Systemic Analytics

www.sa-journal.org

Syst. Anal. Vol. 2, No. 2 (2024) 229–242.

Paper Type: Original Article

A DEA based Performance Measurement Approach with Weak Ordinal Data

Bohlool Ebrahimi[1,*], Dusko Tesic[2]

[1] Department of Industrial Engineering, ACECR, Sharif Unit, Tehran, Iran; b.ebrahimi@aut.ac.ir.
[2] University of Defence in Belgrade, Belgrade, Serbia; tesic.dusko@yahoo.com.

Citation:

Received: 16 February 2024 *Revised: 16 May 2024* *Accepted: 08 September 2024*	Ebrahimi, B., Tesic, D. (2024). A DEA based performance measurement approach with weak ordinal data. *Systemic Analytics, 2* (2), 229-242.

Abstract

Data Envelopment Analysis (DEA) is a mathematical programming for performance evaluation of a set of similar Decision Making Units (DMUs). In DEA model, it is supposed that the values of inputs and outputs are exactly known. But, in many real world problems these values are imprecise in form of ordinal, bounded data, and so on. Until now, different approaches have been proposed to calculate the relative efficiency in presence of ordinal data in DEA. The focus of this paper is on weak ordinal data. The paper briefly reviews the existing methods in this area and explains some drawbacks of these methods. We show that converting ordinal data to some special exact data and ignoring the DEA axioms lead to these drawbacks. In fact, when data are in ordinal format, there are no observed data, and so the inclusion of observation axiom, the first axiom in DEA, is not established. To overcome the drawbacks and because of the necessity of observation axiom, we propose a new algorithm based on generating n random dataset for the ordinal measures such that the relations among the ordinal data will be satisfied. By considering the inclusion of observation axiom, it will be shown that this algorithm leads to the better result comparing with existing approaches. Several numerical examples are used to explain the content of the paper.

Keywords: Data envelopment analysis, Efficiency measure, Imprecise data, Ordinal data.

1 | Introduction

Charnes et al. [1] proposed Data Envelopment Analysis (DEA) model, for the first time, for performance evaluation of several similar Decision Making Units (DMUs) that use multiple inputs to produce multiple outputs. In this model it is supposed that the values of inputs and outputs are exactly known. However, the exact values of inputs and outputs may not be available or cannot be exactly measurable in many real applications. So, these values are imprecise that includes ordinal data (weak and strong), bounded (interval) data, ratio bound data, and so on.

Corresponding Author: b.ebrahimi@aut.ac.ir
https://doi.org/10.31181/sa22202425

For the first time, Cooper et al. [2] used the bounded and weak ordinal data in DEA and named the new model as Imprecise DEA (IDEA), which was a nonlinear and non-convex model. They used the unit-invariant property of DEA and converted the nonlinear model into an equivalent linear model through the scale transformation and variable alterations. Their method has three shortcomings\problems:

I. The high volume of calculations.

II. Necessity of an exact maximum value for scale transformation in interval data.

III. Only the upper bound efficiencies are calculated and the lower bound efficiencies are not considered.

Cooper et al. [3] removed the second problem by introducing some dummy variables. Cooper et al. [4] applied the IDEA approach to evaluate the Korean Mobile Telecommunication Company. Kim et al. [5] used IDEA for performance evaluation in Telephone offices. Lee et al. [6] extended the IDEA concept to the additive DEA model. They used a simple variable alteration to convert the nonlinear IDEA model into an equivalent linear model.

Despotis and Smirlis [7] proposed a method to calculate the lower bound and upper bound of efficiency scores through an appropriate variable alteration. They developed two linear programming to estimate the lower and upper bound efficiencies by considering the pessimistic and optimistic state for each DMU. In fact, the efficiency score for each DMU is an interval in their method. They categorized DMUs in three groups, efficient (the lower bound efficiency score is equal to one), weak efficient (the upper bound efficiency score is equal to one, but the lower bound is less than one) and inefficient (the upper bound efficiency score is less than one).

Zhu [8] showed that the scale transformation (normalization of data) in [3], [4] method is redundant. He used a simple variable alteration to convert the nonlinear and non-convex IDEA model into an equivalent linear model. Next, he converted the bounded data into the exact data and showed that the efficiency score for the exact data is equivalent with the result of solving the nonlinear IDEA model. He converted the weak ordinal data into the bounded data to calculate the relative efficiency with this type of data. The proposed method by Zhu [8] eliminated the volume of calculations of [2] method. Zhu [9] used the method for performance evaluation in Korean Mobile Telecommunication Company. Park [10] reduced the volume of calculations of [3] method by using a simple variable alteration, which was the same as the variable alteration proposed by Zhu [8].

Wang et al. [11] showed that [7] used different production frontiers to calculate the efficiencies. They claimed that a unique production frontier should be used to evaluate all of the DMUs. This frontier attains by considering all DMUs in the best situation. Then, they proposed two linear mathematical programming to obtain the lower bound and upper bound efficiencies (in presence of interval data) by considering a unique (fixed) production frontier for all DMUs. For calculating the efficiency score with ordinal data, they converted this data into interval data. Finally, they proposed a minimax regret-based method to rank interval efficiencies.

Kao [12] emphasized that the efficiency scores should be imprecise in presence of imprecise data. He proposed two mathematical programming to calculate the lower bound and upper bound efficiencies in presence of ordinal and bounded data. His method uses different production frontier for each DMU similar to the [7] method. To consider the ordinal data, he set a lower bound and upper bound for each ordinal data after normalization.

Park [13] used the concept of supremum and infimum and proposed a mathematical programming for calculating the lower bound efficiencies. He also used different production frontiers to calculate the efficiencies. After calculating the efficiencies, he categorized DMUs in three groups: perfectly efficient, potentially efficient and inefficient, similar to the [7].

Kao and Liu [14] argued that if the interval data is wide, then the interval efficiencies obtained by using previous methods is too wide to provide valuable information for a Decision Maker (DM) to make an accurate (good) decision. Therefore, they considered interval data as stochastic data and estimated the distribution of

efficiency score for each DMU by using a simulation method. They showed that regarding bounded data as stochastic data gives more helpful and reliable results than the available interval data approaches. The method was used to obtain the distribution of efficiencies in Taiwan commercial banks to rank them.

Park [15] investigated the dual model of IDEA and its relationships with primal problem based on the duality theory in IDEA, and developed a computational method for it. Marbini et al. [16] investigated the performance evaluation in presence of interval data without sign restrictions. They calculated the lower and upper bounds efficiencies and categorized DMUs in three groups: strictly efficient, weakly efficient and inefficient. Chen et al. [17] presented some models to deal with Likert scale data, discrete and bounded data in DEA. They have used the developed DEA models to evaluate the regional energy efficiency in China.

It should be noted that the IDEA has been used in many real world problems. For example, Asosheh et al. [18] developed an integrated IDEA model for evaluating and ranking Information Technology (IT) projects. They used Balanced Scorecard (BSC) to define the IT projects evaluation criteria, and the integrated IDEA model to obtain most efficient IT project.

Toloo and Nalchigar [19] proposed an integrated IDEA model to find the best supplier in supplier selection problem. They used Zhu [8] approach to consider imprecise data. Karsak and Dursun [20] developed a supplier selection methodology by incorporating Quality Function Deployment (QFD) and DEA in presence of imprecise data. Khalili-Damghani et al. [21] used DEA model to evaluate the performance of combined cycle power plant in the presence of undesirable outputs and uncertain data. They modeled the uncertain data with interval data.

The current study shows that most of existing approaches to calculate the relative efficiency in presence of ordinal data have some drawbacks as follows:

I. Replacing ordinal data with some fixed integer numbers such as zero and one that leads to incorrect efficiency scores. It should be noted that the probability of occurrence of these fixed integer numbers is zero in practice.

II. Ignoring the DEA axioms, so in some cases we are unable to construct the Production Possibility Set (PPS), and so unable to calculate the relative efficiency.

To remove the drawbacks, we will show that treating the ordinal data as stochastic data gives more reasonable results.

The remainder of this paper is organized as follows: Section 2, explains the drawbacks of some important existing methods for ordinal data. Section 3, proposes a new approach to calculate the relative efficiency with ordinal data. Numerical example and conclusion are given in Sections 4 and 5, respectively.

2 | Major Drawbacks of some Proposed Approaches for Ordinal Data

In this section, we explain some major proposed approaches to calculate the relative efficiency in presence of ordinal data in DEA with their drawbacks. It should be noted that, in ordinal data we have only one special relation among the data, and their actual values are unknown. This type of data can be expressed as follows:

$$\begin{aligned} &x_{i1} \le x_{i2} \le \ldots \le x_{in}, \\ &y_{r1} \le y_{r2} \le \ldots \le y_{rn}. \end{aligned} \tag{1}$$

Now consider the CCR model with imprecise data:

$$\begin{aligned} &\max \ \sum_{r=1}^{m} u_r y_{rp}, \\ &\text{s.t.} \\ &\quad \sum_{i=1}^{n} v_i x_{ip} = 1, \end{aligned} \tag{2}$$

$$\sum_{r=1}^{m} u_r y_{rj} - \sum_{i=1}^{n} v_i x_{ij} \le 0 \text{ , } j = 1,...,k,$$
$$x_{ij} \in \theta_i^+, \; y_{rj} \in \theta_r^-,$$
$$u_r, v_i \ge 0.$$

which $x_{ij} \in \theta_i^+$ and $y_{rj} \in \theta_r^-$ represent subset or all of the imprecise data (ordinal or bounded data). Obviously, *Model (2)* is nonlinear and non-convex. As explained in the previous section, different approaches have been developed to solve the model. The drawbacks of some these approaches are explained as follows.

2.1|The Zhu [8, 9] Method

Zhu [8] showed that *Model (2)* can be converted into an equivalent linear model with the following simple variable alterations.

$$\begin{aligned} X_{ij} &= v_i x_{ij}, \quad \text{For all i, j,} \\ Y_{rj} &= u_r y_{rj}, \quad \text{For all r, j.} \end{aligned} \tag{3}$$

He also demonstrated that interval data can be replaced with some exact data. In the other words, when DMU_p is under evaluation, due to maximizing the efficiency of DMU_p, the upper bounds of outputs and lower bounds of inputs are used for DMU_p and for other DMUs, the upper bounds of inputs and lower bounds of outputs are considered. Therefore, [8, 9] ranked DMUs based on their upper bound efficiencies. Some existing methods such as [2–4], [10], also ranked DMUs based on upper bound efficiencies that is not sufficient. To clarify the topic, consider the following example.

Example 1. suppose there are two DMUs with ordinal input and output as presented in *Table 1*. Using the [8] approach implies that the both of these DMUs are efficient. It can be easily seen that DMU_2 dominates DMU_1 and only in one special situation, when $x_{11} = x_{12}$ and $y_{12} = y_{11}$, that occur with zero probability, these DMUs are both efficient.

Table 1. data for 2 DMUs.

DMU No.	Input 1*	Output 1*
1	x_{11}	y_{11}
2	x_{12}	y_{12}

*which $x_{11} \ge x_{12}$ and $y_{12} \ge y_{11}$

2.1.1. Converting ordinal data into exact data

Suppose DMU_p is under evaluation and *Model (2)* has been solved and the following optimal solution for the ordinal data has been obtained.

$$\begin{aligned} &x_{i1}^* \le x_{i2}^* \le ... \le x_{i,p-1}^* \le x_{ip}^* \le x_{i,p+1}^* \le ... \le x_{in}^*, \\ &y_{r1}^* \le y_{r2}^* \le ... \le y_{r,p-1}^* \le y_{rp}^* \le y_{r,p+1}^* \le ... \le y_{rn}^*. \end{aligned} \tag{4}$$

In this condition, by considering the unit-invariant property of DEA, $\rho_i x_{ij}^*$ and $\rho_r y_{rj}^*$ are also optimal solutions $(\rho_i, \rho_r > 0, \text{for all } i, r)$. Thus, we can assume $y_{rp}^* = x_{ip}^* = 1$, and so the optimal solution can be expressed as follows:

$$\begin{aligned} &0 \le x_{i1}^* \le x_{i2}^* \le ... \le x_{i,p-1}^* \le x_{ip}^* = 1 \le x_{i,p+1}^* \le ... \le x_{in}^* \le M, \\ &0 \le y_{r1}^* \le y_{r2}^* \le ... \le y_{r,p-1}^* \le y_{rp}^* = 1 \le y_{r,p+1}^* \le ... \le y_{rn}^* \le M. \end{aligned} \tag{5}$$

M is a positive large enough number [8, 9] proposed that M could be supposed the number of DMUs. By using the *Relation (5)*, [8, 9] converted the ordinal data into interval data as follows (DMU_p is under evaluation):

$$\begin{aligned} &x_{ij} \in [0,1] \ \& \ y_{rj} \in [0,1] \text{ for } DMU_j, \ \forall j \in \{1,2,...,p-1\}, \\ &x_{ij} \in [1,M] \ \& \ y_{rj} \in [1,M] \text{ for } DMU_j, \ \forall j \in \{p+1,...,k\}. \end{aligned} \quad (6)$$

Then, the interval data converted into the following exact data:

$$\begin{aligned} &x_{ij} = 1, \text{ for all } j \le p \ \ \& \ \ x_{ij} = M, \text{for all } j > p, \\ &y_{rj} = 0, \text{for all } j < p \ \ \& \ \ y_{rj} = 1, \text{for all } j \ge p. \end{aligned} \quad (7)$$

In the following example, it will be shown that the efficiency scores are dependent on the value of *M*. Also, it will be shown that the results of the two proposed methods by Zhu [8, 9] are different (the variable alteration method and converting imprecise data into exact data method).

Example 2. Consider three DMUs, each uses one input to produce one output.

Table 2. data for 3 DMUs.

Dmu No.	Input (Ordinal)*	Output (Exact)
1	x_{11}	2
2	x_{12}	3
3	x_{13}	12

* ranking such that $x_{11} \le x_{12} \le x_{13}$.

Using the variable *Alterations (3)* (the first proposed approach in [8, 9] shows that DMU_1 is efficient. Now consider the second approach, when DMU_1 is under evaluation we should use the following exact data for these three DMUs:

Table 3. The exact data for 3 DMUs based on [8, 9], approach.

DMU No.	Input (Ordinal)	Output (Exact)
1	1	2
2	M	3
3	M	12

The data presented in *Table 3* shows that for $M \ge 6$, DMU_1 is efficient and for $M < 6$ it is inefficient. Indeed, based on [9] approach, if we set $M = 3$ then DMU_1 will be inefficient, that is inconsistent with the previous result (the result of the variable alteration method). Also, the *Example (2)* shows that the efficiency score is dependent on the value of *M*.

The following theorem shows that a DMU with a best rank in an ordinal input or output will be always efficient.

Theorem 1. In *Model (2)*, suppose the d_{th} output of DMUs is in ordinal format and DMU_p has the best rank. In the other words, suppose $y_{dp} \ge y_{dj}, \text{for all } j$. In this case, DMU_p is always efficient (the upper bound efficiency score of DMU_p is equal to unity).

Proof: obviously, in the calculation of the relative efficiency score of DMU_p, *Model (2)* could select y_{dp} as a large enough positive number and set $y_{dj}, \text{for all } j \ne p$ as very small numbers such that DMU_p dominates all of the other DMUs. Note that a similar theorem can be presented for ordinal data in inputs. Also, a mathematical proof can be given as follows.

The max-min model to calculate the relative efficiency of DMU_p is as follows:

$$\text{Relative Efficiency of } DMU_p = \underset{u,v\geq\varepsilon}{Max}\left\{\frac{\sum_r u_r y_{rp}}{\sum_i v_i x_{ip}} \Big/ \underset{j}{Max}\left\{\frac{\sum_r u_r y_{rj}}{\sum_i v_i x_{ij}}\right\}\right\}. \quad (8)$$

Now we set $u_r = v_i = 1, \text{ for all } i, r$, thus we have:

$$\text{Relative Efficiency of } DMU_p = \underset{u,v\geq\varepsilon}{Max}\left\{\frac{\sum_r y_{rp}}{\sum_i x_{ip}} \Big/ \underset{j}{Max}\left\{\frac{\sum_r y_{rj}}{\sum_i x_{ij}}\right\}\right\}. \quad (9)$$

Now, it is enough to set $y_{dj} = 1, \text{ for all } j \neq p$ and $y_{dp} = M$, where M is a large positive number such that the following relation is established.

$$\frac{\sum_r y_{rp}}{\sum_i x_{ip}} > \frac{\sum_r y_{rj}}{\sum_i x_{ij}}, \text{ for all } j \neq p. \quad (10)$$

The *Relation (10)* shows that DMU_p is efficient. It should be noted that, according to the efficiency definition in the literature of DEA, DMU_p is efficient if and only if there exists at least a common set of weights $u^* > 0, v^* > 0$, such that $\frac{\sum_r u_r^* y_{rp}}{\sum_i v_i^* x_{ip}} \geq \frac{\sum_r u_r^* y_{rj}}{\sum_i v_i^* x_{ij}}$, for all j., this completes the proof.

Obviously, this result is correct in theory and may not be reasonable in practice.

Overall, the drawbacks of [8, 9] approach can be summarized as follows:

I. Ranking DMUs only based on upper bound efficiencies.

II. Efficiencies are dependent on the values of M.

III. Ignoring the lower bound efficiencies and so leads to wrong ranking (*Example 1*).

IV. Using only 0, 1 and M for all ordinal data and too many zero for ordinal outputs. The probability of occurrence of these data is zero in practice.

V. The result of applying variable alterations approach and converting ordinal data into exact data approach are not equivalent (*Example 2*).

2.2. The Wang et al. [11] Method

Wang et al. [11] used a fixed production frontier for all DMUs and proposed the following two mathematical programming to obtain the lower bound and upper bound of efficiency scores.

The upper bound of efficiency for DMU_p.

$$\begin{aligned} &\text{Max } \theta_p^U = \frac{\sum_{r=1}^{s} u_r y_{rp}^U}{\sum_{i=1}^{m} v_i x_{ip}^L}, \\ &\text{s.t.} \quad \theta_j^U = \frac{\sum_{r=1}^{s} u_r y_{rj}^U}{\sum_{i=1}^{m} v_i x_{ij}^L} \leq 1, \quad j = 1,2,...,n. \\ &\qquad u_r, v_i \geq \varepsilon \text{ for all } i, r. \end{aligned} \quad (11.a)$$

The lower bound of efficiency for DMU_p.

$$\begin{aligned} &\text{Max } \theta_p^L = \frac{\sum_{r=1}^{s} u_r y_{rp}^L}{\sum_{i=1}^{m} v_i x_{ip}^U}, \\ &\text{s.t.} \quad \theta_j^U = \frac{\sum_{r=1}^{s} u_r y_{rj}^U}{\sum_{i=1}^{m} v_i x_{ij}^L} \le 1, \qquad j = 1,2,...,n, \\ &\qquad u_r, v_i \ge \varepsilon \text{ for all } i,r. \end{aligned} \tag{11.b}$$

Obviously, these models can be converted into linear models. To use the models for ordinal data, at first, they obtained the following relation by applying scale transformation.

$$1 \ge \hat{y}_{r1} \ge \hat{y}_{r2} \ge ... \ge \hat{y}_{rn} \ge \sigma_r. \tag{12}$$

which σ_r is a small positive number, reflecting the ratio of the possible minimum of $\{y_{rj} \mid j = 1,2,...,n\}$ to its possible maximum that should be estimated by the DM. Now, ordinal data are converted to the interval data as follows:

$$\hat{y}_{rj} \in [\sigma_r, 1], \text{ for all } j. \tag{13}$$

In fact, in [11] method the ordinal data are regarded to be equal. Let the data presented in *Example 1*, and suppose that the value of σ_1 for input and output estimated by the DM, are 0.1 and 0.05, respectively. In this condition, [11] considered the following data for the evaluation of these two DMUs.

Table 4. converted data by Wang et al. [11] method for 2 DMUs.

DMU No.	Input 1	Output 1
1	[0.1 , 1]	[0.05 , 1]
2	[0.1 , 1]	[0.05 , 1]

Obviously, the two DMUs are the same always with these data, that is not reasonable. As explained in *Example 1*, DMU_2 dominates DMU_1. It seems that the results of [2], [8–10] methods are more reasonable. The reason is that, based on their methods, the upper bound efficiency score of DMU_1 is equal to the efficiency score of DMU_2, but in Wang et al. [11] method both DMUs are the same in all conditions.

2.3 | The Park [13] Method

Park [13] proposed the following mathematical programming to obtain the lower bound of efficiency score of DMU_p in presence of imprecise data.

$$\begin{aligned} &\max \sum_{r=1}^{m} u_r \inf\{y_{rp} \mid y_r \in D_r^+\}, \\ &\text{s.t.} \\ &\sum_{i=1}^{n} v_i \sup\{x_{ip} \mid x_i \in D_i^-\} = 1, \end{aligned} \tag{13}$$

$$\begin{aligned}
&\sum_{r=1}^{m} u_r \inf\{y_{rp} \,|\, y_r \in D_r^+\} - \sum_{i=1}^{n} v_i \sup\{x_{ip} \,|\, x_i \in D_i^-\} \le 0, \\
&\sum_{r=1}^{m} u_r \sup\{y_{rj} \,|\, y_r \in D_r^+\} - \sum_{i=1}^{n} v_i \inf\{x_{ij} \,|\, x_i \in D_i^-\} \le 0 \;, j = 1,...,k, j \ne p, \\
&u_r, v_i \ge \varepsilon, \text{ for all } r, i,
\end{aligned} \tag{14}$$

In this model D_i^- and D_r^+ represent the imprecise data. Inf and sup can be replaced with min and max, respectively. In the model the feasibility is not considered in calculation of the inf and sup. To demonstrate the problem, consider the numerical example used in [13]. In this example, there are 8 telephone offices with three inputs and three outputs. The third output is in ordinal format as follows:

$$D_3^- = \{x_3 \in R^8 \,|\, x_{34} \ge x_{35} \ge x_{33} \ge x_{37} \ge x_{31} \ge x_{36} \ge x_{32} \ge x_{38}\}. \tag{15}$$

According to the Park [13] method, after normalization we have:

$$D_3^{'-} = \{x_3^{'} \in R^8 \,|\, 1 \ge x_{34}^{'} \ge x_{35}^{'} \ge x_{33}^{'} \ge x_{37}^{'} \ge x_{31}^{'} \ge x_{36}^{'} \ge x_{32}^{'} \ge x_{38}^{'} \ge 0\}. \tag{16}$$

Now, when DMU_1 is under evaluation (calculating the lower bound of efficiency), the data for the DMU_1 and other DMUs can be calculated as follows:

$$\begin{aligned}
&\sup\{x_{31}^{'} \,|\, x_3^{'} \in D_3^{'-}\} = \max\{x_{31}^{'} \,|\, x_3^{'} \in D_3^{'-}\} = 1, \\
&\inf\{x_{3j}^{'} \,|\, x_3^{'} \in D_3^{'-}\} = \min\{x_{3j}^{'} \,|\, x_3^{'} \in D_3^{'-}\} = 0; \;\; j = 2,3,...,8.
\end{aligned} \tag{17}$$

Therefore, to calculate the lower bound efficiency score of DMU_1, [13] used $x_3^* = (1,0,0,0,0,0,0,0)$ for ordinal data that is an infeasible solution. As mentioned, the feasibility condition should be considered for calculation of inf and sup. Applying the feasibility situation implies that he should use $x_3^* = (1,0,1,1,1,0,1,0)$. Furthermore, [13] method uses only zero and one for all ordinal data such as [12] method. As mentioned before, the probability of occurrence of these data is zero in practice.

The next example shows that the lower bound efficiencies cannot be calculated by using the [13] method in some situations.

Example 3. Consider the data presented in *Table 2*. For this ordinal data, the [13] method uses only zero and one that has been presented in *Table 5*.

Table 5. The values of ordinal data in calculation of lower bound efficiency with [13] **method.**

The DMU Under Evaluation the Values of Variables	Without Considering The Feasibility			Considering The Feasibility		
	DMU_1	DMU_2	DMU_3	DMU_1	DMU_2	DMU_3
x_{11}^{*}	1	0	0	1	0	0
x_{12}^{*}	0	1	0	1	1	0
x_{13}^{*}	0	0	1	1	1	1
Lower bound efficiencies	Cannot be calculated	Cannot be calculated	Cannot be calculated	0.17	Cannot be calculated	Cannot be calculated

As it can be seen from *Table 5*, there are some DMUs that produce output without consuming any input. With this data, the efficiencies cannot be calculated, and so the ranking of DMUs is not possible.

3|The Proposed Method to Rank Dmus with Ordinal Data

In this section, the cause of the occurrence of the drawbacks (presented in the previous section) and a new ranking method is presented in presence of ordinal data.

The CCR model and PPS are based on some axioms. Suppose we have k DMUs (DMU_j, $j=1,2,3,...,k$) with n inputs ($x_{ij} \geq 0$, $i=1,2,3,...,n$; $j=1,2,3,...,k$) and m outputs (y_{rj}, $r=1,2,3,...,m$; $j=1,2,3,...,k$) such that at least one input and one output is nonzero for each DMU. The axioms are as follows:

I. Inclusion of observation: each observed DMUj belongs to T, (j=1,2,…,k). T is the PPS.

II. Free disposability of inputs and outputs: if $(x,y) \in T, y' \leq y,$ then $(x,y') \in T$ and if $(x,y) \in T, x' \geq x,$ then $(x',y) \in T$.

I. Convexity: if $(x,y)\&(x',y') \in T \Rightarrow \lambda(x,y)+(1-\lambda)(x',y') \in T,$ for all $0 \leq \lambda \leq 1$.

II. Constant returns to scale: if $(x,y) \in T \Rightarrow (\lambda x, \lambda y) \in T,$ for all $0 \geq \lambda$.

III. Minimum extrapolation: T is the intersection of all sets satisfying 1-4.

When the data are in ordinal or bounded format, in fact there is not any observed data and so the first axiom (Inclusion of observation) is not established. In this situation, we cannot build the PPS. To clarify the topic, consider the data presented in *Table 1*. For this data the PPS is all of the first quarter of the area has been shown in *Fig. 1*. It could be seen that the output can be produced without consuming any input. In the other words, $(0,y)$, for all $y \geq 0$, belongs to PPS, that it is not true. This fault shows that observing the DMUs data (first axiom) is necessary and the PPS could not be built correctly without the first axiom.

In all of the previous methods, this problem exists and all methods tried to construct the PPS without observed data. So, by using the unit-invariant property of DEA, the ordinal data replaced with some integer numbers (zero, one or M). Indeed, these methods have not considered the axioms of DEA.

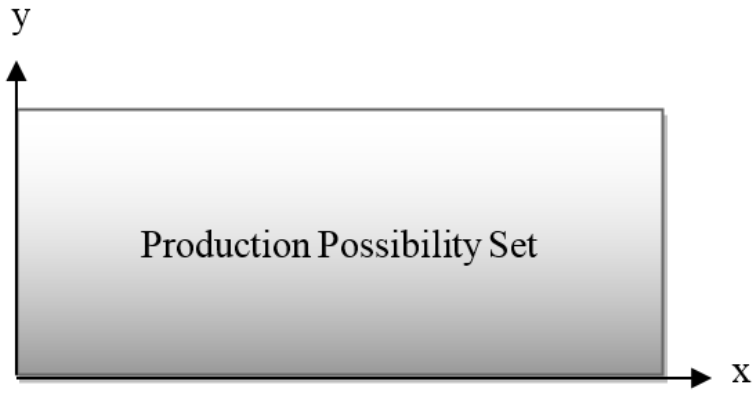

Fig. 1. PPS for the data of Example 1.

In the following, a new method is presented to rank DMUs in presence of ordinal data based on simulation (data generation) method. The values of ordinal data are unknown and only a relation is established among them. For considering the observation axiom, the ordinal data can be regarded as stochastic data. Because the probability distribution of this data is unknown, it can be supposed uniform. In this case the steps of proposed algorithm are as follows:

I. Generating data with the uniform distribution (by considering the relations among ordinal data) with n iterations.

II. Calculating efficiencies for all DMUs with the data obtained from Stage 1. In this stage *n* efficiency scores for each DMU will be obtained.

III. Calculating the average efficiency score and standard deviation of efficiencies. Indeed, the average efficiency is an estimation of the expected value of efficiency.

IV. Ranking DMUs based on average efficiency scores and standard deviation of efficiencies as follows: DMU_{j1} dominates DMU_{j2} if and only if the average efficiency score of DMU_{j1} is greater than DMU_{j2}. If two DMUs have the same average efficiency score, the DMU with less standard deviation has better rank.

The proposed method can be used for other types of imprecise data such as bounded data, ratio bound data, and so on. Obviously, increasing n leads to obtain the more exact average efficiency. If we were able to extract the efficiency distribution, we could calculate the expected value of relative efficiency for each DMU. But, extracting the efficiency distribution is very difficult [14]. In fact, as mentioned, the average efficiency is an estimation of expected relative efficiency, so these values will be approximately equal by increasing the value of n.

To show the revenue of this approach, consider the data presented in *Example 1*. The proposed approaches in Cooper et al. [2–4], [8–11], [13] and some other approaches, are unable to rank these DMUs correctly (to see the drawbacks of some these approaches for this simple example see Section 2). Now, we apply the proposed algorithm in this section for this example. The results are given in *Table 6* for the different values of n.

As it can be seen in *Table 6*, and also we expected, the DMU_2 is always efficient. Also, the DMU_1 is inefficient and its average efficiency score converges to 0.25.

Table 6. The results of proposed algorithm for the data presented in Table 1.

n	Average efficiency score	
	DMU_1	DMU_2
100	0.2686	1
1000	0.2543	1
10000	0.2506	1
100000	0.2505	1
1000000	0.2500	1

Remark 1. in the most of previous methods such as Park [13] and Despotis and Smirlis [7], DMUs are categorized into three groups after calculating the lower bound and upper bound efficiencies as follows:

I. First group: strong efficient, the lower bound of efficiency is equal to one.

II. Second group: potentially efficient, the upper bound of efficiency score is equal to one, but the lower bound efficiency is less than one.

III. Third group: inefficient, the upper bound of efficiency score is less than one.

Based on this categorization, a DMU_1 with interval efficiency of [0.1, 1] has better rank comparing with a DMU_2 with interval efficiency of [0.95, 0.99]. The DMU_1 is efficient in its best situation. The lower bound of its efficiency is 0.1, but the lower bound efficiency of DMU_2 is 0.95. If we suppose that the distribution of efficiencies is uniform, then the expected efficiencies for DMU_1 and DMU_2 will be 0.55 and 0.97, respectively. So, DMU_2 is more reliable and a wise DM prefers DMU_2 to DMU_1. This matter shows that considering the efficiencies distribution and expected efficiencies gives more reasonable results. It should be noted that the expected efficiency differs from the efficiency obtained from the expected values of data.

4 | Numerical Example

In this section, we apply the proposed approach in this paper and some existing approaches to efficiency measure of five DMUs studied in [2]. The DMUs have two inputs (one exact and one interval) and two outputs (one exact and one ordinal), as presented in *Table 7*. The interested reader can refer to Cooper et al. [2] for more information about this data.

Table 7. the values of inputs and outputs for 5 DMUs.

DMUs No.	Inputs		Outputs	
	x_{1j} (Exact)	x_{2j} (Interval)	y_{1j} (Exact)	y_{2j} (Ordinal)*
1	100	[0.6, 0.7]	2000	4
2	150	[0.8, 0.9]	1000	2
3	150	1	1200	5
4	200	[0.7, 0.8]	900	1
5	200	1	600	3

*ranking such that: $y_{23} \geq y_{21} \geq y_{25} \geq y_{22} \geq y_{24}$.

The results of different approaches for this data are summarized in *Table 8* with $\varepsilon = 10^{-6}$. As it can be seen in *Table 8* and *Table 10*, the DMUs have different efficiency scores, and therefore, different ranks.

The presented approaches in [2–4], [8] calculate only the upper bound efficiencies. So, based on these methods, DMU_1 and DMU_3 are efficient. By considering $M = 5$ in [8, 9] approach, the efficiency scores of five DMUs are less than or equal to the upper bound efficiencies. Applying the [11] method implies that DMU_4 has better rank comparing with DMU_3, unlike [2–4], [8], [9], [13], [12]. The ranking based on different methods are summarized in *Table 10*. As explained before, different approaches used different values for ordinal and interval data and also different concept to calculate the relative efficiency. The probability of the occurrence of these values is near zero. Therefore, different ranking are obtained for these 5 DMUs and the DM will be con [fused to rank these five DMUs.

Now, we apply the proposed algorithm in this paper to rank these five DMUs. We generated $n = 1, 10, 100, 1000, 5000$ and 10000 random data for ordinal and interval data and calculated the average efficiency score and standard deviation for these DMUs. The results summarized in *Table 9*. As it can be seen, by increasing the value of *n*, the variations of average efficiencies and standard deviations decrease. The variations are less than 0.01. The results show that generating 1000 or 5000 random data have proper runtime and can leads to reliable ranking. The final rank based on 10000 data generations are given in last column of *Table 10*. The ranking based on 1000 or 5000 data generations is the same as 10000 data generations. As it can be seen, the ranking is as follows: $DMU_1 > DMU_3 > DMU_2 > DMU_4 > DMU_5$.

Table 8. The efficiency score of 5 DMUs calculated from different methods.

DMUs No.	Efficiency Score					
	[2–4]	First Approach [8]	Second Approach *[8]	[11]	[12]	[13]
1	1	1	1	[0.99999, 1]	[1, 1]	[1, 1]
2	0.87499	0.87499	0.87397	[0.33333, 0.74899]	[0.66566, 0.87397]	[0.33333, 0.87499]
3	1	1	1	[0.39999, 0.66587]	[1, 1]	[0.4, 1]
4	0.99999	0.99999	0.99880	[0.33750, 0.85597]	[0.74885, 0.99880]	[0.33748, 0.99999]
5	0.69999	0.69999	0.69856	[0.17999, 0.59858]	[0.59858, 0.69856]	[0.17999, 0.69999]

*we set $M = 5$ based on [9].

*we suppose $\sigma_1 = 0.1$.

Table 9. The results of proposed algorithm for data presented in Table 7.

N	Run Time (S)*	Average Efficiency Score And Standard Deviation									
		DMU_1		DMU_2		DMU_3		DMU_4		DMU_5	
		A.E.S	S.D	A.E.S	S.D	A.E.S	S.D	A.E.S	S.D	A.E.S	S.D
1	0.03	1	0	0.4179	-	1	-	0.3635	-	0.4519	-
10	0.12	1	0	0.4014	0.0787	0.8922	0.1409	0.3837	0.0245	0.3478	0.1325
100	1.12	1	0	0.3918	0.0338	0.9351	0.1023	0.3939	0.0357	0.2846	0.1234
1000	10.84	1	0	0.3952	0.0478	0.9414	0.1014	0.3939	0.0305	0.2947	0.1253
5000	54.08	1	0	0.3944	0.0448	0.9403	0.1020	0.3937	0.0328	0.2945	0.1253
10000	109.22	1	0	0.3950	0.0495	0.9401	0.1034	0.3933	0.0417	0.2941	0.1236

*we used MATLAB 2014.

Table 10. Ranking of the DMUs based on different approaches.

Dmus No.	Ranking					
	[2–4]	[8]	[11]	[12]	[13]	Proposed Algorithm
1	1	1	1	1	1	1
2	3	3	3	3	3	3
3	1	1	4	1	2	2
4	2	2	2	2	3	4
5	4	4	5	4	3	5

5 | Conclusion

DEA has been used widely for performance evaluation of many real-world problems. The values of inputs and outputs are imprecise in most of these problems. For this purpose, many researchers have developed different approaches to deal with imprecise data in DEA. The focus of this paper is on weak ordinal data. The paper briefly reviewed existing approaches and studied some drawbacks such as infeasibility [13], incorrect results [11], different results for same data [8, 9] in presence of ordinal data. Most of these approaches allocated only zero and one for ordinal data and did not consider the DEA axioms. In practice, the probability of occurrence of this data is zero. It was emphasized that the DEA model and PPS are based on some axioms especially the inclusion of observation axiom. When data are in ordinal format, indeed, there is no observed data in hand, and so the first axiom is not established.

With this type of data, we may be unable to determine the PPS correctly, and so the production frontier is not exist in some cases (Section 3). Also in *Theorem 1*, we show that if a DMU has the best rank in an ordinal input or output, then it will be efficient always (it's upper bound of efficiency score is unity). This is not reasonable in practice, but it is correct in theory. To overcome the drawbacks and also to consider the DEA axioms, we proposed a new algorithm based on data generation to estimate the efficiency score in presence of ordinal data. We considered the average efficiency and standard deviation to rank DMUs. It was shown that this algorithm generates more realistic results, especially when all data are in ordinal format (Section 3). We also argued that ranking DMUs based on the lower and upper bound efficiencies is not reasonable. The presented algorithm can be used with the other types of imprecise data such as interval data, ratio bound data, and so on.

Author Contributions

Bohlool Ebrahimi developed the concept of using Data Envelopment Analysis (DEA) with weak ordinal data and performed the theoretical research. Dusko Tesic contributed to the methodology development and provided insights into the mathematical modeling and simulations. Both authors jointly wrote and reviewed the manuscript.

Funding

No specific funding was received for this study from any public, private, or non-profit funding agencies.

Data Availability

All data generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare no conflicts of interest related to this research or its publication.

References

[1] Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research*, *2*(6), 429–444. https://doi.org/10.1016/0377-2217(78)90138-8

[2] Cooper, W. W., Park, K. S., & Yu, G. (1999). IDEA and AR-IDEA: models for dealing with imprecise data in DEA. *Management science*, *45*(4), 597–607. https://pubsonline.informs.org/doi/abs/10.1287/mnsc.45.4.597

[3] Cooper, W. W., Park, K. S., & Yu, G. (2001). IDEA (imprecise data envelopment analysis) with CMDs (column maximum decision making units). *Journal of the operational research society*, *52*(2), 176–181. DOI:10.1057/palgrave.jors.2601070

[4] Cooper, W. W., Park, K. S., & Yu, G. (2001). An illustrative application of IDEA (imprecise data envelopment analysis) to a Korean mobile telecommunication company. *Operations research*, *49*(6), 807–820.

[5] Kim, S. H., Park, C. G., & Park, K. S. (1999). An application of data envelopment analysis in telephone officesevaluation with partial data. *Computers & operations research*, *26*(1), 59–72. https://doi.org/10.1016/S0305-0548(98)00041-0

[6] Lee, Y. K., Sam Park, K., & Kim, S. H. (2002). Identification of inefficiencies in an additive model based IDEA (imprecise data envelopment analysis). *Computers & operations research*, *29*(12), 1661–1676. https://www.sciencedirect.com/science/article/pii/S0305054801000491

[7] Despotis, D. K., & Smirlis, Y. G. (2002). Data envelopment analysis with imprecise data. *European journal of operational research*, *140*(1), 24–36. https://www.sciencedirect.com/science/article/pii/S0377221701002004

[8] Zhu, J. (2003). Imprecise data envelopment analysis (IDEA): a review and improvement with an application. *European journal of operational research*, *144*(3), 513–529. https://www.sciencedirect.com/science/article/pii/S0377221701003927

[9] Zhu, J. (2004). Imprecise DEA via standard linear DEA models with a revisit to a Korean mobile telecommunication company. *Operations research*, *52*(2), 323–329. https://pubsonline.informs.org/doi/abs/10.1287/opre.1030.0072

[10] Park, K. S. (2004). Simplification of the transformations and redundancy of assurance regions in IDEA (imprecise DEA). *Journal of the operational research society*, *55*(12), 1363–1366.

[11] Wang, Y. M., Greatbanks, R., & Yang, J. B. (2005). Interval efficiency assessment using data envelopment analysis. *Fuzzy sets and systems*, *153*(3), 347–370. https://www.sciencedirect.com/science/article/pii/S0165011404005767

[12] Kao, C. (2006). Interval efficiency measures in data envelopment analysis with imprecise data. *European journal of operational research*, *174*(2), 1087–1099. https://www.sciencedirect.com/science/article/pii/S037722170500319X

[13] Park, K. S. (2007). Efficiency bounds and efficiency classifications in DEA with imprecise data. *Journal of the operational research society*, *58*(4), 533–540. https://www.tandfonline.com/doi/abs/10.1057/palgrave.jors.2602178

[14] Kao, C., & Liu, S.-T. (2009). Stochastic data envelopment analysis in measuring the efficiency of Taiwan commercial banks. *European journal of operational research*, *196*(1), 312–322. https://www.sciencedirect.com/science/article/pii/S0377221708002270

[15] Park, K. S. (2010). Duality, efficiency computations and interpretations in imprecise DEA. *European journal of operational research*, *200*(1), 289–296. https://www.sciencedirect.com/science/article/pii/S0377221708010308

[16] Hatami-Marbini, A., Emrouznejad, A., & Agrell, P. J. (2014). Interval data without sign restrictions in DEA. *Applied mathematical modelling*, *38*(7), 2028–2036. https://www.sciencedirect.com/science/article/pii/S0307904X13006392

[17] Chen, Y., Cook, W. D., Du, J., Hu, H., & Zhu, J. (2017). Bounded and discrete data and Likert scales in data envelopment analysis: application to regional energy efficiency in China. *Annals of operations research*, *255*, 347–366. https://link.springer.com/article/10.1007/s10479-015-1827-3

[18] Asosheh, A., Nalchigar, S., & Jamporazmey, M. (2010). Information technology project evaluation: an integrated data envelopment analysis and balanced scorecard approach. *Expert systems with applications*, *37*(8), 5931–5938. https://www.sciencedirect.com/science/article/pii/S0957417410000515

[19] Toloo, M., & Nalchigar, S. (2011). A new DEA method for supplier selection in presence of both cardinal and ordinal data. *Expert systems with applications*, *38*(12), 14726–14731. https://www.sciencedirect.com/science/article/pii/S0957417411007913

[20] Karsak, E. E., & Dursun, M. (2014). An integrated supplier selection methodology incorporating QFD and DEA with imprecise data. *Expert systems with applications*, *41*(16), 6995–7004. https://www.sciencedirect.com/science/article/pii/S0957417414003613

[21] Khalili-Damghani, K., Tavana, M., & Haji-Saami, E. (2015). A data envelopment analysis model with interval data and undesirable output for combined cycle power plant performance assessment. *Expert systems with applications*, *42*(2), 760–773. https://www.sciencedirect.com/science/article/pii/S0957417414005107