An Efficiency Measurement and Benchmarking Model Based on Tobit Regression, GANN-DEA and PSOGA

Mohammad Reza Mirzaei
Ph.D. Candidate, Department of industrial management, Central Tehran Branch, Islamic Azad University, Tehran, Iran.
mohammad_reza_mirzaei@yahoo.com

Mohammad Ali Afshar Kazemi
Associated professor, Department of industrial management, Central Tehran Branch, Islamic Azad University, Tehran, Iran (Corresponding author).
dr.mafshar@gmail.com

Abbas Toloie Eshlaghy
Professor, Department of Industrial Management, Science and Research Branch, Islamic Azad University, Tehran, Iran.
toloie@gmail.com

ABSTRACT

The purpose of this study is designing a model based on Tobit regression, DEA, Artificial Neural Network, Genetic Algorithm and Particle Swarm Optimization to evaluate the efficiency and also benchmarking the efficient and inefficient units. This model has three stages, and it uses the data envelopment analysis combined model with neural network, optimized by genetic algorithm, to evaluate the relative efficiency of 16 regional electric companies of Tavanir. A two-staged approach of data envelopment analysis and Tobit regression has been used to measure the effects of environmental variables on the mean efficiency of companies. Finally we use a hybrid model of particle swarm algorithm and genetic algorithm to benchmark the efficient and inefficient units. The mean efficiency of regional electric companies have increased from 0.8934 to 0.9147, during 2012 to 2017, and regional electric companies of Azarbayjan, Isfahan, Tehran, Khorasan, Semnan, Kerman, Gilan and Yazd, had the highest mean efficiency of 1, and west regional electric companies and Fars had the lowest efficiency of 0.7047 and 0.6025, respectively.

Keywords:

Benchmarking, Efficiency, GANN-DEA, PSOGA, Tobit regression.

1. Introduction

Nowadays, the progress and changes in management knowledge, calls for an evaluation system in organizations, to the extent that the lack of such an evaluation system for evaluating the use of resources and facilities, in order to check goal achievement is deemed an organizational disorder. Ranking the firms and companies is one of the most important measurement tools for examining their strengths and weaknesses, and one of the problems with current ranking methods is their focus on only one main factor like sales or income and their being incomprehensive. In other words, these methods focus on determining the biggest and largest companies instead of determining the best companies. So in evaluation, the method should be comprehensive and cover all work-related aspects.

Data Envelopment Analysis (DEA) is a multi-criterion evaluation technique, which with its efficiency and unique features, has increased the accuracy and transparency of evaluation concept in management area. Also data envelopment analysis has some unique features, i.e. in data envelopment analysis, after evaluating the decision making units (DMU) s, for each inefficient unit, a reference set is defined on the efficiency frontier, so that the unit which is under evaluation reaches the reference set on efficiency frontier by decreasing inputs, and increasing outputs, or increasing inputs and decreasing outputs simultaneously. This valuable feature is unique to this technique and other evaluation techniques don’t have it. However, the reference unit for each inefficient unit, is a combination of available efficient units which are imaginary and artificial and don’t actually exist (Ouenniche, Xu, & Tone, 2017).

Combining data envelopment analysis with neural network has provided more generalizability and the ability to estimate non-linear relationships for data envelopment analysis models with few decision making units (Shokrollahpour, Lotfi, & Zandieh, 2016; Toloie-Eshlaghy, Alborzi, & Ghafari, 2012). In the last decade, its two-staged model with Tobit regression has enabled it to measure the effect of environmental variables on mean efficiency in consecutive years (Çelen, 2013; Y. Wu, Hu, Xiao, & Mao, 2016). Also using the hybrid particle swarm algorithm and genetic algorithm hybrid algorithm, it can benchmark for efficient and inefficient units of data envelopment analysis model with genetic neural network. The main purpose of this study is to design a three-staged model by mixing the mentioned features, and also to make a benchmark using the PSOGA algorithm for efficient and inefficient units.

2. Literature Review

Some studies have examined the efficiency and the factors which influence it on distribution electric companies, and they have also investigated the effect of environmental variables on mean efficiency of distribution electric companies. In Sweden (Hjalmarsson & Veiderpass, 1992), no difference has been reported in the efficiency of public and private companies (Bagdadioglu, Price, & Weyman-Jones, 1996). In 2009, on the positive effects of privatization on the efficiency of distribution electric companies, and they also rejected the relationship between private owners and increased efficiency for the developed countries (Pérez-Reyes & Tovar, 2009). In Eastern Europe, the distribution electric companies of Poland are small and inefficient, the distribution electric companies of the Czech Republic are relatively more efficient and the distribution electric companies of Slovakia and Hungary have an average efficiency. Privatization has had positive effects on technical efficiency of the four countries. In these experimental works, data envelopment analysis has been used a lot more than Stochastic frontier analysis (SFA) for analyzing the efficiency in electric area (Cullmann & von Hirschhausen, 2008). A comparison between distribution electric companies of England and Japan during years 1985 to 1998 demonstrated that after reforms in data envelopment model, the efficiency has increased in companies of England (Hattori, Jamasb, & Pollitt, 2003). Through the studies of the Eastern and Western Germany, and based on some experimental models, it was revealed that, on average the distribution electric companies of Eastern Germany have a higher technical efficiency than their counterparts in Western Germany (Hess & Cullmann, 2007). However, in reviewing the Japanese and American companies, on average Japanese companies are more efficient than American companies (Goto & Tsutsui, 2008). The efficiency of 24 distribution electric companies of China was examined with multi-staged data envelopment analysis and companies Hebei, Chinghai, Ningxia, Beijing and Shanghai were among the most efficient companies during the years 2003 to 2010 (Li, Li, & Zheng, 2014). A model was introduced for evaluating and benchmarking the inefficient units using the average efficient models in 18 regional electric companies (Bongo, Ocampo, Magallano, Manaban, & Ramos, 2018). Among the studies on efficiency of electric companies in Iran, is Emami Meibodi (1998) which studied 30 distribution electric companies in Iran. His study showed that the lack of technical efficiency and scale were equally responsible for the inefficiency of distribution electric companies in Iran, and most companies were functioning in ascending output compared to scale (Meibodi, 1998). Fallahi and Ahmadi (2005) studied 42 distribution electric companies in Iran in 2002, and found that the main factor for the inefficiency of distribution electric companies in Iran has been the lack of scale efficiency, and in output, most companies are ascending retuned to scale (Fallahi & Ahmadi, 2005). Sadjadi and Omrani (2008) studied 38 distribution electric companies in Iran and found that the DEA approach can be relatively more reliable for estimating efficiency and ranking strategies (Sadjadi & Omrani, 2008).

3. Methodology

There is only a few decision making units in regional electric companies in Iran, and also the effect of different environmental variables on the efficiency of regional electric companies is evident, so the preachers of this study designed a new hybrid model, that can rank, evaluate efficiency and benchmark the inefficient units with few decision making units, along with measuring the effect of environmental variables on the mean efficiency of consecutive years, and also make the benchmarking of efficient units in regional electric companies possible. The current study uses the library and documentary methods. For doing this, the panel data 16 of distribution electric company in Iran, during 2012 to 2017 has been used and also the detailed statistics of Iranian electric industry has been acquired from the coordination deputy and financial supervisor of power department. The current study is a practical and developmental research. First the tools used in methodology are introduced, and then in conceptual framework, their application and combination have been examined.

3.1. Data envelopment analysis

Measuring the efficiency was first started by Debra and Koopmans (Debreu, 1951; Koopmans, 1951). Farrell (1957) started measuring the efficiency of a production unit, using a method similar to measuring the efficiency in engineering field. The case which Farrell used for efficiency measurement included an input and an output. Farrell's study included measurements of technical and allocation efficiencies and derivative of efficient production function (Farrell, 1957). Farrell started measuring the efficiency of a production unit, using a method similar to measuring the efficiency in engineering field. The case which Farrell used for efficiency measurement included an input and an output. Farrell's study included measurements of technical and allocation efficiencies and derivative of efficient production function (Abraham Charnes, Cooper, & Rhodes, 1978). This efficiency measurement tool needs a set of inputs and outputs which are used in all the decision making units under evaluation. In 1978, Fire and Lovell objected the efficiency measurement and this resulted in developing a new technical efficiency measurement. The results of Fare and Lovell's studies was named Russell's measure (Russell, 1985). The main difference between Farrell's model and Russell's model is that Farrell's measure is radial but Russell's measure is non-radial, so categorizing the output results of these two models are not necessarily the same. In 1985, they invented a model called addictive model, which considered decreased inputs and increased outputs at the same time. In this model, the lack of inputs and excess of outputs is used, which enables it to measure unit's inefficiency, and assign the efficient unit with zero (A Charnes, Cooper, Lewin, & Seiford, 1995). In the current study, a model introduced by Tone is used, which is based on the excess of inputs and lack of outputs and it includes input-oriented, output-oriented, without orientation, superefficient models that can even have negative variables, and in input variables, Si is the excess of input variables of ith input, and in output variables, Sr represents the lack of output variables of rth output (Tone, 2001). Tone's Model is presented in relation (1):

Relation1. SBM unit-independent model (Tone, 2001)

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

$$Subject\ \ to$$

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

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

$$\sum_{j=1}^{n} \lambda_j = 1 \quad j = 1,2,\ldots,n$$

$$\lambda_j, s_i^-, s_r^+ \geq 0$$

Where, ρ is the efficiency value, m refers to the number of inputs, s is the number of outputs, s_i^- returns to surplus of *i*th input slack, s_r^+ refers to shortage of *j*th output slack, λ_j is the corresponding variable to the standard model constraints, x_{ij} returns to *i*th input of *j*th unit, y_{ij} refers to *i*th output of *j*th unit and R returns to reference set (for inputs is R_i^- and for outputs is R_r^+).

3.2. Data envelopment analysis and Tobit regression

Tobit regression was first introduced by Tobin in 1985. It's used for examining the linear relations, in conditions where a critical point is observed in a dependent variable in either left or right boundaries. In this model, it's wrong to interpret coefficient as it is interpreted in ordinary regression, which is the effect of independent variable on dependent variable. Instead, we should consider it with two combined concepts, the first being the changes of dependent variable when it's higher than the lower boundary, with probability weight being higher than lower boundary, and the second concept is it being higher than the lower boundary with the expected probability weight of dependent variable when it's higher than the lower boundary (Tobin, 1958).

The regression Tobit is presented in relation (2):

$$\delta_{it}^* = z_{it}'\beta + \varepsilon_{it} + \alpha \approx N(0,\sigma^2)$$

$$\delta_{it} = \begin{cases} \delta_{it}^* if\ \ 0<\delta_{it}^*<1 \\ 0 \text{ for other values of } \delta_{it}^* \end{cases}$$

Relation 2.Tobit regression model (Pérez-Reyes & Tovar, 2009)

Here, by environment, we mean the factors that affect the business efficiency, but they are not among the used inputs and we assume they're not under management's control. Some instances of environmental variables include ownership differences, and features such as customer density and business position (Pérez-Reyes & Tovar, 2009). For efficiency measurement of regional distribution electric companies by the two-staged approach of data envelopment analysis and Tobit regression, a possible and feasible approach would be directly entering the environmental variables in LP model. Overall, an environmental variable is entered either in a discrete manner, an output variable, or as a neutral discrete manner, but the two-staged approach consists of solving the data envelopment analysis problem in the first stage, that includes only traditional inputs and outputs. In the second stage, the efficiency scores of the first stage are used as dependent variables and environmental variables are used as independent variables. The environmental variables coefficients sign represents the direction of effects and the standard test of assumptions can be applied to check the relation's integrity. One of the benefits of the two-staged approach is that in the second stage regression, both the continuous and discrete variables are used (Çelen, 2013). In the two-staged approach, the two limited dependent Tobit (2LD), is usually constrained to lie between zero and one. Efficiency scores should vary between 0 and 1, or be equal to 0 or 1. Such efficiency score usually exists, but there is no efficiency score of 0 or any score near that. So it can be assumed that one limited dependent Tobit (1LD) which only considers the upper constraint at 1 for a dependent variable can be more efficient, however, since the two limited dependent uses more existing information than one limited dependent, in final calculations of effect, it is asymptotically more efficient (McDonald, 2009).

3.3. Data envelopment analysis and neural network optimized with genetic algorithm

Artificial neural networks are a data processing system, which consists of many simple processing units heavily connected to each other. Since many factors can affect the neural networks performance, such as hidden layers, the number of neurons in hidden layers, the related weights regarding each layer, transfer functions, normalization, and leaning algorithm, the best neural network architecture is achieved through experiment and trial and error (Munakata, 1998) One of the most famous neural network models is the multi-layer perceptron, which consists of an input layer, an output layer, and a layer between them which is not directly connected to input data and output results. In fact this layer is called the hidden layer. Each unit in the hidden layer and output layer operates as a perceptron, with the difference that the function used is a sigmoid function instead of a step function. The units in input layer are only tasked with distributing the input values to the next layer and therefore they don't perform any calculations (Dreyfus, 2005). We should first define an error function which shows the difference between the actual output and the optimum output. Since we already know the optimum output, this learning is called the "supervised learning". To succeed in learning the network, we must gradually transform the actual output to the optimum output. In other words we should consistently decrease the value of error function. To reach this, the weight factors of connection lines of units are adjusted using the Delta rule. The Delta rule calculates the value of error function and it propagates it from one layer to its previous layer. The weight factors of each unit are adjusted separately and this way the error rate decreases. This approach is simple for units of output layer, because we know the actual and optimum output, but it's not really clear in the middle layer. It's assumed that the weight factors of hidden layers, which are connected to output units with big error rates, should change more to hidden units connected to related units with correct outputs. In fact, mathematic relations show that units' coefficients should change according to the error rate of the unit they're connected to. Therefore the weight factors of connected lines of all layers can be accurately adjusted by back-propagation. This way the error function decreases and the network learns (Dreyfus, 2005). In current study, to evaluate the neural networks performance, Mean Squared Error (MSE) and correlation coefficient (r) is used. In table (1) the neural networks performance evaluation tools are mentioned.

Table 1. Networks performance evaluation tools

Network performance evaluation tool	*Formula*
Mean Squared Error	$MSE = \frac{\sum_{i=1}^{n}(X_{obs,i} - X_{model,i})^2}{n}$
Correlation Coefficient	$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}}$

The neural network NN-DEA is a multi-layered perceptron with a back-propagation algorithm, which has an input layer, a hidden layer and an output layer. The transfer function of hidden layer is a hyperbolic tangent function and the transfer function of output layer is a linear function. In this network, the input includes the sum of inputs and the outputs of each decision making unit, and the expected output (O) is the efficiency of each decision making unit (Athanassopoulos & Curram, 1996; Costa & Markellos, 1997; Emrouznejad & Shale, 2009; Samoilenko & Osei-Bryson, 2010; Shokrollahpour et al., 2016; Toloie-Eshlaghy et al., 2012; D. D. Wu, Yang, & Liang, 2006). Of course probabilistic networks have also been used to evaluate the efficiency of top Arabic ranking banks (Mostafa, 2009). Genetic Algorithm (GA) is a search method for widespread probabilities which follows the natural biologic evolution. Genetic algorithm operates on potential answers and uses the survival of the fittest concept in producing better and better estimates of local answers. In each generation, a new set of estimations are produced through the process of selecting the best member based on their fitness in the scope of the problem and also duplicating through the operators taken from natural genetics. This process ultimately results in the evolution of members which are more consistent with environment than the original members namely their actual parents (Goldberg & Holland, 1988). In 1970, John Holland introduced and created this method (Holland, 1975), but ultimately Goldberg, one of his students, gathered and presented the idea (Goldberg & Holland, 1988). Of course De Jong's researches in his PhD dissertation approved the efficiency of genetic algorithm (De Jong, 1975). The purpose of applying genetic algorithm is optimizing the weight of artificial neural network. So the objective function of genetic algorithm is a function of statistical results of artificial neural network. In each generation, a part of the population is given random values for the purpose of learning through neural network, and then the error rate is calculated using the training data. Then the network parameters are updated considering the values of objective function and the rate of mutation and cross-over. Each chromosome is a set of genes that attempt to optimize the neural network parameters and estimate the water table values. The focus of this method is on a population of points as a set of solutions which prevents the genetic algorithm to fall in local minimums. In the next step, objective functions of first generation chromosomes are calculated. In this step, the genetic algorithm parameters are evolved through training the chromosomes by neural network and evaluating the produced masses in reproduction stage. With new generations, the neural network trained by first generation data may result in error, therefore the training of chromosomes is repeated periodically. The focus of neural genetic search method is on a population of points as a set of solutions which prevents the genetic algorithm to fall in local minimums. Also using the neural network decreases the time consuming complex calculations and applies statistical rules instead of explicit rules. The neural network used in this method is the same perceptron neural network in NN-DEA model, consisting of an input layer, a middle layer, and an output layer, which is combined with genetic algorithm and is called GANN-DEA.

3.4. The hybrid algorithm of particle swarm optimization and genetic algorithm

The particle swarm optimization method or particle swarm optimization in short, is modeled by simulating the social behavior of a bird flock. James Kennedy, social psychologist, and Russell C. Eberhart, electrical engineer, first intended to use social models and existing social relationships in order to create some kind of computational intelligence that didn't need personal special abilities (Eberhart & Kennedy, 1995). This led them to simulate the behavior of birds in searching for seeds. Their work was influenced by Hyper and Gernandre's work to simulate the behavior of birds as a non-linear system (Abd-El-Wahed, Mousa, & El-Shorbagy, 2011). There are entities in particle swarm optimization algorithm which we call particle, and they're distributed in the search-space of the function we intend to minimize (or optimize). Each particle calculates the objective function value in the position it's situated. Then it chooses a direction of movement, using a combination of its current position, its last best known position, and also the data of one or multiple particles which are considered the best particles in the population. All particles choose a direction to move, and after they move a step of the algorithm is finished. These steps are repeated several times until a satisfactory solution is achieved. In fact the particle swarm that searchers for the minimum of a function, operate similar to a bird flock searching for seeds (Abd-El-Wahed et al., 2011; Eberhart & Kennedy, 1995). Considering these cases, a new velocity is calculated for birds in each step. The particle swarm algorithm has two main operators, velocity update and position update. At first, some random birds (solutions) are produced. Then each one of them is assigned a velocity. A new velocity is calculated for the bird based on its current velocity, the distance from its last best known position, and also the distance from other birds' best known positions. Since the obtained velocity is equal to the bird's movement during one step, the bird's new position is achieved in the next step after updating its position. This process is repeated a defined number of times and the solution of the problem is the final position met by all birds (Abd-El-Wahed et al., 2011). Angeline suggested the first approach to mix the concepts of GA with particle swarm optimization which shows that (PSO) function can be improved for specific classes of problems by adding a process similar to what happens in evolutionary algorithms. The selection phase is done before the velocity update and it can be experimentally shown that this algorithm improves the local search ability of (PSO). Since half of particles are replaced by the other half, the solution diversity is decreased by 50 percent in each repetition. The diversity can be replaced by swapping the worst particles and a mutated copy of best particles, and both algorithms are used consistently (Angeline, 1998). Abdol Vahed et al solved different non-linear test functions with PSOGA algorithm and introduced the mentioned algorithm faster and with better solutions(Abd-El-Wahed et al., 2011). Garg (2016) solved two limited mechanical problems (pressure vessel design and welding design) using PSOGA algorithm, and described this approach as efficient (Garg, 2016). Also, in the current study, the combination of GA concepts along with particle swarm optimization which was presented by Garg (2016) has been used. This model consists of two GA and PSO pseudo code, which are described below.

Algorithm1. Pseudo code of Genetic algorithm (GA)

1: Objective function: $f(\mathbf{x})$
2: Define Fitness F (eg .F $\propto f(x)$ for maximization)
3: Initialize population
4: Initial probabilities of crossover (pc) and mutation (pm)
5: **do**
6: Generate new solution by crossover and mutation
7: if pc >rand, Crossover; end if
8: if pm >rand, Mutate; end if
9: Accept the new solution if its fitness increases.
10: Select the current best for the next generation.
11: **While maximum iterations or minimum error criteria is not attained**

Algorithm2. Pseudo code of Particle swarm optimization (PSO)

1: Objective function: $f(x)$, $x = (x1, x2, \ldots, x_n)$;
2: Initialize particle position and velocity for each particle and set k=1.
3: Initialize the particle's best known position to its initial position i.e. $p_{ik} = x_{ik}$.
4: **do**
5Update the best known position (p_{ik}) of each particle and swarm's best known position (p_{gk}).
6: Calculate particle velocity according to the velocity equation $v_{ik+1}=w\cdot v_{ik}+c_1\cdot r_1\cdot(p_{ik}-x_{ik})+c_2\cdot r_2\cdot(p_{gk}-x_{ik})$.
7: Update particle position according to the position equation $x_{ik+1}=x_{ik}+v_{ik+1}$.
8: **While maximum iterations or minimum error criteria is not attained.**

Where r_1 and r_2 represent random numbers between 0 and 1, c_1 and c_2 are constants, p_{ik} represents the best ever position of ith particle, and p_{gk} corresponds to the global best position in the swarm up to kth iteration.

A complete iteration of the PSOGA algorithm is that a pseudo code of the PSO algorithm, after having a iteration, places the answer on the GA algorithm. At the end of a iteration of the GA algorithm, the output is the result of a local answer of hybrid PSOGA algorithm in its iteration. This loop continues until it reaches the stop condition (Garg, 2016).

3.5. Research implementation steps and conceptual framework of this research

Step1: Introducing variables

The panel data 16 of distribution electric company in Iran during the years 2012 to 2017 are divided into two groups in Tobit regression, i.e. before 2013 and after 2013, for privatization imaginary variable. Control (environmental) variables include privatization imaginary variable (DUMPRIVATE) for controlling the ownership structure, the underground network length to overall network length ratio for controlling the network structure (UGR), home customers to all customers ratio for controlling the consumer structure (CONSRESSHARE), and the network load coefficient (the maximum asynchronous load to overall electricity consumption ratio) (LF1) and transformer capacity load coefficient (transformer capacity to electricity demand ratio) (LF2) for controlling the intensity of network use and transformer (LF2) respectively, and circuit density as number of customers to network length ratio (CD2) and customer density as number of customers to the coverage area ratio for controlling the operational environment.

Also the variables under management's control which are put in the so-called box of data envelopment analysis as inputs and the outputs of SBM unit-independent model are: 1-The length of network lines (kilometers), 2-Transformer capacities (Mega Volts-Ampere), 3-Number of employees (people) and 4-Transmission and distribution losses (percent) as input variables and 1-Number of customers (thousand people) and 2-The energy delivered to customers (Million Kilo watts-hour) as output variables.

Step2: Solving the classical model of SBM unit-independent model and Tobit regression

The results of SBM unit-independent model efficiency and environmental variables are put in a two-staged model of data envelopment analysis and Tobit regression and then the effect and final impact of environmental variables on mean efficiency of regional distribution and transmission electric companies are determined. In this research, maximum likelihood of random panel is used. Tobit regression in the current study is described in relation (3):

Relation (3):

$Y_{it} = \alpha + \beta_1 x_{1\,it} + \beta_2 x_{2\,it} + \beta_3 x_{3\,it} + \beta_4\, x4_{it} + \beta_5 x_{5\,it} + \beta_6 x_{6\,it} + \beta_7 x_{7\,it} + u_{it}$

Y_{it} is the final impact of environmental variables on mean efficiency of units, α is Y-intercept, and u_{it} is the remaining error (which is independent and identically distributed N $(0,\ б^2)$). The degree of freedom is 8.

Step3: Making the first GANN-DEA network

Furthermore, we make the GANN-DEA model using the efficiencies obtained through SBM unit-independent model and input and output values, at this point 70 percent of data are used as training data, 15 percent as test data and the remaining 15 percent as validation data.

Step4: Benchmarking the decision making units with PSO-GA

Then for the inefficient units, we present values for slack variables for input and output variables using the hybrid algorithm PSO-GA so that they can be used to improve the efficiency in the next year. We perform this step by putting the data from year 2017 in SBM unit-independent model and we try to transform the efficiency values of hybrid algorithm PSO-GA to efficiency values of GANN-DEA as objective to obtain the values for slack variables for GANN-DEA and for each inefficient unit, a hybrid algorithm PSO-GA is needed. We implement SBM unit-independent model of data envelopment analysis in fitness function of this algorithm and through trial and error and different adjustments for the parameters of this hybrid algorithm, we get the efficiency of GANN-DEA model which is designed by an SBM unit-independent model itself. Out of the best local solutions, the value for slack variables, which is the famous benchmarking of data envelopment analysis model, is obtained. The conceptual framework for this research is illustrated in figure (1).

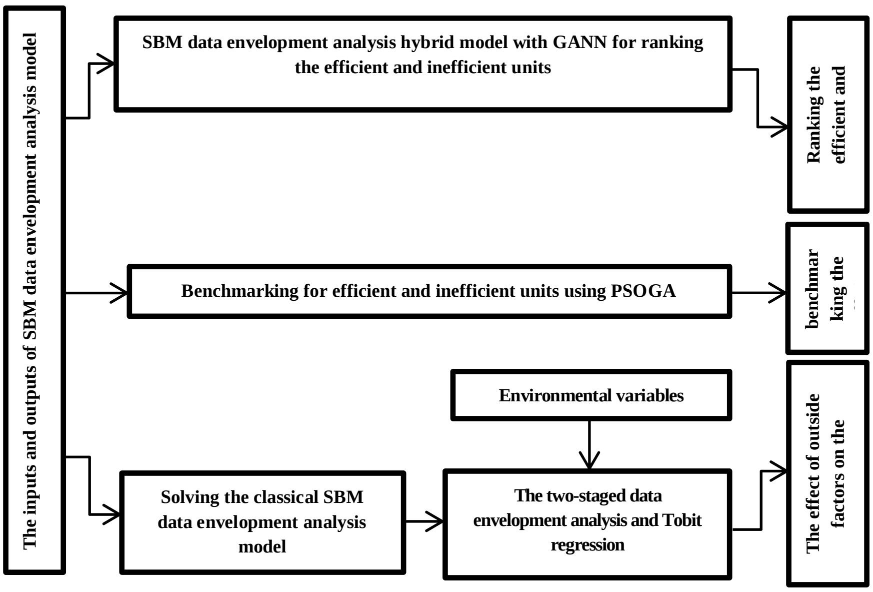

Figure 1. Conceptual framework for this research

Step5: Ranking and benchmarking efficient units with model by Cook and Green

For ranking and presenting a benchmark model for efficient units, we use the decision making units clustering method from data envelopment analysis model by Cook and Green (2005) under "The evaluation of power plants: a hierarchical model" which has been done for electric industry (Cook & Green, 2005). In this method, units which become efficient in each input and output level are left aside as efficient units and the process is repeated to create several levels of efficiency boundaries. Units which become efficient in the first level are recognized as efficient units and are eliminated in post model execution; But we start building virtual units for showing the reference units for these reference units (which have become efficient in the first level); Following that for each factor, we choose the lowest input from inputs and the highest value among outputs, virtual unit, is a unit that has not been actualized but with the set of experienced units, such a unit can be actualized and with adding the virtual unit to decision making units executing the model, a reference unit is defined for reference units.

Step6: Making second GANN-DEA network and PSO-GA algorithm for efficient units

Then we implement and repeat the steps from GANN-DEA model and PSO-GA hybrid algorithm again for efficient units and virtual unit (which is a new unit). With the difference that for designing a GANN-DEA model for efficient units, data from years 2016 and 2017 has been used and also 50 percent of the data is as training data, 25 percent as test data, and the remaining 25 percent as validation data because of the lack of row data, but similar to inefficient units, the data from year 2017 has been used in PSO-GA hybrid algorithm and again for each decision making unit a separate PSO-GA hybrid algorithm is needed.

4. Results

The regional distribution and transmission electric companies of Azarbayjan, Isfahan, Tehran, Khorasan, Semnan, Kerman, Gilan and Yazd have been consistently efficient during the years 2012 through 2017. The efficiency values of regional distribution and transmission electric companies of Tavanir have been obtained using SBM unit-independent model and efficiency variable to scale ratio, during 2012 to 2017 have been stated in table (2).

All the seven environmental variables have not significant impact on mean efficiency. The results of the two-staged Tobit regression and data envelopment analysis model shows the positive or negative impact of environmental variables on the group efficiency along with the final impact of environmental variables on the group efficiency. The final impact of environmental variables and also their positive or negative effect are of significance for managers of any organization. In this study, as a result of limitations in the gathering of statistical data during the years before 2012 in regional electric company of Tavanir, we have sufficed to present short term impacts of environmental variables on group efficiency. Parameters affecting the efficiency of SBM unit-independent model with variable efficiency to scale which are results of the two-staged data envelopment analysis and Tobit regression model are stated in table (3).

To perform GANN-DEA model, we first designed NN-DEA model based on research background and theoretical framework and then we found the best adjustments for NN-DEA and the genetic algorithm that optimizes it through trial and error. Best adjustments for GA include roulette wheel for selection method and the primary population of 20, double coupling at the rate of 0.9, mutation with uniform function and mutation rate of 0.01 and halt condition of 60 minutes or 100 generations.

Comparison of the SBM unit-independent efficiency with the GANN-DEA Model in 2017 has been stated in table (4). Best adjustments for PSO-GA include Cognitive and social components are constants set to 1.5. Inertia weight is 0.4 to 0.9 and GA having a Single point crossover operation with the rate of 0.85. Mutation rate is set to be 0.02. The other randomly control parameters are taken as, minimum number of iteration in GA set to 10, maximum number of iteration in GA set to 20, decreasing rate of no. of individuals that effected by GA set to 10, increasing rate of GA maximum iteration set to 15, first population size in GA set to 1 and last population size in GA set to 20.

The summary of optimization of the two neural networks with genetic algorithm has been stated in table (5) and they are in acceptable range of accuracy

Table 2. The efficiency of SBM unit-independent model and efficiency of variable to scale ratio of regional distribution and transmission companies of Tavanir during 2012 to 2017

Regional electric company	DEA model	Year						Mean SBM
		2012	2013	2014	2015	2016	2017	
Azarbayjan	SBM	1	1	1	1	1	1	1
	Super Efficiency	1.062256	1.064183	1.028367	1.036538	1.031591	1.030269	--
Isfahan	SBM	1	1	1	1	1	1	1
	Super Efficiency	1.066985	1.149197	1.194992	1.17625	1.150807	1.141013	--
Bakhtar	SBM	0.71412	0.753131	0.718155	0.701485	0.771126	0.774004	0.73867
	Super Efficiency	0.430592	0.753131	0.718155	0.701485	0.771126	0.774004	--
Tehran	SBM	1	1	1	1	1	1	1
	Super Efficiency	1.553038	1	1	1	1	1	--
Khorasan	SBM	1	1	1	1	1	1	1
	Super Efficiency	1.356831	1.136353	1.104949	1.081925	1.09131	1.10517	--
Khuzestan	SBM	0.865687	0.804222	0.821384	0.783101	0.817439	1	0.848639
	Super Efficiency	0.517863	0.804222	0.821384	0.783101	0.817439	1.015421	--
Zanjan	SBM	0.836451	0.863951	0.849852	0.858508	0.807127	0.804779	0.836778
	Super Efficiency	0.485461	0.863951	0.849852	0.858508	0.807127	0.804779	--
Semnan	SBM	1	1	1	1	1	1	1
	Super Efficiency	1.49742	1.283835	1.340388	1.377624	1.340355	1.366307	--
Sistan and Baluchestan	SBM	0.999721	0.999702	1	0.999745	0.999422	1	0.999765
	Super Efficiency	1.164043	0.999702	1.013399	0.999745	0.999422	1.01945	--
Gharb	SBM	0.665177	0.711323	0.727271	0.709085	0.699549	0.71636	0.704794
	Super Efficiency	0.454613	0.711323	0.727271	0.709085	0.699549	0.71636	--
Fars	SBM	0.550462	0.583372	0.583682	0.601632	0.626865	0.669565	0.602596
	Super Efficiency	0.44704	0.583372	0.583682	0.601632	0.626865	0.659565	--
Kerman	SBM	1	1	1	1	1	1	1
	Super Efficiency	1.258877	1.167371	1.171332	1.119649	1.177083	1.203153	--
Gilan	SBM	1	1	1	1	1	1	1
	Super Efficiency	1.171777	1.160851	1.116177	1.099063	1.101866	1.095863	--
Mazandaran	SBM	0.808467	0.838839	0.826026	0.842302	0.788371	0.801438	0.817574
	Super Efficiency	0.493216	0.838839	0.826026	0.842302	0.788371	0.801438	--
Hormozgan	SBM	0.855413	0.806408	0.811272	0.800218	0.79706	0.848337	0.819785
	Super Efficiency	0.454285	0.806408	0.811272	0.800218	0.79706	0.848337	--
Yazd	SBM	1	1	1	1	1	1	1
	Super Efficiency	1.078335	1.058751	1.062996	1.060528	1.072376	1.077398	--
Mean	SBM	0.893469	0.897559	0.896103	0.893505	0.894185	0.914704	--

Table3. The factors which affect the efficiency of two-staged data envelopment analysis and Tobit regression model

Variable	The final impact of variables	Statistic t	Significance value (P-value)
α_0	--	***2/935	0/887
CONSRESSHARE	-0/956	***-1/55	30/243
UGR	-0/649	*0/158	-0/321
CD2	-0/891	*-1/559	0/651
CD1	6/704	*0/154	0/075

Variable	The final impact of variables	Statistic t	Significance value (P-value)
LF1	-8/221	**-1/485	-0/545
LF2	7/975	*-1/181	0/142
DUMPRIVATE	0/445	**0/050	0/510
Log Likelihood	-8/399		
Chi-squared statistic	--	139/09	0/000
Likelihood ratio test	--	37/64	0/000
Number of views	112(censored: 51)		
Number of companies	16		
*,**,*** are respectively significance level of 1 percent, 5 percent and 10 percent using a two-tailed test.			

Table 4. Comparison the efficiency of SBM unit-independent with the GANN-DEA Model in 2011

Regional electric company	SBM-DEA Model efficiency in 2017	GANN-DEA efficiency in 2017
Azarbayjan	1	0.994109
Isfahan	1	0.9931097
Bakhtar	0.774004	0.778652
Tehran	1	1.007727
Khorasan	1	1.004362
Khuzestan	1	1.007092
Zanjan	0.804779	0.807656
Semnan	1	1.006518
Sistan and Baluchestan	1	1.002869
Gharb	0.71636	0.710237
Fars	0.669565	0.676048
Kerman	1	1.005444
Gilan	1	0.994384
Mazandaran	0.801438	0.805315
Hormozgan	0.848337	0.84734
Yazd	1	0.994409
Mean	0.91278019	0.914704

Table 5. The summary of optimization of two neural networks GANN-DEA

Optimization summary	Best fitness (neural network including all units)	Mean fitness (neural network including all units)	Best fitness (neural network including efficient units)	Mean fitness (neural network including efficient units)
Generation	4	5	5	7
Lowest mean squared error	0.000413737	0.00149959	3.96483E-05	0.000142828
Mean final Squared error	0.000413737	0.00149959	3.96483E-05	0.000142828

The function of NN-DEA neural networks after being optimized with genetic algorithm have been stated in table (6) and they are in acceptable range of accuracy.

The efficiencies estimated by GANN-DEA and PSOGA algorithm for efficient and inefficient units and also the values of covariate variable for benchmarking the inefficient units by PSOGA hybrid algorithm have been stated in tables (7) and (8).

PSOGA try to obtain the amount of efficiency which is close to the amount of GANN-DEA efficiency, and then the values obtained with the slack variables for benchmarking decision making unit. In this way, the values of the input variable must decrease and the values of the output variables must increase. For example, the regional electric company of Bakhtar has to reduce the length of transmission lines to 4075.862 km/h, decreasing capacity of transformer to 2655.433 MVA, reducing the number of personnel to 255 persons and transmission losses to 0 percent, as well as increasing the number of subscribers to 0 million kWh and energy delivered to subscribers at 0.0000501 million kWh, to reach the frontier curve. The description of other units are according to table (8), but the regional electric companies are efficient in this table and are trying to reach the virtual efficient frontier that is represented within the conceptual framework of this research.

Table 6 . The results of evaluation tools of neural networks

Neural network performance	Output of neural network including all units	Output of neural network including efficient units
Mean squared error	0.000413737	0.000142828
Correlation coefficient	0.99970905	0.999817726

Table 7. Efficiencies of GANN-DEA and PSOGA for inefficient units and values of covariate variable for benchmarking the inefficient units by PSOGA

Regional electric company	GANN-DEA efficiency	PSOGA efficiency	S6(PSOGA)	S5(PSOGA)	S4(PSOGA)	S3(PSOGA)	S2(PSOGA)	S1(PSOGA)
Bakhtar	0.778652	0.774007	5.01E-05	5E-05	0.199862	255.4729	2655.433	4075.862
Fars	0.676048	0.671708	4.97E-05	5.01E-05	0.027146	998.0688	8475.363	6153.824
Gharb	0.710237	0.716365	5.26E-05	5.02E-05	0.542077	89.37919	1933.935	4786.697
Hormozgan	0.847341	0.844972	920.3739	0	0.873357	104.7047	1884.729	1.09E-05
Mazandaran	0.805315	0.803648	5.11E-05	4.98E-05	0.458319	130.4469	1723.022	1136.608
Zanjan	0.807656	0.804789	5.04E-05	5.01E-05	0.591003	155.0925	577.446	1044.265

Table 8. Efficiencies of GANN-DEA and PSOGA for efficient units and values of covariate variable for benchmarking the efficient units by PSOGA

Regional electric company	GANN-DEA efficiency	PSOGA efficiency	S6(PSOGA)	S5(PSOGA)	S4(PSOGA)	S3(PSOGA)	S2(PSOGA)	S1(PSOGA)
Azarbayjan	0.344872	0.340418	787.828	0.009154	0.945692	872.8638	6549.98	5932.302
Isfahan	0.404413	0.404158	5350.029	19122.69	0.510794	521.3247	13542.99	7569.992
Khorasan	0.381241	0.382902	4888.003	24094.56	0.508069	2167.97	35128.99	6882.992
Khuzestan	0.291241	0.29752	0.040636	20010.93	1.166796	1884.068	21156.98	6458.989
Semnan	0.926037	0.928995	7580.16	36352.14	0.191	81.53098	0.001055	0.001321
Sistan and Baluchestan	0.499521	0.495196	34892.85	1.519953	612.2031	1031.076	5333.992	0.439948
Gilan	0.689944	0.682042	948.7077	2275.122	1.12	437.9996	2217.995	604.9967
Yazd	0.742596	0.748482	7539.807	34205	0.001	192.0088	2476.154	1545.257
Tehran	0.291444	0.2952163	0.001	0.001	0.6377	2222	35130.007	6882.999
Kerman	0.560727	0.5673636	7067.6872	30429.006	1.1889	0.00287	5623.0078	6028.922

According to table (8) for example, the regional electric company of Isfahan has to reduce the length of transmission lines to 7569.992 km/h, decreasing capacity of transformer to 13542.99 MVA, reducing the number of personnel to 521.3247 persons and transmission losses to 0.510794 percent, as well as increasing the number of subscribers to 19122.69 million kWh and energy delivered to subscribers at 5350.029 million kWh, to reach the virtual efficient frontier.

5. Discussion and Conclusions

Among the efficient companies in year 2017 (current situation), the best performance belongs to regional electric company of Semnan, and the lowest super efficiency score belongs to regional electric company of Tehran. Ranking the efficiency of efficient companies in 2017, are regional distribution and transmission electric companies of Semnan, Kerman, Isfahan, Khorasan, Gilan, Yazd, Azarbayjan, Sistan and Baluchestan, Khuzestan and Tehran respectively. The mean efficiency scores of regional electric companies have decreased a little bit during 2012 to 2014 but have gone ascending from 2013 onward. The regional distribution and transmission electric company of Khuzestan was first recognized as efficient in 2016, and has had more ascending progress than the other regional electric companies. The regional distribution and transmission electric companies of Bakhtar, Zanjan, Gharb, Fars, Mazandaran and Hormozgan have consistently been recognized as inefficient during 2012 to 2017. Ranking the efficiency of inefficient companies in 2016, are regional distribution and transmission electric companies of Hormozgan with 0.848337 efficiency, Zanjan with 0.804779 efficiency, Mazandaran with 0.801438 efficiency, Bakhtar with 0.774004 efficiency, Gharb with 0.71636 efficiency and Fars with 0.669565 efficiency respectively.

The results of the two-staged Tobit regression and data envelopment analysis model shows the positive or negative impact of environmental variables on the group efficiency along with the final impact of environmental variables on the group efficiency.

NN-DEA models have not been able to benchmark for decision making unites until now, but in the current study, this was possible by the hybrid PSOGA algorithm. Adding benchmarking for efficient units also has other characteristics.

The model used in the current study has new features such as considering the efficiency of several consecutive years for estimating the efficiency of last year, it being less influenced by disturbing data, presenting a benchmark for inefficient units by considering the efficiency of GANN-DEA, and also benchmarking for efficient units after clustering of decision making units considering the efficiency of GANN-DEA and it also provides the ability to measure the impact of environmental variables on mean efficiency for organization evaluators. For further research, we suggest designing data envelopment analysis models of Malcolm Quist and fuzzy models similar to the multi-staged model designed in this study.

References

1) Abd-El-Wahed, W., Mousa, A., & El-Shorbagy, M. (2011). Integrating particle swarm optimization with genetic algorithms for solving nonlinear optimization problems. Journal of Computational and Applied Mathematics, 235(5), 1446-1453 .
2) Angeline, P. J. (1998) .Using selection to improve particle swarm optimization. Paper presented at the Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on.
3) Athanassopoulos, A. D., & Curram ,S. P. (1996). A Comparison of Data Envelopment Analysis and Artificial Neural Networks as Tools for Assessing the Efficiency of Decision Making Units. Journal of the Operational Research Society, 47(8), 1000-1016.
4) Bagdadioglu, N ,.Price, C. M. W., & Weyman-Jones, T. G. (1996). Efficiency and ownership in electricity distribution: a non-parametric model of the Turkish experience. Energy Economics, 18(1-2), 1-23 .
5) Bongo, M. F., Ocampo, L. A., Magallano, Y. A. D., Manaban, G. A & ,.Ramos, E. K. F. (2018). Input–output performance efficiency measurement of an electricity distribution utility using super-efficiency data envelopment analysis. Soft Computing.

6) Çelen, A. (2013). Efficiency and productivity (TFP) of the Turkish electricity distribution companies: An application of two-stage (DEA&Tobit) analysis. Energy Policy, 63, 300-310 .
7) Charnes, A., Cooper, W., Lewin, A., & Seiford, L. (1995). Data Envelopment Analysis: Theory, Methodology and Applications ,Kluwer Publications .
8) Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. European Journal of Operational Research, 2(6), 429-444 .
9) Cook, W. D., & Green, R. H. (2005). Evaluating power plant efficiency: a hierarchical model. Computers & Operations Research, 32(4), 813-823 .
10) Costa, Á., & Markellos, R. N. (1997). Evaluating public transport efficiency with neural network models. Transportation Research Part C: Emerging Technologies, 5(5), 301-312 .
11) Cullmann ,A., & von Hirschhausen, C. (2008). Efficiency analysis of East European electricity distribution in transition: legacy of the past. Journal of Productivity Analysis, 29(2), 155 .
12) De Jong, K. A. (1975). Analysis of the behavior of a class of genetic adaptive systems .
13) Debreu, G. (1951). The Coefficient of Resource Utilization, Econometric, 19, Economics: Principles and Applications: Zaria: AGTAB Publishers Ltd.
14) Dreyfus, G. (2005). Neural networks: methodology and applications: Springer Science & Business Media.
15) Eberhart, R., & Kennedy, J. (1995). A new optimizer using particle swarm theory. Paper presented at the Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on.
16) Emrouznejad, A., & Shale, E. (2009). A combined neural network and DEA for measuring efficiency of large scale datasets. Computers & Industrial Engineering, 56(1), 249-254 .
17) Fallahi, M., & Ahmadi, V. (2005). Cost efficiency analysis of electricity distribution companies in Iran. Journal of Economic Researches, 71, 297-320 .
18) Farrell, M. J. (1957). The measurement of productive efficiency. Journal of the Royal Statistical Society. Series A (General), 120(3), 253-290 .
19) Garg, H. (2016). A hybrid PSO-GA algorithm for constrained optimization problems .Applied Mathematics and Computation, 274, 292-305 .
20) Goldberg, D. E., & Holland, J. H. (1988). Genetic algorithms and machine learning. Machine learning, 3(2), 95-99 .
21) Goto, M., & Tsutsui, M. (2008). Technical efficiency and impacts of deregulation: An analysis of three functions in US electric power utilities during the period from 1992 through 2000. Energy Economics, 30(1), 15-38 .
22) Hattori, T., Jamasb, T., & Pollitt, M. G. (2003). A comparison of UK and Japanese electricity distribution performance 1985-1998 :lessons for incentive regulation .
23) Hess, B., & Cullmann, A. (2007). Efficiency analysis of East and West German electricity distribution companies–Do the “Ossis” really beat the “Wessis”? Utilities Policy, 15(3), 206-214 .
24) Hjalmarsson, L., & Veiderpass, A. (1992). Efficiency and ownership in Swedish electricity retail distribution International Applications of Productivity and Efficiency Analysis (pp. 3-19): Springer.
25) Holland, J. (1975). Adaptation in natural and artificial systems: an introductory analysis with application to biology. Control and artificial intelligence .
26) Koopmans, T. C. (1951). Efficient allocation of resources. Econometrica: Journal of the Econometric Society, 455-465 .
27) Li, J., Li, J., & Zheng, F. (2014). Unified efficiency measurement of electric power supply companies in China. Sustainability, 6(2), 779-793 .
28) McDonald, J. (2009). Using least squares and tobit in second stage DEA efficiency analyses. European Journal of Operational Research, 197(2), 792-798 .
29) Meibodi, A. E. (1998) .Efficiency considerations in the electricity supply industry: The case of Iran: university of Surrey.
30) Mostafa, M. M. (2009). Modeling the efficiency of top Arab banks: A DEA–neural network approach. Expert systems with applications, 36(1), 309-320 .
31) Munakata, T. (1998). Fundamentals of the new artificial intelligence (Vol. 2): Springer.

32) Ouenniche, J., Xu, B., & Tone, K. (2017). DEA IN PERFORMANCE EVALUATION OF CRUDE OIL PREDICTION MODELS. Advances in DEA Theory and Applications: With Extensions to Forecasting Models, 381-403 .
33) Pérez-Reyes, R., & Tovar, B. (2009). Measuring efficiency and productivity change (PTF) in the Peruvian electricity distribution companies after reforms. Energy Policy, 37(6), 2249-2261 .
34) Russell, R. R. (1985). Measures of technical efficiency. Journal of Economic Theory, 35(1), 109-126 .
35) Sadjadi, S., & Omrani, H. (2008). Data envelopment analysis with uncertain data: An application for Iranian electricity distribution companies. Energy Policy, 36(11), 4247-4254 .
36) Samoilenko, S & ,.Osei-Bryson, K.-M. (2010). Determining sources of relative inefficiency in heterogeneous samples: Methodology using Cluster Analysis, DEA and Neural Networks. European Journal of Operational Research, 206(2), 479-487 .
37) Shokrollahpour, E., Lotfi, F. H & ,.Zandieh, M. (2016). An integrated data envelopment analysis–artificial neural network approach for benchmarking of bank branches. Journal of Industrial Engineering International, 12(2), 137-143 .
38) Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. Econometrica, 26(1), 24-36.
39) Toloie-Eshlaghy, A., Alborzi, M., & Ghafari, B. (2012). Assessment of the personnel's efficiency with Neuro/DEA combined model .
40) Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. European Journal of Operational Research, 130(3), 498-509 .
41) Wu, D. D., Yang, Z., & Liang, L. (2006). Using DEA-neural network approach to evaluate branch efficiency of a large Canadian bank. Expert systems with applications, 31(1), 108-115.

Wu, Y., Hu, Y., Xiao, X., & Mao, C. (2016). Efficiency assessment of wind farms in China using two-stage data envelopment analysis. Energy Conversion and Management, 123, 46-55 .