

# Confident-DEA: A Unified Approach For Efficiency Analysis With Cardinal, Bounded And Ordinal Data

Said Al-Gattoufi, (E-mail: [Gattoufi@squ.edu.om](mailto:Gattoufi@squ.edu.om)), Sultan Qaboos University, Sultanate of Oman

## ABSTRACT

*This paper proposes an extension to the existing literature in DEA, the authors call Confident-DEA approach. The proposed new approach involves a bi-level convex optimization model, and hence NP-hard, to which a solution method is suggested. Confident-DEA constitutes a generalization of DEA for dealing with imprecise data and hence a potential method for forecasting efficiency. Imprecision in data is defined as two forms, one is bounded data and the second is cardinal data. Complementing the methodology proposed by Cooper et al (1999) which provides single valued efficiency measures, Confident-DEA provides a range of values for the efficiency measures, e.g. an efficiency confidence interval, reflecting the imprecision in data. For the case of bounded data, a theorem defining the bounds of the efficiency confidence interval is provided. For the general case of imprecise data, that is a mixture of ordinal and cardinal data, a Genetic-Algorithm-based meta-heuristic is used to determine the upper and lower bounds defining the efficiency confidence interval. To the best knowledge of the authors, this is the first work combining Genetic algorithms with DEA. In both cases of imprecision, a Monte-Carlo type simulation is used to determine the distribution of the efficiency measures, taking into account the distribution of the bounded imprecise data over their corresponding intervals. Most of previous DEA works dealing with imprecise data implicitly assumed a uniform distribution. Confident-DEA, on the other hand, allows for any type of distribution and hence expands the scope of the analysis. The bounded data used in the illustrative examples are assumed to have truncated normal distributions. However, the methodology suggested here allows for any other distribution for the data.*

## 1. Introduction

In the standard Data Envelopment Analysis (DEA) approach originally proposed in Charnes, Cooper and Rhodes (1978), Decision Making Units (DMUs) are considered to be economic units using inputs to produce outputs. Although the inputs and outputs are usually assumed to be observable and measurable, in many real life situations these factors are not precisely known except (i) to the extent that the true values lie within prescribed bounds, and/or (ii) to satisfy certain ordinal relations. Data falling in the first category are referred as *bounded data* while the later are known as *ordinal data*. Cooper et al (1999) refers to any mixture of such kind of data with standard single valued data as *imprecise data*. We henceforth use the same terminology.

When data are imprecise, the application of the standard DEA leads to a non-linear program and the piecewise linear efficient frontier defined by that approach is not guaranteed. Moreover, ordinal data cannot be considered in the standard model.

*Imprecise Data Envelopment Analysis (IDEA<sup>1</sup>)*, Cooper et al (1999), treats mixtures involving bounded data in addition to ordinal data and ordinal relations among the weights. However, a major conceptual criticism can

---

<sup>1</sup> The same abbreviation e.g., IDEA, for "Interactive Data Envelopment Analysis" is used in Post and Spronk (1999). The abbreviation proposed by Cooper et al. (1999) is becoming more common in the literature.

be addressed to the IDEA approach. The problem lies with *the derivation of single-valued measures from imprecise multi-valued data. The efficiency measures should, in my opinion, reflect the imprecision in data.* In this setting, a range of values for the efficiency measure is more appropriate than a single-value. Furthermore, this range can be considered as a confidence interval for the efficiency measure and hence the name *Confident-DEA*. The spread of the efficiency interval can be interpreted as an indicator of the degree of volatility for the efficiency of the corresponding DMU. The wider is the efficiency interval, the higher is the volatility of the efficiency and hence the higher is the uncertainty about the relative performance of the corresponding DMU.

*Confident-DEA* extends the standard DEA to the case of imprecise data while overcoming the above mentioned weakness by producing efficiency confidence intervals. Furthermore, it compliments the *IDEA* in the sense that it allows considering stochasticity in data. In the case of bounded data, *Confident-DEA* takes into account the distribution of values of the factors within their corresponding ranges whereas *IDEA* implicitly assumes the uniform distribution.

**2. The Mathematical Modeling**

Given the inputs and outputs to be considered, the following model is the standard *Data Envelopment Analysis* (DEA) model :

$$Max_{\mu, \omega} h_o = \sum_{r=1}^{r=s} \mu_r y_{ro} \tag{1.1}$$

*Subject to :*

$$\sum_{r=1}^{r=s} \mu_r y_{rj} - \sum_{i=1}^{i=m} \omega_i x_{ij} \leq 0; j = 1, 2, \dots, n \tag{1.2}$$

$$\sum_{i=1}^{i=m} \omega_i x_{i0} = 1 \tag{1.3} \tag{1}$$

$$y_r = (y_{rj}) \in D_r^+ \tag{1.4}$$

$$x_i = (x_{ij}) \in D_i^- \tag{1.5}$$

$$\mu = (\mu_r) \in A^+, \mu_r \geq 0 \tag{1.6}$$

$$\omega = (\omega_i) \in A^-, \omega_i \geq 0 \tag{2.7}$$

where  $x$  represents the matrix of input values for each DMU. It specifies the values of inputs used in the production process.  $y$  on the other hand represents the output matrix. It specifies, for each DMU, the values of the different outputs that result from the production process.  $u$  and  $w$  are the coefficient vectors to be determined by solving the model.  $RD_r^+, RD_r^-, RA_r^+$  and  $RA_r^-$  respectively represent domains for the outputs, inputs, output multipliers and input multipliers.

Real world situations often dictate data, the values of which lie within some prescribed bounds. Moreover, the data may be ordinal rather than cardinal in form, and hence known only to be satisfying certain ordinal relations. In Cooper et al (1999) these are labeled “*imprecise data*”. Lastly, the “*data*” may represent the decision-maker’s judgmental restrictions on the relative weights allowed to each or to some of the factors and/or multipliers. This is known in the DEA literature as the *Assurance-Region*. A specific domain for the solution search can be imposed and this is known as the *Cone-Ratio*. The general form presented above allows for all forms of data as well as all forms of restrictions on multipliers. In this paper we deal only with imprecision in Data. Restrictions are out of the scope of the work.

In the case of imprecise data, the model presented above is not linear any longer. The standard DEA approach cannot be applied, and hence the piecewise linear efficient frontier defined that approach is not guaranteed. Formulating the basic DEA model using imprecise data leads in fact to a *non-linear optimization problem*.

The literature of DEA was repeatedly reviewed and analyzed in Gattoufi et al.(2004a and 2004b). The early literature dealing with imprecise data was simply devoted to extend the standard DEA for coping with ordinal data. Golany (1988) presented a model incorporating ordinal relations among the weights of the DEA model. Cook et al (1993) presented a framework for incorporating a single input within the standard DEA framework. In a follow-up work, Cook et al (1996) extended their framework to the case where more than one factor is ordinal. Kim et al (1999) developed a procedure for handling both ordinal data and weights preferences. Lee et al (2002) transformed the non-linear program obtained by considering imprecise data to a linear program, using a series of modification of variables. Cooper et al (1999) developed a unified approach to treating mixtures involving bounded data in addition to ordinal data and ordinal relations among the weights. Their approach, the *Imprecise Data Envelopment Analysis (IDEA)*, extends the standard DEA to cope with imprecise data. In a following-up work, Cooper et al (2001) presented an illustrative application of their unified approach. Formulating the basic DEA model using imprecise data leads to a *non-linear optimization problem*. For the linearization, IDEA proceeds in two steps, scale transformations followed by variable alterations. The transformed model has the form of a standard DEA model. The solution for the original model is obtained from that of the transformed model using the reverse variable alterations and scale transformations.

A common criticism in this respect is that these approaches do not reflect explicitly the imprecision of the data within the assessment efficiency represented by the efficiency coefficients provided. That is, the efficiency measure obtained for each DMU is single-valued regardless the data are single-valued or imprecise.

Three major critics can be addressed as regarding the *IDEA* approach. First, the authors, in order to linearize the model obtained from the application of standard DEA to imprecise data, transformed the status of data to variables. That is, the authors consider the factors of data not precisely known as variables. This leads to an optimization problem where they decide about data as well as about variables. The basic Operations Research methodology requires a clear identification and separation between the decision variables, object of decision for the optimal level they should have, and the parameters represented by the coefficient defined by the data of the problem.

Second, for a variable defined as having bounded data, the *IDEA* approach requires that for the DMU used as anchor for the scale transformation and variable alteration, that is the DMU with the highest range for the corresponding bounded variable, the range is transformed into a single-valued. If this “*approximation*” is not made, the reverse transformations to retrieve the solution for the original problem can not be performed. This reduces some of the generality of the *IDEA* approach. However, this was corrected in Cooper et al (2000) by introducing dummy DMUs in the analysis.

Finally, the major criticism is conceptual in nature. The problem with the existing literature dealing with imprecise data is *the derivation of single-valued measures from imprecise multi-valued data*. The efficiency measure should reflect the imprecision in data and a range of values for the efficiency measure is more appropriate than a single-value. This range can be considered as a confidence interval for the efficiency measure. Later in this study, a new methodology, called *Confident-DEA*, is provided. It extends the standard DEA to the case of imprecise data and produce efficiency confidence intervals.

This paper develops a new approach, *Confident-DEA* that extends and generalizes the *IDEA* approach in the case of single valued and bounded cardinal data in the sense it allows imprecision in data to be reflected in the resulting efficiency measures. This is achieved by providing a range for the efficiency measures, an efficiency confidence interval and hence the name *Confident-DEA*, for each DMU instead of the single valued measure provided by the *IDEA* approach. A generalization to the case of imprecise data, an approach using a Genetic Algorithm based meta-heuristic for determining the bounds of the efficiency confidence interval is proposed in this work.

The upper bound confident efficiency interval for each DMU in the case of bounded data is obtained by solving the following model:

$$\text{Max}_{x,y} \text{Max}_{\mu,\omega} h_0 = \sum_{r=1}^{r=s} \mu_r y_{r0} \quad (2.1)$$

subject to :

$$\sum_{r=1}^{r=s} \mu_r y_{rj} - \sum_{i=1}^{i=m} \omega_i x_{ij} \leq 0; j = 1,2,\dots,n \quad (2.2) \tag{2}$$

$$\sum_{i=1}^{i=m} \omega_i x_{i0} = 1 \quad (2.3)$$

$$y_r = (y_{rj}) \in D_r^+ \quad (2.4)$$

$$x_i = (x_{ij}) \in D_i^- \quad (2.5)$$

$$\mu = (\mu_r) \in A^+, \mu_r \geq 0 \quad (2.6)$$

$$\omega = (\omega_i) \in A^-, \omega_i \geq 0 \quad (2.7)$$

The lower bound is determined by considering the minimization model.

These two models represent a *non-linear convex* problems and can be written in the general form of a *bilevel convex* model, discussed in greater detail by Bard (1998). In the two levels of optimization, multipliers are subjects at the lower level while the factors are subjects at the upper level. The model proceeds by determining the optimal multipliers for a given level of the factors. The general mathematical form of a *bilevel convex* problem, where  $F, G, f$  and  $g$  are convex functions is:

$$\text{Max}_{x \in X} F(x, \omega(x)) \quad (3.1) \tag{3}$$

$$\text{Subject to } G(x, \omega(x)) \leq 0 \quad (3.2)$$

$$\omega(x) = \text{Max}_{y \in Y} f(x, y) \quad (3.3)$$

$$\text{Subject to } g(x, y) \leq 0 \quad (3.4)$$

The well known Max-Min problem is a special case of the general bilevel convex problem. While Jeroslow (1985) proved that the *Max-Min* problem is NP-hard, Hansen et al (1992) proved that the *linear bilevel programming problem* is *strongly NP-hard*. The significant difficulty in solving the general form, convex bilevel optimization problems, justifies the use of heuristics.

### 3. Confident-DEA With Bounded Data

#### 3.1. The Optimistic/Pessimistic Point Of View Theorem

This section considers a situation where data for each DMU is either single valued or bounded . By defining the “*optimistic point of view*” and “*pessimistic point of view*” for each DMU, the *Confident-DEA* model’s underlying theorem will be is proven.

**Definition 1: Optimistic/ Pessimistic point of view for a given DMU:**

- 1- The *optimistic point of view* for each DMU considered is represented by the base-DMU using the minimum allowed quantities of inputs to produce the maximum permitted quantities of outputs. Conversely, its competitors use the maximum of inputs in producing the minimum of outputs.
- 2- The *pessimistic point of view* for each DMU considered represented by the base-DMU using the maximum allowed quantities of inputs in producing the minimum quantities of outputs. Conversely, its competitors use the minimum allowed quantities of inputs to produce the maximum quantities of outputs.

**Theorem 1:**

- 1- Each DMU’s maximum efficiency level is reached when conditions represented by its optimistic point of view apply.
- 2- Each DMU’s minimum efficiency level is reached when conditions represented by its pessimistic point of view apply.

**Proof:**

The proof of the first claim in the theorem is provided next. The second claim can be easily derived by analogy. The proof proceeds in a recurrent way. It first considers the case of two DMUs (DMU<sub>1</sub>) and (DMU<sub>2</sub>), both using the same single input (X<sub>1</sub>) for producing the same single output (Y<sub>1</sub>). The original CCR ratio form of DEA is used for the proof. For this case, the CCR formulation considering DMU<sub>1</sub> as base one is:

$$Max_{u,v} \theta = \frac{u_1 y_{11}}{v_1 x_{11}} \quad (4.1)$$

subject to:

$$\frac{u_1 y_{11}}{v_1 x_{11}} \leq 1 \quad (4.2) \tag{4}$$

$$\frac{u_1 y_{12}}{v_1 x_{12}} \leq 1 \quad (4.3)$$

$$u_1, v_1 \geq \varepsilon \quad (4.4)$$

$$\varepsilon \text{ Non – Archimedian.} \quad (6.5)$$

where *u* and *v* are the weights to be determined. Without loss of generality, we can assume *v*<sub>1</sub>=1 and use *u* as a single variable. This leads to following form derived from (4.1):

$$Max_{u,v} \theta = \frac{u y_{11}}{x_{11}} \quad (5.1)$$

subject to:

$$u \leq \frac{x_{11}}{y_{11}} \quad (5.2) \tag{5}$$

$$u \leq \frac{x_{12}}{y_{12}} \quad (5.3)$$

$$u \geq \varepsilon \quad (5.4)$$

$$\varepsilon \text{ Non – Archimedian.} \quad (5.5)$$

The maximum is then obtained as:

$$u^* = Min \left\{ \frac{x_{11}}{y_{11}}, \frac{x_{12}}{y_{12}} \right\} \tag{6}$$

thus *u*\* = *u*(*x*<sub>11</sub>, *x*<sub>12</sub>, *y*<sub>11</sub>, *y*<sub>12</sub>)

Assume that the bounds for the inputs and outputs are defined as follows:

$$x_{11} \in [x_{11}, \bar{x}_{11}]; x_{12} \in [x_{12}, \bar{x}_{12}] \text{ and } y_{11} \in [y_{11}, \bar{y}_{11}]; y_{12} \in [y_{12}, \bar{y}_{12}] \tag{7}$$

then from (8) we have:  $u^* \leq \text{Min} \left\{ \frac{x_{11}}{y_{11}}, \frac{\overline{x_{12}}}{\overline{y_{12}}} \right\}$

Let:  $\overline{u(x_{11}, y_{11})} = \text{Min} \left\{ \frac{x_{11}}{y_{11}}, \frac{\overline{x_{12}}}{\overline{y_{12}}} \right\}$  (8)

**Case 1:**

If  $\frac{x_{11}}{y_{11}} \leq \frac{\overline{x_{12}}}{\overline{y_{12}}}$  then  $\overline{u(x_{11}, y_{11})} = \frac{x_{11}}{y_{11}}$  which gives, by plugging in (5), an efficiency coefficient for DMU<sub>1</sub>:

$\theta_1 = 1$ . This means that DMU<sub>1</sub> is already efficient compared to DMU<sub>2</sub>. Neither a decrease in the input nor an increase in the output will improve the efficiency, which is already at its maximum level.

**Case 2:**

If  $\frac{x_{11}}{y_{11}} > \frac{\overline{x_{12}}}{\overline{y_{12}}}$  then  $\overline{u(x_{11}, y_{11})} = \frac{\overline{x_{12}}}{\overline{y_{12}}}$  which gives, by plugging in (5), that:

$\theta = \overline{u(x_{11}, y_{11})} \frac{y_{11}}{x_{11}} = \frac{\overline{x_{12}}}{\overline{y_{12}}} \frac{y_{11}}{x_{11}} \leq \frac{\overline{x_{12}}}{\overline{y_{12}}} \frac{\overline{y_{11}}}{\overline{x_{11}}} = \overline{\theta}$  which means that DMU<sub>1</sub>, considered as base DMU, is not

fully efficient and the maximum level of efficiency it can reach is defined by the conditions prescribed by its optimistic point of view e.g.: minimum input and maximum output.

This ends the proof for the case of two DMUs using a single input to produce a single output.

Consider now the general case of *n* DMUs using the same set of *m* inputs to produce the same set of *s* outputs. The CCR ratio form for a generic DMU indexed “*o*” in this general case is:

$$\text{Max}_{u,v} \theta_o = \frac{\sum_{r=1}^{r=s} u_r y_{ro}}{\sum_{i=1}^{i=m} w_i x_{io}} \quad (9.1)$$

Subject to :

$$\frac{\sum_{r=1}^{r=s} u_r y_{rj}}{\sum_{i=1}^{i=m} w_i x_{ij}} \leq 1; j = 1, 2, \dots, n \quad (9.2)$$

$$u, v \geq \varepsilon \quad (9.3)$$

$$\varepsilon \text{ Non – Archimedean} \quad (9.4)$$

Suppose that:

$$x_{ij} \in [\underline{x}_{ij}, \overline{x}_{ij}] \text{ and } y_{rj} \in [\underline{y}_{rj}, \overline{y}_{rj}] \text{ for } i=1,2,\dots,m; j=1,2,\dots,n; r=1,2,\dots,s \tag{10}$$

Let:

$$\theta_o = \frac{\sum_{r=1}^{r=s} u_r y_{ro}}{\sum_{i=1}^{i=m} w_i x_{io}}; \theta_o = \theta_o(u, v, x_o, y_o) \tag{11.1}$$

$$\tag{11}$$

and :

$$\theta_j = \frac{\sum_{r=1}^{r=s} u_r y_{rj}}{\sum_{i=1}^{i=m} w_i x_{ij}}; \theta_j = \theta_j(u, v, x_j, y_j) \tag{11.2}$$

The problem in (9) becomes:

$$Max_{u,v} \theta_o \tag{12.1}$$

Subject to :

$$\theta_j \leq 1; j = 1, 2, \dots, n \tag{12.2}$$

$$u, v \geq \varepsilon \tag{12.3}$$

$$\varepsilon \text{ Non – Archimedian} \tag{12.4}$$

(12)

Let  $\theta_o^*$  be the maximum value for  $\theta_o$  and  $\theta_o^* = \theta_o(u^*, v^*, x_o, y_o)$  where  $(u^*, v^*)$  is the optimal solution of (6.14). Two cases are then to be considered:

**Case 1:**  $\theta_o^* = \theta_o(u^*, v^*, x_o, y_o) = 1$  and  $\theta_j^* = \theta_j(u^*, v^*, x_j, y_j) \leq 1; j = 1, 2, \dots, n$  :

In this case, the base DMU is already efficient and neither a decrease in input nor an increase in output will improve its efficiency because the base DMU is already fully efficient. Its efficiency coefficient is at its maximum level. Thus the theorem holds trivially in this case.

**Case 2:**  $\theta_o^* = \theta_o(u^*, v^*, x_o, y_o) < 1$  and  $\theta_j^* = \theta_j(u^*, v^*, x_j, y_j) \leq 1; j = 1, 2, \dots, n$  :

In this case, there must exist at least one DMU, other than the base-DMU, which is fully efficient.

Otherwise,  $\theta_j^* = \theta_j(u^*, v^*, x_j, y_j) < 1; j = 1, 2, \dots, n$ . That is:

$$\frac{\sum_{r=1}^{r=s} u_r^* y_{rj}}{\sum_{i=1}^{i=m} w_i^* x_{ij}} < 1; j = 1, 2, \dots, n$$

which contradicts with the fact that  $(u^*, v^*)$  is the optimal solution.

Let  $J = \{j': DMU_{j'} \text{ is inefficient}\}$  and  $J^c = \{j'': DMU_{j''} \text{ is efficient}\}$ . That is:

$$\theta_{j'}^* < 1; j' \in J \text{ and } \theta_{j''}^* = 1; j'' \in J^c \tag{13}$$

Let  $\Omega = \left\{ (u^*, v^*) : (u^*, v^*) \text{ is the optimal solution of the CCR model with } DMU_o \text{ as base DMU} \right\}$

If each DMU, except the base one, chooses its values for the inputs and outputs the vector:  $(\bar{x}_j, \bar{y}_j); j \neq o$  where

$\bar{x}_j = (\bar{x}_{1j}, \dots, \bar{x}_{mj}); j \neq o$  and  $\bar{y}_j = (\bar{y}_{1j}, \dots, \bar{y}_{sj}); j \neq o$  then we still have:

$$\begin{cases} \theta_{j'}(u^*, v^*, \bar{x}_{j'}, \bar{y}_{j'}) < 1; j' \in J & (14.1) \\ \theta_{j''}(u^*, v^*, \bar{x}_{j''}, \bar{y}_{j''}) \leq 1; j'' \in J^c & (14.2) \\ \theta_o(u^*, v^*, x_o, y_o) < 1 & (14.3) \end{cases} \tag{14}$$

Therefore,  $(u^*, v^*) \in \Omega$  will remain a feasible solution, although it may not be an optimal one, for the CCR

model at  $(\bar{x}_j, \bar{y}_j); j \neq o$  and  $(x_o, y_o); j = o$ .

If  $DMU_o$  decreases its inputs and increases its outputs, the efficiency coefficient value

$$\theta_o^* = \theta_o^*(u^*, v^*, x_o, y_o) = \frac{\sum_{r=1}^{r=s} u_r^* y_{ro}}{\sum_{i=1}^{i=m} v_r^* x_{io}}$$

will increase as well. Therefore, the maximum is reached at the

conditions defined by the optimistic point of view of  $DMU_o$  considered as the base DMU.

Thus far proof was provided that the upper level of efficiency for each DMU, considered for the relative efficiency in a set of competing DMUs, is reached under conditions prescribed by its optimistic point of view. The proof for the claim that the lower level of efficiency is reached under conditions prescribed by its pessimistic point of view can be obtained analogously. The theorem is proven.

The theorem shows precisely what Charnes and Cooper (1985) describe as natural *desired directions* for production factors in efficiency analysis. They explain that *augmentation* is the desired direction for *outputs* and *diminution* is the desired direction for *inputs*.

### 3.2. An Illustrative Example

An example illustrating the *Confident-DEA* approach in the case of bounded cardinal data is provided next. The results obtained with *Confident-DEA* are compared to those in the *IDEA* approach. Consider eight (8) decision making units (DMUs) competing in the same market using two (2) inputs to produce two (2) outputs. DMU1 uses 124 units of input 1 and an imprecise quantity, at least 40 units and at most 50 units, of input 2 to produce 89.8 units



of output 1 and an imprecise quantity, more than 55 units but less than 65 units, of output 2. Table 1 contains the data for the example.

The data corresponding to the optimistic point of view of DMU 1 is given in Table 2 and data for the pessimistic point of view can be obtained by analogy.

For each DMU, two DEA models are considered and solved. The model in (17) represents the optimistic point of view of DMU1 and results for all the DMUs are given in Table 3.

$$\begin{aligned}
 & \underset{\mu, \omega}{\text{Max}} h_0 = 89.8\mu_1 + 65\mu_2 \\
 & \text{subject to :} \\
 & 89.8\mu_1 + 65\mu_2 - 124\omega_1 - 40\omega_2 \leq 0 \\
 & 99.6\mu_1 + 60\mu_2 - 95\omega_1 - 30\omega_2 \leq 0 \\
 & 87\mu_1 + 70\mu_2 - 92\omega_1 - 70\omega_2 \leq 0 \\
 & 99.4\mu_1 + 95\mu_2 - 61\omega_1 - 100\omega_2 \leq 0 \\
 & 96.4\mu_1 + 65\mu_2 - 63\omega_1 - 80\omega_2 \leq 0 \\
 & 86\mu_1 + 85\mu_2 - 50\omega_1 - 40\omega_2 \leq 0 \\
 & 71\mu_1 + 50\mu_2 - 40\omega_1 - 60\omega_2 \leq 0 \\
 & 98\mu_1 + 95\mu_2 - 16\omega_1 - 20\omega_2 \leq 0 \\
 & 124\omega_1 + 40\omega_2 = 1 \\
 & \mu_1, \mu_2, \omega_1, \omega_2 \geq \varepsilon \\
 & \varepsilon \text{ Non - Archimedian}
 \end{aligned} \tag{15}$$

These models represent the *optimistic point of view* and the *pessimistic point of view*. The solution from the *optimistic point of view* model defines the upper bound for the efficiency measure while the solution from the *pessimistic point of view* model determines its lower bound. Table 3 summarizes the results for all the DMUs.

In order to compare the results obtained from *Confident-DEA* with those obtained from *IDEA*, using the body of data, Table 1 were adjusted the way it shows in Table 4. Running *IDEA* requires in fact single valued data for the anchoring in both steps required by the approach. These are once again the scale transformation and variable alteration. For DMU 4, selected as the anchor for both input  $X_2$  and output  $Y_2$ , the corresponding values of data-point are changed to be single-valued (100) from a bounded one ([90,100] and [95,100]). For DMU 8, which could be selected as anchor for output  $Y_2$ , the corresponding value is also altered to the single value of 100.

Significantly, *IDEA* based measures coincide with the upper bound of efficiency confidence interval the *Confident-DEA* provides for each DMU. That is, *IDEA* is precisely the optimistic point of view case in *Confident-DEA*. Consequently, *Confident-DEA* constitutes a generalization of *IDEA*.

Regarding the benchmarking problem in *IDEA*, Cooper et al. (2001) suggested introducing dummy DMUs to overcome *IDEA*'s shortcoming requiring single-valued data for the DMU used as benchmark in scale transformations and variable alterations. To derive the efficiency measures defined by the initially non-linear problem, the back transformation are made.

**Table 1**  
**An illustrative example with imprecise data**

<b>DMUs</b>	<b>X1</b>	<b>X2</b>	<b>Y1</b>	<b>Y2</b>
<b>1</b>	124	[40 , 50]	89.8	[55 , 65]
<b>2</b>	95	[20 , 30]	99.6	[60 , 70]
<b>3</b>	92	[60 , 70]	87	[70 , 80]
<b>4</b>	61	[90 , 100]	99.4	[95 , 100]
<b>5</b>	63	[70 , 80]	96.4	[65 , 75]
<b>6</b>	50	[30 , 40]	86	[85 , 95]
<b>7</b>	40	[50 , 60]	71	[50 , 60]
<b>8</b>	16	[10 , 20]	98	[95 , 100]

**Table 2**  
**Optimistic point of view of DMU 1**

<b>DMUs</b>	<b>X1</b>	<b>X2</b>	<b>Y1</b>	<b>Y2</b>
<b>1</b>	124	40	89.8	65
<b>2</b>	95	30	99.6	60
<b>3</b>	92	70	87	70
<b>4</b>	61	100	99.4	95
<b>5</b>	63	80	96.4	65
<b>6</b>	50	40	86	85
<b>7</b>	40	60	71	50
<b>8</b>	16	20	98	95

**Table 3**  
**Pessimistic/Optimistic efficiency Coefficients**

<b>DMUs</b>	<b>Pessimistic</b>	<b>Optimistic</b>
<b>1</b>	0.1832	0.4581
<b>2</b>	0.3387	1
<b>3</b>	0.1544	0.2959
<b>4</b>	0.2660	0.2761
<b>5</b>	0.2498	0.2810
<b>6</b>	0.2808	0.6667
<b>7</b>	0.2898	0.2898
<b>8</b>	1	1

Table 4  
Adjusted Data for IDEA and Confident-DEA

DMUs	X1	X2	Y1	Y2
1	124	[40 , 50]	89.8	[55 , 65]
2	95	[20 , 30]	99.6	[60 , 70]
3	92	[60 , 70]	87	[70 , 80]
4	61	100	99.4	100
5	63	[70 , 80]	96.4	[65 , 75]
6	50	[30 , 40]	86	[85 , 95]
7	40	[50 , 60]	71	[50 , 60]
8	16	[10 , 20]	98	100

Table 5  
Comparative Results Confident-DEA and IDEA

DMUs	Pessimistic	Optimistic	IDEA
1	0.1835	0.4587	0.4581
2	0.3388	1	1
3	0.1544	0.2959	0.2959
4	0.2660	0.2660	0.2660
5	0.2498	0.2810	0.2811
6	0.2808	0.6333	0.6333
7	0.2898	0.2898	0.2898
8	1	1	1

Table 6  
Efficiency Confidence Intervals Obtained from Simulations

	DMU1	DMU2	DMU3	DMU4	DMU5	DMU6	DMU7	DMU8
<b>Highest Efficiency</b>	0.4543	1.0000	0.2938	0.2759	0.2810	0.6149	0.2898	1.0000
<b>Lowest Efficiency</b>	0.1882	0.3445	0.1544	0.2660	0.2498	0.2867	0.2898	1.0000

#### 4. Confident-DEA And The Stochastic Imprecise Data: A Simulation Approach

The case of stochastic cardinal data is addressed in this section and a methodology to conduct *Confident-DEA* is proposed. It assumes that the distribution of the values of the bounded factors over the interval is known. When this distribution is ignored, as it is the case in the illustrative example presented above and in Cooper et al (1999 and 2001a), the uniform distribution is implicitly assumed.

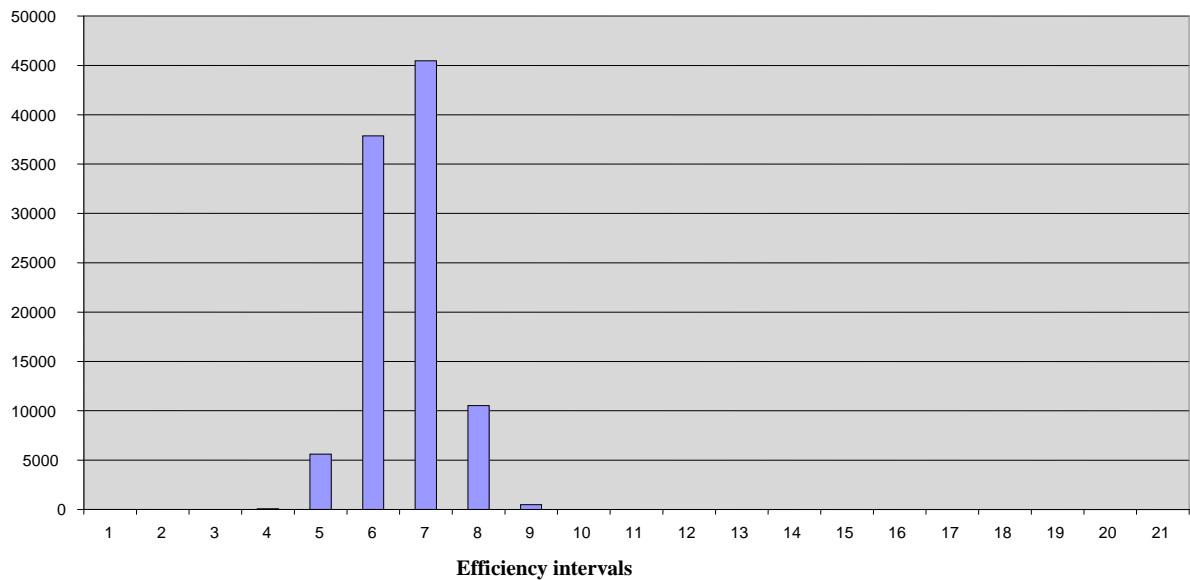
For illustration, assume that the bounded factor values in the above example have a *two-tail truncated normal distribution*. An efficiency histogram is obtained for each DMU by simulating different values for bounded factors. The [0,1] interval is divided into 20 equally large sub-intervals. The 21<sup>st</sup> corresponds to the value one e.g., full efficiency. The histogram represents the number of times each one of the efficiency sub-intervals is hit during

the total number of simulations. In this case the number was set at 100,000. Benchmarks for each intermediary level of efficiency are provided as well. To write the computer codes, *Matlab* software is used.

It should be emphasized that simulation does not yield a unique solution. The benchmarks provided may change if when the simulation is run again. The histogram can be used to define an approximate parametric distribution of efficiency coefficients over the efficiency confidence interval, however this is beyond the limits of this paper.

Table 6 provides the efficiency intervals for the eight DMUs obtained by running the simulation 100,000 times. Although a slight difference can be noticed in the boundaries of some confidence intervals compared with those provided in Table 3, the two sets of efficiency intervals are mutually compatible. The failure of the simulation to capture the exact values provided in Table 6 is a consequence of the truncated normal distribution assumption. This distribution implies that low probabilities are associated with the extreme values, and those values define the exact solutions corresponding to both the optimistic and pessimistic points of view.

**Number of times the corresponding interval is hit**



**Figure 1**  
**Efficiency histogram for DMU 1**

The efficiency histogram for DMU 1 is in Figure 1. This histogram shows that the efficiency level of [0.30,0.35] was obtained 45,463 times out of 100,000. Hence this efficiency level is the most likely, taking into account the normal distribution of values over the predefined interval for  $X_2$  and  $Y_2$ , the factor with bounded data. Hence, the probability for DMU1 reaching this level of efficiency is equal to 0.45463. For each DMU, the efficiency confidence intervals and the distributions of the efficiency values over this interval are defined.

Results for all DMUs are provided in Table 7. The DMUs are listed in columns and the efficiency bounding intervals are shown in rows. Each entry in the table indicates the likelihood of the corresponding DMU having a relative *Confident-DEA* efficiency in the corresponding efficiency interval.

Table 7  
Distribution of the efficiency measures over the *efficiency confidence interval* of each DMU

	DMU1	DMU2	DMU3	DMU4	DMU5	DMU6	DMU7	DMU8
[0.00 , 0.05[	0	0	0	0	0	0	0	0
[0.05 , 0.10[	0	0	0	0	0	0	0	0
[0.10 , 0.15[	0	0	0	0	0	0	0	0
[0.15 , 0.20[	0.00067	0	0.41545	0	0	0	0	0
[0.20 , 0.25[	0.05596	0	0.55693	0	0.99085	0	0	0
[0.25 , 0.30[	0.37858	0	0.02762	1	0.00915	0.01757	1	0
[0.30 , 0.35[	0.45463	0.00001	0	0	0	0.14994	0	0
[0.35 , 0.40[	0.10525	0.00161	0	0	0	0.36844	0	0
[0.40 , 0.45[	0.00490	0.01387	0	0	0	0.32703	0	0
[0.45 , 0.50[	0.00001	0.05742	0	0	0	0.11781	0	0
[0.50 , 0.55[	0	0.14625	0	0	0	0.01805	0	0
[0.55 , 0.60[	0	0.23110	0	0	0	0.00110	0	0
[0.60 , 0.65[	0	0.24098	0	0	0	0.00006	0	0
[0.65 , 0.70[	0	0.17154	0	0	0	0	0	0
[0.70 , 0.75[	0	0.08977	0	0	0	0	0	0
[0.75 , 0.80[	0	0.03454	0	0	0	0	0	0
[0.80 , 0.85[	0	0.01008	0	0	0	0	0	0
[0.85 , 0.90[	0	0.00229	0	0	0	0	0	0
[0.90 , 0.95[	0	0.00043	0	0	0	0	0	0
[0.95 , 1.00[	0	0.00009	0	0	0	0	0	0
1.00	0	0.00002	0	0	0	0	0	1

5. *Confident-DEA: A Genetic-Algorithm-Based Solving Method For A Mixture Of Data*

5.1. **The Context For The Genetic Algorithm:**

In previous sections, the *Confident-DEA* was introduced and illustrated in a context of cardinal data having either the form of single valued or bounded data. This section generalizes the *Confident-DEA* approach to the general situation of imprecise data. As mentioned previously, when imprecision in data is considered, the standard DEA model is not a linear program any longer. Furthermore, it can be seen as *bi-level convex* model, an *NP-hard* problem. This justifies finding heuristics solving methods. Our choice goes to genetic algorithm because, the high predisposition of the model to this meta heuristic.

Holland (1992) and his associates suggested initially in the sixties and seventies the basic principles of Genetic Algorithms. They are inspired by the mechanism of natural selection where stronger individuals are likely to be the winners in a competing environment. Through the genetic evolution method, an optimal, or a satisfactory, solution can be found and represented by the final winner of the genetic game. The name Genetic Algorithm originates from the analogy between the representation of a complex structure by means of a vector of components, and the idea of the genetic structure of *chromosomes* familiar to biologists. A vector, generally a sequence of 0-1 components, represents a chromosome and each component represents a *gene* that reflects a specific elementary characteristic. Manipulations made on chromosomes are called *genetic operators* and the most common are *crossover and mutation*.

The idea of Genetic Algorithm in optimization can be understood as an intelligent neighbouring random search method. While several methods using random sampling have been used, the *Genetic Algorithm* approach is more flexible and provides a new framework for a variety of problems.

The original version, Holland’s version, of the Genetic Algorithm works by maintaining a population of *M* chromosomes considered as potential *parents*. Each chromosome is evaluated using a given function, and assigned a *fitness value*. Each chromosome encodes a solution to the problem and its fitness value is related to the objective function value for that solution. One parent, a chromosome, is selected on a fitness basis (the better the fitness value,

the higher the chance of being chosen), while the other parent is chosen randomly. They are then *mated* by choosing a crossover point X at random, the *offspring* consists of the pre-X section from one parent followed by the post-X section of the other.

The Genetic Algorithm in general allows a *population* composed of many individuals to *evolve* under specified *selection rules* to a state that *maximizes* the *fitness*, a measure of goodness of individuals. It emulates the *survival-of-the-fittest* mechanism in nature. A mating pool is extracted from the original population of individuals or *chromosomes*. The Genetic Algorithm presumes that each chromosome, a potential candidate, can be represented by a set of parameters called *genes* and can be structured by a string of values in binary form. These selected chromosomes constitute the original set of parents.

## **5.2. Description Of The Metaheuristic Proposed:**

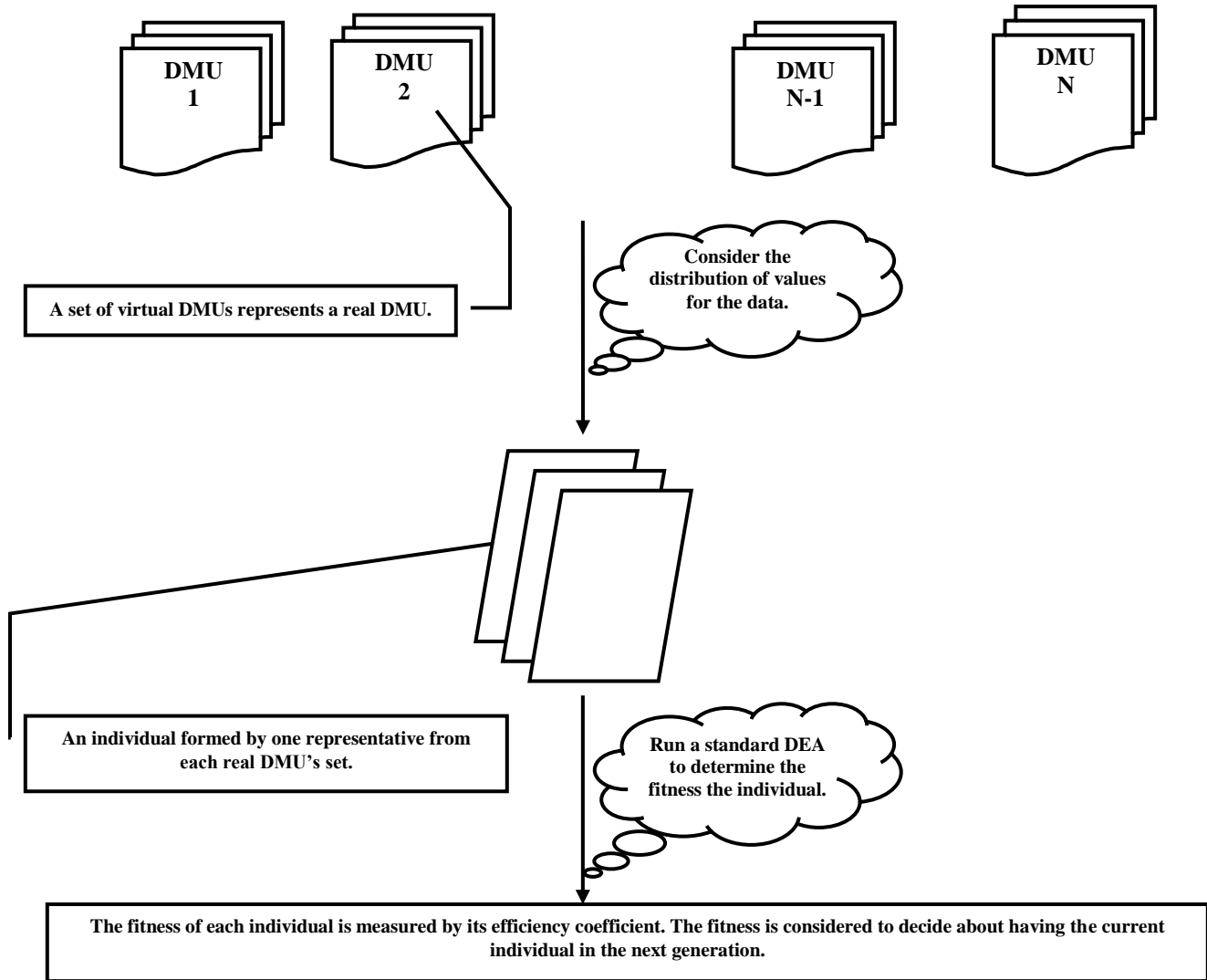
The more general case of *Confident-DEA* proposed in this section uses Genetic-Algorithm to handle a mixture of data involving ordinal, single-valued and bounded. The steps of the meta-heuristic are described in Figure5, Figure 6 and Figure 7.

As any meta-heuristic, the first step is the encoding process that enables representing DMUs in the standard form for Genetic Algorithm use. For each DMU, a string of numbers is defined (continuous or discrete) representing the values of factors. For the factors presumed to be known exactly (single valued), there will be a single-value substring for each. For the bounded factors, each will be represented by a substring containing all possible values obtained from the discretization of the corresponding range. That is, the final string of numbers representing the DMU will be composed of substrings each one representing the possible value(s) for one factor. The key idea in the *Confident-DEA* approach is to represent each DMU by a set of *chromosomes*, binary strings, in which each *gene*, 1 or 0, refers to whether or not the corresponding value is assigned to the corresponding factor.

Each DMU is split into a set of chromosomes, each one representing a *virtual single-valued* alternative for the real imprecise DMU.

For the illustration of the splitting-up process and generation of virtual DMUs, let a DMU using two inputs,  $X_1$  and  $X_2$  to produce two outputs  $Y_1$  and  $Y_2$ . Suppose that  $X_1$  and  $Y_1$  are presumed to be described by exact data while  $X_2$  and  $Y_2$  are described by bounded data.

Let  $X_1= 20$  and  $Y_1= 30$  while  $1 < X_2 < 5$  and  $11 < Y_2 < 15$ .



**Figure 5**  
**Splitting-up process and definition of an “individual”**

The factor’s order is arbitrarily chosen as  $\{X_1 ; Y_1 ; X_2 ; Y_2\}$ . The semi-columns are used only for the purpose of explanation .The string of numbers representing this DMU will then be:  $\{20 ; 30 ; 2 \ 3 \ 4 ; 12 \ 13 \ 14 \}$ . The set of chromosomes representing this DMU will then be:

$$K_1 = \{1 ; 1 ; 1 \ 0 \ 0 ; 1 \ 0 \ 0\}, K_2 = \{1 ; 1 ; 1 \ 0 \ 0 ; 0 \ 1 \ 0\}, K_3 = \{1 ; 1 ; 1 \ 0 \ 0 ; 0 \ 0 \ 1\},$$

$$K_4 = \{1 ; 1 ; 0 \ 1 \ 0 ; 1 \ 0 \ 0\}, K_5 = \{1 ; 1 ; 0 \ 1 \ 0 ; 0 \ 1 \ 0\}, K_6 = \{1 ; 1 ; 0 \ 1 \ 0 ; 0 \ 0 \ 1\},$$

$$K_8 = \{1 ; 1 ; 0 \ 0 \ 1 ; 1 \ 0 \ 0\}, K_8 = \{1 ; 1 ; 0 \ 0 \ 1 ; 0 \ 1 \ 0\}, K_9 = \{1 ; 1 ; 0 \ 0 \ 1 ; 0 \ 0 \ 1\}.$$

The factors’ value of the virtual DMU represented by  $K_1$  are:

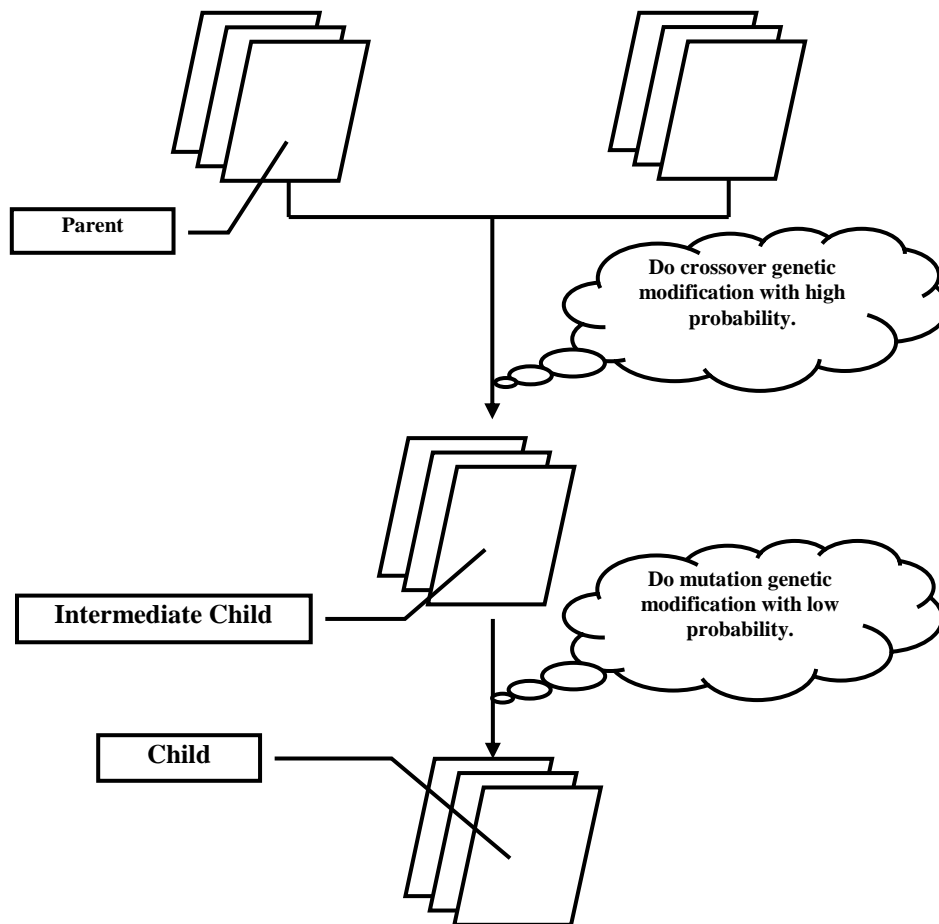
$$X_1 = 20; Y_1 = 30; X_2 = 2; Y_2 = 12.$$

By doing so for all DMUs, a set of virtual DMUs is obtained for each DMU. An *individual* is defined by a set of chromosomes determined by choosing, taking into account the distribution of imprecise values to make the approach stochastic, a representative from each set of virtual DMUs with exact data representing a real DMU. It is important to remark that, unlike the standard Genetic Algorithm procedure, an individual here is represented by a binary matrix rather than a binary string.

Once the encoding is realised, the Genetic Algorithm heuristic for *Confident-DEA* proceeds basically in two phases:

- (i) selection of the initial population using the *Roulette Wheel* method, and
- (ii) creation of the offspring using genetic modifications, to define the next generation.

Multi-point crossover with high probability, around 0.9, and mutation with low probability, in the range 0.001-0.1 are the genetic modification used in the meta-heuristic. The cutting of the matrix-individual to define the crossover points is both vertical and horizontal. The size of the initial population as well as the number of iterations is set up arbitrarily at the beginning.



**Figure 6**  
Genetic modifications: Crossover and Mutation



An initial population using the *Roulette Wheel* selection mechanisms is generated, and it constitutes the mating pool. An individual is a set of chromosomes, each one representing a DMU. All DMUs are represented in each individual and there is a single representative, a chromosome, of each DMU in each individual. The fitness function is the efficiency coefficient of the base-DMU.

Once the initial population is determined, the next phase is the creation of the next generation. This phase proceeds in three steps illustrated in Figure 6: *(i)* the mating of two selected individuals, considered as future parents *(ii)* make crossover with high probability and *(iii)* make mutation with low probability. All genetic modifications are decided based on the fitness of the individual determined by running a standard DEA model. The fitness measure is the efficiency coefficient of the base-DMU and it is computed for the selected individual at each step. Considering the binary matrix representing the individual, the corresponding virtual DMUs are identified. By solving the corresponding DEA model, the fitness, that is the efficiency coefficient of the base-DMU, is determined.

The process continues until a new generation is obtained. This new generation replaces the former generation and the process is initiated again. Iteration stops when the number of generations reaches the predetermined number.

The meta-heuristic proceeds in depth first, which means that all iterations are run for the first base-DMU to determine the lowest level of efficiency, then the iterations are run to determine the highest level. Once done with the first DMU, the process is iterated for the second base-DMU and so on.

Using this Genetic Algorithm based approach, summarized in Figure 7 and Figure 8, an upper bound and a lower bound for the efficiency coefficient of each DMU are defined. Like any heuristic or meta-heuristic, obtaining an optimal solution is not guaranteed.

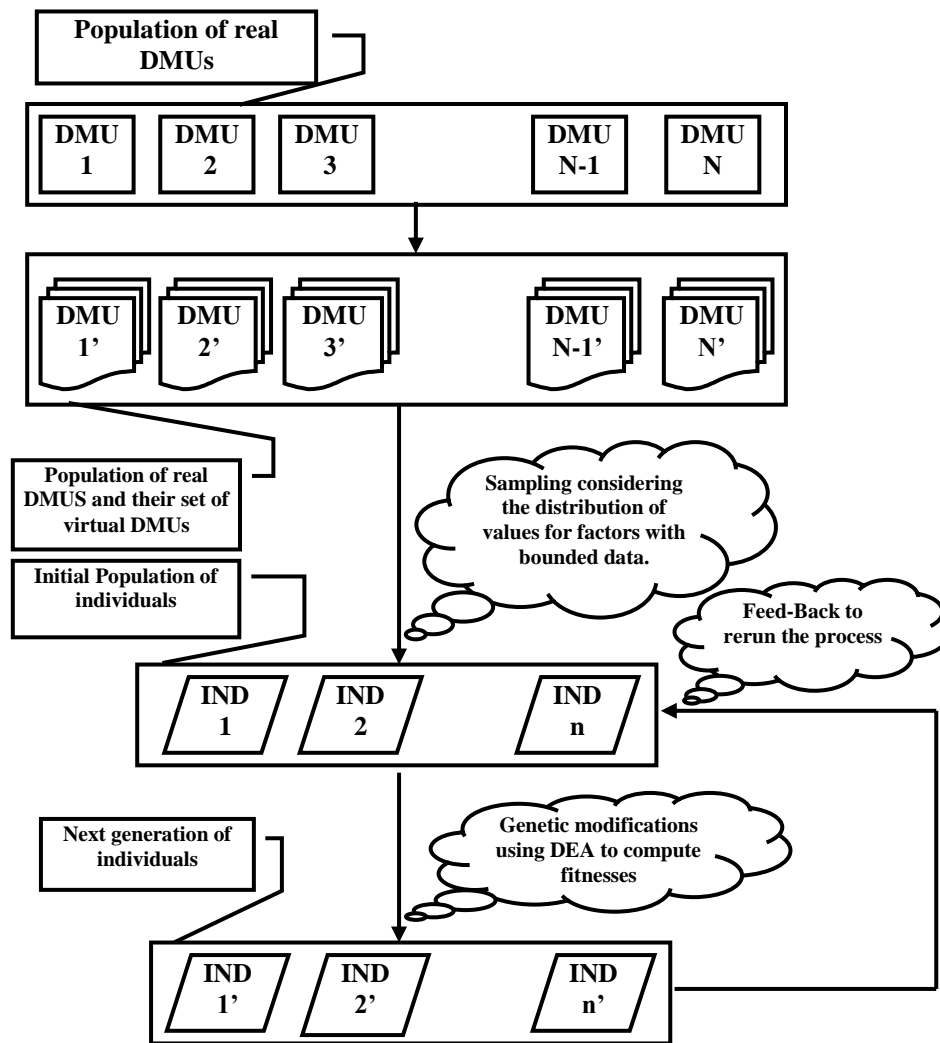


Figure 7  
Methodological Contribution: Marriage of DEA with Genetic Algorithm Procedure

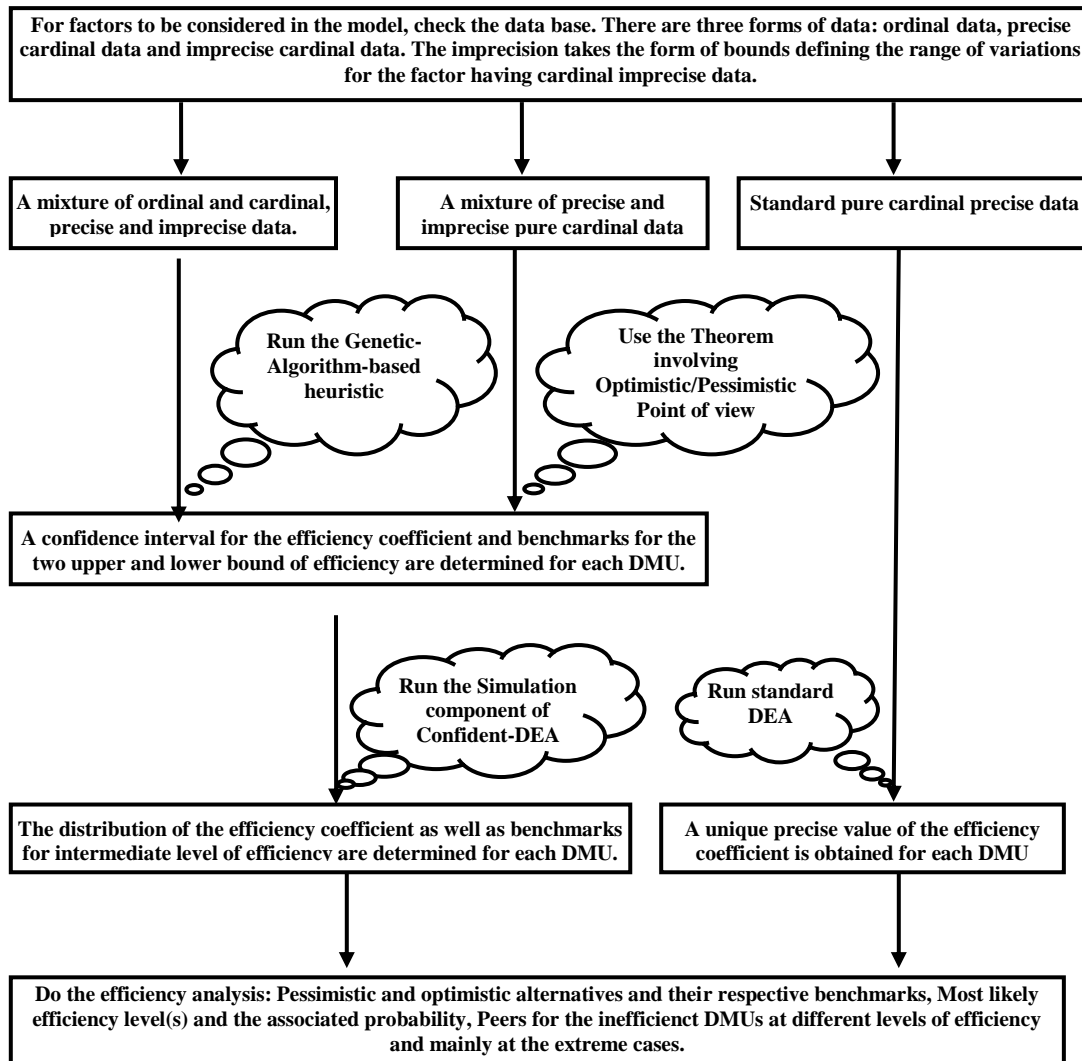
5.4. A Simulation-Based Component Of *Confident-DEA*

The third component of *Confident-DEA* is a simulation based heuristic. It proceeds in three phases:

- (i) define the individuals in the same way described for the Genetic Algorithm based heuristic,
- (ii) run a standard DEA for each individual in order to determine its efficiency coefficient and
- (iii) determine the confidence interval and the distribution of efficiency coefficient for each DMU by using a Monte Carlo type simulation.

Once an individual is chosen, the efficiency coefficient of each one of its virtual DMUs is computed by solving the corresponding standard DEA model. These values are stored for future comparison. In the next iteration, the coefficients obtained are compared with previous results in order to determine the minimum and the maximum efficiency level for each DMU. Once the predetermined number of iterations is reached, the output of the heuristic has three components. First, a confidence efficiency interval for each DMU is determined. Second, benchmarks for

different level of efficiency are identified. Finally, the distribution for the efficiency coefficient is defined based on the frequency histogram number of hits for each predefined sub-interval of [0-1]. The interval [0-1] is in fact pre-divided in a set of sub-intervals with the equal length. This predetermined length reflects the degree of precision in efficiency measure fixed by the modeler. A counter is placed in each sub-interval to record the frequency of efficiency coefficient corresponding to this sub-interval. A histogram is obtained for each DMU and the corresponding efficiency distribution is determined by smoothing the histogram.



**Figure 8**  
**The unified *Confident-DEA* approach to efficiency analysis with single-valued and imprecise data**

### 5.5. An Illustrative Example

To illustrate the methodology and for comparative purposes, consider the data contained in Cooper et al. (2001a) summarized in Table 8. Details about the description of the real-world case, an example involving efficiency evaluations of the branch offices of a mobile telecommunications corporation in Korea, can be found in the cited reference.

**Table 8**  
**Imprecise Data for an Illustrative Example**  
 (Adapted from Cooper et al., 2001)

	$X_1$	$X_2$	$X_3$	$Y_1$	$Y_2$	$Y_3$
<b>DMU 1</b>	124	18.22	4	25.53	89.8	[80;85]
<b>DMU 2</b>	95	9.23	2	18.43	99.6	[85;90]
<b>DMU 3</b>	92	8.07	6	10.29	87	[75;80]
<b>DMU 4</b>	61	5.62	8	8.32	99.4	100
<b>DMU 5</b>	63	5.33	7	7.04	96.4	[70;75]
<b>DMU 6</b>	50	3.53	3	6.42	86	[90;95]
<b>DMU 7</b>	40	3.5	5	2.2	71	[80;85]
<b>DMU 8</b>	16	1.17	1	2.87	98	[95;100]

The GA based heuristics is used to determine the bounds for the efficiency confidence interval. The results are presented in Table 9. This table also contains the efficiency measures obtained by Cooper et al (2001).

**Table 9**  
**Comparative results of IDEA and Simulated Confident-DEA with Imprecise Data**

	DMU 1	DMU 2	DMU 3	DMU 4	DMU 5	DMU 6	DMU 7	DMU 8
<b>Results of IDEA</b>	1	1	0.894	1	0.976	1	0.895	1
<b>Highest Efficiency</b>	1	1	0.8363	0.8723	0.8746	0.9842	0.8113	1
<b>Lowest Efficiency</b>	1	1	0.5941	0.7159	0.2018	0.7414	0.2546	1

One can notice a small deviation from the exact optimal solution as determined by *IDEA*. This is due to the assumption of normality, which gives small weights to extreme values where the highest and lowest values of efficiency coefficients are most likely to be reached.

We define here a new concept: the *efficiency dominance*. This dominance can have three types:

- (i) *First order efficiency dominance*: corresponds to a situation where the lowest value in the efficiency confidence interval of DMU<sub>i</sub> is greater than the highest value in the efficiency confidence interval of DMU<sub>j</sub>. In this situation, DMU<sub>i</sub> is strongly outperforming DMU<sub>j</sub>. We call this situation *First order efficiency dominance*. This is illustrated by the couple DMU2 and DMU4.
- (ii) *Second order efficiency dominance*: corresponds to a situation where the two efficiency confidence interval of DMU<sub>i</sub> and DMU<sub>j</sub> are overlapping. In this situation, DMU<sub>i</sub> is weakly outperforming DMU<sub>j</sub>. We call this situation *second order efficiency dominance*. This is illustrated by the couple DMU5 and DMU6. Studying the distribution over each confidence interval will provide additional information about comparative performances. An interesting situation here is when an interval is totally included in the second. The DMU with narrower efficiency confidence interval range is likely to be considered outperforming since it has less volatile efficiency level. This is illustrated by the couple DMU5 and DMU7. An interesting way of presenting the results is what we call the *Efficiency Cartogram*, presented in Figure 9.

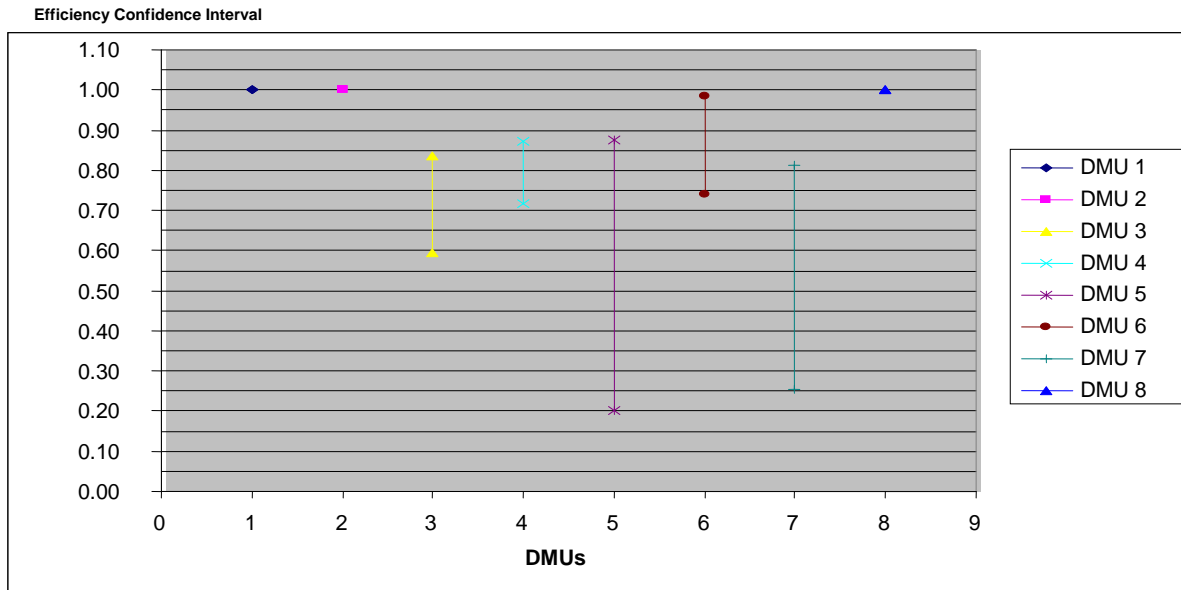


Figure 9  
Efficiency Cartogram

To determine the distribution of the efficiency measures over their corresponding efficiency confidence interval, the simulation component of Confident-DEA was run 100,000 times. The results are provided in the appendix. The efficiency histogram for DMU6 is in Figure 10.

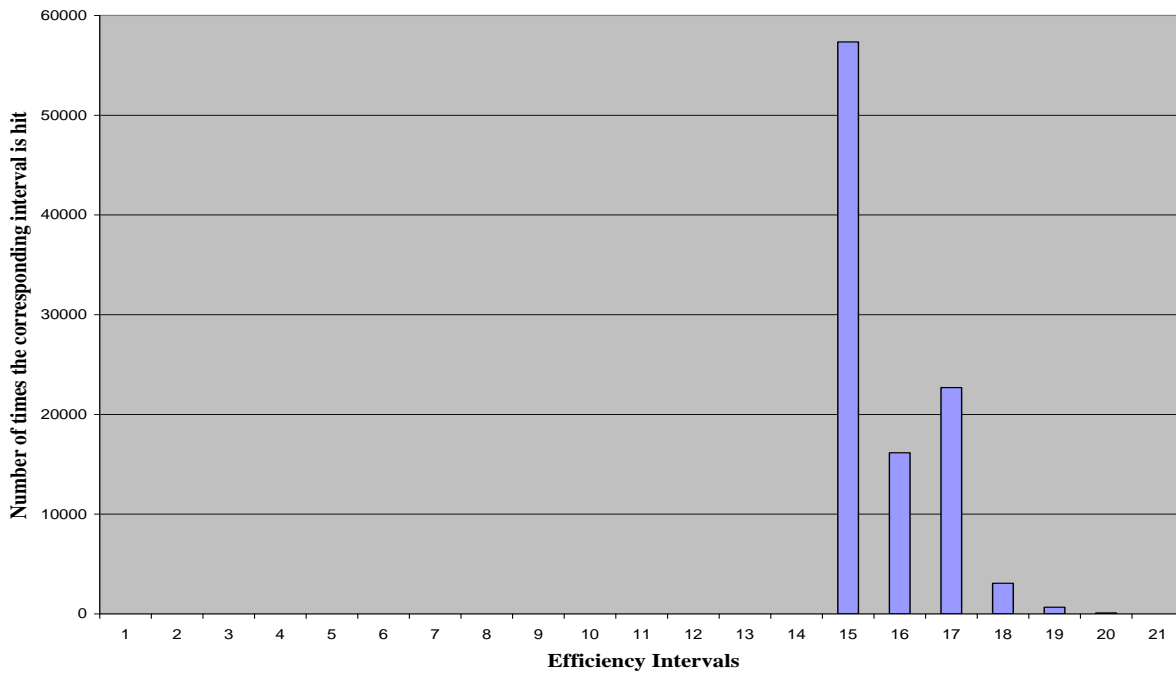


Figure 10  
Efficiency Histogram for DMU 6

As in the case of cardinal bounded data, the Monte Carlo simulation component in for *Confident-DEA* permits in the case of imprecise data the approximation of the distribution of efficiency values over the efficiency confidence interval.

## **6. Conclusions**

A new approach for dealing with the imprecision of data in DEA was presented in this paper. Called *Confident-DEA*, this approach make it possible to reflect the imprecision in data in the final efficiency coefficients by providing a "confidence interval", hence the name *confident-DEA*, for those coefficients. This generalizes and improves the more traditional *IDEA* approach suggested first by Cooper et al. (1999). First, the case of *imprecise cardinal data* e.g., a mixture of single-valued and bounded data, was considered and a theorem was suggested for determining the upper and lower bounds of the efficiency intervals. The resulting range for measures may be considered an *efficiency confidence interval*. The spread of the efficiency confidence interval in any application may be considered as a measure of the "risk" attached to the corresponding DMU: the larger is the spread of the interval, the higher is the uncertainty in the level of the corresponding DMU's efficiency and therefore the higher is the risk attached to the corresponding DMU. Also, the spread can be an indicator of *volatility for the efficiency*. The wider is the spread, the more volatile is the efficiency and hence the less is the stability of the corresponding DMU in terms of efficiency. A potential application for this is analysing the performance of shares in a stock market and providing advices about their future performances. Such work, in the best of our knowledge will represent an originality in the applications of DEA in general and *Confident-DEA* in particular.

A potential application of *Confident-DEA* is in predicting efficiency. Given the relative nature of the efficiency measures, forecasting efficiency measures cannot be done directly by using a time-series methods or econometric modelling. By predicting the production factors, one can generate prediction confidence intervals. These intervals are then considered as imprecise bounded data in a *Confident-DEA* approach to provide the efficiency confidence interval for each DMU. The results obtained from the simulation component can be used to define a parametric approximation for the distribution of efficiency measures over their corresponding confidence intervals.

Once the range for the efficiency was determined, a Monte-Carlo simulation based method was suggested to determine the distribution of the efficiency coefficients over the confidence interval. Significantly, *IDEA* always results in a single valued efficiency measure and implicitly assumes a uniform distribution for the bounded data. *Confident-DEA* on the other hand allows the use of any distribution for the bounded data. Additionally, the simulation component proposes benchmarks, in terms of inputs and outputs, for any DMU considered and for any desired level of efficiency included in its confidence interval. The use of simulation was dictated by the inexistence of analytical results about the relation between the data distribution(s) over their intervals and the distribution of the efficiency coefficients over the efficiency intervals. This remains an open research topic.

Finally, this paper presents an original formulation of the general case of DEA with imprecise data, namely single valued and bounded data as well as ordinal data, in the form of a bilevel convex model. This NP-hard problem has no exact solving method in the literature. A genetic algorithm based solving method is suggested and represents an additional original contribution of the paper. The solving method proceeds in two steps. First each DMU is split up into a set of chromosomes, each one representing a *virtual single-valued* alternative for the real imprecise DMU. An *individual* is defined by a set of chromosomes determined by choosing, taking into account the distribution of imprecise values to make the approach stochastic, a representative from each set of virtual DMUs with exact data representing a real DMU. It is important to remark that, unlike the standard Genetic Algorithm procedure, an individual here is represented by a binary matrix rather than a binary string. The second step is to do the genetic modifications on the individuals. An originality in the genetic algorithm approach proposed here is to consider horizontal and vertical cutting for the crossover, unlike the traditional multiple cutting on the chromosome string commonly used in genetic algorithm. This is believed to improve the efficiency of the algorithm, although it needs to be proven.

It can be affirmed that *Confident-DEA* generalizes *IDEA* in the sense that the efficiency levels identified by *IDEA* for each DMU coincide with the *optimistic point of view* in the *Confident-DEA* approach, in the case of bounded cardinal data. However, although the consistency between the two approaches is confirmed in the general case of imprecise data, more analytical and conceptual work is needed to establish the exact correspondence. An equivalent to the optimistic and pessimistic point of view in the case of ordinal data needs in fact to be defined.

While the Simulation-based component provides distribution for the efficiency coefficients as well as benchmarks for each intermediary level of efficiency, it is not highly efficient in finding the extreme solution. The Genetic-Algorithm-based component provides a better solution for the bounds of the efficiency interval. Combining results from both components will provide more complete and reliable information about the efficiency interval.

Finally, like any heuristic method, the genetic algorithm solving method provides, one should remind, a satisfactory solution rather than an exact solution. Hence, other solving method can be suggested and compared with the results provided by the solving method described and illustrated in this work.

## References

1. Back, T. (1996), *Evolutionary algorithms in theory and practice*, Oxford University Press, New York.
2. Bard, J. F. (1998), *Practical Bilevel Optimization: Algorithms and Applications*, Kluwer Academic Publishers, Dordrecht, The Netherlands.
3. Charnes, A., Cooper, W.W. (1985), Preface to topics in Data Envelopment Analysis, *Annals of Operations Research* 2, 1, 59-94.
4. Charnes, A., Cooper, W.W., and Rhodes, E. (1978), Measuring the efficiency of decision making units *European Journal of Operational Research* 2, 429-444..
5. Cook, W. D., Kress, M. and Seiford, L. M. (1993), On the use of ordinal data in Data Envelopment Analysis, *Journal of Operational Research Society* 44, 2, 133-140.
6. Cook, W. D., Kress, M. and Seiford, L. M. (1996), Data Envelopment Analysis in the presence of both quantitative and qualitative Factors, *Journal of Operational Research Society* 47, 7, 945-953.
7. Cooper, W.W., Park, K. S. and Yu, G. (2001), An illustrative application of IDEA (Imprecise Data Envelopment Analysis) to a Korean Mobile Telecommunication Company, *Operations Research* 49, 6, 807-820.
8. Cooper, W.W., Park, K. S. and Yu, G. (2001), IDEA (Imprecise Data Envelopment Analysis) with CMDS (Column Maximum Decision Making Units), *Journal of Operational Research Society*, 52, 2, 178-181
9. Cooper, W. W., Park, K. S. and Yu, G. (1999), IDEA and AR-IDEA: models for dealing with imprecise data in DEA, *Management Science* 45, 4, 597-607.
10. Dantzig, G. B. (1953), Computational algorithms of the revised simplex method Report, RM 1266, The Rand Corporation, Santa Monica, CA.
11. Gattoufi, S., Oral, M., Kumar, A., Reisman, A., 2004a. Content Analysis of Data Envelopment Analysis Literature and its Comparison with that of other OR/MS Fields. *Journal of the Operational Research Society* 55(9), 911-932.
12. Gattoufi, S., Oral, M., Kumar, A., Reisman, A., 2004b. Epistemology of Data Envelopment Analysis and Comparison with Other Fields of OR/MS for Relevance to Applications. *Socio Economic Planning Sciences* 38(2-3), 123-140.
13. Golany, B. (1988), A note on including ordinal relations among multipliers in Data Envelopment Analysis, *Management Science* 34, 8, 1029-1033.
14. Goldberg, D. E. (1989), *Genetic algorithms in search, optimization and machine learning*, Addison-Wesley Publishing Company, Reading, Massachusetts.
15. Hansen, P., Jaumard, B. and Savard, G. (1992), New Branch-and-Bound rules for linear bilevel programming, *SIAM Journal of Scientific and Statistical Computing*, 13, 5, 1194-1217.
16. Holland, J. H. (1992), *Adaptation in Natural and Artificial Systems*, Second Edition, The MIT Press, Cambridge, Massachusetts.
17. Jeroslow, R. G. (1985), The polynomial hierarchy and a simple model for competitive analysis, *Mathematical Programming*, 32, 146-164.

18. Kim, H. S., Park, C. G. and Park, K. S. (1999), An application of Data Envelopment Analysis in telephone offices evaluation with partial data, *Computers & Operations Research* 26, 1, 59-72.
19. Lee, Y.K., Park, K.S., Kim, S.H. (2002), Identification of inefficiencies in an additive model based IDEA (imprecise data envelopment analysis). *Computers & Operations Research* 29, 12 ,1661-76.
20. Man, K. F., Tang, K. S., Kwong, S. (1999), *Genetic algorithms: concepts and designs*. Springer-Verlag, London, England.
21. Post, T., Spronk, J., (1999), Performance benchmarking using interactive data envelopment analysis, *European Journal of Operational Research*, 115, 3, 472-487.
22. Zhu, J. (2004) Imprecise DEA via Standard Linear DEA Models with a Revisit to a Korean Mobile Telecommunication Company, *Operations Research*, 52, 2, 323-329.

**NOTES**