Applying Data Envelopment Analysis and Discriminant Analysis to Determining the Most Efficient Decision-Making Unit

Kim F. Lam
Department of Management Sciences, City University of Hong Kong, Kowloon, Hong Kong
Email: msblam@cityu.edu.hk

***Abstract*— The use of data-envelopment analysis (DEA) to determine the most efficient decision making unit (DMU) has recently drawn attention in the literature. For some applications of DEA, decision-makers may only want to identify the most efficient DMU rather than determining the efficiencies of all possible DMUs. Some recent approaches combine the use of DEA and discriminant analysis (DA) to rank DMUs and identify the most efficient DMU. However, some of these approaches have drawbacks. This paper addresses those drawbacks and offers suggestions for improvements. This paper introduces a modified model to solve the problem of multiple solutions in DEA in determining the most efficient DMU. The modified model has a new goal, based on the assumption of cluster analysis that objects belonging to the same group should be more similar to each other than to objects from other groups. With this new goal, the modified model selects the solution in which members of the same group, whether efficient or inefficient, are most tightly clustered.**

***Index Terms*— data envelopment analysis, discriminant analysis, goal programming, mixed integer linear programming**

I. Introduction

Data envelopment analysis (DEA) is a method used to measure the relative performance of decision making units (DMUs) within a group. Suppose that there are n DMUs, and that each DMU has m inputs and produces s outputs. Let x_{ij} and y_{rj} represent the ith input and the rth output of DMU_j, respectively, for $i = 1, \ldots, m$, $r = 1, \ldots, s$, and $j = 1, \ldots, n$. The Charnes, Cooper, and Rhodes (CCR) model [1] is as follows:

$$Max \sum_{r=1}^{s} u_r y_{ro} \tag{1}$$

$$\text{s.t.} \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0, \quad j = 1, \ldots, n, \tag{2}$$

$$\sum_{i=1}^{m} v_i x_{io} = 1, \tag{3}$$

where u_r, $v_i \ge 0$, $r = 1, \ldots, s$, and $i = 1, \ldots, m$. The CCR model is solved for each DMU. A DMU is efficient only if its optimal objective value is capable of reaching 1 in the CCR model; otherwise, it is inefficient. Lately researchers [2-6] applied DEA in different applications. Recently, instead of calculating the efficiencies of all possible DMUs, some researchers [7-18] have sought to identify the most efficient DMU. In [8] and [10], two combinations of DEA and DA were used to rank DMUs and determine the most efficient DMU. Following [10], we call these approaches DEA-DA models. One of the primary advantages of applying DEA-DA models is that their solutions retain the fundamental result of DEA; that is, the classification of DMUs as efficient (E) or inefficient (IE). Any deviations from the original classifications imply certain violations of the original DEA model. The first step in solving a DEA-DA model is to classify each DMU as efficient or inefficient, using a classical DEA model. For instance, [8] used the CCR model [1], and [10] used a DEA model with a strong complementary slackness condition [19-21]. The second step is to develop a supporting hyperplane that separates the efficient DMUs from the inefficient DMUs, using a DA model. The supporting hyperplane is developed from the solutions generated in the first step.

The remainder of this paper is organized as follows. The two existing DEA-DA models [8, 10] used to rank DMUs and determine the most efficient DMU are discussed in Section II. In Section III, the modified model is introduced and discussed. Section IV concludes the paper.

II. The Two Existing DEA-DA Models

In a DEA-DA model, each DMU is classified as efficient or inefficient, using a classical DEA model. Then, efficient and inefficient DMUs are separated by a supporting hyperplane developed by a DA model. Two DEA-DA models are discussed below.

In [8], the CCR model [1] is used to classify all decision making units into “efficient” and “inefficient” groups. Next, the following mixed integer linear

Manuscript received November 1, 2019; revised February 26, 2020.

doi: 10.18178/joams.8.2.55-59

programming (MILP1) model is applied to the two groups.

$$\text{Min } \sum_{j=1}^{n} z_j \quad (4)$$

$$\text{s.t. } -\sum_{i=1}^{m} v_i x_{ij} + \sum_{r=1}^{s} u_r y_{rj} + M z_j \geq 0, \quad j \in E, \quad (5)$$

$$-\sum_{i=1}^{m} v_i x_{ij} + \sum_{r=1}^{s} u_r y_{rj} - M z_j \leq -\varepsilon, \quad j \in IE, (6)$$

$$\sum_{i=1}^{m} v_i x_{ie} = 1, \quad (7)$$

$$\sum_{r=1}^{s} u_r y_{re} \geq h, \quad (8)$$

$$\sum_{r=1}^{s} u_r y_{rj} - h\left(\sum_{i=1}^{m} v_i x_{ij}\right) \leq 0, \; j = 1,\ldots,n, j \neq e, (9)$$

$z_j \in \{0,1\}, j = 1,\ldots,n;\; u_r, v_i \geq 0, r = 1,\ldots,s, i = 1,\ldots,m;$ ε is a small positive number, M is a large positive number, and h is a number predetermined by the decision-maker.

The input and output weights obtained from MILP1 for each efficient DMU are used to compute the cross-efficiency scores of the DMUs. The objective of MILP1 is to minimize the number of misclassifications in both the “efficient” group and the “inefficient” group. This objective is consistent with the original classificatory results of the DEA. Furthermore, similar to the super-efficiency model [14] the efficient DMU under evaluation has the largest efficiency in MILP1. However, the value of h must be prescribed before MILP1 can be solved. Although [8] suggested using the super-efficiency model [22] to determine the value of h, trials with different values of h may still be required in some cases. Moreover, MILP1 requires one MILP problem to be solved for each efficient DMU, whereas most other models [9, 11-12] require only one MILP problem to be solved to determine the most efficient DMU.

In [10], DEA with a strong complementary slackness condition [19-20] is first used to classify all decision making units as efficient or inefficient. Then, the following model (MILP2) is applied to the two groups.

$$\text{Min } \varphi = M \sum_{j \in E} z_j + \sum_{j \in IE} z_j \quad (10)$$

$$\text{s.t. } -\sum_{i=1}^{m} v_i x_{ij} + \sum_{r=1}^{s} u_r y_{rj} + \sigma + M z_j \geq 0, j \in E, (11)$$

$$-\sum_{i=1}^{m} v_i x_{ij} + \sum_{r=1}^{s} u_r y_{rj} + \sigma - M z_j \leq -\varepsilon, j \in IE, \quad (12)$$

$$\sum_{i=1}^{m} v_i + \sum_{r=1}^{s} u_r = 1, \quad (13)$$

$$v_i \geq \varepsilon \zeta_i, i = 1,\ldots,m, \quad (14)$$

$$u_r \geq \varepsilon \zeta_i, r = 1,\ldots,s, \quad (15)$$

$$\sum_{i=1}^{m} \zeta_i = m, \quad (16)$$

$$\sum_{i=1}^{m} \zeta_i = s, \quad (17)$$

$$\sigma : URS, v_i \geq 0, \forall i, w_r \geq 0, \forall r,$$

$$z_j\text{: binary } \forall j;\; \zeta_i\text{: binary } \forall i;\; \zeta_r\text{: binary } \forall r.$$

In MILP2, M is a prescribed large number and ε is a prescribed small number. Incorporating the solutions of MILP2, a classification score for each DMU can be computed using the following equation [10]:

$$\rho_j = -\sum_{i=1}^{m} v_i^* x_{ij} + \sum_{r=1}^{s} u_r^* y_{rj} + \sigma^*, j = 1,\ldots,n. \quad (18)$$

The adjusted efficiency scores of all the DMUs can be computed using the classification scores obtained from (18). The DMU with an adjusted efficiency score of 1 (full efficiency) is regarded as the single efficient DMU [10]. All other DMUs have some level of inefficiency, and the DMU with an adjusted efficiency score of 0 is fully inefficient.

According to [10], MILP2 can be used to produce a single efficient DMU and to rank all of the DMUs. However, as MILP2 may have multiple optimal solutions, the single efficient DMU and the ranking of the DMUs may vary depending on which optimal solution is used to compute the classification scores. It is also well known that MILP DA models commonly have multiple optimal solutions. The following numerical example demonstrates that multiple optimal solutions exist in MILP2.

The data set listed in Table I below comprises five DMUs, each with two inputs and one output. Only DMU_1 and DMU_2 are inefficient; the rest are efficient.

With M = 1000 and ε = 0.001, the data in Table I yields three optimal solutions for MILP2. All of the solutions have the same objective value, and their input and output weights are listed in Table II.

The adjusted efficiency scores of the five DMUs for each of the three optimal solutions are computed and listed in Table III. As shown in Table III, each of the three optimal solutions identifies a different DMU as the single efficient DMU, and generates different ranks for the DMUs.

TABLE I. DATA OF THE NUMERICAL EXAMPLE

	DMU_1	DMU_2	DMU_3	DMU_4	DMU_5
Input 1	6	7	9	5	2
Input 2	9	7	1	3	9
Output	1	1	1	1	1
	IE	IE	E	E	E

TABLE II. WEIGHTS AND OPTIMAL OBJECTIVE VALUES

Optimal Solution	Input 1 (v_1)	Input 2 (v_2)	Output (u_1)	σ	φ[a]
1	0.03589	0.08923	0.87488	0	0
2	0.001	0.001	0.998	-0.987	0
3	0.00203	0.001	0.99697	-0.97677	0

[a] Optimal objective value

TABLE III. ADJUSTED EFFICIENCY SCORES FOR THE FIVE DMUS

Optimal Solution	DMU_1	DMU_2	DMU_3	DMU_4	DMU_5
1	0	0.23520	1	0.94244	0.23689
2	0	0.14286	0.71429	1	0.57143
3	0.00404	0	0.23687	0.98788	1

The adjusted efficiency scores of the five DMUs for each of the three optimal solutions are computed and listed in Table III. As shown in Table III, each of the three optimal solutions identifies a different DMU as the single efficient DMU, and generates different ranks for the DMUs.

The objective function of both MILP1 and MILP2 is to minimize the number of DMUs misclassified as either efficient or inefficient. MILP2 differs slightly from MILP1 in placing a larger penalty on misclassified DMUs in the "efficient" group than in the "inefficient" group. However, the most significant difference between MILP1 and MILP2 is that MILP2 uses a common set of weights to generate a supporting hyperplane for all of the DMUs, whereas MILP1 develops a supporting hyperplane for each efficient DMU. MILP1 is similar to the super-efficiency model in that the DMU under evaluation is always the most efficient, whereas this condition is not required by MILP2. However, as mentioned previously, MILP2 may have multiple optimal solutions. In such a case, before MILP2 can be used to identify the best DMU or to rank the DMUs, one optimal solution must be chosen to compute the classification scores for the DMUs. In the next section, an approach to tackling this problem is outlined.

III. A MODIFIED FORMULATION

When multiple optimal solutions exist for MILP2, researchers need to decide which optimal solution should be used to compute the classification scores for the DMUs. To identify the most suitable solution to compute the classification scores of the DMUs, the following modified formulation (MILP3) is proposed:

$$\text{Min} \quad M\sum_{j\in E} z_j + \sum_{j\in IE} z_j + \varepsilon\sum_{j=1}^{n}\left(d_j^+ + d_j^-\right) \tag{19}$$

$$\text{s.t.} \quad -\sum_{i=1}^{m} v_i x_{ij} + \sum_{r=1}^{s} u_r y_{rj} + \sigma + M z_j \geq 0, j \in E, \tag{20}$$

$$-\sum_{i=1}^{m} v_i x_{ij} + \sum_{r=1}^{s} u_r y_{rj} + \sigma - M z_j \leq -\varepsilon, j \in IE, \tag{21}$$

$$\sum_{i=1}^{m} v_i + \sum_{r=1}^{s} u_r = 1, \tag{22}$$

$$\sum_{i=1}^{m} v_i\left(x_{ij} - \overline{x_i^E}\right) + \sum_{r=1}^{s} u_r\left(y_{rj} - \overline{y_r^E}\right) - d_j^+ + d_j^- = 0, \quad j \in E, \tag{23}$$

$$\sum_{i=1}^{m} v_i\left(x_{ij} - \overline{x_i^{IE}}\right) + \sum_{r=1}^{s} u_r\left(y_{rj} - \overline{y_r^{IE}}\right) - d_j^+ + d_j^- = 0, \quad j \in IE, \tag{24}$$

$$v_i \geq \varepsilon\zeta_i, i = 1, \ldots, m, \tag{25}$$

$$u_r \geq \varepsilon\zeta_i, r = 1, \ldots, s, \tag{26}$$

$$\sum_{i=1}^{m} \zeta_i = m, \tag{27}$$

$$\sum_{i=1}^{m} \zeta_i = s, \tag{28}$$

$$\sigma : URS, v_i \geq 0, \forall i, w_r \geq 0, \forall r, d_j^+, d_j^- \geq 0, \forall j,$$

$$z_j\text{: binary } \forall j;\ \zeta_i\text{: binary } \forall i;\ \zeta_r\text{: binary } \forall r.$$

In MILP3, $\overline{x_i^g}$ and $\overline{y_r^g}$ are the average scores of the *i*th input and *r*th output of group g, respectively, where $g \in \{E, IE\}$. As the average classification score of group g is $\sum_{i=1}^{m} v_i \overline{x_i^g} + \sum_{r=1}^{s} u_r \overline{y_r^g} + \sigma$, the variables d_j^+ and d_j^- measure the deviation of the classification score of the *j*th DMU from its group's mean classification score. The third goal of MILP3 is to minimize the sum of the resulting deviations. The justification for minimizing

d_j^+ and d_j^- is that the classification scores of members of the same group should be more similar to each other than to the scores of members of other groups. The objective of minimizing the sum of the deviations of individual classification scores from group mean classification scores was first introduced by [23] in DA.

In practice, MILP3 can also be regarded as a preemptive goal programming model wherein goal 1 is to minimize $\sum_{j\in E} z_j$, goal 2 is to minimize $\sum_{j\in IE} z_j$, and goal 3 is to minimize $\sum_{j=1}^{n}(d_j^+ + d_j^-)$. Applying MILP3 to the data in Table I with M = 1000 and ε = 0.001 gives the second optimal solution listed in Table II.

An advantage of MILP3 over MILP2 is that compared with the multiple solutions of MILP2, the single solution of MILP3 better separates the classification scores of the efficient DMUs from those of the inefficient DMUs. In addition, no prior information, such as past knowledge or managerial restrictions, is required to introduce additional constraints and goals to MILP3. Furthermore, unlike MILP1, which solves one MILP problem for each efficient DMU, MILP3 requires only one MILP problem to be solved.

IV. Conclusion

This paper discusses some of the drawbacks of applying two existing DEA-DA models to rank DMUs and determine the most efficient DMU. Among the main disadvantages of using one of these DEA-DA models to determine the most efficient DMU is that multiple solutions for the model exist. Consequently, the most efficient DMU identified will depend on which optimal solution is used. This paper introduces a modified model to solve the problem of multiple solutions. The modified model has a new goal, based on the assumption of cluster analysis that objects belonging to the same group should be more similar to each other than to objects from other groups. Accordingly, individual classification scores must be closer to their group's mean classification score than to the mean classification score of another group. With this new goal, the modified model selects the solution in which members of the same group, whether efficient or inefficient, are most tightly clustered. This has a tendency to increase the discriminant power to separate the members of the two groups.

Conflict of Interest

The author declares no conflict of interest.

References

[1] A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of the decision making units," *European Journal of Operational Research*, vol. 2, pp. 429-444, 1978.

[2] M. Carrillo and J. M. Jorge, "An alternative neutral approach for cross-efficiency evaluation," *Computers & Industrial Engineering*, vol. 120, pp. 137-145, 2018.

[3] K. F. Lam, "Choosing suitable weights sets for cross evaluations in DEA," *Journal of Advanced Management Science*, vol. 7, pp. 19-24, 2019.

[4] D. Khezrimotlagh, J. Zhu, W. D. Cook, and M. Toloo, "Data envelopment analysis and big data," *European Journal of Operational Research*, vol. 274, no. 3, pp.1047-1054, 2019.

[5] M. Toloo and M. Mirbolouki, "A new project selection method using data envelopment analysis," *Computers & Industrial Engineering*, vol. 138, pp. 106-119, 2019.

[6] G. R. Amin, S. Al-Muharrami, M. Toloo, "A combined goal programming and inverse DEA method for target setting in mergers," *Expert Systems with Applications*, vol. 115, pp. 412-417, 2019.

[7] G. R. Amin and M. Toloo, "Finding the most efficient DMUs in DEA: An improved integrated model," *Computers & Industrial Engineering*, vol. 52, pp. 71-77, 2007.

[8] K. F. Lam, "In the determination of weight sets to compute cross-efficiency ratios in DEA," *Journal of Operational Research Society*, vol. 61, pp. 134-143, 2010.

[9] A. A. Foroughi, "A new mixed integer linear model for selecting the best decision making units in data envelopment analysis," *Computers & Industrial Engineering*, vol. 60, pp. 550-554, 2011.

[10] T. Sueyoshi and M. Goto, "A combined use of DEA (DATA Envelopment Analysis) with strong complementary slackness condition and DEA-DA (Discriminant Analysis)," *Applied Mathematics Letters*, vol. 24, pp. 1051-1056, 2011.

[11] Y. M. Wang and P. Jiang, "Alternative mixed integer linear programming models for identifying the most efficient decision making units in data envelopment analysis." *Computers & Industrial Engineering*, vol. 62, pp. 546-553, 2012.

[12] K. F. Lam, "In the determination of the most efficient decision making unit in data envelopment analysis," *Computers & Industrial Engineering*, vol. 79, pp. 76-84, 2015.

[13] M. Toloo, "Alternative minimax model for finding the most efficient unit in data envelopment analysis," *Computers & Industrial Engineering*, vol. 81, pp. 186-194, 2015.

[14] M. Salahi and M. Toloo, "In the determination of the most efficient decision making unit in data envelopment analysis: A comment," *Computers & Industrial Engineering*, vol. 104, pp. 216-218, 2017.

[15] M. Toloo and M. Salahi, "A powerful discriminative approach for selecting the most efficient unit in DEA," *Computers & Industrial Engineering*, vol. 115, pp. 269-277, 2018.

[16] B. Ebrahimi and M. Khalili, "A new integrated AR-IDEA model to find the best DMU in the presence of both weight restrictions and imprecise data," *Computers & Industrial Engineering*, vol. 125, pp. 357-363, 2018.

[17] M. Toloo and M. Salahi, "A powerful discriminative approach for selecting the most efficient unit in DEA," *Computers & Industrial Engineering*, vol. 115, pp. 269-277, 2018.

[18] M. Toloo, S. Nalchigar, and B. Sohrabi, "Selecting most efficient information system projects in presence of user subjective opinions: a DEA approach," *Central European Journal of Operations Research*, vol. 26 (4), pp. 1027-1051, 2018.

[19] T. Sueyoshi and K. Sekitani, "The measurement of returns to scale under a simultaneous occurrence of multiple solutions in a reference set and a supporting hyperplane," *European Journal of Operational Research*, vol. 181, pp. 549-570, 2007.

[20] T. Sueyoshi and K. Sekitani, "An occurrence of multiple projections in DEA-based measurement of technical efficiency: theoretical comparison among DEA models from desirable properties," *European Journal of Operational Research*, vol. 196, pp.764-794, 2009.

[21] K. F. Lam, "A discriminant analysis and goal programming approach to solve the multiple criteria data envelopment analysis model", *Journal of Advanced Management Science*, vol. 6, pp. 222-226, 2018.

[22] P. Andersen and N. Petersen, "A procedure for ranking efficient units in data envelopment analysis," *Management Science*, vol. 39, pp. 1261-1264, 1993.

[23] K. F. Lam, E. U. Choo, and J. W. Moy, "Minimizing deviations from the group mean: A new linear programming approach for the two-group classification problem," *European Journal of Operational Research*, vol. 88, pp. 358-367, 1996.

Kim Fung Lam received a BBA (1984), an MBA (1986) and a PhD (1992) in management science at Simon Fraser University, Vancouver, B.C., Canada.

He is an Associate Professor of Management Science at City University of Hong Kong, Hong Kong. He has published in journals such as *Annals of Operations Research*, *Computers and Industrial Engineering*, *Computers and Operations Research, European Journal of Operational Research, Journal of the Operational Research Society, Journal of Advanced Management Science* and *Personnel Review*, etc. His research interests include multi-criteria decision making, discriminant analysis, and data envelopment analysis.

Dr. Lam is a member of the International Society on Multiple Criteria Decision Making.