Data Envelopment Analysis with Functional Data using Preference Method

Syyed Nima Hashemi Ghermezi
Department of Industrial Engineering, Faculty of Engineering, Islamic Azad University, Saveh Branch, Saveh, Iran

Taher Taherian
Department of Industrial Engineering, Faculty of Engineering, Islamic Azad University, Saveh Branch, Saveh, Iran

Farhad Hosseinzadeh Lotfi
Department of Mathematics, Science and Research Branch, Islamic Azad University, Tehran, Iran

ABSTRACT
In this paper we assess the efficiency of units using data envelopment analysis (DEA). In first stage, a functional data is converted to a fuzzy bell shape number and then a benchmark point would be chosen for each input or output and using the preference ratio method, the equivalence multiplier of each data would be calculated. In order to simplification of functional data, functional data will be replaced by the equivalence multiplier.

Keywords
Data envelopment analysis (DEA) ,fuzzy bell shape, Preference ratio.

1. INTRODUCTION
One of the ranking methods of fuzzy numbers is the preference ratio method that is presented by Modarres and Sadi-Nazhad(2001) [1]. Suppose A and B as two fuzzy numbers. S(A) and S(B) are defined as the support of A and B, separately and the supremum and infimum of S(A) and S(b) are shown as LA, UA, LB, UB, respectively. The \( \Omega \) is defined below:

\[ \Omega = [\min(LA, LB), \max(UA, UB)] \]

And for \( \theta \in \Omega \), the preference function is defined as follows:

\[ G_k(\theta) = \int \mu_k(x)dx \]

\( k = A, B \)

\[ p(\theta), \] is also defined as follows:

\[ p(\theta) = \begin{cases} A & G_A(\theta) \leq G_B(\theta) \\ B & G_A(\theta) > G_B(\theta) \end{cases} \]

In other words, \( \theta \) belongs to a fuzzy number whichits area ratio under its membership function and after the point \( \theta \), to the total area under the membership function curve is more.

According the previous comments, the preference ration of A and B are defined as follows:

\[ R(A) = \left[ \frac{\Omega_A}{\Omega_A^2} ; \Omega_A = \{ \theta \in \Omega | p(\theta) = A \} \right] \]

\[ R(B) = \left[ \frac{\Omega_B}{\Omega_B^2} ; \Omega_B = \{ \theta \in \Omega | p(\theta) = B \} \right] \]

Which the symbol || shows the interval lengths.

In other words, the fuzzy number with maximum distance from \( \Omega \) will gain the upper rank.In order to rank the fuzzy numbers with the count more than two, Modarresand Sadi-Nazhad (2005) [2] developed an approach wherein all fuzzy numbers compared with a benchmark is investigated. They defined a benchmark for a set of triangular fuzzy numbers and compared each preference ratio of the numbers with them. In order to this comparison, they defined the equivalence and equivalence multipliers follows:

In preference method, two fuzzy number A and B are equivalent if \( R(A) = R(B) = 0.5 \) and in this case, \( A = B \). The equivalence multiplier is also defined as follows:

\( K \) is called the equivalence multiplier of A toward B, whenever \( KA = B \).

Nowadays, DEA is also one of applicable ways to calculate the relation efficiency of units.

Charnes et al (1978) [3] presented the CCR model for first time. The constant return to scale is considered in this model. In other words, increasing the inputs caused increasing the outputs with the same ratio. Banker et al (1984)[4] presented the BCC method wherein the variable return to scale was considered. This means that, any changes in inputs don’t cause the changes in outputs with the same ratio.Since that time until now, data envelopment analysis is used and developed in many centers and organizations like banks, schools, universities, hospitals, insurance agencies, and factories [5-14]. Ranking the units was another development of DEA that many papers have written about it. Andersen and Petersen (1993) [15] presented a procedure for ranking efficient units. They recalcuated the efficiency of efficient units but this time in order to calculate the efficiency of a particular unit, that unit has been eliminated from production possibility set (PPS).

Another advance in DEA is bringing some changes about in type of inputs and outputs. Cook et al (1996) [16-17] presented DEA with ordering data and then in 1997 they considered some inputs sequentially. Cooper et al (1999) [18-19] investigated the interval data in DEA and presented a comprehensive form using the confidence interval concept and converted it to an equivalent linear programming through a set of scale conversions and change of variables and the efficiency value of each decision maker unit obtained from this model was deterministic, less or equal to 1.

helped to calculate the interval efficiency for each DMU in pessimistic and optimistic viewpoint. Their model was applicable for deterministic data at the beginning but they developed their model to be included of interval and fuzzy data. The deficiency of their model was considering just one input and output and their model was using different efficient boundaries to measure efficiency intervals of different DMUs.

Guo and Tanaka (2001)[24], Kahraman and Tolga (1998)[25], Kao and Liu (2000)[26] introduced DEA with fuzzy data. In most important solution, they converted the fuzzy data into interval data per different α cut.

2. MODEL DESCRIPTION

Assume n units as each unit has m inputs and n outputs, being probing and calculating their efficiency. Each unit of input of output has some functional data as each data give different values of dependent variable or response variable for different values of dependent variable. So practically, the data are presented as a set of order pair. Here we present two assumptions. First, the values of independent variable are equal in homologous input or output data. The second, the weight and importance of these values are equal. For example, suppose that the data is evaluated in different periods with symmetric importance. Applying mentioned assumptions, we can just consider the response variables, so the data practically will be a set of single component numbers obtained in different periods. Now the purpose is calculating of such units. Hereinafter, these data will be called, period data or period input and output.

3. CONVERTING A PERIOD DATA TO A FUZZY DATA

In this section, we want to convert a period data to a bell shape fuzzy data. Thetypes of data are numbers obtained from different periods while the results originated from different periods may be different and in this situation, it sounds that the data may be irrational numbers so functional data could be converted to a fuzzy number by defining a value function. It’s clear that each data is a quantitative value of a quality in different periods. It’s clear that, in a period data, each number which demonstrates a quality with more reality will have more values and the average of a period data must have the maximum value and whatever is moved away from the mean the value assigned to the number will be less value. Here the bell shape fuzzy number could be used for each period data that is very similar to a normal distribution with a difference that the area under the distribution curve is equal to √2πσ². Assume m as the mean and σ as the standard deviation of a period data, the bell shape function will be defined as χ(x) = exp(-(x-m)²/2σ²), but the problem is that the numbers located out of the data area will also have a value even if this value may be belittle. So this function will change as follows:

\[ μ(x) = \begin{cases} 0 & x ≤ a \\ \text{EXP}\left(-\frac{(x - m)^2}{2\sigma^2}\right) & a ≤ x ≤ b \\ 0 & x ≥ b \end{cases} \]

Here, a and b are the minimum and maximum value of each period data respectively. Hereinafter, the mentioned function is called, restricted bell shape (RBS) function and it will be shown with (a, m, b, σ).

4. EQUIVALENCE MULTIPLIER CALCULATION

If we want to compare some fuzzy numbers, it may sound that a pairwise comparison can be useful but it’s not practical because being greater in preference method, doesn’t have transitive property and this means if we want to compare three fuzzy numbers, A, B and C, if A is greater than B and B is greater than C, A may be less than C to be obvious, there is a counterexample. Assume that A, B and C are three RBS fuzzy numbers:

A = (0.5000, 1.0000, 4.0000, 0.4000)
B = (0.5981, 0.9968, 3.7879, 0.3987)
C = (0.5490, 0.9982, 3.4000, 0.4000)

<table>
<thead>
<tr>
<th>R(A)</th>
<th>A &gt; B</th>
</tr>
</thead>
<tbody>
<tr>
<td>R(B)</td>
<td>B &gt; C</td>
</tr>
<tr>
<td>R(C)</td>
<td>A &lt; C</td>
</tr>
</tbody>
</table>

As it can be seen, A is greater than B and B greater than C but A is not greater than C so the transitive property can’t be confirmed. To solve this problem, we compare all fuzzy numbers with a benchmark number and the equivalence multiplier of the benchmark will be calculated with respect to all numbers.

Now two algorithms are being presented. First algorithm, calculates the preference ratio of two RBS number and the second algorithm calculates the equivalence multiplier in respect of one RBS number.

Notice that, in first algorithm, φ(x) shows the cumulative normal distribution that help us to calculate the integral of RBS membership function because it’s obvious that if (L, M, U, σ) is a RBS number we have:

\[ \int_L^U \mu_{\text{RBS}}(x) \, dx = \sqrt{\frac{2}{\pi \sigma^2}} \left[ \phi\left(\frac{U-M}{\sigma}\right) - \phi\left(\frac{X-M}{\sigma}\right) \right] \]

\[ G_{\text{RBS}}(X) = \frac{\int_L^U \mu_{\text{RBS}}(x) \, dx}{\int_U^L \mu_{\text{RBS}}(x) \, dx} = \frac{\phi\left(\frac{U-M}{\sigma}\right) - \phi\left(\frac{X-M}{\sigma}\right)}{\phi\left(\frac{U-M}{\sigma}\right) - \phi\left(\frac{L-M}{\sigma}\right)} \]

Algorithm 1:
Assume two fuzzy numbers A and B.
A = (L_A, M_A, U_A, σ_A) and B = (L_B, M_B, U_B, σ_B)

Step 1:
Put \( L = \min \{L_A, L_B\} \), \( U = \max \{U_A, U_B\} \), \( e = \frac{U - L}{100000} \), \( i = L \), \( G(A) = 0 \), \( G(B) = 0 \).

Step 2:
If \( L \geq U \) go to step 6 else go to next step
Step 3:
If \( i \leq L_A \) then put \( A = 1 \)
If \( i > U_B \) then put \( A = 0 \)
If \( L_A < i < U_A \) put: \( A = \frac{[\psi((U_A-M_A)/\sigma_B)-\psi((i-M_A)/\sigma_B)]}{[\psi((U_A-M_A)/\sigma_B)-\psi((L_A-M_A)/\sigma_B)]} \)
Step 4:
If \( i \leq L_B \) then put \( B = 1 \)
If \( i > U_B \) then put \( B = 0 \)
If \( L_B < i < U_B \) put: \( B = \frac{[\psi((U_B-M_B)/\sigma_B)-\psi((i-M_B)/\sigma_B)]}{[\psi((U_B-M_B)/\sigma_B)-\psi((L_B-M_B)/\sigma_B)]} \)
Step 5:
If \( A > B \) then put \( G(A) = G(A) + e \)
If \( A = B \) then put \( G(A) = G(A) + \frac{e}{2} \), \( G(B) = G(B) + \frac{e}{2} \)
Add \( e \) units to \( I \) and go to step 6.
Step 6:
\[ R(A) = \min \left( \frac{G(A)}{U - L} \right), \quad R(B) = \min \left( \frac{G(B)}{U - L} \right) \]

If \((x, y, z, \sigma)\) is a RBS number, we define:
\[ k(x, y, z, \sigma) = (kx, ky, kz, k\sigma) \]
It’s true, because when all of data are multiplied by \( k \), the standard deviation will also be multiplied by \( K \) when \( K \) is a positive number. Assume that \( A \) and \( B \) are two RBS numbers and it’s wanted to calculate \( A \) as \( KA = B \), the prerequisite for equality of \( A \) and \( B \) is their overlaps.

If you attend the fig 2, you will consider that two numbers have overlaps and may be they are equal.
Assume \( R(A) > R(B) \) and \( KA = B \), so if \( K\sigma_A \geq \sigma_B \) then according to fig 2 there will be \( K\sigma_i \leq \sigma_i \) and \( KU_A \geq U_B \) thus \( \frac{U_B}{\sigma_A} \leq K \leq \frac{L_B}{L_A} \).
If \( K\sigma_A \leq \sigma_B \) we’ll have \( \frac{L_B}{\sigma_A} \leq K \leq \frac{U_B}{L_A} \) so we put \( \min \left( \frac{L_B}{\sigma_A}, \frac{U_B}{\sigma_A} \right) \leq K \leq \max \left( \frac{L_B}{\sigma_A}, \frac{U_B}{\sigma_A} \right) \).

Upper and lower limits are also as follows:
\[ KL = \min \left( \frac{L_B}{\sigma_A}, \frac{U_B}{\sigma_A} \right), \quad KU = \max \left( \frac{L_B}{\sigma_A}, \frac{U_B}{\sigma_A} \right) \]

![Fig 2: Overlaps of two RBS numbers](image)

Algorithm 2 is used to calculate \( K \) for the RBS number \( A \) with respect to \( B \).

Algorithm 2: finding the equivalence multiplier of \( A \) with respect to \( B \).

Step 1:
Put switch=0, apply algorithm 1 for \( A \) and \( B \).
If \( A = B \), put \( K = 1 \) and algorithm 2 is finished.
If \( A < B \), swap \( A \) and \( B \) and put switch=1.
Step 2:
Put \( KL = \min \left( \frac{L_B}{\sigma_A}, \frac{U_B}{\sigma_A} \right) \), \( KU = \max \left( \frac{L_B}{\sigma_A}, \frac{U_B}{\sigma_A} \right) \), \( e = 10E-4 \lambda = \frac{KL + KU}{2} \).
Step 3:
Apply algorithm 1 for \( A \) and \( B \) and \( \lambda \).
Step 4:
If the difference between \( R(\lambda A) \) and 0.5 is less than \( e \) then go to step 8 else, go to next step.
Step 5:
If \( R(\lambda A) < 0.5 - e \) then put \( KL = \lambda \)
If \( R(\lambda A) > 0.5 + e \) then put \( KU = \lambda \)
Step 6:
Put \( \lambda = \frac{KL + KU}{2} \).
Step 7:
Apply algorithm 1 for \( A \) and \( B \) and \( \lambda \) and return to step 3.
Step 8:
If switch=0 then put \( K = \lambda e + \frac{1}{\lambda} \).
In data envelopment analysis, we consider a benchmark point for each input and output which they are the period data.
Suppose RBS numbers, \((x_{i1}, x_{m1}, x_{u1}, x_{s1})\) and \((y_{s1}, y_{m1}, y_{u1}, y_{s1})\) as a benchmark for \( i^{th} \) input and \( j^{th} \) output respectively as the first component to fourth, are the minimum value, mean, maximum value and standard deviation, separately.
Also assume \((x_{r1}, x_{m1}, x_{u1}, x_{s1})\) and \((y_{r1}, y_{m1}, y_{u1}, y_{s1})\) as a benchmark for \( i^{th} \) input and \( j^{th} \) output.
It's obvious that \( (x, y, z, r) = (kx, ky, kz, kr) \) because when all of data are multiplied by \( K \), the standard deviation is also be multiplied by \( K \).
To calculate the preference ratio algorithm 1 can be used used and to gain the equivalence multiplier algorithm 2 can be used.
Now the preference method is applied for all of inputs and outputs and we find the equivalence multiplier of benchmarks with respect to related data. Model P1 is a CCR model with RBS data.

\[ p_i : \theta_j = \min \theta \]

\[ \sum_{j=1}^{m} \lambda_j (x_{ij}, x_{m1}, x_{u1}, x_{s1}) \leq 0 (x_{ij}, x_{m1}, x_{u1}, x_{s1}) \]

\[ \sum_{j=1}^{n} \lambda_j (y_{ij}, y_{m1}, y_{u1}, y_{s1}) \geq 0 (y_{ij}, y_{m1}, y_{u1}, y_{s1}) \]

\[ \lambda_j \geq 0 \]

Model P2 can be obtained using two assumptions:

\( K_{x_j} \) is the equivalence multiplier of benchmark \((x_{ij}, x_{m1}, x_{u1}, x_{s1})\) with respect to \((x_{ij}, x_{m1}, x_{u1}, x_{s1})\).

\( K_{y_j} \) is the equivalence multiplier of benchmark \((y_{ij}, y_{m1}, y_{u1}, y_{s1})\) with respect to \((y_{ij}, y_{m1}, y_{u1}, y_{s1})\).
p_2 : \theta_p = \min \theta

st

\sum_{i=1}^{s} \lambda_i K_{x_i}(x_{i0}, x_{i1}, x_{i2}, x_{i3}) \leq 0 \quad K_{x_i}(x_{i0}, x_{i1}, x_{i2}, x_{i3}) \quad i = 1, \ldots, m

\sum_{j=1}^{r} \lambda_j K_{y_j}(y_{j0}, y_{j1}, y_{j2}, y_{j3}) \geq K_{y_j}(y_{j0}, y_{j1}, y_{j2}, y_{j3}) \quad r = 1, \ldots, n

\lambda_i \geq 0 \quad j = 1, \ldots, n

It’s obvious that by changing the benchmark, many models can be derived. With simplification of model, P2 is changed to model P3 that is a simple CCR model.

p_3 : \theta_p = \min \theta

st

\sum_{i=1}^{s} \lambda_i K_{x_i}(x_{i0}, x_{i1}, x_{i2}, x_{i3}) \leq 0 \quad K_{x_i}(x_{i0}, x_{i1}, x_{i2}, x_{i3}) \quad i = 1, \ldots, m

\sum_{j=1}^{r} \lambda_j K_{y_j}(y_{j0}, y_{j1}, y_{j2}, y_{j3}) \geq K_{y_j}(y_{j0}, y_{j1}, y_{j2}, y_{j3}) \quad r = 1, \ldots, s

\lambda_i \geq 0 \quad j = 1, \ldots, n

Now the main question is what is the best benchmark is. Suppose B_1, B_2, \ldots, B_n are some fuzzy numbers that the benchmark A is chosen for them randomly. Following attributes are indefinable:

1- In the equation K_i, A = B_i, if R(A) > R(B_i) then K_i < 1 else K_i > 1.

2- If R(A) < \min_{i=1,\ldots,n}R(B_i) then all of equivalence multipliers are greater than 1 and if (A) > \max_{i=1,\ldots,n}R(B_i), the equivalence multipliers are less than 1.

If \min_{i=1,\ldots,n}R(B_i) < R(A) < \max_{i=1,\ldots,n}R(B_i) then some of equivalence multipliers are less and some are more than 1.

3- If K_i > K_j then R(K_i) > R(K_j).

4- Whatever k is closer to 1, it demonstrates the proximity of fuzzy number and the benchmark.

According to previous notes, each benchmark could be chosen but to have a better analysis, it seems that, it’s better to select the best benchmark; in this case, all numbers will be greater than 1 or they are less than 1 and the advantage is closeness to 1.

In DEA, between inputs, it seems that the best benchmarks the minimum benchmark and between outputs the best is the maximum so the benchmark selection procedure is as follows:

For each input like (x_{i0}, x_{i1}, x_{i2}, x_{i3}), the benchmark will be (x_{i0}, x_{i1}, x_{i2}, x_{i3}) in which we will have:

x_{il} = \min_{i=1,\ldots,n}x_{il}, \quad x_{im} = \min_{i=1,\ldots,n}x_{im}, \quad x_{iu} = \frac{x_{iu} - x_{il}}{6}

The reason of this selection for standard deviation is that in normal distribution and some distributions like normal, 99.73 percent of data are in an interval with 6 standard deviation length.

For each output, the benchmark is (y_{r0}, y_{r1}, y_{r2}, y_{r3}) that we will have:

y_{rl} = \max_{i=1,\ldots,n}y_{rl}, \quad y_{rm} = \max_{i=1,\ldots,n}y_{rm}, \quad y_{ru} = \frac{y_{ru} - y_{rl}}{6}

Of course, it’s obvious that, other benchmarks can be existed.

5. NUMERICAL EXAMPLE

Table 1 shows the data for 5 units as RBS numbers.

<table>
<thead>
<tr>
<th>Table 1. The data for an example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit1</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Input1</td>
</tr>
<tr>
<td>Input2</td>
</tr>
<tr>
<td>Output1</td>
</tr>
<tr>
<td>Output2</td>
</tr>
</tbody>
</table>

In table 2, some benchmarks have chosen for inputs and outputs. The first benchmark has selected the minimum between inputs and maximum between outputs. The second benchmark is always the maximum value of inputs and outputs. The third benchmark considered the minimum value of inputs or outputs and the fourth, has used the random numbers as a benchmark.

<table>
<thead>
<tr>
<th>Table 2. The Benchmarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min-max</td>
</tr>
<tr>
<td>(2.9, 3.0, 3.1, 0.03)</td>
</tr>
<tr>
<td>(1.4, 1.5, 1.6, 0.03)</td>
</tr>
<tr>
<td>(4.4, 5.1, 5.8, 0.23)</td>
</tr>
<tr>
<td>(6.5, 7.4, 8.3, 0.03)</td>
</tr>
</tbody>
</table>

In tables 3, 4, 5 and 6 the equivalence multipliers of the numbers available in table 1 have been calculated.
Then we converted the fuzzy model to a fuzzy simple additive weighting method by preference ratio. The efficiency of units has also been computed considering Min-Max, Min, Max and Random benchmarking. As it can be seen, in benchmark Min-Max, the equivalence multipliers of inputs are greater or equal to 1 and for outputs they are less or equal to 1. The unit 1 between inputs and unit 5 between outputs are the best. In Max benchmark, all data are less or equal to 1 and in Min benchmark, all data are greater or equal to 1. In random benchmark, the equivalence multipliers are greater or less than 1.

Ranking in all outputs and inputs for all of benchmarks are the same. Unit efficiencies are also unique for all of different benchmarks.

6. CONCLUSIONS

In this paper we tried to achieve two main goals. First, we converted and developed the Period Data Envelopment Analysis to a fuzzy models we made the RBS numbers considering the mean, standard deviation, maximum and minimum of data. Then we converted the fuzzy model to a simple model using the preference method. The importance and advantage of this method is in this point that as we know, most of fuzzy solutions are based on different cutting and making the results with different α brings some difficulties in final results and efficiency calculations, but using our method, applying just one simple model in that we use equivalence multiplier instead of main data, we can attain the relation efficiencies which is the final conclusion in easiest manner.

7. REFERENCES


