JJMIE

Volume 15, Number 3, August. 2021
ISSN 1995-6665
Pages 251 - 263

Jordan Journal of Mechanical and Industrial Engineering

Data Envelopment Analysis in the Presence of Correlated Evaluation Variables

Mohammad Bastani[a], Saeedeh Ketabi[b], Reza Maddahi*[c], Roya M.Ahari[a]

[a]*Department of Industrial Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran.*
[b]*Department of Management, University of Isfahan, Isfahan, Iran*
[c] *Department of Mathematics, Najafabad Branch, Islamic Azad University, Najafabad, Iran.*

Received March 19 2020 *Accepted June 21 2021*

Abstract

Data Envelopment Analysis (DEA) is a technique for evaluating homogeneous Decision-Making Units (DMUs) that consume similar inputs to produce similar outputs. An essential principle in this method is to identify inputs and outputs; the identified inputs (outputs) must be independent of each other. However, in the real world, there are situations where there is a correlation between two or more inputs (outputs), and then one of them should be considered in the performance evaluation. This issue can cause problems in practice. The main question, in this case, will be that" Which of these two or more correlated variables should be considered in evaluating DMUs?". In this paper, a method for determining an essential variable using a DEA model is presented. In this way, the basic models of DEA have been integrated with the 0-1 programming to achieve the above objective. The proposed method is then improved by using Centralized Data Envelopment Analysis (CDEA) model, followed by refining the performance evaluation variables. At last, the application of the proposed method has been verified for different examples. Results show that the proposed method selects the appropriate variable from among the correlated variables. Also, improving the method using a centralized approach leads to the selection of a variable that increases the total efficiency. The application and implementation of the proposed method is simple and does not have computational complexity. It also does not need experts' judgment, so it is a cost-effective way.

Keywords: Centralized Data Envelopment Analysis, Correlated variables, Performance evaluation, Filtering variables;

1. Introduction

Data Envelopment Analysis (DEA) was first introduced by Charnes et al. [1], which is a method for evaluating the relative performance of a set of Decision Making Units (DMUs) with multiple inputs and outputs. Manufacturing units, firms, hospital wards, bank branches, and individuals can be mentioned as examples of DMUs. So far, this method has been widely used for efficiency analysis in production and services as well as in the public and private sectors. Emrouznejad and Yang [2]describe an extensive list of DEA-related articles includes developing the theory and methodology and actual applications in various scenarios.

Today, Institutions that provide services or produce products are bound to perform effectively because of the intense competitive environment and limited resources. The performance of these institutions is critically linked to the correct selection of input and output variables [3].Also, the selection of variables is essential because DEA is a nonparametric approach and loses discriminatory power with increasing dimensions of production space. The reason is that when the number of inputs and outputs increases, the observations in the dataset are projected in a large number of orthogonal directions, and the Euclidean distance between observations increases. As a result, many of the observations placing on the frontier; accordingly, DEA loses its discriminatory power [4]. Thus, selecting appropriate input and output variables is one of the key and significant issues in the DEA.

One of the basic principles of the DEA is that efficiency measurement depends on the interrelationships between inputs and outputs. The nonlinear multiplier formula, efficiency, has calculated the ratio of a weighted sum of outputs to a weighted sum of inputs [5].Available DEA models assume that the inputs or outputs of DMUs are independent of each other at all times[6]. However, in terms of Pedraja-Chaparro et al. [7], one of the four factors that influence the results of DEA models is the degree of correlation between inputs and outputs. In many practical applications of DEA, there may be a correlation between two or more selected input variables or output variables. In some of the articles relating to the DEA, to reduce computations and increase efficiency discrimination between DMUs, it is described that if the correlation coefficient between each pair of input or output vectors is a strong and positive, one of the input or output vectors could be omitted[8-16].

Dyson et al. [17]stated this issue as one of the pitfalls in DEA and showed that omitting a highly correlated variable can have a significant impact on the efficiency measurement of some DMUs. In general, whether there is a correlation between inputs, between outputs or between inputs and outputs, is ignored when constructing models by analysts or experts[5].So, the selection of variables between two or more variables cannot be done by the expert and was based on subjective judgment. Therefore, the need to select a

* Corresponding author e-mail: reza.maddahi.esf@gmail.com.

variable is based on a scientific method. Based on the mentioned points, the main question is," Which of the two or more correlated variables should be considered in evaluating DMUs?".

Researchers have proposed different ways to solve these problems. Farzipoor Sean et al. [9] specify the correlation coefficient threshold beyond which the elimination of one or more input vectors has no statistically significant effect on the efficiency mean. They point out that some DEA articles have stated that if the correlation coefficient between each pair of input (output) vectors is 0.9 and above 0.9, one of the inputs (outputs) vectors can be removed.

Kao et al. [18]presented a two-stage approach combining Independent Component Analysis (ICA) and DEA to improve the discriminatory capability of DEA results. They first used ICA to extract input variables to produce independent components (ICs). Then independent components were selected as independent sources of input variables and inputted into the DEA model.

Some researchers have combined DEA and principal component analysis (PCA) to reduce the number of variables. The idea of combining DEA and PCA methodologies was first introduced by [19] and then was developed by[20]and[21]. The PCA-DEA approach is used to replace the main inputs or outputs with a set of uncorrelated components that each of them is linear combinations of the main variables. The obtained uncorrelated components are called principal components (PC) that are obtained from the eigenvalues of the covariance matrix or correlation matrix of the main variables. However, this approach can avoid inaccurate computation of efficiency for DMUs with correlated inputs and outputs but has high computational complexity. It is also often difficult to correct the interpretation of the PCs that are linear combinations of the main variables.

Dario and Simar [22] to reduce the production possibility space dimensionality, integrate highly correlated inputs and outputs into a single input and a single output by eigenvalues. [4]state that the [22] method is very similar to PCA-DEA. Their final model should have only one input and one output; therefore, it is not as public as other methods and has little practical application.

Pastor et al. [23] presented a methodology for analyzing the relevance of a variable about its contribution to efficiency. Two radial DEA formulas are considered, one with the tested variable (candidate) and the other without it. A binomial statistical test specifies that if this variable affects efficiency measure, the candidate variable is important for the production process.

Banker [24] surveys statistical tests to show the importance of input or output variables in the production process.The null hypothesis is that the tested variable does not affect the production process. Simulation studies have been performed, and the results show that these tests are better than Corrected Ordinary Least Squares (COLS) based tests.

Sirvent et al. [25] using Monte Carlo simulation for compared [23] method with [24] testsregarding various factors such as sample size, model size, the specification of returns to scale, and the type and level of inefficiency. The results show that the [23] method is more robust than the [24] tests in terms of the inefficiency distribution and the assumption of a return to scale type.

Jenkins and Anderson [26] describe a systematic statistical method that eliminates variables containing minimum information using partial correlation as a measure of information content. Information in the input or output variable is measured as the variance on the set of production units. Zero variations show that all observed production units have the same value for that variable. They indicated that the removal of highly correlated variables could have the main impact on efficiency scores; therefore, multivariate statistical method using partial correlation measures to determine the relevance of a given variable is useful.

Adler and Yazhemsky [27] demonstrated that PCA-DEA accomplishes better than [26] method, especially when analyzing relatively small datasets. They argue that comparing the methodologies shows that PCA-DEA provides a more powerful tool than the [26] method with more accurate results.

Ruggiero [28] proposed a variable selection method in which a main measure of efficiency is acquired from a set of known production variables and developed guidelines for selection. After that, efficiency regressed against a set of candidate variables. If the coefficients in the regression are statistically significant and have an appropriate sign, the variables are related to the production process. This analysis is repeated. The analysis stops when there are no other variables with appropriate and significant signs coefficients.

Fanchon [29] presented a method that specifies the optimal number of variables and the contribution of each variable to the measure of efficiency. A five-step approach defines a set of variables that best describes the output behavior and then uses the DEA repeatedly to analyze the increase in the number of efficient observations. Two regressions were performed to validate the inserted variables, one with only efficient observations and the other with efficient and inefficient observations. A statistical significance of the regression coefficients represents the validity of the variable. An example in the computer industry for separate efficient and inefficient firms is used to explain the proposed method. The method proposed by [29] is similar to the [28] method[4].

Simar and Wilson [30] propose statistical tests for measuring the relevance of inputs and outputs, as well as tests to consider potentially aggregating inputs and outputs. They use bootstrap methods to obtain the appropriate critical values for these tests. Monte Carlo experiments show the true sizes and power of the proposed tests.

Nataraja and Johnson [4] analyzed the four methods of PCA-DEA,[23],[30], [28] by Monte Carlo simulation to determine the advantages and disadvantages of each approach.

Xia and Chen [6]was used Choquet integral to consider the correlation between the input or output variables by the DEA. First, self-efficacy models based on Choquet integral were applied, which could achieve more efficiency values than existing ones. The idea was then extended to the cross-efficiency models, including the game cross-efficiency models. Based on the regret theory, the optimal DEA analysis was also examined. Various models have been developed to estimate the ranking distances of DMUs. They argue that models of interaction between inputs and outputs can achieve wider ranking intervals.

Ji et al. [31] have developed a new fuzzy DEA model using fuzzy Choquet integral as a cumulative tool to evaluate DMUs' efficiency. This model can be used to evaluate DMUs efficiency with interactive fuzzy inputs or outputs. Finally, numerical examples are used to show the proposed model performance. They state that their study has prepared a theoretical fuzzy DEA framework, but the proposed model has high computational complexity.

Li et al. [32] presented method for choosing DEA Inputs/outputs based on the Akaike's information criteria

(AIC)approach. Wagner and Shimshak [33]developed stepwise procedure to variable selection. Morita and Avkiran [34]used diagonal layout experiments, which is a statistical approach for selecting inputs and outputs in DEA. Some researchers have also used the mixed integer linear programming Approach to select the variable [35-37].The models that have been proposed for issue of correlated inputs/outputs selection in DEA are presented in Table 1.

Reviewing various studies in the literature, reveals that various studies have been conducted to address the issue of correlated inputs and outputs in DEA. The selection of inputs/outputs is a main step in DEA that is typically performed before DEA models are implemented. This issue affects the discriminatory power of DEA and the efficiency score of DMUs. Therefore, the decision to choose from correlated variables is one of the important issues in the literature. An essential aspect of this issue is the development and improvement of practical models for selecting a variable from several correlated variables. This paper proposes DEA-based models for selecting correlated variables inefficiency evaluation of the DMUs.The proposed models in this study are easy to understand for managers and decision makers compared to previous methods and do not require experts' judgment, extensive calculations and statistical analysis. These models also provide valuable management information to managers and decision makers for decision making, and are a good guide for selecting input (output) variables for them. Variables with a correlation of 0.9 and above 0.9 are considered in this paper.

The structure of this research is as follows; In Section 2, the basic DEA models and centralized data envelopment analysis (CDEA) models used in this research are reviewed. In Sections 3, 4, 5 the proposed method is presented. Then, in Sections 6 and 7, a numerical example and a case study are illustrated to describe the method. Finally, the conclusions and recommendations are discussed in Sections8 and 9.

2. Introducing the basic DEA and CDEA models

In this study, a method for determining an essential variable among the correlated variables for DEA andCDEA models is presented. Then the proposed method is used for refining the performance evaluation variables. Therefore, it is necessary to have a brief Introduction with each of these models. Thus in this section, first the basic DEA models and then CDEA model are introduced.

2.1. BCC input and output-oriented models

DEA is a well-known mathematical approach applied to assess the relative efficiency of a set of similar DMUs. This method measures the relative efficiency of each DMU based on its inputs and outputs[38]. The objective function of the DEA model tries to identify the DMUs that produce the maximum outputs with the minimum input. Although this method has been used as a useful tool in management and economics, however, it has recently found many applications in engineering problems[39-40].DEA models can be divided into two categories: input-oriented and output-oriented. The purpose of input-oriented models is to reducing the number of used resources (inputs) by keeping the output constant and output-oriented models seeking to increase the output values by keeping the number of used resources constant[40].In this subsection, the basic input and output-oriented models are reviewed[41]. These models are known as BCC models using the first names of their providers. The input and output BCC envelopment models for the evaluation are shown below in models (1) and (2), respectively. These models, which are called variable returns to scale models, arise with the basic DEA models that assume constant returns to scale and presented by[1].To solve these models, we used GAMS software.

Table 1. Proposed models correlated inputs/outputs selection in DEA

Year and author	Research method								
	ANOVA	DEA	ICA	PCA	eigenvalues	multivariate statistical	Choquet integral	Fuzzy Choquet integral	0-1 programming
Ueda and Hoshiai (1997)		✓		✓					
Jenkins and Anderson(2003)						✓			
Farzipoor Saen et al. (2005)	✓								
Dario and Simar (2007)					✓				
Kao et al. (2011)		✓	✓						
Xia and Chen (2017)							✓		
Ji et al. (2018)								✓	
This paper		✓							✓

$$
\begin{aligned}
& Min\ \ \theta \\
& s.t\ \ \sum_{j=1}^{n}\lambda_j x_{ij} \le \theta x_{ik}, i=1,2,...,m, \\
& \qquad \sum_{j=1}^{n}\lambda_j y_{rj} \ge y_{rk},\ r=1,2,...,s, \qquad (1) \\
& \qquad \sum_{j=1}^{n}\lambda_j = 1, \\
& \qquad \lambda_j \ge 0,\ j=1,2,...,n.
\end{aligned}
$$

$$
\begin{aligned}
& Max\ \ \varphi \\
& s.t\ \ \sum_{j=1}^{n}\lambda_j x_{ij} \le x_{ik},\ i=1,2,...,m, \\
& \qquad \sum_{j=1}^{n}\lambda_j y_{rj} \ge \varphi y_{rk}, r=1,2,...,s, \qquad (2) \\
& \qquad \sum_{j=1}^{n}\lambda_j = 1, \\
& \qquad \lambda_j \ge 0,\ j=1,2,...,n.
\end{aligned}
$$

In the above models it is assumed that there are n DMUs, each DMUused m inputs(x_{ij})to produce s output(y_{rj}).DMU$_k$ is the DMU to be evaluated. Also, if the objectives function value of the above models in the optimal solution equals one, under evaluation unit is efficient, and otherwise, it is called inefficient.

2.2. Centralized Data Envelopment Analysis (CDEA) Model

Since conventional DEA models set separate goals for each DMU and do not consider total input consumption and total output production, models are presented as the centralized resource allocation in which there is a centralized decision-maker who oversees all the units in operation. The main purpose of this model is to optimize total input consumption and output production.

Lozano and Villa [42] present a model called centralized input-oriented resource allocation, in which the centralized decision-maker optimizes the total input consumption. This model ensures that the total output production is not reduced. In the centralized model analysis, all units are projected on the efficient frontier as is common in conventional DEA models, but the process is done in an integrated way rather than in separate ways. In other words, in the centralized model, only one linear programming model is used to project all units on the efficient frontier, whereas in the conventional DEA models, a separate model is used for each unit, and each unit is projected separately on the efficient frontier. Another significant difference between the centralized input-oriented model is that, instead of reducing the inputs of each unit, the goal is to reduce the total input consumption of all units.

CDEA has a wide variety of applications in various sectors, such as fast-food restaurants [43], schools[44], recycling municipalities [45], and public service organizations[46].

The radial centralized input-oriented model consists of two phases. In the first phase, a proportional reduction is sought for all inputs, and the second phase is followed by a further decrease in each input and an increase in each output non-radially. These models are as follows [42]:

$$
\begin{aligned}
& \theta^* = Min\ \ \theta \\
& s.t\ \ \sum_{j=1}^{n}\sum_{l=1}^{n}\lambda_{lj} x_{ij} \le \theta\sum_{j=1}^{n} x_{ij},\ i=1,2,...,m, \\
& \qquad \sum_{j=1}^{n}\sum_{l=1}^{n}\lambda_{lj} y_{rj} \ge \sum_{j=1}^{n} y_{rj},\ r=1,...,s, \qquad (3) \\
& \qquad \sum_{l=1}^{n}\lambda_{lj} = 1, \forall_j \\
& \lambda_{lj} \ge 0,\ \theta free
\end{aligned}
$$

By solving the first phase of the model, θ* will be obtained as the optimal value of model number 3.So, the second phase of the radial centralized input-oriented model is as model (4) [42]:

$$
\begin{aligned}
& Max\ \ \sum_{i=1}^{m} s_i + \sum_{r=1}^{s} t_r \\
& s.t\ \ \sum_{j=1}^{n}\sum_{l=1}^{n}\lambda_{lj} x_{ij} = \theta^*\sum_{j=1}^{n} x_{ij} - s_i,\ i=1,...,m \\
& \qquad \sum_{j=1}^{n}\sum_{l=1}^{n}\lambda_{lj} y_{rj} = \sum_{r=1}^{s} y_{rj} + t_r, r=1,2,...,s, \qquad (4) \\
& \qquad \sum_{l=1}^{n}\lambda_{lj} = 1, \forall_j \\
& \qquad \lambda_{lj}, s_i, t_r \ge 0.
\end{aligned}
$$

In the next section, the proposed method is presented to select an essential variable from the correlated variables.

3. Proposed models in order to select among the correlated variables

In DEA, there are three types of correlation between variables: 1- Correlation between outputs variables on inputs variables. 2- Correlation between outputs variables. 3- Correlation between inputs variable. Correlation between outputs variables on inputs variable is necessary. On the other hand, the inputs (outputs)variables must be independent of each other. In this section, the proposed method is presented in order to select from correlated variables, whether input or output, on several different bases. In the first part, the basic DEA models provide the basis for the first method, and the second part uses the CDEA (second method).The proposed models have been formulated following a few steps.Fig.1 depicts the framework of proposed models.

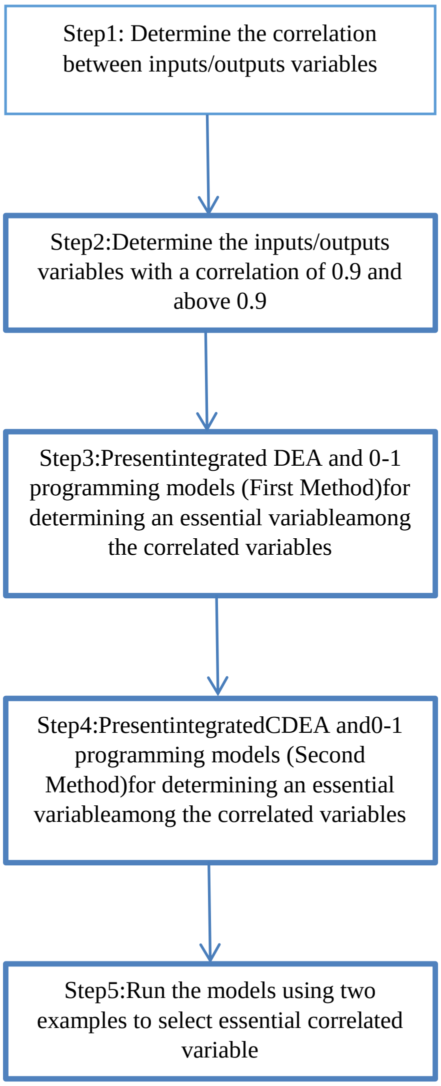

Figure 1. Framework showing the proposed models to select among the correlated variables

3.1. Selection of correlated variables Using Basic DEA Models (First Method)

According to the subjects mentioned in the previous section, there may be two general situations for dealing with correlated variables. First, a situation in which only the correlation between the inputs variables is considered and the select between them is considered. Second, a situation in which only the correlation between the outputs variables is considered and the select between them is considered. In this section, the output-oriented BCC model is used to select from the correlated input variables, and the input-oriented BCC model is used to select from the correlated output variables. The reason for this choice is that in the selection of correlated input variables, the value of the output-oriented efficiency, and in the selection of correlated output variables, the value of the input-oriented efficiency DMUs is desired. This cross selection causes the simultaneous impact of the input correlated variables on the outputs, and the output correlated variables on the inputs are examined in terms of efficiency. To simplify, we first select between two correlated variables in each section, and then this method is generalized to more correlated variables. The following paragraphs explained the proposed models mathematically.

3.1.1. Select between two correlated input variables using basic DEA Models

In this section for more simplicity and understanding of the proposed method, first, the procedure of choice between two correlated inputs variables and then the choice between two correlated output variables are raised. First, suppose that two input variables are correlated. Without reducing the generality of the proposed method, it is assumed for simplicity that the first two input variables are the correlated variables. The method proposed in this research section consisted of two stages. At first, each DMU is allowed to select one of the two above variables as input, by developing an output oriented envelopment model into a 0-1 programming model. In the second stage, a criterion for selecting that variable as the first input in the evaluation of all units is presented(Method A andMethod B). The same process can be used for the case where the two output variables are correlated; the difference is that in this case, the input-oriented envelopment model is used to select the output variable from the two correlated outputs. Consider the following 0-1 programming model:

$$
\begin{aligned}
&\varphi_k^* = Max \ \varphi \\
&s.t \ \sum_{j=1}^{n} \lambda_j x_{ij} \le x_{ik}, \ i = 2,\dots,m, \\
&\sum_{j=1}^{n} \lambda_j x_{1j} \le x_{1k} + \mu M_1, \\
&\sum_{j=1}^{n} \lambda_j x'_{1j} \le x'_{1k} + (1-\mu) M_2, \qquad (5) \\
&\sum_{j=1}^{n} \lambda_j y_{rj} \ge \varphi y_{rk}, r = 1,2,\dots,s, \\
&\sum_{j=1}^{n} \lambda_j = 1, \\
&\lambda_j \ge 0, \ \mu \in \{0,1\}
\end{aligned}
$$

In model (5), M_1 and M_2 considered as two very large positive numbers, so if the μ variable which is 0-1 variable in this model is equal to zero in the optimal solution, the constraint related to the first correlated variable x_{1j} ( )activate in the model and the constraint related to the second correlated variable (x'_{1j})was ineffective in the optimal model and solution. Otherwise, if the μ is equal one in the optimal answer, the opposite will happen.

It is imperative to note that Model (5) alone cannot be used to evaluate DMUs because the homogeneity of the DMUs in this model is violated. That is maybe, the first input genus for two different units considered different that this is inconsistent with DEA principles. In different ways, using the optimal solutions of the model (5),can decide about select the first input from two correlated variables(second stages). Two methods are suggested in this section:

Method A: The number of times that x_{1j} and x'_{1j} have been considered as inputs can be considered as a criterion for their selection as the first input of units.

Method B:It can be used to compare the sum of the efficiency of the units that selected the above variable as inputs. That is, compared the sum of φ_k^* that related to two groups to each other. Suppose that the sets D and D' respectively indicate the set of indexes corresponding to the DMUsthat have chosen x_{1j} and x'_{1j} as the first input variable, then if $\sum_{k\in D}\varphi_k^* < \sum_{k\in D'}\varphi_k^*$, x_{1j} selected as the first input and otherwise x'_{1j} is selected. In output-oriented model if φ_k^* =1, the unit under evaluation is efficient, and if φ_k^* >1 the unit is inefficient, thus less $\sum\varphi_k^*$ is the variable selection criterion.

Therefore, the proposed method in this sectionto select aninput variable in the presence of two types of correlated input variablesis briefly proposed in the following framework(see Fig. 2):

1. Solve model (5) for all DMUs.
2. Determine the suitable input using the optimum objective function values of the model (5) by Method A or Method B.

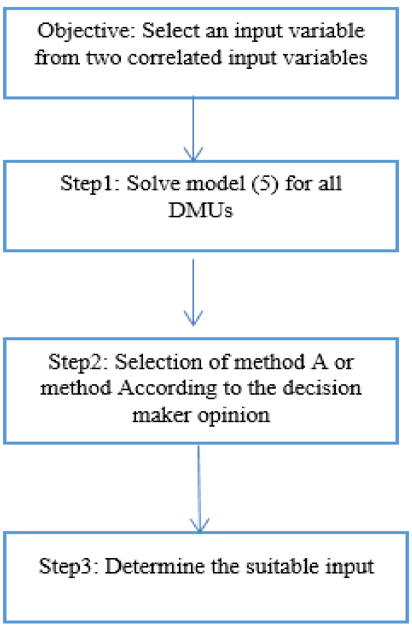

Figure 2. Framework for selecting an input variable from two correlated input variables

3.1.2. Select between two correlated output variables using basic DEA Models

In this section, the proposed model is developed to select from two correlated output variables. If the correlated variables are from the type of outputs, Model (1) can be used as the basis for suitable choosing between two correlated outputs y_{1j} and y'_{1j} . The reason for this choice is that in the selection of correlated output variables, the value of the input-oriented efficiency DMUs is desired. So model (1) changes as follows:

$$\begin{aligned}
&\theta_k^* = Min\ \theta \\
&s.t\ \sum_{j=1}^{n}\lambda_j x_{ij} \le \theta x_{ik},\ i=1,2,\dots,m, \\
&\sum_{j=1}^{n}\lambda_j y_{rj} \ge y_{rk},\ r=2,\dots,s, \\
&\sum_{j=1}^{n}\lambda_j y_{1j} \ge y_{1k} + \mu N_1, \qquad (6)\\
&\sum_{j=1}^{n}\lambda_j y'_{1j} \ge y'_{1k} + (1-\mu)N_2, \\
&\sum_{j=1}^{n}\lambda_j = 1, \\
&\lambda_j \ge 0,\ \mu\in\{0,1\}
\end{aligned}$$

In model (6), N_1 and N_2 considered as two small negatives numbers that with become zero or one of μ 0-1 variable, make the corresponding constraints active or inactive in the model.

3.1.3. Select from t correlated input variables

The process described in the previous sections can be generalized to cases where more than two variables are correlated. For this purpose, suppose among the m variables known as input variables variables are correlated, and the aim is to select only one of them as input. In this case, model (5) can be rewritten as follows:

$$\begin{aligned}
&\varphi_k^* = Max\ \varphi \\
&s.t\ \sum_{j=1}^{n}\lambda_j x_{ij} \le x_{ik},\ i\in ID, \\
&\sum_{j=1}^{n}\lambda_j x_{ij} \le x_{ik} + (1-\mu_i)M_i,\ i\in D, \\
&\sum_{j=1}^{n}\lambda_j y_{rj} \ge \varphi y_{rk}, r=1,2,\dots,s, \qquad (7)\\
&\sum_{j=1}^{n}\lambda_j = 1, \\
&\sum_{i\in D}\mu_i = 1, \\
&\lambda_j \ge 0\ \forall j,\ \mu_i\in\{0,1\}\ \forall i.
\end{aligned}$$

In model (7),D is the set of indexes related to the t correlated input variables, and ID is a set of indexes for independent variables. μ_i variables are the 0-1 variables. Due to constraint, $\sum_{i\in D}\mu_i = 1$ only one of them can be applied in the optimal solution of the model. Also M_i considered as vast positive numbers. Therefore in the optimal solution of model (7), only one constraint as related to the correlated variables is considered. As a result, only one of the correlated variables is considered in performance evaluation. Similarly, model (6) can be extended if the number of correlated outputs is more than two outputs.

4. Selection of correlated variables Using CDEA Models (Second Method)

In Section 3.1, the Basic DEA Models provide the basis for presenting a method for selecting among the correlated variables. As mentioned in section 2.2, there are situations where the CDEA model should be used. So in this section, the selection of correlated variables is based on the CDEA model. The above method is provided for both correlated inputs and correlated outputs.

4.1. Selection from two correlated input variables using the CDEA model

This section presents a 0-1 programming model based on the CDEA model and using it to determine which of the correlated inputs/outputs should be select for evaluating DMUs. Suppose $DMU_j,\ (j = 1, 2, \ldots, n)$،each DMU used m inputs (x_{ij}) to produce s output(y_{rj}).Without reducing the generality of the proposed method and in order to simplify, it is assumed that instead of the first input (x_{1j}), the (x'_{1j}) variable can also be considered for evaluating DMUs but there is correlation between these two input variables, namely(x_{1j}) and (x'_{1j}), it is also not possible to combine these two variables and consider a hybrid variable. In such cases, some DMUs maybe agree to selected (x_{1j}) as the first input and others by selected(x'_{1j}).The main question is, "Which of these two variables should be selected as the first input?".To answer this question, 0-1 programming model is integrated with the CDEA model and recommended as follows:

$$
\begin{aligned}
&\varphi^* = Max\ \ \varphi \\
&s.t\ \ \sum_{j=1}^{n}\sum_{l=1}^{n}\lambda_{lj}x_{ij} \le \sum_{j=1}^{n}x_{ij},\ i = 2,\ldots,m\,(\forall_i), \\
&\qquad \sum_{j=1}^{n}\sum_{l=1}^{n}\lambda_{lj}x_{1j} \le \sum_{j=1}^{n}x_{1j} + M_1(1-\mu), \\
&\qquad \sum_{j=1}^{n}\sum_{j=1}^{n}\lambda_{lj}x'_{1j} \le \sum_{j=1}^{n}x'_{1j} + M_2\mu, \qquad\qquad (8) \\
&\qquad \sum_{j=1}^{n}\sum_{l=1}^{n}\lambda_{lj}y_{rj} \ge \varphi\sum_{j=1}^{n}y_{rj},\ r = 1,2,\ldots,s, \\
&\qquad \sum_{l=1}^{n}\lambda_{lj} = 1, \forall_j \\
&\qquad \lambda_{lj} \ge 0,\ \ \mu \in \{0,1\}
\end{aligned}
$$

Where M_1 and M_2 considered as two vast positive numbers, and the variable μ is a 0-1 variable in the model (8).If it was $\mu^* = 1$ in the optimal solution of this model, the constraint on the first correlated variable (x_{1j}) activated in the model, and the constraint related to input (x'_{1j}) in the model becomes ineffective. Also, ifwas $\mu^* = 0$ the opposite would happen, that is mean (x'_{1j})selected as an input, and the constraint related to (x_{1j}) was disabled in the model. Therefore, firstly, in the optimal solution, one of the two correlated inputs is selected as input. Secondly, from the two correlated input variables, a variable is selected as the input that increases the total efficiency value (efficiency of the CDEA model).

4.2. Selection from two correlated output variables using the CDEA model

A similar method can be used for a state where there are two correlated output variables; Except that in this case, the input-oriented CDEA model is used as follows:

$$
\begin{aligned}
&\theta^* = Min\ \ \theta \\
&s.t\ \ \sum_{j=1}^{n}\sum_{l=1}^{n}\lambda_{lj}x_{ij} \le \theta\sum_{j=1}^{n}x_{ij},\ i = 1,2,\ldots,m, \\
&\qquad \sum_{j=1}^{n}\sum_{l=1}^{n}\lambda_{lj}y_{rj} \ge \sum_{j=1}^{n}y_{rj},\ r = 2,\ldots,s, \\
&\qquad \sum_{j=1}^{n}\lambda_j y_{1j} \ge \sum_{j=1}^{n}y_{1j} + \mu N_1, \qquad\qquad (9) \\
&\qquad \sum_{j=1}^{n}\lambda_j y'_{1j} \ge \sum_{j=1}^{n}y'_{1j} + (1-\mu)N_2, \\
&\qquad \sum_{l=1}^{n}\lambda_{lj} = 1, \forall_j \\
&\qquad \lambda_{lj} \ge 0,\ \mu \in \{0,1\}
\end{aligned}
$$

Where N_1 and N_2 considered two tiny negatives numbers and without reducing the generality of the proposed method, it is assumed that y_{1j} and y'_{1j} were two correlated output variables.

4.3. Extend the model (8) to more than two variables

We generalize this process to cases where more than two variables are correlated. Suppose of the m variables known as input variables, t variables are correlated, and the aim is to select only one of them as input. Then rewrite model (8) as follows:

$$
\begin{aligned}
&\varphi^* = Max\ \ \varphi \\
&s.t\ \ \sum_{j=1}^{n}\sum_{l=1}^{n}\lambda_{lj}x_{ij} \le \sum_{j=1}^{n}x_{ij},\ i \in ID \\
&\qquad \sum_{j=1}^{n}\sum_{l=1}^{n}\lambda_{lj}x_{ij} \le \sum_{j=1}^{n}x_{ij} + M_i(1-\mu_i), i \in D \\
&\qquad \sum_{j=1}^{n}\sum_{l=1}^{n}\lambda_{lj}y_{rj} \ge \varphi\sum_{j=1}^{n}y_{rj}, r = 1,2,\ldots,s, \qquad\qquad (10) \\
&\qquad \sum_{l=1}^{n}\lambda_{lj} = 1, \forall_j \\
&\qquad \sum_{i\in D}\mu_i = 1 \\
&\qquad \lambda_{lj} \ge 0,\ \ \mu \in \{0,1\}
\end{aligned}
$$

In model (10), D is the set of indexes related to the t correlated input variables, and ID is a set of indexes for independent variables. μ_i variables are the 0-1 variables, which, due to constraint, $\sum_{i\in D}\mu_i = 1$ only one of them can be active in the model. Also M_i considered as vast positive numbers. Therefore in the optimal solution of model (10), only one constraint as related to the correlated variables is considered. As a result, only one of the correlated variables is considered. In the same way, model (9) can be extended if the number of correlated outputs is more than two outputs.

Notice that, in the previous sections, only the way to select between correlated inputs and correlated outputs was provided. In the following, a model is presented to select a certain number of variables from the input and output variables based on the proposed model.

5. Refining the variables using the proposed model

The first step for performance evaluation in any research is to identify the variables. If the number of variables was high, the essential variables need to be refined in the next step and used in performance evaluation. There are several methods to do this. These methods include three general categories: 1- Exogenous methods Such as Delphi method and brainstorming that is done using expert opinion.2- Endogenous methods Such as Shannon entropy which uses the manner and information of the data itself to refine it.3- Combined methods that both of experts' opinion and their manner data is used. In this section, by applying the proposed method in this study, an endogenous method is proposed based on the DEA efficiency to refine performance evaluation variables. Fig.3 illustrates this method.

For example, the following studies can be mentioned in this field. Fefer et al. [47] Use the Delphi method to identify critical elements for effective and sustainable tourists. Katcher et al. [48] identify and rate home injury hazard risks for children aged 1–5 years using the modified Delphi method. Mohamadi et al. [49] using a fuzzy screening method to identify the effectual factors in the assessment of contractors, and then the weights of criteria were measured through a combination of Fuzzy AHP and fuzzy Shannon's entropy.

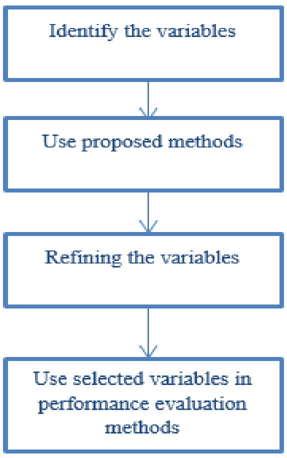

Figure 3. Proposed Framework to refine the variables

In the following, the proposed method in this study is presented to refine the variables using the proposed model. Firstly, the basic DEA Model is the basis for presenting the method, and then, the CDEA model is used. To this end, suppose between m input variables and s the output variables, k_1 variables as input and k_2 variables as the output should be used in performance evaluation. In this case, model (5) can be rewritten as follows:

$$\varphi_k^* = Max\ \varphi$$

$$s.t$$

$$\begin{aligned}
&\sum_{j=1}^{n}\lambda_j x_{ij} \le x_{ik} + M_i(1-\mu_i),\ i=1,2,\dots,m,\\
&\sum_{j=1}^{n}\lambda_j y_{rj} \ge \varphi y_{rk} + N_r(1-\xi_r),\ r=1,2,\dots,s,\\
&\sum_{j=1}^{n}\lambda_j = 1,\\
&\sum_{i=1}^{m}\mu_i = k_1\\
&\sum_{r=1}^{s}\xi_r = k_2\\
&\lambda_j \ge 0\ \forall j,\ \mu_i,\xi_r \in \{0,1\}
\end{aligned} \qquad (11)$$

In model (11), the μ_i and ξ_i variables are 0-1 variables due to constraint $\sum_{i=1}^{m}\mu_i = k_1$ only $k_1$variables, endue to constraint, $\sum_{r=1}^{s}\xi_r = k_2$ only $k_2$variablescan be applied in the optimal solution of the model. Also, M_i are vast positive numbers and N_r are tiny negative numbers. Therefore, in the optimal solution of model (11), only k_1 constraints related to the input variables constraints, and only k_2 constraints related to the output variables constraints are considered. As a result, only k_1 + k_2 variables are included in the performance evaluation.

The proposed method can be applied for the case between m input variables and s output variables, selection of k_1 variables as input and k_2 variables as output is considered in performance evaluation, presented based on the CDEA model. When using a centralized model, the advantages of this model are stated, including solving a model instead of solving (n) models, and selecting variables that increase the total efficiency value are considered in refining the variables. The proposed model is as follows:

$$\varphi^* = Max\ \varphi$$

$$s.t$$

$$\begin{aligned}
&\sum_{j=1}^{n}\sum_{l=1}^{n}\lambda_{lj} x_{ij} \le \sum_{j=1}^{n} x_{ij} + M_i(1-\mu_i), i=1,2,\dots,m\\
&\sum_{j=1}^{n}\sum_{l=1}^{n}\lambda_{lj} y_{rj} \ge \varphi\sum_{j=1}^{n} y_{rj} + N_r(1-\xi_r), r=1,2,\dots,s,\\
&\sum_{l=1}^{n}\lambda_{lj} = 1, \forall_j\\
&\sum_{i=1}^{m}\mu_i = k_1\\
&\sum_{r=1}^{s}\xi_r = k_2\\
&\lambda_{lj} \ge 0,\ \mu_i,\xi_r \in \{0,1\}
\end{aligned} \qquad (12)$$

Also, in the model (12), the μ_i and ξ_i variables are 0-1 variables, according to the constraint $\sum_{i=1}^{m} \mu_i = k_1$ only $k_1$variables, and according to the constraint, $\sum_{r=1}^{s} \xi_r = k_2$ only $k_2$variablescan be active in the model. Also, M_i are vast positive numbers and N_r are minimal negative numbers. Thus, in the optimal solution of model (12), only k_1 constraints related to the input variables constraints, and only k_2 constraints related to the output variables constraints are considered. As a result, only $k_1 + k_2$ variables are included in the performance evaluation.

6. Numerical example

In this section, a numerical example to choose from the correlated variables is provided to describe the presented models. This example is adapted from a journal paper [17].As Table 2 shows, the three inputs(I_1, I_2, and I_3) and two outputs(O_1 and O_2) are considered for each DMUs. The pairwise correlation between the first and second inputs is 1,between the first and third inputs is 0.97,and between the second and third inputs is 0.97 [17].Therefore, all three input variables are correlated and one of them must be selected. In this example, of both methods, the first method (Using basic DEA models)and the second method(Using the CDEA model)are used in order to select from the correlated input variables. Therefore, models (7) and (10) are used to select one of these three inputs as the only DMUs input. The results of applying model (7) for each of the DMUs are presented in Table 3.

Table2. Inputs and outputs values of numerical example

DMUs	O_1	O_2	I_1	I_2	I_3
Unit1	6	7	4	8	4.5
Unit2	10.5	3	6	10	5.5
Unit3	9	2	4	8	4.5
Unit4	8	5	6	10	6.5
Unit5	7	6	5	9	5.5
Unit6	2	8	5	9	4.5
Unit7	12.6	10.5	7	11	7.5
Unit8	4.2	2	2	6	1.5
Unit9	2.25	5.7	3	7	2.5

Table3. Model solution results

DMUs	Selected input	φ_k^* values
Unit1	I_3	1.08
Unit2	I_2	1.086
Unit3	I_1	1
Unit4	I_1	1.425
Unit5	I_3	1.332
Unit6	I_1	1.021
Unit7	I_1	1
Unit8	I_2	1
Unit9	I_3	1

After using the model (7),the input variable does not determine yet, and to determine the desired input from the three correlated inputs must use one of the methods A and B that was introduced in Section 3.1.1.

Method A:Suppose the decision maker is willing to use method A. According to Table 3 input 1 (I_1) 4 times (For Unit 3, Unit 4, Unit 6, Unit 7),Input 2 (I_2) 2 times(For Unit2, Unit8)and input 3 (I_3) 3 times (for Unit 1, Unit 5, Unit 9)have been selected as input. Thus, according to method A, since input 1 has the maximum number of selection, it is selected from the three inputs as the main input for performance evaluation.

Method B:Suppose sets D, D' and D" represent the set of indexes for the DMUs that select I_1, I_2, and I_3 as the first input variable, respectively. According to the values of the third column of Table 3, $\sum_{k\in D'} \varphi_k^* < \sum_{k\in D''} \varphi_k^* < \sum_{k\in D} \varphi_k^*$ that is mean 2.086<3.412<4.446.So I_2 is selected.

In the following, the second method (using the proposed method based on the CDEA model) apply for select one variable as the main input from the three correlated input variables in this example. Therefore, model (10) should be used. After solving this model for the above example data, given that $\mu_1 = 1$ in the optimal solution, as a result,$I_1$is selected as the main input for performance evaluation. The results of using the first method (A and B), and the second method are presented in Table 4.Although the method (A) and the second method have the same answer, for method (A) n models must be solved, but for the second method, it is enough to solve a model.

Table4. Results of the first method (A and B) and the second method

methods	method (A)	method (B)	second method
Selected input	I_1	I_2	I_1

7. Case study

In this section, the proposed methods are examined for a set of real data extracted from the paper[50].The information to compare the efficiency performance of the 14 bank branches is presented in Table 5. The outputs used in their study include 17 bank transactions: loan applications, new passbook loans, life insurance sales, new accounts, closed accounts, travelers checks sold, bonds sold, bonds redeemed, deposits, withdrawals, checks cashed, treasury checks issued, B5 checks, loan payments, passbook loan payments, life insurance payments, mortgage payments. In the following management, reduce the number of outputs based on the complexity and resources required was regarded to be approximately the same. Management proposed reducing the 17 transactions to four transaction types (Output 1, Output 2, Output 3, and Output 4).Finally, this comparison is based on three inputs and four outputs. Input 1(I_1): rent (thousands of dollars),Input 2 (I_2): full time equivalent personnel per branch, Input 3 (I_3): supplies (thousands of dollars) and the output 1(O_1) includes: loan applications, new pass-book loans, life insurance sales, Output 2 (O_2): new accounts, closed accounts, Output 3(O_3): travelers checks sold, bonds sold, bonds redeemed, Output 4(O_4): deposits, withdrawals, checks cashed, treasury checks issued, B5 checks, loan payments, passbook loan payments, life insurance payments, mortgage payments.

Table5. 14 Bank branches information

DMU	Input 1	Input 2	Input 3	Output 1	Output 2	Output 3	Output 4
1	140,000	42,900	87,500	484,000	4,139,100	59,860	2,951,430
2	48,800	17,400	37,900	384,000	1,685,500	139,780	3,336,860
3	36,600	14,200	29,800	209,000	1058,900	65,720	3,570,050
4	47,100	9,300	26,800	157,000	879,400	27,340	2,081,350
5	32,600	4,600	19,600	46,000	370,900	18,920	1,069,100
6	50,800	8,300	18,900	272,000	667,400	34,750	2,660,040
7	40,800	7,500	20,400	53,000	465,700	20,240	1,800,250
8	31,900	9,200	21,400	250,000	642,700	43,280	2,296,740
9	36,400	76,000	21,000	407,000	647,700	32,360	1,981,930
10	25,700	7,900	19,000	72,000	402,500	19,930	2,284,910
11	44,500	8,700	21,700	105,000	482,400	49,320	2,245,160
12	42,300	8,900	25,800	94,000	511,000	26,950	2,303,000
13	40,600	5,500	19,400	84,000	287,400	34,940	1,141,750
14	76,100	11,900	32,800	199,000	694,600	67,160	3,338,390

The correlation matrix of the inputs is as Table6[51]:

Table 6. 14 Bankbranches input variables correlation

	Input 1	Input 2	Input 3
Input 1	1	0.31977	0.93579
Input 2	0.31977	1	0.3679
Input 3	0.93579	0.3679	1

Table 6 shows the correlation values between the input variables. Since the correlation between the first and third inputs (rent and supplies) is greater than 0.9,Therefore, there is a correlation between the rent and supplies variables. Thus, one of these two inputs must be selected for use in the DEA model. Since model (5) is to choice between the two inputs correlated variables, so we use that model. Of course, as mentioned, conventional DEA models must be run separately for each DMU.As a result, since there are 14 bank branches (DMUs) in our case study, the model runs 14 times. The results of selecting the input variable for each DMU are shown in Table (7).

Then the manager or decision maker has to decide on the appropriate variable to choose from among the rent and supplies variables. For this purpose, he/she can use methods A or B. The choice of each of these two methods depends on the approach of the manager or decision maker. In method A, the variable frequency is selected. But in method B,The criterion is the sum of the efficiency resulting from the variable selection. According to the results of Table (7), if we use method A,rent variable 9 times (For Unit14, Unit13, Unit11, Unit9, Unit3, Unit4, Unit6, Unit1, Unit2) and supplies variable 5 times (for Unit8, Unit5, Unit7, Unit10, Unit12) have been selected as input. Therefore, because the frequency of choice of rent variable is higher than the supplies variable, this variable is selected. If the manager or decision maker wants to use method B to select the variable, the $\sum \varphi_k^*$ of each variable must be calculated. This amount according to the results of Table (7) for the rent variable is 9 and for the supplies variable is 5.672. In output-oriented model if φ_k^*=1, the unit under evaluation is efficient, and if φ_k^*>1 the unit is inefficient, thus less $\sum \varphi_k^*$ is the variable selection criterion. Therefore, the supplies variable is selected.

In the following, if wanted by the second method (using the proposed method based on the CDEA model),select one of the above-correlated input variables using the model (8).After solving this mode for case study data since in the optimal solution $\mu = 1$, as a result,I_1(rent variable) is selected. The results of using the first method (A and B), and the second method are presented in Table 8.

Table 7. Model solution results

DMUs	Selected input	φ_k^* values
DMU 1	I_1	1
DMU 2	I_1	1
DMU 3	I_1	1
DMU 4	I_1	1
DMU 5	I_3	1
DMU 6	I_1	1
DMU 7	I_3	1.287
DMU 8	I_3	1.092
DMU 9	I_1	1
DMU 10	I_3	1.089
DMU 11	I_1	1
DMU 12	I_3	1.204
DMU 13	I_1	1
DMU 14	I_1	1

Table 8. Results of the first method (A and B) and the second method

methods	method (A)	method (B)	second method
Selected input	I_1	I_3	I_1

8. Results and Discussion

In this article two general methods for determining an essential variable among the correlated variables have been used. The first method used basic DEA models.For more simplicity and understanding of the proposed method, we first examined the special case where there are two correlated inputs/outputs (section 3.1.1 and section 3.1.2).Then we expanded to more complete case in section 3.1.3.In the first method, the criteria for selecting the essential variable from the correlated variables are two general criteria: 1- A variable that makes more DMUs efficient (Method A). 2- A variable that has the greatest impact on the efficiency of each DMUs (Method B).

The second method used CDEA models. Also in this, for more simplicity and understanding of the second method, we first examined two correlated inputs/outputs selection (section 4.1 and section 4.2). Then we expanded the model to more than two correlated variables in section 4.3. In the second method, the criteria for selecting the essential variable from the correlated variables are the use of total efficiency value of DMUs. That is, a variable is selected that increases the total efficiency of DMUs. The manager or decision maker can also use centralized method based on the case study case. Using this method compared to the previous model has the following advantages:

1. In this method, instead of solving n DEA model, one DEA model is solved. Therefore, the amount of calculations is significantly reduced.

2. In the centralized method, a variable is selected that increases the total efficiency.

It is necessary to mention thatin all two general methods, the criterion for selecting the correlated variable is efficiency.

In the following, the application of the proposed methods has been verified for two examples. Finally, a method was proposed to refine the variables using the proposed models. First, the basic DEA model was the basis for presenting the method and then the CDEA model was used (section 5).

9. Conclusions

Understanding the problems of using standard models and providing solutions to solve these problems is essential for useful research.DEA is a technique that is spreading rapidly. One of the major advantages of DEA is its ability to change information on different inputs and outputs into a single measure of efficiency[52].So, the utility of the DEA model (as a standard and applicable model) depends on its ability to compute the DMUs' relative efficiency using different inputs and outputs. It is a reasonable approach to remove variables that, by their correlations, have the least additional information for the DEA, particularly when we hope that fewer variables lead to better categorization of DMUs[26]. Therefore, determining an essential variable among the correlated variables is very important. Hence in this paper, a method using basic DEA models is presented to solve this problem. In this way, the basic DEA models were integrated with the 0-1 programming model to achieve the above objective. For this purpose, first an algorithm for selecting between two correlated input variables and then for selecting from between two correlated output variables using basic DEA models were presented. Also,the described process was generalized to cases where more than two correlated input/output variables. Then this method was developed to select between two correlated input/output variables and then to select from more than two correlated variables using the CDEA model. Thus, in the optimal solution of the developed CDEA model, from the two correlated inputs variable, a variable is selected as the input that increases the total efficiency value (efficiency derived from the CDEA model).Finally, using the proposed method in this study, a method based on the efficiency of DEA was presented to refine performance evaluation variables. For this purpose, first the basic DEA Model was the basis for presenting the method, and then, the CDEA model was used. The results showed that the proposed method has several advantages. It is easy to use, and implementation, and there is no computational complexity. Computational complexity means the existence of big data and the use of supercomputers to solve models, while the proposed method does not require advanced software equipment to solve the models.Also, the presented method does not have the complexity of statistical topics. It also does not need experts' judgment, so it is a cost-effective way. Another advantage of the proposed method is that it can be easily used for other types of input-oriented or output-oriented DEA models with variable returns to scale or constant returns to scale.

Researchers can use the proposed method in various aspects of future research. Considering different datasets, such as ambiguous, negative, or fuzzy data, can be interesting research topics. Other DEA models with correlated variables can also be studied. The presented method can be used in other DEA application areas such as manufacturing, universities, agriculture, and other organizations. For future research, we suggest a framework considering in the modeling and selection variables one step with the experts. In this case, is possible to conduct a comparison with the method (free of judgment) and was suggested by the experts. Instead of the criteria for selecting the correlated variable in the first method (Method A and Method B), other criteria can be considered as future research and the results can be compared with this research.

References

[1] Charnes, A., Cooper, W.W. and Rhodes, E, "Measuring the efficiency of decision making units". European Journal of Operational Research, Vol. 2 No. 6, 1978, 429–444.

[2] Emrouznejad, A. and Yang, G. liang, "A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016". Socio-Economic Planning Sciences, Elsevier Ltd, Vol. 61, 2018, 4–8.

[3] Alpay, O. and Akturk, E, "Copula approach to select input / output variables for DEA".An International Journal of Optimization and Control: Theories & Applications, Vol. 7 No. 1, 2017, 28–34.

[4] Nataraja, N.R. and Johnson, A.L, "Guidelines for using variable selection techniques in data envelopment analysis". European Journal of Operational Research, Elsevier B.V., Vol. 215 No. 3, 2011, 662–669.

[5] López, F.J., Ho, J.C. and Ruiz-Torres, A.J., "A computational analysis of the impact of correlation and data translation on DEA efficiency scores". Journal of Industrial and Production Engineering, Vol. 33 No. 3, 2016, 192–204.

[6] Xia, M. and Chen, J.X., "Data Envelopment Analysis Based on Choquet Integral". International Journal of Intelligent Systems, Vol. 32 No. 12, 2017, 1312–1331.

[7] Pedraja-Chaparro, F., Salinas-Jiménez, J. and Smith, P., "On the quality of the data envelopment analysis model". Journal of the Operational Research Society, Vol. 50 No. 6, 1999, 636–644.

[8] Dong, F., Mitchell, P.D. and Colquhoun, J., "Measuring farm sustainability using data envelope analysis with principal components: The case of Wisconsin cranberry". Journal of Environmental Management, Elsevier Ltd, Vol. 147, 2015, 175–183.

[9] Farzipoor Sean, R., Memariani, A. and Lotfi, F.H., "The effect of correlation coefficient among multiple input vectors on the efficiency mean in data envelopment analysis". Applied Mathematics and Computation, Vol. 162 No. 2, 2005, 503–521.

[10] Lewin, A.Y., Morey, R.C. and Cook, T.J., "Evaluating the administrative efficiency of courts". Omega, Vol. 10 No. 4, 1982, 401–411.

[11] Seifert, L.M. and Zhu, J., "Identifying excesses and deficits in Chinese industrial productivity (1953-1990): A weighted data envelopment analysis approach". Omega, Vol. 26 No. 2, 1998, 279–296.

[12] Smith, P., "Model misspecification in Data Envelopment Analysis". Annals ofOperations Research, Vol. 73, 1997, 233–252.

[13] Sueyoshi, T., "Estimation of stochastic frontier cost function using data envelopment analysis: An application to the at&t divestiture". Journal of theOperational Research Society, Vol. 42 No. 6, 1991, 463–477.

[14] Thanassoulis, E., "DEA and its use in the regulation of water companies". European Journal of Operational Research, Vol. 127 No. 1, 2000, 1–13.

[15] Charnes, A., Cooper, W.W. and Sueyoshi, T., "A Goal Programming/Constrained Regression Review of the Bell System Breakup". Management Science, Vol. 34 No. 1, 1988, 1–26.

[16] Kao, C., Chang, P. and Hwang, S.N., "Data Envelopment Analysis in Measuring the Efficiency of Forest Management". Journal of Environmental Management, Vol. 38 No. 1, 1993, 73–83.

[17] Dyson, R.G., Allen, R., Camanho, A.S., Podinovski, V. V., Sarrico, C.S. and Shale, E.A., "Pitfalls and protocols in DEA". European Journal of Operational Research, Vol. 132 No. 2, 2001, 245–259.

[18] Kao, L.J., Lu, C.J. and Chiu, C.C., "Efficiency measurement using independent component analysis and data envelopment analysis". European Journal of OperationalResearch, Elsevier B.V., Vol. 210 No. 2, 2011, 310–317.

[19] Ueda, T. and Hoshiai, Y., "Application of principal component analysis for parsimonious summarization of DEA inputs and/or outputs". Journal of the OperationsResearch Society of Japan, Vol. 40 No. 4, 1997, 477–478.

[20] Adler, N. and Golany, B., "Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to Western Europe". European Journal of Operational Research, Vol. 132 No. 2, 2001, 260–273.

[21] Adler, N. and Golany, B., "Including principal component weights to improve discrimination in data envelopment analysis". Journal of the Operational ResearchSociety, Vol. 53 No. 9, 2002, 985–991.

[22] Dario, C., Simar, L., Advanced Robust and Nonparametric Methods in Efficiency Analysis,Springer Verlag, US, 2007, p. XXII -248.

[23] Pastor, J.T., Ruiz, J.L. and Sirvent, I., "A Statistical Test for Nested Radial Dea Models". Operations Research, Vol. 50 No. 4, 2002, 728–735.

[24] Banker, R.D., "Hypothesis tests using data envelopment analysis". Journal of Productivity Analysis, Vol. 7 No. 2–3, 1996, 139–159.

[25] Sirvent, I., Ruiz, J.L., Borrás, F. and Pastors, J.T., "A Monte Carlo evaluation of several tests for the selection of variables in DEA models". International Journal ofInformation Technology and Decision Making, Vol. 4 No. 3, 2005, 325–343.

[26] Jenkins L. and Anderson M., "A multivariate statistical approach to reducing the number of variables in data envelopment analysis". European Journal of OperationalResearch, Vol. 147, 2003, 51–61.

[27] Adler, N. and Yazhemsky, E., "Improving discrimination in data envelopment analysis: PCA-DEA or variable reduction". European Journal of Operational Research, Elsevier B.V., Vol. 202 No. 1, 2010, 273–284.

[28] Ruggiero, J., "Impact assessment of input omission on DEA". International Journalof Information Technology and Decision Making, Vol. 4 No. 3, 2005, 359–368.

[29] Fanchon, P., "Variable selection for dynamic measures of efficiency in the computer industry". International Advances in Economic Research, Vol. 9 No. 3, 2003, 175–188.

[30] Simar, L. and Wilson, P.W., "Testing restrictions in nonparametric efficiency models". Communications in Statistics Part B: Simulation and Computation, Vol. 30 No. 1, 2001, 159–184.

[31] Ji, A.B., Chen, H., Qiao, Y. and Pang, J., "Data envelopment analysis with interactive fuzzy variables". Journal of the Operational Research Society, Taylor & Francis, Vol. 0 No. 0, 2018, 1–9.

[32]Li Y , Shi X , Yang M , Liang L . "Variable selection in data envelopment analysis via akaike's information criteria". Annal Oper Res, 253(1) 2017,453–76.

[33]Wagner J , Shimshak D . "Stepwise selection of variables in data envelopment analysis: procedures and managerial perspectives". Eur J Oper Res, 180(1) ,2007,57–67.

[34]Morita, H. and N. K. Avkiran.,"Selecting inputs and outputs in data envelopment analysis by designing statistical experiments". Journal of the Operations Research Society of Japan,52 (2), 2009,163-173.

[35]Limleamthong, P. and G. Guillen-Gosalbez.,"Mixed-integer programming approach for dimensionality reduction in data envelopment analysis: Application to the sustainability assessment of technologies and solvents". Industrial & Engineering Chemistry Research, 57, 2018, 9866–9878.

[36]Ben´ıtez-Pe˜na, S., P. Bogetoft, and D. R. Morales.,"Feature selection in data envelopment analysis: A mathematical optimization approach". Omega. Volume 96, 2020,102068.

[37]Peyrache, A., Rose, C., & Sicilia, G.,"Variable selection in Data Envelopment Analysis". European Journal of Operational Research. Volume 282, Issue 2, 2020, 644-659.

[38]Bastani M, Ketabi S, Maddahi R, Ahari RM. "Efficient Resource Allocation to Social Security Organization Hospitals in Iran by Using Centralized Data Envelopment Analysis". J Health Man & Info, 7(1), 2020, 59-67.

[39]Al-Refaie, A., Najdawi, R.,Sy, E.,"Using DEA window analysis to measure efficiencies of blowing machines in plastics industry". Jordan Journal of Mechanical and Industrial Engineering. Volume 10 Number 1, 2016, 27- 38.

[40]Mahmoud, I.A., Fahmyaly, M., Mohib, A., Afefy, I.H., "Integration of Benchmarking with Overall Equipment Cost Loss for Industrial Process Improvement". Jordan Journal of Mechanical and Industrial Engineering. 13. 2019, 49 - 59.

[41] Banker, R.D., Charnes, A. and Cooper, W.W.,"Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis".Management Science, Vol. 30 No. 9, 1984, 1078–1092.

[42] Lozano, S. and Villa, G., "Centralized resource allocation using data envelopment analysis". Journal of Productivity Analysis, Vol. 22 No. 1–2, 2004, 143–161.

[43] Giménez-García, V.M., Martínez-Parra, J.L. and Buffa, F.P., "Improving resource utilization in multi-unit networked organizations: The case of a Spanish restaurant chain". Tourism Management, Vol. 28 No. 1, 2007, 262–270.

[44] Mar-Molinero, C., Prior, D., Segovia, M.M. and Portillo, F., "On centralized resource utilization and its reallocation by using DEA". Annals of Operations Research, Vol. 221 No. 1, 2014, 273–283.

[45] Lozano, S., Villa, G. and Adenso-Díaz, B., "Centralised target setting for regional recycling operations using DEA". Omega, Vol. 32 No. 2, 2004, 101–110.

[46] Asmild, M., Paradi, J.C. and Pastor, J.T., "Centralized resource allocation BCC models". Omega, Vol. 37 No. 1, 2009, 40–49.

[47]Fefer, J.P., De-Urioste Stone, S., Daigle, J. and Silka, L., "Using the Delphi Technique to Identify Key Elements for Effective and Sustainable Visitor Use Planning Frameworks". SAGE Open, Vol. 6 No. 2, 2016, 1–16.

[48]Katcher, M. L., Meister, A. N., Sorkness, C. A., Staresinic, A. G., Pierce, S. E., Goodman, B. M., Peterson, N. M., Hatfield, P. M., Schirmer, J. A., "Use of the modified Delphi technique to identify and rate home injury hazard risks and prevention methods for young children". Injury prevention : journal of the International Society for Child and Adolescent Injury Prevention, Vol. 12 No. 3, 2006, 189–194.

[49]Mohamadi, S., Ebrahimi, A. and Alimohammadlou, M., "An application of fuzzy screening, fuzzy AHP and fuzzy Shannon's entropy on identification and prioritisation of effective factors in assessment of contractors in Fars Electric Power Distribution Company, Iran". International Journal of Procurement Management, Vol. 10 No. 2, 2017, 194–226.

[50] Sherman, H.D. and Gold, F., "Evaluation with Data Envelopment Analysis".Vol. 9, 1985, 297–315.

[51] Ji, A., Liu, H., Qiu, H.J. and Lin, H., "Data envelopment analysis with interactive variables". Management Decision, Vol. 53 No. 10, 2015, 2390–2406.

[52] Gilsa, C.V., Lacerda, D.P., Camargo, L.F.R., Souza, I.G. and Cassel, R.A. "Longitudinal evaluation of efficiency in a petrochemical company". Benchmarking: An International Journal, Vol. 24 No. 7,2017, 1786-1813.