Efficiency distribution and expected efficiencies in DEA with imprecise data

Bohlool Ebrahimi

1Satellite Research Institute, Iranian Space Research Center, Tehran, Iran
b_ebrahimi@jdsharif.ac.ir

Abstract
Several methods have been proposed for ranking the decision making units (DMUs) in data envelopment analysis (DEA) with imprecise data. Some methods have only used the upper bound efficiencies to rank DMUs. However, some other methods have considered the both of the lower and upper bound efficiencies to rank DMUs. The current paper shows that these methods did not consider the DEA axioms and may be unable to produce a rational ranking. We show that considering the imprecise data as stochastic and using the expected efficiencies to rank DMUs give better results. Indeed, we propose a new ranking approach, based on considering the DEA axioms for imprecise data that removes the existing drawbacks. Some numerical examples are provided to explain the content of the paper.

Keywords: Data envelopment analysis (DEA), efficiency measure, expected efficiencies, imprecise data

1-Introduction
Charnes et al. (1978) developed data envelopment analysis (DEA) for performance evaluation of several similar decision making units (DMUs). In this model, it is supposed that the values of inputs and outputs are exactly known. However, in many real applications, these values are imprecise. Imprecise data have various types: interval (bounded) data, weak and strong ordinal data, ratio bound data, multiplied order data, and so on. The mathematical representation of these data is given in Park (2007). So far, different approaches have been developed to calculate the relative efficiencies with the imprecise data in DEA. Some methods are given to rank DMUs based on only the upper bound efficiencies (Cooper et al. 1999, 2001; Kim et al. 1999; Lee et al. 2002; Zhu 2003, 2004; Park 2004). Some other methods are developed to calculate the lower and upper bound efficiencies to rank DMUs (Despotis and Smirlis 2002; Wang et al. 2005; Kao 2006; Park 2007).

Cooper et al. (1999) considered the interval and weak ordinal data in DEA and named the new nonlinear model as imprecise DEA (IDEA). They converted the model into an equivalent linear model through the scale transformation and variable alterations. Kim et al. (1999) used IDEA for performance evaluation in Telephone offices. Lee et al. (2002) extended the IDEA concept to the additive DEA model. Despotis and Smirlis (2002) developed two linear programming to estimate the lower and upper bound efficiencies by considering the pessimistic and optimistic state for each DMU.
Zhu (2003) showed that the scale transformation in the Cooper et al. (1999, 2001) approach is redundant. He converted the interval and weak ordinal data into the exact data to estimate the relative efficiencies.

Zhu (2004) used the method for performance evaluation in the Korean Mobile Telecommunication Company. Zhu (2003) and Park (2004) have used the same variable alteration to convert the nonlinear IDEA model into a linear model.

Wang et al. (2005) proposed two linear mathematical programming to obtain the lower bound and upper bound efficiencies by considering a unique production frontier for all DMUs. Kao (2006) emphasized that the efficiency scores should be imprecise in the presence of imprecise data. He proposed two two-level mathematical programming to calculate the lower bound and upper bound efficiencies. Park (2007) used the concept of supremum and infimum and proposed a mathematical programming for calculating the lower bound efficiencies.

Park (2010) investigated the dual model of IDEA and its relationships with primal problem based on the duality theory in IDEA. Marbini et al. (2014) investigated the performance evaluation in the presence of interval data, without sign restrictions. He et al. (2016) developed some DEA models to improve the inputs and outputs of inefficient DMUs such that their upper bound efficiency scores become one, in the presence of interval data. Ebrahimi et al. (2017) proposed a new nonlinear model to efficiency measure in the presence of both general weight restrictions and different types of imprecise data. They proposed a simulation-based genetic algorithm to estimate the efficiencies. Ebrahimi and Rahmani (2017) developed a mixed integer DEA model to find the best BCC-efficient DMUs by solving only one model. Ebrahimi and Khalili (2018) developed a new mixed integer DEA model to find the best DMU in the presence of both weight restrictions and different types of imprecise data. They utilized the model to find the best supplier in the presence of assurance region type I and interval and ordinal data. Ebrahimi et al. (2018) investigated the existing methods to estimate the efficiencies in the presence of interval and ordinal data. They illustrated some drawbacks of the existing methods and proposed a new method to calculate the efficiency scores.

It should be noted that the IDEA approach has been used to efficiency measure in many real-life applications. Farzipoor Saen (2007) applied the proposed method by Zhu (2003) for ranking the suppliers in the supplier selection problem. Asosheh et al. (2010) presented a mixed integer IDEA model to find the most efficient information technology (IT) projects. Ebrahimi et al. (2014) studied the drawbacks of the proposed model by Asosheh et al. (2010). They showed that the model is unable to find the best IT project and proposed a new approach to eliminate the drawbacks. Toloo and Nalchigar (2011) developed a new mixed integer imprecise DEA model to find the best supplier in the supplier selection problem. They used the proposed method of Zhu (2003) to handle the imprecise data and utilized their model to find the best supplier among 18 supplier in the presence of both interval and ordinal data.

Karsak and Dursun (2014) presented a supplier selection methodology by using the IDEA and Quality function deployment (QFD). Toloo (2014) presented a mixed integer programming IDEA model to determine the best supplier in the supplier selection problem. Chen et al. (2017) developed some mathematical DEA models to cope with bounded and Likert scale data. They have used the models for performance evaluation of the regional energy efficiency in China. Khalili-Damghani et al. (2015) applied the DEA model to efficiency measure of combined cycle power plant in the presence of interval data. Baghery et al. (2016) applied the DEA model to prioritize failures in the automotive industry with interval data. Toloo et al. (2018) developed some new DEA models to calculate the lower and upper bound efficiencies in the presence of interval dual-role factors. They used the models to efficiency measure in bank branches.

As literature review shows, different approaches have been developed to calculate the relative efficiency scores in the presence of imprecise data. In the next section, we show that considering the lower bound and upper bound efficiencies to rank DMUs may gives incorrect ranking. We explain the problems in the theory and use some numerical examples to clarify them. Therefore, the main contributions of the paper is as follows:

- Showing the drawbacks of existing methods to rank DMUs in the presence of imprecise data.
Developing a new algorithm to rank DMUs, based on considering the DEA axioms for imprecise data. The algorithm eliminates the drawbacks. It should be here emphasized that the proposed approach in this paper uses a set of exact data instead of imprecise data to calculate the efficiencies and expected efficiencies. To efficiency measure in the presence of random noise in terms of measurement errors, specification errors and also chance constraint DEA models the interested readers can refer to Olesen and Petersen (2016).

The rest of the paper is organized as follows: in section 2, we explain the drawbacks of the existing approaches. Section 3 explains the developed approach of this paper. Numerical examples and conclusions are given in sections 4 and 5, respectively.

2-The problems of the existing methods
This section explains the problems of the existing methods to efficiency measure in the DEA model with imprecise data. First, we study the proposed method by Park (2007). He applied the concept of supremum and infimum and developed the following model (1) to estimate the lower bound efficiency score of DMU

\[
\begin{align*}
\max & \sum_{r=1}^{m} u_r \inf \{ y_{ip} | y_r \in \Theta^+_i \} \\
\text{s.t.} & \quad \sum_{i=1}^{n} v_i \sup \{ x_{ip} | x_i \in \Theta^-_i \} = 1 \\
& \quad \sum_{r=1}^{m} u_r \inf \{ y_{ip} | y_r \in \Theta^+_i \} - \sum_{i=1}^{n} v_i \sup \{ x_{ip} | x_i \in \Theta^-_i \} \leq 0 \\
& \quad \sum_{r=1}^{m} u_r \sup \{ y_{ip} | y_r \in \Theta^+_i \} - \sum_{i=1}^{n} v_i \inf \{ x_{ip} | x_i \in \Theta^-_i \} \leq 0, \quad j = 1,...,k, \quad j \neq p \\
& \quad u_r, v_i \geq 0, \quad \forall r, i
\end{align*}
\]

In this model \( \Theta^-_i \) and \( \Theta^+_i \) represent the imprecise data for inputs and outputs, respectively. It should be noted that the upper bound efficiency score can be calculated by using the mentioned methods in the previous section, such as Cooper et al. (1999, 2001) and Zhu (2003). The model to obtain the upper bound efficiency score is as follows:

\[
\begin{align*}
\max & \sum_{i=1}^{m} u_i y_{ip} \\
\text{s.t.} & \quad \sum_{i=1}^{n} v_i x_{ip} = 1 \\
& \quad \sum_{r=1}^{m} u_i y_{ip} - \sum_{i=1}^{n} v_i x_{ip} \leq 0, \quad j = 1,...,k \\
& \quad x_i = (x_{ij}) \in \Theta^-_i \quad \forall i \\
& \quad y_i = (y_{ij}) \in \Theta^+_r \quad \forall r \\
& \quad u_i, v_i \geq 0 \quad \forall i, r
\end{align*}
\]
Based on the lower and upper bound efficiencies, Park (2007) classified DMUs into three groups, as follows.

**Inefficient**: the upper bound efficiency score is less than one.

**Potentially efficient**: the upper bound efficiency score is equal to unity, but the lower bound is less than one.

**Perfectly efficient**: the lower bound efficiency score is equal to unity.

To solve the model (1), the procedure of calculating the Inf and Sup for ordinal data is as follows. Suppose the \( i^{th} \) inputs of DMUs is in a weak ordinal format, as shown in (3).

\[
x_{i1} \leq x_{i2} \leq \ldots \leq x_{i,p-1} \leq x_{ip} \leq x_{i,p+1} \leq \ldots \leq x_{ik}
\]  

(3)

Since, DEA models have the unit-invariant property, therefore, Park (2007) normalized the data of relation (3) as shown in (4).

\[
0 \leq x_{i1}' \leq x_{i2}' \leq \ldots \leq x_{i,p-1}' \leq x_{ip}' \leq x_{i,p+1}' \leq \ldots \leq x_{ik}' \leq 1
\]  

(4)

Now, the Inf and Sup are calculated as follows (DMU\(_{p}\) is under evaluation):

\[
sup\{x_{ip}'\} = \{0 \leq x_{i1}' \leq x_{i2}' \leq \ldots \leq x_{i,p-1}' \leq x_{ip}' \leq x_{i,p+1}' \leq \ldots \leq x_{ik}' \leq 1\} = 1
\]

\[
inf\{x_{ij}'\} = \{0 \leq x_{i1}' \leq x_{i2}' \leq \ldots \leq x_{i,p-1}' \leq x_{ip}' \leq x_{i,p+1}' \leq \ldots \leq x_{ik}' \leq 1\} = 0, \forall j \neq p
\]  

(5)

In other words, before solving the model (1), the weak ordinal data (3) is replaced with the following integer numbers.

\[
x_{ip} = 1 \& x_{ij} = 0 \forall j \neq p
\]

Obviously, these numbers are infeasible to use in the model (1). Indeed, the correct numbers are 

\[
x_{ib} = 1 \forall b \geq p \& x_{ij} = 0 \forall j \leq p - 1, \text{ to keep the relation (3)}.
\]

To show the problem in more detail, let the numerical example is used in Park (2007). In this example, there are eight telephone offices with three inputs and three outputs. The data of third output is in the weak ordinal format as follows:

\[
D_{3} = \{x_{3} \in R^{3_3} | x_{34} \geq x_{35} \geq x_{33} \geq x_{37} \geq x_{31} \geq x_{36} \geq x_{32} \geq x_{38}\}
\]  

(6)

According to the Park (2007) approach, to evaluate the DMU\(_{1}\) the vector of \( x_{3}^* = (1,0,0,0,0,0,0,0)\) is used instead of weak ordinal data (6), that is infeasible. However, applying the feasibility condition implies that we should use \( x_{3}^* = (1,0,1,1,1,0,1,0)\). Furthermore, Park (2007) method uses only zero and one for all ordinal data. It should be noted that, the probability of occurrence of these data is near to zero in practice.

The above discussion shows that Park (2007) has used a set of infeasible integer numbers instead of weak ordinal data. As a result, the obtained lower bound efficiency score will be incorrect with these data. It should be noted that Park (see the last paragraph on pp. 536) claimed that the efficiency score will be safe and sound if the exact data set is infeasible or not. But, it is so easy to give an example to show that the claim is incorrect.

Moreover, in the following example, we show that the Park (2007) method and also some other existing methods give incorrect result in some cases.
**Example 1**: Consider three DMUs, each uses two weak ordinal inputs to produce one ordinal output, as given in table 1.

<table>
<thead>
<tr>
<th>DMU No.</th>
<th>Input 1 (ordinal)</th>
<th>Input 2 (ordinal)</th>
<th>Output (ordinal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$x_{11}$</td>
<td>$x_{21}$</td>
<td>$y_{11}$</td>
</tr>
<tr>
<td>2</td>
<td>$x_{12}$</td>
<td>$x_{22}$</td>
<td>$y_{12}$</td>
</tr>
<tr>
<td>3</td>
<td>$x_{13}$</td>
<td>$x_{23}$</td>
<td>$y_{13}$</td>
</tr>
</tbody>
</table>

* ranking such that $x_{13} \leq x_{12} \leq x_{11}$, $x_{23} \leq x_{22} \leq x_{21}$, $y_{11} \leq y_{12} \leq y_{13}$

It is easy to see that DMU 3 dominates DMU 2, and DMU 2 dominates DMU 1. In other words, a wise decision maker ranks these DMUs as: DMU 3 $>$ DMU 2 $>$ DMU 1. However, we show that the existing methods give an incorrect ranking for the example.

Applying the proposed methods of Cooper et al. (1999, 2001), Zhu (2003, 2004), Park (2004) and other existing methods to calculate the upper bound efficiency scores, yield that all of the DMUs are efficient. It should be noted that just in one special situation, $x_{13} = x_{12} = x_{11}$ & $x_{23} = x_{22} = x_{21}$ & $y_{11} = y_{12} = y_{13}$, that occurs with zero probability in practice, all of the DMUs are efficient.

Also, the existing methods to rank DMUs based on considering the both of lower and upper bound efficiencies give incorrect results. We use the approach of the Park (2007) to calculate the lower bound efficiencies. The efficiencies are calculated in both states, considering the feasibility condition and without considering the feasibility conditions. The results are summarized in table 2.

<table>
<thead>
<tr>
<th>The DMU under evaluation</th>
<th>Without considering feasibility</th>
<th>Considering feasibility</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DMU 1</td>
<td>DMU 2</td>
</tr>
<tr>
<td>$x_{11}^*$</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$x_{12}^*$</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$x_{13}^*$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$x_{21}^*$</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$x_{22}^*$</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$x_{23}^*$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$y_{11}^*$</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$y_{12}^*$</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$y_{13}^*$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Lower bound efficiencies</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The results show that based on Park (2007) approach, the efficiency scores of the three DMUs is equal to $[0, 1]$. In other words, the DMUs have the same rank, that is unacceptable. It should be noted that the proposed methods by Despotis and Smirlis (2002), Wang et al. (2005) and Kao (2006) also gives a similar result. In other words, all of the existing methods produce incorrect ranking for the example.
In the next example, we find another problem in Park (2007) method. Indeed, we show that the method is unable to calculate the lower bound efficiencies in some cases.

**Example 2:** Consider two DMUs, each uses one ordinal input to produce one precise output.

<table>
<thead>
<tr>
<th>DMU No.</th>
<th>Input (ordinal)</th>
<th>Output (exact)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( x_{11} )</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>( x_{12} )</td>
<td>4</td>
</tr>
</tbody>
</table>

* ranking such that \( x_{11} \leq x_{12} \)

The basic maximin DEA model to calculate the relative efficiency score is as follows:

\[
\max_{u_r \geq 0, v_r \geq 0, \forall r} \left\{ \frac{\sum u_r y_{rp}}{\sum v_r x_{rp}} \right\} \max_i \left\{ \frac{\sum u_r y_{ij}}{\sum v_r x_{ij}} \right\}
\]

(7)

The model can be converted to the linear CCR-DEA model (Khalili et al. 2010).

The result of calculating the \( \text{Inf} \) and \( \text{Sup} \) based on Park (2007) approach is presented in table 4. It is easy to see that it is impossible to calculate the efficiencies for these data by using the model (7), except DMU1 by considering the feasibility condition. So, the ranking of DMUs is not possible.

<table>
<thead>
<tr>
<th>The DMU under evaluation</th>
<th>Without considering feasibility</th>
<th>Considering feasibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>The values of variables</td>
<td>DMU1 ( x_{11} ) ( x_{12} )</td>
<td>DMU1 DMU2</td>
</tr>
<tr>
<td>( x_{11} )^*</td>
<td>1</td>
<td>1\text{ cannot be calculated}</td>
</tr>
<tr>
<td>( x_{12} )^*</td>
<td>0</td>
<td>1\text{ cannot be calculated}</td>
</tr>
<tr>
<td>Lower bound efficiencies</td>
<td>Cannot be calculated</td>
<td>0.5 \text{ cannot be calculated}</td>
</tr>
</tbody>
</table>

Overall, the drawbacks of the existing methods can be summarized as follows:

- Replacing ordinal data with a set of infeasible integer numbers to calculate the lower bound efficiencies. Obviously, the infeasible solutions lead to the incorrect optimal solution in the mathematical programing.
- Produce incorrect ranking, as shown in the example 1.
- In some cases, the existing methods are unable to calculate the lower bound efficiencies, and so unable to rank DMUs.
- Using only zero and one instead of ordinal data that their occurrence possibility is near to zero in practice.

**3-The proposed algorithm**

The existing interval-data methods are capable to find the largest possible efficiency and the smallest possible efficiency scores. However, the probability of their occurrence is too small to have practical meaning. In the following, we propose a new method to rank DMUs in the presence of imprecise data, by considering the efficiency distribution. First, we mention the first axiom of the DEA that helps us to develop this method.
The production possibility set (PPS) and the DEA model are based on some axioms. The first axiom is inclusion of observation. In the presence of imprecise data, there is not precise observed data to construct the PPS. To clarify the topic, consider the data presented in table 3. For this data, the PPS with constant returns to scale (CRS) technology is \( \{ (x, y) \in \mathbb{R}^2 : x \geq 0 \& y \leq 4 \} \) that has been shown in figure 1. For this PPS, the production frontier is the line between (0,0) and (0,4), that is meaningless. In other words, by this PPS it is impossible to calculate the relative efficiencies. This problem shows that the DEA axioms are necessary to build the PPS and determine the production frontier. The previous methods did not consider the axiom that leads to the incorrect results.

To consider the inclusion of observation axiom, the imprecise data can be replaced with random numbers. In other words, we generate a sufficient amount of random data instead of imprecise data. More specially, for ordinal data of table 3, the random data can be generated as follows. First, we generate a random number in interval \([0, 1]\) for \(x_{12}\). Then, the value of \(x_{11}\) is randomly generated in interval of \([0, x_{12}]\). Since, there is no information about the probability distribution of the imprecise data, so it can be assumed uniform. In this case, the steps of the proposed algorithm are as follows:

- Generating uniform random data with \(N\) iterations.
- Calculating efficiencies for all DMUs with the data obtained from previous stage. At this stage, \(N\) efficiency scores for each DMU will be obtained.
- Calculating the average efficiency score of efficiencies. Indeed, the average efficiency is an estimation of the expected value of efficiencies.
- Ranking DMUs based on the expected efficiencies, DMU\(_{j1}\) dominates DMU\(_{j2}\) if and only if the expected efficiency score of DMU\(_{j1}\) is greater than DMU\(_{j2}\).

Obviously, increasing \(N\) leads to obtain the more exact expected efficiencies. If we were able to extract the efficiency distribution, we could calculate the exact value of expected efficiencies. Indeed, the average efficiency is an estimation of expected efficiency, so these values will be approximately equal by increasing the value of \(N\). To show the benefits of the proposed approach we apply it to examples 1 and 2, in the next section.

### 4-Numerical illustration

In this section, to demonstrate the capability and applicability of the proposed method we apply it to the data presented in tables 1 and 3 and discuss the results.

**Example 3:** As discussed in example 1, the existing methods are unable to rank the three DMUs, presented in table 1. We apply the proposed approach in the previous section for this example with \(\varepsilon = 0.001\). The results are provided in table 5, for different values of \(N\). The minimum, maximum, standard deviation (SD) and average efficiencies are calculated for each DMU. As we expected, the
DMU₃ is always efficient, for different values of N. But, the DMU₁ and DMU₂ are inefficient with the average efficiency scores of 0.0914 and 0.336, respectively. A visual inspection reveals that the averages efficiencies do not differ much for N≥1000 for all DMUs. For N≥2000 the differences between the average efficiencies appears at the third digit after the decimal point for all DMUs.

As it can be seen and also we expected, as the value of N increases, the lower bound and upper bound of the efficiencies expand accordingly.

<table>
<thead>
<tr>
<th>N</th>
<th>DMU₁ Ave.</th>
<th>DMU₁ SD</th>
<th>DMU₁ Min</th>
<th>DMU₁ Max</th>
<th>DMU₂ Ave.</th>
<th>DMU₂ SD</th>
<th>DMU₂ Min</th>
<th>DMU₂ Max</th>
<th>DMU₃ Ave.</th>
<th>DMU₃ SD</th>
<th>DMU₃ Min</th>
<th>DMU₃ Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.0418</td>
<td>0.1869</td>
<td>0.0094</td>
<td>0.0908</td>
<td>0.3260</td>
<td>0.2793</td>
<td>0.0516</td>
<td>0.5403</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>100</td>
<td>0.0961</td>
<td>0.1265</td>
<td>4.4268e-05</td>
<td>0.5347</td>
<td>0.3422</td>
<td>0.2574</td>
<td>0.0046</td>
<td>0.9086</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1000</td>
<td>0.0987</td>
<td>0.1166</td>
<td>2.4921e-06</td>
<td>0.7990</td>
<td>0.3361</td>
<td>0.2343</td>
<td>7.6826e-04</td>
<td>0.9717</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2000</td>
<td>0.0912</td>
<td>0.1099</td>
<td>3.1149e-05</td>
<td>0.7607</td>
<td>0.3334</td>
<td>0.2321</td>
<td>6.1184e-04</td>
<td>0.9834</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5000</td>
<td>0.0914</td>
<td>0.1096</td>
<td>2.4148e-06</td>
<td>0.8172</td>
<td>0.3359</td>
<td>0.2320</td>
<td>4.7377e-05</td>
<td>0.9930</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Rank</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It should be noted that considering the efficiency distribution is very important. In most of existing methods, only the lower and upper bound efficiencies are used to rank the DMUs. As it can be seen, for N=5000, the average efficiency score of DMU₂ is near to 0.336 with lower bound of 4.74*10⁻⁵ and upper bound of 0.993. The efficiency distribution of DMU₂ is shown in figure 2 for different values of N. We divided the efficiency score of [0 , 1] into 20 equal segments. The Y-axis is the efficiency frequency for different range of efficiency. The charts show that the efficiency distribution is not uniform. If it was uniform, then the expected efficiency score of DMU₂ was near to 0.5, instead of 0.336.

Also, we calculate the SD of efficiencies that shows the amount of variation or dispersion of efficiencies. A low SD implies that the data points tend to be close to the mean data, while a high SD implies that the data points are spread out over a wider range of values. As we expect, the SD of DMU₃ is zero, since it’s lower bound efficiency score is equal to one, and so this DMU is perfectly efficient. Moreover, the SD of DMU₂ is greater than the SD of DMU₁. This is because the efficiency score of DMU₂ varies between approximately zero and one but, the efficiency score of DMU₁ varies between approximately zero and 0.81. It should be noted that by increasing the value of N the value of SD are reduced. This matter can be seen by considering the figure 2, that for large value of N the curve of efficiency distribution is more smooth.
Example 4: in example 3, it is explained that the Park (2007) method is unable to calculate the efficiency scores to rank the DMUs.

Now, we apply the proposed method for the example. The results summarized in table 6, show that the DMU₁ with the expected efficiency of 0.844 has a better rank in comparing to DMU₂ with the expected efficiency score of 0.757. Also, Since, the variation of efficiency scores of DMU₁ is significantly less than the DMU₂, so it’s SD is less than the SD of DMU₂.
Table 6. The result of the proposed approach for data presented in table 3

<table>
<thead>
<tr>
<th>N</th>
<th>DMU1 Ave.</th>
<th>DMU1 SD</th>
<th>DMU1 Min</th>
<th>DMU1 Max</th>
<th>DMU2 Ave.</th>
<th>DMU2 SD</th>
<th>DMU2 Min</th>
<th>DMU2 Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.8923</td>
<td>0.1826</td>
<td>0.5263</td>
<td>1</td>
<td>0.5603</td>
<td>0.2742</td>
<td>0.0372</td>
<td>1</td>
</tr>
<tr>
<td>100</td>
<td>0.8687</td>
<td>0.1793</td>
<td>0.5014</td>
<td>1</td>
<td>0.7014</td>
<td>0.3003</td>
<td>0.0269</td>
<td>1</td>
</tr>
<tr>
<td>1000</td>
<td>0.8486</td>
<td>0.1810</td>
<td>0.5010</td>
<td>1</td>
<td>0.7511</td>
<td>0.3175</td>
<td>0.018</td>
<td>1</td>
</tr>
<tr>
<td>2000</td>
<td>0.8451</td>
<td>0.1846</td>
<td>0.5001</td>
<td>1</td>
<td>0.7579</td>
<td>0.3219</td>
<td>0.0019</td>
<td>1</td>
</tr>
<tr>
<td>5000</td>
<td>0.8443</td>
<td>0.1851</td>
<td>0.5000</td>
<td>1</td>
<td>0.7573</td>
<td>0.3244</td>
<td>7.1378e-05</td>
<td>1</td>
</tr>
</tbody>
</table>

Rank: 1

This example shows that the efficiencies of average data differ from average efficiencies. Indeed, the expected values of \( x_{11} \) and \( x_{12} \) with the uniform data generation are 0.25 and 0.5, respectively. This average data implies that the two DMUs are both efficient. However, our proposed approach considers the expected efficiencies instead of the efficiencies of expected data. The efficiency distributions of DMU1 and DMU2 are shown in figure 3 and 4 for different values of \( N \), respectively. The chart shows that more than 50% of efficiency frequency are in interval [0.95, 1]. This matter implies that we need to consider the distribution of the efficiency scores in the interval. Most of existing methods rank DMUs based on just the lower and upper bound efficiencies, without considering the efficiency distribution.

![Fig 3. The efficiency distribution of DMU1 for different values of N](image)

Fig 3. The efficiency distribution of DMU1 for different values of \( N \)
Conclusion
The paper explains the drawbacks of the existing methods in the DEA to rank DMUs with imprecise data. We show that the methods did not consider the DEA axioms, so may produce incorrect ranking in some cases. It is shown that the Park (2007) method uses a set of infeasible integer data instead of imprecise data. It is also shown that in some cases, the method is unable to calculate the efficiencies or produces a rational ranking.

It was emphasized that the DEA model and the PPS are based on some axioms, especially the inclusion of observation axiom. It is shown that with imprecise data, we may unable to determine the PPS and the production frontier correctly. Therefore, by considering the DEA axioms, a simple practical algorithm is presented to rank DMUs in the presence of imprecise data.

The proposed approach considers the efficiency distribution and expected efficiencies to rank the DMUs, instead of the lower and upper bound efficiencies. It is explained that the approach covers the mentioned problems and gives more reliable results.

Acknowledgement
The author wish to thank the editor-in-chief and anonymous referees for their helpful comments.

5-Conclusions
The paper explains the drawbacks of the existing methods in the DEA to rank DMUs with imprecise data. We show that the methods did not consider the DEA axioms, so may produce incorrect ranking in some cases. It is shown that the Park (2007) method uses a set of infeasible integer data instead of imprecise data. It is also shown that in some cases, the method is unable to calculate the efficiencies or produces a rational ranking.

It was emphasized that the DEA model and the PPS are based on some axioms, especially the inclusion of observation axiom. It is shown that with imprecise data, we may unable to determine the PPS and the production frontier correctly. Therefore, by considering the DEA axioms, a simple practical algorithm is presented to rank DMUs in the presence of imprecise data.

The proposed approach considers the efficiency distribution and expected efficiencies to rank the DMUs, instead of the lower and upper bound efficiencies. It is explained that the approach covers the mentioned problems and gives more reliable results.

Acknowledgement
The author wish to thank the editor-in-chief and anonymous referees for their helpful comments.
References


