Optimality

www.opt.reapress.com

Opt. Vol. 1, No. 2 (2024) 224–231.

Paper Type: Original Article

An Improved Mixed-Integer DEA Approach to Determine the Most Efficient Unit

Bohlool Ebrahimi[1,*], Szabolcs Fischer[2], Milos Milovancevic[3]

[1] Department of Industrial Engineering, ACECR, Sharif branch, Tehran, Iran; b.ebrahimi@aut.ac.ir.
[2] Department of Transport Infrastructure and Water Resources Engineering, Széchenyi István University, Győr, Hungary; fischersz@sze.hu.
[3] Department of Mechanical Design, Development and Engineering, University of Nis, Nis, Serbia, milovancevic@masfak.ni.ac.rs

Citation:

Received: 04 April 2024 *Revised: 16 July 2024* *Accepted: 16 September 2024*	Ebrahimi, B. (2024). An improved mixed-integer DEA approach to determine the most efficient uni. *Optimality, 1* (2), 224-231.

Abstract

Recently, Lam [1] developed a two-step method to find the most efficient (the best) unit in Data Envelopment Analysis (DEA). The first step finds an appropriate value for the epsilon to use in the second step to find the best unit. Salahi and Toloo [2] showed that the approach of Lam [1] may fail to determine the best unit. To fill this gap, they proposed a new model to find a suitable value for the epsilon in the first step. The current paper shows that in some cases, we may have several most efficient DMUs such that we could not easily discriminate among them to determine one of them as the best DMU. Also, we show that the second step in the proposed approach by Lam and Salahi & Toloo, is redundant. We propose an improved approach that can find the best DMU by solving only one model. As a result, the calculation burden of the new approach is significantly less than the two mentioned approaches. A real numerical example is used to compare the results and show the usefulness of the new approach.

Keywords: Data envelopment analysis, Best DMU, Mixed integer programming, Non-archimedean epsilon.

1 | Introduction

Lam [1] developed the following mixed integer linear Data Envelopment Analysis (DEA) model to find the most efficient Decision-Making Units (DMU).

Corresponding Author: b.ebrahimi@aut.ac.ir

$$
\begin{aligned}
&\max\ h \\
&\text{s.t.} \\
&\quad \sum_{i=1}^{m} v_i x_{ij} + MI_j \le 1 + M\ ,\ j=1,\dots,n \\
&\quad \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} - MI_j \le 0\ ,\ j=1,\dots,n \\
&\quad -\sum_{r=1}^{s} u_r y_{rj} + \sum_{i=1}^{m} v_i x_{ij} + MI_j + h \le M,\ j=1,\dots,n \\
&\quad \sum_{j=1}^{n} I_j = 1 \\
&\quad I_j \in \{0,1\} \\
&\quad u_r, v_i \ge \varepsilon^*,\ \text{for all } i,r.
\end{aligned}
\tag{1}
$$

Where M is a positive large number. It is easy to see that in the optimal solution for only one $k \in \{1,2,\dots,n\}$ we have $I_k^* = 1 \& I_j^* = 0, \text{for all } j \neq k$, which implies the efficiency score of DMU_k is more than or equal to one and the rest less than or equal to one. It should be here noted that a positive h^* implies that the efficiency score of DMU_k is strictly more than one, which shows this DMU is the most efficient DMU. However, we will show that this DMU may not be the only most efficient DMU and we may have several most efficient DMUs. Lam [1] proposed the following model of Amin and Toloo [3] to find a proper value for the epsilon to use in the *Model (1)*.

$$
\begin{aligned}
&\varepsilon^* = \max\ \varepsilon \\
&\text{s.t.} \\
&\quad \sum_{i=1}^{m} v_i x_{ij} \le 1,\ j=1,\dots,n \\
&\quad \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0\ ,\ j=1,\dots,n \\
&\quad u_r, v_i \ge \varepsilon,\ \text{for all } i,r.
\end{aligned}
\tag{2}
$$

Salahi and Toloo [2] discussed that *Model (2)* may give an unsuitable value for the ε^*. In other words, *Model (1)* may be unable to find the most efficient DMU by utilizing the maximum value for the epsilon obtained by solving the *Model (2)*. Therefore, they proposed the following *Model (3)* to find a suitable value for the epsilon to use in *Model (1)*.

$$
\begin{aligned}
&\varepsilon^* = \max\ \varepsilon \\
&\text{s.t.} \\
&\quad \sum_{i=1}^{m} v_i x_{ij} + MI_j \le 1 + M\ ,\ j=1,\dots,n \\
&\quad \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} - MI_j \le 0\ ,\ j=1,\dots,n \\
&\quad -\sum_{r=1}^{s} u_r y_{rj} + \sum_{i=1}^{m} v_i x_{ij} + MI_j + h \le M,\ j=1,\dots,n \\
&\quad \sum_{j=1}^{n} I_j = 1 \\
&\quad I_j \in \{0,1\} \\
&\quad h, u_r, v_i \ge \varepsilon,\ \text{for all } i,r.
\end{aligned}
\tag{3}
$$

They claimed that by using the optimal objective value of *Model (3)*, *Model (1)* gives exactly one unit as the most efficient (best DMU). In the following example, we show that *Model (1)* may give different DMUs as the most efficient DMU.

Example 1. Consider the following three DMUs, each uses a single input to produce two outputs.

Table 1. The data of three DMUs.

DMUs	Input1	Output 1	Output 2
1	2	3	4
2	2	4	3
3	3	1	2

Solving the *Model (3)* with $M=100$ gives $\varepsilon^*=0.1$. Now, solving the *Model (1)* with $M=100$ and $\varepsilon^*=0.1$, gives the following optimal solution.

$$h^*=0.1,\ I_1^*=1, I_2^*=I_3^*=0,\ u_1^*=0.1, u_2^*=0.2, v_1^*=0.5.$$

This optimal solution implies that the efficiency score of the DMU_1 is strictly greater than one and the rest less than or equal to one. In other words, this DMU is the most efficient DMU. We add the constraint $I_1=0$ to the *Model (1)* and resolve it to give the following optimal solution.

$$h^*=0.1,\ I_2^*=1, I_1^*=I_3^*=0,\ u_1^*=0.2, u_2^*=0.1, v_1^*=0.5.$$

The above optimal solution implies that the *Model (1)* has multiple optimal solutions and so the DMU_2 is also the most efficient DMU. As a result, the proposed approach by Lam [1] and Salahi and Toloo [2] cannot discriminate between DMU_1 and DMU_2. Therefore, in some cases we may have several most efficient DMUs, and so the *Model (1)* randomly reports one of them as the best DMU based on the solver that is used to solve it. One method to discriminate between such best DMUs is using weight restrictions.

In the following, we show that the second step (*Model (1)*) is redundant. In other words, the most efficient DMU can be obtained just by solving the *Model (3)*. Indeed, we show that in *Model (3)*, $h^*>0$, which implies this model gives a DMU with an efficiency score of strictly greater than one and the rest less than or equal to one.

Theorem 1. In the *Model (3)*, $h^*>0$. In other words, solving the *Model (3)* gives the most efficient DMU.

Proof: according to Cooper, Seiford, & Tone [4], suppose DMU_k is an extremely efficient DMU and consider the following super-efficiency model of Andersen & Petersen [5] for this DMU.

$$\begin{aligned}
E_k = \max\ & \sum_{r=1}^{s} u_r y_{rk} \\
\text{s.t.}\ & \\
& \sum_{i=1}^{m} v_i x_{ik} = 1, \\
& \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0,\ j=1,\dots,n,\ j\neq k \\
& u_r, v_i \ge \varepsilon,\ \text{for all } i,r.
\end{aligned} \tag{4}$$

Since DMU_k is an extremely efficient DMU, the optimal solution of *Model (4)* implies $E_k^*>1$ with positive weights of $u_r^*, v_i^*>0$. Now, let $h=\varepsilon=\min\left\{E_k^*-1, u_r^*, v_i^*,\ \text{for all } i,r\right\}>0$ and suppose that M is a large enough positive number such that $M \ge \max\left\{E_k^*, h+\sum_{i=1}^{m} v_i^* x_{ij}, j=1,2,\dots,n\right\}$, in this case, we show that $h,\varepsilon,u_r^*,v_i^*>0$ is a feasible solution to *Model (3)* that completes the proof.

Let $I_k = 1, I_j = 0, \text{for all } j \neq k,$ in this case, for the first type of constraints of *Model (3)* we have:

$$\sum_{i=1}^{m} v_i^* x_{ij} + MI_j = \sum_{i=1}^{m} v_i^* x_{ij} \leq M \leq M + 1, \text{for all } j \neq k.$$

$$\sum_{i=1}^{m} v_i^* x_{ik} + MI_k = 1 + M \leq 1 + M.$$

For the second type of constraint s of *Model (3)*, we have:

$$\sum_{r=1}^{s} u_r^* y_{rj} - \sum_{i=1}^{m} v_i^* x_{ij} - MI_j = \sum_{r=1}^{s} u_r^* y_{rj} - \sum_{i=1}^{m} v_i^* x_{ij} \leq 0, \text{ for all } j \neq k.$$

$$\sum_{r=1}^{s} u_r^* y_{rk} - \sum_{i=1}^{m} v_i^* x_{ik} - MI_k = E_k^* - 1 - M \leq E_k^* - 1 - E_k^* = -1 \leq 0.$$

For the third type of constraint s of *Model (3)*, we have:

$$-\sum_{r=1}^{s} u_r^* y_{rj} + \sum_{i=1}^{m} v_i^* x_{ij} + MI_j + h = -\sum_{r=1}^{s} u_r^* y_{rj} + \sum_{i=1}^{m} v_i^* x_{ij} + h \leq \sum_{i=1}^{m} v_i^* x_{ij} + h \leq M.$$

$$-\sum_{r=1}^{s} u_r^* y_{rk} + \sum_{i=1}^{m} v_i^* x_{ik} + MI_k + h = -E_k^* + 1 + M + h \leq -h + M + h = M.$$

Therefore, the weights $u_r^*, v_i^* > 0$ is a feasible solution for *Model (3)* that implies $h^* \geq h > 0$.

Note 1: the optimal value of objective function of *Model (4)* may go to infinity. In such a situation, we could select feasible positive weights with a bonded $E_k > 1$ to prove the *Theorem 1*.

Based on *Theorem 1*, applying *Model (3)* to the data presented in *Table 1* gives the following optimal solution.

$$h^* = 0.1,\ I_1^* = 1, I_2^* = I_3^* = 0,\ u_1^* = 0.1, u_2^* = 0.2, v_1^* = 0.5.$$

This optimal solution implies that the DMU$_1$ is the most efficient DMU. Solving the *Model (3)* with the additional restriction of $d_1 = 0$ gives $h^* = 0.1,\ I_2^* = 1, I_1^* = I_3^* = 0$ which implies DMU$_2$ is another most efficient DMU. These results are consistence with the results of the models of Salahi and Toloo [2].

By considering the discussed issues, in the next section we propose an improved approach to find the best DMU by solving only one model.

2 | The Improved Approach

The main aim of *Model (1)* is to find a DMU with an efficiency score of greater than one and the rest less than or equal to one. For this purpose, we propose the following *Model (5)* to find the most efficient DMU.

$$\begin{aligned}
&\max\ \hat{\varepsilon} \\
&\text{s.t.} \\
&\quad \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} - MI_j \leq 0,\ j = 1,\ldots,n \\
&\quad -\sum_{r=1}^{s} u_r y_{rj} + \sum_{i=1}^{m} v_i x_{ij} + MI_j + \hat{h} \leq M,\ j = 1,\ldots,n \\
&\quad \sum_{j=1}^{n} I_j = 1 \\
&\quad I_j \in \{0,1\} \\
&\quad \hat{h}, u_r, v_i \geq \hat{\varepsilon},\ \text{for all } i, r.
\end{aligned} \tag{5}$$

It is obvious that the *Model (5)* is always feasible. In the following we validate it by some theorem.

Theorem 2. In the *Model (5)*, $0 < \hat{h}^* \leq M$, which implies $\hat{\varepsilon}^* \leq M$.

Proof: since *Model (5)* has fewer constraints compared to *Model (3)*, its optimal solution must be greater than or equal to the optimal objective function of the *Model (3)*. Therefore, $\hat{\varepsilon}^* \geq \varepsilon^*$, and hence $\hat{h}^* \geq h^* > 0$. Now, considering the first and second type constraints of the *Model (5)* implies:

$$\hat{h} \leq \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + M - MI_j \leq MI_j + M - MI_j = M.$$

That completes the proof.

Theorem 3. The optimal objective function of *Model (5)* is strictly positive.

Proof: it is obvious by considering the proof of the *Theorem 1*.

Theorem 4. Solving the *Model (5)* gives a DMU with an efficiency score of strictly greater than one and the rest less than or equal to one.

Proof: let in the optimal solution we have $I_k = 1, I_j = 0, \text{for all } j \neq k,$ so we conclude:

$$\sum_{r=1}^{s} u_r^* y_{rj} - \sum_{i=1}^{m} v_i^* x_{ij} - MI_j = \sum_{r=1}^{s} u_r^* y_{rj} - \sum_{i=1}^{m} v_i^* x_{ij} \leq 0 \Rightarrow \frac{\sum_{r=1}^{s} u_r^* y_{rj}}{\sum_{i=1}^{m} v_i^* x_{ij}} \leq 1, \text{ for all } j \neq k.$$

$$-\sum_{r=1}^{s} u_r^* y_{rk} + \sum_{i=1}^{m} v_i^* x_{ik} + MI_k + \hat{h} \leq M \Rightarrow \sum_{r=1}^{s} u_r^* y_{rk} - \sum_{i=1}^{m} v_i^* x_{ik} \geq \hat{h} > 0 \Rightarrow \frac{\sum_{r=1}^{s} u_r^* y_{rk}}{\sum_{i=1}^{m} v_i^* x_{ik}} > 1.$$

Note 2: our proposed approach solves only one model with fewer constraints to find the best DMU; however, the proposed approach by Lam [1] and Salahi & Toloo [2] solves two models with some more constraints to find the best DMUs. Therefore, our method is more efficient compared with the two mentioned approaches.

In the next section, we will verify *Model (5)* by a real numerical example.

3 | Numerical Example 2

In this real numerical example, we show how our approach can find the most efficient DMU just by solving *Model (5)*. *Table 2* depicts the data set of nineteen Facility Layout Designs (FLDs) taken from Ertay et al. [6].

Table 2. Data of inputs and outputs of 19 FLDs.

DMU No.	Inputs		Outputs			
	Cost ($)	Adjacency Score	Shape Rate	Flexibility	Quality	Hand-Carry Utility
1	20309.56	6405	0.4697	0.0113	0.041	30.89
2	20411.22	5393	0.438	0.0337	0.0484	31.34
3	20280.28	5294	0.4392	0.0308	0.0653	30.26
4	20053.20	4450	0.3776	0.0245	0.0638	28.03
5	19998.75	4370	0.3526	0.0856	0.0484	25.43
6	20193.68	4393	0.3674	0.0717	0.0361	29.11
7	19779.73	2862	0.2854	0.0245	0.0846	25.29
8	19831.00	5473	0.4398	0.0113	0.0125	24.80
9	19608.43	5161	0.2868	0.0674	0.0724	24.45
10	20038.10	6078	0.6624	0.0856	0.0653	26.45
11	20330.68	4516	0.3437	0.0856	0.0638	29.46
12	20155.09	3702	0.3526	0.0856	0.0846	28.07
13	19641.86	5726	0.269	0.0337	0.0361	24.58
14	20575.67	4639	0.3441	0.0856	0.0638	32.20

Table 2. Continued.

DMU No.	Inputs		Outputs			
	Cost ($)	Adjacency Score	Shape Rate	Flexibility	Quality	Hand-Carry Utility
15	20687.50	5646	0.4326	0.0337	0.0452	33.21
16	20779.75	5507	0.3312	0.0856	0.0653	33.60
17	19853.38	3912	0.2847	0.0245	0.0638	31.29
18	19853.38	5974	0.4398	0.0337	0.0179	25.12
19	20355.00	17402	0.4421	0.0856	0.0217	30.02

Solving the *Model (5)* with $M = 100$, gives the following optimal solution.

$\hat{\varepsilon}^* = \hat{h}^* = 0.011.$

$I_{10}^* = 1, I_j^* = 0, \text{for all } j \neq 10.$

$u_1^* = 422.843, u_2^* = 661.407, u_3^* = 0.011, u_4^* = 4.778.$

$v_1^* = 0.015, v_2^* = 0.011.$

This optimal solution implies that the FLD_{10} is the most efficient DMU. Solving the *Model (5)* with the additional constraint of $I_{10} = 0,$ gives $\hat{\varepsilon}^* = \hat{h}^* = 0.008.$ Reducing the optimal objective value shows that there is no alternative solution and hence the FLD_{10} is the best DMU.

Now, we apply the proposed models by Lam [1] and Salahi & Toloo [2] to find the best FLD. Solving the *Model (3)* gives $\varepsilon^* = h^* = 0.000044 \& I_7^* = 1.$ As we proved in *Theorem 1*, this optimal solution implies that the FLD_7 is the most efficient DMU. Anyway, solving *Model (1)* verifies *Theorem 1* and determines this FLD as the most efficient DMU. The optimal weights, efficiency scores and full ranking of FLDs are given in *Table 3*. It should be here noted that, as reported in Lam [1], using $\varepsilon^* = 0$ in the *Model (1)* gives the FLD_{10} as the best DMU.

Table 3. Efficiency scores and rank of 19 FLDs.

FLDs	Model (1) of Lam (2015)		Our Model (5)	
	Eff. Score	Rank	Eff. Score	Rank
1	0.433	14	0.954	10
2	0.530	12	0.988	4
3	0.720	7	0.978	7
4	0.735	5	0.895	13
5	0.560	11	0.950	11
6	0.415	15	0.983	6
7	1.054	1	0.792	18
8	0.140	19	0.882	14
9	0.824	3	0.815	16
10	0.705	9	1.275	1
11	0.724	6	0.976	8
12	1.000	2	1.000	2
13	0.402	16	0.717	19
14	0.714	8	0.999	3
15	0.485	13	0.987	5
16	0.701	10	0.970	9
17	0.757	4	0.848	15
18	0.196	17	0.914	12
19	0.163	18	0.795	17
Optimal weights				
v_1^*		0.000044	0.014928	
v_2^*		0.000044	0.010523	
u_1^*		0.000044	422.843300	
u_2^*		0.000044	661.406900	
u_3^*		12.439940	0.010523	
u_4^*		0.000044	4.777930	

The results in *Table 3* reveal the following points:

I. The *Model (5)* implies that the efficiency score of the FLD_{10} is 1.275, which shows the distance between the efficiency score of the first two top-ranking FLDs is equal to 0.275. This value for the proposed approach by Lam [1] is equal to $1.054-1=0.054$, which shows the discrimination power of the *Model (5)* is more than the *Model (1)*.

II. As reported by Toloo & Salahi [7], all of the proposed approaches by Andersen & Petersen [5], Foroughi [8], Wang and Jiang [9], Toloo [10] and Toloo and Salahi [7] imply that FLD_{10} is the best DMU that verify the result of the *Model (5)*.

As mentioned in *Note 2*, our approach solves only one model to find the best FLD; however, the proposed approach by Lam [1] solves two models with more constraints to obtain the best FLD.

4 | Conclusion

This paper studied the two-step method proposed by Lam [1] and Salahi & Toloo [2] to find the most efficient units. It is mathematically proved that the model of the first step in Salahi & Toloo [2] is enough to find the best DMU, and so the second step is redundant. We improved the model of the first step and, proposed a modified model and showed that it could find the best DMU with considerably less calculation volume. Also, the results of the real numerical example showed that the discrimination power of the modified model is higher than the models of Lam [1]. Also, we compared our result with five existing approaches in the literature that verified the *Model (5)*.

Author Contributions

Bohlool Ebrahimi was responsible for the conceptualization and design of the study, as well as drafting the manuscript. Szabolcs Fischer contributed to the development of the methodology and provided critical revisions. Milos Milovancevic performed the numerical analysis and assisted in the interpretation of results.

Funding

This research received no specific funding from any public, commercial, or not-for-profit organization.

Data Availability

The data and materials supporting this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Lam, K. F. (2015). In the determination of the most efficient decision making unit in data envelopment analysis. *Computers & industrial engineering*, *79*, 76–84. DOI: 10.1016/j.cie.2014.10.027

[2] Salahi, M., & Toloo, M. (2017). In the determination of the most efficient decision making unit in data envelopment analysis: a comment. *Computers & industrial engineering*, *104*, 216–218. DOI: 10.1016/j.cie.2016.12.032

[3] Amin, G. R., & Toloo, M. (2004). A polynomial-time algorithm for finding ε in DEA models. *Computers & operations research*, *31*(5), 803–805. DOI: 10.1016/S0305-0548(03)00072-8

[4] Cooper, W. W., Seiford, L. M., & Tone, K. (2007). *Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software*. Springer, New York, NY. DOI: 10.1007/b109347

[5] Andersen, P., & Petersen, N. C. (1993). A procedure for ranking efficient units in data envelopment analysis. *Management science*, *39*(10), 1261–1264. DOI: 10.1287/mnsc.39.10.1261

[6] Ertay, T., Ruan, D., & Tuzkaya, U. R. (2006). Integrating data envelopment analysis and analytic hierarchy for the facility layout design in manufacturing systems. *Information sciences*, *176*(3), 237–262. DOI: 10.1016/j.ins.2004.12.001

[7] Toloo, M., & Salahi, M. (2018). A powerful discriminative approach for selecting the most efficient unit in DEA. *Computers & industrial engineering*, *115*, 269–277. DOI: 10.1016/j.cie.2017.11.011

[8] Foroughi, A. A. (2011). A new mixed integer linear model for selecting the best decision making units in data envelopment analysis. *Computers & industrial engineering*, *60*(4), 550–554. DOI: 10.1016/j.cie.2010.12.012

[9] Wang, Y.-M., & Jiang, P. (2012). Alternative mixed integer linear programming models for identifying the most efficient decision making unit in data envelopment analysis. *Computers & industrial engineering*, *62*(2), 546–553. DOI: 10.1016/j.cie.2011.11.003

[10] Toloo, M. (2015). Alternative minimax model for finding the most efficient unit in data envelopment analysis. *Computers & industrial engineering*, *81*, 186–194. DOI: 10.1016/j.cie.2014.12.032