Advances in Industrial Engineering, Summer 2019, 53(3): 15-30
DOI: 22059/jieng.2020.307008.1734

RESEARCH PAPER

Designing a Recommendation Model Based on Tobit Regression, GANN-DEA and PSOGA to Evaluate Efficiency and Benchmark Efficient and Inefficient Units

Mohammad Reza Mirzaei [a], Mohammad Ali Afshar Kazemi [a,*], Abbas Toloie Eshlaghy [b]

a. Department of Industrial Management, Central Tehran Branch, Islamic Azad University, Tehran, Iran

[a] Department of Industrial Management, Central Tehran Branch, Islamic Azad University, Tehran, Iran

b. Department of Industrial Management, Science and Research Branch, Islamic Azad University, Tehran, Iran

Received: 26 July 2020, Accepted: 29 July 2020

Abstract

The main purpose of this study is to design a privatized proposal model for Tavanir regional electricity distribution and transmission companies. This proposed model is based on Tobit regression, GANN-DEA and PSOGA to evaluate the efficiency and modeling of efficient and inefficient units. This three-step process is benefited a hybrid data envelopment analysis model with a neural network optimized by a genetic algorithm to evaluate the relative efficiency of 16 Tavanir regional electricity companies. To measure the effect of environmental variables on the average efficiency of companies, two-stage data envelopment analysis and Tobit regression were used. Finally, with a hybrid model of particle mass algorithm and genetic algorithm, we have modeled for efficient and inefficient units. The average efficiency of regional electricity companies during the years 2012 to 2017 has increased from 0.8934 to 0.9147. And companies in regions 1, 2, 4, 5, 8, 12, 13, and 16 have always had the highest efficiency average (one). And the power companies in regions 10 and 11 with the average efficiency values of 0.7047 and 0.6025 had the lowest efficiency values.

Keywords:
Hybrid Algorithm of Particle Swarm Optimization with Modeling;
Tobit Regression;
Efficiency,
Combined Model of Data Eneural Network and Genetic Algorithm

Introduction

In the present era, advances and developments in management knowledge have made the existence of an evaluation system in organizations inevitable and necessary. Ranking of companies and institutions is one of the most important tools for measuring the strengths and weaknesses of organizations, so in its evaluation, its comprehensiveness should be considered in learning all angles of work. With the expansion of data dimensions and the speed with which they are updated, there is a need for models with faster information processing capabilities. Proposer models 1 by analyzing its user behavior recommends the most appropriate items (data or information) and are presented to deal with the problems caused by the large and growing volume of information. These models help the user to get closer to their target faster amid the huge amount of information, and the proposed models have been further developed by soft models of operation research.

[*] Corresponding author: dr.mafshar@gmail.com

Data Envelopment Analysis (DEA) has unique features, such as data envelopment analysis, after evaluating decision-making units (DMUs), identifies a reference point on the efficiency boundary for each inefficient unit. In such a way that the unit under evaluation reaches the reference point on the efficient boundary by decreasing the inputs, increasing the outputs or by decreasing the inputs and increasing its outputs simultaneously.

However, the reference unit for any inefficient unit is a combination of existing efficient units that in reality, does not exist objectively and is artificial and imaginary [1]. In the present study, we present a hybrid proposal model for regional electricity companies that include data envelopment analysis models that have the most repetition in the research history of regional electricity companies.

The combination of data envelopment analysis model and neural network leads to an increase in the ability to generalize and estimate nonlinear relations in data envelopment analysis models with a small number of decision-making units [5,2,3,4]. The combined model of data envelopment analysis with neural network can not model for decision units.

In the present study, this problem is solved by combining particle mass heuristic algorithm and genetic algorithm, and these algorithm models for efficient and inefficient units of the combined model of data envelopment analysis with the genetic neural network. In the last decade, the two-stage model of data envelopment analysis and Tobit regression has been able to measure the effect of environmental variables on the average of the efficiency of consecutive years [6,5]. The main purpose of this study is to design a multi-stage proposal model with a combination of the above features. This model has the ability to model using the PSOGA algorithm for efficient and inefficient units.

The efficiency of these efficient and inefficient units is due to the combination of data envelopment analysis and GANN. Also, a two-stage model of data envelopment analysis and Tobit regression has been used to measure the effect of environmental variables on the average efficiency of consecutive years. The current model has been privatized for Tavanir regional electric companies.

Research background in electricity distribution and transmission companies

Some studies have examined the efficiency and the factors affecting it for electricity distribution companies and examined the effect of environmental variables on the average efficiency of electricity distribution companies. For Sweden [7] and Turkey, there has been no evidence of differences in efficiency between public and private companies [8].

Pérez-Reyes and Tovar in 2009, comments of san hose in 2003, and Motta in 2006 confirm the positive effect of privatization on the efficiency of electricity distribution companies and reject the link between greater efficiency and private ownership for developed countries [9]. A comparison between the UK and Japanese electricity distribution companies from 1985 to 1988, which was accompanied by corrections to the data envelopment analysis model, showed that the efficiency gap between UK firms increased [10].

In the study of East and West Germany, they found that on average, East German electricity distribution companies have higher technical efficiency than their West German counterparts in various experimental models [11]. Of course, in the study of Japanese and American companies, Japanese companies are on average more efficient than American companies [12]. Using the mean of efficient models, a model was developed to evaluate and model inefficient units and has been implemented in 18 regional electricity companies [13]. Among the efficiency studies of power companies in Iran, we can mention the study of Emami Meibodi (1998). In this study, 30 electricity distribution companies in Iran have been studied. Technical and scale inefficiencies have contributed equally to inefficiencies in these companies. Most of these companies operate in the area of upward return to scale [14].

Sajadi and Omrani (2008) evaluated 38 Iranian electricity distribution companies with data envelopment analysis model and estimated the efficiency and ranking of distribution companies [15]. Fallahi and Ahmadi (2005) studied 42 Iranian distribution companies in 2002 and found that scale inefficiency was the most important cause of inefficiency of electricity distribution companies in Iran [16]. And most companies are operating in the efficiency area relative to a growing scale. Salimi and Keramati (2015) in the three-stage model of data envelopment analysis, analyzed the efficiency of 16 regional electricity companies in the years 1385 to 1392 [17]. The small number of decision-making units in the latest division of transmission and distribution companies of Tavanir Region Electricity (16 companies) as well as the effect of exogenous variables on the average efficiency values in different countries, led the authors to develop a hybrid model for Provide efficiency measurement and modeling of companies with a small number of units to be able to measure the impact of environmental variables on efficiency.

The main research question

How to measure the efficiency and modeling of power transmission and distribution companies in Tavanir regions, to design a propositional model with data envelopment analysis that can deal with the expansion of data dimensions and their speed of updating, and Include the maximum number of models in the review literature of these companies.

Materials and methods

The present research is in two ways: library and documentary. To do this, the data of the panel of 16 regional electricity companies in the period 1391 to 1396 (2012 to 2017) has been used. And how to access the information is taken from the detailed statistics of Iran's electricity industry available on the website of Tavanir and the Deputy Minister of Coordination and Financial Supervision of the Ministry of Energy.

For validating the data, we compared the available data with the data of the portal of the National Statistics Center of Iran and no significant difference was observed between the numbers and the accuracy of the data provided to the researchers of the present study was ensured. The present study can be seen from the perspective of an applied-developmental goal.

Data envelopment analysis

The purpose of data envelopment analysis is to measure and compare the relative efficiency of similar organizational units [18]. In the present study, the model proposed by Tan, based on surplus values of inputs and shortages of outputs, is used and includes input-axis, output-axis, off-axis and super-efficiency models [19]. Unlike other data envelopment analysis models that cannot use negative numbers as input or output, this model is able to accept negative variables as input or output, which is why it is called unit-independent [4]. The Tan model is as described in Eq. 1:

$$
\begin{aligned}
& Minimize\ \rho = \frac{1-\frac{1}{m}\sum_{i=1}^{m}(s_i^-/R_i^-)}{1+\frac{1}{m}\sum_{i=1}^{m}(s_i^+/R_i^+)} \\
& Subject\ \ to \\
& \sum_{j=1}^{n} x_{\mathrm{ij}}\lambda_j + s_i^- = x_{iq} \quad i = 1,2,\ldots,m \\
& \sum_{j=1}^{n} y_{\mathrm{ij}}\lambda_j - s_r^+ = y_{rq} \quad r = 1,2,\ldots,s \\
& \sum_{j=1}^{n} \lambda_j = 1 \quad j = 1,2,\ldots,n \\
& \lambda_j, s_i^-, s_r^+ \geq 0
\end{aligned}
\tag{1}
$$

In the above equation ρ is the value of efficiency, m is the number of inputs, s is the number of outputs, s_i^- is the amount of surplus input of ith , s_r^+ is the amount of output shortage of rth, λ_j is the variable corresponding to the constraints of the original model, x_{ij} input ith unit jth, y_{ij} output ith unit jth, and R is a reference set that is divided into inputs R_i^- and outputs R_r^+.

Models have also been introduced to rank efficient units, such as the crossover efficiency method or Anderson and Peterson. In these models, inefficient units are first removed from the model and the mentioned models are implemented among the efficient units.

In these models, decision-making units can take the efficiency value more than one, which that value is called super-efficiency [1]. In this research, the SBM tan super-efficiency model has been used and the tan super-efficiency model has been constructed using the unit-independent SBM model [19]. The unit-independent SBM model has the ability to model for decision-making units.

But after combining with the genetic neural network, since the definite answers give way to local answers (the amount of genetic neural network error), we will no longer be able to use the convex linear space for patterning. In the present study, we have used a hybrid algorithm of particle swarm optimization and a genetic algorithm to eliminate this defect.

Data envelopment analysis and tobit regression

Tobit regression was first proposed by Tobin [21]. The most common regression to examine the effects of variables on efficiency values is Tobit regression. Tobit regression is used when the dependent variable with the positive probability is restricted to the left or right, or both, or is censored at the end of the interval. Because the efficiency values between zeros are one, it probably has one of the angular answers [11]. In this study, we have used the Tobit Random Effects Panel used by Olatubi and Dismukes (2000) [22]. The sign of the coefficients of environmental variables indicates the direction of their effects and the standard test of hypotheses is used to evaluate the correctness of the relations.

Regarding the sampling method for Tobit regression, due to the collection of variable data "Transmission and Distribution Losses" (DEA box input variable) in the period 1391 to 1396, we have had enough of all 112 available samples. (Variable data of "transmission and distribution losses" have not been collected before 2012). In the section of variables, theoretical framework and steps of research, we will introduce the variables used in the proposed research model.

Data envelopment analysis and neural network optimized by genetic algorithm

One of the most popular neural network models is the multilayer perceptron model, which has an input layer, an outer layer, and a layer between them that is not directly connected to the input data and output results. This layer is called the hidden layer. Each unit acts like a perceptron in the hidden layer and the output layer [23]. Mean square error (MSE) and correlation coefficient (r) are used to evaluate the efficiency of neural networks. In the present study, the MSE function has been used for the genetic neural network. Table 1 shows the efficiency evaluation tools of the artificial network.

Table 1. Artificial network efficiency evaluation tools [23]

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 NN-DEA neural network design structure is a multilayer perceptron with an input layer, a hidden layer, and an output layer that has an error back-propagation algorithm.

The latent layer conversion function is a hyperbolic tangent function and the output layer conversion function is a linear function. In this network, network inputs, including the sum of inputs and outputs of each unit, are decisive. And the expected output (O) will be the efficiency of each decision-making unit [28, 27, 26, 25, 4, 5]. Of course, potential networks have also been used to evaluate the efficiency of the top Arab banks [4].

Genetic algorithm (GA) is a pervasive potential search method [29]. The purpose of applying the genetic algorithm is to optimize the weight parameter of the artificial neural network. Therefore, the objective function of the genetic algorithm is a function of the statistical results of the artificial neural network. The NN-DEA network model consists of a multilayer perceptron network that including an input layer, a middle layer and an output layer, which is named GANN-DEA after combining with a genetic algorithm.

Hybrid particle swarm optimization algorithm and genetic algorithm

Angelin proposed the first approach to combining GA concepts with particle swarm optimization, which shows that efficiency (PSO) can be improved for certain classes of problems by adding a selection process similar to what happens in evolutionary algorithms. The selection procedure is performed before the speed update and has been shown experimentally, that this algorithm improves the local search capability (PSO). Also, because half of the particles are replaced by the other half, the diversity of the solution is reduced by up to 50% per repetition. Variation can be replaced by replacing the worst particles with mutated copies of the best particles, and both algorithms are used continuously [30]. Gage (2016) used PSOGA algorithm to solve two mechanical problems (pressure tank design and welding site design), and described the method efficiently [31].

In the present study, we have used the combination of GA concepts with mass particle optimization presented by Gage (2016). The gage model includes two procedures, GA and PSO, which are described as below:

General procedure or quasi-code of genetic algorithm

1 Start: Creating a population of n chromosomes (potential answers of the problem) at random

2 Fitting: Assessing the compatibility of each chromosome (X) using the function f (x)

3 New population: Creating a new population by repeating the following steps until the new population is completed

3.1 Selection: Selection of two parent chromosomes from the population based on their degree of compatibility (greater compatibility = greater probability of selection)

3.2 Intersection: Step parent chromosomes

3.2.1 They mate randomly at a specified probability and give birth to two new offspring.

If the intersection does not occur, the offspring will be exactly the same as the two parent chromosomes.

3.3 Mutation: The resulting offspring chromosomes mutate randomly with a specified probability.

3.4 Acceptance: The created offspring are placed in the new population.

4 Replacement: The new population is used to replicate the algorithm.

5 Test: If the desired conditions are obtained, the algorithm stops and the existing population shows the desired answer.

6 [Repeat looping]: Returns to step 2.

Pseudo-code of particle swarm optimization algorithm

1 n Particle in the dimensions of the problem answer space is created randomly.

2 For all particles such as particle-ith; x the current position of the particle, V_i The velocity of the particle is randomly generated.

3 The following steps are performed until the stop conditions of the algorithm are created.

3.1 The following steps are performed for each of the particles (i = 1, ... , n).

3.2 The fitting function is calculated.

3.3 If the fitting function of the ith particle is better than y_i, then the ith particle replaces y_i and we move on to the next step.

3.4 Otherwise we go to the next step.

3.5 End of loop (per)

4. We select the best particle from the members of the current population that has the best fit function and call it $_i\hat{y}$. For each particle (i = 1, ..., n) the following steps are performed.

4.1 The velocity of each particle is calculated based on Eq. 2.

$$v_{ik+1} = w\, v_{ik} + c_1 \cdot r_1 \cdot \left(p_{ik} - x_{ik}\right) + c_2 \cdot r_2 \cdot \left(p_{gk} - x_{ik}\right) \tag{2}$$

Where w is the coefficient of inertia, r_1 and r_2 are random numbers between zero and one, c_1 and c_2 are the learning coefficients (personal and social, respectively) and p_{gk} is the state of each particle.

4.2 The position of each particle is updated based on Eq. 3.

$$x_{ik+1} = x_{ik} + v_{ik+1} \tag{3}$$

4.3 End of loop (per)

5 End of the loop (until).

The answers obtained from one iteration of the PSO algorithm are provided to the GA algorithm, and after the completion of one iteration of the GA, one iteration of the PSOGA algorithm is completed.

The output is the result of a local response and this loop continues until the stop condition is reached [31].

Variables, theoretical framework, and research implementation steps

In this section, we will describe the steps of constructing the proposed model of Tavanir regional electricity companies.

Step 1: (Define variables)

The data of 16 panels of Tavanir Regional Electricity Distribution Company in the period 1391 to 1396, in the Tobit regression model are divided into two groups before 1392 and after 1392 for the imaginary variable of privatization.

Control variables (environmental) include: 1- Privatization dummy variable (DUMPRIVATE) to control the ownership structure, 2- Ratio of underground network to total network length to control the structure of the network (UGR), 3- The ratio of home customers to all customers to control the consumer structure (CONSRESSHARE), 4- Network load coefficient (maximum asynchronous load ratio to total power consumption) (LF1), 5- Transformer capacity load factor (ratio of transformer capacity to electricity demand) (LF2) to control the intensity of grid and transformer use, respectively.

6- Circuit density as the ratio of the number of customers to the length of the network (CD2) and 7- Customer density is considered as the ratio of the number of customers to the area of coverage to control the operating environment (CD1).

In this research, the method of maximum straightness of the random panel has been used. Also, the control variables that are placed in the well-known data envelopment analysis box as the inputs and outputs of the unit-independent SBM model are: 1- Length of network lines (km), 2- Capacity of transformers (MV), 3- Number of employees (persons) and 4- Transmission and distribution losses (percentage) as input variables and 1- Number of subscribers (thousand people), 2- The energy delivered to the subscribers (million kilowatt hours) is used as output variables.

Step 2: (Solve the two-stage model of SBM data envelopment analysis independent of the classical unit and Tobit regression)

The results of unit-independent SBM model efficiency and environmental variables will be included in a two-stage model of data envelopment analysis and Tobit regression. Then, the effect and the amount of the final effect of environmental variables on the average efficiency of regional electricity distribution and transmission companies are determined. Tobit regression in the present study is as described in Eq. 4:

$$Y_{it} = \alpha + \beta_1 x_{1it} + \beta_2 x_{2\,it} + \beta_3 x_{3\,it} + \beta_4 x\,4_{it} + \beta_5 x_{5\,it} + \beta_6 x_{6\,it} + \beta_7 x_{7\,it} + u_{it} \tag{4}$$

Where,

Y_{it} symbolizes the amount of final influence of environmental variables on the average efficiency of units,

α is the symbol of intercept,

u_{it} is the symbol of the remaining error and has an independent and uniform distribution N (0, $б^2$). The degree of freedom in the hypothesis test is 8.

Step 3: (Build the first GANN-DEA network)

Using the efficiency obtained with the unit-independent SBM model, during the years 2012 to 2017 and the input and output data values of the unit-independent SBM model during these years, we construct the GANN-DEA model. We use 70% of the data at this stage as training data, 15% as test data, and the remaining 15% as generalized data.

Then, the ranking of the units of Tavanir regional electricity companies in 2017 is determined and the separation of efficient and inefficient units is done at this stage.

Step 4: (Modeling for inefficient units with PSOGA)

Using the hybrid PSOGA algorithm, we present the values of the slack variables of inputs and outputs (4 inputs and 2 outputs) to be used to improve the efficiency values in the next year. We perform this step by placing the data values of 2017 in the unit-independent SBM model. And we try to bring the efficiency value of the PSOGA hybrid algorithm closer to the efficiency of the GANN-DEA target, to obtain the values of the slack variables for the GANN-DEA, and for each inefficient unit a hybrid PSOGA algorithm is needed.

The SBM model, independent of the data envelopment analysis unit, is placed in the fitting function of this algorithm, and we try to achieve the efficiency of the GANN-DEA model by trial and error and different settings of the parameters of this complex algorithm. From the best local answers obtained, the values of the slack variables are obtained, which is the well-known modeling of the data envelopment analysis model.

Step 5: (Modeling for efficient units)

To rank and provide a model for efficient units, we use the clustering method of decision-making units introduced by Azar, Daneshvar, Khodadad Hosseini and Azizi (2012) [32] and Toloui Ashlaghi, Afshar Kazemi and Abbasi (2013) [33]. Of course, this model of data envelopment analysis by Cook and Green has also been done under the title "Power Plant Evaluation: A Hierarchical Model" for the electricity industry [34]. Toloei Ashlaghi et al., As well as Azar et al. in this method remove the units that are efficient at the level of each, from the inputs and outputs of the model, and measure the efficiency again for the remaining units to create several levels of the efficiency boundary. (Units that work in the first level will be removed in the run after the model); But to model reference units (which are efficient at the first level), we build a virtual unit; in this way, we select the lowest of all inputs for each index and the highest of all outputs for each virtual unit. A virtual unit is a unit that, although not objectified, it is possible to achieve such a unit with a set of experienced units. After adding the virtual unit to the other decision-making units, we run the model again. Units that have been efficient so far have a efficiency of less than 1, and thus a reference unit can be defined for reference units [33,32], and this step is solved by the SBM model independent of the classical unit.

Step 6: (Build the second GANN-DEA network for efficient units and virtual units)

We implement and repeat the steps of the GANN-DEA model for the efficient units and the virtual unit described in step five. With the difference that for designing GANN-DEA model for efficient units, the data of 2016 and 2017 have been used. Also, 50% of the data was used as training data, 25% of the data was used as test data and the remaining 25% was used as generalized data due to the small data row.

Step 7: (Modeling for efficient units with PSOGA algorithm)

The unit-independent SBM model described in the fifth step, with the data of 1396 we put in the hybrid PSOGA algorithm to modelling for efficient units, and again for each decision-making unit requires a separate PSOGA hybrid algorithm. The Conceptual framework of the present study is shown in Fig. 1.

Research Findings

Among efficient companies in 1396, the best efficiency belongs to 8 Region Electricity Company and the lowest super efficiency score belongs to 4 Region Electricity Company. The average efficiency scores of regional electricity companies have decreased slightly from 1392 to 1394, but from 1392 onwards, it continues its upward trend. Company 6 was recognized as efficient for the first time in 1396, while in previous years it has had higher upward growth than other regional electricity companies. The efficiency values of Tavanir Regional Distribution and Transmission Companies using unit-independent SBM model and variable efficiency compared to the obtained scale are given in Table 2.

Factors affecting the efficiency of the unit-independent SBM model with variable returns to scale, which are the results of the two-stage model of data envelopment analysis and Tobit regression are listed in Table 3.

All seven environmental variables do not have a significant effect on the average efficiency. In this study, due to the limitation in the collection of statistical data in the years before 1391 in Tavanir Regional Electricity Company, we have studied the short-term of environmental variables on group efficiency.

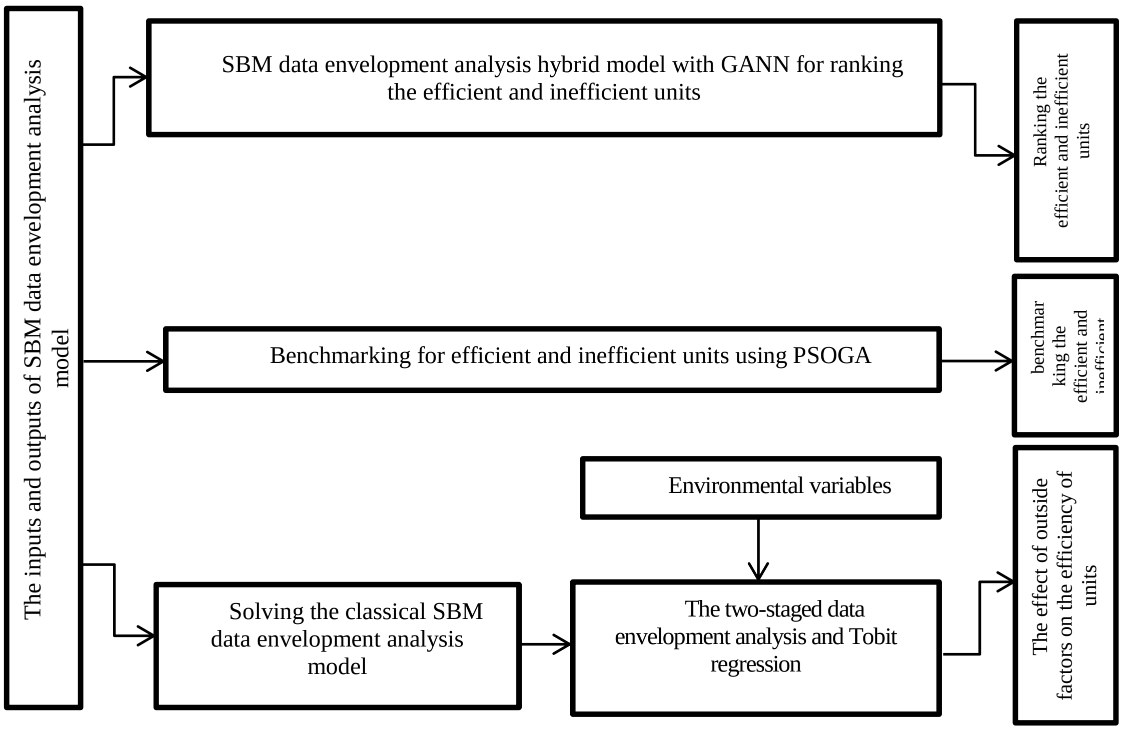

Fig. 1. Conceptual framework of the proposed model based on Tobit regression, GANN-DEA and PSOGA for evaluating the efficiency and modeling of efficient and inefficient units

Table 2. Efficiency of SBM model independent of the unit and variable efficiency compared to the scale of Tavanir Regional Electricity Distribution and Transmission Companies during the years 2012 to 2017

Regional electric company	DEA model	variables		Year						Mean SBM
		input	output	2012	2013	2014	2015	2016	2017	
Company(1)	SBM	1- Length of network lines (km), 2- Capacity of transformers (MV), 3- Number of employees (persons) and 4- Transmission and distribution losses (percentage)	1- Number of subscribers (thousand people), 2- The energy delivered to the subscribers (million kilowatt hours)	1	1	1	1	1	1	1
	Super Efficiency			1.062256	1.064183	1.028367	1.036538	1.031591	1.030269	--
Company(2)	SBM			1	1	1	1	1	1	1
	Super Efficiency			1.066985	1.149197	1.194992	1.17625	1.150807	1.141013	--
Company(3)	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	--
Company(4)	SBM			1	1	1	1	1	1	1
	Super Efficiency			1.553038	1	1	1	1	1	--
Company(5)	SBM			1	1	1	1	1	1	1
	Super Efficiency			1.356831	1.136353	1.104949	1.081925	1.09131	1.10517	--
Company(6)	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	--
Company(7)	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	--
Company(8)	SBM			1	1	1	1	1	1	1
	Super Efficiency			1.49742	1.283835	1.340388	1.377624	1.340355	1.366307	--
Company(9)	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	--
Company(10)	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	--
Company(11)	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	--
Company(12)	SBM			1	1	1	1	1	1	1
	Super Efficiency			1.258877	1.167371	1.171332	1.119649	1.177083	1.203153	--
Company(13)	SBM			1	1	1	1	1	1	1
	Super Efficiency			1.171777	1.160851	1.116177	1.099063	1.101866	1.095863	--
Company(14)	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	--
Company(15)	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	--
Company(16)	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	--

Table 3. Factors affecting the efficiency of the two-stage model of data envelopment analysis and Tobit regression

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
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.			

To implement the GANN-DEA model, we first designed the NN-DEA model according to what has been proposed in the research background and theoretical framework, and then by trial and error, we obtained the best settings for NN-DEA and the genetic algorithm that optimizes it. The best GA settings included roulette wheel for selection and with an initial population of 20, two-point coupling with a rate of 0.9, a mutation with a uniform function and a mutation rate of 0.01 and a 60-minute or 100-generation stop condition were obtained. Comparison of the efficiency values of Tavanir regional electricity distribution and transmission companies in 1396, which was obtained using the unit-independent SBM model and variable efficiency relative to scale, with the efficiency values of GANN-DEA is given in Table 4.

Table 4. Comparison of the efficiency of SBM model of data envelopment analysis unit-independent with GANN-DEA model in 2017

	Regional electric company	SBM-DEA Model efficiency in 2017	GANN-DEA efficiency in 2017
1	Company(1)	1	0.994109
2	Company(2)	1	0.9931097
3	Company(3)	0.774004	0.778652
4	Company(4)	1	1.007727
5	Company(5)	1	1.004362
6	Company(6)	1	1.007092
7	Company(4)	0.804779	0.807656
8	Company(8)	1	1.006518
9	Company(9)	1	1.002869
10	Company(10)	0.71636	0.710237
11	Company(11)	0.669565	0.676048
12	Company(12)	1	1.005444
13	Company(13)	1	0.994384
14	Company(14)	0.801438	0.805315
15	Company(15)	0.848337	0.84734
16	Company(16)	1	0.994409
--	Mean	0.91278019	0.914704

Efficiency estimation of 1396 with GANN-DEA with an average of 0.914704 and efficiency of 2017 with SBM model of data envelopment analysis with an average of 0.91278019 indicates acceptable learning of GANN-DEA network with data for six consecutive years. The efficiency of NN-DEA neural networks after optimization by the genetic algorithm is given in Table 5.

Table 5. Results of evaluation tools the function of genetic 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

A summary of the optimization of the two neural networks with the genetic algorithm is presented in Table 6.

Table 6. Summary of optimization with genetic algorithm for two NN-DEA neural networks

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 lowest mean squared error and the mean squared value of the final error for a genetic neural network, one includes all the units and the other includes the efficient units and the virtual unit are all at an acceptable level.

The best settings of PSOGA hybrid algorithm include perceptual and social component 1.5, maximum inertia coefficient 0.9 and minimum inertia coefficient 0.4, single point coupling with a rate of 0.85, the mutation rate of 0.02, mutation with uniform function, the initial population for each particle in genetic algorithm 10, minimum number of iterations in genetic algorithm 10, the highest number of replications in the genetic algorithm is 20, the lowest number of individuals in genetic algorithm 1, the reduction rate of the number of individuals affected by genetic algorithm (γ) 10 and the maximum increase rate of replication of genetic algorithm (β) 15. The efficiency values estimated by the GANN-DEA and PSOGA algorithms for inefficient and efficient units as well as the slack variable values obtained for modeling inefficient units by the PSOGA hybrid algorithm are presented in Tables 7 and 8.

Table 7. GANN-DEA and PSOGA efficiency values for inefficient units and slack variable values for modeling 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)
Company(3)	0.778652	0.774007	5.01E-05	5E-05	0.199862	255.4729	2655.433	4075.862
Company(11)	0.676048	0.671708	4.97E-05	5.01E-05	0.027146	998.0688	8475.363	6153.824
Company(10)	0.710237	0.716365	5.26E-05	5.02E-05	0.542077	89.37919	1933.935	4786.697
Company(15)	0.847341	0.844972	920.3739	0	0.873357	104.7047	1884.729	1.09E-05
Company(14)	0.805315	0.803648	5.11E-05	4.98E-05	0.458319	130.4469	1723.022	1136.608
Company(7)	0.807656	0.804789	5.04E-05	5.01E-05	0.591003	155.0925	577.446	1044.265

Table 8. GANN-DEA and PSOGA efficiency values for efficient units and slack variable values for modeling 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)
Company(1)	0.344872	0.340418	787.828	0.009154	0.945692	872.8638	6549.98	5932.302
Company(2)	0.404413	0.404158	5350.029	19122.69	0.510794	521.3247	13542.99	7569.992
Company(5)	0.381241	0.382902	4888.003	24094.56	0.508069	2167.97	35128.99	6882.992
Company(6)	0.291241	0.29752	0.040636	20010.93	1.166796	1884.068	21156.98	6458.989
Company(8)	0.926037	0.928995	7580.16	36352.14	0.191	81.53098	0.001055	0.001321
Company(9)	0.499521	0.495196	34892.85	1.519953	612.2031	1031.076	5333.992	0.439948
Company(13)	0.689944	0.682042	948.7077	2275.122	1.12	437.9996	2217.995	604.9967
Company(16)	0.742596	0.748482	7539.807	34205	0.001	192.0088	2476.154	1545.257
Company(4)	0.291444	0.2952163	0.001	0.001	0.6377	2222	35130.007	6882.999
Company(12)	0.560727	0.5673636	7067.6872	30429.006	1.1889	0.00287	5623.0078	6028.922

With the PSOGA algorithm, we try to obtain the performance value in a way that is close to the GANN-DEA performance value. The values obtained with the slack variables will be the patterns of the decision units. By decreasing the values of the input variables and increasing the values of the output variables.

For example, the regional electricity company 3 in Table 7 should reduce the length of grid lines by 4075.862 km, the capacity of transformers by 2655.433 mV, the number of employees by 255 and the transmission and distribution losses by 0.199862%. And increase the number of subscribers to 0.00005 thousand people and the energy delivered to the subscribers by 0.0000501 million kilowatt-hours to reach the efficient border.

Of course, the same interpretation applies to Table 8, but the regional power companies in this table are efficient and try to reach the virtual efficiency limit, which has the lowest values of input variables and the highest values of output variables, Reaches.

NN-DEA models have not been able to modeling the decision-making units in research to date. However, in the present study, this was made possible by the PSOGA hybrid algorithm. The addition of modeling for efficient units is another feature of the present study that was not possible to date for the NN-DEA model. Also, NN-DEA network optimization has been done for the first time in the current research, which has a faster rate of NN-DEA network convergence.

Considering that the data collection of Tavanir Regional Electricity Transmission and Distribution Companies are updated every two months or even less, the above-proposed model will be able to evaluate the new conditions with training once available from the existing 6-year data. And provide the proposed values for increasing or decreasing data on network lines length(km), transformer capacity (megavolt amperes), number of employees (person) and transmission and distribution losses(percentage), number of subscribers(thousand people), and energy delivered to subscribers (million kWh).

Tobit regression in the proposed model will be able to determine the final effect of environmental variables (exogenous) on the average of the efficiency of units and the direction of their effect (positive or negative).

Research Limitations

In the present study, due to the limitation of collecting statistical data in the years before 1391 in Tavanir Regional Electricity Company, we have been satisfied with a short-term study of environmental variables on group efficiency.

Conclusion and suggestions

The proposed model based on Tobit regression, GANN-DEA and PSOGA for evaluating the efficiency and modeling of efficient and inefficient units, in addition to having the features and capabilities of classical data envelopment analysis models, is a tool that has new capabilities, including considering the efficiency of several consecutive years to estimate the efficiency of the last year, very little impact on disturbance data, providing a model for inefficient units with considering GANN-DEA efficiency, and also modeling for efficient units considering GANN-DEA efficiency.

It can also measure the impact of environmental variables that are not under the control of the organization's senior managers on the average the efficiency of the unit-independent SBM and provide for the organization's evaluators. For future research, it is suggested that Malcom Quist, multi-stage and fuzzy data envelopment analysis models be designed similar to the multi-stage proposal model of this research. Of course, bio-efficiency models of data envelopment analysis with very similarity to the current research will be feasible. Corporate development budget index and climate indicators (such as humidity, salt content and particulate matter) are also suggested for designing propositional models similar to the present research.

References

[1] Mehregan, M .(2004). *Quantitative models in evaluating organizations performance (DEA).* Chapter 2 and 3, Tehran University Press Faculty Publishing.

[2] Ajalli, M., & Safari, H. (2011). Analysis of the Technical Efficiency of the Decision Making Units Making Use of the Synthetic Model of Performance Predictor Neural Networks, and Data Envelopment Analysis (Case Study: Gas National Co. Of Iran), *Journal of Industrial Engineering* Vol, 45, No.1.

[3] Rezaeian, J., & Asgarinezhad, A. (2014). Performance Evaluation of Mazandaran Water and Wastewater by Data Envelopment Analysis and Artificial Neural Network, *Journal of Industrial Engineering* Vol, 48, No.2.

[4] Toloie-Eshlaghy, A., Alborzi, M., & Ghafari, B. (2012). Assessment of the personnel's efficiency with Neuro/DEA combined model. *Elixir Mgmt. Arts* Vol. 43, No.1. PP.6605-6617.

[5] 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.

[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] 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.

[8] Hjalmarsson, L., & Veiderpass, A. (1992). Efficiency and ownership in Swedish electricity retail distribution. *International Applications of Productivity and Efficiency Analysis* (pp. 3-19): Springer.

[9] 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.

[10] 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.

[11] Hattori, T., Jamasb, T., & Pollitt, M. G. (2003). A comparison of UK and Japanese electricity distribution performance 1985-1998: lessons for incentive regulation.

[12] 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.

[13] 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.

[14] 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*. doi:10.1007/s00500-018-3007-2

[15] Meibodi, A. E. (1998). *Efficiency considerations in the electricity supply industry: The case of Iran* .Chapter 1, university of Surrey.

[16] Sadjadi, S., & Omrani, H. (2008). Data envelopment analysis with uncertain data: An application for Iranian electricity distribution companies. *Energy Policy,* 36(11), 4247-4254.

[17] Fallahi, M., & Ahmadi, V. (2005). Cost efficiency analysis of electricity distribution companies in Iran. *Journal of Economic Researches*, 71, 297-320.

[18] Russell, R. R. (1985). Measures of technical efficiency. *Journal of Economic Theory*, 35(1), 109-126.

[19] Charnes, A., Cooper, W., Lewin, A., & Seiford, L. (1995). *Data Envelopment Analysis: Theory, Methodology and Applications,* Kluwer Publications.

[20] Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research*, 130(3), 498-509.

[21] Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica*, 26(1), 24-36.

[22] Olatubi, W. and Dismukes D. (2000), "A Data Envelopment Analysis of the Levels and Determinants of Coal-fired Electric Power Generation Performance", *Utilities Policy*, 9, PP. 47–59.

[23] Dreyfus, G. (2005). *Neural networks: methodology and applications*: Springer Science & Business Media.

[24] Alborzi, Mahmood. (2014). *Translated of neural computing: an introduction-R Beale & T Jackson.* Chapter 4, Institute of Scientific Publications of Sharif University of Technology: Tehran.

[25] 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.

[26] Costa, Á., & Markellos, R. N. (1997). Evaluating public transport efficiency with neural network models. *Transportation Research Part C: Emerging Technologies,* 5(5), 301-312.

[27] 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.

[28] 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.

[29] Alborzi, Mahmood. (2014). *Genetic algorithm*. Chapter 3-5, Institute of Scientific Publications of Sharif University of Technology: Tehran.

[30] 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.

[31] Garg, H. (2016). A hybrid PSO-GA algorithm for constrained optimization problems. *Applied Mathematics and Computation*, 274, 292-305.

[32] Toloie Ashlaghi, A, Afshar Kazemi, M.A. And Abbasi, F. (2013) Evaluation of the performance of insurance companies based on a balanced scorecard card and data envelopment analysis technique and providing a development path for inefficient companies. *Journal of Business Management* (17), 65-82.

[33] Azar, A. Daneshvar, M. Khodadad Hosseiny, S. H. Azizi, S.(2012). Designing a multilevel performance evaluation model: A robust data envelopment analysis approach. *Enterprise Resource Management Research*, Volume II (3), 1-22.

[34] Cook WD and Green RH.(2005). Evaluating power company efficiency: a hierarchical model. *Computers & Operations Research*; 32: 813-823.