Available online at http://ijim.srbiau.ac.ir/
Int. J. Industrial Mathematics (ISSN 2008-5621)
Vol. 10, No. 1, 2018 Article ID IJIM-00876, 11 pages
Research Article

Science and Research Branch (IAU)

Cost Efficiency Measures In Data Envelopment Analysis With Nonhomogeneous DMUs

M. Barat *, G. Tohidi †‡, M. Sanei §

Received Date: 2016-05-21 Revised Date: 2017-06-15 Accepted Date: 2017-08-26

Abstract

In the conventional data envelopment analysis (DEA), it is assumed that all decision making units (DMUs) using the same input and output measures, means that DMUs are homogeneous. In some settings, however, this usual assumption of DEA might be violated. A related problem is the problem of *missing data* where a DMU produces a certain output or consumes a certain input but the values are not available. To address this problem there are some approaches which assign a value (e.g. zero or average of existing values) to the missing data. On the other hand, there are situations where the missing output or input can be produced or consumed by the DMU but for some reasons, an output is not created or the DMU does not have accessibility to an input, hence assigning an artificial value to the nonexistent factor is inappropriate. As some recent studies have focused on addressing the problem of nonhomogeneity among inputs and outputs measures, it has become increasingly important to undrestand its cost structure. This study develops a new DEA methodology to assess cost efficiency (CE) of DMUs in the situation of nonhomogeneous DMUs with different outputs configurations. Via proceeding in three-step procedure both CE scores and subgroup CE scores of DMUs is derived. A numerical example containing a set of 47 steel fabrication plants is used to show the applicability of the model.

Keywords : Cost efficiency; Data Envelopment Analysis; Nonhomogeneous DMUs; (Mono)morphism.

1 Introduction

Data Envelopment Analysis (DEA), initially introduced by Charnes [3], provides a non-parametric linear programming (LP) methodology for evaluating the relative efficiency of each of a set of homogeneous Decision Making Units (DMUs). Homogeneity means that while the amounts of inputs and outputs may vary from one DMU to the other, all DMUs use some common inputs to produce some common outputs in the sense that belonging to the same production technology. Nevertheless, in some settings, although all DMUs use the same technology, the assumption of homogeneity among DMUs is violated. For example, in assessing the efficiency of assurance branches, homogeneity would require all the branches to undertake the same activities. Whilst, largest branches would engage in most of the activities, the smaller one may undertake some of the activities. As another example, consider a case of evaluating different departments of a university, where not all have the same inputs,

*Department of Mathematics, College of Science, Central Tehran Branch, Islamic Azad University, Tehran, Iran.
†Corresponding author. ghatohidi@yahoo.com Tel: +98(912)2951833.
‡Department of Mathematics, College of Science, Central Tehran Branch, Islamic Azad University, Tehran, Iran.
§Department of Mathematics, College of Science, Central Tehran Branch, Islamic Azad University, Tehran, Iran.

and hence they are not homogeneous.

A question that may arise in the absence of homogeneity would be "How to fairly compare a DMU to the others?". A long-existing and related problem is the problem of missing data; certain output is produced or certain input is consumed by a DMU, but its value is not known. Some techniques such as creating a value for the blank entries or using zero as a dummy for the missing output or input have been developed to deal with missing data. In some settings, however, a DMU has chosen not to produce a certain output/consumed a certain input or for some reasons cannot produce/consumed one or some of the outputs/inputs; hence the issue of missing data for certain DMUs change to the concern that the output is not produced or the input is not consumed. As an example, consider a set of universities, where all not have the same departments. Consequently, those universities which do not have the science department cannot directly be compared with those that have it. It is clear that in the cases which a DMU has chosen not to produce or for any reason cannot produce a certain output, substituting a zero value or artificially assigning some other value for the nonexistent entries is not appropriate since that DMU does not consume any resources to produce that output. On the other hand, it is not fair to compare those DMUs produce further outputs with those which produce the smaller one. Same discussion has been hold for the case where the input mix can be different for some DMUs as compared to others. A variety of arguments that have been made to deal with such problems will be discussed in the next section.

In the traditional DEA models, it is usually assumed that unit price and unit cost information have variability or are not available; hence DEA models focus on the technical aspects of production. When information on prices and costs are exactly known, the concept of allocative efficiency (AE) emerges to identify the types of inefficiency. In the current paper, we develop a new cost-based measure for assessing the efficiency of a set of DMUs in a setting where DMUs belong to the same technology but the necessity of the homogeneity among them is relaxed.

The reminder of the paper unfolds as follows. Section 2 gives a brief review of methods aimed at measuring cost efficiency (CE). In addition in this section, the background of non-homogeneity in DEA is studied. Section 3 is devoted to the CE model in the presence of non-homogeneity. A numerical example and managerial implication are discussed in Section 4. Finally, Section 5 concludes with remarks.

2 Literature Review

Many attempts have been made in the past to study the efficiency of a set of DMUs, where the requirement of homogeneity is relaxed. A related problem that has been attracted much attention among researchers is missing data problem. Some approaches have been proposed to cope with the missing data in DEA. Kousmanen [12] set the blank entries of the output and input to be zero and a very large number values respectively to reduce the interference of the units with missing data on the efficiency evaluation of other units. It should be noted that the question of blank output entries is closely related to the dealing with zeros in the data matrices (Thompson [21]). Another technique for dealing with the missing data which is easy but reduces the statistical power and affects the accuracy of estimate parameters such as correlations is the deletion of the missing data (see [15]). Replacement the missing data with suitable values is another trick which is followed by some researchers. Smirlis et al. (2006), proposed using the interval DEA, through which missing data are substituted by a lower and an upper bound, and as a result, the efficiency bound can be estimated. Zha [26] developed modified DEA models to estimate the appropriate value of missing data in its interval, based on DEA and Inter-dimensional Similarity Halo Effect and determined the estimated value of missing data by the General Impression of original DEA efficiency.

Cook et al. [5] argued that there is a difference between the situation where a DMU commits the resources to produce an output but fails to do so, or else a non-zero amount exists but is unknown and the situation where the DMU intentionally does not produce that output. They developed a DEA model structure to address certain types of such problems, wherein a subset of DMUs produce one less than the full set of outputs produced by the others. They proposed three steps to obtain the efficiency of each DMU. They first split the outputs into some subgroups and then

Table 1: Product line by DMU group.

	Outputs					
Group	O1	O2	O3	O4	O5	O6
N_1	x	x	x		x	
N_2		x	x	x	x	x
N_3			x		x	x
N_4	x		x		x	

in step 1 the inputs are split between the output subgroups. In step 2, a standard DEA analysis is carried out to each of the output subgroups and its dedicated inputs and of course those DMUs which produce that subgroup. In step 3 the overall efficiency score of each DMU is derived by taking a weighted average of the subgroup scores obtained in step 2. Cook et al. [4] extended the proposed models to the general settings where a set of DMUs produces a certain set of products, but not all products are produced by all DMUs. They then use the models to examine the efficiencies of a set of manufacturing plants. Li et al. [14] investigated the problem of lack of homogeneity on the input side and extended the earlier researches of Cook et al. [5, 4] to cover the case where different input configurations across a set of DMUs. They developed a DEA-based methodology to deal with this situation and applied it to a set of 31 provinces in China in which one of the inputs is the quantity of natural resources available to the region and not all regions have the same natural resources.

An important concept in applied production analysis of organizations which can be traced back to Farrell [8] and Debreu [7] is measuring the CE of the companies. Fare et al. [9] developed these concepts into implementable forms of CE in DEA. Tone [22] identified the shortcomings of the cost and allocative efficiency as used in DEA literature. In response, a new scheme to the CE is proposed by Tone [22] and extended to decomposition of CE by Tone and Tsutsui [25]. CE was also encountered by Tone and Sahoo [24] when examining the performance of Life Insurance Corporation of India, and by Tone and Sahoo [25] in discussing the issue of cost elasticity. Fukuyama and Weber [10] developed a variant of the CE model based on the framework of the directional distance function (DDF) by Chamber et al. [1, 2]. Sahoo et al. [18] extended the models of Tone [22] to new directional measures of value-based CE. For more information about the different aspects of CE, the enthusiastic reader would referred to Ray and Kim [17], Cooper et al. [6], Kousmanen and Post [13], Sengputa and Sahoo [19], Jahanshahloo et al. [11], and Mostafaee and Saljooghi [16] among others.

To the best of authors' knowledge, there is not any reference that discusses CE in the presence of non-homogeneity. In the next section, a new methodology is developed for dealing with the problem of measuring the CE of a set of DMUs where the requirement of homogeneity on the output side is relaxed.

3 A cost model for Nonhomogeneous DMUs with different outputs configuration

Consider a situation where a set of n DMUs produces a certain kind of products but not all of the products are produced by all the DMUs. Similar to the definitions of Cook et al. [4], suppose that the DMUs are organized into P mutually exclusive groups denoted by $\{N_p\}_{p=1}^{P}$ wherein those in N_k produce some different products as those in N_t with all groups using the same inputs. Then constitute mutually exclusive output subgroups R_k, $k = 1, \ldots, K$, where R_k represents the subset of outputs with the property that all of its members appear as the outputs of exactly the same set of DMUs. More precisely, if outputs $r_1, r_2 \in R_k$, then the DMU profiles of these two outputs are identical. For example consider a set of DMUs with four inputs and different outputs configuration. DMUs have been grouped toghether into 4 DMU groups, N_1, N_2, N_3 and N_4. For the men-

Table 2: Data on 47 Plants-Outputs.

	Sheet Steel	Flat Bar	Pipes/Cylinders	Ducts	Structural Steel	Storage Tanks
DMU	**O1**	**O2**	**O3**	**O4**	**O5**	**O6**
1	70	103	100	-	60	-
2	-	125	90	123	48	133
3	50	110	105	-	170	-
4	80	80	110	-	82	-
5	-	-	60	-	100	150
6	40	95	120	-	151	-
7	100	-	200	-	64	-
8	-	-	180	-	104	66
9	65	150	125	-	93	-
10	40	110	70	-	79	-
11	70	117	122	-	132	-
12	-	-	89	-	80	189
13	88	-	57	-	150	-
14	48	-	146	-	162	-
15	-	-	220	-	111	73
16	99	-	89	-	56	-
17	-	-	88	-	41	161
18	-	55	132	129	112	113
19	80	97	142	-	82	-
20	97	-	209	-	106	-
21	-	-	55	-	157	130
22	-	-	93	-	163	55
23	59	-	218	-	79	-
24	61	-	58	-	75	-
25	68	-	110	-	48	-
26	-	-	86	-	109	69
27	-	65	166	41	183	137
28	-	-	228	-	199	71
29	-	-	95	-	110	54
30	50	-	77	-	89	-
31	-	138	206	68	102	74
32	36	106	167	-	130	-
33	-	84	98	45	176	69
34	-	62	120	57	58	154
35	24	135	185	-	112	-
36	-	-	144	-	196	78
37	58	-	178	-	147	-
38	-	123	206	63	195	57
39	41	110	225	-	53	-
40	-	-	188	-	60	127
41	70	-	140	-	150	-
42	-	-	55	-	70	191
43	45	-	124	-	139	-
44	63	-	161	-	125	-
45	85	-	81	-	90	-
46	42	78	69	-	82	-
47	25	-	184	-	162	-

Table 3: Data on 47 Plants-Iutputs.

	Labor	Shears	Presses	Torches
DMU	**X1**	**X2**	**X3**	**X4**
1	30	5	3	15
2	40	4	6.5	18
3	35	5.2	4.2	10
4	38	7	7.6	9
5	28	9	5.5	13
6	37	4.2	3.8	17
7	31	6	4.1	11
8	35	5	7	15
9	25	6.2	4.8	19
10	30	3	3.2	21
11	25	4	6	12
12	45	5	3.3	23
13	35	4.1	5	25
14	32	5.3	3.5	11
15	26	7.7	4.3	16
16	19	5.3	6.2	12
17	25	8	3	9
18	32	6	2.8	7
19	33	2.8	3.9	13
20	27	3.3	4.3	22
21	25	7.9	5	16
22	34	5	5.4	20
23	45	4	4.1	12
24	24	5.1	3.4	19
25	33	8.6	2.7	10
26	21	9.8	5.5	5
27	25	7	3.1	23
28	38	4.5	2.4	10
29	33	3.2	4.6	24
30	27	6.4	3	7
31	20	5.8	5.1	18
32	39	8.4	3.8	16
33	42	6.5	2.4	8
34	44	4.3	3	22
35	26	3.7	6.7	20
36	43	7.5	7.1	8
37	35	6.8	4.7	14
38	22	3.9	3.2	25
39	41	6.7	2.5	21
40	21	5.2	4.9	10
41	33	3.5	5.9	7
42	20	4.7	7.2	23
43	39	9.5	5.4	6
44	48	6.7	6.2	18
45	31	3.6	4.7	23
46	28	9.2	2.6	5
47	36	7.6	5.7	10

Table 4: Output Subgroup Scores, Proposed CE Scores, Conventional CE Scores.

DMUs	K=1	K=2	K=3	K=4	K=5	Proposed CE	Conventional CE
1	0.63213	0.52381	0.48890			0.55	0.85
2		0.58803	0.59539	0.32459	0.57322	0.52	0.96
3	0.71880	0.80883	0.35158			0.63	0.99
4	0.56706	0.39402	0.49219			0.48	0.75
5			0.48096		0.22422	0.35	0.89
6	0.48891	0.54445	0.29779			0.44	0.77
7	1		0.31805			0.66	1
8			0.19694		0.21286	0.20	0.69
9	0.97383	0.44943	0.60264			0.68	1
10	0.59538	0.30019	0.49182			0.46	0.71
11	0.76252	0.86084	0.53541			0.72	1
12			0.28015		0.20551	0.24	0.74
13	0.54690		0.25741			0.40	0.80
14	0.45321		0.32306			0.39	0.97
15			0.26945		0.28494	0.28	0.92
16	0.68770		0.47458			0.58	1
17			0.44302		0.29682	0.37	1
18		0.69355	1	0.36841	0.78568	0.71	1
19	0.75133	0.55730	0.53653			0.62	0.95
20	0.85648		0.31725			0.59	1
21			0.46649		0.24007	0.35	0.98
22			0.20785		0.16198	0.18	0.70
23	0.52133		0.25735			0.39	0.81
24	0.49082		0.18085			0.34	0.65
25	0.65321		0.16772			0.41	0.63
26			0.31296		0.20175	0.26	0.81
27		1	0.44524	0.50746	0.37076	0.58	1
28			0.30074		0.24119	0.27	1
29			0.13707		0.16337	0.15	0.49
30	0.56609		0.21091			0.39	0.74
31		0.57896	0.52460	1	1	0.78	1
32	0.32810	0.55114	0.31424			0.40	0.72
33		0.83971	0.30181	0.54941	0.35395	0.51	0.85
34		0.64678	0.47375	0.28620	0.26253	0.42	0.78
35	0.30246	0.87839	0.35050			0.51	0.89
36			0.26869		0.19428	0.23	0.84
37	0.42601		0.28475			0.36	0.85
38		0.80485	0.35177	0.83741	0.69928	0.67	1
39	0.38212	0.71151	0.31801			0.47	0.80
40			0.40749		0.41722	0.41	1
41	0.73216		0.29877			0.52	1
42			0.34047		0.28863	0.31	1
43	0.38092		0.23203			0.31	0.73
44	0.36291		0.19260			0.28	0.60
45	0.59172		0.16944			0.38	0.70
46	0.70935	0.35978	0.38292			0.48	0.688
47	0.19111		0.30479			0.25	0.822

tioned DMU profles the K output sets are

$$R_1 = \{1\},\ R_2 = \{2\},\ R_3 = \{3, 5\},$$

$$R_4 = \{4\},\ R_5 = \{6\}.$$

Cook et al. [5] claimed that for evaluating the efficiency of a given DMU (say j_o) three stages should be preceded. In stage 1, an LP should be run to decide what portion of each input i will be allocated to each of the output subgroups R_k. In fact, they proposed the following model

$$\begin{array}{ll}
\max & \sum_{R_k \in L_{N_{p^o}}} \sum_{r \in R_k} \mu_r y_{rj_o} \\
\text{s.t.} & \sum_{R_k \in L_{N_{p^o}}} \left(\sum_i \gamma_{iR_k p^o} x_{ij_o} \right) = 1, \\
& \sum_{r \in R_k} \mu_r y_{rj} - \sum_i \gamma_{iR_k p} x_{ij} \leq 0, \\
& \quad \forall j \in N_p, R_k \in L_{N_p}, p = 1, \ldots, P, \\
& \sum_{R_k \in L_{N_p}} \gamma_{iR_k p} = \upsilon_i, \\
& \quad \forall i, p = 1, \ldots, P, \\
& \upsilon_i a_{iR_k p} \leq \gamma_{iR_k p} \leq \upsilon_i b_{iR_k p}, \\
& \quad \forall i, ; R_k \in L_{N_p}, ; p = 1, \ldots, P, \\
& \mu_r, \upsilon_i, \gamma_{iR_k p} \geq \varepsilon, \\
& \quad \forall i, r, R_k, p = 1, \ldots, P,
\end{array} \tag{3.1}$$

where L_{N_p} denotes those R_k forming the full output set for any DMU in N_p.

$\alpha_{iR_k p^o} = \gamma_{iR_k p^o}/\upsilon_i$, where $\alpha_{iR_k p^o}$ is the proportion of input i to be assigned to output subgroup R_k. Parameters $a_{iR_k p}$ and $b_{iR_k p}$ place lower and upper limits on the size of alpha variables. Stage 2 is dedicated to the evaluation of the efficiency of the DMU in terms of each of its subgroups R_k using the conventional CCR DEA model, and in stage 3 a weighted average of these subgroup scores is taken to get the overall efficiency of the DMU.

As is mentioned in Section 2, Tone [22] revealed that the traditional Farrell-Debereu CE measure suffers from a serious shortcoming that is caused by the supposed production possibility set (PPS). Tone [22] defined a new PPS which considers unit inputs cost in addition to the technical factors $X = (x_1, \ldots, x_m) \in R^{m \times n}$ and $Y = (y_1, \ldots, y_s) \in R^{s \times n}$. In fact, the proposed PPS is;

$$P_c = \left\{ (\bar{x}, y) | \, \bar{x} \geq \bar{X}\lambda, y \leq Y\lambda, \lambda \geq 0 \right\},$$

where,

$$\bar{X} = (\bar{x}_1, \ldots, \bar{x}_m)$$

with

$$\bar{x}_j = (c_{1j} x_{1j}, \ldots, c_{mj} x_{mj})^T.$$

Considering the new cost-based PPS, Tone [22] defined the new CE measure as $\bar{\gamma}^* = e\bar{x}_o^*/e\bar{x}_o$, where $e \in R^m$ is a row vector with all elements being equal to 1, and $\bar{x}_o^*$ is the optimal solution of the following cost model,

$$\begin{array}{ll}
\min & e\bar{x} \\
\text{s.t.} & \bar{x} \geq \bar{X}\lambda, \\
& y_o \leq Y\lambda \\
& \lambda \geq .
\end{array} \tag{3.2}$$

The dual form of (3.2) is expressed with non-negative vectors $v = (v_1, \ldots, v_m)$, and $u = (u_1, \ldots, u_s)$ of variables as follows,

$$\begin{array}{ll}
\max & uy_o \\
\text{s.t.} & v \leq e, \\
& -v\bar{X} + uY \leq 0, \\
& v, u \geq 0.
\end{array} \tag{3.3}$$

In models (3.2) and (3.3) it is assumed that a set of homogeneous DMUs is under evaluation. Now, to demonstrate how to consider non-homogeneity in the model, a new methodology is presented. Using the same notation as Cook et al. [4], suppose a set of n DMUs, which fall into P mutually exclusive groups, and their outputs fall into R_k, $k = 1, \ldots, K$ mutually exclusive output subgroups, both described in this section. Moreover, assume that information on costs is known exactly, with the presumption that the costs from DMU to DMU are not necessarily the same. To evaluate the CE of a given DMU, we proceed in three stages.

Stage 1: In this stage, we propose a cost-based model for deriving the split of each input i to be allocated to each of the output subgroups $R_k \in L_{N_p}$; this proportion is denoted by $\alpha_{iR_k p}$. To do so, we revised the cost model 3.2 in a manner that the objective and constraints organized according to the output subgroups. In fact, the following model captures the idea,

$$\begin{aligned}
\max \quad & \sum_{R_k \in L_{N_{p^o}}} \sum_{r \in R_k} u_r y_{ro} \\
\text{s.t.} \quad & \sum_{R_k \in L_{N_p}} \Big(\sum_{r \in R_k} u_r y_{rj} - \sum_i v_i \alpha_{iR_kp} \bar{x}_{ij} \Big) \le 0, \\
& \quad \forall j \in N_p,\ p = 1, \ldots, P, \\
& v_i \le 1, \quad \forall i \\
& \sum_{R_k \in L_{N_p}} \alpha_{iR_kp} = 1, \quad \forall i,\ p = 1, \ldots, P, \\
& a_{iR_kp} \le \alpha_{iR_kp} \le b_{iR_kp}, \\
& \quad \forall i,\ R_k,\ p = 1, \ldots, P, \\
& u_r,\ v_i,\ \alpha_{iR_kp} \ge 0, \quad \forall i,\ R_k,\ p.
\end{aligned} \tag{3.4}$$

The first type constraints require that the multipliers are chosen for a DMU_o satisfy the condition that when they are applied to any other DMU, the corresponding constraint does not exceed zero. In fact, if we consider the corresponding efficiency ratio, (of outputs to inputs) this ratio should not exceed unity. The third-type constraints denote that the resource splitting variables values, α_{iR_kp}, assigned to output subgroups corresponding to any set p, sum to unity for each i. The fourth-type constraints set the upper and lower bound on the α variables.

Now, it should be emphasized that we can manipulate the first-type constraints and impose the efficiency condition to any output subgroup. Specifically, we can impose a type of constraints such that the resource-splitting variables, α_{iR_kp}, be selected in a manner that allows the efficiency ratio corresponding to the outputs in R_k does not exceed unity. In other words, we can split the first-type constraints of model 3.2, for each of the output subgroups. Specifically, we propose the following cost-based model,

$$\begin{aligned}
\max \quad & \sum_{R_k \in L_{N_{p^o}}} \sum_{r \in R_k} u_r y_{ro} \\
\text{s.t.} \quad & \sum_{r \in R_k} u_r y_{rj} - \sum_i v_i \alpha_{iR_kp} \bar{x}_{ij} \le 0, \\
& \quad \forall j \in N_p,\ R_k, p = 1, \ldots, P, \\
& v_i \le 1, \quad \forall i \\
& \sum_{R_k \in L_{N_p}} \alpha_{iR_kp} = 1, \quad \forall i,\ p = 1, \ldots, P, \\
& a_{iR_kp} \le \alpha_{iR_kp} \le b_{iR_kp}, \\
& \quad \forall i,\ R_k,\ p = 1, \ldots, P, \\
& u_r,\ v_i,\ \alpha_{iR_kp} \ge 0, \quad \forall i,\ R_k,\ p.
\end{aligned} \tag{3.5}$$

Model (3.5) is nonlinear through the product of v_i, and α_{iR_kp}. Making the change of variables $\gamma_{iR_kp} = v_i \alpha_{iR_kp}$, problem (3.5) becomes,

$$\begin{aligned}
\max \quad & \sum_{R_k \in L_{N_{p^o}}} \sum_{r \in R_k} u_r y_{ro} \\
\text{s.t.} \quad & \sum_{r \in R_k} u_r y_{rj} - \sum_i \gamma_{iR_kp} \bar{x}_{ij} \le 0, \\
& \quad \forall j \in N_p,\ R_k, p = 1, \ldots, P, \\
& v_i \le 1, \quad \forall i \\
& \sum_{R_k \in L_{N_p}} \gamma_{iR_kp} = v_i, \quad \forall i,\ p = 1, \ldots, P, \\
& v_i a_{iR_kp} \le \gamma_{iR_kp} \le v_i b_{iR_kp}, \\
& \quad \forall i,\ R_k,\ p = 1, \ldots, P, \\
& u_r,\ v_i,\ \gamma_{iR_kp} \ge 0, \quad \forall i,\ R_k,\ p.
\end{aligned} \tag{3.6}$$

Stage 2: After applying stage 1, the optimal proportion of inputs, for each DMU_o, to be assigned to each output subgroup is derived. These proportions are given by $\alpha_{iR_kp^o} = \gamma_{iR_kp^o}/v_i$. Using these proportions, we can assign the acceptable proportion of inputs to each output subgroup R_k. Then, the cost model (3.5) can be applied to each subgroup R_k of DMU_o. Specifically, the following cost model should be solved for each DMU_o, and each subgroup R_k of DMU_o corresponding to the set N_{P^o} that contains o as a member;

$$\begin{aligned}
\max \quad & \sum_{r \in R_{k^o}} u_r y_{ro} \\
\text{s.t.} \quad & v_i \le 1, \quad \forall i \\
& \sum_{r \in R_{k^o}} u_r y_{rj} - \sum_i v_i \hat{x}_{ij}^{k^o} \le 0, \\
& \quad j \in N_p, N_p \in M_{R_k}, \\
& u_r, v_i \ge \varepsilon,
\end{aligned} \tag{3.7}$$

where, $M_{R_k} = \{N_p \text{ such that } R_k \in L_{N_p}\}$, and $\hat{x}_{ij}^{k^o} = \alpha_{iR_kp^o} \bar{x}_{ij}$.

Stage 3: In this stage, the CE score of DMU_o is derived by the arithmetic mean of subgroups scores obtained in stage 2.

Theorem 3.1 *A DMU can be cost efficient if and only if all of its output subunit subgroups are cost efficient as well.*

Proof. On the contrary, assume that a DMU is cost efficient and at least one of its output subgroups (subgroup t) is not cost efficient (specifically, if the objective of stage 2 is denoted by e_{R_k}, $k = 1, \ldots, K$, $e_{R_t} < 1$). According to the proposed methodology, $e_o = \sum_{R_k \in L_{N_p}} e_{R_t}/|L_{N_p}|$, and $e_{R_k} \leq 1$. Since $e_{R_t} < 1$ then $e_o < 1$, which violates the being assumption of the efficiency of DMU$_o$, hence all of the output subgroups are cost efficient. On the other hand, suppose that all of the output subgroups of under-evaluating DMU are cost efficient. Since the efficiency of the DMU is the weighted average of these efficient subgroups, the cost efficiency score of DMU is one, and the DMU is cost efficient. This completes the proof.

In the next section, we apply the proposed model to a data set originates from a study by Cook et al. [4].

4 Numerical example

In this section, a set of 47 steel fabrication plants is considered in the evaluation process. The steel plants data is derived from a study by Cook et al. [4]. The production is characterized by 6 outputs; Sheet steel products (ladders, guards, bumpers, and conveyors); Flat bar products used mainly in building construction (brackets, base plates, headers, and posts); Pipes and cylinders (storm drains, plumbing products, etc.); Furnace and air conditioning ducts; Structural steel (e.g., joists and support beams); and Tanks (residential and industrial), and four inputs; plant labor; shears and saws; presses and rolling equipment; and cutting torches and welding equipment. Appendix, Table 2 and Table 3, present the data for the 47 steel fabrication plants.

It should be noted that some plants have chosen not to produce certain products. Plants with the same product lines have been grouped together into four DMU groups, N_1 to N_4. As can be observed, plants in N_1 manufacture products 1, 2, 3, 5; those in N_2 produce products 2, 3, 4, 5, 6; and those in N_3 and N_4 make products 3, 5, 6, and 1, 3, 5, respectively.

Before following the methodology presented in the previous section it should be noted that after applying model (3.6), an appropriate allocating of the inputs to the subunits that make up the DMU is at hand. As what is mentioned in Cook et al. [4] survey data from a sample of the plants suggested the following ranges to bound the values of α :

$$\begin{aligned} N_1 &: (0.15,\ 0.80) \\ N_2 &: (0.10,\ 0.60) \\ N_3 &: (0.20,\ 0.90) \\ N_4 &: (0.20,\ 0.90) \end{aligned}$$

It can be easily interpreted from the given ranges that the more subunits that N_p contains, the narrower are the ranges.

Additionally, we consider $c_{ij} = 1, i = 1, \ldots, 4, j = 1, \ldots, 47$. Applying model (3.6) to the data of Tables 2 and 3, the values of α_{iR_kp} for each DMU in N_p, $p = 1, \ldots, 4$, have been derived. As is mentioned earlier, α_{iR_kp} are determined to allocate to subgroup R_k the appropriate amount of input x_{ij_o}. Then, using the appropriately adjusted data, Model (3.7) is applied to each DMU in M_{R_k}. To derive a CE score for each DMU $j_o \in N_p$, $p = 1, \ldots, 4,$, the relevant subunit scores are combined using the arithmetic mean of the subunit scores. The resulting cost scores are presented along with their relevant subunit scores in Appendix, Table 4.

The seventh column of Table 4 shows the CE scores derived by the proposed methodology. None of the DMUs are rendered CE efficient by the proposed models, since a DMU is CE efficient if and only if it is CE efficient in all of its output subgroups. Results show that DMU31 with 0.78 is designated the highest CE score and DMU29 with 0.15 received the lowest CE score by the proposed models.

It is worth comparing the CE results obtained using the proposed model with what would have carried out by conventional CE model, model (3.2), by inserting zeros in the data for any missing output. The last column of Table 4, shows the CE scores, derived by applying the conventional CE model. It should be noted that by replacing all blank spaces with zeros, 14 out of 47 DMUs are rendered cost efficient. On the other hand, all the CE scores derived by the proposed model are less than the conventional CE scores. Moreover, DMU29 which has the lowest CE score by the proposed method, has the lowest CE score, 0.49, by the conventional CE model and DMU31 which has the highest CE score by the new methodology is CE efficient by the conventional CE model, as well.

5 Conclusion

This paper has dealt with the measurement of the CE in a setting where DMUs are nonhomogeneous. In such situation, the usual assumption in DEA where each DMU produces the same output measures although in a differing amount from one DMU to the next is violated. It is worth noting that this problem differs from the problem of 'missing' outputs where the missing item exists but its value is not available to the analyst, or that the missing item can be created by the DMU but for some reasons, none was produced.

In this paper, we develop a new methodology for CE analysis dealing with the situation where the usual assumption of homogeneity does not hold. To address this problem we need to proceed in three stages, where the DMUs are viewed as consisting of subunits which the DMUs in each subunit produce the same outputs. Numerical example containing a set of 47 steel fabrication plants is used to show the applicability of the model.

References

[1] R. G. Chambers, Y. Chung, R. Fare, Benefit and distance functions, *Journal of Economic Theory* 70 (1996) 407-418.

[2] R. G. Chambers, Y. Chung, R. Fare, Profit, distance functions and Nerlovian efficiency, *Journal of Optimization Theory and Applications* 98 (1998) 351-364.

[3] A. Charnes, W. W. Cooper, E. Rhodes, Measuring the efficiency of decision making units,*European Journal of Operational Research* 2 (1978) 429-444.

[4] W. D. Cook, J. Harrison, R. Imanirad, P. Rouse, J. Zhu, Data Envelopment Analysis with Nonhomogeneous DMUs, *Operations Research* 61 (2013) 666-676.

[5] W. D. Cook, J. Harrison, P. Rouse, J. Zhu, Relative efficiency measurement: The problem of a missing output in a subset of decision making units, *European Journal of Operations Research* 220 (2012) 79-84.

[6] W. W. Cooper, R. G. Thompson, R. M. Thrall, Extensions and new developments in DEA, *Annals of Operations Research* 66 (1996) 1-45.

[7] G. Debreu, The coefficient of resource utilization, *Econometrica* 19 (1951) 273-292.

[8] M. J. Farrell, The measurement of productive efficiency, *Journal of the Royal Statistical Society Series AI* 120 (1957) 253-281.

[9] R. Fare, S. Grosskopf, C. A. K. Lovell, Measurement of efficiency of production, *Kluwer, Boston*, (1985).

[10] H. Fukuyama, W. L. Weber, Economic inefficiency measurement of input spending when decision-making units face different input prices, *Journal of Operational Research Society* 55 (2004) 1102-1110.

[11] G. R. Jahanshahloo, M. Soleimani-damaneh, A. Mostafaee, A Simplified version of the DEA cost efficiency model, *European Journal of Operational Research* 184 (2008) 814-815.

[12] T. Kuosmanen, Modeling black data entries in Data Envelopment Analysis, *Working Paper, Wageningen University, Netherlands*, (2002).

[13] T. Kuosmanen, T. Post, Measuring economic efficiency with incomplete price information: With an application to European commercial banks, *European Journal of Operational Research* 134 (2001) 43-58.

[14] W. H. Li, L. Liang, W. D. Cook, J. Zhu, DEA Models for Non-Homogeneous DMUs with Different Input Configurations, *European Journal of Operations Research* 254 (2016) 946-956.

[15] R. J. A. Little, D. B. Rubin, Statistical analysis with missing data, *Wiley, Canada, second ed.* (2002).

[16] A. Mostafaee, F. H. Saljooghi, Cost efficiency measures in data envelopment analysis with data uncertainty, *European Journal of Operational Research* 202 (2010) 595-603.

[17] S. C. Ray, H. J. Kim, Cost efficiency in the United-States Steel-Industry a nonparametric analysis using data envelopment analysis,

European Journal of Operational Research 8 (1995) 654-671.

[18] B. K. Sahoo, M. Mehdiloozad, K. Tone, Cost, revenue and profit efficiency measurement in DEA: A directional distance function approach, *European Journal of Operational Research* 237 (2014) 921-931.

[19] J. K. Sengupta, B. K. Sahoo, Efficiency models in data envelopment analysis: Techniques of evaluation of productivity of firms in a growing economy, *Palgrave Macmillan, London*, (2006).

[20] Y. G. Smirlis, E. K. Maragos, D. K. Despotis, Data envelopment analysis with missing values: An interval DEA approach, *Applied Mathematics and Computation* 177 (2006) 1-10.

[21] R. G. Thompson, P. S. Dharamapala, R. M. Thrall, Importance for DEA of zeros in data, multipliers, and solutions, *Journal of Productivity Analysis* 4 (1993) 370-390.

[22] K. Tone, A strange case of the cost and allocative efficiencies in DEA, *Journal of the Operational Research Society* 53 (2002) 1225-1231.

[23] K. Tone, B. K. Sahoo, Cost-elasticity: A reexamination in DEA, *Annals of Operations Research (forthcoming)*, (2005b).

[24] K. Tone, B. K. Sahoo, Evaluating cost efficiency and returns to scale in the life assurance corporation of India using data envelopment analysis, *Socio-Economic Planning Sciences* 39 (2005a) 261-285.

[25] K. Tone, M. Tsutsui, Decomposition of cost efficiency and its application to Japan-US electric utility comparisons, *Research Report Series I, GRIPS,* (2004).

[26] Y. Zha, A. Song, C. Xu, H. Yang, Dealing with missing data based on data envelopment analysis and halo effect, *Applied Mathematical Modelling* 37 (2013) 6135-6145.

Mona Barat is a faculty member in the Department of Mathematics, Islamic Azad University, Mahshahr Branch, Mahshahr, Iran. She is also a PhD candidate in the department of Mathematics, Islamic Azad University, Central Tehran Branch, Tehran, Iran. Her research interests include Data Envelopment Analysis and Operations Research.

Ghasem Tohidi is Associate Professor in Mathematics department at the Islamic Azad University Central Tehran Branch in Iran. His teaching has been in operation research and data envelopment analysis (DEA). His major research interests are multi objective programming and DEA. His publications have been appeared in several journals, including Computers Industrial Engineering, Journal of the Operational Research Society, Applied Mathematics and Computation, Applied Mathematical Modelling, Journal of the Operational Research Society of Japan, International Journal of Intelligent Systems, Journal of Intelligent Fuzzy Systems, and Asia-Pacific Journal of Operational Research, among others.

Masoud Sanei is an Associate Professor in the Department of Applied Mathematics, Islamic Azad University, Central Tehran Branch, Tehran, Iran. He received his PhD degree in Applied Mathematics (Operations Research) from Islamic Azad University, Science and Research Branch, Tehran, Iran, in 2004. His research interests are in the areas of operation research such as data envelopment analysis and supply chain management. He has several papers in journals and conference proceedings.