Research Article

An Extended Three-Stage DEA Model with Interval Inputs and Outputs

Guo-Qing Cheng*, Liang Wang, Ying-Ming Wang

Decision Sciences Institute, Fuzhou University, No. 2, Xueyuan Road, University Town, Fuzhou, 350116, PR. China

1. INTRODUCTION

Data envelopment analysis (DEA) [1] was first proposed by the famous American operations researchers Cooper and Rhodes. The advantage lies in the ability to evaluate the efficiency of similar decision-making units (DMUs) with multiple input and output indicators and complex relationships. Since it was proposed, it is a popular approach and has been widely discussed in the literature [2–5]. The traditional three-stage DEA model established on the DEA model firstly proposed by Fried et al. [6], which takes into account the influence of external environment and random errors on the efficiency calculation based on Stochastic Frontier Analysis (SFA). It uses the DEA-SFA-DEA method to build three-stage DEA model [7]. The first stage is to use the input-oriented A. Charnes, W.W. Cooper, E. Rhodes (CCR) or R. D. Banker, A. Charnes, W. W. Cooper (BCC) model to get the efficiency and investment relaxation of each DMU [8–10]. In the second stage, because the technical efficiency is affected by external environmental factors, random interference, and management inefficiency co-effects, so the cost-oriented SFA model is established with input slack as the dependent variable and external environmental factors as the independent variables [11–13], thereby eliminating the impact of the external environmental factors; in the third stage, the newly obtained input–output values are used for efficiency measurement.

The traditional three-stage DEA model evaluation is to measure all DMUs whose inputs and outputs are exact data, and only needs to use linear programming. However, in actual production activities, it is impossible to obtain certain indicators data due to the accuracy problems of some measurement methods, technical limitations, and lack of information [14–16]. If these uncertain factors are ignored, the relative effectiveness of these DMUs will still be evaluated using the DEA model established on the basis of certain values, and biased or even wrong information will be obtained, which will bring some errors to management decisions. Based on the research results, the traditional three-stage DEA model cannot achieve the efficiency measurement of uncertain data [17]. Therefore, it seems convenient and necessary to consider the uncertain DEA model.

For the uncertain DEA model, Cooper et al.1 introduced the concept of uncertain data to DEA for the first time, and proposed an evaluation method based on interval efficiency. Since then, a series of interval DEA methods have emerged, mainly including variable replacement methods, interval efficiency methods, and integration methods [18,19]. The interval DEA method for variable replacement was first proposed by Cooper et al. The method is to replace the interval data to exact data for each indicator, and then obtain the efficiency value of DMU. The interval efficiency method was first proposed by Despotis and Smirlis [20]. It mainly uses the combination of the maximum and minimum values for the interval of the unit under evaluation and the reference unit to obtain the maximum and minimum efficiency of the unit under evaluation. The variable replacement method can be regarded as a part of the interval efficiency method [21,22]. The efficiency value of DMU obtained by the variable replacement method is actually the maximum value of DMU efficiency obtained by the interval efficiency method [23,24]. However, interval DEA method cannot eliminate the influence of external environmental factors on the efficiency evaluation. So, there is still a need from the interval DEA method to develop a new model that keeps original advantage and considers the influence of environmental factors.

*Corresponding author. Email: 330721700@qq.com
On such motivation basis, this study proposes a new three-stage interval DEA method, which combines three-stage DEA model with interval DEA method. Compared with the traditional interval DEA method, the new three-stage interval DEA method takes into account the influence of external environmental factors. Compared with three-stage DEA model, the new three-stage interval DEA method can evaluate the input and output of interval data. The effectiveness is classified based on the interval efficiency by three-stage interval DEA method. Additionally, for the problem of DMUs' natural differences [21], this paper combines the three-stage interval DEA efficiency model with cluster analysis to classify. Poorly performing in the same category can select best performing as reference targets for improvement. Afterward, a study with numerical example is presented to expose the advantages of the new proposed methods.

The remainder of this paper is organized as follows. Section 2 briefly introduces the traditional three-stage DEA model and interval DEA method in order to make our proposed method understood easily. Section 3 develops three-stage interval DEA model and a new method to identify benchmarks by cluster analysis. Section 4 conducts numerical example with related comparisons to illustrate the superiority, validity, and feasibility of the proposed method regarding previous ones. The conclusions and future works are offered in Section 5.

2. BACKGROUND

In this section, the traditional three-stage DEA model and the interval DEA method will be introduced respectively to facilitate unfamiliar readers can understand the proposed model more easily and clearly.

2.1. The Traditional Three-Stage DEA Model

We first introduce the three-stage DEA model proposed by Fried et al. [6]. In the first stage, the traditional DEA model is used to analyze efficiency. In the second stage, the SFA method is used to correct the effects of environmental variable and random error. In the third stage, the adjusted input data and original output data are used for DEA efficiency measurement again [25,26].

The first stage: Assumed that there are $n$ DMUs to be evaluated [27,28], where each DMU contains $m$ inputs and $s$ outputs. The $i$th input factors of $j$th DMU is $x_{ij}$, and the $r$th output factors of $j$th DMU is $y_{jr}$. The initial DMUs' performance evaluation is conducted using a traditional DEA model. The traditional DEA model can be written as follows:

$$\max \theta_i = \sum_{r=1}^{s} u_r y_{rj}$$

$$\sum_{i=1}^{m} v_i x_{ij} = 1$$

s.t. $$\sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} u_r y_{rj} \geq 0$$

$$j = 1, 2, \cdots, n$$

$$v_i, u_r \geq \varepsilon, \forall i, r$$

where $\varepsilon$ is a non-Archimedes infinitesimal; $v_i$ is the $i$th input indicator weight; $u_r$ is the $r$th output indicator weight. $\theta_i$ represents the relative efficiency value of the evaluated DMU$_j$.

With the addition of slack variable, the dual form of the above model can be expressed as [24]

$$\min \theta_i - \varepsilon \left( \sum_{r=1}^{n} S^+_r + \sum_{r=1}^{m} S^-_r \right)$$

$$\sum_{j=1}^{n} \lambda_j x_{ij} + S^-_j = \theta_i y_{rj}$$

s.t. $$\sum_{j=1}^{n} \lambda_j y_{rj} - S^+_r = \nu_{rj}$$

$$j = 1, 2, \cdots, n; i = 1, \cdots, m; r = 1, \cdots, s$$

$$\lambda_j, S^+_j, S^-_j \geq 0, \forall j, i, r$$

where $S^+_j$ is the slack variable of the $r$th output of DMU$_j$; $S^-_r$ is the slack variable of the $i$th input of DMU$_j$; and $\theta_i$ is the relative efficiency of DMU$_j$.

The second stage: SFA regression analysis of environmental variables is used to overcome the shortcomings of the traditional DEA model [29,30]. The input slack variables corresponding to the first-stage solution are decomposed into a function with three variables of environmental impact factor, random error factor, and management inefficiency factor [31,32]. The model construction of the SFA regression function is based on the method proposed by Fried et al. [2] as follows:

$$S_i = f(Z_i; \hat{\beta}) + \nu_i + \mu_i;$$

$$i = 1, 2, \cdots, m; t = 1, 2, \cdots, n$$

In Eq. (3), $S_i$ represents the slack variables of the $i$th DMU on the $t$th input indicator; $Z_i$ represents the environmental variables of individual DMU; $\hat{\beta}$ is the coefficients of environmental variables; $f(Z_i; \hat{\beta})$ represents the influence of environmental variables on input slack variables; $\nu_i$ is the random error; $\mu_i$ is the management inefficiency of truncated normal distribution; $\nu_i + \mu_i$ is the mixed error term. According to the regression results, adjusting the selected input variable [33,34] the adjustment formula is following as

$$X_{ni} = X_{ni} + [\max(f(Z_i; \hat{\beta})) - f(Z_i; \hat{\beta})]$$

$$i = 1, 2, \cdots, m; t = 1, 2, \cdots, n$$

where $X_{ni}$ is the new input variable values after homogenization; $X_{ni}$ is the values before adjustment for each DMU. $[\max(f(Z_i; \hat{\beta})) - f(Z_i; \hat{\beta})]$ represents the influence of the adjusted environmental factors; $[\max(\nu_i) - \nu_i]$ represents the influence of the adjusted random error factors. These two items adjust the external environmental factors and luck of all DMUs to the same situation.

The third stage: This stage is a repetition of the first stage, the adjusted input data and the original output data is used to calculate the efficiency value of each DMU. At this time, the efficiency value of DMUs is obtained by eliminating environmental variables and random errors [35,36]. Thus, the efficiency obtained in the third stage will be more realistic in reflecting the managerial efficiency.
The traditional three-stage DEA model can eliminate the influence of environmental factors, and it is easy to evaluate the exact data. However, it has no effective solution when input and output data are in the form of intervals. Therefore, it seems necessary and convenient to develop a new DEA method to overcome such a limitation.

2.2. The Interval DEA Method

Without loss of generality, it is assumed that all the input and output data \( x_j \) and \( y_j \) cannot be exactly obtained. Due to uncertainty, it is only known to lie within the upper and lower bounds represented by the intervals, where expressed as \( x_j \in [x_{i,j}^L, x_{i,j}^U] \) and \( y_j \in [y_{j,r}^L, y_{j,r}^U] \).

The interval DEA model is to calculate the maximum efficiency value and the minimum efficiency value of each evaluated units [37]. According to the different combinations of the maximum and minimum values of the input–output interval index of the evaluated DMU and reference unit, and then form an efficiency interval [38]. This section analyzes the relative effectiveness of each DMU. Therefore, the interval DEA model is obtained by deforming Eq. (1) as follows:

\[
\max \theta_{ij} = \sum_{i=1}^{s} u_i \left[ x_{i,j}^U, x_{i,j}^L \right] \\
\sum_{i=1}^{m} v_i \left[ x_{i,j}^U, x_{i,j}^L \right] = 1 \\
\sum_{j=1}^{n} \left[ v_{j} \left( x_{i,j}^L \right) - u_i \left( y_{j,r}^L \right) \right] \geq 0 \\
\sum_{j=1}^{n} \left[ v_{j} \left( x_{i,j}^U \right) - u_i \left( y_{j,r}^L \right) \right] \geq 0 \\
\sum_{j=1}^{n} v_{j} \left( x_{i,j}^L \right) \geq u_i \left( y_{j,r}^L \right), \quad \forall i, r
\]

In order to solve such an uncertain situation and obtain the interval efficiency value, we firstly consider the best solution for DMU. Using the minimum input value \( x_{i,j}^L \) and the maximum output value \( y_{j,r}^U \) as the input–output value of DMU [39]. The other DMU \( j = 1, 2, \cdots, n, j \neq j_0 \) are opposite, using the maximum input value and the minimum output value [40]. From this, the model for solving the highest value of efficiency of DMU is

\[
\max \bar{\theta}^U = \sum_{i=1}^{s} u_i y_{i,j_0}^U \\
\sum_{i=1}^{m} v_i x_{i,j_0}^L = 1 \\
\sum_{j=1}^{n} \left[ v_{j} x_{i,j_0}^L - u_i y_{j,r}^L \right] \geq 0 \\
\sum_{j=1}^{n} \left[ v_{j} x_{i,j_0}^U - u_i y_{j,r}^L \right] \geq 0 \\
\sum_{j=1}^{n} v_{j} x_{i,j_0}^L \geq u_i y_{j,r}^L, \quad \forall i, r
\]

Secondly, we consider the worst situation for DMU. Using the maximum input value \( x_{i,j_0}^U \) and the minimum output value \( y_{j_0}^L \) as the input–output value of DMU. The other DMU \( j = 1, 2, \cdots, n, j \neq j_0 \) are opposite, using the minimum input value and the maximum output value [41]. From this, the model for solving the lowest value of efficiency of DMU is

\[
\max \theta^L = \sum_{i=1}^{s} u_i y_{i,j_0}^L \\
\sum_{i=1}^{m} v_i x_{i,j_0}^L = 1 \\
\sum_{j=1}^{n} \left[ v_{j} x_{i,j_0}^L - u_i y_{j,r}^L \right] \geq 0 \\
\sum_{j=1}^{n} \left[ v_{j} x_{i,j_0}^U - u_i y_{j,r}^L \right] \geq 0 \\
\sum_{j=1}^{n} v_{j} x_{i,j_0}^U \geq u_i y_{j,r}^L, \quad \forall i, r
\]

Motivated by the limitations pointed out in Introduction previously, the three-stage interval DEA model is proposed in this section, which cannot only eliminate the influence of environmental factors, but also deal with interval input–output data. Then, we combine cluster analysis and three-stage interval DEA model to identify benchmarks for poorly performing DMUs. The proposed three-stage interval DEA model and identification of benchmarks by cluster analysis are introduced as below.

3. THE PROPOSED MODEL

3.1. Three-Stage Interval DEA Model

The first stage: Let the input data \( x_j \) and output data \( y_j \) of DMU be interval as defined in Section 2.2, where \( x_j \in \left[ x_{i,j}^L, x_{i,j}^U \right] \) and \( y_j \in \left[ y_{j,r}^L, y_{j,r}^U \right] \). According to Eqs. (6) and (7), the interval efficiency value \( \theta_{ij}^{L, U} \) is obtained.

According to the previous research results [7,26], all DMUs could be classified in the following three categories:

\[
E^+ = \{ DMU | \theta_{ij}^{L} = 1, j \in (1, 2, \cdots, n) \} \\
E = \{ DMU | \theta_{ij}^{L} < 1, \theta_{ij}^{U} = 1, j \in (1, 2, \cdots, n) \} \\
E^- = \{ DMU | \theta_{ij}^{L} < 1, j \in (1, 2, \cdots, n) \}
\]

DMU \( j \) in \( E^+ \) means that the \( j \)th DMU is fully efficient in any case; DMU \( j \in E \) means that the \( j \)th DMU is efficient in the best situation, but it is inefficient in the worst situation, which indicates that this is partially efficient. DMU \( j \in E^- \) means that the \( j \)th DMU is
completely inefficient in any case. The interval efficiencies calculated in the first stage and the third stage classify the DMUs in three categories according to the above classification rules.

While calculating the interval efficiency value, $x_{ij}$ generates one input slack variable according to Eqs. (6) and (7) for the best situation and the worst situation respectively. Further analysis of these slack variables will be carried out in the second stage.

The second stage: Different from the past, the input–output variables discussed in this section are interval data. Corresponding to the best efficiency situation, the input indicator of DMU generates an input slack variable, which is assumed to be $S^U_i$. Corresponding to the worst efficiency situation, it is assumed to be $S^L_i$. Therefore, we firstly consider placing all DMUs at the best efficiency situation as follows:

$$S^U_i = f_i^U \left( Z_i; \hat{\beta}_i^U \right) + v^U_i + \mu^U_i; \quad i = 1, 2, \cdots, m; \quad t = 1, 2, \cdots, n$$

Secondly, we consider placing all DMUs at the worst efficiency situation as follows:

$$S^L_i = f_i^L \left( Z_i; \hat{\beta}_i^L \right) + v^L_i + \mu^L_i; \quad i = 1, 2, \cdots, m; \quad t = 1, 2, \cdots, n$$

According to the adjustment Eq. (4) and the above two SFA regression analysis Eqs. (9) and (10), corresponding to each DMU of the best efficiency situation, its adjustment formula is expressed as

$$X'^{U}_{it} = X^U_{it} + \left[ \max(f_i^U(Z_i; \hat{\beta}_i^U)) - (Z_i; \hat{\beta}_i^U) \right] - f_i^U(Z_i; \hat{\beta}_i^U) + \left[ \max(v^U_i) - v^U_i \right]; \quad i = 1, 2, \cdots, m; \quad t = 1, 2, \cdots, n$$

where $X'^{U}_{it}$ is the new adjusted input value under the best efficiency, $X^U_{it}$ is the original input value under the best efficiency. Similarly, corresponding to each DMU of the worst efficiency situation, its adjustment formula is expressed as

$$X'^{L}_{it} = X^L_{it} + \left[ \max(f_i^L(Z_i; \hat{\beta}_i^L)) - (Z_i; \hat{\beta}_i^L) \right] - f_i^L(Z_i; \hat{\beta}_i^L) + \left[ \max(v^L_i) - v^L_i \right]; \quad i = 1, 2, \cdots, m; \quad t = 1, 2, \cdots, n$$

Therefore, there are two adjustment values corresponding to an input variable. After adjustment, the maximum and minimum values of the new input can be obtained, and this is the new interval input data assumed $x'^{U}_{ij}$, which is following as

$$x'^{U}_{ij} \in \left[ x_{ij}^{L}, x_{ij}^{U} \right];$$

$$x'^{L}_{ij} = \min \left( x'^{U}_{ij}, x'^{L}_{ij} \right);$$

$$x'^{U}_{ij} = \max \left( x'^{U}_{ij}, x'^{L}_{ij} \right)$$

The third stage: This stage is also a repetition of the first stage, the adjusted interval input data and the original output data is used to calculate the interval efficiency value of each DMU by Eqs. (6) and (7). The interval efficiency of each DMU is the adjusted efficiency eliminating the influence of environmental factor and random error. At this time, the interval efficiency value is more fair and effective compared with the initial interval efficiency value.

### 3.2. Identification of Benchmarks by Cluster Analysis

In traditional DEA, the improvement target of an ineffective unit is the linear combination of effective units in its reference set. These ineffective units may be naturally different from the units in their reference set. Some researchers have suggested using cluster analysis, principal components, and multidimensional scaling to classify DMUs more accurately into similar groups or clusters [42].

From the perspective of how environmental factors affect the interval efficiency of DMUs, computing the degree of interval efficiency change before and after eliminating external environmental factors. We can obtain the similarity of those DMUs.

**Definition 1.** The degree of efficiency change vector for DMU $j$ is the ratio of the $j$th interval efficiency after eliminating the influence of external environmental factors to the original interval efficiency, which can be expressed as

$$\left( w^j, w^U_j \right), j \in \{1, 2, \cdots n\}$$

where $w^j$ and $w^U_j$ are the degree of change of the minimum and maximum interval efficiency values; $\theta^j, \theta^U_j$ are the minimum and maximum values of interval efficiency after eliminating the influence of external environmental factors; $\theta^j, \theta^U_j$ are the interval efficiency.

**Property 1.** The degree of efficiency change is strictly positive.

**Proof.** By Definition 1,

$$\therefore 0 < \theta^L_j, \theta^L_j, \theta^U_j, \theta^U_j \leq 1$$

$$\therefore w^L_j = \frac{\theta^L_j}{\theta^U_j} > 0,$$

$$w^U_j = \frac{\theta^U_j}{\theta^U_j} > 0$$

**Property 2.** When $0 < w^L_j, w^U_j < 1$, it indicates that the environmental factors have a inhibitory effect on the efficiency value; when $w^L_j = 1$, it indicates that the environmental factors have no effect on the minimum efficiency, when $w^U_j = 1$, it indicates that the environmental factors have no effect on the maximum efficiency; when $w^L_j, w^U_j > 1$, indicating that environmental factors play a role in promoting the efficiency value.

**Proof.**

$$\therefore w^L_j = \frac{\theta^L_j}{\theta^U_j}, w^U_j = \frac{\theta^U_j}{\theta^U_j}$$

$$\therefore \text{when } 0 < w^L_j, w^U_j < 1, \theta^L_j < \theta^L_j, \theta^U_j < \theta^U_j$$

$$\therefore \text{when } w^L_j = 1, \theta^L_j = \theta^L_j$$
when \( w^U_j = 1, \theta^L_j = \theta^U_j \)
\[ \therefore \text{when } w^L_j, w^U_j > 1, \theta^L_j > \theta^U_j, \theta^L_j > \theta^U_j \]

**Property 3.** When \( 0 < w_j^L, w_j^U < 1 \), the smaller the \( w_j^L, w_j^U \) value, the greater the impact of environmental factor; when \( w_j^L, w_j^U > 1 \), the larger the \( w_j^L, w_j^U \) value, the greater the impact of environmental factor.

**Proof.** \( \therefore w_j^L = \frac{\theta^L_j}{\theta^U_j} \), \( w_j^U = \frac{\theta^L_j}{\theta^U_j} \); \( \therefore \theta^L_j = \theta^U_j \times \theta^U_j = w_j^L \times \theta^U_j \).
\[ \therefore \text{when } 0 < w_j^L, w_j^U < 1, \theta_j^L - \theta_j^U = \theta_j^L - \theta_j^U = \left( 1 - w_j^L \right) \theta_j^U, \]
\[ \theta_j^L - \theta_j^U = \theta_j^U - w_j^L \times \theta_j^U = \left( 1 - w_j^U \right) \theta_j^U, \]
\[ \therefore \text{the smaller the } w_j^L, w_j^U \text{ value } \therefore \text{the greater } \left( 1 - w_j^L \right), \left( 1 - w_j^U \right), \]
\[ \therefore \text{the impact of environmental factor is greater; } \]
\[ \therefore \text{when } w_j^L, w_j^U > 1, \theta_j^L - \theta_j^U = \theta_j^L - \theta_j^U = \left( w_j^L - 1 \right) \theta_j^U, \]
\[ \theta_j^L - \theta_j^U = \theta_j^U - \theta_j^U = \left( w_j^U - 1 \right) \theta_j^U, \]
\[ \therefore \text{the larger the } w_j^L, w_j^U \text{ value, the greater } \left( w_j^L - 1 \right), \left( w_j^U - 1 \right), \]
\[ \therefore \text{the impact of environmental factor is greater.} \]

According to Definition 1, by calculating the degree of interval efficiency change vector, it can be found that DMUs have a similar degree of efficiency change before and after the external environmental factors’ influence. If the changes in efficiency are similarly affected by the environment between DMUs, it means that the relationship between the input and output of DMUs and the external environmental factors has a natural similarity. Thus, using these degrees of interval efficiency change vector as the elements can cluster with inherently similar DMUs. And DMU with the highest column mean in a given cluster can be used as the primary benchmark for improvement by other DMUs in that cluster. The specific steps to identify the benchmark through cluster analysis are following as

**Step 1:** Calculate the degree of efficiency change vectors according to Eq. (14).

**Step 2:** According to the calculated vectors by Step 1 as elements, the system clustering is performed by the class average method.

**Step 3:** Find the best efficient DMU in each category obtained by clustering as an improvement benchmark for other DMUs.

### 4. NUMERICAL EXAMPLE

This section aims at showing the three-stage interval DEA model and the improvement benchmarks its superiority, validity, and feasibility. The paper selects all nineteen set of data from reference [34,43], as shown in Table 1. There are four input indicators, two output indicators, and three environmental variable indicators to evaluate the utilization efficiency of power grid equipment [44].

The data we refer to here is exact data. For this set of data, we let input and output data of each DMU increase and decrease with random size of 0% to 10% its value respectively, forming interval input and output data, as shown in Table 2.

4.1. Analysis of the Results for Three-Stage Interval DEA

The first stage uses interval input–output data to evaluate the initial interval efficiency of all DMUs with the three-stage interval DEA model described in the Eqs. (6) and (7). The results obtained are shown in Table 3, which contain several environmental factors and random errors.

From Table 3, some conclusions can be known. There are two DMUs belong to category \( E^+ \), fourteen DMUs belong to category \( E^- \), and three DMUs belong to category \( E^0 \). We can see that before excluding environmental factors and random factors, only DMU2 and DMU15 belong to category \( E^+ \), which means DMU2 and DMU15 are completely efficient. Meanwhile DMU5, DMU6, and DMU9 belong to category \( E^- \) are completely inefficient. The others are all partially efficient.

At this stage, while calculating the maximum and minimum efficiency of each DMU, it also obtains the slack variable value of each input. Further results analysis of these slack variables will be carried out in the second stage.

The dependent variable of the second stage SFA regression analysis is the slack variable corresponding to the input indicator in the first-stage DEA interval efficiency analysis. According to Eqs. (9) and (10), the slack variables are decomposed into a function with three variables of environmental impact factor, random error factor and management inefficiency factor. Each input value of each DMU has its own corresponding two slack variables \( S_i^L \) and \( S_i^U \). We have four input indicators for each DMU, so there are eight regression results. Since each DMU has to be done once, there are nineteen DMUs, here we only show the regression results of DMU1, as shown in Table 4.

It can be seen from Table 4 that the three environment variables have different effects on the slack variables of the four input indicators. Then adjust the original input value according to Eqs. (11–13), to obtain DMU’s new input interval data. In the same way, repeat the above steps to adjust the input of the remaining eighteen DMUs to obtain complete new input interval data. After the second stage of SFA regression analysis, all DMUs have been adjusted to the same external environment and luck level.

In third stage, using the adjusted interval input data and initial output data, the interval DEA efficiency method is used to measure the interval efficiency again. The results are shown in Table 5.

After eliminating environmental factors and random errors, from the perspective of efficiency range, we can see that DMU2, DMU3, and DMU13 belong to category \( E^+ \), which means they are complete efficient. Where DMU7, DMU11, DMU15, DMU16, and DMU17 belong to category \( E^- \) are partial efficient. The others are all inefficient.

We draw the interval efficiency values of the first stage and the third stage in Figure 1. It shows the comparison between two sets interval efficiency obtained by three-stage interval DEA model and directly calculated by the interval DEA method respectively.

From the Figure 1, we can clearly see the overlapping and changing parts of the interval efficiency between the first and third stages. The elimination of environment variables and random errors, DMU2 is
Table 1 | Empirical sample data from 19 cities’ power supply bureaus.

| DMU | Power Supply Bureau | Input Indexes | Output Indexes | Environmental Factors |
|-----|---------------------|---------------|----------------|-----------------------|
|     |                     | Total Length of 500kV Transmission Line/km | Total Length of 220kV Transmission Line/km | Capacity of 500kV Power Substations/tenMVA | Capacity of 220kV Power Substations/ten MVA | The City’s Peak Load on Electricity/tenMW | The City’s Electricity Consumption/Hundred Million kwh | The Total Number of GDP/Hundred Million Yuan | The Resident Population/Ten Thousand | The Power Supply Area/km² |
| 1   | Chaozhou            | 130.002       | 526.159        | 100                    | 240                    | 139.7                        | 75.1                        | 780.34                        | 270                        | 301.58                     |
| 2   | Dongguan            | 526.927       | 1060.703       | 1700.1                 | 2016                   | 1279.3                       | 660.99                      | 5490.02                       | 829.23                      | 2535.64                     |
| 3   | Foshan              | 330.684       | 1302.686       | 1300                   | 1524                   | 1005.5                       | 564.13                      | 7010.17                       | 726.18                      | 3833.41                     |
| 4   | Heyuan              | 212.233       | 675.12         | 100                    | 264                    | 141                          | 73.83                       | 680.33                        | 301.01                      | 9888.23                     |
| 5   | Huizhou             | 1065.831      | 1672.235       | 675                    | 951                    | 485.2                        | 271.97                      | 2678.35                       | 467.4                       | 9455.63                     |
| 6   | Jiangmen            | 1116.784      | 1198.185       | 500                    | 723                    | 384                          | 227.87                      | 2000.18                       | 448.27                      | 9568.9                      |
| 7   | Jieyang             | 286.661       | 768            | 200                    | 387                    | 246.4                        | 158.92                      | 1605.35                       | 595.59                      | 5269                        |
| 8   | Maoming             | 232.46        | 1093.67        | 150                    | 318                    | 140.8                        | 94.77                       | 2160.17                       | 596.76                      | 11458                       |
| 9   | Meizhou             | 604.88        | 1037.617       | 200                    | 267                    | 142.4                        | 76.51                       | 800.01                        | 429.41                      | 16110.81                    |
| 10  | Qingyuan            | 662.687       | 2051.424       | 350                    | 447                    | 284                          | 173.89                      | 7153                          | 376.6                       | 19037                       |
| 11  | Shantou             | 313.098       | 714.164        | 250                    | 534                    | 307.6                        | 174.2                       | 1565.9                        | 544.81                      | 2121.27                     |
| 12  | Shanwei             | 296.12        | 331.309        | 150                    | 114                    | 82.3                         | 43.8                        | 671.75                        | 296.9                       | 4881.2                      |
| 13  | Shaoguan            | 154.082       | 1066.511       | 150                    | 339                    | 230                          | 119.3                       | 1010.07                       | 286.87                      | 18930.5                     |
| 14  | Yangjiang           | 503.13        | 649.69         | 250                    | 210                    | 142.8                        | 90.71                       | 1039.84                       | 247                         | 7813.4                      |
| 15  | Yunfu               | 0             | 542.956        | 0                      | 141                    | 92.1                         | 54.85                       | 602.3                         | 241.65                      | 7777.55                     |
| 16  | Zhanjiang           | 1479.46       | 711.005        | 150                    | 279                    | 161.7                        | 109.47                      | 2060.01                       | 710.92                      | 12922.72                    |
| 17  | Zhaoying            | 107.84        | 628.828        | 350.4                  | 369                    | 239.1                        | 156.24                      | 1660.07                       | 398.23                      | 15205.27                    |
| 18  | Zhongshan           | 364.052       | 781.636        | 500                    | 906                    | 463.7                        | 237.6                       | 2638.93                       | 315.3                       | 3598.28                     |
| 19  | Zhuhai              | 111.42        | 562.962        | 200                    | 592                    | 199.4                        | 134.3224                    | 1662.38                       | 158.26                      | 1711.24                     |

DMU, decision-making unit.
Table 2: The interval input and output data.

| DMU | Power Supply Bureau | Input Indexes | Output Indexes |
|-----|---------------------|---------------|----------------|
|     | Upper Bound | Lower Bound | Upper Bound | Lower Bound | Upper Bound | Lower Bound |
|     | Upper Bound | Lower Bound | Upper Bound | Lower Bound | Upper Bound | Lower Bound | Upper Bound | Lower Bound | Upper Bound | Lower Bound | Upper Bound | Lower Bound |
| 1   | 121.504     | 142.442     | 491.764     | 575.229     | 98         | 107         | 238         | 259         | 132.5       | 140.9       | 73.65       | 81.83       |
| 2   | 477.621     | 551.054     | 1043.359    | 1158.404    | 1637       | 1843        | 1817        | 2029        | 1187.2      | 1406.7      | 654.44      | 663.91      |
| 3   | 322.732     | 355.945     | 1199.167    | 1377.902    | 1205       | 1374        | 1381        | 1527        | 969.9       | 1103.2      | 532.69      | 607.71      |
| 4   | 196.118     | 227.952     | 645.412     | 692.512     | 98         | 110         | 246         | 285         | 136.1       | 153.5       | 71.53       | 75.15       |
| 5   | 986.566     | 1077.120    | 1546.492    | 1710.474    | 657        | 737         | 900         | 1035        | 463.1       | 505.3       | 262.75      | 277.69      |
| 6   | 1040.668    | 1168.520    | 1190.494    | 1290.125    | 489        | 519         | 658         | 768         | 375.6       | 388.8       | 216.25      | 248.52      |
| 7   | 820.579     | 895.485     | 716.452     | 822.928     | 198        | 213         | 382         | 395         | 238.8       | 264.3       | 148.93      | 160.53      |
| 8   | 213.315     | 236.528     | 1023.450    | 1139.500    | 247        | 272         | 312         | 348         | 129.8       | 150.6       | 91.07       | 95.29       |
| 9   | 594.986     | 645.164     | 997.071     | 1122.301    | 208        | 231         | 256         | 278         | 142.3       | 154.4       | 72.67       | 79.81       |
| 10  | 603.417     | 696.919     | 1986.306    | 2218.521    | 509        | 530         | 405         | 463         | 257.8       | 305.9       | 156.54      | 188.00      |
| 11  | 291.097     | 317.907     | 657.811     | 775.030     | 234        | 251         | 412         | 462         | 306.3       | 319.2       | 165.74      | 189.78      |
| 12  | 267.886     | 312.137     | 314.557     | 352.369     | 136        | 162         | 103         | 122         | 76.5        | 88.3        | 43.20       | 45.51       |
| 13  | 143.609     | 154.645     | 965.097     | 1113.860    | 139        | 162         | 305         | 365         | 224.8       | 236.2       | 108.24      | 130.25      |
| 14  | 462.417     | 540.795     | 645.791     | 706.002     | 240        | 265         | 203         | 224         | 133.2       | 149.6       | 84.24       | 96.32       |
| 15  | 0.000       | 0.000       | 508.685     | 562.235     | 0          | 0           | 138         | 145         | 86.4        | 94.3        | 52.97       | 59.98       |
| 16  | 143.125     | 156.031     | 640.118     | 726.944     | 146        | 154         | 260         | 294         | 158.8       | 175.1       | 108.10      | 117.47      |
| 17  | 103.538     | 112.316     | 587.800     | 666.872     | 335        | 358         | 354         | 391         | 220.8       | 261.4       | 146.14      | 169.26      |
| 18  | 357.472     | 373.349     | 751.368     | 792.750     | 460        | 549         | 838         | 959         | 458.7       | 472.1       | 228.14      | 255.42      |
| 19  | 111.191     | 121.712     | 561.547     | 586.669     | 199        | 211         | 559         | 627         | 197.4       | 209.2       | 123.10      | 138.65      |

DMU, decision-making unit.
Table 3. First-stage interval efficiency and category.

| DMU | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| $h_0^U$ | 0.709 | 1.000 | 0.942 | 0.652 | 0.576 | 0.637 | 0.850 | 0.624 | 0.628 | 0.738 |
| $h_0^L$ | 1.000 | 1.000 | 1.000 | 1.000 | 0.894 | 1.000 | 1.000 | 0.814 | 0.933 | 1.000 |
| Category | E | E+ | E | E | E | E | E | E | E | E |

| DMU | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|-----|----|----|----|----|----|----|----|----|----|
| $h_0^U$ | 0.932 | 0.796 | 0.829 | 0.793 | 1.000 | 0.807 | 0.844 | 0.798 | 0.785 |
| $h_0^L$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| Category | E | E+ | E | E | E+ | E | E | E | E |

DMU, decision-making unit.

Table 4. SFA regression results.

| Indicator | $s_{11}^U$ | $s_{12}^U$ | $s_{13}^U$ | $s_{14}^U$ | $s_{11}^L$ | $s_{12}^L$ | $s_{13}^L$ | $s_{14}^L$ |
|-----------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| $\hat{\beta}_{0}^U$ | -212.86 | -78.98 | 0.00 | 0.00 | -212.86 | -157.38 | -137.50 | 0.00 | 0.00 |
| $\hat{\beta}_{1}^U$ | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| $\hat{\beta}_{2}^U$ | 0.15 | -0.27 | 0.00 | 0.00 | 0.20 | 0.16 | 0.00 | 0.00 | 0.00 |
| $\hat{\beta}_{3}^U$ | 0.01 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| $\gamma$ | 1.00 | 0.26 | 0.10 | 0.10 | 1.00 | 1.00 | 0.10 | 0.10 |

SFA, Stochastic Frontier Analysis.

Table 5. Third-stage interval efficiency and category.

| DMU | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| $h_0^U$ | 0.580 | 1.000 | 1.000 | 0.576 | 0.700 | 0.753 | 0.941 | 0.604 | 0.520 | 0.794 |
| $h_0^L$ | 0.828 | 1.000 | 1.000 | 0.768 | 0.736 | 0.832 | 1.000 | 0.680 | 0.624 | 0.929 |
| Category | E | E+ | E+ | E | E | E | E | E | E | E |
| Average | 0.704 | 1.000 | 1.000 | 0.672 | 0.718 | 0.793 | 0.971 | 0.642 | 0.572 | 0.862 |

| DMU | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|-----|----|----|----|----|----|----|----|----|----|
| $h_0^U$ | 0.908 | 0.491 | 1.000 | 0.645 | 0.633 | 0.887 | 0.821 | 0.934 | 0.783 |
| $h_0^L$ | 1.000 | 0.625 | 1.000 | 0.777 | 1.000 | 1.000 | 1.000 | 0.947 | 0.969 |
| Category | E | E | E | E | E | E | E | E | E |
| Average | 0.954 | 0.558 | 1.000 | 0.711 | 0.817 | 0.944 | 0.911 | 0.941 | 0.876 |

DMU, decision-making unit.

still complete efficient, indicating that it is not affected by environmental factors and luck. DMU15 has changed from complete efficient to partial efficient; DMU1, DMU4, DMU6, DMU10, DMU12, DMU14, DMU18, and DMU19 has changed from partial efficient to complete inefficient, indicating that its efficiency before adjustment is overestimated, the reason is that it is greatly affected by favorable environmental factors and luck. The efficiency value of the other DMUs has increased, indicating that the efficiency value is lower before adjustment, the reason is that it is affected by different degrees of adverse environmental factors and luck.

At the same time, we rank the interval efficiency values of the third stage by taking the average of the upper and lower bounds of the interval as the criterion show in Table 5. The efficiency value by the traditional three-stage DEA model could be obtained from reference [34,43]. Comparing the ranking result with three-stage interval DEA model and the traditional three-stage DEA model in Table 6. By calculating the correlation coefficient, the fit of these two sets of rank results reaches 0.875. It shows that the overall trends of the two rank results are basically consistent.

4.2. Analysis of the Results for Identification of Benchmarks

Step 1: According to Eq. (14), the degree of efficiency change vector $(w^E_j, w^U_j)$ can be obtained, as shown in Table 7.

Step 2: According to the calculated vectors by Step 1 as elements, the system clustering is carried out according to the class average method, and the hierarchical diagram of the clustering results for nineteen DMUs is shown in Figure 2.

Step 3: It can be seen from Figure 2 that a total of five clusters are identified in this analysis. The best performing DMU in each cluster is used by the other DMUs in these clusters as the main benchmark for improvement. For example, DMU2, DMU3, DMU13, DMU7, DMU10, DMU11, DMU16, DMU19, and DMU19 in the same cluster, its efficiency value changes similarly to the environmental impact. Among them, DMU2, DMU3, and DMU13 belong to a complete efficient set, so it is considered to be the best performing DMU of its cluster. It can be regarded as an improvement benchmark by
other DMUs of its cluster, and other DMUs can be adjusted and improved according to its external environment to improve self-efficiency value.

The traditional improvement benchmark directly means that all noncompletely effective DMUs are improved according to the set of completely effective DMUs. Compared with this, the method proposed in this paper clusters DMUs with similar relationships into one category and finds improvement benchmark in each category, which is more scientific.

4.3. Discussions

From the numerical examples, the main novelty and advantages of our proposed method are summarized as follows:

1. The proposed three-stage interval DEA model provides a novel way to deal with the interval input and output data. In the proposed method, the reasonable and effective way of coping with the effects of excluding external environmental factors and random errors is proposed. This is the distinct superiority and difference between the proposed method and other DEA models.

2. The improved benchmark is considered in the proposed method, its influence, and importance has been illustrated through the provided numerical examples. It can cluster DMUs which are naturally similar to provide improvement targets for poorly performing DMUs.

Like each coin has two sides, except for the aforementioned advantages, the proposed method has limitations in current version, i.e., it does not consider the preferences of decision makers during the decision process. Actually, the preferences are quite common in our daily life, which are practical and inevitable issues in the real-world situation, particularly under risk and uncertain environment. Although it is a limitation in the proposed method, it is one of the promising and solid future research directions, which can make the decision further close to the real-world situation.
5. CONCLUSIONS AND FUTURE WORKS

The traditional three-stage DEA model can evaluate the relative efficiency of a group of DMUs with multiple inputs and outputs. It can also consider the impact of external environmental factors and random errors on efficiency measurement. However, this method cannot perform efficiency measurement where the input and output data are interval data. At the same time, the interval DEA model can deal with the situation where the input and output data are interval data, but the method has not yet considered the effects of excluding external environmental factors and random errors. Regarding aforementioned limitations, we have provided a three-stage interval DEA model based on three-stage DEA model together with interval DEA model. It has been compared from different perspectives with the traditional three-stage DEA model and has shown a better performance on managing input–output values as interval data and more reliable decision results comparing with the traditional interval DEA model that do not exclude the influence of environmental factors on efficiency measurement.

In addition, from the perspective of environmental factors, this article combines the three-stage interval DEA model with clustering techniques to cluster naturally similar DMUs into one category, and provides an improved benchmarks for poorly performing DMUs. Compared with the previously improved reference unit, the improved benchmark proposed by this method considers the influence of environmental factors, and provides a more easily achieved goal for DMUs with poor performance. Finally, examples illustrate the advantages, potentials, and applications of the model proposed in this paper.

Based on the analysis on this study, it is found that the proposed method not only improves the current studies, but also implies several promising and solid future research directions, i.e., (1) at present, only the case is considered where the input and output data is interval data. Where the environmental data is interval data, the case is worthy of our in-depth study. (2) This study considers the input–output indicators of DMUs as the type of interval data and explores the three-stage interval DEA problem. But the research on different data types needs to be extended to ordinal data and even bounded ratio data. (3) Making fewer changes to maximize the efficiency improvement is also our future research direction. It is a promising future research direction, which can enable the decision models and methodologies close to the real-world situation and easy to be accepted by decision maker and experts.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Guo-Qing Cheng: Conceptualization, Data curation, Formal analysis, Writing—original draft, Writing—review & editing. Liang Wang: Conceptualization, Data curation, Formal analysis, Writing—original draft, Writing—review & editing. Ying-Ming Wang: Conceptualization, Data curation, Formal analysis, Writing—original draft, Writing—review & editing.

ACKNOWLEDGMENTS

This work was partly supported by the National Natural Science Foundation of China (Project No. 71901071, 61773123).

REFERENCES

[1] A. Charnes, W.W. Cooper, E. Rhodes, Measuring the efficiency of decision making units, Eur. J. Oper. Res. 2 (1978), 429–444.
[2] J.T. Bastos, Y.J. Shen, E. Hermans, et al., Bootstrapping DEA scores for road safety strategic analysis in Brazil, Int. J. Comput. Intell. Syst. 8 (2015), 29–38.
[3] Ö. Coşgun, G.O. Kaya, Data envelopment analysis application in Turkish energy market, Int. J. Comput. Intell. Syst. 7 (2014), 636–649.
[4] X.L. Geng, X.M. Gong, X.N. Chu, Component oriented remanufacturing decision-making for complex product using DEA and interval 2-tuple linguistic TOPSIS, Int. J. Comput. Intell. Syst. 9 (2016), 984–1000.
[5] J. Zheng, Y.M. Wang, K. Zhang, A case retrieval method combined with similarity measurement and DEA model for alternative generation, Int. J. Comput. Intell. Syst. 11 (2018), 1123–1141.
[6] H.O. Fried, C.A.K. Lovell, S.S. Schmidt, et al., Accounting for environmental effects and statistical noise in data envelopment analysis, J. Prod. Anal. 17 (2002), 157–174.
[7] J.P. Guo, Y.H. Wu, Evaluation on the decision making units in interval DEA, Syst. Eng. Theory Methodol. Appl. 13 (2004), 339–342.
[8] N. Sun, Z.X. Ma, S.Y. Ma, Research on efficiency of generalized super-efficiency interval DEA, Math. Pract. Theory. 44 (2014), 175–180.
[9] M. Feng, X.Y. Li, Evaluating the efficiency of industrial environmental regulation in China: a three-stage data envelopment analysis approach, J. Clean. Prod. 242 (2020), 118535.
[10] P.P. Xia, J. Wu, X. Ji, P.F. Xi, A DEA-based empirical analysis for dynamic performance of China’s regional coke production chain, Sci. Total Environ. 717 (2020), 136890.
