A heuristic method for choosing 'virtual best' DMUs to enhance discrimination power of augmented DEA model

Marzieh Sadat Rezaei¹, Abdorrahman Haeri²*

1. School of Industrial Engineering, Iran University of Science & Technology, Tehran, Iran, E-mail: m.rezaee90@yahoo.com.
2. Assistant Professor of School of Industrial Engineering, Iran University of Science & Technology, Tehran, Iran, Email: ahaeri@iust.ac.ir.

*Corresponding author: Phone: +98-21-73225019, Fax: +98-21-73225098.

Abstract

Despite its intrinsic advantageous features as a tool for increasing discrimination power of the basic DEA (data envelopment analysis) model, augmented DEA has two main drawbacks including the presence of unrealistic efficiency scores and the presence of great distance between its efficiency scores and scores obtained by primary model. In this regard, this paper extends a heuristic method for dealing with both issues and improving the power of augmented DEA model in performance evaluation. Since different virtual DMUs lead to different results for ranking, the hierarchical clustering algorithm is applied in this study to select the best virtual DMUs in order to reduce the possibility of having inappropriate efficiency scores. Finally, to demonstrate the superiority of the proposed approach over previous approaches in literature, two numerical examples are provided.

Keywords: Data envelopment analysis; Augmented DEA; Performance evaluation, Hierarchical clustering; Virtual DMUs

1. Introduction

Nowadays, the necessity of an appropriate performance evaluation tool due to the today's competitive markets is perceived more than ever to achieve the goals of the organizations. The measure of efficiency as a capability to do work well, successfully and without waste, is necessary in evaluating the performance of resources used in an organization. This evaluating also has a key role in providing periodic feedback to managers which can reveal the scope for improvement and lead to the success of an organization [1].

The DEA method has several applications consists of ranking efficiency scores (Rezaie et al. [2]; Haeri et al [3]) and providing improvement plan (Haeri and Rezaie [4]; Haeri and Ghousi [5]; Rezaee et al [6]). Over the recent years, DEA as a powerful technique for performance evaluation has drawn great attention, while beside its popularity, the basic type of DEA model has some drawbacks. For instance, it uses variable weights which are derived directly from the data and the weights are chosen in a manner that gives the most favorable set of weights to each DMU. In other words, the DMU is permitted to utilize its best multiplier weights to increase its efficiency. As a result, more than one efficient unit is usually obtained which cannot be discriminated. Regarding such a problem, several forms of DEA model have been developed in order to enhance the discrimination power and ability of the basic DEA model to rank the efficient DMUs. Appalla [7] extended an augmented DEA model to increase discrimination power of basic DEA by introducing a new virtual DMU generated by
choosing the best values of each factor from the existing DMU base. However, the proposed augmented DEA by Appalla [7] has the following two main drawbacks.

- The presence of unrealistic efficiency scores
- The presence of great distance between the efficiency scores obtained by the basic DEA and the scores obtained through augmented DEA

The above two notable issues indicate that presenting appropriate efficiency scores is a critical problem that must be addressed in performance evaluation; hence, many researches have been conducted to compute real and appropriate scores for efficiency.

Another challenge with DEA model is to find a way to incorporate judgment into it. Golany and Roll [8] proposed the inserting of engineering such as standards (i.e., virtual DMUs) into the evaluation with the aim of having a set of standard data to compare DMUs through simply expanding the reference set. The incorporation of standards is very worthwhile since it enhances the potential that DMUs before were efficient will be indicated to be inefficient. In addition, Golany and Roll [8] expressed that creating standards is a challenge, however, one of the best attribute of DEA is that it identifies that ‘excellence can be acquired from different combinations of inputs and outputs’.

Data mining is regarded as a new area where organizations can obtain the competitive advantage. By the use of data mining process, useful information can be extracted from large databases which is important and vital in today's business and marketing since this extracted information can assist decision makers in making better and more intelligent decisions ([9], [10]). Data mining consists of a number of common classes of tasks in which cluster analysis is considered as a main task of that as well as a common method for data analysis which seeks to classify a set of elements, so that elements in the same cluster are more similar to each other than to those in other clusters [11].

According to the enumerated matters, the present work proposes a new heuristic method to enhance discrimination power of augmented DEA model by creating a set of potential virtual DMUs based on employing a data mining approach of DMU’s data. To do so, cluster analysis as a popular data mining approach is applied. By using cluster analysis, the analysis of data set improves and with the help of obtained information through clustering, virtual DMUs can be created in a more coherent manner compared to the previous approaches in the literature. This prevents the problem of having inappropriate efficiency scores. The main questions that this study is going to answer are listed as follows:

1. What values should be consider for input and output factors of virtual DMUs?
2. How many virtual DMUs should be added?

The remaining structure of this paper is as follows. After the introduction, section 2 expresses a review of the literature regarding augmented DEA. Section 3 states the DEA model along with the application of Augmented DEA. The extended approach is introduced in Section 4. In Section 5, a case study and a numerical example are presented to show how the extended
heuristic method performs compared to the previous approaches in literature. Section 6 is devoted to results and discussion. Finally section 7 includes conclusion of this research.

2. Literature review

Over the past years, various models and techniques have been developed to enhance the discrimination power of DEA model by using augmented DEA. Shokr et al. [12] developed an augmented common weight data envelopment analysis model (ACWDEA) for material selection in high-tech industries where both qualitative and quantitative criteria are involved. Their proposed model has more discrimination power and it is able to produce full ranking vectors. In addition, it is capable to determine weight of qualitative and quantitative criteria precisely. Shen et al. [13] introduced an augmented DEA model to eliminate the poor discriminatory power of the basic DEA considering the distances to both the efficient and anti-efficient frontiers. In the provided model, standard model of DEA and inverted model of DEA as two linear programming models were solved concurrently to present further information regarding frontiers. Wu et al. [14] proposed a new approach for evaluation and selection of suppliers named augmented imprecise DEA (AIDEA) which was able to rank supplier effectively in the presence of imprecise data such as ordinal data and interval data. Additionally, AIDEA model was capable to increase discriminatory power however, it gave the weight flexibility to the DMUs and allowed inappropriate DMUs to become false positive candidates. Ghasemi et al. [15] extended an augmented form of DEA model to improve discriminating capability in DEA considering deviation variable framework of the variable returns to scale method. Wu and Blackhurst [16] provided an augmented DEA model for evaluating and selecting of suppliers. The provided approach included virtual DMUs produced by choosing the best value of one factor and the average values of the remaining factors from the inputs and outputs, and weight constraints proposed to decrease the possibility of getting unsuitable input and output factor weights. Hou et al. [17] proposed a novel model to improve DMUs’ evaluation through introducing two virtual DMUs namely ideal point and anti-ideal point DMUs. The former was based on efficiency while the latter was based on fairness. Noorizadeh et al. [18] presented an augmented DEA model by considering both non-discretionary inputs and dual-role factors. The proposed model was applied only for choosing a DMU from efficient DMUs. Kianfar et al. [19] combined clustering with AHP to eliminate the poor discrimination capability of the DEA in prioritizing efficient DMUs. Hatefi and Razmi [20] provided an augmented imprecise DEA to evaluate a set of suppliers. Mahdiloo et al. [21] provided an approach to rank suppliers in the presence of both undesirable outputs and dual-role factors. The proposed model was run with a virtual DMU which improved the discrimination power of primary model.

Some of the previous research used augmented DEA as a performance evaluation tool for different DMUs. For instance, Rezaie et al. [2] used augmented DEA to improve discriminatory power of DEA to can rank organizational resources properly. They introduced a virtual DMU high outputs and low inputs. Haeri [22] applied augmented DEA to boost discriminatory power of DEA to rank Photovoltaic Solar cells technologies correctly. Based on this technique, two virtual DMUs which are assumed and have the best and the worst efficiency scores are added to the existing DMU base. Rezaee et al. [23] employed an
augmented DEA model to increase discrimination capability of DEA (CCR) model in evaluating the automotive vendors' performance. In the provided model, they introduced two virtual DMUs that have the highest and the lowest efficiency scores and added them to the basic model. Geng et al. [24] presented a two-phase remanufacturing decision-making method for complex products. In the first phase, an augmented DEA is utilized to assess the efficiencies of the pre-selected components. Ouellette and Yan [25] developed a dynamic version of DEA model. They used an augmented DEA model to measure technical and allocative efficiencies. Table 1 indicates a comprehensive classification about different approaches of handling virtual DMUs in literature. Khalili-Damghani and Fadaei [26] introduce two virtual DMUs termed an ideal virtual DMU and anti-ideal virtual DMU to enhance discriminatory power of DEA.

In other studies (Allen et al. [27]; Pedraja-Chaparro et al. [28]), weight restrictions were considered to decrease efficient DMUs and improve discrimination among DMUs. Golany and Roll [8] stated that weight restriction and virtual DMUs, impact on the efficiency scores in the same direction. In other words, by tightening the bounds on weights or adding virtual DMUs, efficiency scores cannot improve. Virtual DMUs have the ability to turn DMUs previously considered efficient into inefficient ones and even turn DMUs previously taken into account inefficient into lower efficiency scores. Dyson and Thanassoulis [29] developed an approach termed direct weight restrictions for increasing discrimination among DMUs. In this approach the restrictions impose numerical limits on the weights. Charnes et al. [30] proposed the cone ratio model which results in at least one efficient DMU. Thompson et al. [31] proposed "The Assurance Region I" which results in at least one efficient DMU and "The Assurance Region II" which imposes restrictions on the ratio between input and output weights. In this case there is no certainty that there will be at least one efficient DMU. Bal et al. [32] developed an approach according to the dispersion of weights in DEA model to increase the discrimination among efficient DMUs. Hatami-Marbini et al. [33] used dual weight constraints to remove low discrimination power in DEA. Liu [34] proposed an approach for a fuzzy two-stage DEA model, where the weights are restricted in ranges. He used assurance region to reduce weight flexibility. Ennen and Batool [35] applied weights restrictions for inputs and outputs in the DEA procedure to increase the ability of DEA to differentiate performance levels. Wang et al. [36] developed a method to rank DMUs by imposing an appropriate minimum weight constraint on all input and output factors. Ebrahimi et al. [37] presented a modified type of DEA model to evaluate the performance in real-life systems which includes various types of weight restrictions and imprecise data. The developed approach eliminates the drawbacks of existing models, and provides more reliable outcomes.

The assurance region approach can be considered as another approach for increasing discrimination capability of DEA model. In this method, weights of input and output factors are restricted by upper and lower bounds. Haeri and Rezaie [4] proposed a three-step assurance region method to calculate upper and lower bounds of the input and output factors. In the first step of the proposed method, the basic (unbounded) DEA model is run. In the second step, optimal solutions of the unbounded DEA model are used to compute the average
weights of input and output factors. Finally, in the third step, two sets of constraints are added to the basic DEA model to restrict weights between upper and lower bounds.

2.1. Research gaps and contributions

According to the literature review on augmented DEA, the research gap and main contribution of this work are explained as follow:

First, with respect to the previous studies, it can be realized that augmented DEA has mostly been used for increasing the discrimination power of DEA model whereas two issues are not considered. The first problem is that the new obtained efficiency scores using augmented DEA model are usually unreal. Moreover, the distance between the efficiency scores of the basic DEA and the scores obtained by augmented DEA model is great. Accordingly, in this paper a heuristic method is extended to cope with both issues and improve the discrimination power of augmented DEA model in performance evaluation.

Second, different virtual DMUs lead to different results for DMUs ranking, therefore, the proposed method in this study tries to choose the best virtual DMUs in order to reduce the possibility of having inappropriate efficiency scores. As shown in Table 1, most of the previous studies have taken into account only one virtual DMU for increasing the capability of DEA which is mostly created by selecting the best values of each factor from all DMUs. In other words, for each output factor, the maximum value and for each input factor, the minimum value are selected. On the other hand, there are some articles that have considered two virtual DMUs with the best and worst performance and there are also few researches that have taken into account the number of virtual DMUs equal to the number of input and output factors. To the best of the authors’ knowledge, there is not any reference that employs a data mining approach of DMU’s data (i.e., clustering analysis) to determine the number as well as input and output factors’ value of virtual DMUs.

Third, since adding a new DMU, is the same as adding a weight restriction to the dual model (Weight Model) so in this study the computational effort for weight restrictions is eliminated and the discrimination power is improved only by creating new virtual DMUs.

3. Augmented Data envelopment analysis

DEA as a non-parametric method was initially introduced by Charnes et al. [38] and it is mainly used for evaluating a set of DMUs’ efficiency which are regarded as independent units to convert same inputs into same outputs. Generally, smaller value for input and larger value for output are preferable. In this method, the efficiency of a DMU is computed as follow:

\[
Efficiency = \frac{\text{weighted sum of outputs}}{\text{weighted sum of inputs}}
\]  

(1)

Each DMU can pick its best weights to increase its efficiency score. A DMU with an efficiency score of one is efficient and a DMU with efficiency score that is less than one is inefficient. The basic type of DEA model for evaluating a set of DMUs' efficiency is named as CCR model which has two drawbacks including lack of discrimination among efficient
DMUs and assigning unreal weights to input and output factors, in other words a small (or zero) weight is given to an important factor or a big weight is given to factor with less importance. To cope with aforementioned problems, Appalla [7] provided an augmented version of basic DEA model termed augmented DEA. In this approach a new virtual DMU called the ‘virtual best’ DMU is defined. Since, the lower value is more desirable for inputs and the higher value is more desirable for outputs, the input and output factors of virtual ideal DMU are generated by choosing the minimum value of each input factor and the maximum value of each output factor from the existing DMU base. Based on this method, the efficient frontier of the model is changed and hence the efficiency of each DMU is achieved according to the efficient frontier of the ‘virtual best’ DMU.

3.1. Basic DEA model

The following notations are used for formulating the DEA models:

\[ j \quad \text{Index of DMUs, } j = \{1,2, \ldots, n\} \]
\[ r \quad \text{Index of outputs, } r = \{1,2, \ldots, s\} \]
\[ i \quad \text{Index of outputs, } i = \{1,2, \ldots, m\} \]
\[ x_{ij} \quad \text{The amount of input } i \text{ for DMU } j \]
\[ y_{rij} \quad \text{The amount of output } r \text{ for DMU } j \]
\[ y^*_{ri} \quad \text{The maximum value of output } r \text{ in the DMU base} \]
\[ x^*_{i} \quad \text{The minimum value of input } i \text{ in the DMU base} \]
\[ u_r \quad \text{Weight of output } r \]
\[ v_i \quad \text{weight of input } i \]

The efficiency of \( DMU_o \) can be obtained by solving the following model:

\[
\begin{align*}
\text{Max } & \sum_{r=1}^{s} u_r y_{ro} \\
\text{Subject to,} & \\
\sum_{i=1}^{m} v_i x_{io} & = 1 \\
\sum_{r=1}^{s} u_r y_{rij} - \sum_{i=1}^{m} v_i x_{ij} & \leq 0 & j = 1, \ldots, n \\
u_r & \geq 0 \\
v_i & \geq 0
\end{align*}
\] (2)

3.2. Augmented DEA model

The mathematical model of the augmented DEA model using the above-mentioned notations is as follow:
Max \( \sum_{r=1}^{s} u_r y_{r} \)

Subject to.
\[
\sum_{i=1}^{m} v_i x_{i} = 1
\]
\[
\sum_{r=1}^{s} u_r y_{r} - \sum_{i=1}^{m} v_i x_{i} \leq \cdot \quad j = \cdot, \ldots, n
\]
\[
\sum_{r=1}^{s} u_r y_{r}^* - \sum_{i=1}^{m} v_i x_{i}^* \leq 0
\]

\( y_{r}^* \) and \( x_{i}^* \) form the input and output factors of the ‘virtual best’ DMU. The DMU’s efficiency is measured by using the above mathematical model and can then be used to rank the DMUs.

4. Proposed approach

In this paper a heuristic method is developed to enhance the discrimination power of augmented DEA model in performance evaluation. The major idea of this method is choosing the best virtual DMUs in order to reduce the possibility of having inappropriate efficiency scores. DEA analysis combined with clustering analysis is a very interesting tool for creating a set of potential virtual DMUs. By grouping DMUs into clusters, the visualization of input and output factors as well as the analysis of data set are improved and with the help of obtained information through clustering, virtual DMUs can be created in a more coherent manner compared to the previous approaches in the literature. This prevents the problem of having inappropriate efficiency scores. Therefore, in this section, a four-step algorithm is introduced to generate a set of potential virtual DMUs and so enhance the discrimination power of augmented DEA by employing the clustering approach of DMU’s data. For this purpose, first, cluster analysis is applied to group DMUs into clusters. In order to perform cluster analysis, the SPSS statistical software is used. Then, \( n \) virtual DMUs are created by choosing the best values of each factor from \( n \) optimal clusters determined in the previous step. Next, the new produced virtual DMUs are added to the existing DMUs base and finally, the DEA (CCR) model is run to obtain new efficiency scores for DMUs. The proposed algorithm proceeds the following steps.

4.1. Step 1. Implement the clustering algorithm on the existing DMU’s data

The purpose of this step is to group DMUs into clusters based on their inputs and outputs. In this regard, clustering as a popular data mining approach is used in this study to classify a set of DMUs into a number of different groups such that:

- DMUs in each group are similar to each other.
- DMUs of one group are different from the DMUs of other groups.
Consequently, for a given DMUs set the greater the differences between groups the more
homogeneous each group is and vice versa. To the best of the authors’ knowledge, this paper
is the first to create virtual DMUs by employing clustering approach, so this step of the
proposed approach is new and can be considered as one of the innovations of this research.

In order to perform cluster analysis, the SPSS statistical software is used in this step. SPSS as
a widely used tool for data analysis in social science covers many statistical analyses tests,
filters and prepares data for an analysis, creates various charts, carries out analyses
relationships between two and more factors, classifies data and creates clusters. This software
offers three methods for the cluster analysis (K-Means Cluster, Hierarchical Cluster, Two-
Step Cluster) according to the research needs. Furthermore, SPSS carries out an analysis and
concludes with more precision when working with complex relationships in data.
Additionally, it provides graphics that have more analytical features which facilitates
discussion of the resulting output from clustering process [39]. Clustering algorithm of
DMU’s data using SPSS can be described stepwise as follows:

4.1.1. Choose the type of clustering technique

Since the task of clustering is based on individual’s perception, there are a lot of methods that
can be utilized for achieving this goal. Hierarchical clustering is known as one of the most
popular clustering algorithms. In this paper, hierarchical clustering method is adopted for the
following reasons [40]:

- The main feature of hierarchical clustering is analyzing grouping in the data,
simultaneously over a variety of scales.

- The results of hierarchical clustering are typically provided in a dendrogram which is
a diagram that can be utilized as a visualization tool in monitoring the hierarchical
relationship between objects and controlling the decision making process.

- In hierarchical clustering results are reproducible.

- By interpreting the dendrogram you can stop at whatever number of clusters you find
appropriate.

Hierarchical clustering technique aims to create a hierarchy of clusters and can be divided
into two main types including agglomerative and divisive. Agglomerative clustering begins
with individual elements and aggregate them into clusters while divisive clustering begins
with all data and divide them into partitions. To determine which clusters for agglomerative
must be merged or where a cluster for divisive must be separated, a measure of distance
between pairs of elements is essential. This study uses the agglomerative hierarchical
clustering method to group DMUs into clusters.

4.1.2. Select a measure of similarity

The distance between the two points (DMUs) is considered as a measure of similarity. There
are some commonly used metrics such as Manhattan distance, Mahalanobis distance,
Maximum distance, Euclidean distance and Squared Euclidean distance. Euclidean distance or the squared Euclidean distance are the most common distance measure in published papers. Therefore, in this study the squared Euclidean distance between two DMUs is applied as a measure of similarity which is stated in below.

\[ d_{euc}^2(x, y) = \sum_{i=1}^{p} (x_i - y_i)^2 \]  \hspace{1cm} (11)

**4.1.3. Select the type of clustering method for the selected technique**

To calculate distances between two clusters, various methods have been proposed such as nearest neighbor, farthest neighbor, average linkage method, ward's method and centroid method. The different algorithms for hierarchical clustering vary mostly according to how the distance between the two clusters is calculated. In centroid method used in this study, each cluster is replaced by an average point (DMU) which is the centroid of that cluster.

**4.1.4. Determine the number of clusters**

SPSS has an output viewer window that contains all output we generate. In the first step of the cluster procedure, the proximity matrix will be produced in SPSS' Output. This matrix gives the squared Euclidean distance that was calculated between DMUs. Agglomeration Schedule that follows the proximity matrix in the output, shows the clusters obtained at each stage using hierarchical clustering.

**4.1.5. Determine the number of optimal clusters**

There are various indicators for determining the number of optimal clusters. R-SQUARED (RS) is one of the most widely used statistics for evaluating the cluster solution and determining the number of optimal clusters. RS measures that which groups are different from each other and which groups are homogeneous which is stated in below.

\[ RS = \frac{SS_b}{SS_t} = \frac{\text{sum of squares between – group}}{SS_t} = \frac{\text{sum of squares between – group}}{SS_b + SS_w \text{ (sum of squares within – group)}} \]  \hspace{1cm} (12)

\[ 0 \leq RS \leq 1 \]  \hspace{1cm} (13)

The values of RS range from 0 to 1, with 0 showing no differences among clusters and 1 showing maximum differences among clusters.

**4.2. Sep 2. Create n virtual DMUs by choosing the best values of each factor from n optimal clusters**

In this step, virtual DMUs are generated based on the optimal clusters. Clusters with the maximum RS are considered as basis to produce virtual DMUs. Thus, given that there are n optimal clusters, n virtual DMUs will be generated. Virtual DMUs are generated by choosing the best values of each factor from the optimal clusters. Since the higher value is more desirable for outputs and the lower value is more desirable for inputs so the input and output
factors of virtual ideal DMU are created by selecting the minimum value of each input factor and the maximum value of each output factor from the optimal clusters.

4.3. **Step 3. Add the created virtual DMUs to the existing DMU's data**

The efficient frontier of the basic DEA model is changed by adding a new virtual best DMU and hence the DMU's efficiency is achieved according to the efficient frontier of the new virtual DMU. Therefore, the discrimination power of classical DEA is improved in such a way that DMUs previously taken into account efficient will be indicated to be inefficient.

4.4. **Step 4. Run the DEA (CCR) model and calculate new efficiency scores**

Finally, to obtain new efficiency scores for DMUs by considering virtual DMUs, DEA (CCR) model is run.

The logic of each step is explained in Table 2.

Business Process Model and Notation (BPMN) is a graphical illustration of very complex processes, as a means of understanding, analyzing and making positive changes to processes. Usage of BPMN will help to visualize the processes and make better decisions. The most important superiority of BPMN over other techniques is that it’s a standard with well-defined set of rules. Hence, it makes collaboration much easier since it is familiar for many business analysts. Additionally, this standard is supported by most modeling tools which makes it much easier to share and edit if even utilizing various software. All mentioned matters make BPMN the most popular business process modeling technique at the present time. Accordingly, to make the process of the heuristic method proposed in this study for choosing 'virtual best' DMUs more understandable, BPMN is used (see Figure 1). Figure 2 indicates all the BPMN elements which are targeted in the transformation from above-mentioned steps to BPMN.

Explanation about symbols of Figure 2 is stated as follows. Task represents the lowest level activity within a process flow. Start and end events indicate the occurrence and result of a process, respectively. Data-based exclusive gateway creates alternative flows in a process so that only one of the paths can be chosen. Parallel gateway creates parallel paths so that no decision is required. Sequence flow links two elements of a process and indicates in which order the activities are performed. Data object provides information that activities require and finally, data store displays information banks related to the process.
5. Numerical examples

To show the application and superiority of the extended approach over the previous approaches, two numerical examples are borrowed from the literature. The data collection procedure of this study is expressed as below:

First, the purpose of data collection is determined. The purpose of this paper is to enhance the discrimination power of augmented DEA model in performance evaluation by creating a set of potential virtual DMUs based on employing clustering approach of DMU's data.

Second, sources of data are determined according to the purpose of the paper. Various sources such as online sites and relevant articles are investigated to determine suitable sources.

Third, the data is collected. The data sources for input and output factors are obtained from secondary sources. Relevant providers of secondary data are Wu and Blackhurst [16] and Cook and Kress [41].

In the first example, the results of new approach are compared with results of previous approaches that are used in Wu and Blackhurst [16] and Appalla [7] papers. In the second example, the data from Cook and Kress [41] is considered.

5.1. Example 1

The dataset of first example has been taken from a global company which provides communication and aviation electronics. It maintains headquarters and manufacturing operations in the United States with extra places in Europe, Mexico and Australia with more than 19,000 staff worldwide. The company has emphasized on improving the supplier performance. In this regard, Wu and Blackhurst [16] provided an approach for Supplier evaluation and selection of this company. In this case application there are 10 DMUs (suppliers), two inputs namely Price and Proprietary design partnerships and two output factors namely Delivery performance and Quality. The input and output factors have been selected by consulting with the managers in strategic sourcing division of the company which were seemed to be very important by the company. The inputs and outputs selected in this study are not inherently related. For example, price as an input factor represents the amount paid by a buyer while the input of proprietary design partnerships indicates agreements established between the supplier and the company concerning the use of technology in product design. On the other hand, the outputs of quality and delivery performance are benefits derived by the buyer. In this paper, this case application is used to demonstrate the superiority of this paper's approach to the approach used in Wu and Blackhurst [16]. The normalized values of the inputs and outputs for the suppliers are given in Table 3. In the first step, the basic DEA model is utilized to measure the efficiency scores. The results and efficiency scores are illustrated in Table 3.
As shown in Table 3, there are more than one efficient DMU which means the basic type of DEA model can't prioritize DMUs properly and the results of this model are not so useful. In order to solve this problem, the augmented DEA model introduced in section 3.2, is used which eliminates this drawback through the creation of a virtual DMU.

Table 4 shows that in the basic DEA context, S01, S02, S03, S05 and S09 are ranked as the best DMUs while in the augmented DEA context, S02 is not among the top five DMUs. On the other hand, the basic DEA ranks S07 and S10 as the worst DMUs while the augmented DEA ranks S06 and S08 besides S07 and S10 as the worst DMUs which indicate that the basic DEA performance is not on a par with the augmented DEA model. In addition, as can be seen in Table 4, there is great distance between the efficiency scores of S02, S06 and S08 in the basic DEA model and their scores in the augmented DEA model. In other words, the obtained efficiency scores for these DMUs through augmented DEA are unreal. To overcome this problem, the four-step algorithm described in Section 4 is implemented using SPSS statistical software. With respect to this algorithm, first, a hierarchical agglomerative clustering is applied to group 10 DMUs into clusters based on their inputs and outputs. To do so, the squared Euclidean distance and centroid method are applied to calculate the distance between two DMUs and the distance between two clusters, respectively. Table 5 indicates the proximity matrix which provides the squared Euclidean distance calculated between 10 DMUs. As shown in Table 5, S07 and S10 have the smallest distance (approximately .010).

Table 6 illustrates the Agglomeration Schedule which presents how the hierarchical cluster analysis increasingly clusters the DMUs. As Table 6 represents, each row displays a stage at which two DMUs are merged to form a cluster, through an algorithm controlled by the distance and centroid linkage. The number of stages is one less than the number of DMUs. The schedule gives all of the stages in which the clusters are merged until there is only one cluster remaining after the last stage. The coefficients at each stage illustrate the distance of the two clusters being merged.

“Stage Cluster First Appears” in Table 6 shows the clusters created from previous stages and first appear in this stage. For instance, at the first stage of Agglomeration Schedule (see Table 6), S7 is combined with S10 since these two DMUs have the smallest squared Euclidean distance. The two zeros under Cluster 1 and Cluster 2 indicate that neither DMU has been previously clustered. The cluster created by joining S7 and S10 next appears in stage 8. In stage 8, number 6 under Cluster 1 and number 1 under Cluster 2 indicate that the clusters created in stages 1 and 6 are joined in this stage. The resulting cluster next appears in stage 9.

Hierarchical clustering results are usually illustrated in dendrogram or Icicle plot. Figure 3 is a dendrogram of the results of Table 6. Dendrogram as a tree diagram is often utilized to show the arrangement of the clusters generated by hierarchical clustering. In this diagram, the distance or dissimilarity between clusters are represented by the horizontal axis and the elements and clusters are represented by the vertical axis. In order to interpret a dendrogram, the height at which any two elements are joined together should be considered. For instance, as can be observed from Figure 3, the height of the link that joins S7 and S10 together is the smallest, so they are the most similar. The next two most similar elements are S1 and S6.
Compared to dendrogram, it is easier in an icicle plot to read which elements belong to which clusters since the element labels are revealed exactly where the clusters are specified. Figure 4 presents icicle plot from clustering of 10 DMUs. The column placed between two DMUs indicates the number of common clusters between them. As can be observed from Figure 4, the highest common column is between S7 and S10, this means that S7 and S10 belong to the first stage and they are the most similar. After that the highest common cluster is between S1 and S6 which belong to the second stage.

Table 7 shows the clusters obtained at each stage using hierarchical clustering. According to Table 7, six statuses can be considered for clustering. In continue, RS as the most widely used statistic is applied to determine the number of optimal clusters. According to Table 8, statuses 1 and 2 have the maximum differences among clusters, therefore they can be considered as basis to produce virtual DMUs. As mentioned previously, the higher value is more desirable for outputs and the lower value is more desirable for inputs. So the input and output factors of virtual ideal DMU are generated by choosing the minimum value of each input factor and the maximum value of each output factor from the existing clusters in statuses 1 and 2.

Additionally, since there are five clusters in status 1 and four clusters in status 2, five virtual DMUs and four virtual DMUs are derived from the data of clusters respectively. As shown in Table 9, five virtual DMUs are created for status 1 and four virtual DMUs are created for status 2. The new produced virtual DMUs are added to the existing DMUs base and then the DEA (CCR) model is run to obtain new efficiency scores for 10 DMUs. A comparison between the basic DEA model, augmented DEA, Wu and Blackhurst [16] approach and heuristic method is provided in Table 10. For receiving the best result, efficiency scores are calculated in both statues.

As shown in Table 10, the number of efficient DMUs in status 2 is less than the number of efficient DMUs in status 1; therefore status 2 is chosen as the best result. Table 10 includes some interesting findings in connection with the comparison of four above approaches. These findings are as follows:

- It is interesting to note that the heuristic method shows the greater discriminatory power compared to Appalla [7] approach (augmented DEA model).
- The distance between the efficiency scores of the basic DEA model and the efficiency scores of heuristic method (statues 2) has been minimized and the discrimination power of Appalla [7] approach (augmented DEA model) has been improved through the proposed approach.
- In Wu and Blackhurst [16] method, two enhancements have been added to the basic DEA model including virtual DMUs and weight constraints which require more computational effort while in the heuristic method, computational effort for weight constraints has been eliminated and just virtual DMUs have been added to the basic DEA model.

5.2. Example 2
To represent the applicability of the extended method, the data from Cook and Kress [41] is considered. In this case application, there are 12 DMUs, three inputs and two outputs. The normalized values of each factor have been given in Table 11. The basic DEA model is applied to the data set which results in more than one efficient DMU (see Table 11). To solve this problem, the augmented DEA model is used.

As can be seen Table 11, there is great distance between the efficiency scores of the basic DEA model and scores obtained through augmented DEA. To overcome this problem, the heuristic method proposed in this study is exploited. For this purpose, the hierarchical clustering is first performed. Table 12 shows the clusters obtained at each stage using hierarchical clustering. According to Table 12, two statuses can be considered for clustering. For determining the number of optimal clusters, RS is used.

As shown in Table 13, status 1 has the maximum differences among clusters, therefore it can be considered as basis to produce virtual DMUs. Three virtual DMUs are created for this case with respect to the input and output data of clusters (see Table 14). Three virtual DMUs are added to the existing DMUs base and then the DEA (CCR) model is run to obtain new efficiency scores for 12 DMUs. The comparison study results are shown in Table 15.

Table 15 shows the enhanced discriminatory power of heuristic method over Appalla [7] approach (augmented DEA) to rank DMUs.

6. Results and discussion

This study aims to extend the application of augmented DEA through creating a set of potential virtual DMUs based on the cluster analysis of DMU's data in order to reduce the possibility of having inappropriate efficiency scores. The proposed heuristic method in this study can address the problems of the augmented DEA model where DMUs may have unrealistic efficiency scores or great distance with the efficiency scores obtained by the basic DEA. In order to show the application of the proposed method in performance evaluation context, it is implemented on two numerical examples which are taken from the literature.

In total, the DMUs in the first and second examples can be classified into groups in terms of efficiency scores obtained through the basic DEA model (see Table 16). As can be observed from Table 16, considering efficiency scores, 10 DMUs in example 1 are classified into four groups and 12 DMUs in example 2 are classified into six groups. The specific point in Table 16 is that in the second example, the range of efficiency scores is more varied in comparison to the first example.

In the first example, the algorithm of the proposed method is implemented on a dataset from a communication and aviation electronics company that has emphasized on the supplier performance. This example consists of 10 DMUs (suppliers), two inputs and two output. According to the algorithm explained in Section 4, first, a hierarchical agglomerative clustering is applied to group 10 suppliers into clusters based on their inputs and outputs which results in creation of six statuses for clustering. Among the created statuses, statuses 1 and 2 are determined as optimal clusters since they have the maximum values of RS.
(RS_{statues 1} = 0.941 and RS_{statues 2} = 0.886) compared to the other statues. Then, five virtual DMUs based on the five clusters in status 1 and four virtual DMUs based on the four clusters in status 2 are created by choosing the best values of each factor from the existing clusters in each status. Next, the new created virtual DMUs are added to the existing DMU's data and finally the DEA (CCR) model is run to obtain new efficiency scores for 10 suppliers. According to the efficiency results, the number of efficient DMUs in status 2 is less than the number of efficient DMUs in status 1; therefore status 2 is chosen as the best result for comparison experiments. As Table 10 represents, the full rankings are achieved through augmented DEA and the heuristic method and the discrimination powers of these two approaches are very high compared to the basic DEA model as a distinctive rank is assigned to each DMU. However, it can be concluded from contents of Table 10 that the efficiency scores obtained from the heuristic method are more consistent with the results of the basic DEA model and therefore they are considered to be more realistic in comparison to efficiency scores obtained by augmented DEA. Figure 5 provides a comparative result of the performance of three approaches towards the efficiency scores of 10 suppliers. As can be observed from Figure 5, there is a considerable distance between the efficiency scores obtained by the basic DEA model and the efficiency scores of augmented DEA, while by applying the heuristic method and adding a set of potential virtual DMUs to the existing DMU's data this great distance has been reduced and therefore, the discrimination power of augmented DEA has been improved through the proposed approach.

In the second example, the algorithm of the proposed heuristic method is implemented on a dataset from Cook and Kress [41] including 12 DMUs, three inputs and two outputs. First, 12 DMUs are grouped into clusters considering their inputs and outputs which results in creation of two statuses for clustering. Status 1 is selected as optimal clusters since it has the maximum value of RS (RS_{statues 1} = 0.897) compared to Status 2 (RS_{statues 1} = 0.797). Then, three virtual DMUs are created based on the three clusters in status 1 by choosing the best values of each factor from the existing clusters in this status. Next, the new created virtual DMUs are added to the existing DMU's data and finally the DEA (CCR) model is run to obtain new efficiency scores for 12 DMUs. Even though augmented DEA increases discriminatory ability in basic DEA, the efficiency scores obtained by augmented DEA do not reflect the real efficiency of the DMUs obtained by CCR DEA and there is a considerable distance between their efficiency scores. As Figure 6 represents, the heuristic method has advantages over the augmented DEA model since through this method, DMUs do not get unrealistic efficiency scores and contrary to augmented DEA, there is less distance between efficiency scores obtained by this method and efficiency scores of CCR DEA and this is due to the correct selection of virtual DMUs.

7. Conclusions

Today, issues that involve decision support systems and efficiency analysis inside a company require special consideration and several tools have been introduced to assist managers. One of these tools is DEA which its application is expanding in new developments and in research. The problem is that the basic DEA model is usually blamed because of low discriminating power. To overcome this problem, many studies have looked into ways to
insert engineering such as virtual DMUs into DEA with the aim of having a set of standard data to evaluate DMUs through simply expanding the reference set. Since different virtual DMUs lead to different results for DMUs ranking, the current study is an attempt to create the best virtual DMUs in order to reduce the possibility of having inappropriate efficiency scores. The main contribution of this study in comparison with the different approaches of handling virtual DMUs is the usage of hierarchical clustering method in order to create a set of potential virtual DMUs in a more coherent manner. The proposed heuristic method in this study can address the problems of adding virtual DMUs to the basic DEA model where DMUs may have unrealistic efficiency scores or great distance with the efficiency scores obtained by the basic DEA. The applicability of proposed heuristic method is tested by two numerical examples taken from the literature. The results of proposed heuristic method are compared with the existing methods in past works. The results show that using hierarchical clustering method in DEA for creating virtual DMUs will improve DMU's efficiency over the previous approaches.

Regardless of the type and number of DMUs, the proposed model of this study is applicable to all problems that their discrimination powers need to be enhanced. This property makes the usage of the proposed approach more general and it fits with a large number of real-life and managerial applications. The managerial insights of the proposed method lie in its use to provide a better evaluation ability to organizations to achieve more profitability. Performance evaluation has been known as one of the management functions in organizations. However, in order to evaluate performance, it is essential to apply appropriate evaluation tools. The proposed heuristic method in this study allows managers to compare performers for making important evaluation decisions. In addition, incorporating heavy computational efforts such as weight constraints into DEA in order to improve discriminating or ranking of efficient performers may not be interesting in terms of managerial perspective. The proposed method of this study eliminates such computational efforts and incorporates standards which increases the ability for organizations to evaluate and rank performers.
References

1. Zegordi, S. H., & Omid, A. "Efficiency assessment of Iranian handmade carpet company by network DEA", Scientia Iranica. Transaction E, Industrial Engineering, 25(1), 483-491(2018).

2. Rezaie, K., Haeri, A., Amalnick, M.S. and et al. "Using augmented dea to calculate efficiency scores of organizational resources", In Computer Modelling and Simulation (UKSim), 2011 UkSim 13th International Conference on (pp. 365-369). IEEE (2011, March).

3. Haeri, A., Rezaie, K., & Amalnick, M. S. "Using multi-objective DEA to assess the overall and partial performance of hierarchical resource Utilization", Research Journal of Applied Sciences, Engineering and Technology, 5(4), 1213-1224 (2013).

4. Haeri, A., & Rezaie, K. "Using data envelopment analysis to investigate the efficiency of resource utilisation and to develop an improvement plan", International journal of productivity and quality management, 13(1), 39-66 (2014).

5. Haeri, A., & Ghoussi, R. "Using data envelopment analysis (DEA) to improve the sales performance in Iranian agricultural clusters by utilizing business networks and business development services providers (BDSPs)", Journal of Industrial and Systems Engineering, 9(3), 82-95 (2016).

6. Rezaee, M. S., Haeri, A., & Noori, S. "Using data envelopment analysis to evaluate the performances of food production companies based on EFQM's criteria and to present an improvement plan", International Journal of Business Excellence, 14(2), 256-274 (2018).

7. Appalla, R.K. "An augmented DEA for supplier evaluation", Arizona State University, Thesis (2003).

8. Golany, B. and Roll, Y. "Incorporating standards via DEA", In Data envelopment analysis: Theory, methodology, and applications (pp. 313-328). Springer, Dordrecht (1994).

9. Moslehi, F., Haeri, A., Gholamian, M. "A novel selective clustering framework for appropriate labeling of the clusters based on K-means algorithm", Scientia Iranica, (2019).

10. Beheshti-Nia, M., Mousavi, Z. (2017). "A new classification method based on pairwise SVM for facial age estimation", Journal of Industrial and Systems Engineering, 10(1), pp. 91-107.

11. Granato, D., Santos, J. S., Escher, G. B., and et al. "Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective", Trends in Food Science & Technology, 72, 83-90 (2018).

12. Shokr, I., Amalnick, M.S. and Torabi, S.A. "An Augmented Common Weight Data Envelopment Analysis for Material Selection in High-tech Industries", International Journal of Supply and Operations Management, 3(2), pp.1234 (2016).
13. Shen, W.F., Zhang, D.Q., Liu, W.B. and et al. "Increasing discrimination of DEA evaluation by utilizing distances to anti-efficient frontiers", *Computers & Operations Research*, 75, 163-173 (2016).

14. Wu, T., Shunk, D., Blackhurst, J. and et al. "AIDEA: A methodology for supplier evaluation and selection in a supplier-based manufacturing environment", *International journal of manufacturing technology and management, 11*(2), pp.174-192 (2007).

15. Ghasemi, M.R., Ignatius, J. and Rezaee, B. "Improving discriminating power in data envelopment models based on deviation variables framework", *European Journal of Operational Research, 278*(2), pp.442-447 (2019).

16. Wu, T. and Blackhurst, J. "Supplier evaluation and selection: an augmented DEA approach", *International Journal of Production Research, 47*(16), pp.4593-4608 (2009).

17. Hou, Q., Wang, M., & Zhou, X. "Improved DEA Cross Efficiency Evaluation Method Based on Ideal and Anti-Ideal Points", *Discrete Dynamics in Nature and Society*, (2018).

18. Noorizadeh, A., Mahdiloo, M. and Farzipoor Saen, R. "A data envelopment analysis model for selecting suppliers in the presence of both dual-role factors and non-discretionary inputs", *International Journal of Information and Decision Sciences, 4*(4), pp.371-389 (2012).

19. Kianfar, K., Ahadzadeh Namin, M., Alam Tabriz, A., and et al. "Hybrid cluster and data envelopment analysis with interval data", *Scientia Iranica, 25*(5), 2904-2911 (2018).

20. Hatefi, S.M. and Razmi, J. "An integrated methodology for supplier selection and order allocation in the presence of imprecise data", *International Journal of Industrial and Systems Engineering, 15*(1), pp.51-68 (2013).

21. Mahdiloo, M., Noorizadeh, A. and Saen, R.F. "A new model for suppliers ranking in the presence of both dual-role factors and undesirable outputs", *International Journal of Logistics Systems and Management, 15*(1), pp.93-107 (2013).

22. Haeri, A. "Evaluation and comparison of crystalline silicon and thin-film photovoltaic solar cells technologies using data envelopment analysis", *Journal of Materials Science: Materials in Electronics, 28*(23), pp.8183-18192 (2017).

23. Rezaee, M. S., Haeri, A., & Noori, S. "Automotive Vendor's Performance Evaluation and Improvement Plan Presentation by Using a Data Envelopment Analysis", *International Journal of Engineering, 31*(2), 374-381(2018).

24. Geng, X., Gong, X. and Chu, X. "Component oriented remanufacturing decision-making for complex product using DEA and interval 2-tuple linguistic TOPSIS", *International Journal of Computational Intelligence Systems, 9*(5), 984-1000 (2016).

25. Ouellette, P. and Yan, L. "Investment and dynamic DEA", *Journal of Productivity Analysis, 29*(3), pp.235-247 (2008).
26. Khalili-Damghani, K., & Fadaei, M. "A comprehensive common weights data envelopment analysis model: ideal and anti-ideal virtual decision making units approach", *Journal of Industrial and Systems Engineering*, 11(3), 281-306 (2018).

27. Allen, R., Athanassopoulos, A., Dyson, R.G. and et al. "Weights restrictions and value judgements in data envelopment analysis: evolution, development and future directions", *Annals of operations research*, 73, pp.13-34 (1997).

28. Pedraja-Chaparro, F., Salinas-Jimenez, J. and Smith, P. "On the role of weight restrictions in data envelopment analysis", *Journal of Productivity Analysis*, 8(2), pp.215-230 (1997).

29. Dyson, R.G. and Thanassoulis, E. "Reducing weight flexibility in data envelopment analysis", *Journal of the Operational Research Society*, 39(6), pp.563-576 (1988).

30. Charnes, A., Cooper, W.W., Huang, Z.M. and et al. "Polyhedral cone-ratio DEA models with an illustrative application to large commercial banks", *Journal of econometrics*, 46(1-2), pp.73-91 (1990).

31. Thompson, R.G., Singleton Jr, F.D., Thrall, R.M. and et al. "Comparative site evaluations for locating a high-energy physics lab in Texas", *interfaces*, 16(6), pp.35-49 (1986).

32. Bal, H., Örkcü, H.H. and Çelebioğlu, S. "A new method based on the dispersion of weights in data envelopment analysis", *Computers & Industrial Engineering*, 54(3), pp.502-512 (2008).

33. Hatami-Marbini, A., Rostamy-Malkhalifeh, M., Agrell, P.J., and et al. "Extended symmetric and asymmetric weight assignment methods in data envelopment analysis", *Computers & Industrial Engineering*, 87, pp.621-631 (2015).

34. Liu, S.T. "Restricting weight flexibility in fuzzy two-stage DEA", *Computers & Industrial Engineering*, 74, pp.149-160 (2014).

35. Ennen, D. and Batool, I. "Airport efficiency in Pakistan-A Data Envelopment Analysis with weight restrictions", *Journal of Air Transport Management*, 69, pp.205-212 (2018).

36. Wang, Y.M., Luo, Y. and Liang, L. "Ranking decision making units by imposing a minimum weight restriction in the data envelopment analysis", *Journal of Computational and Applied Mathematics*, 223(1), pp.469-484 (2009).

37. Ebrahimi, B., Rahmani, M. and Ghodsypour, S.H. "A new simulation-based genetic algorithm to efficiency measure in IDEA with weight restrictions", *Measurement*, 108, pp.26-33 (2017).

38. Charnes, A., Cooper, W.W. and Rhodes, E. "Measuring the efficiency of decision making units", *European journal of operational research*, 2(6), pp.429-444 (1978).

39. Paura, L., & Arhipova, I. "Advantages and Disadvantages of Professional and Free Software for Teaching Statistics", *Information Technology and Management Science*, 15(1), 9-64 (2012).
40. Kamis, N. H., Chiclana, F., & Levesley, J. "Geo-uninorm consistency control module for preference similarity network hierarchical clustering based consensus model", *Knowledge-Based Systems*, 162, 103-114 (2018).

41. Cook, W.D. and Kress, M. "Characterizing an equitable allocation of shared costs: A DEA approach", *European Journal of Operational Research*, 119(3), pp.652-661 (1999).

**Figure Captions**

**Figure 1.** The process of choosing 'virtual best' DMUs

**Figure 2.** Overview of applied BPMN elements

**Figure 3.** Dendrogram using centroid linkage

**Figure 4.** Icicle plot

**Figure 5.** Comparative results of the performance of three approaches towards the efficiency scores of 10 DMUs

**Figure 6.** Comparative results of the performance of three approaches towards the efficiency scores of 12 DMUs

**Table Captions**

**Table 1.** Different approaches of handling virtual DMUs in literature

**Table 2.** The logic of each step

**Table 3.** Input and output data and efficiency scores of 10 DMUs using the basic DEA model

**Table 4.** A comparison between the basic DEA model and augmented DEA

**Table 5.** Squared Euclidean distance calculated between 10 DMUs

**Table 6.** Agglomeration Schedule of 10 DMUs

**Table 7.** Clusters obtained at each stage using hierarchical clustering

**Table 8.** Six selected statuses

**Table 9.** Created virtual DMUs for the first numerical example

**Table 10.** Comparison results of the four approaches

**Table 11.** Input and output data and efficiency scores of 12 DMUs

**Table 12.** Clusters obtained at each stage using hierarchical clustering

**Table 13.** Two selected statuses

**Table 14.** Three Virtual DMUs for the second numerical example

**Table 15.** Comparison results
Table 16. Detected groups for the first and second examples based on efficiency scores by the basic DEA model
Choosing Virtual Best DMUs

1. Consider SPSS statistical software
2. Group DMUs into clusters
3. Perform hierarchical cluster analysis
4. Select hierarchical agglomerative clustering
5. Select hierarchical divisive clustering
6. Create an Agglomeration Schedule based on proximity matrix
7. Produce proximity matrix
8. Select a measure for computing distance between two DMUs
9. Select a method for computing distance between two clusters
10. Determine the number of optimal clusters using R-SQUARED (R^2)
11. Obtain clusters at each stage
12. Generate n virtual DMUs based on n optimal clusters
13. Select the maximum value of each input factor of optimal clusters
14. Select the maximum value of each output factor of optimal clusters
15. Run DEA (CCR) model to obtain new efficiency scores
16. Add new produced virtual DMUs to the existing DMUs base

Figure 1
Figure 2

Dendrogram using Centroid Linkage
Rescaled Distance Cluster Combine

Figure 3
Figure 4

Figure 5
Table 1

| Ref.                          | The number of virtual DMUs | Input and output factors' value of virtual DMUs                                                                 |
|-------------------------------|-----------------------------|---------------------------------------------------------------------------------------------------------------|
| Rezaie et al. [2]             | One virtual DMU            | The best values of each factor from all DMUs (high outputs and low inputs)                                   |
| Appalla [7]                   | One virtual DMU            | The best values of each factor from all DMUs (high outputs and low inputs)                                   |
| Wu et al. [14]                | One virtual DMU            | The best values of each factor from all DMUs (high outputs and low inputs)                                   |
| Wu and Blackhurst [16]        | Equal to the number of input and output factors | The best value of one factor and the average values of the remaining factors from all DMUs                  |
| Noorizadeh et al. [18]        | One virtual DMU            | The best values of each factor from the efficient DMUs                                                      |
| Hatefi and Razmi [20]         | One virtual DMU            | The best values of each factor from all DMUs (high outputs and low inputs)                                   |
| Mahdiloo et al. [21]          | One virtual DMU            | The best values of each factor from the efficient DMUs                                                      |
| Haeri [22]                    | Two virtual DMUs           | 1<sup>st</sup> virtual DMU: The best values of each factor from all DMUs (high outputs and low inputs)       |
|                               |                             | 2<sup>nd</sup> virtual DMU: The worst values of each factor from all DMUs (low outputs and high inputs)      |
| Rezaee et al. [23]            | Two virtual DMUs           | 1<sup>st</sup> virtual DMU: The best values of each factor from all DMUs (high outputs and low inputs)       |
|                               |                             | 2<sup>nd</sup> virtual DMU: The worst values of each factor from all DMUs (low outputs and high inputs)      |
| Geng et al. [24]              | Equal to the number of input and output factors | The best value of one factor and the average values of the remaining factors from all DMUs                  |
| Khalili-Damghani and Fadaei [26] | Two virtual DMUs         | Ideal virtual DMU: The best target formed according to the observed DMUs                                      |
|                               |                             | anti-ideal virtual DMU: The worst target formed according to the observed DMUs                               |
| This paper                    | Equal to the number of optimal clusters | The best values of each factor from the optimal clusters                                                     |

Table 2

| Step   | Logic                                                                 |
|--------|----------------------------------------------------------------------|
| Step 1 | The aim of this step is to group DMUs such that each cluster is as homogeneous as possible |
| Step 2 | Optimal clusters provide the ability to create the best virtual DMUs |

Figure 6
Step Logic
Step 3 The addition of virtual DMU increases the capability of the basic DEA
Step 4 The purpose of this step is to calculate new efficiency scores based on the added virtual DMUs

**Table 3**

| DMUs | Proprietary design partnerships | Price | Quality | Delivery performance | Efficiency scores |
|------|--------------------------------|-------|---------|----------------------|------------------|
| S01  | 1                              | 0.0715| 0.4285  | 0.98                 | 1                |
| S02  | 0.998                          | 0.1173| 0.7143  | 0.991                | 1                |
| S03  | 0.336                          | 0.7105| 0.8571  | 0.98                 | 1                |
| S04  | 0.65                           | 1      | 1       | 0.999                | 0.66             |
| S05  | 0.336                          | 0.1801| 0.6428  | 0.985                | 1                |
| S06  | 1                              | 0.1801| 0.431   | 0.98                 | 0.68             |
| S07  | 0.998                          | 0.8992| 0.2585  | 0.995                | 0.34             |
| S08  | 0.999                          | 0.2111| 0.9286  | 0.98                 | 0.94             |
| S09  | 0.336                          | 0.2124| 0.4     | 0.99                 | 1                |
| S10  | 0.999                          | 0.952 | 0.3448  | 0.99                 | 0.34             |

**Table 4**

| DMUs | The basic DEA model | Augmented DEA model |
|------|---------------------|---------------------|
| S01  | 1                   | 0.98                |
| S02  | 1                   | 0.98                |
| S03  | 1                   | 0.98                |
| S04  | 0.66                | 0.52                |
| S05  | 1                   | 0.99                |
| S06  | 0.58                | 0.39                |
| S07  | 0.34                | 0.34                |
| S08  | 0.94                | 0.34                |
| S09  | 1                   | 0.99                |
| S10  | 0.34                | 0.33                |

**Table 5**

| Case | 1:S1 | 2:S2 | 3:S3 | 4:S4 | 5:S5 | 6:S6 | 7:S7 | 8:S8 | 9:S9 | 10:S10 |
|------|------|------|------|------|------|------|------|------|------|--------|
| 1:S1 | 0    | 0.084| 1.033| 1.312| 0.499| 0.012| 0.714| 0.270| 0.462| 0.782 |
| 2:S2 | 0.084| 0    | 0.811| 0.982| 0.447| 0.084| 0.819| 0.055| 0.546| 0.833 |
| 3:S3 | 1.033| 0.811| 0    | 0.203| 0.327| 0.904| 0.832| 0.694| 0.457| 0.760 |
| 4:S4 | 1.312| 0.982| 0.203| 0    | 0.899| 1.119| 0.681| 0.750| 1.079| 0.553 |
| 5:S5 | 0.499| 0.447| 0.327| 0.899| 0    | 0.486| 1.103| 0.522| 0.060| 1.124 |
| 6:S6 | 0.012| 0.084| 0.904| 1.119| 0.486| 0    | 0.547| 0.249| 0.443| 0.603 |
| 7:S7 | 0.714| 0.819| 0.832| 0.681| 1.103| 0.547| 0    | 0.923| 0.930| 0.010 |
| 8:S8 | 0.270| 0.055| 0.694| 0.750| 0.552| 0.294| 0.923| 0    | 0.719| 0.890 |
| 9:S9 | 0.462| 0.546| 0.457| 1.079| 0.060| 0.443| 0.930| 0.719| 0    | 0.990 |
| 10:S10|0.782|0.833|0.760|0.553|1.124|0.603|0.010|0.890|0.990|0.000 |

**Table 6**

| Stage | Cluster Combined | Coefficients | Stage Cluster First Appears | Next Stage |
|-------|------------------|--------------|-----------------------------|-----------|

26
| Stage | Cluster Combined | Coefficients | Stage Cluster First Appears | Next Stage |
|-------|------------------|--------------|-----------------------------|------------|
|       | Cluster 1 | Cluster 2 |                | Cluster 1 | Cluster 2 |
| 1     | 7      | 10     | 0.010           | 0         | 0         | 8         |
| 2     | 1      | 6      | 0.012           | 0         | 0         | 5         |
| 3     | 2      | 8      | 0.055           | 0         | 0         | 5         |
| 4     | 5      | 9      | 0.060           | 0         | 0         | 7         |
| 5     | 1      | 2      | 0.155           | 2         | 3         | 7         |
| 6     | 3      | 4      | 0.203           | 0         | 0         | 8         |
| 7     | 1      | 5      | 0.453           | 5         | 4         | 9         |
| 8     | 3      | 7      | 0.653           | 6         | 1         | 9         |
| 9     | 1      | 3      | 0.532           | 7         | 8         | 0         |

Table 7

| Stage | Clusters                          | RS   |
|-------|-----------------------------------|------|
| 1     | (S7,S10)                          | 0.941|
| 2     | (S1,S6) (S2,S8)                   | 0.886|
| 3     | (S5,S9) (S1,S6,S2,S8)             | 0.447|
| 4     | (S3,S4) (S7,S10)                  | 0.675|
| 5     | (S1,S6,S2,S8,S5,S9)               | 0.712|
| 6     | (S3,S4,S7,S10)                    | 0.658|

Table 8

| Status | Clusters | RS   |
|--------|----------|------|
| 1      | (S7,S10) | 0.998|
| 2      | (S1,S6) | 0.999|
| 3      | (S2,S8) | 0.999|
| 4      | (S5,S9) | 0.999|
| 5      | (S7,S10) | 0.999|
| 6      | (S1,S6,S2,S8) | 0.999|

Table 9

| Status 1 | Output | Input | Status 2 | Output | Input |
|----------|--------|-------|----------|--------|-------|
| clusters | Proprietary design partnership | Price | Quality | Delivery performance | clusters | Proprietary design partnership | Price | Quality | Delivery performance |
|          | 0.998  | 0.8992| 0.2585  | 0.995  | 1      | 0.0715 | 0.4285 | 0.98    |          |
| (S7,S10) | 0.999  | 0.952 | 0.3448  | 0.99   |        |        |        |         |          |
| Virtual 1| 0.998  | 0.8992| 0.3448  | 0.995  | 1      | 0.1801 | 0.431  | 0.98    |          |
| (S1,S6)  | 0.999  | 0.952 | 0.3448  | 0.99   |        |        |        |         |          |
| Virtual 2| 1      | 0.0715| 0.4285  | 0.98   | 0.998  | 0.8992 | 0.2585 | 0.995   |          |
| (S2,S8)  | 0.998  | 0.952 | 0.3448  | 0.99   |        |        |        |         |          |
| Virtual 3| 0.998  | 0.1173| 0.9286  | 0.991  | 0.999  | 0.1173 | 0.9286 | 0.991   |          |
| (S5,S9)  | 0.336  | 0.1801| 0.6428  | 0.985  | 0.1801 | 0.6428 | 0.985  | 0.336   | 0.2124 | 0.4   | 0.99   |
| Clusters          | Proprietary design partner | Price | Quality | Delivery performance |
|-------------------|----------------------------|-------|--------|----------------------|
| Status 1          |                            |       |        |                      |
| Output            |                           | 0.336 | 0.2124 | 0.4                  |
| Input             |                           | 0.99  |        |                      |
| Status 2          |                            |       |        |                      |
| Output            |                           | 0.336 | 0.1801 | 0.6428               |
| Input             |                           | 0.99  |        |                      |

(Virtual 4)  
(S3,S4)  
(S3,S4)

| Clusters          | Proprietary design partner | Price | Quality | Delivery performance |
|-------------------|----------------------------|-------|--------|----------------------|
| Status 1          |                            |       |        |                      |
| Output            |                           | 0.336 | 0.7105 | 0.8571               |
| Input             |                           | 0.98  |        |                      |
| Status 2          |                            |       |        |                      |
| Output            |                           | 0.336 | 0.7105 | 1                    |
| Input             |                           | 1     |        | 0.999               |

Table 10

| DMUs | CCR DEA mode | Augmented DEA (Appalla [7]) | Wu and Blackhurst [16] approach | The heuristic method |
|------|--------------|-----------------------------|----------------------------------|----------------------|
|      | Status 1     | Status 2                    |                                  |                      |
| 1    | 1            | 0.98                        | 0.95                             | 1                    |
| 2    | 1            | 0.6                         | 0.85                             | 0.92                 |
| 3    | 1            | 0.98                        | 0.98                             | 0.98                 |
| 4    | 0.66         | 0.52                        | 0.66                             | 0.59                 |
| 5    | 1            | 0.99                        | 1                                | 1                    |
| 6    | 0.68         | 0.39                        | 0.61                             | 0.68                 |
| 7    | 0.34         | 0.34                        | 0.34                             | 0.34                 |
| 8    | 0.94         | 0.33                        | 0.86                             | 0.81                 |
| 9    | 1            | 0.99                        | 0.99                             | 0.99                 |
| 10   | 0.34         | 0.33                        | 0.33                             | 0.34                 |

Table 11

| DMUs | Input | Output | CCR DEA model | Augmented DEA (Appalla [7]) |
|------|-------|--------|---------------|-----------------------------|
|      | x1    | x2     | x3           | y1             | y2             |               |
| S01  | 0.2889| 0.0285 | 0.0034       | 0.0519         | 0.6248         | 0.7498        |
| S02  | 0.2454| 0.0176 | 0.0025       | 0.057          | 0.5075         | 0.919         |
| S03  | 0.3492| 0.0218 | 0.0017       | 0.0586         | 0.4849         | 0.6797        |
| S04  | 0.2312| 0.0092 | 0.0034       | 0.0544         | 0.5528         | 1             |
| S05  | 0.2479| 0.0092 | 0.0008       | 0.0586         | 0.3685         | 1             |
| S06  | 0.2973| 0.0201 | 0.0101       | 0.0653         | 0.892          | 0.9615        |
| S07  | 0.4481| 0.0109 | 0.0042       | 0.0561         | 0.3786         | 0.8173        |
| S08  | 0.227 | 0.0235 | 0       | 0.0611         | 0.4899         | 1             |
| S09  | 0.2663| 0.0168 | 0          | 0.0586         | 0.8953         | 1             |
| S10  | 0.3677| 0.0494 | 0.0008     | 0.0578         | 0.8936         | 0.723         |
| S11  | 0.2663| 0.0168 | 0          | 0.0168         | 0.2889         | 0.3227        |
| S12  | 0.3677| 0.0494 | 0.0008     | 0.0829         | 1              | 1             |

Table 12

| Stage | Clusters        |
|-------|-----------------|
| 1     | (S2,S8)         |
| 2     | (S6,S9)         |
| 3     | (S2,S8,S4)      |
| 4     | (S6,S9,S10)     |
| 5     | (S5,S11)        |
| Stage | Clusters |
|-------|----------|
| 6     | (S2, S8, S4, S3) |
| 7     | (S1, S2, S8, S4, S3) |
| 8     | (S6, S9, S10, S12) |
| 9     | (S5, S11, S7) |
| 10    | (S1, S2, S8, S4, S3, S5, S11, S7) |
| 11    | (S1, S2, S8, S4, S3, S5, S11, S7, S6, S9, S10, S12) |

Table 13

| Status | Clusters | RS |
|--------|----------|----|
| 1      | (S1, S2, S8, S4, S3) (S6, S9, S10, S12) (S5, S11, S7) | 0.897 |
| 2      | (S6, S9, S10, S12) (S1, S2, S8, S4, S3, S5, S11, S7) | 0.797 |

Table 14

| Clusters   | X1  | X2  | X3  | Y1   | Y2   |
|------------|-----|-----|-----|------|------|
| (S1, S2, S8, S4, S3) | 0.2889 | 0.0285 | 0.0034 | 0.0519 | 0.6248 |
|            | 0.2454 | 0.0176 | 0.0025 | 0.057 | 0.5075 |
|            | 0.227 | 0.0235 | 0 | 0.0611 | 0.4899 |
|            | 0.2312 | 0.0092 | 0.0034 | 0.0544 | 0.5528 |
|            | 0.3492 | 0.0218 | 0.0017 | 0.0586 | 0.4849 |
| Virtual 1  | 0.227 | 0.0092 | 0 | 0.0611 | 0.6248 |
|            | 0.2973 | 0.0201 | 0.0101 | 0.0653 | 0.892 |
|            | 0.2663 | 0.0168 | 0 | 0.0586 | 0.8953 |
| (S6, S9, S10, S12) | 0.3677 | 0.0494 | 0.0008 | 0.0578 | 0.8936 |
|            | 0.3677 | 0.0494 | 0.0008 | 0.0829 | 1 |
| Virtual 2  | 0.2663 | 0.0168 | 0 | 0.0829 | 1 |
|            | 0.2479 | 0.0092 | 0.0008 | 0.0586 | 0.3685 |
| (S5, S11, S7) | 0.2663 | 0.0168 | 0 | 0.0168 | 0.2889 |
|            | 0.4481 | 0.0109 | 0.0042 | 0.0561 | 0.3786 |
| Virtual 3  | 0.2479 | 0.009 | 0 | 0.0586 | 0.3786 |

Table 15

| DMUs | CCR DEA model | Augmented DEA (Appalla [7]) | The heuristic method |
|------|---------------|-----------------------------|---------------------|
| 1    | 0.7498        | 0.4919                      | 0.5773              |
| 2    | 0.919         | 0.6353                      | 0.7455              |
| 3    | 0.6797        | 0.4595                      | 0.541               |
| 4    | 1             | 0.6566                      | 0.8904              |
| 5    | 1             | 0.7071                      | 0.9589              |
| 6    | 0.9615        | 0.6809                      | 0.799               |
| 7    | 0.8173        | 0.5726                      | 0.7766              |
| 8    | 1             | 0.7374                      | 0.8653              |
| 9    | 1             | 0.763                       | 0.8953              |
| 10   | 0.723         | 0.5517                      | 0.6473              |
| 11   | 0.3227        | 0.2462                      | 0.2889              |
| 12   | 1             | 0.6173                      | 0.7244              |

Table 16

| Groups            | Example 1                                                                 | Example 2                                                                 |
|-------------------|---------------------------------------------------------------------------|---------------------------------------------------------------------------|
| 1<sup>st</sup>    | Five efficient DMUs of S1, S2, S3, S5 and S9 (ES = 1).                    | Five efficient DMUs of 4, 5, 8, 9 and 12 (ES = 1).                         |
| 2<sup>nd</sup>    | S8 whose efficiency score is equal to 0.94.                                | Two DMUs of 2 and 6 whose efficiency scores are 0.919 and 0.9615, respectively. |
### Groups

| Group | Example 1                                                                 | Example 2                                           |
|-------|--------------------------------------------------------------------------|-----------------------------------------------------|
| 3rd   | Two DMUs of S4 and S6 whose efficiency scores are 0.66 and 0.68, respectively. | DMU7 whose efficiency score is equal to 0.8173.     |
| 4th   | Two DMUs of S7 and S10 (ES = 0.34).                                      | Two DMUs of 1 and 10 whose efficiency scores are 0.7498 and 0.723, respectively. |
| 5th   | DMU3 whose efficiency score is equal to 0.6797.                           | DMU11 whose efficiency score is equal to 0.3227.    |
| 6th   | DMU11 whose efficiency score is equal to 0.3227.                          |                                                     |

### Biographies

**Marzieh Sadat Rezaei** received her BSc degree in industrial engineering from Iran University of Science and Technology, Iran in 2014. She pursued MSc degree in 2014 and received it industrial engineering from Iran University of Science and Technology in 2016. Her research interests include Data Envelopment Analysis, Performance measurement, Healthcare Operation Management and Supply Chain Network Design.

**Abdorrahman Haeri** is an Assistant Professor of Industrial Engineering at Iran University of Science and Technology, Iran. He received his PhD from the Department of Industrial Engineering, College of Engineering, and University of Tehran, Iran in 2013. He holds an MSc in Industrial Engineering, Sharif University of Technology, Iran in 2008. His main areas of teaching and research interests include data envelopment analysis, network design and data mining. He has published several papers in international conferences and academic journals including Safety Science, Scientometrics, Energy Sources, Part B: Economics, Planning, and Policy, Journal of Engineering Manufacture: Part B, Int. J. Production Research, Int. J. Advanced Manufacturing and Technology, Int. J. Productivity and Quality Management (IJPQM).