Available online at http://sanad.iau.ir/journal/ijim/
Int. J. Industrial Mathematics (ISSN 2008-5621)
Vol.17, No.3, 2025 Article ID IJIM-1629, 18 pages
Research Article

Science and Research Branch (IAU)

A Novel TOPSIS-Least Squares Approach for Classifying DMUs: A Comparative Analysis with DEA

F. Zarabi [1], F. Rezai Balf [2 *]
[1,2] Department of Mathematics, Sari. C., Islamic Azad University, Sari, Iran.

Submission Date:2025/04/07, Revised Date:2025/08/16, Date of Acceptance:2025/09/04

Abstract
Data Envelopment Analysis (DEA) is a non-parametric mathematical programming technique used to evaluates the performance of a set of homogeneous Decision-Making Units (DMUs) and assess their efficiency. DEA differentiates efficient units from inefficient ones by establishing a production or efficiency frontier. Any unit located on this frontier is considered efficient, while others are classified as inefficient. This paper presents a novel approach to distinguishing between efficient and inefficient units, and compares its results with the DEA technique. Our method is based on the combination of two techniques: TOPSIS and Least Squares (TLS). First, the TOPSIS method is employed, where the distance of each DMU from the positive ideal solution is considered as an input index, and its distance from the negative ideal solution is regarded as an output index. Then, using the Least Squares method, the units are separated into two groups: efficient and inefficient. The proposed TLS approach offers several advantages over the traditional CCR-DEA model, including computational similarity, flexibility, and transparency in classification. These features make it a valuable alternative for scenarios where DEA may not provide sufficient adaptability.

Keywords: Data Envelopment Analysis, Classifying, TOPSIS, Least Squares.

* Corresponding author: Email: fr.balf@iau.ir

1. Introduction

Data Envelopment Analysis (DEA), originating from Farrell's [1] seminal work and operationalized by Charnes, Cooper, and Rhodes [2], is a prominent technique for estimating production frontiers and assessing the relative efficiency of homogeneous Decision-Making Units (DMUs). A DMU is defined as an entity that transforms a set of inputs into a set of outputs. Essentially, DEA is a non-parametric linear programming method that identifies best practices within a group of organizations or firms, relative to the observed best performers in that group. DEA effectively measures the relative efficiency of DMUs with multiple inputs and outputs, categorizing them into two groups: efficient (either weakly or strongly) and inefficient units, based on the efficiency frontier (Amin et al. [3]). To appreciate the significance of the DEA technique, one need only examine the past four decades, which have witnessed the publication of numerous studies applying DEA across various disciplines.

In the context of improving a DMU's efficiency [4], it is essential to first detect inefficiencies and subsequently devise appropriate strategies to address them. Consequently, identifying and classifying firms has become a crucial step, enabling decision-makers (DMs) to select one or more superior or efficient DMUs [5]. The selection of an appropriate classification method depends on the performance indicators and insights of DMs, as well as the specific context of the problem. Given the importance of resource constraints in production, DMs must make precise management decisions and implement necessary adjustments in the allocation of input resources to optimize outputs [6].

The classification of DMUs is influenced by various factors, with one of the most critical being returns to scale (RTS) and the shape of the production frontier [7-11].

While numerous studies and applications highlight DEA's effectiveness [12-18], two key challenges remain. First, DEA models may not fully align with the axioms of returns to scale (RTS) and convexity. In production economics, RTS is defined as the maximal proportional increase in outputs resulting from a given proportional increase in inputs. However, the convexity assumption imposed in DEA is not always supported by empirical evidence, especially in cases involving specific input-output relationships. While non-convex DEA models, such as the free Disposal Hall (FDH) model [19], exist, they do not provide a comprehensive solution for addressing non-convexity. Second, the efficiency frontier in DEA is highly sensitive when classifying units as efficient or inefficient. Even minor errors in input or output values can result in significant misclassifications, particularly when identifying efficient units [20,21].

Although methods like discriminant analysis (DA) have been proposed for classification [22,23], this paper introduces a novel approach for dividing units into efficient and inefficient groups without relying on DEA, DA, or similar techniques. Notably, our method pays greater attention to certain inefficient units often overlooked by DEA and reclassifies some DEA-efficient units as inefficient. This approach allows decision-makers to focus more on inefficient units, fostering their growth and productivity.

This paper focuses on firms with multiple inputs and outputs, aiming to classify them into two groups through a two-step process. In the first step, using the TOPSIS method, we calculate each firm's distance from the positive and negative ideal solutions. For each obtained distance, we create a New DMU (NDMU) comprising an input index (distance from the positive ideal solution) and output index (distance from the negative ideal solution). In the second step, we employ Least Squares (LS) method to identify a linear function that separates the set of NDMUs into two distinct categories: one containing efficient NDMUs, those located on the side of the positive ideal solution, and the other includes inefficient NDMUs, which are on the side of the negative ideal solution. Hosseinzadeh Lotfi et, al. [24] introduced a novel hybrid approach based on TOPSIS and least squares support vector machines for decision making in DEA.

The remainder of this paper is organized as follows. Section 2 provides a concise review of the literature on DEA, TOPSIS, and Least Squares (TLS). Section 3 outlines the mathematical algorithm and methodology. Section 4 presents a numerical example and a real-world application of the proposed method, involving 20 and 28 DMUs, respectively. Finally, Section 5 concludes the paper and suggests directions for future research.

2. Literature Review

In this section, we focus on three fundamental topics: Data Envelopment Analysis (DEA), the TOPSIS method, and the Least Squares (LS) approach.

2.1. The CCR Model in DEA

DEA provides a comprehensive framework of methodologies and techniques to assess the efficiency of units operating with multiple inputs and outputs. Its strength lies in leveraging linear programming for performance evaluation. Using the inputs and outputs of the DMUs and the principles underpinning their operations, Farrell [1], introduced the concept of a production possibility set and identified a segment of its frontier as a production function. In economic terms, the production function represents the process by which multiple inputs are transformed into a single output. The frontier of this production function denotes the maximum attainable outputs, often referred to as the efficiency frontier for DMUs with multiple inputs and outputs. Units positioned on this frontier are classified as efficient, while those falling below it are deemed inefficient. Building on Farrell's seminal work, Charnes et al. [2], extended the concept into a structured mathematical framework.

DEA defines efficiency as the ratio of weighted outputs to weighted inputs, formulating the problem as one of fractional linear programming. Through the Charnes-Cooper transformation, this fractional model is converted into a standard linear programming format for practical implementation. In this paper, we analyze a set of n observed DMUs, where each unit utilizes m inputs to generate s outputs. Let each DMU_j, $j=1,...,n,$ contain input vector $x_j=(x_{1j},...,x_{mj})\in R^m \geq 0$ and output vector $y_j=(y_{1j},...,y_{sj})\in R^s \geq 0$. Let $u_1,...,u_s$ are weights for outputs and $v_1,...,v_m$ for inputs, too. The efficiency score for unit j, in present constant return to scale (CRS), is defined as $e_j=e(x_j,y_j)=\frac{\sum_{r=1}^{s}u_r y_{rj}}{\sum_{i=1}^{m}v_i x_{ij}}$, so that $e(x_j,y_j)=e(\lambda x_j,\lambda y_j),\ \lambda>0$.

The mathematical ratio model proposed by Charnes, Cooper and Rhodes (CCR) in 1978 is presented below, providing a framework to determine the relative efficiency of unit k.

$$
\begin{aligned}
\max\ & e_k = \frac{\sum_{r=1}^{s} u_r y_{rk}}{\sum_{i=1}^{m} v_i x_{ik}} \\
s.t.\ & \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \le 1, \quad j = 1, \ldots, n, \\
& u_r, v_i \ge 0, \qquad r = 1, \ldots, s;\ i = 1, \ldots, m.
\end{aligned}
\tag{1}
$$

The fractional linear form (1), of the DEA model represents a nonlinear problem. Charnes et al. [25], demonstrated that this fractional programming problem can be transformed as an equivalent linear programming formulation by $\sum_{i=1}^{m} v_i x_{ik} = \frac{1}{t}$ for $t > 0$.

$$
\begin{aligned}
\max\ & e_k = \sum_{r=1}^{s} u_r y_{rk} \\
s.t.\ & \sum_{i=1}^{m} v_i x_{ik} = 1, \\
& \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0, \quad j = 1, \ldots, n, \\
& u_r, v_i \ge 0, \qquad r = 1, \ldots, s;\ i = 1, \ldots, m.
\end{aligned}
\tag{2}
$$

This section presents some propositions and definitions (Charnes et al., [25]).

Proposition. Assume $\mathrm{e}_k^* = \max e_k$ is the optimal objective value when evaluating unit k in model (2). Then $e_k^* \le 1$.

Definition 2.1 *(Efficient/Inefficient).* Consider the unit k or DMU_k. If, in its evaluation, the value of the objective function (2) is equal to 1, then it is considered efficient. Otherwise, it is called inefficient.

Definition 2.2 *(Domination).* Let (x_A, y_A) and (x_B, y_B) represent two units A and B, respectively. We say that DMU_A dominates DMU_B if and only if $x_A \le x_B$ and $y_A \ge y_B$ with strict inequality holding for at least one component.

2.2. TOPSIS

The TOPSIS approach, introduced by Hwang and Yoon [26], in 1981, is a widely used and well-established method for solving multiple criteria decision-making (MCDM) problems. The TOPSIS method is particularly advantageous for decision-making scenarios where a balance between conflicting criteria (benefit and cost) is essential. Its simplicity, logical structure, and ability to visualize the relative performance of alternatives make it a powerful tool in various field, including engineering, economics, and management.

It provides a systematic and logical framework for ranking and selecting alternatives by comparing them to an ideal solution. The essence of TOPSIS is to identify solutions that are simultaneously closest to the positive ideal solution and farthest from the negative ideal solution.

Consider a MCDM problem represented in matrix form as $A_i = (x_{i1},\dots,x_{ij},\dots,x_{in}),\ i=1,\dots,m$, where x_{ij} is i^{th} alternative A_i, with respect to j^{th} criteria x_j. The procedure of TOPSIS can be related in a series of steps:

Step 1: Determine the positive and Negative Ideal Solutions

The positive ideal solution (A^+) and the negative ideal solution (A^-) are determined. These solutions may correspond to real alternatives or hypothetical, virtual alternatives. The definitions are given as:

$$\begin{aligned} A^+ &= (\alpha_1^+,\dots,\alpha_n^+) = \left\{(\max_i \alpha_{ij} \mid j\in I),(\min_i \alpha_{ij} \mid j\in J)\right\}, \\ A^- &= (\alpha_1^-,\dots,\alpha_n^-) = \left\{(\min_i \alpha_{ij} \mid j\in I),(\max_i \alpha_{ij} \mid j\in J)\right\}, \end{aligned} \tag{3}$$

where I is associated with benefit attribute, and J is associated with cost attribute. It is worth mentioning that if the elements of the decision-making matrix belong to the interval [0,1], then the positive ideal solution and the negative ideal solution can be determined as:

$$A^+ = (\alpha_j^+) = \begin{cases} 1, & j\in I \\ 0, & j\in J \end{cases}, \text{ and } A^- = (\alpha_j^-) = \begin{cases} 0, & j\in I \\ 1, & j\in J \end{cases}.$$

Step 2: Calculate the Euclidean Distances

The Euclidean distances of each alternative (A_i) from the positive ideal solution (A^+) and the negative ideal solution (A^-) are computed as:

$$d_i^+ = \sqrt{\sum_{j=1}^{n} (x_{ij}-\alpha_j^+)^2}, \qquad d_i^- = \sqrt{\sum_{j=1}^{n} (x_{ij}-\alpha_j^-)^2}, \tag{4}$$

Step 3: Compute the Relative Closeness to the Ideal Solution

The relative closeness (R_i) of each alternative (A_i) is calculated using the formula:

$$R_i = \frac{d_i^-}{d_i^- + d_i^+},\ R\in[0,1]. \tag{5}$$

Step 4: Rank the Alternatives

Finally, the alternatives are ranked based on their relative closeness index (R_i) in descending order. The alternative with the highest R_i is considered the best choice. This ranking can be used to prioritize DMUs effectively.

2.3. Least Squares

2.3.1. The Least Squares Method

The Least Squares Method is a widely used and well-established procedure for finding a function that best fits a set of scattered data points. The main objective of this method is to minimize the sum of squared differences (residuals) between the observed values and the values predicted by the model.

Consider a set of data points (x_i, y_i), where $i = 1,...,n$, which are treated as two-dimensional coordinates on the xoy plane. The goal is to find a model, such as a line, or any other type of function $f(x)$, that describes these points as well as possible. This best fit means minimizing the error of the model in predicting the values of y. The error for each data point is defined as the difference between the observed value (y_i) and the value predicted be the model ($f(x_i)$):

$$e_i = y_i - f(x_i). \tag{6}$$

$$E = \sum_{i=1}^{n} (y_i - f(x_i))^2. \tag{7}$$

2.3.2. Fitting a Straight Line

The central problem in this section is to determine the best-fitting straight line $y = ax + b$, for a given set of observed data points (x_i, y_i), $i \in 1,...,n$. The error associated with this model can be expressed as the sum of squared differences between the observed values and those predicted by the line. Specially, we define the error function as:

$$E(a,b) = \sum_{i=1}^{n} (y_i - (ax_i + b))^2. \tag{8}$$

The objective is to fine the values of a and b that minimize the error. To achieve this, we set the partial derivatives of the error function with respect to a and b equal to zero:

$$\frac{\partial E}{\partial a} = 0, \qquad \frac{\partial E}{\partial b} = 0. \tag{9}$$

By differentiating the error function $E(a,b)$, the following equations are obtained:

$$\begin{cases} (\sum_{i=1}^{n} x_i^2)a + (\sum_{i=1}^{n} x_i)b = \sum_{i=1}^{n} x_i y_i \\ (\sum_{i=1}^{n} x_i)a + \qquad nb \qquad = \sum_{i=1}^{n} y_i \end{cases} \tag{10}$$

Then the formulas for a and b are:

$$a = \frac{n\sum_{i=1}^{n}(x_i y_i) - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

$$b = \frac{(\sum_{i=1}^{n} x_i^2)(\sum_{i=1}^{n} y_i) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} x_i y_i)}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \tag{11}$$

As a result, the optimal values of a and b that minimize the error are calculated. Consequently, the fitted line is expressed as:

$$y = ax + b. \tag{12}$$

3. Proposed Methodology

Data Envelopment Analysis (DEA) is one of the most prominent and widely utilized techniques for categorizing decision-making units (DMUs) into efficient and inefficient groups. The methodology proposed in this paper seeks to enhance the classification of n decision-making units (x_i, y_i), $i = 1,...,n$, by integrating the TOPSIS method with the Least Squares approach, providing a more nuanced distinction between efficient and inefficient units.

Initially, the positive and negative ideal solutions are determined using a set of n DMUs. These ideal solutions are then incorporated into the set, increasing the total number of DMUs to $n+2$. Subsequently, the process initiated by employing the TOPSIS method to generate a new set of $n+2$ units (NDMUs), in the form of (d_i^+, d_i^-), $i = 1,...,n+2$, designed based on their respective distances from the positive and negative ideal solutions. To simplify the representation of subsequent formulas and relationships, the points (d_i^+, d_i^-) are denoted as $(\hat{x}_i, \hat{y}_i)$, which are treated as two-dimensional coordinates on the xoy plane. Next, the Least Squares method is employed to derive the regression line for this new set $(\hat{x}_i, \hat{y}_i)$, $j = 1,...,n+2$ as $y = ax + b$. It is worth emphasizing that the new units corresponding to the positive and negative ideal are denoted by $(\hat{x}_{n+1}, \hat{y}_{n+1}) = (0, d_{A^+A^-})$ and $(\hat{x}_{n+2}, \hat{y}_{n+2}) = (d_{A^+A^-}, 0)$, respectively.

To determine the normal line $y = Ax + \lambda B$, for classifying NDMUs into two groups, efficient and inefficient, and to ensure fair competition among them for inclusion in the efficient group, we consider $(\hat{x}_{n+1}, \hat{y}_{n+1})$ as the primary criterion for efficient group. Hence, this normal line is calculated for the set of points $(\hat{x}_i, \hat{y}_i)$, $i = 1,...,n$, and $(0, d_{A^+A^-}) \equiv (\hat{x}_{n+1}, \hat{y}_{n+1})$. The presence of $(\hat{x}_{n+1}, \hat{y}_{n+1})$ among the new units $(\hat{x}_i, \hat{y}_i)$, $i = 1,...,n$, may significantly influence the determination of the intercept for the normal line during fitting. This selection is specifically designed to assign units with higher priorities, which are closer to the positive ideal solution, to the efficient group. This can lead to a situation where the separating line fails to effectively classify the new units into efficient and inefficient groups. Specifically, all n new units could fall into the opposite half-space of $(\hat{x}_{n+1}, \hat{y}_{n+1})$, leaving no unit capable of competing as part of the efficient group. To address this issue, an arbitrary coefficient λ, is introduced to adjust the intercept. The inclusion of the coefficient λ in the intercept of the separating in between the groups enhances its flexibility in identifying a greater number of units as efficient. Accordingly, to achieve this objective, the parameter λ is decreased if the value B is positive, and increased if B is negative. By varying the values of λ, it becomes possible to identify and classify multiple layers can be tailored based on managerial preferences and decision-making strategies. Therefore, we compute the intercept of the normal line on the regression line, ensuring that it has a defined slope as $y = Ax + \lambda B$, with the goal of minimizing the associated error function (13).

$$\min \quad F(B) = \sum_{i=1}^{n+1} (\hat{y}_i - (A\hat{x}_i + \lambda B))^2 \tag{13}$$

Thus, by considering $F'(B) = 0$, concludes that,

$$\lambda B = \frac{\sum_{i=1}^{n+1} \hat{y}_i - A\sum_{i=1}^{n+1} \hat{x}_i}{n+1}, \tag{14}$$

where $A = \dfrac{-1}{a}$ and a is slope of regression line.

To summarize, the normal line $y = Ax + B$ to the regression line can be determined using the following equation:

$$\begin{aligned} &step\,1: \quad A = \frac{(n+2)\sum_{i=1}^{n+2} \hat{x}_i^2 - (\sum_{i=1}^{n+2} \hat{x}_i)^2}{\sum_{i=1}^{n+2} \hat{x}_i \sum_{i=1}^{n+2} \hat{y}_i - (n+2)\sum_{i=1}^{n+2} \hat{x}_i \hat{y}_i} \\ &step\,2: \quad B = \frac{\sum_{i=1}^{n+1} \hat{y}_i - A\sum_{i=1}^{n+1} \hat{x}_i}{n+1} \end{aligned} \tag{15}$$

Let $H^+ = \{(x,y) \mid y > Ax + \lambda B\}$, and $H^- = \{(x,y) \mid y < Ax + \lambda B\}$, then it obviously, $(0, d_{A^+A^-}) \in H^+$.

Then it can be shown that $|\lambda B| < d_{A^+A^-}$ or $\dfrac{-d_{A^+A^-}}{B} < \lambda < \dfrac{d_{A^+A^-}}{B}$.

The primary objective of this study is to classify these newly generated units $(\hat{x}_i, \hat{y}_i)$, $j = 1, \dots, n$ instead of the initial ones (x_i, y_i), $j = 1, \dots, n$.

Algorithm: Classification of Units into Efficient and Inefficient Groups Using TOPSIS-LS

Step 1: Initialize Data:

1.1. Let the set of initial DMUs be represented by $D = \{DMU_1, DMU_2, ..., DMU_n\}$.

Step 2: Determine Ideal Solutions:

2.1. Positive Ideal Solution (PIS): Identify the beat possible solution, PIS, for all DMUs.

2.2. Negative Ideal Solution (NIS): Identify the worst possible solution, NIS, for all DMUs.

Step 3: Expand the DMU set:

3.1. Include the ideal solutions into the set D, increasing the total number of DMUs to $n+2$, where the new set is $D' = D \cup \{PIS, NIS\}$

Step 4: Apply TOPSIS Method:

4.1. Using the expanded set D', apply the TOPSIS method to determine the relative closeness of each DMU to the ideal solutions, generating a new set of DMUs (NDMUs), denoted as $NDMUs = \{NDMU_1, NDMU_2, ..., NDMU_{n+2}\}$.

4.2. For each unit in $NDMUs$, calculate the distances to both the positive and negative ideal solutions.

Step 5: Represent Units in Two-Dimensional Space:

5.1. Represent each DMU in $NDMUs$ as a point on a two-dimensional plane (x, y), where the coordinates represent distances to the positive and negative ideal solutions.

Step 6: Apply Least Squares Method:

6.1. Fit a regression line $y = ax + b$ to the set of points representing $NDMUs$, where a is the slope, b is the intercept.

6.2. Derive the normal line to the regression line with the parameter λ as $y = Ax + \lambda B$, where $A = \frac{-1}{a}$.

Step 7: Classification of DMUs:

7.1. To classify the NDMUs into efficient and inefficient groups, calculate the normal line's intercept and slope.

7.2. If B is positive, decrease the parameter λ; if B is negative, increase it to adjust the intercept of the normal line and enhance its flexibility in categorizing NDMUs.

7.3. Adjust the normal line by introducing an arbitrary coefficient λ to refine the intercept and better identify efficient units.

Step 8: Ensure Fair Classification:

8.1. Ensure that the classification of NDMUs is fair by ensuring that the regression line's slop and intercept accurately reflect the true relationship between the NDMUs and the ideal solutions.

8.2. If the regression line fails to effectively separate efficient units, modify the normal line parameters to ensure the appropriate classification.

Step 9: Final Classification:

9.1. Classify the NDMUs as efficient or inefficient based on the final position relative to the normal line.

9.2. Efficient units: Those on the optimal side of the normal line.

9.3. Inefficient units: Those on the opposite side.

9.4. Align the categorization of units in set $\{NDMU_1, NDMU_2, ..., NDMU_n\}$ to corresponding precisely with the categorization of units in set $\{DMU_1, DMU_2, ..., DMU_n\}$.

4. Examples

In this section, we provide two illustrative examples: the first is a numerical example involving 20 Decision-Making Units (DMUs) drawn from the study by Rung-Wei et al. [6], and the second is an application example featuring 28 DMUs as reported by Charnes et al. [2].

4.1. Numerical example

We consider a set of 20 DMUs, each characterized by two distinct inputs and a single uniform output, with the output for all units standardized to a value of 1, as summarized in Table

Table 1. The set of 20 DMUs each with two inputs and one output

DMUs	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
x_1	1	2	3	5	2	3	3	4	5	4	6	7	7	7	8	9	10	11	10	11
x_2	5	3	2	1	5	4	8	8	9	10	5	5	4	3	4	2	3	3	1	2
y	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 2. The Classification of DMUs by TOPSIS-LS (TLS) and CCR-DEA for 20 DMUs

NDMUs	$\hat{x}$	$\hat{y}$	TOPSIS-LS $(\lambda=1)$	DMUs	e^*	CCR-DEA
1	4.0	11.2	E	1	1.0000	E
2	2.2	11.4	E	2	1.0000	E
3	2.2	11.3	E	3	1.0000	E
4	4.0	10.8	E	4	1.0000	E
5	4.1	10.3	E	5	0.7778	IE
6	3.6	10.0	E	6	0.7143	IE
7	7.3	8.2	IE	7	0.5000	IE
8	7.6	7.3	IE	8	0.4375	IE
9	8.9	6.1	IE	9	0.3684	IE
10	9.5	7.0	IE	10	0.3889	IE

11	6.4	7.1	IE	11	0.4545	IE
12	7.2	6.4	IE	12	0.4167	IE
13	6.7	7.2	IE	13	0.4667	IE
14	6.3	8.1	IE	14	0.5385	IE
15	7.6	6.7	IE	15	0.4375	IE
16	8.1	8.2	IE	16	0.5385	IE
17	9.2	7.1	IE	17	0.4375	IE
18	10.2	7.0	IE	18	0.4118	IE
19	9.0	9.1	IE	19	0.9955	IE
20	10.0	8.0	IE	20	0.4995	IE
A^{+}	0	13.5	-	-	-	-
A^{-}	13.5	0	-	-	-	-

According to Table 1, the positive and negative ideal solutions are identified as specific points, denoted as $A^{+}=(1,1,1)$ and $A^{-}=(11,10,1)$, respectively. The results for the 20 NDMUs and DMUs, including the processes d^{+}, d^{-}, classification by TLS, efficiency scores e^{*} from Model (2), and classification by DEA, are summarized in Table 2. In the DEA evaluation, four units (DMUs 1, 2, 3, and 4) are classified as efficient, while the remaining units are deemed inefficient. However, under the TLS evaluation with parameter $\lambda=1$, it is observed that six units (DMUs 1, 2, 3, 4, 5, and 6) fall into the efficient group, while the rest are classified as inefficient.

To calculate the slope of the regression line, all points in the NDMUs set are used. However, to determine the intercept of the normal line to the regression line, unit $(\hat{x}_{22},\hat{y}_{22})=(13.5,0)$ is excluded, while unit $(\hat{x}_{21},\hat{y}_{21})=(0,13.5)$ is retained. For determining the normal line, Table 2 provides the following results:

$$\sum_{i=1}^{22}\hat{x}_i=147.6,\ \sum_{i=1}^{22}\hat{x}_i^{\,2}=1204.64,\ \sum_{i=1}^{22}\hat{x}_i\hat{y}_i=1059.05,\ \sum_{i=1}^{21}\hat{x}_i=134.1,$$
$$\sum_{i=1}^{21}\hat{y}_i=\sum_{i=1}^{22}\hat{y}_i=182,$$

From relation (15), we obtain: $A=1.323$, $B=0.218$. Therefore, the normal line is expressed as $y=1.323x+0.218\lambda$ with $\lambda=1$, (See Fig. 1). For NDMUs located in the half-space with $(\hat{x}_{21},\hat{y}_{21})=(0,13.5)$, they must satisfy the inequality $y\geq 1.323x+0.218$, with $\lambda=1$ to be classified as efficient. Figure 2 illustrates the efficiency frontier in DEA, which consists of four extremes efficient DMUs. These efficient units dominate all other DMUs, thereby categorizing the dominated DMUs into the inefficient group. In Table 2, the label "E" denotes efficient units, while "IE" represents inefficient units.

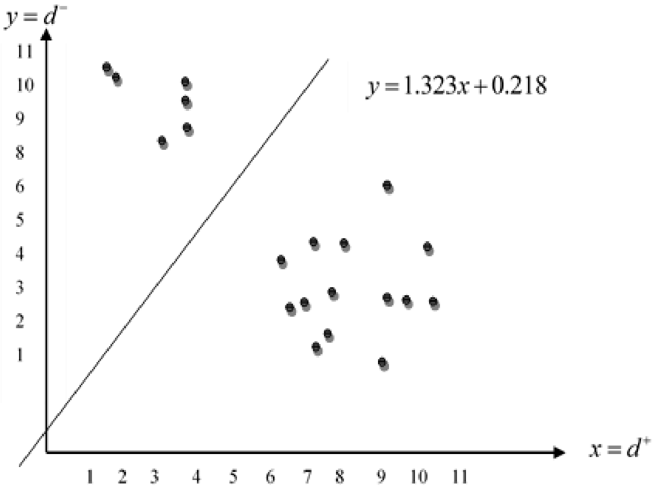

Fig. 1. Classifying NDMUs to two groups efficient and inefficient

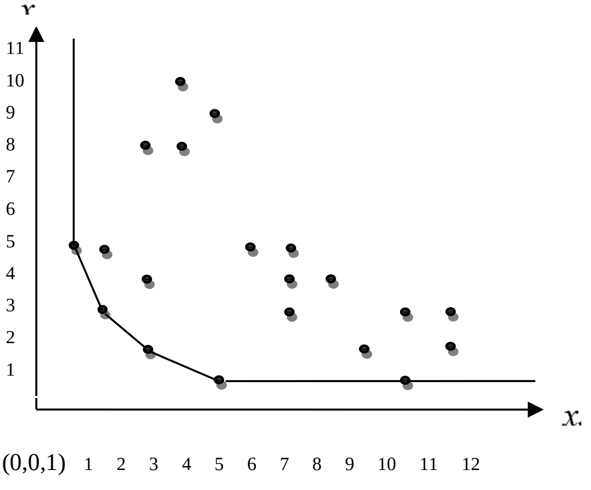

Fig. 2. Classification of DMUs by DEA

4.2. Application example

Consider a dataset of 28 DMUs, each characterized by three inputs and three outputs, as presented in Table 3. These data were originally reported by Charnes et al. [2], and represent 28 Chinese cities (DMUs) in 1983, provided in normalized form. The three outputs are gross industrial output value, profit and taxes, and retail sales, while the three inputs are labor, working funds, and investment.

With reference to Table 3, since data fall within the interval [0,1], the positive ideal point and negative ideal point are defined as $A^+(0,0,0,1,1,1)$ and $A^-(1,1,1,0,0,0)$, respectively. The results of the TLS method and the CCR-DEA Model (2), including the efficiency measure for the 28 DMUs, are summarized in Table 4. To find the normal line, the following calculations have been performed.

$\sum_{i=1}^{30}\hat{x}_i = 45.9,\ \sum_{i=1}^{30}\hat{x}_i^{\,2} = 73.6,\ \sum_{i=1}^{30}\hat{x}_i\hat{y}_i = 63.7,\ \sum_{i=1}^{29}\hat{x}_i = \sum_{i=1}^{30}\hat{y}_i = 43.4,\ \sum_{i=1}^{29}\hat{x}_i = 43.4,$

Relation (15) for TLS method is expressed as $y = 1.2771x - 0.4143\lambda$. For NDMUs located in the half-space defined by H^+, they must also satisfy in inequalityto $y \geq 1.2771x - 0.4143\lambda$ be classified as efficient.

The column 4,5 and 6 of Table 4 illustrate the classification of the 28 DMUs into two groups efficient and inefficient based on the TLS method, according to the parameter λ. For $\lambda = 1$, any NDMUs not belong to the efficient group. In case $\lambda = 1.1$, DMUs 8, 15, 21, 22, 23, 24, 25, and 26 are classified as efficient, so that most of them also being identified as efficient by the DEA method.

Table 3. The data for 28 Chinese cities

DMUs	Input 1	Input 2	Input 3	Output 1	Output 2	Output 2
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0.7701	0.6121	0.6248	0.3693	0.3418	0.7677
3	0.5553	0.4905	0.5542	0.3378	0.2551	0.4350
4	0.4183	0.3239	0.1903	0.1707	0.0852	0.3088
5	0.4098	0.3374	0.1826	0.1833	0.1285	0.2918
6	0.3705	0.3027	0.3075	0.1749	0.1192	0.5557
7	0.3065	0.2626	0.1572	0.0971	0.0542	0.2202
8	0.3932	0.2921	0.1814	0.1464	0.8853	0.3251
9	0.2553	0.1757	0.1489	0.1259	0.0848	0.2199
10	0.2420	0.2184	0.1486	0.0894	0.0491	0.1912
11	0.2684	0.2116	0.1498	0.1085	0.0717	0.2738
12	0.2200	0.1422	0.0867	0.0670	0.0421	0.2147
13	0.1857	0.1509	0.1550	0.0728	0.0495	0.1089
14	0.2262	0.2019	0.1365	0.1242	0.0935	0.2235
15	0.1770	0.1324	0.0800	0.1144	0.0733	0.2157
16	0.1494	0.1591	0.1198	0.0724	0.0739	0.1092
17	0.1577	0.1153	0.0778	0.0711	0.0425	0.1454
18	0.1516	0.1031	0.0702	0.0759	0.0720	0.0930
19	0.1795	0.1360	0.0897	0.0922	0.1085	0.1198
20	0.1430	0.1134	0.1080	0.0564	0.0465	0.1139
21	0.1608	0.0966	0.0749	0.1278	0.0409	0.2415
22	0.2017	0.1480	0.1072	0.1223	0.0804	0.2230
23	0.1138	0.0569	0.0700	0.0769	0.0234	0.1691
24	0.1387	0.1031	0.0703	0.1282	0.0545	0.1786
25	0.0959	0.0719	0.0509	0.0891	0.0351	0.1172
26	0.1348	0.0693	0.0456	0.0886	0.0233	0.2065
27	0.0416	0.0363	0.0886	0.0215	0.0074	0.0225
28	0.1445	0.0843	0.0502	0.0470	0.0199	0.1553
A^+	0	0	0	1	1	1
A^-	1	1	1	0	0	0

When $\lambda = 1.2$, three additional DMUs, namely 1, 6 and 19, are included in the efficient group. Moreover, Table 4 indicates that DMUs 1, 6, 8, 21, 23, 24, 25, and 26 are classified as efficient by the DEA method. This demonstrates a strong alignment between the classification produced by the DEA and TLS methods, as both yield a similar grouping of DMUs into efficient and inefficient categories. The consistency in results highlights the robustness of the TLS method in effectively identifying efficiency among DMUs.

Table 4. The Classification of DMUs by TOPSIS-LS (TLS) and CCR-DEA for 28 Chinese cities

NDMUs	$\hat{x}$	$\hat{y}$	TOPSIS-LS method			DMUs	e^*	CCR-DEA method
			$\lambda=1$	$\lambda=1.1$	$\lambda=1.2$			
1	1.7321	1.7321	IE	IE	E	1	1.00	E
2	1.4977	1.0893	IE	IE	IE	2	0.72	IE
3	1.4726	1.0122	IE	IE	IE	3	0.66	IE
4	1.5226	1.2581	IE	IE	IE	4	0.52	IE
5	1.4977	1.2613	IE	IE	IE	5	0.58	IE
6	1.4062	1.3098	IE	IE	E	6	1.00	E
7	1.5829	1.3401	IE	IE	IE	7	0.49	IE
8	1.2125	1.5653	IE	E	E	8	1.00	E
9	1.5259	1.4247	IE	IE	IE	9	0.63	IE
10	1.5862	1.3990	IE	IE	IE	10	0.54	IE
11	1.5242	1.4041	IE	IE	IE	11	0.70	IE
12	1.5751	1.4936	IE	IE	IE	12	0.65	IE
13	1.6243	1.4552	IE	IE	IE	13	0.45	IE
14	1.5173	1.4337	IE	IE	IE	14	0.72	IE
15	1.5210	1.5302	IE	E	E	15	0.87	IE
16	1.6042	1.4927	IE	IE	IE	16	0.60	IE
17	1.5982	1.5397	IE	IE	IE	17	0.64	IE
18	1.6051	1.5519	IE	IE	IE	18	0.77	IE
19	1.5660	1.5110	IE	IE	E	19	0.66	IE
20	1.6216	1.5279	IE	IE	IE	20	0.57	IE
21	1.5155	1.5661	IE	E	E	21	1.00	E
22	1.5145	1.4938	IE	E	E	22	0.79	IE
23	1.5866	1.6046	IE	E	E	23	1.00	E
24	1.5374	1.569	IE	E	E	24	1.00	E
25	1.5991	1.6132	IE	E	E	25	1.00	E
26	1.5618	1.6052	IE	E	E	26	1.00	E
27	1.7056	1.6367	IE	IE	IE	27	0.54	IE
28	1.6164	1.5809	IE	IE	IE	28	0.70	IE
A^+	0.0000	2.4495	-	-	-	-	-	-
A^-	2.4495	0.0000	-	-	-	-	-	-

5. Analytical Comparison: TOPSIS-LS vs. CCR Model

Building upon the algorithm presented, a thorough and detailed comparison between the TOPSIS-LS and the CCR-DEA method can be conducted. The key aspects of this comparison are highlighted below.

1. Difference in Classification Approach

1.1. CCR-DEA:

The CCR model classifies DMUs by solving a mathematical programming problem to define an efficiency frontier. Units on this frontier are classified as efficient, while others are considered inefficient.

1.2. TOPSIS-LS:

The proposed method uses a combination of TOPSIS and Least Squares. Instead of defining an efficiency frontier, it employs a normal line based on distances to positive and negative ideal solution (PIS and NIS) for classification. This approach provides more flexibility in differentiating efficient and inefficient units.

2. Type of Model

2.1. CCR-DEA:

While non-parametric, the CCR model relies on linear programming to calculate efficiency scores.

2.2. TOPSIS-LS:

The proposed method is fully non-parametric and does not require any specified model. It uses distances and regression techniques to classify units.

3. Computational Complexity

3.1. CCR-DEA:

The CCR model involves solving linear programming problems for each DMU, which becomes computationally expensive as the number of DMUs increases.

3.2. TOPSIS-LS:

The combination of TOPSIS and regression method is computationally simpler and more efficient, especially when dealing with a large number of DMUs.

4. Flexibility in Classification

4.1. CCR-DEA:

The classification boundary is fixed and solely determined by the efficiency frontier, leaving no room for adjustment.

4.2. TOPSIS-LS:

The proposed method allows for adjusting the parameter of the normal line (e.g., intercept), making it possible to refine the classification boundary for greater accuracy.

5. Application Contexts

5.1. CCR-DEA:

Primarily suitable for evaluating relative efficiency based on input-output data, often used in benchmarking studies.

5.2. TOPSIS-LS:

Better suited for scenarios requiring flexible classification, particularly when input/output relationships are unclear or when a more tailored distinction between DMUs is required.

6. Advantages of the Proposed Method

6.1. Eliminates the dependency on mathematical programming models.

6.2. Simpler and easier to implement with adjustable parameters.

6.3. The integration of TOPSIS and Least Squares provides more transparent and precise classification of DMUs.

6.4. Particularly effective in cases with diverse data or when DEA's fixed efficiency frontier is too restrictive.

6. Conclusions

The accurate identification and classification of efficient and inefficient firms hold significant importance in management sciences. Occasionally, a decision-maker worries that some firms may not be properly classified. Decision-makers often face concerns about the potential misclassification of firms, which could lead to suboptimal decisions and outcomes.

In this paper, we proposed a novel method aimed at addressing this issue, particularly in cases where traditional approaches, such as Data Envelopment Analysis (DEA), may incorrectly classify firms. For instance, a firm deemed inefficient by DEA might be considered efficient under the proposed method, and vice versa.

This study has thoroughly examined the similarities and distinctions between the TLS method and DEA, specifically in terms of their ability to classify decision-making units (DMUs) as efficient or inefficient. The findings suggest that the TLS method offers a distinct advantage by enabling the classification of efficient DMUs into multiple layers based on the efficiency parameter, providing a more granular understanding of efficiency.

The authors believe that the proposed method can significantly enhance future research efforts in this area. In particular, further investigation into the behavior of this method under various conditions, such as imprecise, bounded, ordinal, or fuzzy data, could yield valuable insights. These extensions could pave the way for broader applications and a deeper understanding of the proposed method's potential in complex decision-making environments.

References

[1] Farrell M J (1957) The Measurement of Productive Efficiency. Journal of Royal Statistical Society, Series A, 120: (3) 253-290.

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

[3] Amin GR, Emrouznejad A, Rezaei S (2011) Some clarifications on the DEA clustering approach, European Journal of Operational Research 215: 498-501.

[4] Hosseinzadeh Lotfi F., Didehkhani H. (2021). Developing a Model for Determining Practical Super Efficiency and Improvement Based on Artificial DMUs. Research Annals of Industrial and Systems Engineering, Vol. 2, No. 1, pp. 63-77.

[5] Hosseinzadeh Lotfi F., Jahanshahloo G. R., Moghaddas Z. (2013). Ranking of DMUs by Using TOPSIS and Different Ranking Models in DEA. International Journal of Industrial Mathematics, Vol. 5, No. 3, pp. 119-132.

[6] Rung-Wei P, Yuh-Yuan G, Miin-Shen Y (2009) A new clustering approach using data envelopment analysis, European Journal of Operational Research 199: 276-284.

[7] Golany B, Yu G (1994) Estimating returns to scale in DEA. European Journal of Operational Research 103: 28-37.

[8] Banker RD, Bardhan I, Cooper WW (1996a) A note on returns to scale in DEA. European Journal of Operational Research 88: 583-585.

[9] Seiford LM, Zhu J (1998a) On alternative optimal solutions in the estimation of returns to scale in DEA. European Journal of Operational Research 108: (1) 149-152.

[10] Jahanshahloo GR, Hosseinzadeh Lotfi F, Zhiani Rezai H, Rezai Balf F (2007) Finding strong defining hyperplanes of Production Possibility Set European Journal of Operational Research 177: 42-54.

[11] Ouellette P, Quesnel J-P, Vigeant S (2012) Measuring returns to scale in DEA models when the firm is regulated, European Journal of Operational Research 220: (216) 571-576.

[12] Zhongbao Zh, Helu X, Qianying J, Wenbin L (2018) DEA frontier improvement and portfolio rebalancing: An application of China mutual funds on considering sustainability information disclosure, European Journal of Operational Research 269: (116) 111-131.

[13] Panagiotis M, Michael A, Talias, Ioannis M (2015) Combining stochastic DEA with Bayesian analysis to obtain statistical properties of the efficiency scores: An application to Greek public hospitals, European Journal of Operational Research 243: (116) 302-311.

[14] Tammie AN, David M, Jeffrey AA (2009) Employing super-efficiency analysis as an alternative to DEA: An application in outpatient substance abuse treatment, European Journal of Operational Research 196: (31) 1097-1106.

[15] Sungmook L, Kwang WO, Joe Z (2014) Use of DEA cross-efficiency evaluation in portfolio selection: An application to Korean stock market, European Journal of Operational Research 236: (11) 361-368.

[16] Edirisinghe NCP, Zhang X (2010) Input/output selection in DEA under expert information, with application to financial markets, European Journal of Operational Research 207: (316) 1669-1678.

[17] Sahoo BK, Luptacik M, Mahlberg B (2011) Alternative measures of environmental technology structure in DEA: An application, European Journal of Operational Research 215: (316) 750-762.

[18] Odeck J (2000) Assessing the relative efficiency and productivity growth of vehicle inspection services: An application of DEA and Malmquist indices, European Journal of Operational Research 126: (31) 501-514.

[19] Tulkens H (1993) On FDH Analysis: Some Methodological Issues and Applications to Retail Banking, Courts and Urban Transit, Journal of Productivity Analysis 4: (1-2) 183-210.

[20] Tsionas EG (2003) Combining DEA and stochastic frontier models: An empirical Bayes approach, European Journal of Operational Research 147: (316) 499-510.

[21] Charnes A, Cooper WW, Sueyoshi T (1986a) Least square/ridge regression and goal programming/constrained regression alternatives. European Journal of Operational Research 27: (2) 146–157.

[22] Sueyoshi T (1999) DEA-discriminant analysis in the view of goal programming, European Journal of Operational Research 115: 564-582.

[23] Jahanshahloo GR, Hosseinzadeh Lotfi F, Rezai Balf F, Zhiani Rezai H (2007) Discriminant analysis of interval data using Monte Carlo method in assessment of overlap Applied Mathematics and Computation 191: 521-532.

[24] Hosseinzadeh Lotfi F., Zegordi S. H. (2018). A novel hybrid approach based on TOPSIS and least squares support vector machines for decision making in DEA. *Applied Soft Computing*, 67, 521-533.

[25] Charnes A, Cooper WW, Rhodes E (1981) Evaluating program and managerial efficiency: An application of data envelopment analysis to program follow through. Management Science 27: 668-697.

[26] Hwang CL, Yoon KP (1981) Multiple Attribute Decision Making: Methods and Applications. Berlin/Heidelberg/New York: Springer-Verlag.