International Journal of Management Science
Vol 14, No 2, November 2008

Imprecise DEA Efficiency Assessments : Characterizations and Methods*

Kyung Sam Park**
Korea University, Business School
Anam-5Ga, Seongbuk, Seoul 136-701, Korea

(Received: May 26, 2008 / Accepted: June 16, 2008)

ABSTRACT

Data envelopment analysis (DEA) has proven to be a useful tool for assessing efficiency or productivity of organizations which is of vital practical importance in managerial decision making. While DEA assumes exact input and output data, the development of imprecise DEA (IDEA) broadens the scope of applications to efficiency evaluations involving imprecise information which implies various forms of ordinal and bounded data possibly or often occurring in practice. The primary purpose of this article is to characterize the variable efficiency in IDEA. Since DEA describes a pair of primal and dual models, also called envelopment and multiplier models, we can basically consider two IDEA models: One incorporates imprecise data into envelopment model and the other includes the same imprecise data in multiplier model. The issues of rising importance are thus the relationships between the two models and how to solve them. The groundwork we will make includes a duality study which makes it possible to characterize the efficiency solutions from the two models. This also relates to why we take into account the variable efficiency and its bounds in IDEA that some of the published IDEA studies have made. We also present computational aspects of the efficiency bounds and how to interpret the efficiency solutions.

Keywords: DEA, Duality, Efficiency Evaluation, Imprecise Data, Linear Programming

1. Introduction

This paper is concerned with the use of imprecise data in data envelopment analysis (DEA). Imprecise data implies that some data are known only to the extent that the true values lie within prescribed bounds while other data are known only in terms of

* The author is grateful for financial supports from the Korea Research Foundation, Grant No. R05-2004-000-11138-0, and from a Korea University Grant.
** Corresponding author, E- mail: sampark@korea.ac.kr

ordinal relations. Cook *et al.* [6, 7] showed how DEA could be extended to treat ordinal data. To deal with all aspects of imprecise data in DEA, Cooper *et al.* [9] proposed a body of concepts and methods that go by the name of imprecise data envelopment analysis (IDEA). There have since been a number of refinements, extensions, and applications [5, 8, 10-16, 19, 20, 26, 27]. These studies have also developed different methods for solving a nonlinear IDEA problem because some inputs and outputs are unknown decision variables with values to be determined in the model.

Although the computational algorithms are different, they result in the same efficiency scores and, hence, the same efficiency classifications into efficient and inefficient groups. This results from the fact that they target the same model that is formed by incorporating imprecise data into the dual (or multiplier) side of DEA. This brings us into play a question of what happens to other models that can be thought of as embedding the same imprecise data in the primal (or envelopment) side of DEA. This development still needs to be made to enable us to obtain other measures of efficiency from envelopment models synthesized from imprecise data. Still further studies could then be made to understand the relationships that characterize the efficiency solutions for these two different, but related, formulations.

We develop a duality theory for the two models whereby we uncover the characteristics of their efficiency solutions. The starting point for our developments is a fresh view of DEA that leads to an understanding of how imprecise data force a reconsideration of central tenets in DEA. Basically, the exact data assumed in DEA imply that every DMU utilizes and produces exact amounts of inputs and outputs so a fixed measure of efficiency can be obtained along with the efficient and inefficient partitions. In contrast, the presence of imprecise data involves variable resources and outputs. This leads to a variable measure on efficiency. To realize that the variable efficiency lies within an upper and lower bound, we develop two distinct approaches, which in turn characterize the solutions to the multiplier and envelopment models in the presence of imprecise data. We also present generalized methods to obtain upper and lower bounds on efficiency. This gives rise to a new scheme of efficiency classifications that is more detailed than the customary two-group partition. Note that our developments are all made in a general form to treat arbitrary types of imprecise data in DEA.

It should be noted that there have been other studies related to ours. Despotis and Smirlis [12] treated bounded or interval data in DEA and showed an efficiency classification analogous to ours. Kao and Liu [14, 15] dealt with a slightly different situation in which fuzzy data were involved in DEA, referred to as fuzzy DEA. Despite the different kinds of data, the fuzzy approach generates interval data from the given fuzzy data so it immediately follows that the resulting efficiency partition will also be similar to ours. In fact, when dealing with interval data alone, it is relatively easy to obtain the efficiency bound that results in the efficiency partition. Recently, Kao [13] developed a different method to handle ordinal data in DEA and arrive at an upper and lower bound on efficiency. He also pointed out that the original IDEA model of Cooper *et al.* [9] yielded an upper bound. However, he confined his attention to some special imprecise data, pure interval and pure ordinal data, and did not look at the envelopment model involving imprecise data. Unlike this, our study targets imprecise data in a more general form, including any combinations of bounded and ordinal data, and characterizes the solutions for both multiplier and envelopment models in the situation of arbitrary imprecise data.

It is also important to note that Chen *et al.* [5] have investigated not only a similar three-group classification but also the existence of duality gap between the multiplier and envelopment models when imprecise data are imposed. They showed that, for some special cases where the information set on a factor was specified entirely by weak orders or prescribed bounds, the multiplier model could yield an upper bound and the envelopment model a lower bound on efficiency. Our approach uses a unique instrument, linear program, no matter what types of imprecise data are given. Moreover, they did not provide the underlying characteristics and dualities of the two models generally and rigorously.

The paper is organized as follows. First, the multiplier and envelopment DEA models with imprecise data are shown. We then define two distinct approaches for establishing a certain bound on the variable efficiency. This is followed by the duality results which characterize the efficiency solutions to the multiplier and envelopment models. After this has been done we present generalized methods to achieve the upper and lower bound. We then show how to interpret the efficiency solutions. Finally, we conclude the paper with a summary and a sketch of further research opportunities.

2. Preliminaries and definitions

The IDEA model proposed by Cooper *et al.* [9] can be represented by

$$\left.\begin{aligned} z_o^* &= \max \textstyle\sum_{r=1}^{s} \mu_r y_{ro} \\ \text{s.t.} \quad & \textstyle\sum_{r=1}^{s} \mu_r \mathbf{y}_r - \sum_{i=1}^{m} \omega_i \mathbf{x}_i \leq \mathbf{0} \\ & \textstyle\sum_{i=1}^{m} \omega_i x_{io} = 1 \\ & \mu = (\mu_r) \geq \varepsilon; \ \omega = (\omega_i) \geq \varepsilon \end{aligned}\right\} \quad (1.1)$$

$$\left.\begin{aligned} \mathbf{y}_r \in D_r^+, \quad r = 1, \cdots, s \\ \mathbf{x}_i \in D_i^-, \quad i = 1, \cdots, m \end{aligned}\right\} \quad (1.2)$$

Here, $\mathbf{y}_r = (y_{r1}, \cdots, y_{rn})^{\mathrm{T}}$ and $\mathbf{x}_i = (x_{i1}, \cdots, x_{in})^{\mathrm{T}}$ respectively represent the column vectors of outputs produced and inputs consumed by n DMUs under consideration. The y_{ro}, x_{io} data represent the outputs and inputs for DMU$_o$, as the DMU$_j$ to be evaluated, $j = 1, \cdots, n$. The sets of variables μ and ω are multipliers associated with outputs and inputs and $\varepsilon > 0$ is a non-Archimedean element. The sets D_r^+, D_i^- in (1.2) represent the constraint sets of imprecise data for $\mathbf{y}_r$ and $\mathbf{x}_i$, respectively. As mentioned before, imprecise data encompass both ordinal and bounded data. More generally, each of the D_r^+, D_i^- sets can be assumed to be a convex polyhedron in $\Re^n$, formed by a system of linear inequalities representing arbitrary linear-type imprecise data. Throughout we assume $\mathbf{y}_r \geq 0$, $\mathbf{y}_r \neq 0$ $\forall r$ and $\mathbf{x}_i \geq 0$, $\mathbf{x}_i \neq 0$ $\forall i$.

Model (1) is a nonlinear program since the input and output data are known imprecisely and their values are to be determined. However, several methods have been developed to transform the nonlinear model into a linear program [10, 11, 19, 26]. The optimal solution z_o^* to (1) then represents the efficiency score of the DMU$_o$ to be evaluated. A two-group efficiency classification can then be made in which DMU$_o$ is classified as efficient if $z_o^* = 1$, otherwise it is classified as inefficient.

IDEA approach (1) incorporates imprecise data into the multiplier side of DEA. This is one possible way, however, because one can think of another approach in which the same imprecise data are included in the envelopment model as follows:

$$\left.\begin{aligned} f_o^* &= \min \theta - \varepsilon(\mathbf{1s}^+ + \mathbf{1s}^-) \\ \text{s.t.} \quad & \lambda \mathbf{y}_r - s_r^+ = y_{ro} \quad \forall r \\ & \lambda \mathbf{x}_i + s_i^- = \theta x_{io} \quad \forall i \\ & \mathbf{s}^+, \mathbf{s}^-, \lambda \geq \mathbf{0} \end{aligned}\right\} \quad (2.1)$$

$$\left.\begin{aligned} \mathbf{y}_r \in D_r^+, \quad r = 1, \cdots, s \\ \mathbf{x}_i \in D_i^-, \quad i = 1, \cdots, m \end{aligned}\right\} \quad (2.2)$$

where $\lambda = (\lambda_j) \in \Re^n$, $\mathbf{s}^+ = (s_r^+)^T \in \Re^s$, $\mathbf{s}^- = (s_i^-)^T \in \Re^m$, and $\mathbf{1} = (1, \cdots, 1)$ is a vector of ones of appropriate size. The imprecise data sets D_r^+, D_i^- in (2.2) are the same as in (1.2).

If all the $\mathbf{y}_r$, $\mathbf{x}_i$ data are exact, then models (1) and (2) become a dual pair and $z_o^* = f_o^*$ holds by the dual theorem of linear programming. This duality does not hold for the nonlinear problems in (1) and (2) and, hence, different efficiency solutions will yield different efficiency classifications from these two models. Of particular interest now is how to obtain the optimal solution to the newly constructed model in (2) and the relationships that characterize the efficiency solutions for these two different models. We address these issues below. A point to be noted is that the previous IDEA studies have dealt primarily with model (1) only and developed methods for achieving its solution. We may therefore say that the half of what is needed has already been done but the other half is not taken into account.

3. Two distinct approaches

We begin with the following definitions to guide our developments:

- **Definition 1** (Potential efficiency): The DMU$_o$ to be evaluated is *potentially efficient* if and only if there exist (i) at least one set of variable values satisfying $\mu, \omega \geq \varepsilon$, and (ii) at least one set of input-output data for DMU$_o$ satisfying $\mathbf{y}_r \in D_r^+$, $\mathbf{x}_i \in D_i^-$ $\forall r, i$, which together render DMU$_o$ efficient (i.e., 100% efficiency).

- **Definition 2** (Perfect efficiency): The DMU$_o$ is *perfectly efficient* if and only if the maximal efficiency of DMU$_o$ is 100%, for (i) some variable values satisfying $\mu, \omega \geq \varepsilon$, and for (ii) *all* data variable values satisfying $\mathbf{y}_r \in D_r^+$, $\mathbf{x}_i \in D_i^-$ $\forall r, i$.

The defined potential efficiency is to be measured in a manner that is efficient for some positive variable values and for *some* data variable values satisfying the data constraints given in advance. In contrast, perfect efficiency is measured in a more stringent manner than potential efficiency. The only difference between Definitions 1 and 2 is in part (ii): Definition 1 requires the efficiency to be obtained for *some* data variable values while Definition 2 refers to the efficiency for *all* data variable values in the feasible region specified by imprecise data. We can represent these definitions in the following mathematical formulations:

- MIDEA-P model:

$$\begin{aligned} u_o^* &= \max \sum_{r=1}^{s} \mu_r y_{ro} \\ \text{s.t.} \quad & \sum_{r=1}^{s} \mu_r \mathbf{y}_r - \sum_{i=1}^{m} \omega_i \mathbf{x}_i \le \mathbf{0} \\ & \sum_{i=1}^{m} \omega_i x_{io} = 1 \\ & \mu, \omega \ge \varepsilon \\ \text{for some} & \begin{Bmatrix} \mathbf{y}_r \in D_r^+, & r = 1, \cdots, s \\ \mathbf{x}_i \in D_i^-, & i = 1, \cdots, m \end{Bmatrix} \end{aligned} \tag{3}$$

- MIDEA-F model:

Model (3) but with l_o^* in place of u_o^* and "for all" in place of "for some." (4)

We refer to (3) as MIDEA-P model in the sense that imprecise data are incorporated into the multiplier DEA side and this is to determine potential efficiency. Similarly (4) is referred to as MIDEA-F model where F means perfect (or Full) efficiency.

Consider a data matrix $[\mathbf{Y}^p, \mathbf{X}^p] = [(\mathbf{y}_r^p), (\mathbf{x}_i^p)]$ chosen from the given imprecise data, $\mathbf{y}_r^p \in D_r^+$, $\mathbf{x}_i^p \in D_i^-$ $\forall r, i$. The symbol p indicates a choice-making strategy for imprecise data. Let $M(p) = \{(\mu, \omega) \mid \mu\mathbf{Y}^p - \omega\mathbf{X}^p \le 0;\ \omega\mathbf{x}_o^p = 1;\ \mu, \omega \ge \varepsilon\}$ be the feasible region of multipliers μ, ω. Then, clearly $l_o^* \le u_o^*$ because the size of $M(p \equiv$ "for some") in (3) is larger than $M(p \equiv$ "for all") in (4). Moreover, (3) can be assumed to have the largest set of M which contains the possible values of multiplier variables, while (4) is supposed to have the smallest set. We can thus use u_o^* as an upper bound for the efficiency that the evaluated DMU_o can have, and l_o^* as a lower bound.

Note that the upper and lower bound on efficiency we just mentioned is within the technology function as defined in the condition part of M applied to both (3) and (4). This technology function is based on the ordinary DEA formulation, referred to as the CCR (Charnes, Cooper, and Rhodes [4]) model, in which all the $\mathbf{y}_r$, $\mathbf{x}_i$ data are assumed to be exact. With the assumption of exact data it is not necessary to consider any optimization strategy for the data. The only consideration that is mandatory is thus the optimization of multiplier variables from which $l_o^* = u_o^*$ holds. We also note that the same strategy is applied to the same multiplier variables in models (3) and (4), as denoted in part (i) of Definitions 1 and 2.

However, since we assume imprecise data in DEA, we need to undertake another task in order to treat the variables involving imprecise data. According to the way we take into account the data variables, it is thus possible to obtain different efficiency ratings for each DMU. Based on the possible optimization principles for models (3) and (4), we then find that true efficiency, which is unknown, lies within the upper and lower bounds represented by u_o^* and l_o^*. This represents a new development for DEA with imprecise data. Hence we develop it in more detail as follows.

4. Duality Results

First, we study duality for *potential efficiency*, which establishes a link between the multiplier and envelopment models involving imprecise data and characterizes the efficiency solution to the original IDEA model in (1). To accomplish this, we first define the following new model:

- EIDEA-P model:

$$
\begin{aligned}
\alpha_o^* = \min\ & \theta - \varepsilon(\mathbf{1s}^+ + \mathbf{1s}^-) \quad \text{s.t.} \\
& \lambda \quad \mathbf{y}_r - s_r^+ = y_{ro} \quad \forall r \\
& \lambda \mathbf{x}_i + s_i^- = \theta x_{io} \quad \forall i \\
& \mathbf{s}^+, \mathbf{s}^-, \lambda \geq \mathbf{0} \\
\text{for all } & \begin{Bmatrix} \mathbf{y}_r \in D_r^+,\ r = 1, \cdots, s \\ \mathbf{x}_i \in D_i^-,\ i = 1, \cdots, m \end{Bmatrix}
\end{aligned}
\tag{5}
$$

The notations in this model are as defined in model (2). We refer to (5) as EIDEA-P model in the sense that imprecise data are incorporated into the envelopment DEA model. Note that model (5) has an explicit choice-making strategy for imprecise data, "for all" permissible data, while model (2) has no such a strategy explicitly for the same imprecise data. The following proposition shows the duality result for potential efficiency:

Proposition 1: The optimal objective values of (3) and (5) are equal, $u_o^* = \alpha_o^*$, and hence the dual to MIDEA-P model (3) is EIDEA-P model (5).

Proof: Because of the notion of maximization "for some" permissible data, MIDEA-P model (3) can be replaced by the following equivalent:

$$
\begin{aligned}
u_o^* = \max_{\substack{\mathbf{y}_r \in D_r^+ \ \forall r \\ \mathbf{x}_i \in D_i^- \ \forall i}} u_o &= \max_{\mu,\omega} \sum_{r=1}^{s} \mu_r y_{ro} \\
\text{s.t.} \qquad \sum_{r=1}^{s} \mu_r \mathbf{y}_r &- \sum_{i=1}^{m} \omega_i \mathbf{x}_i \le \mathbf{0} \\
\sum_{i=1}^{m} \omega_i x_{io} &= 1 \\
\mu, \omega &\ge \varepsilon
\end{aligned}
\tag{6}
$$

Now consider

$$
\begin{aligned}
\varphi_o^* = \max_{\substack{\mathbf{y}_r \in D_r^+ \ \forall r \\ \mathbf{x}_i \in D_i^- \ \forall i}} \varphi_o &= \min_{\substack{\theta,\lambda \\ \mathbf{s}^+,\mathbf{s}^-}} \theta - \varepsilon(\mathbf{1s}^+ + \mathbf{1s}^-) \\
\text{s.t.} \qquad & \lambda \mathbf{y}_r - s_r^+ = y_{ro} \quad \forall r \\
& \lambda \mathbf{x}_i + s_i^- = \theta x_{io} \quad \forall i \\
& \mathbf{s}^+,\ \mathbf{s}^-,\ \lambda \ge \mathbf{0}
\end{aligned}
\tag{7}
$$

We show $u_o^* = \varphi_o^*$ in (6) and (7) as follows: Assume that there exists a matrix of input-output data values, $[(\mathbf{y}_r{}^p), (\mathbf{x}_i{}^p)]$ satisfying $\mathbf{y}_r{}^p \in D_r{}^+$, $\mathbf{x}_i{}^p \in D_i{}^-$ $\forall r, i$. These exact data can be used simultaneously in (6) and (7) to obtain

$$
\begin{aligned}
u_o^{(p)} \equiv \max \sum_{r=1}^{s} \mu_r y_{ro}^p \quad & \text{s.t.} \\
\sum_{r=1}^{s} \mu_r \mathbf{y}_r^p - \sum_{i=1}^{m} \omega_i \mathbf{x}_i^p &\le \mathbf{0} \\
\sum_{i=1}^{m} \omega_i x_{io}^p &= 1 \\
\mu, \omega &\ge \varepsilon
\end{aligned}
$$

and

$$
\begin{aligned}
\varphi_o^{(p)} \equiv \min \theta - \varepsilon(\mathbf{1s}^+ + \mathbf{1s}^-) \quad & \text{s.t.} \\
\lambda \mathbf{y}_r^p - s_r^+ = y_{ro}^p \quad & \forall r \\
\lambda \mathbf{x}_i^p + s_i^- = \theta x_{io}^p \quad & \forall i \\
\mathbf{s}^+, \mathbf{s}^-, \lambda \ge \mathbf{0} &
\end{aligned}
$$

We now have a pair of primal and dual models in DEA, and no problem of existence of finite optima for the latter primal model because the data for DMU_o appear on both sides of the constraints. Hence we have $u_o^{(p)} = \varphi_o^{(p)}$ by the dual theorem of linear programming. It follows that $u_o^{(p)} = \varphi_o^{(p)}$ for the same p. Moreover, we have $[(\mathbf{y}_r^*), (\mathbf{x}_i^*)]$ such that $u_o^* = \max u_o^{(p)} = \max \varphi_o^{(p)} = \varphi_o^*$, since the same objective and the same data constraints are used in the left-hand side of both (6) and (7). Therefore, $u_o^* = \varphi_o^*$ holds in (6) and (7).

Finally, it is obvious that $\alpha_o^* = \varphi_o^*$ holds in (5) and (7) because of the notion of the max-min structure in (7). Consequently, $u_o^* = \alpha_o^*$ holds and hence the dual to MIDEA-P model (3) is EIDEA-P model (5), and vice versa. □

Proposition 1 implies that, per the envelopment DEA model, if we take an ultraconservative strategy in choosing exact data from the given imprecise data, then we obtain the same upper bound on efficiency as in MIDEA-P model (3). Note that we have used "maximization" in the objective of data variables in both (6) and (7) to obtain potential efficiency. However, the meanings are quit different from each other. As defined in MIDEA-P model (3), "maximization" for the data variables taken in (6) represents the most optimistic strategy to reflect "for some" permissible data. In sharp contrast, "maximization" in (7) signifies an ultraconservative strategy because the immediately following objective is to be minimized. This strategy reflects "for all" permissible data as shown in EIDEA-P model (5).

Proposition 2: Both optimal objective values in (1) and (5) are equal, $z_o^* = \alpha_o^*$.

Proof: By definition, we have $z_o^* \le u_o^*$ in (1) and (6). Now reconsider (6) and assume that there exists a matrix of data values $[(\mathbf{y}_r^*), (\mathbf{x}_i^*)]$ such that we obtain the maximal value of $u_o = \sum \mu_r y_{ro}$ to be maximized subject to the constraints in (6). This yields a value for u_o^* together with optimal multipliers (μ^*, ω^*). We then find that the set of optimal solutions $[(\mathbf{y}_r^*), (\mathbf{x}_i^*)]$, (μ^*, ω^*) to (6) is feasible in model (1) since the same constraints are used for both (1) and (6). It follows that $z_o^* \ge u_o^*$ and hence we have $z_o^* = u_o^*$. Since $u_o^* = \alpha_o^*$ in Proposition 1, we finally have $z_o^* = \alpha_o^*$ in (1) and (5) □

Proposition 2 means that the original IDEA model in (1) follows the strategy implied by Definition 1 so it yields the same upper bound. This also implies that if we incorporate imprecise data directly (i.e., without imposing any choice-making strategy explicitly for imprecise data) into the multiplier DEA model, then it generates the upper bound.

Next, we provide the duality result for *perfect efficiency* in Definition 2 or MIDEA-F model (4). We define

- EIDEA-F model:

$$\begin{aligned} \beta_o^* &= \min\ \theta - \varepsilon(\mathbf{1s}^+ + \mathbf{1s}^-) \quad \text{s.t.} \\ &\lambda \mathbf{y}_r - s_r^+ = y_{ro} \quad \forall r \\ &\lambda \mathbf{x}_i + s_i^- = \theta x_{io} \quad \forall i \\ &\mathbf{s}^+, \mathbf{s}^-, \lambda \geq \mathbf{0} \\ \text{for some } &\begin{Bmatrix} \mathbf{y}_r \in D_r^+,\ r = 1, \cdots, s \\ \mathbf{x}_i \in D_i^-,\ i = 1, \cdots, m \end{Bmatrix} \end{aligned} \tag{8}$$

Proposition 3: The optimal objective values of (4) and (8) are equal, $l_o^* = \beta_o^*$, and hence the dual to MIDEA-F model (4) is EIDEA-F model (8).

Proof: Because of the notion of maximization "for all" permissible data, MIDEA-F model (4) can be replaced by the following equivalent:

$$\begin{aligned} l_o^* = \min_{\substack{\mathbf{y}_r \in D_r^+\ \forall r \\ \mathbf{x}_i \in D_i^-\ \forall i}} l_o &= \max_{\mu,\omega} \sum\nolimits_{r=1}^{s} \mu_r y_{ro} \\ \text{s.t.} \qquad &\sum\nolimits_{r=1}^{s} \mu_r \mathbf{y}_r - \sum\nolimits_{i=1}^{m} \omega_i \mathbf{x}_i \leq \mathbf{0} \\ &\sum\nolimits_{i=1}^{m} \omega_i x_{io} = 1 \\ &\mu, \omega \geq \varepsilon \end{aligned} \tag{9}$$

We now consider

$$\begin{aligned} \eta_o^* = \min_{\substack{\mathbf{y}_r \in D_r^+\ \forall r \\ \mathbf{x}_i \in D_i^-\ \forall i}} \eta_o &= \min_{\substack{\theta,\lambda \\ \mathbf{s}^+,\mathbf{s}^-}} \theta - \varepsilon(\mathbf{1s}^+ + \mathbf{1s}^-) \\ \text{s.t.} \qquad &\lambda \mathbf{y}_r - s_r^+ = y_{ro} \quad \forall r \\ &\lambda \mathbf{x}_i + s_i^- = \theta x_{io} \quad \forall i \\ &\mathbf{s}^+, \mathbf{s}^-, \lambda \geq \mathbf{0} \end{aligned} \tag{10}$$

We then have $l_o^* = \eta_o^*$ in (9) and (10), which can be proven similarly as $u_o^* = \varphi_o^*$ is proven for (6) and (7) in Proposition 1.

Finally, it is obvious that $\beta_o^* = \eta_o^*$ holds in (8) and (10) because of the notion of the min-min structure in (10). Consequently, $l_o^* = \beta_o^*$ holds and hence the dual to MIDEA-F model (4) is EIDEA-F model (8), and vice versa. □

Proposition 4: Both optimal objective values in (2) and (8) are equal, $f_o^* = \beta_o^*$.
Proof: This can be proven similarly as done in Proposition 2. □

Note again that the meaning of "minimization" in the objective of data variables in both (9) and (10) is different. As defined in MIDEA-F model (4), "minimization" for the data variables taken in (9) represents an ultraconservative strategy to reflect "for all" permissible data. In contrast, "minimization" in (10) signifies the most optimistic strategy because the immediately following objective is to be minimized, thus reflecting "for some" permissible data as in EIDEA-F model (8). Therefore, we know from Proposition 3 that, per the envelopment DEA model, if we take the most optimistic strategy in choosing exact data from the given imprecise data, then we obtain the same lower bound as in MIDEA-F model (4). Proposition 4 means that if we incorporate imprecise data directly (i.e., without imposing any choice-making strategy explicitly for imprecise data) into the envelopment DEA model, then it also yields the same lower bound.

5. Efficiency Classifications

We set aside computational aspects of potential and perfect efficiencies for the moment and turn to efficiency classifications based upon the results of the previous sections. Let t_o be the true unknown efficiency for DMU$_o$. We then recognize that the true efficiency lies between the lower and upper bound on efficiency, $0 \leq l_o^* \leq t_o \leq u_o^* \leq 1$. Of course, $l_o^* = u_o^*$ if all the data are known exactly. Based on this, we can classify DMUs into three categories: (i) inefficient, (ii) potentially (but not perfectly) efficient, and (iii) perfectly (or fully) efficient. Specifically,

(i) DMU$_o$ is inefficient when $u_o^* < 1$

(ii) DMU$_o$ is potentially efficient when $u_o^* = 1$, $l_o^* < 1$

(iii) DMU$_o$ is perfectly efficient when $l_o^* = 1$.

Therefore, it is evident that any DMU in the perfectly efficient group is *always efficient* within the given imprecise data, and hence can be regarded as a best performer. Every DMU in the potentially efficient group is *sometimes efficient* and, hence, is a good performer. Least performers are in the inefficient group that consists of *always inefficient* DMUs. It is also clear that DMU_o cannot be efficient in the usual DEA sense unless this DMU is at least potentially efficient and that DMU_o will always be efficient if it is perfectly efficient.

As noted in the introduction, we are here addressing some critical problems underlying the previous IDEA studies, which include the insufficient efficiency classification revealed in [6-11, 16, 19, 26, 27] and the limited attention to imprecise data in [5, 12-15]. In addition, our classification scheme can lead to further prioritizations of efficiency performance. For instance, assume we have two potentially efficient DMUs and their efficiency bounds are respectively given by [0.8, 1] and [0.9, 1]. The former has bigger variation of efficiency than the latter. It follows that the latter DMU is better than the former in terms of sturdiness of efficiency. In fact, the assertion that a perfectly efficient DMU is better than a potentially efficient DMU can also be confirmed in terms of the efficiency sturdiness.

6. Computational aspects

To accomplish our efficiency classifications, we now present how we can obtain the efficiency bound. Based on our duality study in Section 4, we can use the original IDEA model in (1) to obtain the upper bound and the MIDEA-F model in (4) to result in the lower bound. Both models are nonlinear programming problems so we show how to reduce these problems to linear programming equivalents in order to arrive at their exact optimal solutions.

6.1 Calculation of the upper bound

We recall model (1) and specify the D_r^+, D_i^- sets as $D_r^+ = \{\mathbf{y}_r \in \Re^n \mid \mathbf{A}_r \mathbf{y}_r \leq \mathbf{a}_r\}$ for each output r and $D_i^- = \{\mathbf{x}_i \in \Re^n \mid \mathbf{B}_i \mathbf{x}_i \leq \mathbf{b}_i\}$ for each input i. Thus these sets each become convex polyhedrons in $\Re^n$ formed by systems of linear inequalities representing arbitrary linear-type imprecise data. The number of rows of the $\mathbf{A}_r$, $\mathbf{B}_i$ matrices corresponds to the number of constraints on the $\mathbf{y}_r$, $\mathbf{x}_i$ variables, respectively. We can then rewrite model (1) as

$$\begin{aligned}
&\max \sum_{r=1}^{s} \mu_r y_{ro} \\
&\text{s.t. } \sum_{r=1}^{s} \mu_r \mathbf{y}_r - \sum_{i=1}^{m} \omega_i \mathbf{x}_i \leq \mathbf{0} \\
&\qquad \sum_{i=1}^{m} \omega_i x_{io} = 1 \\
&\qquad \mu_r, \omega_i \geq \varepsilon, \quad \forall r, i \\
&\qquad \mathbf{A}_r \mathbf{y}_r \leq \mathbf{a}_r, \quad r = 1, \cdots, s \\
&\qquad \mathbf{B}_i \mathbf{x}_i \leq \mathbf{b}_i, \quad i = 1, \cdots, m
\end{aligned} \tag{11}$$

As shown in [9-11, 19, 26], using the transformation technique

$$\begin{aligned}
\mathbf{Y}_r &= (Y_{r1}, \cdots, Y_{rn})^{\mathrm{T}} = \mu_r \mathbf{y}_r, \quad \forall r \\
\mathbf{X}_i &= (X_{i1}, \cdots, X_{in})^{\mathrm{T}} = \omega_i \mathbf{x}_i, \quad \forall i
\end{aligned} \tag{12}$$

we can reduce model (11) to the following linear program:

$$\begin{aligned}
&\max \sum_{r=1}^{s} Y_{ro} \\
&\text{s.t. } \sum_{r=1}^{s} \mathbf{Y}_r - \sum_{i=1}^{m} \mathbf{X}_i \leq \mathbf{0} \\
&\qquad \sum_{i=1}^{m} X_{io} = 1 \\
&\qquad \mu_r, \omega_i \geq \varepsilon, \quad \forall r, i \\
&\qquad \mathbf{A}_r \mathbf{Y}_r \leq \mu_r \mathbf{a}_r, \quad \forall r \\
&\qquad \mathbf{B}_i \mathbf{X}_i \leq \omega_i \mathbf{b}_i, \quad \forall i
\end{aligned} \tag{13}$$

with the variables all constrained to be nonnegative.

Even though we achieve a linear program successfully using the existing method, we still have a technical problem. Looking at model (13), we find that the concrete value of ε needs to be specified en route to computing the efficiency score. This implies that the score depends on the epsilon value selected. To address this problem, we now modify (13) to

$$
\begin{array}{llllllll}
\max & \sum_{r=1}^{s}\mathbf{e}_o\mathbf{Y}_r & & & & & & \\
\text{s.t.} & \sum_{r=1}^{s}\mathbf{I}\mathbf{Y}_r & & -\sum_{i=1}^{m}\mathbf{I}\mathbf{X}_i & & \leq & \mathbf{0} & \\
 & & & \sum_{i=1}^{m}\mathbf{e}_o\mathbf{X}_i & & = & 1 & \\
 & \mathbf{A}_r\mathbf{Y}_r & -\mu_r\mathbf{a}_r & & & \leq & \mathbf{0}, & \forall r \\
 & & & \mathbf{B}_i\mathbf{X}_i & -\omega_i\mathbf{b}_i & \leq & \mathbf{0}, & \forall i \\
 & & -\mu_r & & & \leq & -\varepsilon, & \forall r \\
 & & & & -\omega_i & \leq & -\varepsilon, & \forall i
\end{array} \tag{14}
$$

The $\mathbf{e}_o \in \Re^n$ is the row unit vector with one in the oth place. If $\text{DMU}_{o=1}$ is evaluated, then $\mathbf{e}_o = (1, 0, \cdots, 0)$. If $\text{DMU}_{o=2}$ is evaluated, $\mathbf{e}_o = (0, 1, 0, \cdots, 0)$. The $\mathbf{I}$ is the $n \times n$ identity matrix.

Dual to linear program (14) becomes

$$
\begin{array}{llllll}
\min & \theta - \varepsilon(\sum_{r=1}^{s}s_r^{+} + \sum_{i=1}^{m}s_i^{-}) & & & & \\
\text{s.t.} & \lambda\mathbf{I} \;\; +\mathbf{p}_r\mathbf{A}_r & & \geq & \mathbf{e}_o^{\mathrm{T}}, & \forall r \\
 & \qquad\;\; \mathbf{p}_r\mathbf{a}_r & +s_r^{+} & = & 0, & \forall r \\
 & \lambda\mathbf{I} \;\; -\mathbf{q}_i\mathbf{B}_i & & \leq & \theta\mathbf{e}_o^{\mathrm{T}}, & \forall i \\
 & \qquad\;\; \mathbf{q}_i\mathbf{b}_i & +s_i^{-} & = & 0, & \forall i
\end{array} \tag{15}
$$

where all variables, $\lambda = (\lambda_j)$, $\mathbf{p}_r$, $\mathbf{q}_i$, s_r^{+}, s_i^{-} are nonnegative except for θ.

We can now utilize the two-stage DEA method of Arnold *et al.* [1] for solving problem (15). Therefore, without specifying the ε value, we can achieve the efficiency score θ^{*}, slacks, and a set of the λ_j^{*} values, which implies the peer group and returns to scale information. Indeed, the earlier IDEA method does not pursue further the dual to the transformed linear program in (13) so it suffers from ε. Moreover, in (13) a different objective function and a new constraint are present for each DMU to be evaluated. Referring to model (15), we now have the same objective function and the same constraints for all DMUs, but we need to change the data in the right-hand side according to the change of DMU_o, as in the ordinary DEA computation. Therefore, the current method improves the previous generalized method.

6.2 Calculation of the lower bound

Now, we present how to solve MIDEA-F problem (4) for obtaining the lower bound on efficiency. Park [20] has proposed a method to solve this problem, using the concept and techniques of the generalized linear programming developed by Soyster [22, 23], and demonstrated its application to the imprecise data involved in a telecommunication company. We thus provide a brief summary of the method as follows.

The nonlinear programming problem in (4) is reduced to a linear program,

$$\begin{aligned}\max\ l_o &= \sum_{r=1}^{s} \mu_r y_{ro}^* \\ \text{s.t.}\quad & \sum_{r=1}^{s} \mu_r y_{ro}^* - \sum_{i=1}^{m} \omega_i x_{io}^* \le 0 \\ & \sum_{r=1}^{s} \mu_r y_{rj}^* - \sum_{i=1}^{m} \omega_i x_{io}^* \le 0 \quad \forall j \ne o \\ & \sum_{i=1}^{m} \omega_i x_{io}^* = 1 \\ & \mu_r,\ \omega_i \ge \varepsilon \quad \forall r,\ i\end{aligned} \tag{16}$$

where

$$\begin{aligned} y_{ro}^* &= \inf\{y_{ro} \mid \mathbf{y}_r \in D_r^+\}, && \text{for all } r \\ x_{io}^* &= \sup\{x_{io} \mid \mathbf{x}_i \in D_i^-\}, && \text{for all } i \\ y_{rj}^* &= \sup\{y_{rj} \mid \mathbf{y}_r \in D_r^+\}, && \text{for all } j \ne o \text{ and all } r \\ x_{ij}^* &= \inf\{x_{ij} \mid \mathbf{x}_i \in D_i^-\}, && \text{for all } j \ne o \text{ and all } i. \end{aligned}$$

The exact data y_{rj}^*, x_{ij}^* are to be selected from the given imprecise data D_r^+, D_i^-, respectively. To obtain the lower bound on efficiency, the worst possible data are chosen for the DMU_o being evaluated, while the best possible data are selected for the other DMUs.

When all the D_r^+, D_i^- sets are closed and bounded, the sup and inf can be replaced by max and min, respectively, so that no problem exists in obtaining the exact data for use in model (16). If these sets are parallelepipeds in $\Re^n$, then the exact data are easily determined without solving the auxiliary linear programming problems. Trouble may be encountered, however, in the case of unbounded sets of imprecise data (e.g., ordinal data), where the maximum may go to infinity. To overcome this, it is useful to note the idea explored in Cooper *et al.* [11] where it is shown that, by virtue of the *column rule* of linear programming, a linear normalization to the original data in D_r^+ and D_i^- can be used without changing the optimal value of the objective l_o^*. We can therefore utilize this normalization concept in the corresponding D_r^+ or D_i^- so that we can avoid the problem associated with such unbounded sets and can obtain a set of exact data for use with model (16). See Park [20] for this issue in detail.

7. Interpretations

In this section, we provide how to interpret the efficiency results either from model (15) or (16), which includes a redial measure of efficiency (θ^*), slacks, and the returns-to-scale classification. Their interpretations are clear in the ordinary DEA with exact data. For instance, the radial measure refers to the proportion of inefficiency present in all inputs so a percentage reduction is needed for all the inputs to improve efficiency. A positive slack for a particular input stands for a further inefficiency in that input. The returns-to-scale classification has to do with the relative scale sizes of DMUs in terms of input usages and output productions. However, such interpretations may alter when imprecise data are incorporated into DEA, and they depend on the type or nature of imprecise data.

The nature of imprecise data is problem specific and will depend upon particulars of the problem in question such as prior knowledge or experiences with the factors involved. For example, on a qualitative factor such as ease of use, skill level of labors or managers, may take the form of only *ordinal* relations as when an expert may say DMU_1 is best, DMU_2 is second, and so on. These DMUs might alternatively be categorized into several groups with respect to performances that are regarded as good, moderate, or poor. If such ordinal data are imposed in DEA, then one can only regard the values of radial efficiency as relations of "better" and "worse." This implies a preference relation between DMUs in terms of their performances, since the ordinal data can be viewed as a preference relation. The case of positive slacks, however, leads to somewhat more discriminating treatment in that one can treat these values in a cardinal manner as for the factors with exact data, as in the ordinary DEA, but not for the ordinal factor. The returns-to-scale classification has nothing to do with the scale sizes of DMUs since we cannot treat the given ordinal data in a cardinal manner. More strictly, this classification is meaningless when ordinal data are imposed in DEA.

The other forms of imprecise data occur in practice. For example, on a qualitative factor, an expert may say that the performance of DMU_1 is more than two times and less than three times that of DMU_2. Alternatively for the same factor the expert may say that the performance of DMU_2 is between 70% and 80% of the performance level of DMU_1. Like ordinal data, this kind of bounded data comes from subjective judgments and represents preference values rather than objective data. Thus the interpretations similar to those for ordinal data can apply to the situation with these bounded data. Note that, unlike the case of ordinal data, one may treat the redial measure and slacks in a cardinal manner: For example, based on the radial measure, a 10% increase in the given 80% performance level is required and, based on the positive slack, a further increase of 2% in this factor's level needs to be made in order to become 90% in the end. This interpretation is possible but it is not directly operational to improve from 80% to 90%. Namely the problem of how to improve it still remains. It is also hard to relate such percentage data to the scale sizes of DMUs. The only one clear trait underlying these percentage data is more-the-better in terms of output.

Even some quantitative factors may involve imprecise data. For example, the manpower employed in a DMU could fail to be constant and may be so volatile that they can only be said to fluctuate within upper and lower bounds. Although this is an example of bounded data, the nature is quite different from the above bounded data. The current bounded data are based primarily upon objective information so we can apply the interpretations like those in the ordinary DEA to the case of these bounded data.

8. Concluding remarks

We have here developed a new approach for the treatment of imprecise data in DEA and this approach provides a more complete capability to account for IDEA. Through a duality study, we show a complete characterization of the efficiency solutions for all possible IDEA models. As a result, we know that incorporating imprecise data directly into the multiplier DEA model measures potential efficiency (or an upper bound on efficiency), while direct incorporation of the same imprecise data into the envelopment model measures perfect efficiency (or a lower bound). Based on these duality results, we present generalized methods to obtain the upper and lower bounds on efficiency, which accomplishes a more detailed and useful classification of efficiency. An improved method for obtaining the upper bound is developed in the manner that extends the previous generalized method. We also demonstrate how to interpret the resulting efficiency and emphasize that the interpretations depend on the type or nature of imprecise data. Possible areas of application are numerous, including project selection, location and policy analysis, vendor and knowledge worker performance analysis, among others.

We would be remiss if we did not remark that this all adds flexibility to the use of imprecise data. It has been assumed in the body of this paper that imprecise data are represented in linear constraints, as in the earlier works on the use of imprecise data in DEA that are cited in the bibliography. In fact this assumption can be relaxed in most of our developments. For instance, the duality results we achieved can be maintained even in the case of nonlinear imprecise data, such as ellipsoids and hyperspheres with associated centers and radiuses, where we can think of the center as the nominal or average value measured for a factor and the radius as the allowance of measurement errors. The solution method we developed for obtaining perfect efficiency can also be applied to such nonlinear imprecise data, since the generalized linear programming method proposed by Soyster [22] allows us to use convex activity sets.

The models and methods developed here are directed to deterministic uses of DEA. They are not intended to cover stochastic approaches to DEA as in the *chance constrained programming* formulations of Olesen and Petersen [17, 18] or the statistical characterizations provided by Banker [2]. See also Simar and Wilson [21] for *bootstrapping* approaches to such statistical characterizations. One feature underlying imprecise data could involve uncertainty in which case the associated stochastic nature of the data may require recourse to stochastic approaches for DEA. Of particular interest may be the chance of realization or specification of the relationships that need to be identified in order to facilitate the exchange of information between the imprecise DEA and stochastic DEA areas. Thus, further work on this topic is warranted that could also lead to new tools and still further extensions.

Finally, revisiting the work of Cooper *et al.* [9-11] we can identify some paths for further research. Besides imprecise data in DEA they also dealt with *assurance region* (AR) conditions on the multiplier variables, as in Thompson *et al.* [24, 25] and the combined variable-data transformations employed in the *cone-ratio envelopment* of

Charnes *et al.* [3]. Hence they developed one unified approach referred to as AR-IDEA. In contrast, we do not deal with AR bounds in the present paper because we want to focus on analyzing technical efficiency in terms of potential and perfect efficiency, separately. Thus, a series of extensions is warranted in that AR bounds could be incorporated into our developments and our efficiency classifications could then be extended to further prioritize the efficiency status of DMUs.

References

[1] Arnold, V., I. Bardhan, W. W. Cooper, and A. Gallegos, Primal and dual optimality in computer codes using two-stage solution procedures in DEA, In: Aranson, J., Zionts, S. (Eds.), *Operations Research: Methods, Models and Applications* Quorum Books, Westport, CT, (1998), 57-96.

[2] Banker, R. D., Maximum likelihood, consistency and data envelopment analysis: A statistical foundation, *Management Science* 39 (1993), 1265-1273.

[3] Charnes, A., W. W. Cooper, Z. M. Huang, and D. B. Sun, Polyhedral cone-ratio DEA models with an illustrative application to large commercial banks, *Journal of Econometrics* 46 (1990), 73-91.

[4] Charnes, A., W. W. Cooper, and E. Rhodes, Measuring the efficiency of decision making units, *European Journal of Operational Research* 2 (1978), 429-444.

[5] Chen, Y., L. M. Seiford, and J. Zhu, Imprecise data envelopment analysis, Unpublished paper, Worcester Polytechnic Institute, Worcester, MA, 2000.

[6] Cook, W. D., M. Kress, and L. M. Seiford, On the use of ordinal data in data envelopment analysis, *Journal of the Operational Research Society* 44 (1993), 133-140.

[7] Cook, W. D., M. Kress, and L. M. Seiford, Data envelopment analysis in the presence of both quantitative and qualitative factors, *Journal of the Operational Research Society* 47 (1996), 945-953.

[8] Cook, W. D. and J. Zhu, Rank order data in DEA: A general framework, *European Journal of Operational Research* 174 (2005), 1021-1038.

[9] Cooper, W. W., K. S. Park, and G. Yu, IDEA and AR-IDEA: Models for dealing with imprecise data in DEA, *Management Science* 45 (1999), 597-607.

[10] Cooper, W. W., K. S. Park, and G. Yu, An illustrative application of IDEA (imprecise data envelopment analysis) to a Korean mobile telecommunication company, *Operations Research* 49 (2001), 807-820.

[11] Cooper, W. W., K. S. Park, and G. Yu, IDEA (imprecise data envelopment analysis) with CMDs (column maximum decision making units), *Journal of the Operational Research Society* 52 (2001), 176-181.

[12] Despotis, D. K. and Y. G. Smirlis, Data envelopment analysis with imprecise data, *European Journal of Operational Research* 140 (2002), 24-36.

[13] Kao, C., Interval efficiency measures in data envelopment analysis with imprecise data, *European Journal of Operational Research* 174 (2006), 1087-1099.

[14] Kao, C. and S. T. Liu, Fuzzy efficiency measures in data envelopment analysis, *Fuzzy Sets and Systems* 113 (2000), 427-437.

[15] Kao, C. and S. T. Liu, Data envelopment analysis with missing data: An application to university libraries in Taiwan, *Journal of the Operational Research Society* 51 (2000), 897-905.

[16] Kim, S. H., C. K. Park, and K. S. Park, An application of data envelopment analysis in telephone offices evaluation with partial data, *Computers and Operations Research* 26 (1999), 59-72.

[17] Olesen, O. B. and N. C. Petersen, Chance constrained efficiency evaluation, *Management Science* 41 (1995), 442-457.

[18] Olesen, O. B., N. C. Petersen, Probabilistic bounds on the virtual multipliers in data envelopment analysis: Polyhedral cone constraints, *Journal of Productivity Analysis* 12 (1999), 103-134.

[19] Park, K. S., Simplification of the transformations and redundancy of assurance regions in IDEA (imprecise DEA), *Journal of the Operational Research Society* 55 (2004), 1363-1366.

[20] Park, K. S., Efficiency bounds and efficiency classifications in DEA with imprecise data, *Journal of the Operational Research Society* 58 (2007), 533-540.

[21] Simar, L., P. W. Wilson, Statistical inference in nonparametric frontier models: The state of the art, *Journal of Productivity Analysis* 13 (2000), 49-78.

[22] Soyster, A. L., Convex programming with set-inclusive constraints and applications to inexact linear programming, *Operations Research* 21 (1973), 1154-1157.

[23] Soyster, A. L., Inexact linear programming with generalized resource sets, *European Journal of Operational Research* 3 (1979), 316-321.

[24] Thompson, R. G., P. S. Dharmapala, and R. M. Thrall, Linked-cone DEA profit ratios and technical efficiency with application to Illinois coal mines, *International Journal of Production Economics* 39 (1995), 99-115.

[25] Thompson, R. G., L. N. Langemeier, C. T. Lee, E. Lee, and R. M. Thrall, The role of multiplier bounds in efficiency analysis with applications to Kansas farming, *Journal of Econometrics* 46 (1990), 93-108.

[26] Zhu, J., Imprecise data envelopment analysis (IDEA): A review and improvement with an application, *European Journal of Operational Research* 144 (2003), 513-529.

[27] Zhu, J., Imprecise DEA via standard linear DEA models with a revisit to a Korean mobile telecommunication company, *Operations Research* 52 (2004), 323-329.