Estimation in DEA by Possibility Regression with the Upper and Lower Approximations

Tomoe Entani, Hidetomo Ichihashi
Graduate School of Engineering, Osaka Prefecture University
1-1 Gakuen-cho, Sakai, Osaka, 599-8531, Japan
E-mail {entani, ichi}@ie.osakafu-u.ac.jp

Hideo Tanaka
Graduate School of Management and Information, Toyohashi Sozo College
20-1 Matsushita, Ushikawa-cho, Toyohashi, Aichi, 440-8511, Japan
E-mail tanakah@sozo.ac.jp

Abstract

The exponential possibility regression is applied to estimate interval efficiency values. Interval efficiency value is obtained by interval DEA. It shows the possible efficiency and its upper and lower limits are obtained from the opimistic and pessimistic viewpoints for an object to be analyzed. The interval efficiency value for the object is calculated relatively by using all given inputs and outputs without assuming an identical single model for all DMUs. We analyze the interval efficiency values by regression analysis so that the identical model can be obtained. The obtained linear model in regression analysis can represent the relation between the given inputs and outputs and the obtained interval efficiency values in interval DEA. When new objects are added, we can estimate their interval efficiency values using the models obtained by regression analysis.

1. Introduction

DEA is a nonparametric technique for measuring and evaluating the relative efficiency of DMUs (Decision Making Units) with common inputs and outputs [1][6]. Input and output data produce the efficiency values. The ratios of outputs to inputs are taken as evaluation values and called efficiency values. While the efficiency value of conventional DEA is only from the optimistic viewpoint, the upper and lower limits of the interval efficiency value are obtained from the optimistic and pessimistic viewpoints respectively [2]. The interval efficiency value indicates the possible efficiency. In this paper, we obtain an identical model by regression analysis though there is no identical model in DEA. The obtained interval efficiency values and the given inputs and outputs in DEA are considered as the interval outputs and inputs in regression analysis respectively.

Regression analysis is known as one of the evaluation methods. The linear system or polynominal system is assumed, through which the inputs produce the output. The system estimates output obtained from the given inputs as close to the given output as possible. Therefore the estimated interval output shows the possible efficiency value for the given inputs. In the case where the outputs are given as intervals, interval regression with the upper and lower approximation models has been proposed [4][5]. Based on the possibility concept, the estimated interval output by the upper approximation model includes the given interval output. On the other hand, the estimated interval output by the lower one is included in the given interval output, based on the necessity concept. In other words, the upper and lower approximation models can be defined as the least upper bound and the greatest lower bound, respectively. However, the relation between input variables is not concerned because the coefficients of the assumed system are independent in interval regression analysis. The inputs in regression analysis that are the given inputs and outputs in DEA are related each other. Therefore the exponential possibility distribution of coefficients which is similar to the normal distribution in probability theory is introduced to interval regression analysis in this paper. The purpose of the study is to obtain the approximation models that are identical for all DMUs and to characterize the relations between the given inputs and outputs and the obtained interval efficiency values in DEA. With the obtained linear models, the interval efficiency values for new DMUs can be estimated.

2. Interval efficiency value by interval DEA

In DEA the maximum ratio of output to input is assumed as the efficiency which is calculated from the optimistic viewpoint for each DMU. DMU_o is evaluated taking all other DMUs into consideration. A DEA model with an interval efficiency value consisting of the efficiency values obtained from both the optimistic and pessimistic viewpoints was proposed [2].

Since DEA is based on a relative evaluation, the upper limit of the interval efficiency value is formulated as follows.

$$\begin{aligned} \theta_o{}^* &= \max_{\boldsymbol{u},\boldsymbol{v}} \frac{\boldsymbol{u}^t\boldsymbol{y}_o/\boldsymbol{v}^t\boldsymbol{x}_o}{\max\limits_j \left(\boldsymbol{u}^t\boldsymbol{y}_j/\boldsymbol{v}^t\boldsymbol{x}_j\right)} \\ \text{s.t.} \quad & \boldsymbol{u} \geq \boldsymbol{0}, \ \ \boldsymbol{v} \geq \boldsymbol{0} \end{aligned} \tag{1}$$

where the decision variables are the input and output weight vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, the given input and output vectors for DMU_j are $\boldsymbol{x}_j = (x_{j1}, ..., x_{jm})^t$ and $\boldsymbol{y}_j = (y_{j1}, ..., y_{jl})^t$ whose elements are positive and the number of DMU is n.

The optimal value of (1) is equivalent to that of CCR (Charnes Cooper Rhodes) model [1][6]. CCR model is the basic DEA model and formulated as follows.

$$\begin{aligned} & \theta_o{}^* = \max_{\boldsymbol{u}} \boldsymbol{u}^t\boldsymbol{y}_o \\ \text{s.t.} \quad & \boldsymbol{v}^t\boldsymbol{x}_o = 1 \\ & -\boldsymbol{v}^t X + \boldsymbol{u}^t Y \leq \boldsymbol{0} \\ & \boldsymbol{u} \geq \boldsymbol{0}, \boldsymbol{v} \geq \boldsymbol{0} \end{aligned} \tag{2}$$

where $X \in \Re^{m\times n}$ is an input matrix consisting of all input vectors, $Y \in \Re^{l\times n}$ is an output matrix consisting of all output vectors. By replaceing maximization in (1) with minimization, the lower limit of interval efficiency value for DMU_o can be defined as follows.

$$\begin{aligned} \theta_{o*} &= \min_{\boldsymbol{u},\boldsymbol{v}} \frac{\boldsymbol{u}^t\boldsymbol{y}_o/\boldsymbol{v}^t\boldsymbol{x}_o}{\max\limits_j \left(\boldsymbol{u}^t\boldsymbol{y}_j/\boldsymbol{v}^t\boldsymbol{x}_j\right)} \\ \text{s.t.} \quad & \boldsymbol{u} \geq \boldsymbol{0}, \ \ \boldsymbol{v} \geq \boldsymbol{0} \end{aligned} \tag{3}$$

We obtain the lower limit of efficiency value directly by

$$\theta_{o*} = \min_{p,r} \frac{y_{op}/x_{or}}{\max\limits_j (y_{jp}/x_{jr})} \tag{4}$$

The optimal value in (4) can be said that the optimal weight vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in (3) have the entry 1 respectively and all other entries are 0. It is shown in [2] that the optimal values in (3) and (4) are equal. Usually (4) is easier to solve than (3), because we do not have to solve any LP problems in (4).

The interval efficiency value denoted as $[\theta_{o*}, \theta_o{}^*]$ represents the possible efficiency for DMU_o. Its upper and lower limits are obtained from the optimistic and pessimistic viewpoints respectively.

3. Estimation of interval efficiency value by exponential possibility regression

3.1. Exponential possibility regression with interval output

To deal with the interactive coefficients, quadratic membership functions of fuzzy coefficients defined by a positive definite matrix are introduced to interval regression analysis. A fuzzy coefficient vector can be expressed as the possibility distribution defined by the exponential membership function $\pi_A(\boldsymbol{a})$;

$$\pi_A(\boldsymbol{a}) = \exp\left\{-(\boldsymbol{a}-\boldsymbol{a}_c)^t D_A{}^{-1}(\boldsymbol{a}-\boldsymbol{a}_c)\right\} \tag{5}$$

where $\boldsymbol{a}_c$ is a center vector, D_A is a symmetric positive definite matrix, i.e., $D_A > 0$ and A is parametrically denoted as $\left(\boldsymbol{a}, D_A\right)_e$. Let us assume a regression model as the following possibilistic linear system.

$$Y = \boldsymbol{x}^t\boldsymbol{A} \tag{6}$$

where $\boldsymbol{x} = (1, x_1, \cdots, x_s)^t$ is an input vector, $\boldsymbol{A} = (A_0, \cdots, A_s)$ is a fuzzy coefficient vector and Y is the corresponding fuzzy output. By using the extension principle, the possibility distribution of Y can be written as follows.

$$\pi_Y(y) = \max_{\{\boldsymbol{a}|y=\boldsymbol{x}^t\boldsymbol{a}\}} \pi_A(\boldsymbol{a}) \tag{7}$$

By Lagrangean multiplier method, this optimization problem is solved and the possibility distribution of Y is obtained as follows.

$$\pi_Y(y) = \exp\left\{-(y-\boldsymbol{x}^t\boldsymbol{a}_c)^2/(\boldsymbol{x}^t D_{\mathrm{A}}\boldsymbol{x})\right\} \tag{8}$$

where $\boldsymbol{x}^t\boldsymbol{a}_c$ and $\boldsymbol{x}^t D_{\mathrm{A}}\boldsymbol{x}$ are the center and the spread of the fuzzy output in (6), also (8) can be represented parametrically as $Y = (\boldsymbol{x}^t\boldsymbol{a}_c, \boldsymbol{x}^t D_{\mathrm{A}}\boldsymbol{x})_e$.

To deal with interval outputs, we consider h-level set of fuzzy output which is denoted as follows.

$$[Y]_h = \{y|\pi_Y(y) \geq h\} = [y_*, y^*]$$

The bounds of $[Y]_h$ are calculated by letting $\pi_Y(y) = h$. Taking logarithm, we obtain the bounds of h-level set of fuzzy output y_* and y^* as follows.

$$\begin{array}{l} y^* = \boldsymbol{x}^t \boldsymbol{a}_c + \sqrt{(\boldsymbol{x}^t D_{\mathrm{A}} \boldsymbol{x})(-\log h)} \\ y_* = \boldsymbol{x}^t \boldsymbol{a}_c - \sqrt{(\boldsymbol{x}^t D_{\mathrm{A}} \boldsymbol{x})(-\log h)} \end{array} \tag{9}$$

Denoting intervals as $A = [\underline{a}, \overline{a}]$ and $B = [\underline{b}, \overline{b}]$, the inclusion relation is defined as follows.

$$A \supseteq B \Leftrightarrow \underline{a} \leq \underline{b} \quad \text{and} \quad \overline{b} \leq \overline{a}$$

To approximate the given interval outputs $Y_i = [y_{i*}, y_i^*]$, we consider two models **(a)** and **(b)** as follows: **(a)** upper approximation model ($\overline{Y_i}$), which includes the given interval output and **(b)** lower approximation model ($\underline{Y_i}$), which is included in the given interval output.

(a) For the input $\boldsymbol{x}_i$, the upper fuzzy output $\overline{Y_i}$ is assumed as $\boldsymbol{x}_i{}^t \overline{\boldsymbol{A}}$ and its possibility distribution is given as (8). The interval output Y_i is included in the h_i level set of estimated output $\overline{Y_i}$. Let the h_i level set of $\overline{Y_i}$ be $[\overline{Y_i}]_{h_i} = [\overline{y_{i*}}, \overline{y_i^*}]$ given as (9). Then the following relation is assumed.

$$[\overline{Y_i}]_{h_i} = [\overline{y_{i*}}, \overline{y_i^*}] \supseteq Y_i = [y_{i*}, y_i^*] \tag{10}$$

Therefore the constraint conditions for the upper approximation model are as follows.

$$\begin{array}{l} (\boldsymbol{x}_i{}^t D_{\overline{\mathrm{A}}} \boldsymbol{x}_i)(-\log h_i) \geq (y_i^* - \boldsymbol{x}_i{}^t \boldsymbol{a}_c)^2 \\ (\boldsymbol{x}_i{}^t D_{\overline{\mathrm{A}}} \boldsymbol{x}_i)(-\log h_i) \geq (y_{i*} - \boldsymbol{x}_i{}^t \boldsymbol{a}_c)^2 \\ \quad (i = 1, ..., n) \\ D_{\overline{\mathrm{A}}} > \boldsymbol{0} \end{array} \tag{11}$$

The sum of spreads of the upper fuzzy output $\overline{Y_i}$ is minimized to find the least upper bound.

$$\min_{\boldsymbol{a}_c, D_{\overline{\mathrm{A}}}} \sum_{i=1}^n \boldsymbol{x}_i{}^t D_{\overline{\mathrm{A}}} \boldsymbol{x}_i \tag{12}$$

(b) For the input, the lower fuzzy output is assumed as $\boldsymbol{x}_i{}^t \underline{\boldsymbol{A}}$ and its possibility distribution is given as (8). The h_i level set of $\underline{Y_i}$ is included in the obtained interval output. The following relation is assumed.

$$Y_i = [y_{i*}, y_i^*] \supseteq [\underline{Y_i}]_{h_i} = [\underline{y_{i*}}, \underline{y_i^*}] \tag{13}$$

The constraint conditions for the lower approximation model are as follows.

$$\begin{array}{l} (\boldsymbol{x}_i{}^t D_{\underline{\mathrm{A}}} \boldsymbol{x}_i)(-\log h_i) \leq (y_i^* - \boldsymbol{x}_i{}^t \boldsymbol{a}_c)^2 \\ (\boldsymbol{x}_i{}^t D_{\underline{\mathrm{A}}} \boldsymbol{x}_i)(-\log h_i) \leq (y_{i*} - \boldsymbol{x}_i{}^t \boldsymbol{a}_c)^2 \\ D_{\underline{\mathrm{A}}} > \boldsymbol{0} \end{array} \tag{14}$$

The sum of spreads of the lower fuzzy output $\underline{Y_i}$ is maximized to find the greatest lower bound.

$$\max_{\boldsymbol{a}_c, D_{\underline{\mathrm{A}}}} \sum_{i=1}^n \boldsymbol{x}_i{}^t D_{\underline{\mathrm{A}}} \boldsymbol{x}_i \tag{15}$$

The upper and lower approximation models are obtained from the minimization problem (12) with constraints (11) and the maximization problem (15) with constraints (14) respectively. These are non-linear optimization problems but can be changed into LP problems. We divide them into two problems, i.e., the one is to obtain center vector $\boldsymbol{a}_c$, and the other is to obtain symmetric positive matrices $D_{\overline{\mathrm{A}}}$ and $D_{\underline{\mathrm{A}}}$. The procedures for two probelms are as follows:

1. We obtain the center vector $\boldsymbol{a}_c{}^*$ from the following problem.

$$\begin{array}{ll} \min_{\boldsymbol{a}_c} & \sum_{i=1}^n (y_i^c - \boldsymbol{x}_i{}^t \boldsymbol{a}_c)^2 \\ \text{s.t.} & y_{i*} \leq \boldsymbol{x}_i{}^t \boldsymbol{a}_c \leq y_i^* \quad (i = 1, ..., n) \\ & \boldsymbol{a}_c \geq \boldsymbol{0} \end{array} \tag{16}$$

   where y_i^c is the center of the interval output Y_i and $\boldsymbol{x}_i{}^t \boldsymbol{a}_c$ is the estimated center value of output for the input $\boldsymbol{x}_i$. The sum of deviations of the estimated outputs from the center values of the given outputs is minimized. It is not guaranteed that there exists a linear model in the constraint conditions in (16). If (16) is not feasible, more complex model such a polynominal one is suggested. By increasing the number of terms of the polynomials, we can find a polynomial system which is included in the given interval outputs.

2. In order that the matrix in exponential function determined by LP problem should be positive definite, we apply the following corollary instead of the necessary and sufficient conditions that D_A is positive definite [3][4]. In previous research [5], the orthogonal conditions formed by combination of input variables are used. Since there are many combinations, it is difficult to select the best combination. To remedy this weakness, instead of the orthogonal conditions, we use the corollary of Gersgorin theorem as a sufficient condition to constrain the matrix to be positive definite.

   [Corollary of Gersgorin Theorem]
   A sufficient condition for a real symmetric matrix $D = [d_{ij}] \in R^{(s+1)\times(s+1)}$ being positive definite is as follows.

$$d_{ii} > \sum_{j=0, j\neq i}^{s} (d_{ij}^+ + d_{ij}^-), \quad i = 0, ..., s$$

where $d_{ij} = {d_{ij}}^+ - d_{ij}^-$ with $d_{ij}^+ \geq 0$ and $d_{ij}^- \geq 0$. By this sufficient condition, it is easy to obtain a positive definite matrix in exponential function.

In order that the upper approximation model might include the lower one, the upper and lower approximation models are determined simultaneously by assuming that the centers of both models are equal and the spread of the upper model is larger than that of the lower one for any inputs.

It is formulated as follows.

$$\begin{array}{l}
\min\limits_{D_{\overline{A}}, D_{\underline{A}}} \quad \sum_{i=1}^{n} \left(\boldsymbol{x}_i{}^t D_{\overline{A}} \boldsymbol{x}_i - \boldsymbol{x}_i{}^t D_{\underline{A}} \boldsymbol{x}_i \right) \\
\text{s.t.} \\
(\boldsymbol{x}_i{}^t D_{\overline{A}} \boldsymbol{x}_i)(-\log h_i) \\
\geq \max\limits_i \left\{ (\overline{y}_i - \boldsymbol{x}_i{}^t \boldsymbol{a}_c{}^*)^2, (\underline{y}_i - \boldsymbol{x}_i{}^t \boldsymbol{a}_c{}^*)^2 \right\} \\
(\boldsymbol{x}_i{}^t D_{\underline{A}} \boldsymbol{x}_i)(-\log h_i) \\
\leq \min\limits_i \left\{ (\overline{y}_i - \boldsymbol{x}_i{}^t \boldsymbol{a}_c{}^*)^2, (\underline{y}_i - \boldsymbol{x}_i{}^t \boldsymbol{a}_c{}^*)^2 \right\} \\
\quad (i = 1, ..., n) \\
\overline{d}_{jj} - \underline{d}_{jj} \\
\geq \sum_{k=0, k \neq j}^{s} \left\{ (\overline{d}_{jk}^+ - \underline{d}_{jk}^+) + (\overline{d}_{jk}^- - \underline{d}_{jk}^-) \right\} + \varepsilon \\
\underline{d}_{jj} \geq \sum_{k=0, k \neq j}^{s} (\underline{d}_{jk}^+ + \underline{d}_{jk}^-) + \varepsilon \\
\quad (j = 0, ..., s)
\end{array} \tag{17}$$

(12) and (15) are integrated into an objective function. The first and second constraint conditions are formed from (11) and (14). The last two conditions show that the upper approximation model should always include the lower one, that is $D_{\overline{A}} - D_{\underline{A}} \geq \mathbf{0}$, and $D_{\underline{A}} \geq \mathbf{0}$.

3.2. Estimating the interval efficiency value from the inputs and outputs in DEA

We estimate the interval efficiency value determined by (2) and (4). We assume regression models with respect to the upper and lower approximations as the following possibilistic linear systems for the interval efficiency value from the given inputs and outputs.

$$\overline{\Theta} = \boldsymbol{z}^t \overline{\boldsymbol{A}}, \;\; \underline{\Theta} = \boldsymbol{z}^t \underline{\boldsymbol{A}}$$

where $\boldsymbol{z}^t = (1, \boldsymbol{x}^t, \boldsymbol{y}^t)$ is an input vector in regression model, $\overline{\boldsymbol{A}}$ and $\underline{\boldsymbol{A}}$ are fuzzy coefficient vectors.

In the same way as 3.1, at first we obtain the center vector $\boldsymbol{a}_c{}^*$ as follows.

$$\begin{array}{ll}
 & \min\limits_{\boldsymbol{a}_c} \;\; \sum_{i=1}^{n} (\theta_i^c - \boldsymbol{a}_c{}^t \boldsymbol{z}_i)^2 \\
\text{s.t.} & \theta_{i*} \leq \boldsymbol{a}_c{}^t \boldsymbol{z}_i \leq \theta_i{}^* \;\; (i = 1, ..., n) \\
 & \boldsymbol{a}_c \geq \mathbf{0}
\end{array} \tag{18}$$

where θ_i^c is the center of the inerval efficiency value $[\theta_{i*}, \theta_i{}^*]$ and $\boldsymbol{a}_c{}^t \boldsymbol{z}_i$ is the estimated center value of the interval efficiency value for DMU_i.

Then we obtain the upper and lower models simultaneously under the condition that the upper model include the lower one. It is formulated as follows.

$$\begin{array}{l}
\min\limits_{D_{\overline{A}}, D_{\underline{A}}} \quad \sum_{i=1}^{n} \left(\boldsymbol{z}_i{}^t D_{\overline{A}} \boldsymbol{z}_i - \boldsymbol{z}_i{}^t D_{\underline{A}} \boldsymbol{z}_i \right) \\
\text{s.t.} \\
(\boldsymbol{z}_i{}^t D_{\overline{A}} \boldsymbol{z}_i)(-\log h_i) \\
\geq \max\limits_i \left\{ (\theta_i{}^* - \boldsymbol{z}_i{}^t \boldsymbol{a}_c{}^*)^2, (\theta_{i*} - \boldsymbol{z}_i{}^t \boldsymbol{a}_c{}^*)^2 \right\} \\
(\boldsymbol{z}_i{}^t D_{\underline{A}} \boldsymbol{z}_i)(-\log h_i) \\
\leq \min\limits_i \left\{ (\theta_i{}^* - \boldsymbol{z}_i{}^t \boldsymbol{a}_c{}^*)^2, (\theta_{i*} - \boldsymbol{z}_i{}^t \boldsymbol{a}_c{}^*)^2 \right\} \\
\quad (i = 1, ..., n) \\
\overline{d}_{jj} - \underline{d}_{jj} \\
\geq \sum_{k=0, k \neq j}^{l+m} \left\{ (\overline{d}_{jk}^+ - \underline{d}_{jk}^+) + (\overline{d}_{jk}^- - \underline{d}_{jk}^-) \right\} + \varepsilon \\
\underline{d}_{jj} \geq \sum_{k=0, k \neq j}^{l+m} (\underline{d}_{jk}^+ + \underline{d}_{jk}^-) + \varepsilon \\
\quad (j = 0, ..., l+m)
\end{array} \tag{19}$$

4. Numerical example

4.1. Estimation of the interval efficiency value

There are 10 DMUs with one-input and two-output data shown in Table 1. The interval efficiency values determined by (2) and (4) are also shown in Table 1.

DMU	x_1	y_1	y_2	interval efficiency $[\theta_{i*}, \theta_i{}^*]$
A	3	1	8	[0.056,1.000]
B	4	2	3	[0.083,0.522]
C	1	2	6	[0.286,1.000]
D	5	3	3	[0.100,0.652]
E	2	3	7	[0.250,1.000]
F	1	4	2	[0.250,0.750]
G	3	4	5	[0.222,0.957]
H	6	5	2	[0.056,0.826]
I	1	6	2	[0.250,1.000]
J	5	7	1	[0.033,1.000]

Table 1. Data

The center vector of the estimated interval efficiency value is obtained by (18) and shows the input and output weights characters of all DMUs.

$$\boldsymbol{a}_c{}^* = (0.197, -0.032, 0.063, 0.048)^t$$

With using the obtained center vector, by solving (19) at $h_i = 0.5 (i = 1, ..., 10)$, the coefficient matrices are obtained as follows.

$$D^*_{\overline{\mathrm{A}}} = \begin{pmatrix} 0.03135 & -0.00047 & -0.00596 & 0.00256 \\ -0.00047 & 0.00594 & -0.00056 & -0.00413 \\ -0.00596 & -0.00056 & 0.00655 & 0.00000 \\ 0.00256 & -0.00413 & 0.00000 & 0.00671 \end{pmatrix}$$

$$D^*_{\underline{\mathrm{A}}} = \begin{pmatrix} 0.01076 & -0.00281 & -0.00512 & -0.00283 \\ -0.00281 & 0.00356 & -0.00070 & 0.00004 \\ -0.00512 & -0.00070 & 0.00850 & -0.00268 \\ -0.00283 & 0.00004 & -0.00268 & 0.00555 \end{pmatrix}$$

With the obtained center $\boldsymbol{a}_c{}^*$ and coefficient matrices $D^*_{\overline{\mathrm{A}}}$, $D^*_{\underline{\mathrm{A}}}$, the upper and lower fuzzy outputs determined by their centers and spreads are denoted as follows.

$$\overline{\Theta} = (\boldsymbol{z}^t \boldsymbol{a}_c{}^*, \boldsymbol{z}^t D^*_{\overline{\mathrm{A}}} \boldsymbol{z}), \quad \underline{\Theta} = (\boldsymbol{z}^t \boldsymbol{a}_c{}^*, \boldsymbol{z}^t D^*_{\underline{\mathrm{A}}} \boldsymbol{z})$$

In Figure 1, the obtained interval efficiency values and 0.5-level sets of the upper and lower fuzzy outputs are illustrated by normal and bold lines respectively and their centers are illustrated as $\bullet$ and $-$. We note that the upper and lower limits of the upper approximation models may be greater than one and smaller than zero respectively. All the obtained interval efficiency values are included in 0.5-level sets of the upper fuzzy outputs and include those of the lower ones.

We apply the following interval regression using the centers obtained by (18), to show that exponential possibility regression model can approximate the obtained interval efficiency values more precisely than interval one.

$$\begin{array}{ll} & \min\limits_{\overline{\boldsymbol{c}}, \underline{\boldsymbol{c}}} \ \sum_{i=1}^n (\boldsymbol{z}_i{}^t \overline{\boldsymbol{c}} - \boldsymbol{z}_i{}^t \underline{\boldsymbol{c}}) \\ \text{s.t.} & \boldsymbol{z}_i{}^t \boldsymbol{a}_c^* - \boldsymbol{z}_i{}^t \overline{\boldsymbol{c}} \leq \theta_{i*} \leq \boldsymbol{z}_i{}^t \boldsymbol{a}_c^* - \boldsymbol{z}_i{}^t \underline{\boldsymbol{c}} \\ & \boldsymbol{z}_i{}^t \boldsymbol{a}_c^* + \boldsymbol{z}_i{}^t \underline{\boldsymbol{c}} \leq \theta_i{}^* \leq \boldsymbol{z}_i{}^t \boldsymbol{a}_c^* + \boldsymbol{z}_i{}^t \overline{\boldsymbol{c}} \\ & \quad (i = 1, ..., n) \\ & \overline{\boldsymbol{c}} \geq \underline{\boldsymbol{c}} \end{array} \tag{20}$$

where the estimated interval efficiency values of the upper and lower approximation models are obtained simultaneously so as to include and to be included in the obtained interval efficiency value $[\theta_{i*}, \theta_i{}^*]$ respectively. By solving (20), $\overline{\boldsymbol{c}}^*$ and $\underline{\boldsymbol{c}}^*$, which determine the spreads of the estimated outputs, are obtained as follows.

$$\overline{\boldsymbol{c}}^* = (-0.098, 0.006, 0.071, 0.062)^t$$
$$\underline{\boldsymbol{c}}^* = (-0.072, 0.011, 0.060, 0.031)^t$$

The upper and lower approximations obtained by (20) are shown as dotted lines in Figure 1.

We take the sum of the deviations of the bounds of the upper approxiamtion model from those of the lower one as its measure of fitness. They indicate how close the upper approximation output is to the lower one. In this example, they are 1.344 and 2.408 for exponential and interval regression models respectively. We find that the fuzzy efficiency values in exponentiail regression model are more precisely than those in interval regression model. The inputs in regression analysis that are the given inputs and outputs in DEA are interactive each other, therefore exponential possibility regression where interavtive coefficients can be dealt with is more suitable than interval one. It is favorable to obtain the precise approximation models, when we estimate the interval efficiency values for new DMUs with them.

4.2. Interval efficiency values for new DMUs

For example, we add three DMUs (K,L,M) shown in Table 2 and estimate their interval efficiency values. We assume that $h = 0.5$. It should be noted that the h-level set of their estimated upper fuzzy efficiency values might be greater than 1.000, though the maximum of efficiency values on DEA must be constrainted to be 1.000. As for K and L, which are quite similar to the existing G and D respectively, their efficiency values are estimated by the obtained models fairly well. However, as for M which is not similar to the existing ones, the estimated and obtained efficiency values are somewhat different. The proposed approximation models are determined based on the existing DMUs therefore the more similar the added DMU to the existing ones is, the better the estimation is.

5. Conclusion

We applied exponential possibility regression to interval DEA where the interval efficiency value is obtained by the input and output data. The given inputs and outputs and the obtianed interval efficiency value for a DMU in interval DEA were assumed as inputs and interval output respectively in regression analysis. We formulated the upper and lower possibility regression models for interval outputs. We used two steps to obtain the fuzzy output by solving LP problems. The first step is to determine the center in the sense of least squares regression. Then using the obtained center, the second step is to determine the spread matrix of the approximation

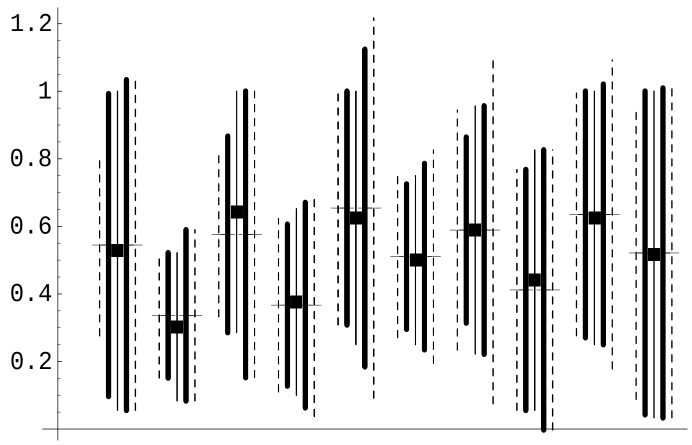

Figure 1. Interval efficiency values obtained by interval DEA and estimated by exponential and interval regression models

				upper approximation		lower approximation		interval efficiency by DEA
DMU	x_1	y_1	y_2	Θ	h-level set	$\underline{\Theta}$	h-level set	$[\theta_*, \theta^*]$
K	4	4	6	(0.605,0.377)	[0.093,1.116]	(0.605,0.077)	[0.374,0.835]	[0.167,1.000]
L	5	4	3	(0.429,0.201)	[0.056,0.802]	(0.429,0.083)	[0.190,0.669]	[0.100,0.783]
M	6	9	10	(1.043,1.239)	[0.117,1.990]	(1.043,0.356)	[0.246,1.540]	[0.250,1.000]

Table 2. Interval efficiency values obtained by interval DEA and estimated by regression analysis

model. With the spread matrix, the relations between inputs can be taken into consideration and squares of spread can be dealt with by LP problem. By numerical example we showed that exponential possibility regression is more fitting to the obtained interval efficiency values than interval regression because the inputs and outputs in interval DEA are interactive. With using the obtained upper and lower approximation models that are identical for existing DMUs, the interval efficiency values for new DMUs can be estimated without calculating interval DEA. If the added DMUs are similar to the existing ones, we can obtain the reliable estimations of their interval efficiency values.

References

[1] A.Charnes, W.W.Cooper and E.Rhodes, ”Measuring the Efficiency of Decision Making Unit”, *European Journal of Operational Research*, 1978, pp 429-444.

[2] T.Entani, Y.Maeda and H.Tanaka, Dual Models of Interval DEA and Its Extension to Interval Data, *European Journal of Operational Research*, (accepted for publication).

[3] S.Gersgori, ”Uber die abgrenzug der eigenwerter einer matrix”, *Izv. Akad. Nauk*, SSSR 7 1931, pp.749-754.

[4] H.Tanaka and P.Guo, *Possibilitic Data Analysis for Operations Research*, Physica-Verlag, Heidelberg, 1999.

[5] H.Tanaka, H.Ishibuchi and S.Yoshikawa, ”Exponential possibility regression analysis”, *Fuzzy sets and Systems*, Vol.69, No.6 1995, pp 305-318.

[6] K.Tone, *Mesurement and Improvement of Efficiency by DEA*, Nikkagiren, Japan, 1993 (in Japanese).