European Journal of Operational Research 202 (2010) 595–603

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Interfaces with Other Disciplines

Cost efficiency measures in data envelopment analysis with data uncertainty

A. Mostafaee [a,*], F.H. Saljooghi [b]

[a] Department of Mathematics, Islamic Azad University, North Tehran Branch, Tehran, Iran
[b] Department of Mathematics, Sistan and Balouchestan University, Zahedan, Iran

ARTICLE INFO

Article history:
Received 18 June 2008
Accepted 8 June 2009
Available online 14 June 2009

Keywords:
DEA
Cost efficiency
Uncertain data
Range data
Imprecise data
Two-level mathematical programming
Extreme point
DMU

ABSTRACT

This paper extends the classical cost efficiency (CE) models to include data uncertainty. We believe that many research situations are best described by the intermediate case, where some uncertain input and output data are available. In such cases, the classical cost efficiency models cannot be used, because input and output data appear in the form of ranges. When the data are imprecise in the form of ranges, the cost efficiency measure calculated from the data should be uncertain as well. So, in the current paper, we develop a method for the estimation of upper and lower bounds for the cost efficiency measure in situations of uncertain input and output data. Also, we develop the theory of efficiency measurement so as to accommodate incomplete price information by deriving upper and lower bounds for the cost efficiency measure. The practical application of these bounds is illustrated by a numerical example.

1. Introduction

Data envelopment analysis (DEA) is a non-parametric technique for evaluating decision making units (DMUs) based on the production possibility set. Also, traditional DEA is used for measuring the efficiency of a set of DMUs where the input and output data of the DMUs are known exactly [3]. Zhu [16,17], Wang et al. [15], Kao [10], Entani et al. [5], and Despotis and Smirlis [4] developed the theory of efficiency measurement where the data are imprecise. In these references there is no discussion concerning cost efficiency with imprecise data. In fact these references focused on the technical–physical aspects of production for use in situations where unit price and unit cost information are not available, or where their uses are limited because of variability in the prices and costs that might need to be considered. It is worthwhile to note that in many real application of DEA, the cost efficiency analysis is required, when some information on prices and costs are available. Technology and cost are the wheels that drive modern enterprises; some enterprises have advantages in terms of technology and others in cost. Hence, the management is eager to know how and to what extent their resources are being effectively and efficiently utilized, compared to other similar enterprises in the same or a similar field. There are some DEA models that deal with cost efficiency (CE) analysis when the data are known exactly. In fact, cost efficiency evaluates the ability to produce current outputs at minimal cost. See, e.g., [6,8,9,14] for more details concerning cost efficiency analysis with deterministic data. In these references, there is no discussion concerning imprecise data, whereas in most cases in industry it is usually known from experience that inputs and outputs vary over certain ranges in a short period of time. It should be mentioned that the uncertainty concerns the researchers only, who conduct efficiency analysis. Moreover, nothing is known of the distribution of the data owing to insufficient information. The only thing available to the decision maker is the two extreme points of the range. The same is true about input prices: there exist many factors in the market beyond the control of the management, which affect input prices. Factors such as interest rate, inflation, and currently exchange rate impose uncertainty of prices on the decision makers. Also, exact knowledge of prices is difficult and prices may be subject to variations in the short term. Estimation of cost efficiency is one of the vital topics in DEA. Although there are many papers for estimating cost efficiency in DEA models (see, for example, [6,8,9,14]), there are only few papers which concern the estimation of cost efficiency in the presence of imprecise data: Jahanshahloo et al. [7], Kuosmanen and Post [11–13], and Camanho and Dyson [2]. For instance Jahanshahloo et al. [7] provide some models for the treatment of ordinal data in cost efficiency analysis. In [7] the models have multiplier forms with additional weight restrictions. The main idea in constructing these models is based on the weighted enumeration of the number of inputs/outputs of each unit which are categorized on the same scale rate. Kuosmanen and Post [11,12] derived upper and lower bounds for overall cost efficiency assuming incomplete price data in the

* Corresponding author.
E-mail address: mostafaee_m@yahoo.com (A. Mostafaee).

0377-2217/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2009.06.007

form of a convex polyhedral cone. Although they presented and proved a model for determining the lower bound of CE, they did not utilize the model in empirical application of their CE concepts and resorted to the Free Disposable Hull technology. For computing the lower bound, they defined set W^V. It is observed that set W^V may be non-convex, which will make it more complicated to operationalize the model they presented for computing the lower bound of CE. So, in the current paper we modify the models for obtaining the upper and lower bounds of CE, which is interesting from theory and practical point of view. In [13], Kuosmanen and Post assume that the firm analysis does not know the prices until after the production plan is fixed. Specifically, they assume that the price are random variables with domain $D \subseteq R^q_+$ and joint distribution function $F : D \to [0,1]$. They presented an approach based on first order stochastic dominance that dealt with uncertainty related to input–output prices. Anyway, we can combine the method proposed in the current paper and the method provided in [13], to obtain the economic efficiency measures. Camanho and Dyson [2] discussed the assessment of CE in complex scenarios of price uncertainty. They assumed that input prices appear in the form of ranges. The upper bound of the CE estimate is obtained with the incorporation of weight restrictions in a standard DEA model, the model which they provided (see Model (7) in [2]) is computationally expensive, because the number of constraints (7a) are $2 \times \mathscr{C}^m_2$. Let n be the number of observations. In order to obtain the lower bound of CE for n DMUs, as they mentioned, it is required to solve n^2 linear programming models. Note that the proposed model may be infeasible, and also computationally expensive.

In this paper, we propose a pair of two-level mathematical programming problems to obtain the upper and lower bounds of cost efficiency when some of the input and output data appear in the form of ranges. In turn, the resulting two-level mathematical programming problems are transformed into equivalent linear programs. Also, we consider situations in which the input and output data as well as input prices appear in the form of ranges, and we obtain the lower and upper bound for cost efficiency in these situations. We prove that when the input prices can be represent by a convex set, the upper and lower bounds of CE are obtained in extreme points of the convex set.

The rest of the paper unfolds as follows: in Section 2, some DEA CE models are reviewed and a CE model in the multiplier form is provided, which is necessary in the next sections. In Section 3, for obtaining the lower and upper bounds of CE, a pair of two-level mathematical programming problems are provided. In Section 4, the provided two-level mathematical programming problems are transformed into equivalent linear ones. Section 5, includes the main results. In fact, in this section the theory of CE is generalized to the situations in which the input prices are also imprecise in the form of ranges. In Section 6 we compare our work with other existing works. Section 7 contains an illustrative numerical example, and Section 8 gives some conclusions.

2. Preliminaries

Assume that we deal with a set of DMUs consisting of $\text{DMU}_j; j = 1, \ldots, n$, with input–output vectors $(x_j, y_j); j = 1, \ldots, n$, in which $x_j = (x_{1j}, \ldots, x_{mj})^T$ and $y_j = (y_{1j}, \ldots, y_{sj})^T$. Define $X = [x_1, x_2, \ldots, x_n]$ and $Y = [y_1, y_2, \ldots, y_n]$ as $m \times n$ and $s \times n$ matrices of inputs and outputs, respectively. Without the loss of generality, we assume that the input and output data x_{ij} and $y_{rj} (i = 1, \ldots, m; r = 1, \ldots, s; j = 1, \ldots, n)$ cannot be exactly obtained due to the existence of uncertainty. They are only known to lie within the upper and lower bounds represented by the ranges $\left[x^L_{ij}, x^U_{ij}\right]$ and $\left[y^L_{rj}, y^U_{rj}\right]$, where $x^L_{ij} > 0$ and $y^L_{rj} > 0$.

In order to obtain a measure of cost efficiency, when the input and output data are known exactly, Färe et al. [6] provide the following LP model:

$$\theta_o = \min\left\{\frac{w_o x}{w_o x_o} : X\lambda = x, Y\lambda \geqslant y_o, \lambda \geqslant 0\right\}. \tag{1}$$

In the above model $w_o \in \mathbb{R}^m_+$ is a user-specified row vector of the prices of the inputs of DMU_o, the unit under assessment. The variables of Model (1) are x and λ. $(\lambda_o = 1, \lambda_j = 0; j \neq o, x = x_o)$ is a feasible solution to (1) which implies that this model is feasible and bounded, and $\theta_o \in (0,1]$. Note that Model (1) has $m + s$ constraints where the RHS values of m constraints are zero, and this can lead to strong degeneracy and hence to great complexity. Regarding part (iii) of the main theorem in [8] one can use the following model instead of model (1), to determine the cost efficiency

$$\vartheta_o = \min\left\{\frac{w_o X\lambda}{w_o x_o} : Y\lambda \geqslant y_o, \lambda \geqslant 0\right\}. \tag{2}$$

Denoting the dual variables associated with the constraint sets as u, the measure of cost efficiency can be obtained by solving the dual of Model (2) as follows (see [9]):

$$\max\left\{u^T y_o : u^T Y \leqslant \frac{w_o X}{w_o x_o}, u \geqslant 0\right\}. \tag{3}$$

In Model (3) the variable is u vector. Regarding the constraints, the optimal objective value of this model is not greater than one.

3. Uncertain cost efficiency models based on uncertain data

When the input and output data are in the form of ranges, the cost efficiency measure calculated from the data should be uncertain as well. In order to deal with such uncertain situations, the following two-level mathematical program is proposed to generate the upper bound of cost efficiency range, for each DMU_o:

$$CE^U_o = \max_{\left\{\begin{array}{l} x^L_j \leqslant x_j \leqslant x^U_j, \\ y^L_j \leqslant y_j \leqslant y^U_j, \\ j = 1, \ldots, n \end{array}\right\}} \min\left\{\frac{w_o X\lambda}{w_o x_o} : Y\lambda \geqslant y_o, \lambda \geqslant 0\right\}. \tag{4}$$

The inner program, i.e., the second-level program, calculates the cost efficiency measure for each set of (x_j, y_j) defined by the outer program, i.e., the first-level program, while the outer program determines the set of (x_j, y_j) that produces the highest cost efficiency measure. The objective value of Model (4) is the upper bound of the cost efficiency measure for DMU_o. The inner program in (4) is a linear program and the dual associated with it is (3). Substituting (3) in (4) gives

$$CE^U_o = \max_{\left\{\begin{array}{l} x^L_j \leqslant x_j \leqslant x^U_j, \\ y^L_j \leqslant y_j \leqslant y^U_j, \\ j = 1, \ldots, n \end{array}\right\}} \max\left\{u^T y_o : u^T Y \leqslant \frac{w_o X}{w_o x_o}, u \geqslant 0\right\}. \tag{5}$$

In the above model, the inner program and the outer program have the same objective function of maximization. So, they can be combined into a one-level program by considering all constraints of the two programs at the same time, as follows:

$$\begin{aligned} CE^U_o = \max \quad & u^T y_o, \\ & u^T y_j \leqslant \frac{w_o x_j}{w_o x_o}, \quad j = 1, \ldots, n, \\ & x^L_j \leqslant x_j \leqslant x^U_j, \quad j = 1, \ldots, n, \\ & y^L_j \leqslant y_j \leqslant y^U_j, \quad j = 1, \ldots, n, \\ & u \geqslant 0. \end{aligned} \tag{6}$$