Journal of Mathematics and Modeling in Finance (JMMF)
Vol. 6, No. 1, Winter & Spring 2026
Research paper

A New Clusterless DEA Cross-Efficiency Evaluation in the Presence of Negative Data and its Application in Portfolio Selection

Kumari, Reenu[1]

[1] Maharaja Surajmal Institute of Technology, Janakpuri, Delhi, India
reenu_kumari@msit.in

Abstract:
This study examines the impact of clustered Decision-Making Units (DMUs) in DEA cross-efficiency evaluation, taking into account variables with both positive and negative values. The Range Directional Measure (RDM) model is often used when DMUs have both positive and negative data. However, its application frequently results in clustered DMUs, where multiple units obtain identical RDM cross-efficiency scores, thereby reducing discriminatory power and limiting the reliability of rankings. To overcome this drawback and enhance ranking reliability, we characterize clustered DMUs as scenarios in which identical scores lead to groups of DMUs being termed as 'clustered'. The necessity of this research lies in improving the robustness of efficiency analysis, especially in situations where both positive and negative data are involved. We then present an algorithm to identify potential clusters and propose a novel clusterless cross-efficiency evaluation method, which restores discrimination and provides more credible performance analysis in decision-making contexts. To demonstrate the practical relevance and advantages of the proposed method, a case study on stock selection in Iran's stock market portfolio is provided.

Keywords: Decision-making units, Cross-efficiency, Clusters, Portfolio, Stock selection
Classification: 91G10,90C05,90C08

1 Introduction

The cross-efficiency technique, proposed by Sexton et al. [1], is a powerful methodology to rank the DMUs. This allows the assessment of DMUs' overall efficiencies through both self-assessment and peer evaluation. Self-assessment enables the DMUs' efficiencies to be assessed using the most beneficial weights, allowing each DMU to attain its highest relative efficiency. Conversely, peer evaluation assesses each DMU's efficiency using weights determined by the other DMUs. Despite its widespread application, the method has several drawbacks, including the non-uniqueness of the DEA optimal weights, which may limit the utility of cross-efficiency assessment. Doyle and Green [2] address the non-uniqueness issue and propose several secondary

[1] Corresponding author
Received: 24/10/2025 Accepted: 05/01/2026
https://doi.org/10.22054/jmmf.2026.89074.1229

goals to resolve it. They present aggressive and benign formulations, which represent opposing strategies. The benevolent formulation aims to determine optimal weights that maximize the efficiency of the DMU in question and the average efficiency of the other DMUs. In contrast, the optimal weights for the aggressive angle seek to maximize the efficiency of the evaluated DMU while reducing the average efficiency of the other units. More detailed literature on the DEA cross-efficiency and its extensions can be found in [3–5].

The above literature ignores an inherent flaw in the cross-efficiency assessment. The cross-efficiency assessment relies on the CCR model introduced by Charnes et al. [6], which prevents it from handling negative inputs or outputs. Yet many practical contexts include negative values in either inputs or outputs. For example, when we evaluate the operational efficiency of bank branches, we typically use staff count as an input and transaction volume as an output, and both variables are positive for every branch. However, other outputs like changes in accounts and deposits may be negative for certain banks [7]. Portela et al. [7] proposed a range directional measure (RDM) using the directional distance function (DDF). The RDM model can manage inputs and outputs with both positive and negative values by using the range of possible improvement as the direction vector, i.e., the difference between the initial evaluated value and the maximum or minimum value. Lin [8] introduced a method for cross-efficiency evaluation using the DDF, leveraging the RDM model as presented by Portela et al. [7] and duality theory within the context of VRS technology. The approach by Lin [8] is adept at handling scenarios where negative values exist in the input–output data, and it effectively resolves the challenge of negative cross-efficiency. Soltainfar and Sharafi, [9] initially introduced a new non-radial model to assess the performance of DMUs when negative data is present. Subsequently, using this model, they proposed a hybrid MADM-DEA approach that incorporates the fuzzy VIKOR method. Chen et al. [10] integrated prospect theory to formulate a novel cross-efficiency aggregation method, referred to as the APC method. Chen et al. [11] extend DEA cross-efficiency by incorporating prospect theory and a distance entropy function to capture subjective risk preferences. Kao and Liu [12] proposed the slacks-based efficiency measure to address the problem of negative cross-efficiencies in conventional cross-efficiency evaluation.

DEA cross-efficiency can be applied in portfolio selection to determine an optimal distribution of assets, with the objective of reducing risk while attaining a desired level of expected return. Initially, Lim et al. [13] suggested combining DEA cross-efficiency with the MV model to create the DEA M-V cross-efficiency model and address the problem of ganging of DMUs. Gong et al. [14] presented the regret cross-efficiency model to assess the DEA cross-efficiency scores of assets within a fuzzy multi-objective framework. Amin and Hajjami [15] explore the significance of alternative optimal solutions in DEA cross-efficiency evaluation for portfolio selection. They demonstrate that integrating alternative optimal solutions into the construction of a cross-efficiency matrix enhances the outcomes of the MV (mean-variance) portfolio selection method. Kumari et al. [16] introduced a novel method for cross-efficiency aggregation and applied the method to portfolio selection problems. Shrivastava et al. [17] integrate cross-efficiency with cumulative prospect theory to explore the psychological facets of decision-makers in portfolio selection. Deng and Feng [18] concentrate on integrating DEA prospect cross-efficiency assessment into fuzzy portfolio optimization within a new mean-variance-maverick (MVM) model. Banihashemi and Sanei [19] used RDM cross-efficiency in portfolio selection by using OWA operator weights for aggregation of cross-efficiencies. Kumari et al. [20] proposed a method for stock categorization using the column and row average value of the cross-efficiency matrix.

While the DEA cross-efficiency technique has been used effectively in ranking DMUs, there are still some limitations that restrict its application. One such well-known problem is the ganging-together issue of cross-efficiencies as discussed by Amin and Oukil [21]. They show that the ganging phenomenon can significantly influence the cross-efficiency evaluation in favor of some DMUs. Therefore, it would be unjust to use these cross-efficiency scores as a basis for any decision-making. We extend the idea of clustered or ganged DMUs presented by Amin and Oukil [21] for problems where positive as well as negative data can be present.

1.1 Research Gap and Motivation

To the best of our knowledge, this is the first study to examine the impact of clustered DMUs in the context of negative input–output data. We illustrate this through a practical application that highlights how clustered cross-efficiency assessment impacts the stock selection. To address this issue, we propose a novel cross-evaluation approach in which cross-efficiency values are computed after excluding the clustered DMUs/stocks from the calculation. Furthermore, we provide an algorithm to systematically detect and identify such clustered DMUs.
In summary, this study highlights the following research gaps:

(i) Limited attention has been given in the literature to cross-efficiency analysis when negative data are involved.

(ii) Conventional DEA cross-efficiency suffers from the drawback that clustered DMUs can distort cross-evaluation results.

(iii) From a practical perspective, investors and decision-makers require reliable evaluation tools that not only identify the most efficient stocks/DMUs but also ensure fairness in rankings, unaffected by distortions caused by clustering. Existing methods do not fully satisfy this requirement.

The remainder of this study is structured as follows: Section 2 discusses the DEA technique, which effectively handles negative data. Section 3 outlines the RDM cross-efficiency evaluation designed to address negative data. Section 4 introduces the concept of clustered DMUs, along with an efficient algorithm for cluster identification and an evaluation method for clusterless cross-efficiency. In Section 5, we discuss the impact of the introduced clusterless cross-efficiency assessment on the top-performing stocks in the Tehran stock market. Finally, Section 6 provides the concluding remarks.

2 DEA in the presence of negative input-output

In conventional DEA methods, each DMU is represented by non-negative input and output vectors, where inputs are transformed into outputs. However, these methods are not applicable when DMUs involve both positive and negative input and/or output values. To address this issue, Portela et al. [7] introduced the RDM model. Their approach is defined by an ideal point and a directional vector, where the directional vector is constructed from the possible improvement range—i.e., the difference between the initial evaluated value and its corresponding maximum or minimum. Building on the ideas underlying the RDM model, Sharp et al. [22] developed a modified slack-based measure (MSBM) model, and Emrouznejad et al. [23] introduced the semi-oriented radial measure (SORM) model, which are specifically designed to manage datasets containing both positive and negative inputs and outputs. It is important to note that the MSBM model was developed specifically for "naturally negative" inputs, as defined by Sharp et al. [22]. Because of this narrow focus, its applicability is more limited compared to the RDM and SORM models. Furthermore, a limitation of SORM relative to RDM is its tendency to increase the dimensionality of the model by treating the negative component of a variable as a separate dimension, which can make the evaluation process more complex and computationally demanding. Therefore, in this study, we employ the RDM model to assess the performance of the DMUs, as presented below.

Consider a set of n DMUs, denoted by DMU_j for $j = 1, 2, \ldots, n$, where each DMU utilizes m inputs x_{ij} $(i = 1, 2, \ldots, m)$ to produce s outputs y_{rj} $(r = 1, 2, \ldots, s)$. Let DMU_k $(k = 1, 2, \ldots, n)$ represent the specific DMU under evaluation. The RDM (Range Directional Measure) model can then be formulated as:

Model I

$$\begin{aligned}
&\max \ \beta_k \\
&\text{s.t.} \\
&\sum_{j=1}^{n} \lambda_j x_{ij} \le x_{ik} - \beta_k d_{ik}^{-}, \quad i = 1, \ldots, m, \\
&\sum_{j=1}^{n} \lambda_j y_{rj} \ge y_{rk} + \beta_k d_{rk}^{+}, \quad r = 1, \ldots, s, \\
&\sum_{j=1}^{n} \lambda_j = 1, \ \ \lambda_j \ge 0, \\
&j = 1, \ldots, n, \beta_k \ \text{ is free.}
\end{aligned}$$

Where d_{ik}^{-} and d_{rk}^{+} are the potential improvement of non-negative ranges for inputs and outputs is represented as direction vectors and defined as follows:

$$d_{ik}^{-} = x_{ik} - \min_j \{x_{ij}\}, \ \ i = 1, 2, \ldots, m \ \text{ and } \ d_{rk}^{+} = \max_j \{y_{rj}\} - y_{rk}, \ \ r = 1, 2, \ldots, s$$

The value β_k represents the inefficiency of DMU$_k$. Thus, $(1 - \beta_k)$ is the efficiency measure, indicating the distance between the observed and the desired input.

3 RDM cross-efficiency assessment

To evaluate the cross-efficiency of DMUs in the presence of negative data, determining optimal weight values is essential. Hence, to establish a cross-efficiency structure analogous to classical DEA, the RDM model is converted into its dual form. Based on standard linear programming duality and normalization using direction vector constraints, Lin [8] proposed a DDF-based cross-efficiency evaluation method capable of handling negative input-output variables. The formulation of Lin's model [8] is presented as follows:
Model II

$$\begin{aligned}
\min \quad & \sum_{i=1}^{m} \omega_{ik}(x_{ik} + d_{ik}^{-}) - \sum_{r=1}^{s} \mu_{rk}(y_{rk} - d_{rk}^{+}) + \psi \\
\text{s.t.} \quad & \sum_{i=1}^{m} \omega_{ik} x_{ij} - \sum_{r=1}^{s} \mu_{rk} y_{rj} + \psi \ge 0, \quad j = 1, 2, \ldots, n, \\
& \sum_{i=1}^{m} \omega_{ik} d_{ik}^{-} + \sum_{r=1}^{s} \mu_{rk} d_{rk}^{+} = 1, \\
& \omega_{ik}, \mu_{rk} \ge 0, \quad \psi \in \mathbb{R}.
\end{aligned}$$

Where ω_{ik} and μ_{rk} are the weights for the i^{th} input and r^{th} output, of k^{th} DMU. The optimal solution of Model II provides the optimal weights corresponding to k^{th}

DMU, denoted as ω_{ik}^* and μ_{rk}^*. These weights are used to obtain the cross-efficiency of DMU$_j$ and are defined as follows:

$$\theta_{kj} = \frac{\sum_{i=1}^{m} \omega_{ik}^* \left(-x_{ij} + d_{ik}^-\right) + \sum_{r=1}^{s} \mu_{rk}^* \left(y_{rj} + d_{rk}^+\right) - \psi^*}{\sum_{i=1}^{m} \omega_{ik}^* d_{ik}^- + \sum_{r=1}^{s} \mu_{rk}^* d_{rk}^+}. \tag{1}$$

A matrix of cross-efficiencies is obtained as $E = (\theta_{kj}),\ k, j = 1, 2, \ldots, n$, where θ_{kj} is the cross-efficiency of DMU$_j$ evaluated using the optimal weights of DMU$_k$. The final assessment of each DMU is determined by averaging the cross-efficiency scores across each column of the cross-efficiency matrix E, and is defined as:

$$\bar{\theta}_j = \frac{1}{n} \sum_{k=1}^{n} \theta_{kj}, \quad j = 1, 2, \ldots, n$$

It is important to note that each θ_{kj} is referred to as a cross-efficiency score, while the average, $\bar{\theta}_j$, is also termed as cross-efficiency in the literature. Generally, "cross-efficiency" refers to the average score. Hence, $\bar{\theta}_j$ is the final RDM cross-efficiency value.

Remark 3.1. The step-by-step derivations underlying Model II—including the dual transformation, the normalization of directional weights, and the establishment of Pareto–Koopmans efficiency conditions—are fundamental to the theoretical structure of the range-based DEA framework. Lin [8] provides the complete theoretical development linking Model I to the normalized cross-efficiency formulation presented as Model II. In order to remain concise and avoid repeating these well-established derivations, the present study relies on Lin's formulation and uses Model II directly as the adopted cross-efficiency measure.

One drawback of using DEA cross-evaluation is the occurrence of clustering or ganging DMUs. The clustered DMUs can significantly bias the cross-efficiency evaluation in favor of specific DMUs [21]. Thus, it would be unsuitable to utilize these cross-efficiency scores for cross-evaluation purposes. We will define the concepts of clustered DMU and clusterless cross-efficiency assessment in the next section.

4 Proposed Methodology: The clusterless cross-efficiency assessment

In this section, we formally define clustered DMUs and present an algorithm for cluster detection. Subsequently, we compute the clusterless cross-efficiency scores of DMUs for performance evaluation. The proposed methodology involves three main steps: (i) detection of clusters, (ii) evaluation of clusterless cross-efficiency, and (iii) Statistical validation of clusters.

4.1 Detection of Clusters

Let $C = \{1, 2, \ldots, c\}$ represent a set of possible clustered DMUs. The purpose of this step is to identify groups of DMUs that demonstrate identical cross-efficiency scores. Such DMUs are said to be clustered, as their performance evaluations are indistinguishable across all peer appraisals.

Definition 4.1 (Clustered DMUs)**.** A group of at least two DMUs indexed in C is said to form a cluster if and only if their cross-efficiency values are identical across all evaluations. More formally, for any pair of DMUs $p, q \in C : \theta_{pj} = \theta_{qj} \ \ \forall j = 1, 2, \ldots$. In this case, DMU_p and DMU_q belong to the same cluster.

This definition implies that clustered DMUs behave identically under the cross-efficiency evaluation. Therefore, retaining all of them in the ranking process may bias the results by overweighting their identical performance. To avoid this, it becomes necessary to detect clusters and treat them appropriately. We propose an algorithm to detect all clusters, with its flowchart shown in Figure 1 and the step-by-step procedure detailed in Algorithm 1.

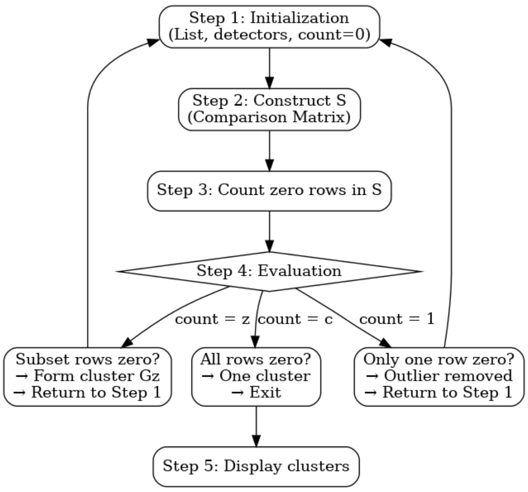

Figure 1: Flowchart for Cluster Detection

This algorithm systematically partitions the set of DMUs into one or more clusters, depending on their similarity patterns.

Remark 4.2. The proposed process is independent of the presence of multiple optimal solutions, as clustering is determined by the relative efficiency values rather than by the multiplicity of optimal weight vectors. It is important to note, however, that under one optimal solution, a particular cluster may exist, whereas under another optimal solution, the identified clusters may differ or may not appear at all. Therefore, careful consideration of clusters is essential in the analysis

Algorithm 2 Cluster Detection Algorithm

Input : Sub-matrix $M_C = (R_t)_{t \in C}$, where R_t is the t-th row of the cross-efficiency matrix R.

Output : A list of clusters of DMUs.

Step 1: Initialization
Set List $\leftarrow C$, assign detector $D(\hat{\imath})$ for $\hat{\imath} = 1, \ldots, c$, initialize count $\leftarrow 0$.

Step 2: Construct Comparison Matrix
Construct matrix S, where each row is the difference between a candidate row and the reference row.

Step 3: Count Matches
Set count $\leftarrow$ number of rows in S that are identically zero.

Step 4: Evaluation

```
while List ≠ ∅ do
    if count = c then
        Form one cluster with all DMUs in List and exit.
    else
        if count = 1 then
            Remove the outlier DMU: List ← List \ {î}, c ← c − 1.
        else
            Identify subset of z identical rows, form cluster G_z, remove them: List ←
              List \ G_z, c ← c − z.
```

Step 5: Display
Output the list of all identified clusters.

4.2 Evaluation of Clusterless Cross-Efficiency

Once clusters are detected, the next step is to compute clusterless cross-efficiency scores. The purpose of this step is to prevent clusters of identical DMUs from disproportionately influencing the performance evaluation. If all members of a cluster were retained, their repeated presence could bias the overall average, leading to unfair rankings. For a cluster containing c DMUs, we retain only one representative DMU and exclude the remaining $c - 1$ DMUs from the evaluation. This procedure is repeated across all clusters, such that exactly one representative from each cluster is included. DMUs that do not belong to any cluster are automatically considered in the calculation. Finally, the cross-efficiency scores are averaged over the set of selected representative DMUs and the non-clustered DMUs. The resulting values represent the clusterless cross-efficiency scores, which are then used for ranking and performance analysis. This procedure ensures that each distinct performance pattern is represented exactly once, thereby improving fairness and reducing bias in the final ranking. The detailed procedure is presented as a flowchart in Figure 2, while Algorithm 2 outlines the computational steps for the cluster detection process.

Algorithm 2 generates unbiased cross-efficiency scores by selecting a single repre-

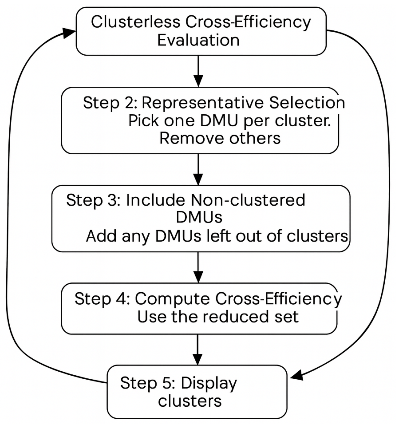

Figure 2: Flowchart for clusterless evaluation

sentative from each cluster while including all non-clustered DMUs. This approach ensures fair and balanced evaluation, with the final scores accurately reflecting the relative performance of all DMUs and reducing the impact of cluster size dominance.

4.3 Statistical validation of clusters

To ensure the reliability of the obtained clusters, the Silhouette score proposed by [24] is employed as the primary statistical validation metric. The Silhouette Score evaluates the degree of separation and cohesion among clusters by comparing the average distance of each DMU to other members of its own cluster with the distance to the nearest neighbouring cluster.

The Silhouette Score for each DMU_k is computed using two distance measures: the *intra-cluster dissimilarity* and the *nearest-cluster dissimilarity*. First, the average distance between DMU_k and all other DMUs within the same cluster is calculated and denoted as $a(k)$. Next, for every other cluster, the average distance between DMU_k and all DMUs in that cluster is computed; the smallest of these values is taken as $b(k)$, representing the dissimilarity to the nearest neighbouring cluster. The following steps describe the computation of the silhouette coefficient $s(k)$ for each DMU_k.

Step 1. Compute the Euclidean distance between DMU_k and DMU_j as

$$d(k,j) = \sqrt{\sum_{l=1}^{n} (\theta_{kl} - \theta_{jl})^2}$$

Algorithm 3 Clusterless cross-efficiency evaluation

Input: Clusters $C_1, C_2, \ldots, C_{\hat{i}}$ identified by Algorithm 1
Output: Clusterless cross-efficiency scores for all DMUs

Step 1: Representative Selection
For each cluster $C_{\hat{i}}$ containing c DMUs:
Select one representative DMU
Discard the remaining $c-1$ DMUs

Step 2: Inclusion of Non-clustered DMUs
Include all DMUs not belonging to any cluster

Step 3: Aggregation
Compute the average cross-efficiency scores using the reduced set of representatives DMUs and non-clustered DMUs

Step 2. Calculate within-cluster dissimilarity of DMU_k as

$$a(k) = \frac{1}{|C(k)|-1} \sum_{\substack{j \in C(k) \\ j \neq k}} d(k,j),$$

where $C(k)$ denotes the cluster containing DMU_k.

Step 3. The between-cluster dissimilarity is obtained by computing the average distance from DMU_k to each external cluster and selecting the minimum as

$$b(k) = \min_{C' \neq C(k)} \left(\frac{1}{|C'|} \sum_{j \in C'} d(k,j) \right).$$

Step 4. The silhouette value of DMU_k is then given by

$$s(k) = \frac{b(k) - a(k)}{\max\{a(k),\, b(k)\}}.$$

A value of $s(k) \approx 1$ indicates that DMU_k is well aligned with its assigned cluster and clearly separated from all other clusters, thereby supporting the validity of the clustering structure.

In the subsequent section, we demonstrate the proposed methodology using stock market data. We first detect clusters among the stocks to identify identical performance patterns. Then, using the clusterless cross-efficiency approach, we evaluate the performance of the representative stocks. This allows us to examine the impact of clustering on the final ranking.

5 An application to the Iranian stock Market

In portfolio selection, diversifying investments across a wide range of stocks is essential to reduce risk and avoid significant losses. This requires an accurate ranking of stocks to ensure balanced and effective portfolio construction. However, the presence of clustered stocks in cross-efficiency evaluation weakens discrimination, creating ambiguity in rankings and increasing the risk of misjudgment. This section, therefore, highlights the influence of clustered stocks on portfolio selection and underscores the need for methods that ensure fairer and more reliable rankings to support diversification.

We utilize a dataset of 20 Iranian Stock companies from Banihashemi and Sanei [19] with one input as variance and one output as a return. The mean return is taken as an output, as the goal is to maximize gains, while variance is treated as an input, reflecting risk that should be minimized. The input-output data of stocks taken from [19] are presented in Table 1.

Table 1: Input and output values of stocks from [19].

Stocks	1	2	3	4	5	6	7	8	9	10
Input	6.534	10.474	3.720	4.256	32.259	70.764	57.497	19.609	21.496	67.378
Output	7.285	7.388	-2.193	10.853	12.517	9.052	52.511	-3.676	3.537	7.570
Stocks	11	12	13	14	15	16	17	18	19	20
Input	14.171	29.002	42.133	12.420	1.611	11.429	25.358	4.856	28.464	1.560
Output	6.896	1.888	18.737	1.302	1.231	14.741	3.896	2.967	32.677	2.022

5.1 Results and Discussion

First, we solve Model II and then apply equation 1 to construct the RDM cross-efficiency matrix. The resulting cross-efficiency matrix is presented in Table 2. This matrix provides the basis for assessing the relative performance and benchmarking of the decision-making units. All computations are performed in MATLAB 2022a, selected for its reliability, efficient handling of large-scale numerical operations, and availability of built-in optimization toolboxes that facilitate precise and reproducible results

To compute the RDM cross-efficiency (traditional) for each DMU, we take the simple average of each column in the cross-efficiency matrix. Because the matrix contains DMUs that form clusters of similar or highly correlated stocks, the resulting efficiency values are referred to as clustered cross-efficiency scores. These clustered scores reflect both the individual performance of each DMU and the influence of its cluster, capturing patterns of similarity among the DMUs.

As previously highlighted, clustered cross-efficiency can bias the evaluation process and may lead to misjudgments if directly applied in portfolio selection. Since clustered DMUs tend to dominate the efficiency scores, relying on them could

Table 2: The RDM cross-efficiency matrix

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.887	0.829	0.778	1.000	0.652	0.417	0.874	0.608	0.659	0.423	0.774	0.585	0.627	0.723	0.859	0.945	0.629	0.840	1.000	0.873
2	0.887	0.829	0.778	1.000	0.652	0.417	0.874	0.608	0.659	0.423	0.774	0.585	0.627	0.723	0.859	0.945	0.629	0.840	1.000	0.873
3	0.821	0.679	0.817	1.000	0.359	0.187	0.276	0.438	0.442	0.194	0.581	0.359	0.303	0.582	0.981	0.720	0.399	0.837	0.468	1.000
4	0.887	0.829	0.778	1.000	0.652	0.417	0.874	0.608	0.659	0.423	0.774	0.585	0.627	0.723	0.859	0.945	0.629	0.840	1.000	0.873
5	0.887	0.829	0.778	1.000	0.652	0.417	0.874	0.608	0.659	0.423	0.774	0.585	0.627	0.723	0.859	0.945	0.629	0.840	1.000	0.873
6	0.786	0.746	0.680	0.878	0.627	0.421	1.000	0.558	0.611	0.425	0.705	0.551	0.621	0.652	0.745	0.858	0.589	0.738	1.000	0.757
7	0.786	0.746	0.680	0.878	0.627	0.421	1.000	0.558	0.611	0.425	0.705	0.551	0.621	0.652	0.745	0.858	0.589	0.738	1.000	0.757
8	0.887	0.829	0.778	1.000	0.652	0.417	0.874	0.608	0.659	0.423	0.774	0.585	0.627	0.723	0.859	0.945	0.629	0.840	1.000	0.873
9	0.887	0.829	0.778	1.000	0.652	0.417	0.874	0.608	0.659	0.423	0.774	0.585	0.627	0.723	0.859	0.945	0.629	0.840	1.000	0.873
10	0.786	0.746	0.680	0.878	0.627	0.421	1.000	0.558	0.611	0.425	0.705	0.551	0.621	0.652	0.745	0.858	0.589	0.738	1.000	0.757
11	0.887	0.829	0.778	1.000	0.652	0.417	0.874	0.608	0.659	0.423	0.774	0.585	0.627	0.723	0.859	0.945	0.629	0.840	1.000	0.873
12	0.887	0.829	0.778	1.000	0.652	0.417	0.874	0.608	0.659	0.423	0.774	0.585	0.627	0.723	0.859	0.945	0.629	0.840	1.000	0.873
13	0.887	0.829	0.778	1.000	0.652	0.417	0.874	0.608	0.659	0.423	0.774	0.585	0.627	0.723	0.859	0.945	0.629	0.840	1.000	0.873
14	0.887	0.829	0.778	1.000	0.652	0.417	0.874	0.608	0.659	0.423	0.774	0.585	0.627	0.723	0.859	0.945	0.629	0.840	1.000	0.873
15	0.821	0.679	0.817	1.000	0.359	0.187	0.276	0.438	0.442	0.194	0.581	0.359	0.303	0.582	0.981	0.720	0.399	0.837	0.468	1.000
16	0.887	0.829	0.778	1.000	0.652	0.417	0.874	0.608	0.659	0.423	0.774	0.585	0.627	0.723	0.859	0.945	0.629	0.840	1.000	0.873
17	0.887	0.829	0.778	1.000	0.652	0.417	0.874	0.608	0.659	0.423	0.774	0.585	0.627	0.723	0.859	0.945	0.629	0.840	1.000	0.873
18	0.887	0.829	0.778	1.000	0.652	0.417	0.874	0.608	0.659	0.423	0.774	0.585	0.627	0.723	0.859	0.945	0.629	0.840	1.000	0.873
19	0.887	0.829	0.778	1.000	0.652	0.417	0.874	0.608	0.659	0.423	0.774	0.585	0.627	0.723	0.859	0.945	0.629	0.840	1.000	0.873
20	0.821	0.679	0.817	1.000	0.359	0.187	0.276	0.438	0.442	0.194	0.581	0.359	0.303	0.582	0.981	0.720	0.399	0.837	0.468	1.000

distort the true performance of individual assets. Therefore, it becomes crucial to identify such clusters and carefully assess their impact before making portfolio decisions.

To achieve this, we employ Algorithm 1 (presented in Section 4) on the RDM cross-efficiency matrix to systematically detect all existing clusters. When Algorithm 1 is applied to the RDM cross-efficiency matrix, the following three distinct clusters are identified:

(i) $C_1 = \{1, 2, 4, 5, 8, 9, 11, 12, 13, 14, 16, 17, 18, 19\}$

(ii) $C_2 = \{6, 7, 10\}$,

(iii) $C_3 = \{3, 15, 20\}$.

The categorization of stocks into these clusters is visually represented in Figure 3, which clearly illustrates the groupings based on their similarities in the RDM matrix.

To compute the clusterless cross-efficiency of stocks, we propose the removal of $c - 1$ clustered DMUs, as described in Algorithm 2, before calculating the column-wise averages of the cross-efficiency matrix, where c denotes the number of stocks in a given cluster. From each identified cluster, only one stock is retained as the representative stock, and its cross-efficiency score is used in place of the entire cluster. Thus, the selected representative captures the performance of all stocks within the same cluster. In contrast, stocks that do not belong to any cluster represent only themselves. For example, in the RDM cross-efficiency matrix, cluster $C_1 = \{1, 2, 4, 5, 8, 9, 11, 12, 13, 14, 16, 17, 18, 19\}$ consists of 14 stocks. Instead of including all 14 rows, we retain only one representative (say stock 1), and eliminate the remaining $(14 - 1 = 13)$ rows corresponding to the other stocks in C_1. Similarly, for cluster $C_2 = \{6, 7, 10\}$, we retain stock 6 and remove 2 rows, while for cluster $C_3 = \{3, 15, 20\}$, stock 3 is chosen as the representative.

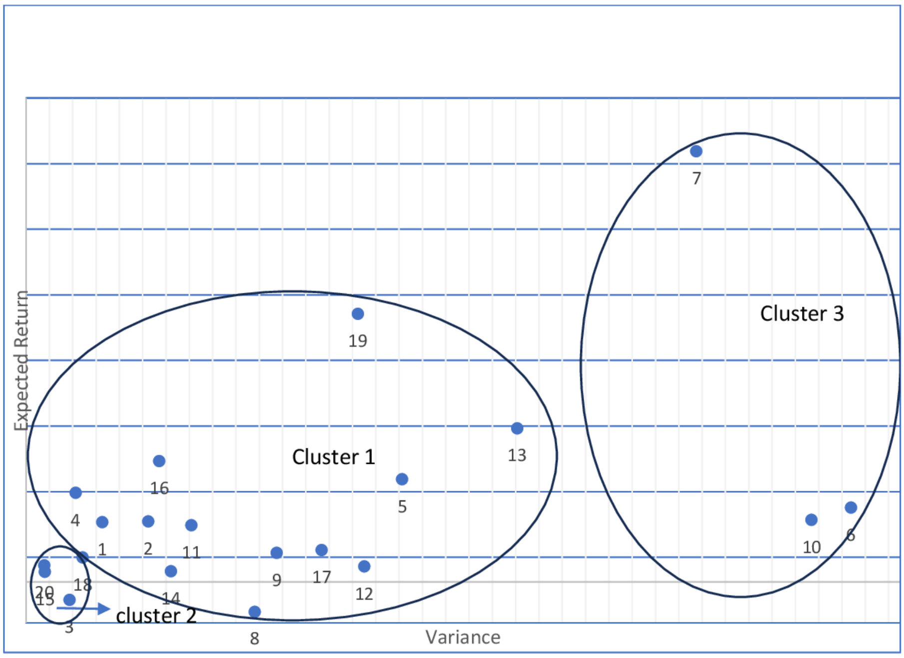

Figure 3: Stock categorization based on different clusters

After this reduction, the clusterless cross-efficiency is computed by averaging each column of the modified matrix. The results are illustrated in Table 3. For ease of comparison, both the clustered (traditional) cross-efficiency values and the clusterless cross-efficiency values are reported side by side in Table 4.

Table 3: Clusterless cross-efficiency scores based on representative stocks

cluster	stocks																			
Representative Stocks ↓	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.887	0.829	0.778	1.000	0.652	0.417	0.874	0.608	0.659	0.423	0.774	0.585	0.627	0.723	0.859	0.945	0.629	0.841	1.000	0.873
3	0.821	0.679	0.817	1.000	0.359	0.187	0.276	0.438	0.442	0.194	0.581	0.359	0.303	0.582	0.981	0.720	0.399	0.837	0.468	1.000
6	0.786	0.746	0.680	0.878	0.627	0.421	1.000	0.558	0.611	0.425	0.705	0.551	0.621	0.652	0.745	0.858	0.589	0.738	1.000	0.757
Clusterless cross efficiency	0.831	0.752	0.759	0.960	0.546	0.342	0.717	0.535	0.570	0.347	0.686	0.498	0.517	0.652	0.862	0.841	0.539	0.805	0.823	0.877

As shown in Table 4, the values of clusterless cross-efficiency are lower than those of the traditional clustered cross-efficiency. This reduction arises because, in the clusterless approach, the supportive influence of multiple clustered stocks is removed, leaving only one representative from each cluster.

Table 4: Clustered and clusterless cross-efficiency values of the stocks

Stocks	1	2	3	4	5	6	7	8	9	10
Clustered cross-efficiency	0.862	0.794	0.769	0.982	0.604	0.383	0.803	0.575	0.619	0.389
Clusterless cross-efficiency	0.831	0.752	0.759	0.960	0.546	0.342	0.717	0.535	0.570	0.347
Stocks	11	12	13	14	15	16	17	18	19	20
Clustered cross-efficiency	0.734	0.546	0.577	0.691	0.860	0.892	0.589	0.824	0.920	0.875
Clusterless cross-efficiency	0.686	0.498	0.517	0.652	0.862	0.841	0.539	0.805	0.823	0.877

These results demonstrate that the cross-efficiency–based distance structure leads to highly distinct and well-separated clusters, supporting the suitability of cross-efficiency scores as an effective basis for unsupervised grouping of DMUs.

5.2 Impact on portfolio selection

In addition, we conduct a comparative analysis of the top five stocks identified for portfolio inclusion under both evaluation approaches—clustered cross-efficiency and clusterless cross-efficiency. Stocks with higher clustered cross-efficiency scores are assigned higher rankings, with the top five stocks corresponding to those with the highest clustered cross-efficiency values. A similar procedure is applied for selecting the top five stocks based on clusterless cross-efficiency. This comparison highlights the differences in stock selection when the influence of clustered DMUs is considered versus when it is mitigated. The detailed results of this comparison are reported in Table 5, providing insights into how cluster adjustment can alter portfolio composition and lead to a more balanced selection of assets. From Table

Table 5: Top 5 stocks identified using clustered and clusterless cross-efficiency

Rank	**Stocks**	**Clustered cross-efficiency**	**Stocks**	**Cluster**	**Clusterless cross-efficiency**
1	4	0.982	4	C1	0.960
2	19	0.920	20	C2	0.877
3	16	0.892	15	C2	0.862
4	20	0.875	16	C1	0.841
5	1	0.862	1	C1	0.831

5, it is evident that the 4th stock consistently retains its leading position in both evaluation scenarios. However, its efficiency decreases slightly from 0.982 under clustered cross-efficiency to 0.960 in the clusterless evaluation, reflecting the loss of supportive influence from other clustered stocks. In contrast, the 19th stock, which initially had an efficiency of 0.920, experiences a notable decline to 0.823 once cluster support is removed, resulting in its exclusion from the top five rankings. The 20th stock shows only a marginal change in efficiency between the two approaches, since it belongs to a relatively small cluster of just three stocks, thereby limiting the impact of cluster removal. Interestingly, the 15th stock appears among the top five in the clusterless cross-efficiency ranking, whereas it does not hold a top-five position in the clustered evaluation, highlighting the shifts in stock selection that arise when cluster effects are eliminated.

Furthermore, investors may also choose stocks from different clusters when constructing portfolios. It is important to note that a stock's cross-efficiency score generally decreases once the supportive influence of its clustered DMUs is removed. Conversely, stocks that are not part of any cluster may gain relative importance in the clusterless cross-efficiency evaluation.

Remark 5.1. In the illustrative example, each stock attains a silhouette score of 1, the maximum possible value of the silhouette coefficient, indicating perfect clustering quality. This result shows that every stock is optimally aligned with its assigned cluster and is maximally separated from the other clusters. For instance, the members of cluster 2 (3,15,20) have identical cross-efficiency vectors, which yields a within-cluster distance of $a(k) = 0$. In contrast, the nearest-cluster distances remain strictly positive because the cross-efficiency profiles of the clusters differ substantially. For stock 3, the distance to cluster 1 is 1.1332, while the distance to cluster 3 is 0.34618. Therefore, the silhouette coefficient is computed as

$$s(3) = \frac{b(3) - a(3)}{\max\{a(3),\, b(3)\}} = \frac{0.34618 - 0}{\max(0,\, 0.34618)} = 1.$$

By applying the same procedure, all 20 stocks in the dataset obtain a silhouette coefficient of 1, confirming that the detected clusters are perfectly validated.

In the illustrative example, each stock received a silhouette score of 1, the maximum possible value of the silhouette coefficient. This outcome indicates that all stocks are perfectly matched to their assigned clusters and are maximally separated from other clusters. Specifically, the within-cluster distance for each stock is zero because, for example, the members of Cluster 2 (DMU 3, 15, and 20) have identical cross-efficiency scores. As a result, the average distance between a DMU and all other DMUs in its own cluster, denoted as $a(k)$, becomes zero. Meanwhile, the nearest-cluster distance, denoted by $b(k)$, remains positive because the clusters' efficiency profiles differ substantially.

Therefore, the clustered cross-efficiency framework provides a statistically reliable foundation for ranking stocks and constructing diversified portfolios.

5.3 Comparative Analysis and Limitation

In this section, we present a comparative analysis of the proposed clusterless cross-efficiency approach against existing studies on portfolio selection problems.

Cross-efficiency has been applied to portfolio selection in various ways throughout the literature. However, our proposed clusterless cross-efficiency evaluation offers a distinct and straightforward alternative compared to existing approaches. For instance, Amin and Oukil [21] examine portfolio selection using gangless cross-evaluation, yet their method does not account for cases involving negative data. Similarly, Wu [25] adopts a different framework based on row and column efficiency evaluations, categorizing stocks according to these measures. Given the fundamental differences between Wu's approach and ours, a direct comparison is not appropriate. Lim et al. [13] introduce a unique application of cross-efficiency in portfolio selection by integrating the MV portfolio optimization model with cross-efficiency. Their model aims to identify stocks (or DMUs) that minimize risk while achieving a predetermined expected return. Amin and Hajjami [15] employ DEA MV cross-efficiency evaluation and discuss the impact of alternative optimal solutions in portfolio selection. In contrast to [13, 15], where the presence of clustered DMUs can influence the cross-efficiency used, our proposed method demonstrates that portfolio selection can be affected when clusterless cross-efficiency is employed. It would be interesting to investigate how the clusterless cross-efficiency and MV model can be combined for more effective portfolio selection, which is a matter of future research. Several authors apply cross-efficiency in fuzzy multi-objective frameworks, including Mashayekhi and Omrani [26], Chen et al. [27], and Chen et al. [28], among others. In the future, the proposed clusterless cross-efficiency values could be utilized in fuzzy multi-objective portfolio selection.

While our proposed method offers both simplicity and robustness, it is not without limitations. First, the approach identifies which stocks should be included in a portfolio, but does not determine the optimal allocation or percentage of investment for each stock. This limitation may be addressed by integrating the derived cross-efficiency values with the Markowitz Mean–Variance (MV) model, which we leave as a direction for future research. Second, when clusters contain an equal number of DMUs, the elimination of clustered DMUs inevitably reduces the cross-efficiency scores. Moreover, based on our extensive review of the literature, we find no existing methodology that allows for a direct comparison with the outcomes of the approach proposed in this study.

6 Conclusion

This study addressed a critical limitation in DEA cross-efficiency evaluation, namely the formation of clustered DMUs, which undermines the discriminatory power and ranking reliability of efficiency assessment. To overcome this drawback, we proposed the clusterless cross-efficiency evaluation method. By removing the effect of clustering, the technique ensures fairer comparisons among DMUs, reduces the possibility of misjudgment, and generates rankings that are both reliable and diverse. This advancement enhances the decision-making process by providing a stronger basis for identifying top-performing units. The superiority of the proposed approach is particularly evident in portfolio selection, where accurate ranking of stocks is crucial.

Although the proposed method successfully addresses the clustering problem, certain limitations remain. In particular, the current study focuses primarily on clustered DMUs defined within efficiency-based evaluations, and further exploration is required to incorporate both efficient and inefficient units into this framework. Extending the clusterless evaluation method to broader decision-making contexts also holds promise for producing unique and more robust rankings across diverse applications. Furthermore, a preliminary integration of the proposed method with mean-variance optimization could significantly enrich its contribution, offering a hybrid framework that balances efficiency with risk-return trade-offs. This direction represents a valuable avenue for future research, ensuring the continued relevance and practical utility of the method.

Acknowledgment

The author expresses sincere appreciation to the academic community and peers whose insights in the areas of portfolio optimization, efficiency analysis, and multi-criteria decision making have inspired this research. We also sincerely thank the reviewer for their valuable comments and constructive feedback.

Bibliography

[1] T.R. Sexton, R.H. Silkman, and A.J. Hogan, "Data envelopment analysis: Critique and extensions", New directions for program evaluation, 1986, 32, 73-105.

[2] J. Doyle, R. Green, "Efficiency and cross-efficiency in DEA: Derivations, meanings and uses. Journal of the Operational Research Society, 1994, 45, 567-578.

[3] J. Wu, J. Sun, and L. Liang, "Methods and applications of DEA cross-efficiency: Review and future perspectives", Frontiers of Engineering Management,2021 8(2), 199-211.

[4] M. Wu, X. Hou, and J. Fan, "Improvement of cross-efficiency based on the TODIM method", Soft Computing, 2022, 26(17), 8427-8439.

[5] M. Davtalab-Olyaie, and M. Asgharian, "On Pareto-optimality in the cross-efficiency evaluation" European Journal of Operational Research, 2021, 288(1), 247-257.

[6] A. Charnes, W.W. Cooper, and E. Rhodes, "Measuring the efficiency of decision-making units", European journal of operational research, 1978, 2(6), 429-444.

[7] M.S. Portela, E. Thanassoulis, and G. Simpson, "Negative data in DEA: A directional distance approach applied to bank branches", Journal of the Operational Research Society, 2004,55, 1111-1121.

[8] R. Lin, "Cross-efficiency evaluation capable of dealing with negative data: A directional distance function-based approach", Journal of the Operational Research Society,2020, 71(3), 505-516.

[9] M. Soltanifar, and H. Sharafi, H "A modified DEA cross-efficiency method with negative data and its application in supplier selection", Journal of Combinatorial Optimization, 2022,43(1), 265-296.

[10] L. Chen, Y. M. Wang, and Y. Huang, "Cross-efficiency aggregation method based on the prospect consensus process", Annals of Operations Research, 2020,288(1), 115-135.

[11] X. Chen, X.Liu, Q. Zhu, and Z. Wang. "DEA cross-efficiency models with prospect theory and distance entropy: An empirical study on high-tech industries." Expert Systems with Applications, 2024, 244,122941.

[12] C. Kao, and S.T., Liu, 2020. “A slacks-based measure model for calculating cross efficiency in data envelopment analysis”, Omega,2020 95, p.102192.

[13] S. Lim, K.W. Oh, and J. Zhu, “Use of DEA cross-efficiency evaluation in portfolio selection: An application to Korean stock market”, European Journal of operational research, 2014,236(1), 361-368.

[14] X. Gong, C. Yu, L. Min, and Z. Ge, “Regret theory-based fuzzy multi-objective portfolio selection model involving DEA cross-efficiency and higher moments”, Applied Soft Computing, 2021, 100, 106958.

[15] G.R. Amin, and M. Hajjami “Improving DEA cross-efficiency optimization in portfolio selection”, Expert Systems with Applications, 2021,168, 114280.

[16] R. Kumari, A. Aggarwal, A. Gupta. “Cross-efficiency aggregation by ordered visibility graph averaging: method, and application in portfolio selection” IMA Journal of Management Mathematics, 2025 Jan;36(1):83-110.

[17] S. Srivastava, A. Aggarwal, and P. Bansal, “Efficiency evaluation of assets and optimal portfolio generation by cross-efficiency and cumulative prospect theory”, Computational Economics, 2024,63(1), 129-158.

[18] X. Deng, and W. Fang, “A novel mean-variance-maverick DEA prospect cross-efficiency approach for fuzzy portfolio selection”, Journal of Intelligent & Fuzzy Systems, 2019,37(6), 8113-8130.

[19] S. Banihashemi, and M. Sanei, “Cross efficiency evaluation with negative data in selecting the best of a portfolio using OWA operator weights”, Int J Data Envel Anal. 2015, 3, 633-651.

[20] Kumari, R., Gupta, A., & Aggarwal, A. “Stock assessment using cumulative prospect theory in DEA cross-efficiency model: A case study of the Indian stock market” Croatian Operational Research Review, 2024, 15(1), 1-11.

[21] G.R. Amin, and A. Oukil, “Gangless cross-evaluation in DEA: an application to stock selection”, RAIRO-Operations Research, 2019 53(2), 645-655.

[22] J.A. Sharp, W. Meng & W. Liu, ”A modified slacks-based measure model for data envelopment analysis with ‘natural’negative outputs and inputs”, Journal of the Operational Research Society, 2007,58(12), 1672-1677.

[23] A. Emrouznejad, A.L. Anouze, & E. Thanassoulis, ”A semi-oriented radial measure for measuring the efficiency of decision making units with negative data, using DEA”, European journal of Operational Research, 2010, 200(1), 297-304.

[24] P.J. Rousseeuw, P. J., ”Silhouettes: a graphical aid to the interpretation and validation of cluster analysis”. Journal of computational and applied mathematics, 1987, 20, 53-65.

[25] D. Wu, “Robust decision support system for asset assessment and management”, IEEE Systems Journal, 11(3), 2016,1486-1491.

[26] Z. Mashayekhi, and H.Omrani “An integrated multi-objective Markowitz-DEA cross-efficiency model with fuzzy returns for portfolio selection problem”, Appl Soft Comput., 2016, 38,1–9.

[27] W. Chen, S.S.,Li, J. Zhang, and M.K. Mehlawat, “A comprehensive model for fuzzy multi-objective portfolio selection based on the DEA cross-efficiency model”, Soft Computing, 2020,24, 2515-2526.

[28] W. Chen, S.S. Li, M.K., Mehlawat, L. Jia, and A. Kumar, “Portfolio selection using data envelopment analysis cross-efficiency evaluation with undesirable fuzzy inputs and outputs”, International Journal of Fuzzy Systems, 2021,23, 1478-1509.

How to Cite: Kumari, Reenu[1], *A New Clusterless DEA Cross-Efficiency Evaluation in the Presence of Negative Data and its Application in Portfolio Selection*, Journal of Mathematics and Modeling in Finance (JMMF), Vol. 6, No. 1, Pages:207–224, (2026).