358	0	23
B47	B48			1.000			(0)	04
B48	B48			1.000			49	02
B49	B49			1.000			42	03
B50	B48	B36		0.994	0.006		0	18
B51	B48			1.000			0	09
B52	B48	B36		0.956	0.044		0	07
B53	B36	B48		0.026	0.974		0	16
B54	B49	B36	B48	0.260	0.080	0.660	0	08
B55	B49	B36	B48	0.766	0.021	0.213	0	25

(0) – efficient but no peer to inefficient banks

Source: Author's Calculation

6.2.2. Efficiency analysis of banks based on Model-B

In DEA, a banks is said to be efficient if the efficient score is equal to 1 and there exits one optimal solution with weights of the inputs and outputs greater than zero ($\square^*$ and $\square^*$) [19]. Average efficiency scores of Model–B is 0.2015. This indicates that banks have potential for saving by 80% (0.7985). In other words, commercial sectors banks in India uses only 20.15% of the resources to produce the given output. Inefficient banks can improve if they utilize these unused resources properly ($\approx 80\%$). Four banks namely, B36, B47, B48 and B49 are efficient and rest fifty one banks are inefficient.

6.2.2. Benchmarking and Ranking of banks based on Model–B

A set of corresponding efficient banks act as a reference banks or peers for inefficient banks. For improvement, inefficient banks can follow their reference banks. Benchmarking for inefficient banks are done based on Model–B. Four banks are efficient in Model–B; these banks act as a peer for fifty one inefficient banks. Table 6.2.3 shows the reference set and corresponding weights of the inefficient banks. For example, B01 is an inefficient bank it has three reference banks namely, B49, B48 and B36 with 0.100, 0.268 and 0.632 weights respectively. B01 can follow any of these three banks for improving. Similarly, other inefficient banks have their own reference banks. Ranking of banks are done based on efficient scores for inefficient banks and for efficient banks based on their peer count that is, first rank is given for efficient banks which act as a peer for maximum number of inefficient banks. B36 is peer for 50 inefficient banks, it was given first rank, B48 second rank with 49 peers, B49 stands third with 42 peers, B47 is an efficient bank but no peer for any inefficient banks in the fourth rank. B01 stands last (rank 55) with the least efficient score of 0.0047 and it was preceded by B21 with efficient score 0.0126.

7. Conclusion

The present study aimed at analyzing the efficiency of commercial banks in India for the period 2012. For such purposes, broadly chosen non-parametric optimizing technique is DEA, since it is easy to calculate and the method does not need pre-defined function like other parametric frontier techniques. One of the limitations of DEA is the selection of number of variables; when the number of variables is increased then discrimination power of efficient and inefficient DMUs decrease. In real life applications, it may not be possible to get less number of variables. To overcome this difficulty, an integrated PCA with DEA is used in the present study. Efficiency of banks is found using variable – return – scale (VRS) input oriented model by both traditional DEA and PCA-DEA. On comparing 4 models, Model-B has low efficient score, least number of efficient banks, highest SD and MAD value and distribution of efficiency score spread widely from 0 to 1. The key findings of this study are

- When variables are large PCA-DEA is preferable than traditional DEA.
- Based on PCA – DEA (Model–B) it is found that commercial banks in India use only 20% of the resources to produce the given output. Inefficient banks can improve if they utilize these unused ($\approx$ *80%*) resources properly.
- Bank B36 is peer for 50 inefficient banks; it was given first rank and followed by bank B48 second rank with 49 peers.
- Bank B01 stands last (rank 55) with the least efficient score of 0.0047 and it was preceded by B21 with efficient score 0.0126.

References

1. Abbasi, F., Hajioseini, H and Mirizadeh. M, "Resource allocation with DEA ranking: evidence from Iran medical universities, performance management and measurement with data envelopment analysis", Proceedings of the 8th international conference of DEA, Olayan school of Business, American University of Beirut, Lebanon, 2010.
2. Adler, N and Golany, B, "Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to western Europe", European journal of operational research, vol.132, pp. 260–273, 2001.
3. Adler, N and Golany, B, "Including principal component weights to improve discrimination in data envelopment analysis, Journal of operational research society, vol.53, pp.985–991, 2002.
4. Ahmadvand. A., Abtahy. Z and Bashiri. M, "Considering undesirable variables in PCA-DEA method: A case of road safety evaluation in Iran", Journal of industrial engineering, vol.7, pp. 43-50, 2001.
5. Ali-Hussain, M and Saed, A, "Econometric estimation of agricultural production function in Jordan: 1981-1996", Journal of economic and administrative sciences, vol.17, pp.90-103, 2001.
6. Arnade, C, "Using a programming approach to measure international agricultural efficiency and productivity", Journal of agricultural economics, vol.49 (1), pp.67-84, 1998.
7. Avkiran, N. K, "Productivity analysis in the services sector with data envelopment analysis", Technical report, University of Queensland, Australia, 1999a.
8. Barr, R. S., Seiford, L. M and Siems, T. F, "Forecasting bank failure: A nonparametric frontier estimation approach", Recherches economiques de Louvain, vol. 60, pp. 417-429, 1994.
9. Barros, C. P, "Measurement of efficiency of Portuguese seaport authorities with DEA", International journal of transport economics, vol. 30 (3), pp. 335–354, 2003.
10. Baysal, M. E, "Performance evaluation in fuzzy decision making environment: an efficiency analysis in higher education system in Turkey, Dissertation, Gazi University institute of science and technology, pp. 87-92 , 2005.
11. Beasley, J. E, "Determining teaching and research efficiencies", Journal of the operational research society, vol. 46 (4), pp. 441–452, 1995.
12. Belloumi, M and Matoussi, M, "Measuring agricultural productivity growth in MENA countries", Journal of development and agricultural economics, vol.1 (4), pp.103-113, 2009.
13. Bowlin, W. F. "Evaluating the efficiency of US air force real-property maintenance activities", Journal of Operation Research Society, Vol. **38** (2), pp. 127-35, 1987.
14. Buyukkayikci, H and Sahin I, "The comparative analysis of patient's costs with reimbursement in surgical services (in Turkish)", Hacettepe Saglik Idaresi Dergisi, vol. 5 (3), pp.119:138, 2000.
15. Chandraskeran, R. and Madhanagopal, R, "Investigating efficiency and benchmarking public sector commercial banks in India by PCA-DEA approach", Proceedings of national conference on recent advances in Mathematical analysis and applications, K.S.R college of Arts and Science, Tiruchengode, Tamil Nadu, India, pp.75-85, 2013.

16. Charnes, A., Cooper, W. W. and Rhodes, E. “Measuring the Efficiency of Decision Making Units”, European Journal of Operational Research, Vol.2, no.6, pp.429-444, 1978.
17. Chen, Y., “Banking Panics: The Role of the First-Come, First-Served Rule and Information Externalities”, Journal of Political Economy1999, vol. 107 (5), pp. 946-968, 1999.
18. Coli, M., Nissi, E, and Rapposelli, A, “Efficiency evaluation in an airline company: some empirical results”, Journal of applied science, vol.11, pp. 737–742, 2011.
19. Cooper, W. W., Seiford, L. M. and Tone, K, “Data envelopment analysis, a comprehensive text with models, applications, references and DEA-solver software”, second edition, USA: springer, 2007.
20. Cullinane, K., Song, D. W., Ji, P and Wang, T.F, “An application of DEA windows analysis to container port production efficiency”, Review of network economics, vol.3 (2), pp. 184-206, 2004.
21. Cullinane, K., Wang, T. F, “The efficiency of European container ports: across-sectional data envelopment analysis”, International journal of logistics: research and applications, vol.9 (1), pp 19-31, 2006.
22. Ersoy, K., Kavuncubasi, S., Ozcan, Y and Harris, J “Technical efficiencies of Turkish hospitals: DEA approach”, Journal of medical systems, vol. 21 (2), pp.67-74, 1997.
23. Farrell, M. J. “The measurement of productive efficiency”, Journal of Royal Statistical Society. Serial A, Vol. 120, pp.253–281, 1957.
24. Golany, B and Roll, Y, “An application procedure for data envelopment analysis”, Management science, vol.17, pp. 237–250, 1989.
25. Jacobs, R., Smith P. C., and Street, A, “Measuring efficiency in health care”, Cambridge University Press, New York, 2006.
26. Johnes, J. and Johnes, G. (1995) Research funding and performance in UK university departments of economics-a frontier analysis, Economics of Education Review 14 (3):301–314.
27. Khodabakhshi, M, “New input and output orientation models in analysis and application”. Ph.D Thesis, 2003.
28. Ling-Jing Kao., Chi-Jie Lu and Chih-Chou Chiu, “Efficiency measurement using independent component analysis and data envelopment analysis”, European journal of operation research, vol. 210 (2), pp. 310-317, 2011.
29. Lin, E.T, (2008), “Route-based performance evaluation of Taiwanese domestic airlines using data envelopment analysis: a comment”, Transportation research, vol. 44, pp.894-899, 2008.
30. Martinez-Budria, E., Diaz-Armas, R., Navarro-Ibanez, M and Ravelo-Mesa, T, “A study of the efficiency of Spanish port authorities using data envelopment analysis”, International journal of transport economics, vol. 26 (2), pp.237–253, 1999.
31. Martic, M. M, Novakovic, M. S and Baggia, A, “Data envelopment analysis – basic models and their utilization”, Organizacija, vol.2 (2), pp.37- 43, 2009.
32. Min, H., Park, B.I, “Evaluating the inter-temporal efficiency trends of international container terminals using data envelopment analysis”, International Journal of integrated supply management, vol. 1 (3), pp 258-277, 2005.
33. Mostafa, M. M, “Modeling the competitive market efficiency of Egyptian companies: A probabilistic neural network analysis”, Expert systems with applications, vol. 36, pp. 8839 - 8848, 2009.
34. Neetubala and SunilKumar, “How efficient are public sector banks in India? an empirical investigation”, Journal of banking and financial services and insurance research, vol.1 (3), pp. 39-62, 2011.
35. Ozcan, Y, “Health care benchmarking and performance evaluation: an assessment using data envelopment analysis”, Springer publishing, New York, 2008.
36. Permachandra, I. M, “A note on DEA vs. principle component analysis: An improvement to Joe Zhu’s approach”, European journal of operation research, vol. 132, pp. 553-560, 20011.
37. Pjevcevic, D., Radonjic, A., Hrle, Z and Colic, V, “DEA window analysis for measuring port efficiencies in Serbia”, Traffic management, vol. 42 (1), pp. 63 - 72, 2011.
38. Rhodes, E. L and Southwick, L, “Determinants of efficiency in public and private universities”, working paper, School of environmental and public affairs, Indiana University, Bloomington, 1986.
39. Roll, Y and Hayuth, Y, “Port performance comparison applying data envelopment analysis”, Maritime policy and management, vol. 20 (2), pp.153–161, 1993.
40. Sahin, I, “Total factor productivity analysis of Turkish social security hospitals that were transferred to the health ministry (in Turkish)”, Iktisat Isletme ve Finans, vol. 24, pp.9-40, 2009.
41. Sahin, I. and Y. Ozcan, “Public sector hospital efficiency for provincial markets (in Turkey)”, Journal of medical systems, vol. 24 (6), pp.307-320, 2000.
42. Shanmugam, R, Johnson, C. H, “At a crossroad of data envelopment and principal component analyses”. Omega, vol. 35, pp. 351-364, 2007.
43. Sealey, C and Lindley J. T, “Inputs, outputs and a theory of production and cost at depository financial institution”, Journal of finance, vol.32, pp.1251–1266, 1977.
44. Tongzon, J, “Efficiency measurement of selected Australian and other international ports using data envelopment analysis”, Transportation research part A, vol. 35 (2), pp. 107-122, 2001.
45. Tsai, H. C., Chen, C. M., and Tzeng, G. H, “The comparative productivity efficiency for global telecoms”, International journal of production economics, vol.103, pp.509–526, 2006.
46. Ueda, T and Hoshiai, Y., “Application of principal component analysis for parsimonious summarization of DEA inputs and/or outputs”, Journal of operational research society Japan, vol.40, pp. 446–478, 1997.
47. Valentine, V. F and Gray, R, “The measurement of port efficiency using data envelopment analysis”, in Proceedings of the 9th world conference on transport research, Seoul, July 2001.
48. Wheelock, D. C and Wilson, P, “Explaining bank failures: deposit insurance, regulation and efficiency”, Review of Economics and Statistics, vol. 77, pp. 689-700, 1995.

Authors Biographies

Dr. R. Chandrasekaran obtained his B.Sc., M.Sc., and Ph.D. in Statistics from the University of Madras, Chennai, India. He retired in May 2012 as Associate Professor and Head of the Department of Statistics, from Madras Christian College, Tambaram, Chennai, India, 600059. He has taught Statistics and Computer Applications at the postgraduate level and guided students for M.Phil. and Ph.D. degrees for the past 36 years. As a Statistician and subject expert, he continues to officiate as Resource person and Consultant. He has presented over 35 papers in National / International Conferences and published over 30 papers in the Journals/Proceedings of the Conferences. Also, he has guided a scholar for Ph.D. on Neural Networks and four more scholars are working for their Ph.D. under him. So far, he has guided 19 M.Phil. Scholars. He has officiated as Resource person for over 25 Refresher Courses in Statistics, Mathematics and Computer Science for college teachers/professors and also for Faculty Development Programmes for engineering and management faculties. He has organized 7 workshops/seminars and delivered over 25 invited talks in conferences/workshops. He has chaired several sessions in National and International conferences. He is Life member of several associations and societies. He has been using *Statistical Tools and Packages* in many research areas of applications. He has specialized in the areas of designing **Sample Surveys** and **Market Research** over past three decades. He is also sharing his expertise by officiating as a Consultant **/ Subject Expert / Sample Survey Expert** for various educational ***institutions/agencies/companies***. His areas of Interest include Applied Multivariate Analysis, Sample Surveys, Pattern Recognition, Theoretical Computer Science, Artificial Neural Networks, Financial Statistics, and Biostatistics.

R. Madhana Gopal received B.Sc (Statistics), M.Sc (Statistics), M.Phil (Statistics) and M.B.A degree from Bharathiyar University, Coimbatore, Tamil Nadu, India in 2002, 2004, 2006 and 2010. Currently, working as an Assistant professor and since 2010, Ph.D research scholar in Madras Christian College, Chennai, Tamil Nadu, India. Earlier worked as a guest lecturer in government Arts College, Coimbatore. Published many research papers in various international journals. His research interests include Multivariate Analysis, Data Mining, Fuzzy Logic and Frontier Analysis.