An integrated data envelopment analysis-artificial neural network approach for benchmarking of bank branches

Abstract

Efficiency and quality of services in banking industry is crucial in today's economy and society. The competition in this section has become increasingly intense, with the improvement in technology. Therefore, performance analyses in the banking sector attract more attention. Data envelopment analysis (DEA) is a pioneer approach in literature for efficiency measurement and benchmarks.

DEA uses past data for calculating benchmarks, hence forecasting the benchmarks for future is the main weakness of DEA method.

This paper integrates DEA and artificial neural networks to calculate the relative efficiency and benchmarks of one of Iranian commercial bank branches. Therefore, each branch could have strategy to improve the efficiency and eliminate the cause of inefficiencies due to, the five years forecast.

Keywords: Data envelopment analysis; Artificial neural network; Benchmarking

1. Introduction

Since banking industry is highly competitive, the performance assessment has received more attention recently. The banking sector is in a race to see who can offer the best services. These include intensified competition in the market place. Therefore, bank management wants to identify and eliminate the underlying causes of inefficiencies to help the firms improving efficiency. In literature data envelopment analysis (DEA) is a leading approach for the performance analysis and finding benchmarks. Various models of DEA are widely used for evaluating bank efficiency such as, Sherman and Gold (1985), Soteriou and Zenios (1999), Golany and Storbeck (1999), Athanassopoulos and Giokas (2000), and, etc. thick frontier approach (TFA) as in Berger and Humphrey (1991), Clark (1996) and Deyoung (1998), free disposal hull (FDH) as in Tulkens (1993) and Chang (1999),stochastic frontier approach (SFA), also called econometric frontier approach (EFA) as in Kaparakis, et al. (1994), Berger and Humphrey (1997), Hao, et al.(2001), and distribution free approach (DFA) as in Berger, et al. (1993), Akhavein, et al. (1997) and Deyoung (1997).

As DEA can hardly predict the performance of other decision-making units, Wang (2003) used artificial neural network (ANN) to assist in estimating efficiency.

Athnassopoulos and Curram (1996) firstly introduced the combination of neural networks and DEA for classification and/or prediction. They used DEA in bank with multi-output: 4 inputs, 3 outputs to monitor training cases in a study. The comparison between DEA and ANN demonstrates that DEA is superior to ANN for measurement purpose. Azadeh, et al. (2006), Azadeh, et al. (2007) and Azadeh, et al. (2007) utilized a highly flexible ANN algorithm to measure and rank the performance of decision-making units (DMUs). They defined an application of algorithm in efficiency calculation of Iran steam power plants in 2004. Results demonstrate that the proposed algorithm estimates the values of efficiency closer to the ideal efficiency. Finally, they displayed that the result of proposed algorithm is more robust than the conventional approach because better performance patterns are explored. Furthermore, they proposed a method to integrate their pervious algorithm (Azadeh, et al. 2007, 2008).
Wu, et al. (2006) combined DEA and ANN for measuring performance of a large Canadian bank. They displayed that the DEA-ANN method produces a more robust frontier and helps to identify more efficient units. Furthermore, for inefficient units, it provides the guidance on how to improve their performance to different efficiency ratings. Finally, they concluded there is no need for assumptions about the production function in ANN approach (the major drawback of the parametric approach) and it is highly flexible. Weakness of DEA in forecasting is the reason to use ANN (Wu, et al. 2006). The classic DEA methods did not have ability to demonstrate benchmarks for future. ANN has been viewed as a useful tool for managers in predicting system. There are few researches address the combination of DEA and ANN, Bose and Patel. 2015 overcame the regression model shortcoming of a single dependent variable by using neural network, they described a combinatorial method for a data generation procedure and training and prediction to overcome this data-based shortcoming

we describe a procedure that uses two-stage DEA, a data generation procedure (DGP) and ANN training and prediction to overcome this data-based shortcoming

such as He-BoongKwon, 2016 who used, CCR model and ANN for measuring performance and predicting model of Class I railroads in the United States. In another work, Doaei, et al. 2017 predicted Malaysian manufacturing firm's efficiency by applying a combined method of Data Envelopment Analysis (DEA) and ANN.

This paper integrates DEA and neural networks to forecast inputs and outputs for examining efficiency. Therefore, benchmarks are based on the future data and inefficient MDUs have better performance patterns to improve their efficiencies.

The paper is organized as follows. Section 2 briefly reviews neural network and DEA. Section 3 demonstrates the models and methodology utilized in this paper. The DEA results and further discussion is given in Section 4. Finally, our conclusions and future work are offered in Section 5.

2. Problem definition

2.1. Data envelopment analysis

DEA is a non-parametric method, which uses linear programming to calculate the efficiency in a given set of decision-making units (DMUs).

The DMUs that make up a frontier envelope scored 1. The less efficient firms and the relative efficiency of the firms are calculated in terms of scores on a scale of 0 to 1.

Envelopment surface that represents best practice, can give an indication of how inefficient DMU can improve to be efficient. DEA provides a comprehensive analysis of relative efficiencies for multiple input-multiple output situations by evaluating each DMU's performance relative to an envelopment surface composed of efficient DMUs. Units that lie on the surface are known efficient in DEA while those units that do not, are named inefficient. The efficient reference set is included DMUs which are the peer group for the inefficient units.

The projection of inefficient units on envelopment surface is called benchmark.

Benchmarks are the indication of how the inefficient DMU can improve to be efficient. Benchmark means if the evaluated DMU had this inputs and outputs could have been efficient.

Assume input and outputs for $j = 1,...,n$ DMUs (X_j, Y_j) where, $X_j = (x_{1j},...,x_{ij},...,x_{mj})$ is a vector of observed inputs and $Y_j = (y_{1j},...,y_{rj},...,y_{sj})$ is a vector of observed outputs for DMU_j.

The production possibility set is as bellow:

$$T \quad \{(X,Y) \mid Y \quad 0 \; \text{can be produced from} X \quad 0\}$$

The input possibility $L(Y)$, for each Y, and the output possibility $P(X)$, for each X, are defined as below:

$$L(Y) = \{X \,|(\mathrm{X},\mathrm{Y}) \in \mathrm{T}\}$$

$$P(X) = \{X \,|(\mathrm{X},\mathrm{Y}) \in \mathrm{T}\}$$

For achieving production possibility set, T the following proprieties were postulated:

1. Convexity:

$$If\,(X_j, Y_j) \in T, j = 1,...,n, and\ \lambda_j \geq 0\, are\, nonnegative\, scalars\, such\, that \sum_{j=1}^{n} \lambda_j = 1, then$$

$$\left(\sum_{j=1}^{n} \lambda_j X_j, \sum_{j=1}^{n} \lambda_j Y_j \right) \in T$$

2. Inefficiency postulate:

$$(a) If\,(X,Y) \in T\, and\, \overline{X} \geq X, then\,(\overline{X},Y) \in T$$
$$(b) If\,(X,Y) \in T\, and\, \overline{Y} \leq Y, then\,(X,\overline{Y}) \in T$$

3. Ray unboundedness:

$$If\,(X,Y) \in T\, then\,(KX,KY) \in T\ for\, any\, k > 0$$

4. Minimum extrapolation: T is the intersection set of T satisfying postulate 1,2, and 3 and subject to condition that each of observed vectors $(X_j, Y_j) \in T, j = 1,...,n$.

With mentioned assumptions Tv is as bellow:

$$Tv = \{\begin{pmatrix} X \\ Y \end{pmatrix} | \mathrm{X} \geq \sum_{j=1}^{n} \mathrm{X}_{\mathrm{ij}} \lambda_j\ \& Y \leq \sum_{j=1}^{n} Y_{ij} \lambda_j\ \& \sum_{j=1}^{n} \lambda_j\ \& \lambda \geq 0\}$$

Different models for calculating efficiency were introduced, the oldest model is BCC (Banker- Charnes and Cooper (1984)) model:

Input oriented BCC Model:

$$Minimze\ \ \theta - \varepsilon(1s^{+} + 1s^{-})$$
$$Subject\ \ to,$$

$$\sum_{j=1}^{n} x_{ij} \lambda_j + s_i^{-} = \theta\, x_{iq} \qquad i = 1,2,...,m$$

$$\sum_{j=1}^{n} y_{ij} \lambda_j - s_r^{+} = y_{rq} \qquad r = 1,2,...,s$$

$$\sum_{j=1}^{n} \lambda_j = 1 \qquad j = 1,...,n$$

$$\lambda_j, s_i^{-}, s_r^{+} \geq 0 \quad i = 1,2,...,m \quad r = 1,2,...,s \quad j = 1,...,n$$

A DMU is called efficient, if it has $\theta^* = 1, s_i^{-*} = 0, s_r^{+*} = 0$. Otherwise, it is called inefficient.

For inefficient DMU's (ex. DMU_q), the DEA model calculates the benchmark. The benchmark is as follow:

$$\begin{pmatrix} X_q.\theta^* - s^{-*} \\ Y_q + s^{+*} \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{n} \lambda_j^* X_j \\ \sum_{j=1}^{n} \lambda_j^* Y_j \end{pmatrix}$$

Benchmarks are like alerts for designing new strategies or changing old strategies. For each DMU two parts should be taken to account:

1. Eliminate the distance between each DMU and peer group
2. Display the frontier in specific time horizon

As the benchmarks are based on the past data, they couldn't help to show the frontier in specific time horizon. Therefore, ANN is used to indicate the envelope surface.

2.2. Artificial neural networks

The original inspiration for the structure of the neural networks comes from the workings of the brain. The key factor of this paradigm is the novel structure of the information-processing system. The system consists of large number of highly interconnected processing neurons working together to solve specific problems. In similar with people, ANNs learn by example. Neural network is training by adjusting the weights between neurons so that an input leads to a target output.

The fast growth of ANN over the last decade, has introduced a new dimension into the field of performance measurement especially in business application. One of the major application areas of ANNs is forecasting (Sharda, 1994). Many different ANN models have been proposed since 1980s. Multi-layer perceptron (MLP), Hopfield networks, and Kohonen's self-organizing networks are the most influential models.

The MLP networks are used in several problems especially in forecasting because of their inherent capability of arbitrary input – output mapping. Several layers of nodes are included in MLP. The information receiver layer is an input layer, which is the lowest layer. The last or the highest layer is an output layer in which the problem solution is obtained. The hidden layers are the intermediate layers where the input and output layers separated. Acyclic arcs from a lower layer to a higher layer connected the nodes in adjacent layers. Fig. 1. shows an example of a fully connected MLP with one hidden layer.

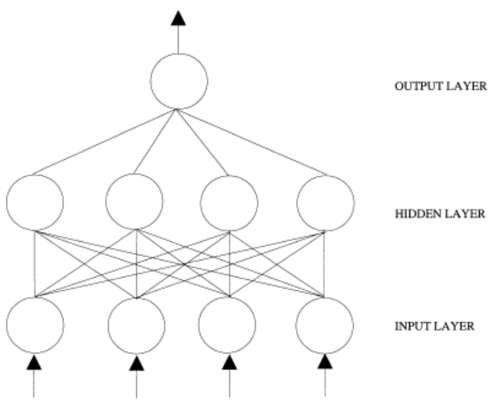

Fig. 1. The structure of three layers MLP network.

Mostly multilayer network trained using the backpropagation (BP) algorithm for forecasting. BP neural networks consist of a collection of inputs and processing units known as neurons.
BP networks are a class of feed-forward neural networks, which refers to the direction of information flow from the input to the output layer, with supervised learning rules. In such a learning each network's forecasts are compared with the known correct answer and the weights are adjusted based on the resulting forecast error to minimize the error function.
For example for forecasting the value of $x(t+1)$ in $x(1)...x(t)$ time series, $x(t-k+1)...x(t)$ is chosen as the inputs to multilayer network and the output will be the forecast. The network uses data, which extracted from the historical time series for training and testing on large training and testing sets.

Before an ANN can be used to perform any desired task, it must be trained to do so. Basically, training is the process of demonstrating the arc weights, which are the key factors of an ANN. Arcs and nodes are saving the learned knowledge in the form of arc weights and node biases. The MLP training is a method of training, in which the desired response of the network (target value) for each input pattern (example) is always available

The steps in training process are usually as following. Firstly, examples of the training set are entered into the input nodes. Secondly, the activation values of the input nodes are weighted and accumulated at each node in the first hidden layer. Lastly, activation value is obtained by an activation function, which is transforming the total into activation value. The value becomes an input into the nodes in the next layer. This process works until the output activation values are found. The training algorithm is trying to the weights that minimize the mean squared errors (MSE) or the sum of squared errors (SSE).

3. ANN-DEA

In this research multilayer ANN is applied to forecast the input and outputs of each DMU in five years. After preliminary analyses and trial, Levenberg–Marquardt algorithm, the fastest training algorithm is chosen for proposed MLP network. Levenberg–Marquardt algorithm, can be considered as a trust-region modification of the Gauss–Newton algorithm.
Two operations must be considered in MLP networks: training and prediction.

MLP uses two data sets, the training set for the training of the MLP and test set for the prediction.
Arbitrary values of the weights, which might be random numbers, are the beginning of training mode. In each epoch, the iteration of the complete training set, the network adjusts the weights. Adjusting the weights is based on reducing error.
The prediction mode begins with information flow from inputs to outputs. The network produces an estimation of output according to the input values. The resulting error demonstrates the quality of prediction of the trained network.

Fig. 2. Shows the two samples of test and train regression charts for proposed ANN.

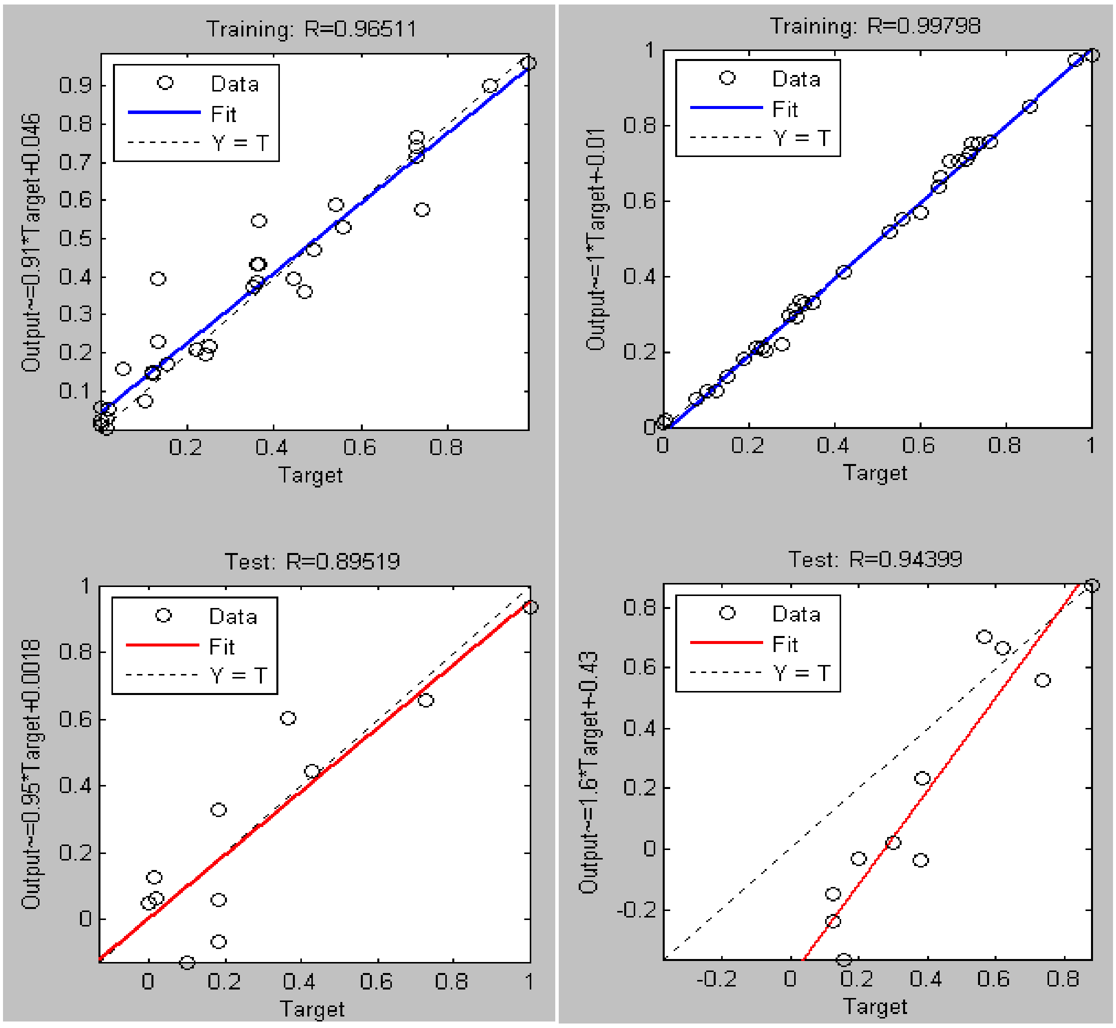

Fig. 2. Training and testing regression charts

Fig.2. Displays the good quality of the trained network prediction.
After forecasting inputs and outputs by ANN, the DEA model must be selected for calculating the efficiency and benchmarks.

Since some inputs and outputs in this study could be negative, the selected DEA model for efficiency measurement and benchmarking should not be sensitive to negative data. One of the best models, which could deal with negative data is SBM model

The SBM model is as follow:

$$Minimize\ \rho = \frac{1 - \frac{1}{m}\sum_{i=1}^{m}(s_i^- / R_i^-)}{1 + \frac{1}{s}\sum_{r=1}^{s}(s_r^+ / R_r^+)}$$

Subject to,

$$\sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = x_{iq} \qquad i = 1,2,\ldots,m$$

$$\sum_{j=1}^{n} y_{ij}\lambda_j - s_r^+ = y_{rq} \qquad r = 1,2,\ldots,s$$

$$\sum_{j=1}^{n} \lambda_j = 1 \quad j = 1,\ldots,n$$

$$\lambda_j, s_i^-, s_r^+ \geq 0 \quad i = 1,2,\ldots,m \quad r = 1,2,\ldots,s \quad j = 1,\ldots,n$$

Where, $R_i^- = \max\{x_{ij} : j = 1,\ldots,n\} - \min\{x_{ij} : j = 1,\ldots,n\}$. The variables *s*+ and *s*– measure the distance of inputs $X\lambda$ and outputs $Y\lambda$ of a virtual unit from those of the unit evaluated (X_q). The numerator and the denominator of the objective function of model measure the average distance of inputs and outputs, respectively, from the efficiency threshold. For variable returns to scale, condition $\sum_{j=1}^{n}\lambda_j = 1$ is added.

The stages involved in the proposed algorithm are illustrated in Fig. 3.

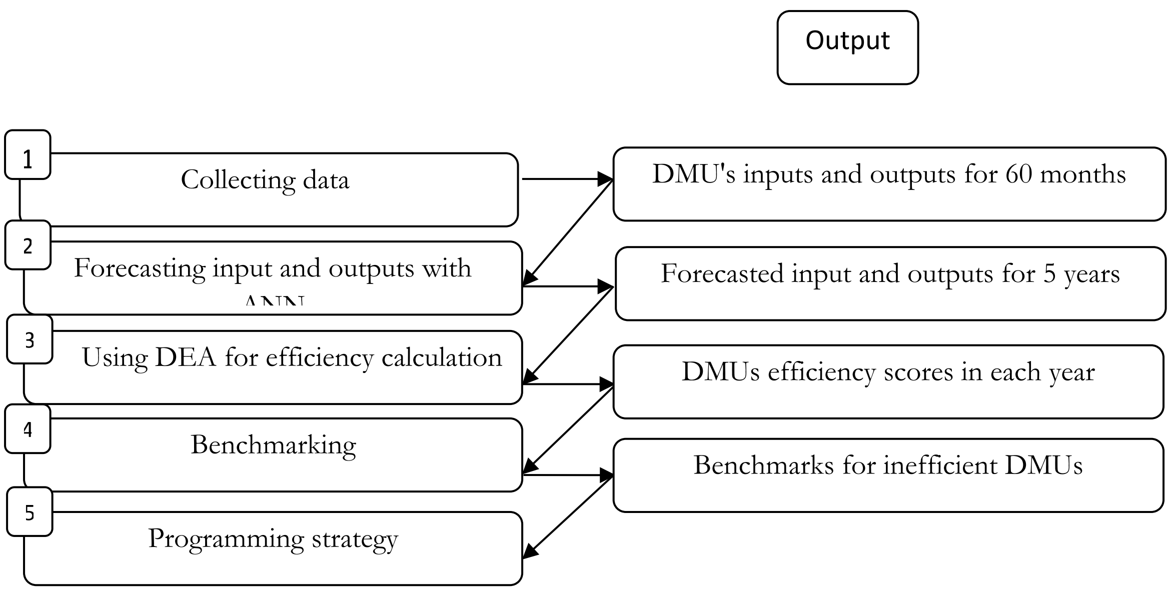

Fig. 3. The steps of ANN-DEA

4. Computational results

100 branch of one of Iranian commercial banks were selected and the related data were collected. The data cover the period of March to February in year 2006 to 2011. Each branch demonstrates a decision-making unit (DMU) and uses two inputs to produce seven outputs as it shown in Table 1.

Table 1- Inputs and outputs of branches

Inputs	Outputs
1- Deposit's paid profit (X_1)	1- Income condominium (Y_1)
2- Expenses (personnel & official) (X_2)	2- Fee (Commission) (Y_2)
	3- Other income (Y_3)
	4- Main deposits (Y_4)
	5- Other deposits (Y_5)
	6- Current deposit (Y_6)
	7- Loan granted account (Y_7)

After implementing ANN and computing the efficiencies by SBM model, for selected11th DMU the efficiency and benchmark is as follow:

$$\rho = 0.005$$

$$Benchmark: \begin{pmatrix} \sum_{j=1}^{100} \lambda_j X_j \\ \sum_{j=1}^{100} \lambda_j Y_j \end{pmatrix}$$

As it shown in Table 2, the 11th DMU should decrease its inputs and increase outputs within 5 years' time horizon. Hence, the 11th DMU will be efficient in five years.

Table 2. Distance between DMU11 and its benchmark

Inputs and outputs	Benchmark	DMU11	Difference
X_1	3.81E+09	7.27E+09	-3.47E+09
X_2	7.21E+08	7.21E+08	0
Y_1	7.13E+09	1.64E+09	5.49E+09
Y_2	3.31E+09	6.4E+07	3.25E+09
Y_3	7.12E+11	6.14E+11	9.83E+10
Y_4	1.35E+12	1.09E+12	2.64E+11
Y_5	9.14E+10	2.28E+10	6.86E+10
Y_6	8.32E+09	0	8.32E+09
Y_7	6.6E+11	1.14E+11	5.46E+11

For DMU 38, the benchmark is displayed in Table 3. For being efficient in five years, the 38th DMU should increase the inputs and outputs.

0.737

Table 3. Distance between DMU38 and its benchmark

Inputs and outputs	Benchmark	DMU38	Difference
X_1	1.71E+12	1.54E+09	1.71E+12
X_2	2.92E+09	8.41E+08	2.08E+09
Y_1	2.25E+11	9.2E+08	2.24E+11
Y_2	4.70E+11	1.03E+08	4.70E+11
Y_3	6.39E+12	5.35E+11	5.86E+12
Y_4	1.87E+12	3.55E+11	1.52E+12
Y_5	4.72E+10	3.23E+09	4.40E+10
Y_6	5.44E+10	0	5.44E+10
Y_7	6.28E+12	4.81E+10	6.23E+12

The efficient DMUs are included in reference set are the peer group for the inefficient units. Therefore, the benchmark and DMU are the same.
For super-efficient DMUs, like DMU1 the benchmark is as Table 4.
1

Table 4. Distance between DMU1 and its benchmark

Inputs and outputs	Benchmark	DMU1	Difference
X_1	3.81E+09	3.81E+09	0
X_2	1.46E+09	1.46E+09	0
Y_1	7.13E+09	7.13E+09	0
Y_2	3.31E+09	3.31E+09	0
Y_3	7.12E+11	7.12E+11	0
Y_4	1.35E+12	1.35E+12	0
Y_5	9.14E+10	9.14E+10	0
Y_6	8.32E+09	8.32E+09	0
Y_7	6.60E+11	6.60E+11	0

Yearly prediction could help each bank branch to have a strategic improvement plan. Hence, the bank management can plan due to this guide, and reach the five-year goal.

5. Conclusions and future works

This paper presents an ANN-DEA study to the branches in one of Iranian commercial banks. The result helps DMUs to improve their efficiency and gives them a useful strategic plan for future development. Through the DEA weakness ANN-DEA approach guides worse performers on how to improve their performance to different efficiency ratings for future. We can also list the following directions for future research: First, Ranking DMUs can be considered for future work. Second, Malmquist productivity index can be used for calculating the DMU's progress or regress.