Expert Systems With Applications 223 (2023) 119902

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Analytic hierarchy process and data envelopment analysis: A match made in heaven

Madjid Tavana [a,b,*], Mehdi Soltanifar [c], Francisco J. Santos-Arteaga [d], Hamid Sharafi [e]

[a] *Business Systems and Analytics Department, Distinguished Chair of Business Analytics, La Salle University, Philadelphia, USA*
[b] *Business Information Systems Department, Faculty of Business Administration and Economics, University of Paderborn, Paderborn, Germany*
[c] *Department of Mathematics, Semnan Branch, Islamic Azad University, Semnan, Iran*
[d] *Departamento de Análisis Económico y Economía Cuantitativa, Universidad Complutense de Madrid, Madrid, Spain*
[e] *Department of Mathematics, Science and Research Branch, Islamic Azad University, Tehran, Iran*

A R T I C L E I N F O

Keywords:
Data envelopment analysis
Analytic hierarchy process
Pairwise comparisons
Performance evaluation
Rank absolute deviation
MACBETH

A B S T R A C T

Analytic hierarchy process (AHP) and data envelopment analysis (DEA) are two popular decision science methods with many business, science, and engineering applications. AHP is a Multi-Attribute Decision-Making (MADM) method for prioritizing alternatives, and DEA is a non-parametric method for estimating production frontiers. Each method has known strengths and weaknesses, and the strengths in one method can overcome the weaknesses in the other. We study several strategies to diminish the weaknesses of DEA with the strengths of pairwise comparisons in AHP. We tackle the low discrimination power inherent to conventional DEA methods. AHP, with its pairwise comparison capability, has been consistently used to increase the discrimination power and accuracy in DEA. We propose and evaluate several new hybrid MADM-DEA models of different computational complexity and consistency, including combinations of the best-worst method (BWM) and its variants with DEA as well as a novel method composed of Measuring Attractiveness by a Categorical Based Evaluation Technique (MACBETH) and DEA. We further develop a new technique for evaluating the similarity among multiple ranking results in MADM. The new simple but powerful technique is called Rank Absolute Deviation (RAD) and is inspired by the mean absolute deviation method. Several numerical examples and a real-world problem are used to demonstrate the applicability and efficacy of the new BWM-DEA, MACBETH-DEA, and RAD methods proposed in this study. We illustrate how less computationally demanding MADM-DEA techniques provide rankings that are highly correlated with the benchmark DEA-AHP and different consensus ranking models.

1. Introduction

Analytic Hierarchy Process (AHP) is a structured method for making complex decisions grounded in mathematics and psychology. AHP was first proposed by Saaty (1980) and has been extensively studied and used for alternative ranking, prioritization, and selection. On the other hand, Data Envelopment Analysis (DEA) is a non-parametric technique in operations research, economics, and management science for estimating the Production Possibility Set (PPS) frontiers. DEA is used to measure the relative efficiency of Decision-Making Units (DMUs). By computing the relative efficiency of DMUs, DEA provides a ranking of the different alternatives whose behavior is measured against a reference benchmark defined by the most efficient one. These efficiency values are derived from an optimization problem that evaluates the capacity of DMUs to produce a given set of outputs using a series of inputs.

Bouyssou (1999) reviewed the equivalence between the concept of efficiency in DEA and convex efficiency in Multiple-Criteria Decision-Making (MCDM) environments. The author emphasized the set dependence of the rankings generated by DEA as well as the importance of incorporating preference information – such as the relative importance of criteria – within the corresponding optimization problems, as is generally done by standard MCDM methods. In this regard, Sarkis (2000) showed how incorporating the value judgments of decision-

* Corresponding author at: Business Systems and Analytics Department, Distinguished Chair of Business Systems and Analytics, La Salle University, Philadelphia, PA 19141, United States.
E-mail addresses: tavana@lasalle.edu (M. Tavana), soltanifar@khayam.ut.ac.ir (M. Soltanifar), fransant@ucm.es (F.J. Santos-Arteaga).
URL: http://tavana.us/ (M. Tavana).

https://doi.org/10.1016/j.eswa.2023.119902
Received 14 July 2022; Received in revised form 24 December 2022; Accepted 15 March 2023
Available online 21 March 2023
0957-4174/

makers (DMs) into a DEA framework provides comparable results to those derived from traditional MCDM approaches. The author also highlighted the fact that DEA approaches require less information from DMs while providing useful rankings of the alternatives. Opricovic and Tzeng (2003) analyzed the main differences between DEA and MCDM techniques, particularly the fact that the former does not consider the preferences of DMs. They highlighted the fact that despite these differences, DEA complements the screening of alternatives within MCDM frameworks.

Although Farrell (1957) introduced the idea of measuring relative efficiency, the first DEA models were formulated by Charnes et al. (1978), under the Constant Returns to Scale (CRS) assumption called the CCR model, and Banker et al. (1984), under the Variable Returns to Scale (VRS) assumption called the BCC model. Many extensions of DEA have been developed over four decades of research (Hatami-Marbini et al., 2022). The conventional DEA suggests each DMU selects its most desirable value of the variables solving the corresponding optimization problem. Applying the subsequent weights allows the DMUs to achieve their maximum performance. It is common for many DMUs to achieve maximum efficiency (equal to 1). Different methodologies have been defined in case several DMUs are regarded as efficient. For instance, ranking methods can improve DEA's discrimination power in differentiating between the DMUs with maximum efficiency (Labijak-Kowalska and Kadziński, 2021). These authors illustrated numerically the use of fifteen selected DEA approaches, including DEA-AHP combinations, to rank DMUs and the significant impact on the ranking that follows from the choice of a given method.

Even though AHP and DEA have unique strengths and weaknesses and have evolved independently, research has shown that they can be integrated to solve real-world problems effectively and efficiently (Ho, 2008; Dutta et al., 2022). As a result, many researchers have attempted to establish a relationship between the two methods and offset the weaknesses of one with the strengths of the other.

This integration can be examined from two different strategic perspectives. The first strategy is to overcome the weaknesses of AHP with the strengths of DEA by using DEA to find the relative weights and local priorities in AHP. Ramanathan (2006) used DEA and calculated the relative weights of the elements within an AHP hierarchy and showed that the rank-reversal phenomenon does not occur in the integrated approach. Ramanathan's (2006) approach eliminated a major flaw in AHP with DEA. Sevkli et al. (2007) embedded the DEA method into AHP and used the hybrid DEAHP method effectively for supplier selection and showed that the DEAHP outperforms the AHP method. Citing Ramanathan (2006), these authors illustrated how DEA correctly calculates the true weights for a consistent judgment matrix. Furthermore, DEAHP does not suffer from rank reversal when irrelevant alternatives are added or removed. Starcevic et al. (2019) used the AHP results as multiple outputs in a DEA model for terrain vehicle selection. Liu and Hai (2005) proposed a new AHP method called Voting AHP (VAHP) and calculated the relative weights of the criteria with DEA instead of forming pairwise comparison matrices in AHP. Hadi-Venchehand and Niazi-Motlagh (2011) addressed some of the shortcomings of Liu and Hai's (2005) VAHP method and applied an extended VAHP method to a supplier selection problem. Soltanifar and Hosseinzadeh Lotfi (2011) used the extended VAHP method to rank efficient DMUs in DEA. This method and several other hybrid methods are also discussed in Tavana et al. (2021). Wang et al. (2008) proposed a method to eliminate some of the shortcomings of AHP and used said method to evaluate bridge risks.

The second strategy is to overcome the weaknesses of DEA with the strengths of AHP. Several methods have been developed for implementing this strategy. These methods include the qualitative data to quantitative data (Lin et al., 2011), ranking the efficient units (Joblonsky, 2007), weighing the degree of improvement (the amount of change) in initial inputs and outputs of DMUs in the target setting (Lozano and Villa, 2009), restricting the input and output weights (Takamura and Tone, 2003), restricting the virtual weights of inputs and outputs for each DMU (Shang and Sueyoshi, 1995), weighting the inputs and outputs in the DEA structure (Kim, 2000), reflecting information on the hierarchical structure of input and output data in the performance assessment of DMUs (Pakkar, 2016), achieving common weights in DEA (Pakkar, 2015), and estimating the missing data for DMUs (Farzipoor Saen et al., 2005). Azadeh et al. (2011) evaluated the effective personnel operation indicators by management, which are usually qualitative, and converted them into quantitative forms using AHP. Then they evaluated and optimized the ranking and efficiency of the organization by DEA.

We propose and evaluate several new hybrid MADM-DEA methods of different computational complexity and consistency. These methods have been designed to reduce the number of computations required relative to recent developments of the DEA-AHP model (Alirezayee et al., 2012). This is particularly the case in terms of the number of pairwise comparisons. The hybrid techniques proposed do not require DM to compute the relative weights of the criteria, as is the case in DEA-AHP, increasing the consistency of the analysis. In particular, the DEA-AHP model will be used as a benchmark to compare the techniques introduced, which consist of combinations of the Best-Worst Method (BWM), Best Method (BM), and Worst Method (WM) with DEA. We also introduce and evaluate a novel, more complex, and computationally demanding approach composed of Measuring Attractiveness by a Categorical Based Evaluation Technique (MACBETH) and DEA (MACBETH-DEA).

We compute the maximize agreement heuristic (MAH) and design a new technique for evaluating the similarity among multiple ranking results in MADM. The new simple but powerful technique is called Rank Absolute Deviation (RAD) and is inspired by the mean absolute deviation method (Yager and Alajlan, 2014). We demonstrate the applicability and efficacy of the new MADM methods and MACBETH-DEA with several examples and a real-world case study that applies the RAD developed in this study.

In particular, we illustrate how the rankings generated by the hybrid MADM-DEA models proposed do not differ significantly from the MAH and RAD consensus rankings. That is, hybrid techniques that are less demanding computationally display a high degree of correlation and similarity with the benchmark DEA-AHP and consensus ranking models. We will also analyze the ranking variability induced by the more demanding – in computational and consistency terms – MACBETH-DEA technique.

The remainder of this paper is organized as follows: Section 2 reviews the DEA ranking methods considered. Section 3 defines the new BWM-DEA, BM-DEA, WM-DEA, and MACBETH-DEA methods. In Section 4, the proposed new methods are reviewed and compared with other popular ranking methods through a numerical example. Section 5 introduces a new approach for selecting the results of a ranking model on a problem under investigation inspired by one of the new MADM methods. Section 6 presents a real-world problem to demonstrate the applicability and efficacy of the hybrid ranking methods proposed in this study. We conclude in Section 7 discussing the limitations of the models presented and future research directions.

2. Ranking methods in DEA

The DEA technique allows each DMU to select a set of weights for inputs and outputs, indicating that the DMU is in the most favorable position relative to the other units. This benevolent view allows to obtain several DMUs with a maximum relative efficiency score of 1. Conventional DEA models do not differentiate between efficient units (units with a relative efficiency score of 1). In response, researchers have proposed ranking methods to distinguish among the efficient DMUs. These ranking methods can be categorized into six groups. The first group uses super-efficiency models, including those proposed by Hashimoto (1997), Andersen and Petersen (1993), Mehrabian et al. (1999), Sueyoshi (1999), and Tone (2002). In the super-efficiency models, the efficient DMU is removed from the performance evaluation process, and the changes in the performance frontier are studied. The second group uses target setting, where one unit has a higher rank in the group when introduced as a target for other units. This idea was used by Torgersen et al. (1996). The third group uses multivariate statistical techniques to rank the DMUs (Sinuany-Stern et al., 1994; Friedman and Sinuany-Stern, 1997; Sinuany-Stern and Friedman, 1998). The fourth group ranks inefficient units by their degree of inefficiency (Bardhan et al., 1996). Simple and easy-to-interpret scalar efficiency measures are provided using the concept of efficiency dominance and in the form of mixed integer programming models. The efficiency of DMUs is therefore determined through the comparisons made with actually observed performances, allowing for the generation of rankings among the efficient DMUs. The fifth group uses a combination of DEA models and multi-criteria optimization for ranking (Golany, 1988; Li and Reeves, 1999; Hosseinzadeh Lotfi et al., 2010a, 2010b). Finally, the sixth group uses the cross-efficiency matrix for ranking DMUs (Sexton et al., 1986; Doyle and Green, 1995).

In this section, we present cross-efficiency models in greater detail along with their advantages and disadvantages. In the cross-efficiency models, the self-evaluation process is replaced with the peer evaluation process. Two main qualities of cross-evaluation are generally emphasized when suggesting its implementation as an extension of traditional DEA models. Both qualities follow from the fact that DMUs are rated by the weighting schemes of all DMUs composing the sample analyzed. As a result, cross-evaluation delivers unique orderings while eliminating unrealistic weighting scenarios without requiring experts to impose weight constraints exogenously (Anderson et al., 2002).

Recent developments focusing on the capacity of cross-efficiency models to obtain full rankings include the introduction of behavioral decision-making features, such as accounting for the satisfaction and consensus of DMs (Wu et al., 2021) and the inclusion of interval and fuzzy data (Wang et al., 2021). The combination of AHP and cross-efficiency DEA into hybrid models, exploiting their respective complementarities, has been consistently developed in the literature. For instance, An et al. (2018) combined DEA and AHP to fully rank DMUs considering all possible interval cross efficiencies among DMUs, which, at the same time, were used to define interval preference relations characterizing the ranking. Similarly, Akbarian (2020) introduced an approach to rank DMUs combining the interval cross-efficiencies derived from DEA and interval AHP.

Assume that a set of n homogeneous DMUs is available with m inputs and s outputs. Also, assume that for a given DMU_j $(j = 1, 2, \ldots, n)$, $x_{ij}(i = 1, 2\ldots m)$ is the i^{th} input value and $y_{rj}(r = 1, 2\ldots s)$ is the r^{th} output value. Cross-efficiency in DEA is usually evaluated through two steps. In the first step, one of the basic DEA models, such as the CCR model (Charnes et al., 1978), is constructed for each DMU_o, with the subscript referring to the unit being analyzed, to calculate the relative efficiency and the weights of the inputs, $v_{io},\ i = 1,2,\ldots,m$, and outputs, $u_{ro},\ r = 1,2,\ldots,s$, for achieving this relative efficiency. For instance, the calculation based on the CCR Model (1) is described below.

$$
\begin{aligned}
& E_{oo}^{*} = \max \sum_{r=1}^{s} u_{ro} y_{ro} \\
& \text{s.t.} \\
& \sum_{i=1}^{m} v_{io} x_{io} = 1 \\
& \sum_{r=1}^{s} u_{ro} y_{rj} - \sum_{i=1}^{m} v_{io} x_{ij} \leqslant 0;\ j = 1, 2, \ldots, n \\
& u_{ro} \geqslant 0;\ r = 1, 2, \ldots, s \\
& v_{io} \geqslant 0;\ i = 1, 2, \ldots, m
\end{aligned} \tag{1}
$$

After solving Model (1) and finding the optimal solution, the cross-efficiency of the other DMUs will be calculated based on the efficiency of DMU_o using Eq. (2).

$$E_{oj} = \frac{\sum_{r=1}^{s} u_{ro}^{*} y_{rj}}{\sum_{i=1}^{m} v_{io}^{*} x_{ij}};\ j = 1, 2, \ldots, n \tag{2}$$

where (*) represents the optimal values of Model (1). In the second step, the cross-efficiency value of each DMU is calculated using Eq. (3) to determine the arithmetic mean of the cross-efficiency values computed by Eq. (2).

$$\overline{E}_j = \frac{1}{n} \sum_{o=1}^{n} E_{oj};\ j = 1, 2, \ldots, n \tag{3}$$

Despite its advantages, the use of cross-efficiency matrices has two major drawbacks. First, Doyle and Green (1994) showed that the existence of other optimal solutions in the DEA models changes DMU rankings. In other words, choosing different optimal solutions will change the ranking of other DMUs. Sexton et al. (1986) and Doyle and Green (1995) suggested using secondary goal models to reduce the effects of this flaw. They used both benevolent and aggressive perspectives and introduced secondary goal models and weight selection. We will define both perspectives when analyzing the hybrid techniques presented throughout the manuscript. For completeness, they have both been described in the Appendix A section. Soltanifar and Shahghobadi (2013) proposed a new method to select the best secondary goal model according to the benevolent perspective by considering different secondary goal models and introducing new ones. Davtalab-Olyaie (2018) defined a new criterion to propose secondary goal models.

The second flaw lies in using the arithmetic mean (instead of other metrics, such as the geometric, Winsorized, or interquartile mean) to aggregate the results of the cross-efficiency matrix. A central tendency measure is needed to describe the performance of each DMU. In this regard, the appropriate description depends on the type of observations and the characteristics of the problem being discussed. In the traditional cross-efficiency method, the arithmetic mean is proposed without considering these specific characteristics while other measures of central tendency may constitute a more appropriate description of the efficiency of the DMU under evaluation. Zerafat Angiz et al. (2013) addressed this concern by proposing a new method for aggregating cross-efficiency matrix results based on the preferred voting process. In this paper, we use MADM and pairwise comparisons to address these flaws.

3. Ranking methods based on pairwise comparisons

In this section, we study four ranking methods based on pairwise comparisons:

3.1. AHP-DEa

AHP has been proposed to weaken the flaws of the cross-efficiency method by using the pairwise comparison between DMUs to calculate the relative weights. Sinuany-Stern et al. (2000) proposed a two-step process and two DEA models to evaluate the performance of DMUs relative to each other. They then used AHP to calculate the final weights of these units through pairwise comparison among the DMUs. Alirezayee and Rafiee Sani (2010) claimed that Sinuany-Stern et al.'s (2000) method has several shortcomings, including inconsistent DEA rankings in multi-input and multi-output problems. The author showed how an inefficient unit in the Sinuany-Stern et al.'s (2000) method may rank higher than an efficient unit and cited poor discrimination power in problems with many inputs and outputs.

To eliminate these weaknesses, Rafiee Sani (2010) introduced a new process where the envelopment form was used to calculate the pairwise comparisons of the units with a new format each time by removing a DMU from the PPS. Finally, the AHP method was applied to the obtained pairwise comparison matrices. A similar envelopment form and PPS were proposed by Rezaeitaziani and Barkhordariahmadi (2015) to achieve improved results in AHP-DEA. Alirezayee et al. (2012) showed that the method of Alirezayee and Rafiee Sani (2010) needs to solve $n(n-1)+n$ linear programming models for a problem with n DMUs despite eliminating many shortcomings of the Sinuany-Stern et al.'s (2000) method. Consequently, they proposed a new method that required solving only n linear programming problems. In their method, after solving Model (1) for all DMUs and considering Eq. (2), the matrix of pairwise comparisons $\left([a_{ij}]_{n\times n}\right)$ is formed through Eq. (4).

$$\begin{cases} a_{ij} = \dfrac{E_{ii}+E_{ji}}{E_{jj}+E_{ij}} \\ a_{ji} = \dfrac{1}{a_{ij}} \end{cases} \quad i,j = 1,2,\cdots n \tag{4}$$

Finally, the weight of each DMU is obtained by calculating relative weights via least-squares, logarithmic least squares, eigenvectors, or approximation methods. Ehsanifar (2014) used this method to rank different cars.

In all of the above versions of the AHP-DEA method, the structure of pairwise comparison matrices in AHP has been used to rank DMUs in DEA. Recent research developments introduced in the literature on AHP-DEA include the analysis of interval cross-efficiencies (An et al., 2018) and the incorporation of fuzzy variables (Yilmaz et al., 2022). Other MADM techniques that implement a pairwise comparison process will be considered to develop new ranking methods in DEA.

3.2. Best-worst method-DEA (BWM-DEA)

One of the methods for weighting criteria is the best-worst method (BWM) proposed by Rezaei (2015) and implemented by many researchers, including Ahmadi et al. (2017), Delice & Can (2020), Liang et al. (2020), Rezaei (2015), and Rezaei et al. (2015, 2016). In this method, after determining the best and worst criteria, the pairwise comparison of other criteria with these two becomes the basis for presenting a mathematical programming model. Rezaei (2015) proposed the linear and non-linear versions of the method and discussed the inconsistency ratio. The weight of the criteria is extracted from solving a mathematical programming model, while the inconsistency ratio checks the validity of the comparisons. The algorithm of this method is defined as follows:

Step 1. Identify the influential criteria for the purpose of the problem by interacting with the DM: $C_1, C_2, ..., C_m$.

Step 2. Determine the best (C_B) and the worst (C_W) criteria among the final ones obtained based on the DM's preferences.

Step 3. Determine the DM's preferences for the best criterion over others based on Saaty's 9-point scale $\left(a_{Bj},\ j = 1,2,...,m\right)$.

Step 4. Determine the DM's preferences for the other criteria over the worst one based on Saaty's 9-point scale $\left(a_{jW}, j = 1,2,...,m\right)$.

Step 5. Obtain the weights of the criteria by solving Model (5).

$$\begin{aligned} & \min \xi^L \\ & \text{s.t.} \\ & \left|w_B - a_{Bj}w_j\right| \leqslant \xi^L, \quad \forall j \\ & \left|w_j - a_{jW}w_W\right| \leqslant \xi^L, \quad \forall j \\ & \sum_j w_j = 1 \\ & w_j \geqslant 0, \quad \forall j \end{aligned} \tag{5}$$

Theorem 1. *Model (5) is always feasible.*

Proof. This is achieved by taking $w_B, \xi^L = 1$ and all other variables equal to zero. Therefore, we conclude that Model (5) is feasible, and this ends the proof. □.

Model (5) is known as the linear version of the BWM method (Rezaei, 2016). It should be noted that nonlinear, random, and multiplicative versions of this method are also provided, but we use the linear version, which is the simplest and the most common version of the BWM method. If $\left(w_1^*, w_2^*, ..., w_m^*\right)$ are the optimal solutions of Model (5), ξ^{L*} is considered the inconsistency rate of the system. The closer the inconsistency rate is to zero, the greater the confidence in the judgments made by the DMs.

Omrani et al. (2020) presented the DEA-BWM multi-objective model by combining the concepts of DEA and BWM, which reduced the freedom of action of input and output weights by considering DM preferences. They also proposed goal programming for finding the common set of weights of inputs and outputs based on the decision of the DM. Fan et al. (2020) used the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) and BWM methods to improve the cross-efficiency method in DEA. After forming a cross-efficiency matrix in a DEA problem, they calculated the ideal and anti-ideal options and the Euclidean distance of the other options in the matrix from these two options. They extracted the final weights of the DMUs based on these distances and the BWM method. The integration of BWM and DEA with fuzzy data can be found in Chen and Ming (2020).

We will establish this relationship differently by considering a set of homogenous DMUs that use m inputs $x_{ij}, 1 \leqslant i \leqslant m,\ 1 \leqslant j \leqslant n$ to produce s output of $y_{rj},\ 1 \leqslant r \leqslant s,\ 1 \leqslant j \leqslant n$. The algorithm of our proposed BWM-AHP method is defined as follows.

Step 1. Calculate the best and the worst DMUs using Eqs. (6) and (7), respectively. Note that these units may be virtual units and may not exist externally.

$$\text{Best DMU} = \left(x_1^-, \cdots, x_m^-, y_1^+, \cdots, y_s^+\right);\ x_i^- = \min_{1 \leqslant j \leqslant n} x_{ij}\ \&\ y_r^+ = \max_{1 \leqslant j \leqslant n} y_{rj} \tag{6}$$

$$\text{Worst DMU} = \left(x_1^+, \cdots, x_m^+, y_1^-, \cdots, y_s^-\right);\ x_i^+ = \max_{1 \leqslant j \leqslant n} x_{ij}\ \&\ y_r^- = \min_{1 \leqslant j \leqslant n} y_{rj} \tag{7}$$

Step 2. After solving Model (1) for all DMUs, solve Models (8) and (9) for the best DMU and the worst DMU, respectively.

$$\begin{aligned} & E_{BB} = \max \sum_{r=1}^{s} u_r y_r^+ \\ & \text{s.t.} \\ & \sum_{i=1}^{m} v_i x_i^- = 1 \\ & \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leqslant 0;\ j = 1,2,\dots,n \\ & u_r \geqslant 0;\ r = 1,2,\dots,s \\ & v_i \geqslant 0;\ i = 1,2,\dots,m \end{aligned} \tag{8}$$

$$\begin{aligned} & E_{WW} = \max \sum_{r=1}^{s} u_r y_r^- \\ & \text{s.t.} \\ & \sum_{i=1}^{m} v_i x_i^+ = 1 \\ & \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leqslant 0;\ j = 1,2,\dots,n \\ & u_r \geqslant 0;\ r = 1,2,\dots,s \\ & v_i \geqslant 0;\ i = 1,2,\dots,m \end{aligned} \tag{9}$$

Step 3. Find pairwise comparisons of the Best DMU (BDMU) with other DMUs and the other DMUs with the Worst DMU (WDMU) through Eqs. (10) and (11), respectively, using Eq (2).

$$a_{Bj} = \frac{E_{BB}+E_{jB}}{E_{jj}+E_{Bj}} \quad \forall j \tag{10}$$

$$a_{jW} = \frac{E_{jj}+E_{Wj}}{E_{WW}+E_{jW}} \quad \forall j \tag{11}$$

Step 4. Obtain the weights of the DMUs by solving Model (5) and

Table 1
The semantic scale in MACBETH.

Semantic scale	Null	Very Weak	Weak	Moderate	Strong	Very Strong	Extreme
Their significances	Indifference between the alternatives	An alternative is very weakly more attractive than another	An alternative is weakly more attractive than another	An alternative is moderately more attractive than another	An alternative is strongly more attractive than another	An alternative is very strongly more attractive than another	An alternative is extremely more attractive than another
Numerical scale	0	1	2	3	4	5	6

rank the DMUs based on these weights.

3.3. Best method-DEA (BM-DEA) and worst method-DEA (WM-DEA)

Leal (2020) provided a simple version of the AHP method, considering only pairwise comparisons of the best option with the other options. Based on this idea, the BM-DEA method is presented as follows.

Step 1. Calculate the best DMU using Eq. (6). Note that this unit may be a virtual unit and may not exist externally.

Step 2. After solving Model (1) for all DMUs, solve Model (8) for BDMU.

Step 3. Define pairwise comparisons between BDMU and the other DMUs through Eq. (10) using Eq. (2).

Step 4. Obtain the weights of the DMUs from Eq. (12) and rank the DMUs based on these weights.

$$w_j = \frac{1/a_{Bj}}{\sum_{k=1}^{n} 1/a_{Bk}} \qquad \forall j \tag{12}$$

Considering WDMU and the idea presented by Leal (2020), the WM-DEA method can also be described as follows.

Step 1. Calculate the worst DMU using Eq. (7). Note that this unit may be a virtual unit and may not exist externally.

Step 2. After solving Model (1) for all DMUs, solve Model (9) for WDMU.

Step 3. Define pairwise comparisons between the other DMUs and WDMU through Eq. (11) using Eq. (2).

Step 4. Obtain the weights of the DMUs from Eq. (13) and rank the DMUs based on these weights.

$$w_j = \frac{a_{jW}}{\sum_{k=1}^{n} a_{kW}} \qquad \forall j \tag{13}$$

3.4. MACBETH-DEA method

MACBETH is a multi-attribute compensatory decision method developed by Bana e Costa and Vansnick (1994). This method quickly attracted the attention of researchers and was used in various decision-making problems (Bana e Costa et al., 2012; Ishizaka and Siraj, 2018). Applications of this method include performance analysis of online bookstores (Ertugrul and Qztas, 2016), evaluation and selection of flexible manufacturing systems (Karande and Chakraborty, 2013a), supplier performance assessment (Akyuz et al., 2018; Karande and Chakraborty, 2013b), and the evaluation of steam boiler alternatives (Kundakcı, 2019), among others.

The idea in MACBETH is to prioritize and rank criteria and alternatives based on pairwise comparisons and a distance scale of priority information. In this method, DMs are asked to make pairwise comparisons between two criteria or two alternatives using a seven-point scale, i.e., extreme, very strong, strong, moderate, weak, very weak, and null (Bana e Costa et al., 2002, 2012). The performance scores are usually qualitative arbitrations which are further quantified proportionately on a 0–100 scale. In this way, DMs can determine the relative and absolute weights of several alternatives based on the relative weights of multiple criteria. The M-MACBETH Software is available at (https://m-macbeth.com/) to facilitate all the necessary calculations. Although DMs generally use this method to prioritize qualitative criteria, it is also used for quantitative criteria. In particular, quantitative performance levels are also transformed into commensurate MACBETH measures with reference levels named good and neutral.

The algorithm for the new MACBETH-DEA method is presented as follows:

Step 1. Calculate the two reference levels, good and neutral, using Eqs. (6) and (7), respectively. Note that these units may be virtual DMUs and may not exist externally.

Step 2. After solving Model (1) for all DMUs, solve Models (8) and (9) for the good (BDMU) and neutral (WDMU) levels, respectively.

Step 3. Define the pairwise comparison matrix through Eqs. (4), (10), (11), and (14) using Eq. (2).

$$a_{BW} = \frac{E_{BB} + E_{WB}}{E_{WW} + E_{BW}} \tag{14}$$

Step 4. Obtain the MACBETH scores of the DMUs by solving Model (15) and ranking the DMUs based on these scores.

$$\begin{gathered} min[v(BDMU) - v(WDMU)] \\ s.t. \\ v(WDMU) = 0\ (arbitrary\ score) \\ v(DMU_i) - v(DMU_j) = 0\ ;\ i,j \in \{1,2,\dots,n;B,W\} \&\ a_{ij} = 1 \\ v(DMU_i) - v(DMU_j) \geqslant v(DMU_k) - v(DMU_l) + a_{ij} - a_{kl};\ i,j,k,l \in \{1,2,\dots,n;B,W\} \&\ a_{ij} > a_{kl} \geqslant 1 \end{gathered} \tag{15}$$

If the judgments made are inconsistent, problem (15) will be infeasible (Bana et al., 1999). Therefore, after forming the pairwise comparison matrix, we divide the maximum distance between the pairwise comparisons – up to 1 – by 6 and call the resulting value d. We then consider 1 equal to the zero scale and the values between 1 and $1 + d$ equal to the scale 1. This process is repeated until the values equivalent to scale six are determined from Table 1. Finally, we input the scales in M-MACBETH software and derive the final ranking.

Due to the proximity of the pairwise comparison values to each other and Model (15) constraints, the infeasibility of this Model is high. We incorporate the deviational variables d_{ijkl}, $\forall i,j,k,l \in \{1,2,\dots,n;B,W\}$, into Model (15) to increase the flexibility of its constraints, eliminating this latter shortcoming and delivering a model immediately applicable within a DEA environment. Thus, we propose Model (16) for the MACBETH-DEA method.

$$\begin{aligned} & min[v(BDMU) - v(WDMU)] + \sum_i \sum_j \sum_k \sum_l d_{ijkl} \\ & \text{s.t.} \\ & v(WDMU) = 0\ (arbitrary\ score) \\ & v(DMU_i) - v(DMU_j) = 0\ ;\ i,j \in \{1,2,\dots,n;B,W\} \&\ a_{ij} = 1 \\ & v(DMU_i) - v(DMU_j) + d_{ijkl} \geqslant v(DMU_k) - v(DMU_l) + a_{ij} - a_{kl};\ i,j,k,l \in \{1,2,\dots,n;B,W\} \&\ a_{ij} > a_{kl} \geqslant 1 \\ & d_{ijkl} \geqslant 0;\ \forall i,j,k,l\ \text{ such that } a_{ij} > a_{kl} \geqslant 1 \end{aligned} \tag{16}$$

Theorem 2. *Model (16) is always feasible.*

Proof. This is achieved by taking $d_{ijkl} = a_{ij} - a_{kl}; \forall i, j, k, l$, such that $a_{ij} > a_{kl} \geqslant 1$, and all other variables equal to zero. Therefore, we conclude that Model (16) is feasible, and this ends the proof. □.

If $(v^*(DMU_1), v^*(DMU_2), \dots, v^*(DMU_n), \boldsymbol{d})$ is the optimal solution of Model (16), then the weight of DMUs is given by $(v^*(DMU_1), v^*(DMU_2), \dots, v^*(DMU_n))$ and $\psi^* = \sum_i \sum_j \sum_k \sum_l d^*_{ijkl}$ will be considered as the inconsistency rate of the system. The closer the inconsistency rate is to zero, the greater the confidence in the judgments made by the experts.

A remark is due regarding the assumption that all the cross efficient comparisons introduced in the constraints are higher than one. That is, the inequalities $a_{ij} > a_{kl} \geqslant 1$ imposed within the third set of constraints in Equations (15) and (16) cannot hold for all elements of a pairwise comparison matrix, since, if this were the case, we should have $a_{ji} < a_{lk} \leqslant 1$. As illustrated in Equation (4), if a_{ij} and a_{ji} are not equal to 1, one of them will be greater than 1. MACBETH calculates the score of each DMU solely based on the latter pairwise comparisons.

4. Survey on new ranking methods

In this paper, an important relationship has been established between DEA and MADM. Different MADM methods based on pairwise comparisons have been used to rank DMUs in DEA. In a scenario with n DMUs, the AHP-DEA method requires solving n linear programming problems, calculating $(n^2 - n)/2$ pairwise comparisons, and using one of the methods available to compute the relative weights (i.e., least-squares, logarithmic least squares, eigenvector, or approximation techniques). Note that each method used for calculating relative weights has its own computational complexity. Applying the BWM-DEA method requires solving $n + 3$ linear programming problems and calculating $2n + 1$ pairwise comparisons, but there is no need to perform any extra calculations to find the relative weights. The introduction of BDMU and WDMU makes the comparison structure of DMUs more manageable than the ideal and anti-ideal alternatives used in the cross-efficiency matrix defined by Fan et al. (2020). In the BM-DEA and WM-DEA methods, we only need to solve $n + 1$ linear programming problems and calculate n pairwise comparisons. In these latter methods, there is no need to calculate the relative weights.

The decrease in the number of pairwise comparisons required by BWM-DEA relative to AHP-DEA implies that the inconsistency of the final judgments should be lower. The AHP-DEA method requires performing the calculations relevant to obtain the inconsistency ratio. Saaty (1980) suggests that the results of the pairwise comparisons are acceptable if the inconsistency ratio is less than or equal to 0.1. Aguarón et al. (2020) suggested using the geometric consistency index method to reduce the inconsistency in AHP. On the other hand, the optimal value of Model (5) includes all the consistent information required by BWM-DEA. In particular, a value of the objective function close to zero indicates consistency in the judgments. Therefore, there are no inconsistencies in the final judgments produced by the BM-DEA and WM-DEA methods. In the MACBETH-DEA case, the use of pairwise

Table 2
A comparison of different ranking methods.

Methods	Linear programs	Pairwise comparisons	Additional calculations to find relative weights	Inconsistency potential
AHP-DEA	*n*	*Yes*	*High*	
BWM-DEA	$n+3$	$2n+1$	*No*	*Medium*
BM-DEA	$n+1$	*n*	*No*	*No*
WM-DEA	$n+1$	*n*	*No*	*No*
MACBETH-DEA	$n+3$	$(n+2)(n+1)/2$	*No*	*High*

Table 3
The input and output data for six DMUs in Sexton et al. (1986).

DMU	Input1	Input2	Output1	Output2
DMU1	150	0.2	14,000	3500
DMU2	400	0.7	14,000	21,000
DMU3	320	1.2	42,000	10,500
DMU4	520	2	28,000	42,000
DMU5	350	1.2	19,000	25,000
DMU6	320	0.7	14,000	15,000

comparisons is slightly different. This method is similar to BWM-DEA in terms of computational volume and similar to AHP-DEA in terms of the information required. A summary of the comparison between the methods proposed and reviewed, assuming a total of *n* DMU is given in Table 2.

To better compare the proposed methods with each other and other ranking methods, we consider the following example with six DMUs, two inputs, and two outputs (Sexton et al., 1986). The inputs and outputs of each DMU are listed in Table 3.

The results of the most popular ranking methods in DEA and those proposed in this study applied to Sexton et al.'s (1986) example are presented in Fig. 1. This figure shows that the traditional CCR models rank four efficient DMUs first. The super-efficiency methods rank the efficient units based on their respective models and the inefficient ones based on their efficiency scores. The remaining methods presented rank all DMUs. Whenever appropriate, the results are described from both the optimistic (benevolent) and the pessimistic (aggressive) perspectives. For simplicity, these DEA hybrid methods are called AHP-Benevolent, AHP-Aggressive, BWM-Benevolent, BWM-Aggressive, BM-Benevolent, BM-Aggressive, WM-Benevolent, WM-Aggressive, MACBETH-Benevolent, and MACBETH-Aggressive. One of the most remarkable features of Fig. 1 is the increase in ranking variability observed as the benevolent, and aggressive versions of the hybrid models are introduced, illustrating the effect of the reference solutions in determining the ranking delivered by these techniques.

The variety of ranking methods considered implies that DMUs are evaluated from very different perspectives. The results further show how all super-efficiency methods rank inefficient DMUs fifth and sixth since these DMUs were ranked based on their efficiency scores while the other units were ranked based on their super-efficiency scores. In contrast, DMUs are ranked according to their best efficiency score and their efficiency score compared to other DMUs in all methods considering pairwise comparisons.

5. Selecting a ranking method

Each ranking method ranks DMUs from a different perspective. The million-dollar question is, which ranking method should a DM choose? One way to deal with this question is to combine ranking results using techniques such as the Copeland method (Momeni, 2016), preferential voting (Soltanifar and Shahghobadi, 2013), or other similar methods. However, we are not only interested in a consensus ranking; we are searching for suitable rankings within a given set of possible ones, which requires defining a similarity score among rankings akin to a standard correlation coefficient. To this end, we propose a simple yet powerful ranking method called Rank Absolute Deviation (RAD), inspired by the concept of mean absolute deviation in statistics. The premise of this technique is to select the ranking methods producing similar results to those of the competing methods. In Section 6, we will illustrate the similarities between the RAD method proposed and a commonly implemented one such as MAH. The latter is described in Section 5.2.

5.1. The rank absolute deviation method

Let us assume that *K* ranking methods are utilized and denoted by $DMU_i^k (i = 1, 2, \ldots, n; k = 1, 2, \ldots, K)$ the i^{th} priority DMU from the viewpoint of the k^{th} method. We use Eq. (16) to find the priority matrix for

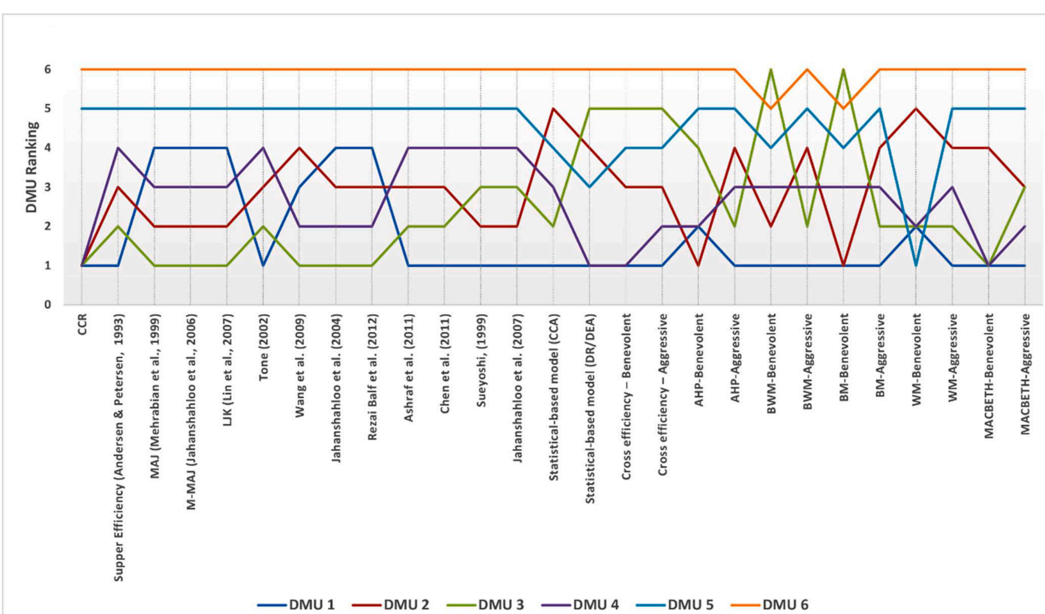

Fig. 1. DMU rankings for Sexton et al.'s (1986) problem solved with different methods.

Table 4
Statistical summary of input and output data.

Statistical Parameters	Inputs			Outputs			
	Staffing score	Interests Paid*	Arrears*	Total Resources*	Loans offered*	Charges received*	Interests received*
Max	34.35	2,274,754,862	37,853,546,276	135,894,969,932	347,812,000,000	381,190,572	4,801,018,177
Min	2.67	112,787,061	1,882,957	10,480,341,806	8,481,463,453	3,790,591	2,209,178
Mean	9	503,891,934	2,752,188,729	40,781,792,487	59,898,992,728	53,420,530	305,850,430
Standard Deviation	6.441687786	405817622	5934430277	27880295240	56589540867	57329316.25	647402877.6

*Thousand Rials.

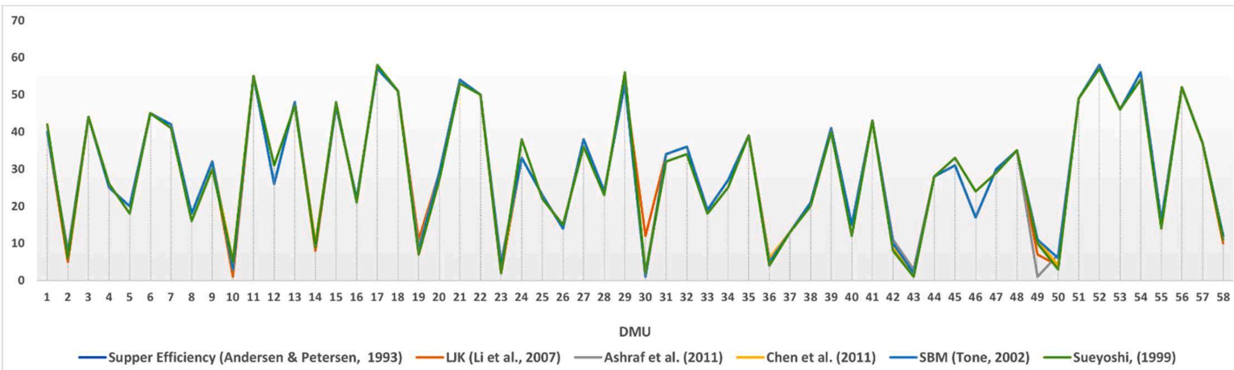

Fig. 2. Comparison of the results of different ranking methods on bank Branches.

each method. That is, a binary pairwise comparison matrix of the DMU rankings is constructed for each method. A zero is used if a DMU in a row of the matrix is ranked less than or equal to the DMU in the column, and a one is used if the DMU in the row is ranked greater than the DMU in the column of the matrix. Note that the diagonal values are always equal to zero.

$$R_k = \left[a_{ii'}^k\right]_{n\times n},\quad a_{ii'}^k = \begin{cases} 0 & \text{if } DMU_i^k \leqslant DMU_{i'}^k \\ 1 & \text{if } DMU_i^k > DMU_{i'}^k \end{cases};\ i,i' = 1,2,\dots,n;\ k = 1,2,\dots,K \tag{17}$$

Next, we use Eq. (17) to obtain the RAD score between each method and other methods. The ranking with the least RAD score is the most similar ranking to others. That is, if the rankings of method k have the least distances (are completely similar) to the rankings of all other ranking methods, the similarity is maximum, and the RAD score will be equal to zero. However, if the rankings of method k have the most distances (are completely dissimilar) to the rankings of all other ranking methods, the similarity is minimum, and the RAD score will be equal to $(K-1)\times(n^2-n)$.

Denote by ρ_k the total deviation score of the ranking method k compared to other methods.

$$\rho_k = \sum_{\substack{k'=1\\k'\neq k}}^{K}\sum_{i=1}^{n}\sum_{i'=1}^{n}\left|a_{ii'}^k - a_{ii'}^{k'}\right|, k = 1,2,\dots K \tag{18}$$

As noted before, the diagonal values are always equal to zero and $\rho_k \in [0,(K-1)\times(n^2-n)]$. Thus, we use Eq. (18) to normalize ρ_k and derive the RAD score for each method:

$$RAD_k = \frac{\rho_k}{(K-1)\times(n^2-n)}, k = 1,2,\dots,K \tag{19}$$

This normalization is applied to standardize the RAD score ($0\leqslant RAD_k\leqslant 1$) and select the ranking method with the least RAD score according to Eq. (19):

$$RAD_{k^*} = \min_{1\leqslant k\leqslant K} RAD_k \tag{20}$$

A numerical example is presented in Appendix B to demonstrate the working details of the RAD method and its comparison with the mean absolute deviation. Note that the introduction or omission of a ranking method does not modify the distances existing among the other methods but affects their RAD scores since the distances with respect to the rankings introduced or omitted would differ across methods. Thus, ranking modifications could arise since the score is determined by the distances among the rankings of all the techniques considered, and adding or omitting methods modifies these values.

5.2. The maximize agreement heuristic method

Beck and Lin (1983) introduced the MAH method to maximize the consensus of different raters in decision-making problems by ranking the alternatives through the resulting Final Consensus Ranking (FCR), which has been applied within a wide range of MADM scenarios (Kengpol and Tuominen, 2006). We build on the guidelines provided by Beck and Lin (1983) to define consensus among the methods described.

Consider k MADM methods used to rank n alternatives. Define an agreement matrix, A, where a_{ij} refers to the number of methods in which alternative i is preferred to j. Suppose we add up the column values for each alternative i. In that case, we obtain a positive preference column vector, P, where each element describes the number of times alternative i is preferred to every other alternative:

$$P_i = \sum_{j=1}^{n} a_{ij},\quad i = 1,2,\dots,n. \tag{21}$$

On the other hand, if we add the values of each alternative j for all rows, we obtain a negative preference row vector, N, where each element describes the number of times the other alternatives are preferred to j:

$$N_i = \sum_{j=1}^{n} a_{ji},\quad i = 1,2,\dots,n. \tag{22}$$

The outcome of both equations determines the consensus criterion. If alternative i displays a value of zero in the corresponding entry of the negative preference vector, it should not be ranked below any other

Table 5
Peoples Bank branch ranking according to ten ranking methods.

Bank Branches	AHP-Benevolent	AHP-Aggressive	BWM-Benevolent	BWM-Aggressive	BM-Benevolent	BM-Aggressive	WM-Benevolent	WM-Aggressive	MACBETH-Benevolent	MACBETH-Aggressive
DMU1	41	41	38	38	43	43	55	55	49	42
DMU2	7	4	9	5	20	5	21	7	15	11
DMU3	40	40	44	44	47	47	37	38	42	43
DMU4	32	32	26	27	32	34	42	43	27	32
DMU5	25	26	14	21	23	28	39	40	21	21
DMU6	37	37	48	49	51	51	32	33	33	35
DMU7	44	44	40	41	45	45	46	46	52	41
DMU8	10	15	12	19	22	26	7	17	16	5
DMU9	24	27	32	32	37	39	5	15	28	25
DMU10	19	8	19	9	26	18	23	5	15	23
DMU11	58	58	56	56	56	56	58	58	58	58
DMU12	28	28	13	20	26	27	33	34	31	27
DMU13	48	45	53	53	53	53	41	42	40	50
DMU14	11	11	10	11	20	20	19	6	15	9
DMU15	54	54	46	46	49	49	52	51	54	53
DMU16	34	34	17	22	25	29	44	44	37	34
DMU17	55	56	58	58	58	58	38	39	45	57
DMU18	51	50	45	45	48	48	53	52	46	51
DMU19	2	5	15	6	24	16	2	2	15	1
DMU20	31	31	30	31	36	38	34	35	32	30
DMU21	39	39	55	55	55	55	26	28	43	39
DMU22	53	53	49	49	51	51	54	54	55	52
DMU23	4	1	31	3	37	3	14	1	15	13
DMU24	47	48	43	43	46	46	27	29	36	37
DMU25	17	21	5	14	17	23	24	27	26	19
DMU26	3	13	3	10	15	19	9	19	2	2
DMU27	12	16	40	40	45	45	3	12	3	14
DMU28	23	23	25	26	31	33	6	16	24	29
DMU29	57	57	52	52	52	52	50	53	57	56
DMU30	1	3	2	1	1	1	1	4	15	4
DMU31	30	30	33	33	39	40	15	23	29	31
DMU32	27	25	35	35	40	41	16	24	25	26
DMU33	13	17	18	23	26	30	4	13	17	12
DMU34	21	20	27	28	33	35	12	22	23	18
DMU35	42	42	36	36	41	42	47	47	50	45
DMU36	20	9	20	7	27	16	29	10	15	22
DMU37	22	22	4	13	16	22	31	32	15	25
DMU38	26	24	24	25	30	32	30	31	22	16
DMU39	35	35	39	39	44	44	10	20	35	33
DMU40	17	19	6	16	18	24	20	26	19	17
DMU41	43	43	51	51	52	52	40	41	44	46
DMU42	14	7	22	12	28	21	25	9	15	10
DMU43	9	2	1	2	1	2	43	11	15	20
DMU44	38	38	29	30	35	37	49	49	41	40
DMU45	29	29	23	24	28	31	35	36	30	28
DMU46	15	18	7	15	18	24	17	25	18	8
DMU47	36	36	28	29	34	36	36	37	39	38
DMU48	46	47	34	34	40	41	48	48	48	44
DMU49	16	10	21	8	27	17	22	8	1	15
DMU50	5	6	16	4	24	4	14	3	15	6
DMU51	52	52	47	47	50	50	57	57	56	54
DMU52	56	55	57	57	57	57	45	45	47	55
DMU53	49	49	42	42	46	46	51	50	51	49
DMU54	50	51	50	50	52	52	56	56	53	48
DMU55	8	14	8	18	19	25	8	18	20	7
DMU56	45	46	54	54	54	54	28	30	38	47
DMU57	33	33	37	37	41	42	11	21	34	36
DMU58	6	12	11	17	21	24	18	14	15	3

alternative in the consensus ranking. On the other hand, assume that alternative i displays a value of zero in the corresponding entry of the positive preference vector. This implies that this alternative should not be ranked above any other alternative in the consensus ranking.

It, therefore, follows that the value $(P_i - N_i)$ determines the ranking of the alternatives displaying no zero-entries within the negative or positive preference vectors. Thus, if an alternative i displays the maximum value of $|P_i - N_i|$ for a positive $(P_i - N_i)$, it should be placed at the top of the positions available within the FCR. On the other hand, if an alternative i displays the maximum value of $|P_i - N_i|$ for a negative $(P_i - N_i)$, it should be place at the bottom of the positions available within the FCR.

The main steps determining the FCR are summarized below:

***Step* 1.** Define the agreement matrix A, with the parameter n referring to the number of alternatives being considered.

***Step* 2.** The entries of the N and P preference vectors are computed using Equations (20) and (21).

***Step* 3.** Alternatives displaying a zero-value-entry in P are placed at the bottom of the consensus ranking. Similarly, those alternatives displaying a zero-value entry in N are placed at the top of the ranking.

***Step* 4.** The remaining alternatives are ranked based on the value of $P_i - N_i$. Whenever the maximum difference is positive, the corresponding alternative is placed in the highest position available among the remaining ones. However, when the maximum difference is negative,

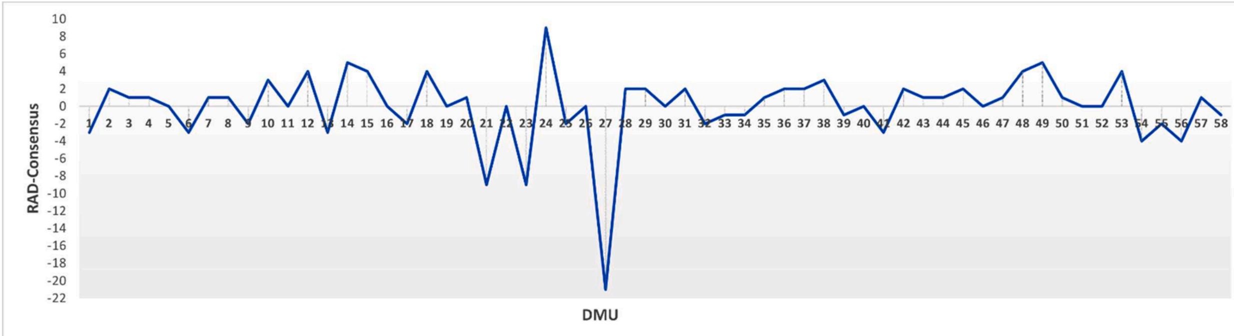

Fig. 3. Position differences between the rankings derived from the RAD and MAH consensus methods.

the corresponding alternative is placed in the lowest position available among the remaining ones. Ties are broken arbitrarily. The row and column corresponding to alternative i are eliminated from the agreement matrix.

Step 5. Set/Rename $n-1 = n$.

***Step* 6.** If $n > 1$, go back to Step 2. If $n = 1$, place the final alternative in the highest position available among the remaining ones and stop.

6. A real-world application

In this application, we use the proposed methods to rank 58 branches of the Peoples Bank[1] in Iran. We consider three inputs and four outputs. A brief description of the inputs and outputs is presented next. The variables selected are standard to the efficiency models applied to study financial intermediaries (Khalili-Damghani et al., 2016), with inputs reflecting the inflow of funds into the bank and outputs considering the main consequences of the use of these funds. The data was retrieved through 2018 and provided by the managers of the 58 bank branches studied.

6.1. Inputs:

- **Staffing score (Input 1):** This input is a composite score reflecting a series of quantitative and qualitative criteria used by the human resources department at the bank. The staffing score is the weighted sum of the workforce size, level of training, and education provided to the bank employees.
- **Interests paid (Input 2):** Interests paid are the interests banks pay for savings deposits or money borrowed from the government to attract new customers and maintain the ones they have. These interests are expenses for the bank
- **Arrears (Input 3):** Banks grant loans to their personal and business account holders. Some customers may experience financial difficulties and pay late or never pay back their loans. Bank arrears are obligations or liabilities that have not received payment by their due dates. These late payments or unpaid loan obligations are expenses for the bank.

6.2. Outputs:

- **Total resources (Output 1):** Banks collect deposits from their personal or business account holders. In addition, banks receive government incentives or low-interest loans, investment properties, and stocks and bonds. Total Resources are considered assets for the bank.
- **Loans offered (Output 2):** The amounts of bank loans offered to personal or business customers are considered output in a bank.
- **Charges received (Output 3)**: Banks collect service charges or bank fees for services such as ATM charges or credit card maintenance fees. These bank charges are revenues for the bank.
- **Interests received (Output 4):** Banks lend money to their personal or business customers and, in turn, receive interest on these loans. These interests are revenues for the bank.

Table 4 provides a statistical summary of the input and output data. The whole set of input and output data for the 58 branches of the Peoples Bank is presented in Appendix C.

6.3. Numerical results

Fig. 2 compares the ranking results for the 58 branches of the Peoples Bank for different super-efficiency and slack-based methods previously discussed in the DEA literature and Sexton's example within Section 4. In the ranking methods displayed in this figure, the results are based on an efficiency score or a super-efficiency score obtained by each DMU. What is certain – despite the similarity of the rankings – is that each method ranks the DMUs from a different perspective and may lead to different results from other methods. In this paper, we have examined those ranking methods in which a DMU is ranked not only by considering its best performance score but also by considering its performance score compared to other DMUs. We believe that this feature can increase confidence in the results.

Table 5 presents the rankings obtained for the 58 branches of the Peoples Bank according to the hybrid DEA methods proposed in this study.

The ten methods presented in Table 5 all apply pairwise comparisons. Next, we compare the ten ranking results through the RAD method. We use Eq. (18) and determine the following RAD scores ($0 \leqslant RAD \leqslant 1$) for the ten methods in the order of most similar to least similar to others:

1. $RAD_{\text{AHP-Benevolent}} = 0.1310$
2. $RAD_{\text{AHP-Aggressive}} = 0.1328$
3. $RAD_{\text{BM-Aggressive}} = 0.1422$
4. $RAD_{\text{BWM-Aggressive}} = 0.1424$
5. $RAD_{\text{MACBETH-Aggressive}} = 0.1448$
6. $RAD_{\text{MACBETH-Benevolent}} = 0.1598$
7. $RAD_{\text{BM-Benevolent}} = 0.1663$
8. $RAD_{\text{BWM-Benevolent}} = 0.1675$
9. $RAD_{\text{WM-Aggressive}} = 0.1945$
10. $RAD_{\text{WM-Benevolent}} = 0.2482$

The AHP-Benevolent method was selected in the current study case since $RAD_{\text{AHP - Benevolent}} = \min_k RAD_k$. It should be noted that this choice varies in different applications depending on the ranking scores. In this regard, one of the main objectives of the manuscript is to highlight the differences in the rankings delivered by the set of hybrid models. Note that we are operating with vectors when dealing with the BWM, BM, and

[1] The name is changed to protect the anonymity of the bank.

Table 6
Spearman ranks correlation coefficients among the different hybrid techniques.

	AHP-Benevolent	AHP-Aggressive	BWM-Benevolent	BWM-Aggressive	BM-Benevolent	BM-Aggressive	WM-Benevolent	WM-Aggressive	MACBETH-Benevolent	MACBETH-Aggressive	MAH Consensus
AHP-Benevolent	1.000	0.979	0.852	0.896	0.855	0.896	0.790	0.879	0.938	0.978	0.970
AHP-Aggressive	0.979	1.000	0.831	0.921	0.834	0.921	0.737	0.900	0.949	0.944	0.961
BWM-Benevolent	0.852	0.831	1.000	0.930	0.999	0.929	0.516	0.608	0.780	0.849	0.908
BWM-Aggressive	0.896	0.921	0.930	1.000	0.931	1.000	0.541	0.738	0.848	0.867	0.944
BM-Benevolent	0.855	0.834	0.999	0.931	1.000	0.931	0.523	0.614	0.784	0.853	0.911
BM-Aggressive	0.896	0.921	0.929	1.000	0.931	1.000	0.542	0.740	0.848	0.867	0.945
WM-Benevolent	0.790	0.737	0.516	0.541	0.523	0.542	1.000	0.875	0.763	0.794	0.736
WM-Aggressive	0.879	0.900	0.608	0.738	0.614	0.740	0.875	1.000	0.895	0.845	0.852
MACBETH-Benevolent	0.938	0.949	0.780	0.848	0.784	0.848	0.763	0.895	1.000	0.923	0.917
MACBETH-Aggressive	0.978	0.944	0.849	0.867	0.853	0.867	0.794	0.845	0.923	1.000	0.962
MAH Consensus	0.970	0.961	0.908	0.944	0.911	0.945	0.736	0.852	0.917	0.962	1.000

WM. Equations (10) and (11) define the pairwise comparisons in these models and describe vectors, with the best and worst DMUs acting as the benchmarks determining the comparisons with other DMUs. It also makes intuitive sense that the values in Equations (10) and (11) are higher than one, which allows for the implementation of the BWM directly.

However, when moving to MACBETH, the constraints shift from vectors of pairwise comparisons – relative to the best and neutral or worse references – to matrices comparing all elements. It is, therefore, natural that the results obtained from different DEA ranking models differ since each model ranks DMUs from a concrete formal perspective. For example, BWM ranks DMUs using $2n+1$ pairwise comparisons, which the BM and WM techniques reduce to a total of n. MACBETH requires a substantially larger number of pairwise comparisons, namely, $(n+2)(n+1)/2$. This will increase both the sensitivity of the model to the optimal weights selected from Model (1) when defining the benevolent and aggressive perspectives and the potential inconsistency of the results. As a consequence, DMUs 8, 27, 30, 49, and 58 display extreme ranks under MACBETH, whose compatibility restrictions introduce modifications that deviate the ranking from the consensus ones described in the paper.

All in all, the rankings derived from these models will present considerable differences. We elaborate further on these features in the next section.

6.4. Comparing hybrid ranking techniques

We now define two distinct sets of results based on the different rankings obtained. First, we compare the results derived from RAD – which selects the AHP-Benevolent method as the one most similar to others – with those of a standard MAH consensus approach by computing the differences in positions between both rankings for each DMU. Fig. 3 illustrates the similarity between both ranking patterns in these terms.

Note how both methods generally rank DMUs within a range of (-2,3) positions of difference. This similarity is preserved through both rankings with the exceptions of DMUs 27 – which displays the largest difference between rankings –, 24, 23, and 21 with a difference of 9 positions each. More importantly, both techniques coincide in ranking the first three DMUs, namely, 30, 19, and 26.

Additional intuition can be obtained by computing the Spearman correlation among the different hybrid methods. The results are presented in Table 6 and validate the intuition derived from the previous analyses. The WM-Benevolent and WM-Aggressive hybrid methods display relatively lower correlation coefficients than the other techniques, which are highly correlated among themselves and with the consensus ranking.

The second set of results builds on Fig. 4, which illustrates the differences in ranking positions between the BM-aggressive and AHP-aggressive hybrid methods relative to the MAH consensus ranking, respectively. Note how the BM-aggressive hybrid method displays a higher variability relative to the consensus when compared to the AHP-aggressive technique. A similar result arises when considering the AHP-benevolent hybrid method, which defines the RAD ranking. However, the differences are not substantial, as the correlation analysis has shown.

We have illustrated how the RAD and consensus methods deliver very similar rankings, coinciding with the first three alternatives. When considering the different hybrid techniques, AHP-Benevolent displays the highest similarity in terms of RAD. However, BM-Aggressive provides a good approximation – and is computationally and consistently simpler –. We have validated this idea through correlation analyses, illustrating how most techniques deliver highly correlated rankings. The visual analysis of the ranking differences between techniques provides a more intuitive approach, emphasizing the idea that a simpler method such as BM-Aggressive delivers a highly correlated ranking with the main consensus approaches. However, other than when considering the

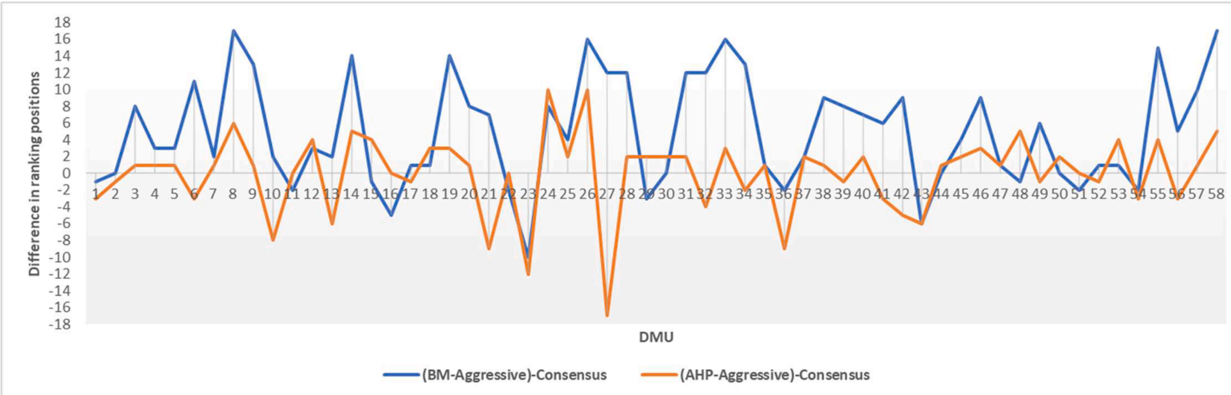

Fig. 4. Ranking position differences between the BM-aggressive and AHP-aggressive hybrids and the MAH consensus method.

Table B1
Ranking matrix.

DMU	Method1 (M1)	Method2 (M2)	Method3 (M3)	Method4 (M4)
DMU1	2	1	2	5
DMU2	1	3	1	2
DMU3	3	5	4	4
DMU4	4	2	3	1
DMU5	5	4	5	3

Table B2
Rank deviation matrices.

	Method1 (M1)					Method3 (M3)				
DMU	DMU1	DMU2	DMU3	DMU4	DMU5	DMU1	DMU2	DMU3	DMU4	DMU5
DMU1	0	0	1	1	1	0	0	1	1	1
DMU2	1	0	1	1	1	1	0	1	1	1
DMU3	0	0	0	1	1	0	0	0	0	1
DMU4	0	0	0	0	1	0	0	1	0	1
DMU5	0	0	0	0	0	0	0	0	0	0
DMU1	0	1	1	1	1	0	0	0	0	0
DMU2	0	0	1	0	1	1	0	1	0	1
DMU3	0	0	0	0	0	1	0	0	0	0
DMU4	0	1	1	0	1	1	1	1	0	1
DMU5	0	0	0	1	0	1	0	1	0	0
	Method2 (M2)					Method4 (M4)				

Table B3
RAD Results.

Ranking method	ρ_k	RAD_k	Method ranking
Method (M1)	22	0.367	2
Method (M2)	24	0.400	3
Method (M3)	**18**	**0.300**	**1**
Method (M4)	32	0.533	4

Table C1
Input and output data.

Bank Branches	Inputs			Outputs			
	Staffing score	Interests Paid	Arrears	Total Resources	Loans offered	Charges received	Interests received
DMU1	33.98	1079358382	2086168200	91898780446	220469000000	72724247	9.06E+08
DMU2	13.15	747456958	121796222	56646389489	101334000000	37738410	1458250548
DMU3	6.04	311866711	164783100	23684649429	31746964459	26690662	8.73E+07
DMU4	18.64	1031120275	1182161972	73363125327	158690000000	47258172	9.47E+08
DMU5	4.54	209183080	820113418	15257878659	44842531204	29136145	2.24E+08
DMU6	11	378307741	213817809	30281651888	50609705055	52000292	2.88E+08
DMU7	16.31	730377583	6749810197	70908856674	83225550325	84525013	2.56E+08
DMU8	4.26	213340711	58233105	20277684747	34481739099	13546827	1.67E+08
DMU9	4.81	285196512	95456828	26345268063	22850208765	25795686	3.67E+07
DMU10	20.94	2274754862	37853546276	129300039762	347812000000	55040708	4801018177
DMU11	10.39	464246324	6331145727	33996467114	42914415082	54642451	1.67E+08
DMU12	10.15	888386856	264435033	65359066067	38180659776	77909563	1.11E+08
DMU13	9.65	569571327	1003502567	42636854807	34238312340	54213650	2.53E+08
DMU14	7.51	647567555	812334511	54235717371	79434731566	30712579	667839912
DMU15	12.18	732310630	4088977937	50134201956	67928353656	31106259	1.68E+08
DMU16	12.85	503878040	6803910835	39942883553	120058000000	71588992	1.83E+08
DMU17	4.38	215973701	138806808	15520778468	15425158663	14189010	4.70E+07
DMU18	4.44	267680990	1003173193	19581326749	13766492606	19006713	1.23E+08
DMU19	4.66	240327459	32400507	34199132981	22633744058	55525368	65953223
DMU20	9.51	539621093	355824384	48437419923	57869744453	111001842	9.31E+07
DMU21	4.85	258694781	390618794	18649320014	19261077765	44564778	3.02E+07
DMU22	6.88	324327450	1761559332	25520659249	31392470715	48668845	1.01E+08
DMU23	12.21	1151126508	100239722	101099612459	102374000000	381190572	310491516
DMU24	4.65	303011951	20000000000	22745817944	21613603741	17890127	3.46E+08
DMU25	6.53	513373118	233162000	45493103310	18653120532	25633080	8.38E+07
DMU26	9.48	951481305	1717749529	76699002954	85410611350	97544683	3.31E+08
DMU27	5.56	308080148	35381332	26689014980	20275988867	16247415	9.93E+07
DMU28	2.67	204131635	60354794	15429139143	8481463453	4969623	1.14E+08
DMU29	34.35	1694107282	15111611772	135894969932	161761000000	91378152	4.05E+08
DMU30	6.12	431865745	1882957	36024728534	36565325297	49149445	159263836
DMU31	5.67	282195319	82385214	24246450096	27052155548	25889624	2.01E+08
DMU32	6.84	213896431	73706784	20041692851	30392894598	9503699	2.50E+08
DMU33	4.76	237636396	156214731	27364531815	30782208801	6190712	1.57E+08
DMU34	5.19	224872448	31112538	18090303486	26674432747	29419107	2.42E+08
DMU35	26.3	763670673	8699223086	82141983526	147523000000	123174178	8.18E+07
DMU36	8.85	147531933	4610966754	26626433427	56106390968	91150781	1.05E+07
DMU37	12.2	410878532	9329126361	43696861637	125781000000	58704233	2.21E+06
DMU38	7.5	333582109	2302732316	27473760066	60630371332	89521688	2.25E+08
DMU39	4.29	197983646	76618478	19240770614	18018796009	12879448	1.37E+07
DMU40	6.7	454272734	4328571212	41402957147	67720690717	60916995	5.27E+07
DMU41	4.89	172163458	428165612	12505693070	13787129165	26931434	1.50E+08
DMU42	4.05	141605450	362391084	14108905095	34017180746	3790591	424057749
DMU43	4.8	112787061	6087334	10480341806	53704561791	9836718	340379388
DMU44	5.92	197859023	1736948341	17177674685	47443669311	10912422	1.93E+07
DMU45	11.23	1008144971	799888409	69519520671	50342813036	150201949	1.21E+08
DMU46	12.04	1410965027	712052821	88431130203	73380424398	107951057	3.88E+08
DMU47	10.55	718757712	3106946729	59083530442	67457583720	79202081	1.72E+08
DMU48	5.72	299039079	3118411405	19769139505	44277668582	24740449	8.18E+06
DMU49	6.47	454971307	3999989615	35186243162	59473584399	177263231	16672265
DMU50	6.03	145330526	329595121	44294011009	26545626886	21646027	2732551
DMU51	9.31	354774612	1033254371	26038732430	52324771345	20958898	2.47E+08
DMU52	7.04	383065651	96678903	23740214769	18214008786	34995971	1.50E+08
DMU53	8.12	606265050	2443557421	37809295924	47931834559	33380294	3.23E+08
DMU54	8.78	176162886	880865398	16161780072	34771456232	12587454	3.65E+07
DMU55	13.22	332839827	394388652	34253388199	65670887687	84799279	1.95E+08
DMU56	4.89	258281304	270616937	19096094735	21953697669	3920575	1.41E+08
DMU57	5.43	408549570	117132994	29237028213	25417664084	17678876	1.17E+08
DMU58	6.89	306922751	506358795	31871953609	52445102264	28653660	620927500

initial alternatives composing the rankings, differences naturally arise among techniques.

7. Conclusion and future research directions

This study has established an important and noteworthy relationship between DEA and MADM. We have discussed the strengths and weaknesses of each ranking method and studied several strategies to diminish the weaknesses of a method with the strengths of the other. We have introduced several new hybrid MADM-DEA techniques of different computational complexity and consistency and showed that the two methods are a match made in heaven.

Each method ranks the business units from a different perspective. Thus, we have developed a new RAD technique to evaluate the similarity among multiple ranking results in MADM. Unlike preferential voting, the MAH, and the Copeland method, which combine multiple rankings into one consensus ranking, RAD defines a similarity score designed to search for more suitable rankings. The central premise of RAD is the selection of the methods producing similar results for making informed decisions with confidence. Several examples and a real-world case study have been used to demonstrate the applicability and effectiveness of the new hybrid MADM-DEA methods, including MACBETH-DEA and RAD.

We have focused on highlighting how simpler hybrid MADM-DEA techniques deliver similar rankings to the consensus and benchmark

DEA-AHP ones, the latter being more computationally and consistently demanding, with MACBETH-DEA providing additional evidence on this quality. These findings imply that computational complexity and strict consistency requirements may lead to substantial variations in the ranking, far from the consensus one, which can be approached via simpler methods. Clearly, in the current setting, the consensus ranking is based on a majority of similar techniques that reduce the relative importance of the MACBETH-DEA hybrid. However, the high correlation displayed with DEA-AHP provides additional intuition validating the results obtained. Despite this fact, the conclusions presented require additional validation through the introduction and analysis of further models.

As suggestions for future research, expanding this established relationship to other MADM techniques, including outranking methods such as ÉLimination Et Choix Traduisant la REalité (ELECTRE) and the Preference Ranking Organization METHod for Enrichment of Evaluations (PROMETHEE), and developing a fuzzy version of these models for decision-making problems under uncertainty could be considered valuable extensions of this study.

CRediT authorship contribution statement

Madjid Tavana: Conceptualization, Methodology, Formal analysis, Writing – review & editing, Validation. **Mehdi Soltanifar:** Conceptualization, Methodology, Formal analysis, Writing – review & editing, Visualization. **Francisco J. Santos-Arteaga:** Methodology, Formal analysis, Writing – review & editing. **Hamid Sharafi:** Investigation, Writing – review & editing, Software.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

Dr. Francisco J. Santos-Arteaga is grateful for the support received from the María Zambrano contract from the Universidad Complutense de Madrid financed by the Ministerio de Universidades with funding from the European Union Next Generation program.

Appendix A

Differentiating between benevolent and aggressive DEA perspectives

The optimistic (benevolent) and pessimistic (aggressive) perspectives are determined by choice of optimal weights derived from Model (1). If Model (1) provides a unique optimal solution, both perspectives, benevolent and aggressive, coincide. However, if Model (1) has multiple optimal solutions, as is usually the case, the choice of each optimal solution can affect other units' scores. If the optimal solution chosen is such that the DMU is evaluated in the best conditions, we have selected the benevolent view. On the other hand, if the optimal solution chosen is such that the DMU is evaluated in the worst-case conditions, we have chosen an aggressive view. In summary, the following models have been used in the DEA literature to define the benevolent and aggressive perspectives after solving Model (1).

Benevolent perspective for DMU_p	Aggressive perspective for DMU_p
$E_{op} = \max \sum_{r=1}^{s} u_{ro} y_{rp}$	$E_{op} = \min \sum_{r=1}^{s} u_{ro} y_{rp}$
s.t. $\sum_{i=1}^{m} v_{io} x_{ip} = 1$	s.t. $\sum_{i=1}^{m} v_{io} x_{ip} = 1$
$\sum_{r=1}^{s} u_{ro} y_{ro} - E_{oo}^{*} \sum_{i=1}^{m} v_{io} x_{io} = 0$	$\sum_{r=1}^{s} u_{ro} y_{ro} - E_{oo}^{*} \sum_{i=1}^{m} v_{io} x_{io} = 0$
$\sum_{r=1}^{s} u_{ro} y_{rj} - \sum_{i=1}^{m} v_{io} x_{ij} \leqslant 0;\ j = 1,2,\ldots,n$	$\sum_{r=1}^{s} u_{ro} y_{rj} - \sum_{i=1}^{m} v_{io} x_{ij} \leqslant 0;\ j = 1,2,\ldots,n$
$u_{ro} \geqslant 0;\ r = 1,2,\ldots,s$	$u_{ro} \geqslant 0;\ r = 1,2,\ldots,s$
$v_{io} \geqslant 0;\ i = 1,2,\ldots,m$	$v_{io} \geqslant 0;\ i = 1,2,\ldots,m$

Appendix B

The RAD solution for a simple example

In this section, we illustrate the behavior of the RAD method with a simple example. Suppose five DMUs ($n = 5$) are ranked by four different methods ($K = 4$), as shown in Table B1.

For each ranking method, a matrix is formed based on Eq. (16). Table B2 shows these matrices for the different ranking methods.

We can now present the results in Table B3 based on Equations (17) and (18). As shown in Table B3, the third method (M3) produces the lowest rank deviation relative to the other three ranking methods, and therefore, we should choose this method over the other three potential ranking methods.

Appendix C

Peoples Bank data

References

Aguarón, J., Escobar, M. T., & Moreno-Jiménez, J. M. (2020). Reducing inconsistency measured by the geometric consistency index in the analytic hierarchy process. *European Journal of Operational Research, 288*(2), 576–583.
Ahmadi, H. B., Kusi-Sarpong, S., & Rezaei, J. (2017). Assessing the social sustainability of supply chains using Best Worst Method. *Resources, Conservation and Recycling, 126*, 99–106.
Akbarian, D. (2020). A new DEA ranking system based on interval cross efficiency and interval analytic hierarchy process methods. *International Journal of Management and Decision Making, 19*(3), 344–363.
Akyuz, G., Tosun, O., & Aka, S. (2018). Multi criteria decision-making approach for evaluation of supplier performance with MACBETH method. *International Journal of Information and Decision Sciences, 10*(3), 249–262.
Alirezayee, M. R., & Rafiee Sani, M. R. (2010). A development on AHP/DEA methodology for ranking decision making units. *Industrial Management Journal, 2(2): 5*, 83–184.
Alirezayee, M. R., Rakhshan, S. A., & Rafiee Sani, M. R. (2012). Cross AHP/DEA method for ranking Decision Making Units. *Industrial Management Journal, 4*(1), 51–68.
An, Q., Meng, F., & Xiong, B. (2018). Interval cross efficiency for fully ranking decision making units using DEA/AHP approach. *Annals of Operations Research, 271*(2), 297–317.
Andersen, P., & Petersen, N. C. (1993). A procedure for ranking efficient units in data envelopment analysis. *Management Science, 39*(10), 1261–1294.
Anderson, T. R., Hollingsworth, K. B., & Inman, L. B. (2002). The fixed weighting nature of a cross-evaluation model. *Journal of Productivity Analysis, 18*(1), 249–255.
Azadeh, A., Ghaderi, S. F., Mirjalili, M., & Moghaddam, M. (2011). Integration of analytic hierarchy process and data envelopment analysis for assessment and optimization of personnel productivity in a large industrial bank. *Expert Systems with Applications, 38*(5), 5212–5225.
Bana e Costa, C. A., Correa, E. C., De Corte, J.-M., & Vansnick, J.-C. (2002). Facilitating bid evaluation in public call for tenders: A socio-technical approach. *Omega, 30*(3), 227–242.
Bana e Costa, C. A., De Corte, J.-M., & Vansnick, J.-C. (2012). Macbeth. *International Journal of Information Technology & Decision Making, 11*(2), 359–387.
Bana e Costa, C. A., & Vansnick, J. C. (1994). MACBETH- an interactive path towards the construction of cardinal value functions. *International Transactions in Operational Research, 1*, 489–500.
Bana e Costa C.A., & Vansnick J.C. (1999). The MACBETH Approach: Basic Ideas, Software, and an Application. In: Meskens N., Roubens M. (eds) Advances in Decision Analysis. Mathematical Modelling: Theory and Applications, 4. Springer, Dordrecht. doi: 10.1007/978-94-017-0647-6_9.
Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models estimating technical and scale in efficiencies in data envelopment analysis. *Management Science, 30*(9), 1078–1092.
Bardhan, I., Bowlin, W. F., Cooper, W. W., & Sueyoshi, T. (1996). Models for efficiency dominance in data envelopment analysis. Part I: Additive models and MED measures. *Journal of the Operations Research Society of Japan, 39*, 322–332.
Beck, M. P., & Lin, B. W. (1983). Some heuristics for the consensus ranking problem. *Computers & Operations Research, 10*(1), 1–7.
Bouyssou, D. (1999). Using DEA as a tool for MCDM: Some remarks. *Journal of the operational Research Society, 50*(9), 974–978.
Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*(6), 429–444.
Chen, Z., & Ming, X. (2020). A rough–fuzzy approach integrating best–worst method and data envelopment analysis to multi-criteria selection of smart product service module. *Applied Soft Computing, 94*, Article 106479. https://doi.org/10.1016/j.asoc.2020.106479
Davtalab-Olyaie, M. (2018). A secondary goal in DEA cross-efficiency evaluation: A "one home run is much better than two doubles" criterion. *Journal of the Operational Research Society, 70*(5), 807–816.
Delice, E. K., & Can, G. F. (2020). A new approach for ergonomic risk assessment integrating KEMIRA, best–worst and MCDM methods. *Soft Computing, 24*(19), 15093–15110.
Doyle, J., & Green, R. (1994). Efficiency and cross efficiency in DEA: Derivations, meanings and the uses. *Journal of the Operational Research Society, 45*(5), 567–578.
Doyle, J. R., & Green, R. H. (1995). Cross-evaluation in DEA: Improving discrimination among DMUs. *INFOR, 33*, 205–222.
Dutta, P., Jaikumar, B., & Arora, M. S. (2022). Applications of data envelopment analysis in supplier selection between 2000 and 2020: A literature review. *Annals of Operations Research, 315*(2), 1399–1454.
Ehsanifar, M. (2014). DEA/AHP and its application in Full Ranking of Decision Making Units. *International Journal of Business Quantitative Economics and Applied Management Research, 1*(4), 49–59.
Ertugrul, I., & Qztas, G. Z. (2016). Performance analysis of online bookstores by using MACBETH and PROMETHEE methods. *Journal of Internet Applications & Management (İnternet Uygulamaları ve Yonetimi Dergisi), 7*(2), 21–38.
Fan, J. P., Guo, Z. W., & Wu, M. Q. (2020). An Improvement of DEA Cross-Efficiency Aggregation Based on BWM-TOPSIS. *IEEE Access, 8*, 37334–37342.
Farrell, M. J. (1957). The measurement of productivity efficiency. *Journal of the Royal Statistical Society, Series A: General, 120*(3), 253–281.
Farzipoor Saen, R., Memariani, A., & Hosseinzadeh Lotfi, F. (2005). Determining relative efficiency of slightly non-homogeneous decision making units by data envelopment analysis: A case study in IROST. *Applied Mathematics and Computation, 165*, 313–328.
Friedman, L., & Sinuany-Stern, Z. (1997). Scaling units via the canonical correlation analysis and the data envelopment analysis. *European Journal of Operational Research, 100*(3), 629–637.
Golany, B. (1988). An interactive MOLP procedure for the extension of data envelopment analysis to effectiveness analysis. *Journal of the Operational Research Society, 39*(8), 725–734.
Hadi-Vencheh, A., & Niazi-Motlagh, M. (2011). An improved voting analytic hierarchy process–data envelopment analysis methodology for suppliers selection. *International Journal of Computer Integrated Manufacturing, 24*(3), 189–197.
Hashimoto, A. (1997). A ranked voting system using a DEA/AR exclusion model: A note. *European Journal of Operational Research, 97*(3), 600–604.
Hatami-Marbini, A., Toloo, M., Amini, M. R., & Azar, A. (2022). Extending a fuzzy network data envelopment analysis model to measure maturity levels of a performance based-budgeting system: A case study. *Expert Systems with Applications, 200*, Article 116884.
Ho, W. (2008). Integrated analytic hierarchy process and its applications–A literature review. *European Journal of operational research, 186*(1), 211–228.
Hosseinzadeh Lotfi, F., Jahanshahloo, G. R., Soltanifar, M., Ebrahimnejad, A., & Mansourzadeh, S. M. (2010a). Relationship between MOLP and DEA based on output-orientated CCR dual model. *Expert Systems with Applications, 37*(6), 4331–4336.
Hosseinzadeh Lotfi, F., Jahanshahloo, G. R., Ebrahimnejad, A., Soltanifar, M., & Mansourzadeh, S. M. (2010b). Target setting in the general combined-oriented CCR model using an interactive MOLP method. *Journal of Computational and Applied Mathematics, 234*(1), 1–9.
Ishizaka, A., & Siraj, S. (2018). Are multi-criteria decision-making tools useful? An experimental comparative study of three methods. *European Journal of Operational Research, 264*(2), 462–471.
Joblonsky, J. (2007). Measuring the efficiency of production units by AHP models. *Mathematical and Computer Modelling, 46*, 1091–1098.
Karande, P., & Chakraborty, S. (2013a). Evaluation and selection of flexible manufacturing systems using MACBETH method. *International Journal of Services and Operations Management, 16*(1), 123–144.
Karande, P., & Chakraborty, S. (2013b). Using MACBETH method for supplier selection in manufacturing environment. *International Journal of Industrial Engineering Computations, 4*(2), 259–279.
Kengpol, A., & Tuominen, M. (2006). A framework for group decision support systems: An application in the evaluation of information technology for logistics firms. *International Journal of Production Economics, 101*(1), 159–171.
Khalili-Damghani, K., Tavana, M., & Santos-Arteaga, F. J. (2016). A comprehensive fuzzy DEA model for emerging market assessment and selection decisions. *Applied Soft Computing, 38*, 676–702.
Kim, T.S. (2000). Extended Topics in the Integration of Data Envelopment Analysis and the Analytic Hierarchy Process in Decision Making, Ph.D. Thesis, Agricultural and Mechanical College, Louisiana State University, Baton Rouge, LA.
Kundakcı, N. (2019). An integrated method using MACBETH and EDAS methods for evaluating steam boiler alternatives. *Journal of Multi-Criteria Decision Analysis, 26*(1–2), 27–34.
Labijak-Kowalska, A., & Kadziński, M. (2021). Experimental comparison of results provided by ranking methods in Data Envelopment Analysis. *Expert Systems with Applications, 173*, Article 114739.
Leal, J. E. (2020). AHP-express: A simplified version of the analytical hierarchy process method. *MethodsX, 7*, Article 100748.
Li, X.-B., & Reeves, G. R. (1999). Multiple criteria approach to data envelopment analysis. *European Journal of Operational Research, 115*, 507–517.
Liang, F., Brunelli, M., & Rezaei, J. (2020). Consistency issues in the best worst method: Measurements and thresholds. *Omega, 96*, Article 102175.
Lin, M. I., Lee, Y. D., & Ho, T. N. (2011). Applying integrated DEA/AHP to evaluate the economic performance of local governments in China. *European Journal of Operational Research, 209*, 129–140.
Liu, F.-H.-F., & Hai, H. L. (2005). The voting analytic hierarchy process method for selecting supplier. *International Journal of Production Economics, 97*(3), 308–317.
Lozano, S., & Villa, G. (2009). Multiobjective target setting in data envelopment analysis using AHP. *Computers & Operations Research, 36*, 549–564.
Mehrabian, S., Alirezaee, M. R., & Jahanshahloo, G. R. (1999). A complete efficiency ranking of decision-making units in data envelopment analysis. *Computational Optimization and Applications, 14*, 261–266.
Momeni, M. (2016). *New Topics In Operations Research*. Momeni Press.
Omrani, H., Alizadeh, A., & Naghizadeh, F. (2020). Incorporating decision makers' preferences into DEA and common weight DEA models based on the best–worst method (BWM). *Soft Computing, 24*(6), 3989–4002.
Opricovic, S., & Tzeng, G. H. (2003). Comparing DEA and MCDM method. In T. Tanino, T. Tanaka, & M. Inuiguchi (Eds.), *Multi-Objective Programming and Goal Programming. Advances in Soft Computing* (vol 21, pp. 227–232). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-36510-5_32.
Pakkar, M. (2015). An integrated approach based on DEA and AHP. *Computational Management Science, 12*(1), 153–169.
Pakkar, M. S. (2016). Using DEA and AHP for hierarchical structures of data. *Industrial Engineering & Management Systems, 15*(1), 49–62.
Ramanathan, R. (2006). Data envelopment analysis for weight derivation and aggregation in the analytic hierarchy process. *Computers and Operations Research, 33*, 1289–1307.
Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega, 53*, 49–57.
Rezaei, J. (2016). Best-worst multi-criteria decision-making method: Some properties and a linear model. *Omega, 64*, 126–130.

Rezaei, J., Nispeling, T., Sarkis, J., & Tavasszy, L. (2016). A supplier selection life cycle approach integrating traditional and environmental criteria using the best worst method. *Journal of Cleaner Production, 135*, 577–588.

Rezaei, J., Wang, J., & Tavasszy, L. (2015). Linking supplier development to supplier segmentation using Best Worst Method. *Expert Systems with Applications, 42*(23), 9152–9164.

Rezaeitaziani, T., & Barkhordariahmadi, M. (2015). A two-stage model for ranking DMUs using DEA/AHP. *International Journal of Industrial Mathematics, 7*(2), 161–169.

Saaty, T. L. (1980). *The Analytic Hierarchy Process*. New York: Mcgraw Hill.

Sarkis, J. (2000). A comparative analysis of DEA as a discrete alternative multiple criteria decision tool. *European Journal of Operational Research, 123*(3), 543–557.

Sevkli, M., Koh, L. S. C., Zaim, S., Demirbag, M., & Tatoglu, E. (2007). An application of data envelopment analytic hierarchy process for supplier selection: A case study of BEKO in Turkey. *International Journal of Production Research, 45*(9), 1973–2003.

Sexton, T. R., Silkman, R. H., & Hogan, A. J. (1986). Data envelopment analysis: Critique and extensions. *New Directions for Program Evaluation, 1986*(32), 73–105.

Shang, J., & Sueyoshi, T. (1995). A Unified Framework for the Selection of a Flexible Manufacturing System. *European Journal of Operational Research, 85*(2), 297–315.

Sinuany-Stern, Z., & Friedman, L. (1998). Data envelopment analysis and the discriminant analysis of ratios for ranking units. *European Journal of Operational Research, 111*, 470–478.

Sinuany-Stern, Z., Mehrez, A., & Barboy, A. (1994). Academic departments efficiency via data envelopment analysis. *Computers and Operations Research, 21*(5), 543–556.

Sinuany-Stern, Z., Mehrez, A., & Hadad, Y. (2000). An AHP/DEA methodology for ranking decision making units. *International Transactions in Operational Research, 7*(2), 109–124.

Soltanifar, M., & Lotfi, F. H. (2011). The voting analytic hierarchy process method for discriminating among efficient decision making units in data envelopment analysis. *Computers & Industrial Engineering, 60*(4), 585–592.

Soltanifar, M., & Shahghobadi, S. (2013). Selecting a benevolent secondary goal model in data envelopment analysis cross-efficiency evaluation by a voting model. *Socio-Economic Planning Sciences, 47*(1), 65–74.

Starcevic, S., Bojovic, N., Junevicius, R., & Skarickij, V. (2019). Analytical hierarchy process method and data envelopment analysis application in terrain vehicle selection. *Transport, 34*(5), 600–616.

Sueyoshi, T. (1999). Data envelopment analysis non-parametric ranking test and index measurement: Slack-adjusted DEA and an application to Japanese agriculture cooperatives. *Omega, 27*(3), 315–326.

Takamura, Y., & Tone, K. (2003). A comparative site evaluation study for relocating Japanese government agencies out of Tokyo. *Socio-Economic Planning Sciences, 37*, 85–102.

Tavana, M., Soltanifar, M., & Santos-Arteaga, F. J. (2021). Analytical hierarchy process: Revolution and evolution. *Annals of Operations Research, 1–29*. https://doi.org/10.1007/s10479-021-04432-2

Tone, K. (2002). A slacks-based measure of super-efficiency in data envelopment analysis. *European Journal of Operational Research, 143*(1), 32–41.

Torgersen, A. M., Forsund, F. R., & Kittelsen, S. A. C. (1996). Slack-adjusted efficiency measures and ranking of efficient units. *Journal of Productivity Analysis, 7*, 379–398.

Wang, Y.-M., Liu, J., & Elhag, T. M. S. (2008). An integrated AHP–DEA methodology for bridge risk assessment. *Computers & Industrial Engineering, 54*(3), 513–525.

Wang, Y., Wu, D., Qian, W., & Li, H. (2021). Cross-efficiency intervals integrated ranking approach based on the generalized Fermat-Torricelli point. *Computers & Industrial Engineering, 162*, Article 107786.

Wu, D., Wang, Y., Liu, Y., & Wu, J. (2021). DEA cross-efficiency ranking method considering satisfaction and consensus degree. *International Transactions in Operational Research, 28*(6), 3470–3492.

Yager, R. R., & Alajlan, N. (2014). A note on mean absolute deviation. *Information Sciences, 279*, 632–641.

Yilmaz, M. K., Kusakci, A. O., Aksoy, M., & Hacioglu, U. (2022). The evaluation of operational efficiencies of Turkish airports: An integrated spherical fuzzy AHP/DEA approach. *Applied Soft Computing, 119*, Article 108620.

Zerafat Angiz, M., Mustafa, A., & Kamali, M. J. (2013). Cross-ranking of Decision Making Units in Data Envelopment Analysis. *Applied Mathematical Modelling, 37*, 398–405.