Applied Mathematics and Computation 270 (2015) 527–533

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

On variable reductions in data envelopment analysis with an illustrative application to a gas company

Mehdi Toloo [a,*], Seddigheh Babaee [b]

[a] Faculty of Economics, Technical University of Ostrava, Ostrava, Czech Republic
[b] Transportation Research Institute, Hasselt University, Diepenbeek, Belgium

ARTICLE INFO

Keywords:
Data envelopment analysis
Stable interval
Variable reduction
Radial and non-radial models
Gas company

ABSTRACT

Data envelopment analysis (DEA) is a non-parametric data oriented method for evaluating relative efficiency of the number of decision making units (DMUs) based on pre-selected inputs and outputs. In some real DEA applications, the large number of inputs and outputs, in comparison with the number of DMUs, is a pitfall that could have major influence on the efficiency scores. Recently, an approach was introduced which aggregates collected inputs and outputs in order to reduce the number of inputs and outputs iteratively. The purpose of this paper is to show that there are three drawbacks in this approach: *instability* due to existence of an infinitesimal epsilon, *iteratively* which can be improved to just one iteration, and providing *non-radial* inputs and outputs and then capturing them. In order to illustrate the applicability of the improved approach, a real data set involving 14 large branches of National Iranian Gas Company (NIGC) is utilized.

1. Introduction

Data envelopment analysis (DEA) is a well-known non-parametric mathematical approach originated by Charnes et al. [5], which is used for evaluating the relative efficiency of a set of similar decision making units (DMUs). In general, DMUs use multiple inputs to produce multiple outputs. In primal DEA model, the CCR model (which was initially proposed by Charnes, Cooper and Rhodes in 1978) relative efficiency score of a DMU is measured as the maximum value of the weighted sum of outputs over the weighted sum of inputs subject to the condition that this ratio must be less than or equal to one for all DMUs. Moreover, the CCR model is provided based on Constant Returns to Scale (CRS) assumption. However, Banker et al. [7] extended it to a model, known as the BCC model, under Variable Returns to Scale (VRS) assumption. Conventional DEA models categorize the set of DMUs into two main groups: efficient and inefficient.

An important pitfall in some DEA applications happens when the number of inputs and outputs is relatively large in comparison with the number of DMUs. In this case, the number of efficient DMUs increases; therefore the results of such an evaluation are not acceptable. Suppose that there are n DMUs with m inputs and s outputs; there is a rule of thumb in DEA literature, the following formula (1), which provides a guidance for determining the number of inputs, outputs, and DMUs (for a deeper discussion of the rule of thumb we refer the readers to [6]).

$$n \geq \max\{3(m+s), m \times s\} \tag{1}$$

As a result, the selection of acceptable number of inputs and outputs is critical for successful DEA applications.

* Corresponding author at: VŠB-TU Ostrava, Faculty of Economics, Sokolská tř. 33, 701 21 Ostrava1, Ostrava, Czech Republic. Tel.: +420 59732 2009.
E-mail address: mehdi.toloo@vsb.cz, m_toloo@yahoo.com (M. Toloo).

http://dx.doi.org/10.1016/j.amc.2015.06.122
0096-3003/

Many studies have been accomplished to deal with increasing the discrimination power of DEA models [10–12]. There are some other techniques toward reducing the dimensionality when there is an excessive number of inputs and outputs in relation to the number of DMUs [1,8,9,16]. They introduced some techniques. Amirteimoori et al. [4] proposed an approach, which aggregates selected inputs and outputs for reducing the number of performance measures. In other words, their approach aggregates inputs or outputs that are highly linearly correlated to increase the discrimination power. Nevertheless, there are three noticeable drawbacks in their approach: First, the proposed models are highly sensitive to the epsilon variations and hence the different values of epsilon lead to the different aggregated inputs or outputs. Second, their models must be run iteratively to obtain an acceptable number of aggregated inputs and outputs. In order to aggregate m_1 inputs and s_1 outputs, the approach must solve $(m_1 - 1) + (s_1 - 1)$ models. Third, although their proposed models non-radially aggregate the inputs and outputs, the radial CCR model is utilized to evaluate the efficiency. As a result, their method fails to measure the correct radial efficiency score thus it might return incorrect results. This paper addresses these issues and proposes a new approach to overcome the weaknesses of their methodology.

The remainder of this paper is organized as follows: Section 2 gives a literature review on the related works. In Section 3, we first analyze the method of Amirteimoori et al. [4], next utilize a real data set to illustrate three drawbacks in this method, and then capture these drawbacks. Conclusions and remarks are provided in the last section.

2. Preliminary

In DEA, each observation, $\text{DMU}_j (j = 1, \dots, n)$ is characterized by a pair of semi-positive input and output vectors, $(\boldsymbol{x}_j, \boldsymbol{y}_j) \in \mathbb{R}_+^{m+s}$. The production technology is determined by Production Possibility Set (PPS), $T = \{(\boldsymbol{x}, \boldsymbol{y}) : \boldsymbol{x} \in \mathbb{R}_+^m \text{ can produce } \boldsymbol{y} \in \mathbb{R}_+^s\}$, which is the intersection of all sets $S \subseteq \mathbb{R}_+^{m+s}$ that satisfy the principal axioms: feasibility, free disposability, convexity, and CRS. Under these axioms, the minimum extrapolation PPS can be explicitly stated as:

$$T = \left\{ (\boldsymbol{x}, \boldsymbol{y}) : \boldsymbol{x} \geq \sum_{j=1}^{n} \lambda_j \boldsymbol{x}_j,\ \boldsymbol{y} \leq \sum_{j=1}^{n} \lambda_j \boldsymbol{y}_j, \boldsymbol{\lambda} \geq 0_n \right\}$$

where $\boldsymbol{\lambda}$ is a semi-positive vector and 0_n is origin in $\mathbb{R}^n$ space. To evaluate the performance of DMU_o, the under evaluation unit, the CCR model is as below:

$$\begin{aligned} \max \ & \sum_{r=1}^{s} u_r y_{ro} + u_o \\ \text{s.t.} \ & \\ & \sum_{i=1}^{m} v_i x_{io} = 1 \\ & \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0 \quad j = 1, \dots, n \\ & v_i \geq \varepsilon \quad i = 1, \dots, m \\ & u_r \geq \varepsilon \quad r = 1, \dots, s \end{aligned} \tag{2}$$

where v_i and u_r are the weights for ith input and rth output respectively, and $\varepsilon > 0$ is a non-Archimedean and sufficiently small number (see Amin and Toloo [3] for the more details). Note that model (2) has degeneracy and having an initial basic feasible solution helps us to solve this model efficiently (for a deeper discussion about finding initial basic feasible solution in DEA model we refer the readers to the body of the paper [13]).

The dual of model (2) is as follows:

$$\begin{aligned} \max \ & \theta - \varepsilon \left(\sum_{i=1}^{m} s_i^x + \sum_{r=1}^{s} s_r^y \right) \\ \text{s.t.} \ & \\ & \sum_{j=1}^{n} \lambda_j x_{ij} + s_i^x = \theta x_{io}, \quad i = 1, \dots, m \\ & \sum_{j=1}^{n} \lambda_j y_{rj} - s_r^y = y_{ro}, \quad r = 1, \dots, s \\ & \lambda_j \geq 0 \ \forall j, \quad s_i^x \geq 0 \ \forall i, \quad s_r^y \geq 0 \ \forall r \end{aligned} \tag{3}$$

The optimal objective value of model (3) (and also model (2)) varies between 0 and 1. DMU_o is efficient if and only if it obtains the efficiency score of one; otherwise it is inefficient. It follows that the efficiency score distinguishes between efficient and inefficient DMUs by establishing whether or not a DMU is located on the efficient frontier (PPS frontier). Moreover, in DEA