

---

## Performance assessment of decision making units in different levels using a hierarchical clustering DEA method

---

B. Dalvand, \*F. Hosseinzadeh Lotfi, G. R. Jahanshahloo

Department of Mathematics, Science and Research Branch, Islamic Azad University, Tehran, Iran

Received: 9<sup>th</sup> Feb 2017, Accepted: 22<sup>nd</sup> Feb 2017, Published: 15<sup>th</sup> April 2017

### Abstract

Data envelopment analysis (DEA) is a performance assessment technique which mathematically provide a covering frontier referred to as the efficiency frontier. In DEA theory the units are supposed to be homogenous and comparable, whereas, this assumption is not always hold in practical issues when there is high-level units. In this paper, to have a fair assessment and augment the competition between the Decision Making Units (DMUs) we provide an algorithm based on hierarchical clustering and evaluate DMUs in difference levels. Finally, we apply suggested procedure to evaluate the performance of Iranian commercial banks.

**Key Words:** Data envelopment analysis, Hierarchical agglomerative clustering, Data mining

### 1. Introduction

Data Envelopment Analysis (DEA) as a non-parametric efficiency measurement method was first described in Charnes et. al. (1978) and then extended by Banker et. al (1984), for performance assessment of a set of peer Decision Making Units (DMUs) each of which uses multiple inputs to produce multiple outputs. In some cases there is no global standard for evaluating the performance of the under evaluated firms, if any, may be far from the realities of society under evaluation. In fact, some existing absolute standards adopted by advanced countries may be defined taking into account their own society. While, measuring a society in a developing country according to these standards may led to very poor or even disappointing efficiency score. Accordingly, one of the strengths of DEA is calculating the relative efficiency of DMUs. DEA technique, occasionally called frontier analysis, establishes a covering frontier called efficiency frontier considering a Production Possibility Set (PPS). The DMUs located on the frontier as the standards of performance are defined efficient in a sense of DEA that DMUs not on the efficient frontier could try to achieve. Classical DEA models assign an efficiency score to inefficient units according to their distances to the efficient frontier. Actually, the main aim of calculating the relative efficiency is to create a competitive environment between the members of the society underevaluation. While in practical issues we may face a case which the selected PPS contains extraordinary units that

because of having access to favorable technical, economic, political and technical conditions, are known efficient with a much higher efficiency level than the others. So, it is difficult or even impossible for some inefficient units to achieve such high-level units. On the other hand, considering the new PPS after eliminating these high-level units led to obtaining optimistic score efficiency for remaining units. It seems that a fair strategy in such cases is that inefficient DMUs improve their performance hierarchically. It means, addition to taking into account the best of society, an inefficient DMU be evaluated compared with units which located on lower levels and are more compatible with it. To handle this situations, Cooper et. al (2002) suggested a hierarchical category so that all DMUs are divided into several categories based on desired criteria and then each DMU is evaluated comparing with DMUs in the same and higher category levels. Alirezaee and Mir-Hassani (2006) presented an algorithm and partition all DMUs into some finite disjoint sets corresponding to different frontiers. The main contribution of current paper is to augment the competition between the units by partitioning all DMUs into different levels in terms of size. To do so, we implement hierarchical agglomerative clustering as one of the commonly clustering technique in data mining and evaluate each unit comparing with members of the clusters that it belongs to.

The outline of the paper is organized as follows. In Section 2, we present an introduction to data envelopment analysis. Section 3 suggests a new method based on hierarchical agglomerative clustering for measuring the efficiency of DMUs in different levels. In section 4, we illustrate the proposed approach with an application to Iranian commercial banks. The paper ends with some conclusions and future research directions in Section 5.

### 2. Data envelopment analysis

Consider a set of  $n$  DMUs which  $DMU_j$  ( $j=1, \dots, n$ ) consumes the input vector  $X_j=(x_{1j}, \dots, x_{mj})$  to produce the output vector  $Y_j=(y_{1j}, \dots, y_{sj})$ . In DEA literature, Production Possibility Set (PPS) or Production Technology is a set of all input-output vector  $(X, Y)$  so that the output vector  $Y$  can be produced by the input vector  $X$ . The PPS spanned by all DMUs is defined as

$$PPS = \left\{ (X, Y) \mid X \geq \sum_{j=1}^n \lambda_j X_j, Y \leq \sum_{j=1}^n \lambda_j Y_j, \lambda_j \geq 0, j = 1, \dots, n \right\}$$

PPS provides a covering frontier called efficiency frontier. Any DMU located on the efficiency frontier is defined “efficient”, otherwise, it is defined “inefficient” (Charnes et al. (1978)).

In order to estimate the relative efficiency of DMU<sub>p</sub> (X<sub>p</sub>, Y<sub>p</sub>), we solve the following slacks-based measure (SBM) model proposed by Tone (2001) :

$$\begin{aligned} \min \quad & \theta = \frac{1 - \frac{1}{m} \sum_{i=1}^m \frac{s_i^-}{x_{ip}}}{1 + \frac{1}{s} \sum_{r=1}^s \frac{s_r^+}{y_{rp}}} \\ \text{s.t.} \quad & \sum_{j=1}^n x_{ij} \lambda_j + s_i^- = x_{ip}, \quad i = 1, \dots, m \\ & \sum_{j=1}^n y_{rj} \lambda_j - s_r^+ = y_{rp}, \quad r = 1, \dots, s \\ & \lambda_j \geq 0, \quad j = 1, \dots, n \\ & s_i^- \geq 0, \quad i = 1, \dots, m \\ & s_r^+ \geq 0, \quad r = 1, \dots, s \end{aligned} \tag{1}$$

For each DMU<sub>p</sub> we have  $\theta \leq 1$ . DMU<sub>p</sub> is an efficient unit if and only if  $\theta = 1$ , otherwise, DMU<sub>p</sub> is inefficient.

**3. The proposed algorithm**

Despite accepting the DEA basic assumption of homogeneity for DMUs in input and output indexes, heterogeneity in size and activities cause unfair evaluation. The problem show up when high-level units on the efficient frontier are suggested as reference set for inefficient units. So, practically, it is difficult or even impossible for some inefficient ones to develop their performance, depending on how close they are to the efficient frontier.

In this section, to reduce the impact of this heterogeneity on evaluation results, we suggest an algorithm based on hierarchical clustering. Hierarchical clustering is a method of cluster analysis in data mining and statistics which are subdivided into agglomerative (bottom-up) and divisive (top-down) methods. Hierarchical agglomerative clustering treats each observation as a singleton cluster of its own at the outset and then consecutively the two closest clusters are merged into a new combined cluster until all clusters have been aggregated into a single cluster

that contains all observations. Hierarchical divisive clustering begin with all observation as the first cluster and then in succeeding steps, with the most dissimilar clusters being divided into finer groupings until individual observations are reached. Here, we consider agglomerative one as the most commonly used technique in computer programs that apply hierarchical clustering. Our proposed procedure for *n* DMUs with *m* inputs and *s* outputs can be summarized in the following three steps:

Step1. Determine the  $n \times (m + s)$  matrix  $\begin{bmatrix} X & | & Y \end{bmatrix}$  referred to as input-output matrix.

Step2. To cluster all DMUs, apply hierarchical agglomerative clustering for input-output matrix.

Step3. Evaluate each DMU related to cluster that it belongs to. In this steps, we obtain the score efficiency and reference set corresponding each cluster.

Now, we apply the above procedure considering a simple example by which the obtained results can be readily seen in two-dimension space. We have 10 DMUs which utilize one input, whose value is normalized to 1, in order to produce two outputs (Y<sub>1</sub>, Y<sub>2</sub>) whose data as given in Table 1. The respective PPS in output space (following Farrell (1957)) is portrayed in Fig 1 (a). As depicted in Fig 1 (b), after applying hierarchical agglomerative clustering, all DMUs will be classified into the 19 clusters. Clusters 1-10 are singleton clusters each of which contains DMU1-DMU10 respectively. The number of each cluster has specified on it. Resultant clusters 11-19 are listed in Table2. The new PPS corresponding to these nine clusters (11-19) are shown with ellipsis in Fig 1 (a). The corresponding hierarchical clustering dendrogram is shown in fig 2 that illustrates the arrangement of the clusters.

Table1 Input-output data

DMU1	DMU2	DMU3	DMU4	DMU5	DMU6	DMU7	DMU8	DMU9	DMU10
1	1	1	1	1	1	1	1	1	1
4	3	5	4.5	2	6	6	5	1	2.5
0.5	6	5	2	3	3	1	4	2.5	3.5

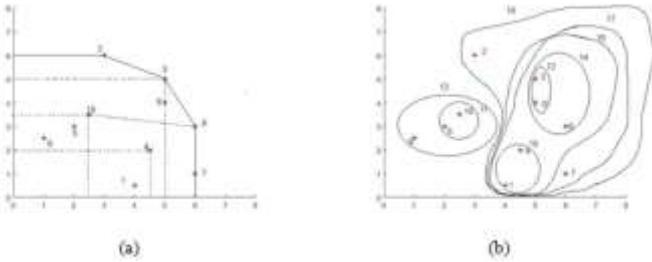


Fig. 1. (a) PPS in output space constructed by given DMUs (b) Resultant merged clusters derived from hierarchical agglomerative clustering

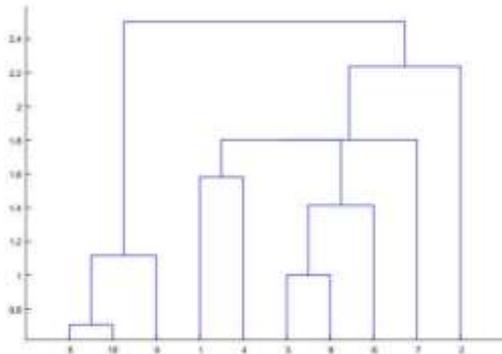


Fig 2. Hierarchical clustering dendrogram of 10 DMUs

Table 2 Resultant clusters along with their members for 10 DMUs

Cluster	Cluster members (DMU)
11	5,10
12	3,8
13	9,5,10
14	6,3,8
15	1,4
16	6,3,8,1,4
17	7, 6,3,8,1,4
18	2, 7, 6,3,8,1,4
19	9,5,10, 2,7, 6,3,8,1,4

Following third step of the suggested approach, we solve model (1) and evaluate each DMU comparing with members of the clusters that it belongs to. Each rows of Table 3 give the clusters consist of each DMU along with corresponding score efficiency and reference sets, respectively.

For example, DMU1 is classified into clusters 15,16,17,18 and 19 and its score efficiency relative to five corresponding PPS is 0.39, 0.18, 0.18, 0.17 and 0.17 respectively. The produced reference sets are given in the third column of Table 3. This results show that although this unit is inefficient in all levels, it has better performance at lower levels. As the management strategy, DMU1 should improve its performance hierarchically by considering reference sets {DMU4}, {DMU3}, {DMU3}, {DMU2,DMU3}

and {DMU2,DMU3} corresponding the lower to higher cluster respectively.

Table 3 The obtained score efficiency and reference set corresponding each cluster for 10 DMUs

DMU	clusters	Efficiency score	Reference set
D1	15,16,17,18,19	0.39,0.18,0.18,0.17,0.17	{D4}, {D3}, {D3}, {D2,D3}, {D2,D3}
D2	18,19	1,1	{D2}
D3	12,14,16,17,18,19	1,1,1,1,1,1	{D3}
D4	15,16,17,18,19	1,0.55,0.55,0.55,0.55	{D4},D3,D3,{D2,D3},{D2,D3}
D5	11,13,19	0.83,0.83,0.48	{D10}, {D10}, {D3}
D6	14,17,18,19	1,1,1,1	{D6}
D7	17,18,19	0.5,0.5,0.5	{D6},..., {D6}
D8	12,14,16,17,18,19	0.89,0.89,0.89,0.89,0.89,0.89	{D3},..., {D3}
D9	13,19	0.51,0.28	{D10}, {D6}
D10	11,13,19	1,1,0.58	{D10}, {D10}, {D3}

**4. Application to Iranian commercial banks**

In this section, we consider a case study presented by Bagherzadeh Valami (2009) consisting of 24 Iranian commercial banks with three inputs (Payable interest, Personnel, Non-performing loans) and five outputs (The total sum of the four main deposits, Other deposits, Loans granted, Received interest, Fee). Table 4 contains input-output data.

Table 4 Input-output data for 24 DMUs

DMU	x1	x2	x3	y1	y2	y3	y4	y5
1	3105.05	35.47	121910	3109668	179168	1508247	112273.11	1087.79
2	1113.92	21.23	14310	145446	11175	312514	7584.57	839.46
3	1476.28	20	10735	335492	161235	322235	34195.15	167.82
4	1962.49	13.58	2109	265643	27191	251604	1218.22	15.95
5	4521.03	26.08	41890	802090	501363	1763008	7017.64	672.82
6	552.42	21.58	18935	1510181	254988	519720	10065.04	89
7	4155.88	14.46	6513	391820	4791	180814	2315.37	134.13
8	8727.27	18.37	48683	1172029	106362	640433	46433.47	1297.05
9	309.69	28.31	52819	270421	265635	389094	2563.16	465.45
10	858.96	18.43	264	296173	65334	1068621	28712.99	454.82
11	2497.29	27.03	9363	434360	651496	1119173	11520.54	2103.71
12	5714.57	12.49	991	594269	8740	2072894	57808.34	4.8
13	1016.43	22.81	280852	437615	425490	2127581	11081132	1948.88
14	10176.79	15.54	12558	803008	9355	2733094	36242.9	7.53
15	2671.41	19.18	982	674783	53412	2927788	3288.39	511.1
16	3528.29	21.11	52888	454623	33596	287705	4650.81	123.99
17	4833.09	21.81	20616	550557	28837	447227	2063	97.47
18	2546.77	21.32	18918	333411	28294	181693	1276.63	333.37
19	2134.71	36.97	57294	338997	235339	602678	6487.29	116.56
20	3706.76	28.42	29362	35875	100579	352019	1650.86	48.02
21	3078.37	25.76	41104	777689	483724	1749329	1162.84	925
22	492	13.15	19201	338822	3767	27233	159.35	97.12
23	3317.51	28.55	17575	634772	776270	699471	89577.41	1567.93
24	780.46	29.52	31492	138304	12248	297669	2825.85	79.09

Table 5 shows 47 resultant clusters from the implementation of hierarchical agglomerative clustering method along with their members for 24 DMUs.

Table 5 Resultant clusters along with their members for 24 DMUs

Cluster	Cluster members (DMU)
25	02,24
26	05,21
27	07,18
28	03,20
29	04,7,18
30	04,07,16,18
31	03,04,07,16,18,20
32	02,03,04,07,16,18,20,24
33	02,03,04,07,09,16,18,20,24
34	02,03,04,07,09,16,18,20,22,24
35	02,03,04,07,09,16,17,18,20,22,24
36	02,03,04,07,09,16,17,18,19,20,22,24
37	14,15
38	06,08
39	12,13
40	11,23
41	02,03,04,07,09,10,16,17,18,19,20,22,24
42	05,12,13,21
43	02,03,04,07,09,10,11,16,17,18,19,20,22,23,24
44	02,03,04,06,07,08,09,10,11,16,17,18,19,20,22,23,24
45	05,12,13,14,15,21
46	02,03,04,05,06,07,08,09,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24
47	All

After evaluating each DMU comparing with members of the clusters that it belongs to, obtained different score efficiencies and reference sets are listed in Table 6. As can be seen, DMUs {1, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 21, 23} have the best performance in all levels. While DMU20 is inefficient in all levels and has the worst performance among all units. Such unit should adopt a management policy, by considering the associated introduced benchmarks in different levels, and improve its system hierarchically. For example, according to the fourth column, DMU3 is introduced as the only target point of this DMU at the first levels.

Table 6 obtained score efficiency and reference set corresponding each cluster for 24 DMUs

DMU	clusters	Efficiency score	Reference set (DMU)
01	47	1	{01}
02	25,32,33,34,35,36,41,43,44,46,47	1,1,1,1,1,1,0.14,0.12,0.12,0.12	{02}, {02}, {02}, {02}, {02}, {02}, {02}, {09,11,23}, {06,09,10,11}, {06,09,10,11}, {06,09,10,11}
03	28,31,32,33,34,35,36,41,43,44,46,47	1,1,1,1,1,1,1,0.42,0.42,0.41,0.41	{03}, {03}, {03}, {03}, {03}, {03}, {03}, {10,11,23}, {10,11,23}, {10,11,23}, {10,11,23}
04	29,30,31,32,33,34,35,36,41,43,44,46,47	1,1,1,1,1,1,1,0.15,0.14,0.09,0.08,0.08	{04}, {04}, {04}, {04}, {04}, {04}, {04}, {02,07}, {07,10,23}, {06,10,11}, {10,11,12}, {10,11,12}
05	26,46,47	1,1,1	{05}
06	38,44,46,47	1,1,1,1	{06}
07	27,29,30,31,32,33,34,35,36,41,43,44,46,47	1,1,1,1,1,1,1,1,1,1,1,1,0.06,0.05,0.05	{07}, {07}, {07}, {07}, {07}, {07}, {07}, {07}, {07}, {07}, {07}, {07}, {07}, {07}, {06,10,11}, {11,12,23}, {11,12,23}
08	38,44,46,47	1,1,1,1	{08}
09	33,34,35,36,41,43,44,46,47	1,1,1,1,1,1,1,1,1	{09}
10	41,43,44,46,47	1,1,1,1,1	{10}
11	43,44,46,47	1,1,1,1	{11}
12	39,42,45,46,47	1,1,1,1,1	{12}
13	45,46,47	1,1,1	{13}
14	37,45,46,47	1,1,1,1	{14}
15	37,45,46,47	1,1,1,1	{15}
16	30,31,32,33,34,35,36,41,43,44,46,47	1,1,1,1,0.46,0.46,0.46,0.43,0.11,0.11,0.11,0.11	{16}, {16}, {16}, {16}, {03,07,22}, {03,07,22}, {03,07,22}, {07,10}, {23}, {23}, {23}, {23}
17	35,36,41,43,44,46,47	1,1,1,1,0.08,0.07,0.07,0.07	{17}, {17}, {17}, {17}, {08,23}, {12,13,23}, {12,13,23}
18	27,29,30,31,32,33,34,35,36,41,43,44,46,47	1,1,1,1,0.21,0.21,0.21,0.21,0.21,0.13,0.06,0.06,0.06,0.06	{18}, {18}, {18}, {18}, {02,03,07}, {02,03,07}, {02,03,07}, {02,03,07}, {02,03,07}, {02,03,07}, {03,07,10}, {23}, {23}, {12,13,23}, {12,13,23}
19	36,41,43,44,46,47	1,0.27,0.19,0.19,0.16,0.16	{19}, {03,09,10}, {10,23}, {10,23}, {10,13,23}, {10,13,23}
20	28,31,32,33,34,35,36,41,43,44,46,47	0.13,0.13,0.13,0.13,0.13,0.13,0.13,0.13,0.1,0.05,0.05,0.05,0.05	{03}, {03}, {03}, {03}, {03}, {03}, {03}, {03}, {10}, {23}, {23}, {12,13,23}, {12,13,23}
21	26,42,45,46,47	1,1,1,1,1	{21}
22	34,35,36,41,43,44,46,47	1,1,1,1,1,0.07,0.07,0.07	{22}, {22}, {22}, {22}, {22}, {06,23}, {06,09,11,13}, {06,09,11,13}
23	40,43,44,46,47	1,1,1,1,1	{23}
24	25,32,33,34,35,36,41,43,44,46,47	1,1,0.15,0.15,0.15,0.15,0.14,0.1,2,0.1,0.1,0.1	{24}, {24}, {03,09}, {03,09}, {03,09}, {03,09}, {03,09}, {09,10}, {09,11,23}, {06,23}, {06,13,23}, {06,13,23}

5. Conclusion and future research directions

The classic Data Envelopment Analysis (DEA) models developed with the assumption that we deal with a set of peer Decision Making Units (DMUs), whereas, we may face heterogeneity among DMUs in the presence of high-level in the actual world. So, in this contribution, we implemented the hierarchical agglomerative clustering and examined the efficiency of all units in different cluster to have a fair evaluation and reduce the effect of heterogeneity on the score efficiency.

Our research can be extended in several ways. First, the analysis in this paper is based on technical efficiency score of DMUs related to different levels. While in situations where prices are known at each DMU, another kind of relative efficiency in DEA literature such as the cost, revenue and profit efficiency can be taken in to account for more curate analysis. Second, there are many families of data clustering algorithm, each of which classifies data based on a specific criteria, e.g. partitioning techniques of clustering like k-means and k-medoids and density-based clustering algorithms like DBSCAN and SNN. However, the analysis of obtained results through

implementing different method to choose the best cluster may be an interesting future research undertaking. Third, we believe that future researches are necessary to develop an appropriate clustering and evaluation method in the presence of imprecise or negative data. Also, in case of classification problems with DMUs whose activity continues across a multiple periods, the manager will deal with multi-period production system and it is required to investigate the effective factors on efficiencies and technology changes.

### **References**

- Alirezaee, M.R. and Mir-Hassani, S.A.,2006. A Layer DEA Model for Measuring and Improving the Efficiency in the Presence of Special Decision Making Units, *IUST International Journal of Engineering Science*, 17, 3-4, 75-79.
- Bagherzadeh Valami, H., 2009, Group performance evaluation, an application of data envelopment analysis, *Journal of Computational and Applied Mathematics* 230 ,485\_490.
- Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale efficiencies in data envelopment analysis. *Management Science*, 30, 1078–1092.
- Charnes, A., Cooper, W. W., Rhodes, E., 1978. Measuring the efficiency of decision making units. *Eur. J Oper Res.* 2 , 429-444.
- Cooper, W.W., Seiford, L.M., Tone, K., 2002. *Data Envelopment Analysis. A Comprehensive Text with Models, Applications, References and DEA-Solver Software.* Kluwer Academic Publishers: Boston, MA.
- Farrell, M.J. The measurement of productive efficiency,1957. *Journal of the Royal Statistical Society Series A, General* 120, 253\_281.
- Tone K. ,2001, A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research*, 130, 498–509.