A New Weighted Restriction Method for Data Envelopment Analysis

Asama Supanimitcharoenporn, Siraprapa Manomat, and Orathai Polsen

Abstract—Data Envelopment Analysis (DEA) is one method to evaluate the relative efficiency of Decision Making Units (DMUs). One of the drawbacks of this method is optimization becoming extremely weighted referring to advantaged inputs or outputs becoming more disadvantaged in terms of DMUs ranking. Since yielding many numbers of DMUs as efficient. The main purpose of this study is to overcome the above problem by using weight restriction. In this study, the author developed a new absolute weight restriction method using Chebyshev’s inequality to bound input-output weights to achieve this goal. Also, the performance of the proposed method was examined and compared to the modified CVDEA.

Keywords—Data Envelopment Analysis, weight restriction, absolute weights restriction, Chebyshev’s inequality

I. INTRODUCTION

Data Envelopment Analysis (DEA) is a mathematical programming technique to evaluate the relative efficiency of Decision Making Units (DMUs) with multiple inputs and outputs. DEA has been widely used and applied in many areas. For example, in engineering, it is useful to compare suppliers for selection, see [1]-[4]. In logistics, it is interesting to evaluate the efficiency of enterprises, see [5], [6]. Moreover, in industry, it is often of interest to evaluate an efficiency of machinery for performance improvement, and rank assessment, see [7].

The DEA method was developed by [8], they used frontier analysis principles which are a concept of [9]. Efficiency scores calculated from the ratio of the weighted sums of inputs to the weighted sums of outputs. The DEA obtains the optimal weights for all inputs and outputs for a nonlinear optimization model with the constraints of lower and upper bounds on weights. This method attaches the constraint of lower and upper bound input-output weights in the DEA model. After that [11],[12] divided weight restriction into 3 groups as follows assurance region of type I, assurance region of type II, and absolute weights restrictions which valuator must determine the bounds on weights. This method uses the bounds on weights considered by human value judgement.

Weight restriction using coefficient of variation of inputs and outputs for a nonlinear optimization model was suggested by [13]. Later [14] modified the model of [13] called Modified CVDEA (MO-CVDEA) which continued to use coefficient for variation to weight restriction. This method was developed for the betterment of dispersion of input-output weights based on the minimization of coefficient of variation. The drawback of this method was the nonlinear optimization model which cause the problem to the construction of dual problem.

In this study, the author has developed an absolute weights restriction method using Chebyshev’s inequality. This method is conducted without human interaction referring to bounds on weights. As a result, the weights do not become zero, the relative efficiency of DMUs can be ranked. Moreover, this model suitable for linear optimization.

II. DATA ENVELOPMENT ANALYSIS

Assuming that there are n DMUs each with m inputs and s outputs, the relative efficiency of a particular DMU \( k \) \((k \in \{1, 2, \ldots, n\})\) is obtained by solving the following fractional programing problem.

Max \( \theta_k = \frac{\sum_{j=1}^{s} u_j Y_{jk}}{\sum_{i=1}^{m} v_i X_{ik}} \)

s.t \( \sum_{i=1}^{m} u_i Y_{ij} \leq 1 \), \( j = 1, 2, \ldots, n \) \( (1) \)

\( \sum_{j=1}^{s} v_j X_{ij} \)

\( u_i \geq 0 \), \( r = 1, 2, \ldots, s \)

\( v_j \geq 0 \), \( i = 1, 2, \ldots, m \),

where \( j \) is the DMU index, \( j = 1, \ldots, n \), \( r \) is the output index, \( r = 1, \ldots, s \), \( i \) is the input index, \( i = 1, \ldots, m \). \( Y_{ij} \) is the value of the \( r \)th output for the \( j \)th DMU, \( X_{ij} \) is the value of the \( r \)th input for the \( j \)th DMU, \( u_i \) is the weight given to the \( r \)th
output, $v_i$ is the weight given to the $i$th input, and $\theta_k$ is the relative efficiency of DMU$_k$, the DMU under evaluation or the target DMU. In this model, $\theta_k = 1$ means DMU$_k$ is efficient.

This fractional program can be converted into a linear programing where the optimal value of the objective function indicates the relative efficiency of DMU$_k$. The reformulated linear programing, also known as the Constant Returns to Scale (CRS) model, is as follows:

Max $\theta_k = \sum_{i=1}^{m} u_i Y_{ik}$

s.t $\sum_{i=1}^{m} v_i X_{ik} = 1$ (2)

$\sum_{i=1}^{m} u_i Y_{ij} - \sum_{i=1}^{m} v_i X_{ij} \leq 0$, $j = 1,2,...,n$

$u_i \geq 0$, $r = 1,2,...,s$

$v_i \geq 0$, $i = 1,2,...,m$

III. MODIFIED MODELS BASED ON COEFFICIENT OF VARIATION

A modified model was suggested by [14]. Basically, this method uses coefficient of variation of weights into optimization objective which minimizing coefficient of variation of weights.

Let $\overline{v}_i = \frac{1}{n} \sum_{j=1}^{n} v_{ij}$ and $\overline{\eta}_j = \frac{1}{n} \sum_{i=1}^{n} \eta_{ij}$ are optimal weight of input and output variables respectively which are obtained from (2). The following model, which is a nonlinear optimization model, is based on the CRS model. In this model, they considered the differences of weight and its mean as follows:

Max $\sum_{i=1}^{m} u_i Y_{ik} = \frac{\sum_{i=1}^{m} (u_i - \overline{u}_i)^2}{s-1} - \frac{\sum_{i=1}^{m} (v_i - \overline{v}_i)^2}{m-1}$

s.t $\sum_{i=1}^{m} v_i X_{ik} = 1$ (3)

$\sum_{i=1}^{m} u_i Y_{ij} - \sum_{i=1}^{m} v_i X_{ij} \leq 0$, $j = 1,2,...,n$

$u_i \geq 0$, $r = 1,2,...,s$

$v_i \geq 0$, $i = 1,2,...,m$

where $u_{ik}$ is the weight given to the $r$th output of $k$th DMU, $v_{ik}$ is the weight given to the $i$th input of $k$th DMU.

The optimal weights from model (2) are used to evaluate the efficiency score ($W_k$) of DMU$_k$, as follows:

$W_k = \sum_{i=1}^{m} u_i Y_{ik}$

where $W_k = 1$ means DMU$_k$ is efficient.

IV. ABSOLUTE WEIGHTS RESTRICTION USING CHEBYSHEV’S INEQUALITY

This study developed an absolute weights restriction by using Chebyshev’s inequality to bound the weights, called CIDEA. The bounds of weights based on the CRS model defined as follows:

Max $Z_k = \sum_{i=1}^{m} u_i Y_{ik}$

s.t $\sum_{i=1}^{m} v_i X_{ik} = 1$

$\sum_{i=1}^{m} u_i Y_{ij} - \sum_{i=1}^{m} v_i X_{ij} \leq 0$, $j = 1,2,...,n$

$\delta_i \leq v_i \leq \tau_i$

$\rho_r \leq u_r \leq \eta_r$

$u_r \geq 0$, $r = 1,2,...,s$

$v_i \geq 0$, $i = 1,2,...,m$

where $\delta_i$ and $\tau_i$ are lower bounds, and upper bounds of weight of $i$th inputs respectively. $\rho_r$ and $\eta_r$ are lower bounds, and upper bounds of weight of $r$th outputs respectively.

The $\delta_i$ and $\tau_i$ are calculated from $\overline{v}_i = \frac{1}{n} \sum_{j=1}^{n} v_{ij}$ and $sd(V_i) = \sqrt{\frac{\sum_{j=1}^{n} (v_{ij} - \overline{v}_i)^2}{n-1}}$. Also, the $\rho_r$ and $\eta_r$ are calculated from $\overline{\eta}_r = \frac{1}{n} \sum_{i=1}^{n} \eta_{ij}$ and $sd(U_r) = \sqrt{\frac{\sum_{i=1}^{n} (u_{ij} - \overline{\eta}_r)^2}{n-1}}$, $v_{ij}$ and $u_{ij}$ are optimal weights of input and output variables respectively which are obtained from (2), and $l$ is a constant from Chebyshev’s inequality for each weight. Let $V_i$ is a random variable represents the $i$th input weight, then

$P(|V_i - \overline{v}_i| \geq l \cdot sd(V_i)) \leq \frac{1}{l^2}$

So that

$P(\overline{v}_i - l \cdot sd(V_i) \leq V_i \leq \overline{v}_i + l \cdot sd(V_i)) \geq 1 - \frac{1}{l^2}$

The above expression is used to compute the bound of weights in the case of inputs. If the lower bound weight is less than zero, it will be defined as $10^{-5}$, which was suggested by [15]. The output bounds can be calculated in the same way.

The optimal weights in model (4) as $Z_k$ is the relative efficiency of DMU$_k$. In this model, $Z_k = 1$ means DMU$_k$ is efficient.
V. NUMERICAL EXAMPLE

In this study a numerical example was used to compare the CIDEA with the MO-CVDEA approach. The purpose is to study the dispersion of weights and find out the best method for ranking DMUs.

Numerical Example. The input and output variables are drawn randomly from Uniform [10,200] distribution with three inputs and three outputs variables for ten DMUs (see Table I).

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>DATA OF TEN DMUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMU</td>
<td>( x_1 )</td>
</tr>
<tr>
<td>1</td>
<td>52</td>
</tr>
<tr>
<td>2</td>
<td>89</td>
</tr>
<tr>
<td>3</td>
<td>58</td>
</tr>
<tr>
<td>4</td>
<td>135</td>
</tr>
<tr>
<td>5</td>
<td>182</td>
</tr>
<tr>
<td>6</td>
<td>132</td>
</tr>
<tr>
<td>7</td>
<td>98</td>
</tr>
<tr>
<td>8</td>
<td>152</td>
</tr>
<tr>
<td>9</td>
<td>162</td>
</tr>
<tr>
<td>10</td>
<td>178</td>
</tr>
</tbody>
</table>

The data in Table I was analyzed by using the DEA model (2) and the results are shown in Table II and Table III. It can be seen in Table II that there are seven DMUs as efficient identifies DMU 2, DMU 3, DMU 4, DMU 5, DMU 7, DMU 8, and DMU 10. Table III presents that the DEA model obtains a high number of zero optimal weights, showing that it does not emphasize input and output variables.

<table>
<thead>
<tr>
<th>TABLE II</th>
<th>EFFICIENCY SCORES OF THE DEA MODEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMU</td>
<td>1</td>
</tr>
<tr>
<td>Efficiency</td>
<td>0.7194</td>
</tr>
</tbody>
</table>

In Table IV, the efficiency scores of CIDEA models (the bounds on weight at \( l = 0.4 \), 0.45, 0.5, 0.6, 0.7 and 1) and MO-CVDEA models are shown. Table IV shows that four DMUs are efficient in the bounds at \( l = 0.4 \) of CIDEA model and six DMUs are efficient in the MO-CVDEA model. Therefore, the CIDEA model (\( l = 0.4 \)) is relatively better in ranking DMUs than the MO-DEA model and the worst case of the CIDEA model (\( l = 1 \)) is still able to show indifferent ranking with the MO-CVDEA model. The CIDEA model at \( l = 1 \) was the worst case and at \( l = 0.4 \) was the best case of the CIDEA model.

In Table V, it is seen that the weight of the 3th input variable of DMU 6, DMU 7, and DMU 10 in MO-CVDEA model are zero. It means this model still does not emphasize that input.
Tables VIII and IX show that there are high correlations between the efficiency scores of DMUs and hence a similarity between the ranks obtained by the CIDEA and the DEA models for each case.

Next, the comparison of weight variances corresponding to the CIDEA models and the MO-CVDEA models which the following hypotheses were tested at significance level 0.05 (α = 0.05). (see Table X)

Table X and XI show the results of the homogeneity test. The comparison of the weight variances corresponding to the CIDEA models and the MO-CVDEA models which the following hypotheses were then tested. (see Table XI)

Table VI and Table VII show the weights of input and output variables of the CIDEA models (l = 1 and l = 0.4), indicating the weights of input and output variables are not zero.

For each model, using the following hypotheses, it is proven that the two approaches can obtain the similar rankings to DMUs.

- **H₀**: The DMU efficiency scores for the CIDEA model are uncorrelated to the DMU efficiency scores for the DEA model.
- **H₁**: The DMU efficiency scores for the CIDEA model are correlated to the DMU efficiency scores for the DEA model.

In testing the hypotheses, the Spearman and Kendall correlations were used (see Table VIII and IX).
In summary, the CIDEA approach attaches importance to all input-output variables while the DEA and the MO-CVDEA ignored some variables. From the results of efficiency scores, it is clear to see the CIDEA approach can reduce the number of efficient DMUs better than the MO-CVDEA approach.

VI. CONCLUSION

The CIDEA models were used to overcome the problem of DEA. Since the DEA models yielded several zero input-output weights or provided extreme values, this lead to many efficient DMUs not becoming ranked. Several advantages of this model over the MO-CVDEA were discovered. Firstly, the efficiency scores can be calculated within 2 stages instead of 3 iteration stages. Secondly, since the input-output weights of this model do not become zero, this model emphasize all input-output variables. Moreover this model is linear optimization model so it can be successfully formulated to handle dual problems.

ACKNOWLEDGMENT

The authors are grateful to the anonymous referees for valuable comments and suggestions.

REFERENCES