Decision Support Systems 54 (2012) 330–335

Contents lists available at SciVerse ScienceDirect

Decision Support Systems

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

Undesirable factors in integer-valued DEA: Evaluating the operational efficiencies of city bus systems considering safety records

Chien-Ming Chen [a], Juan Du [b], Jiazhen Huo [b], Joe Zhu [c,*]

[a] Nanyang Business School, Nanyang Technological University, Singapore 639798, Singapore
[b] School of Economics and Management, Tongji University, 1239 Siping Road, Shanghai 200092, PR China
[c] School of Business, Worcester Polytechnic Institute, Worcester, MA 01609, USA

ARTICLE INFO

Article history:
Received 1 November 2011
Received in revised form 29 April 2012
Accepted 21 May 2012
Available online 28 May 2012

Keywords:
Data envelopment analysis
Efficiency
Integer-valued data
Slacks
Super-efficiency
Undesirable factors

ABSTRACT

In conventional data envelopment analysis (DEA) methods, all inputs and outputs are assumed to be continuous. However, in many practical situations, firms may generate both desirable and undesirable outputs, and some of which may only take integer values (e.g., the number of traffic accidents and deaths in a transportation system). The efficiency evaluation results can be inaccurate if these conditions are not incorporated in the model. In this paper we propose an integer DEA model with undesirable inputs and outputs. The proposed model is developed based on the additive DEA model, in which input and output slacks are used to compute efficiency scores. We also develop an integer super-efficiency model to discriminate the performance of efficient firms. As an illustration, we apply the proposed models to the longitudinal data from a city bus company in Taiwan.

1. Introduction

First introduced by Charnes et al. [5], data envelopment analysis (DEA) is an effective approach to measuring the relative efficiency of peer decision-making units (DMUs) with multiple inputs and outputs. In recent years, DEA has been applied to DMUs in various settings, such as efficiency measurement for network systems [17], and failure mode and effects analysis [10]. Standard DEA models assume that real values for all inputs and outputs, and that all outputs are desirable (i.e., more is always preferred to less). Once the efficient frontier is identified, DEA improves the performance of inefficient DMUs to reach the frontier either by increasing the current output levels or by decreasing the current input levels. However, we can see in many real-world cases where these assumptions are violated. For example, outputs from a health-care service may involve the numbers of post-surgery deaths and medical malpractices, which are undesirable and only take non-negative integer values. In this paper, we develop integer DEA models to handle situations where some of the inputs or outputs are undesirable and integer-valued. We provide an application of the models in evaluating a bus company's efficiency, where the number of accidents is considered an undesirable integer-valued output. Our models can also be used to deal with problems where some inputs are expected to be maximized. For example, Seiford and Zhu [24] study the efficiency of a waste treatment process, in which the amount of waste to be treated (an undesirable/irregular input), is expected to be increased rather than to be decreased.

The integer restriction in DEA has received some but limited attention in literature. Lozano and Villa [21] propose the mixed integer linear programming (MILP) model to restrict the computed targets to integers. Kuosmanen and Kazemi Matin [18] later improve Lozano and Villa's [21] model based upon a new axiomatic foundation for production models involving integers. On the other hand, there are several studies dealing with undesirable inputs/outputs in DEA methods. Färe et al. [16] develop a non-linear DEA program that utilizes Farrell-type of efficiency measure to simultaneously increase desirable outputs and decrease the undesirable outputs by the same scaling factor. Some other researchers such as Scheel [22] and Seiford and Zhu [24], propose different data transformation approaches to transform undesirable outputs/inputs into desirable ones, so that the standard DEA models can be used. However, as discussed in [19], the approaches based upon data transformation may unexpectedly lead to distorted results, in that the ranking and the reference target for a DMU may depend on the transformation method used. As an alternative to the data transformation approach, one may treat the undesirable outputs as inputs [4], or utilize input and output slacks directly in producing an efficiency measure through a slacks-based measure (SBM) approach [27,29]. Chung et al. [11], and Färe and Grosskopf [15] use directional distance functions (DDF) to handle undesirable cases. As indicated in [25], the DDF approach is a special case of the weighted additive models. Liu et al. [20] present a systematic classification of DEA models that consider

* Corresponding author. Tel.: +1 508 831 5467; fax: +1 508 831 5720.
E-mail addresses: cmchen@ntu.edu.sg (C-M. Chen), dujuan@tongji.edu.cn (J. Du), huojiazhen@163.com (J. Huo), jzhu@wpi.edu (J. Zhu).

0167-9236/$ – see front matter
doi:10.1016/j.dss.2012.05.040

undesirable factors, including slacks-based models, radial models, and models with Russell measurement.

In this paper, we use the slacks-based measure (SBM) proposed in [20,27], and formulate additive efficiency and super-efficiency DEA models to deal with integer-valued undesirable data. Our slacks-based measure of super-efficiency analysis is capable of deriving a full ranking of efficient DMUs. Being able to distinguish the performance of efficient DMUs can not only provide decision makers with better insights into the performance of peer DMUs, but also help carry out further analysis for managerial decisions, such as resource-allocation decision (see, e.g., [9]). We illustrate our approach and models through an empirical study of Kaohsiung city bus transit from 1994 to 2009, in which the number of bus accidents represents an integer-valued undesirable output.

The rest of this paper is organized as follows. Section 2 proposes an additive DEA model subject to integer restriction and undesirable factors. Section 3 presents an additive super-efficiency model to further distinguish efficient DMUs. We present an empirical application to Kaohsiung Municipal Bus in Section 4. Section 5 concludes with a summary of our contributions.

2. Additive DEA model

We should begin by noting that one may deal with integer-valued variables by rounding the referent targets obtained from the standard DEA model to the nearest integers. However, it has been proven that this simple rounding approach may result in misleading efficiency evaluations and reference targets, especially for those DMUs with relatively small input and output scales [18].

Next we introduce the notations used in this paper. Suppose that there are n DMUs producing the same set of outputs by consuming the same set of inputs. Unit j is denoted by $DMU_j (j=1, \ldots, n)$. We use X_j and Y_j to denote the input and output vectors of DMU_j. We can categorize input and output variables according to whether a variable is continuous or integer-valued, and whether a variable is a regular (desirable) one or undesirable one. Hence in total we have eight variable sets (i.e., inputs or outputs, continuous or integer-valued, desirable or undesirable). The eight variable sets include desirable real-valued inputs and outputs, desirable integer-valued inputs and outputs, undesirable real-valued inputs and outputs, and undesirable integer-valued inputs and outputs. As now inputs and outputs are categorized into eight different groups according to their attributes, we use m_{GR}, s_{GR}, m_{BR}, s_{BR}, m_{GI}, s_{GI}, m_{BI}, s_{BI} to respectively represent the number of variables in these eight variable sets. Specifically, the subscripts "G" and "B" stand for "good" and "bad" inputs/outputs, respectively; the subscripts "R" and "I" stand for "real-valued" and "integer-valued" variables, respectively. Finally, we use "m" as the index set for input variables, and "s" for output variables.

We assume that all inputs and outputs are non-negative; i.e., for $j=1, \ldots, n$, it holds that $X_j^{GR}=(x_{ij}^{GR})_{m_{GR}\times 1}\geq 0$, $X_j^{GI}=(x_{ij}^{GI})_{m_{GI}\times 1}\geq 0$, $X_j^{BR}=(x_{ij}^{BR})_{m_{BR}\times 1}\geq 0$, $X_j^{BI}=(x_{ij}^{BI})_{m_{BI}\times 1}\geq 0$, $Y_j^{GR}=(y_{rj}^{GR})_{s_{GR}\times 1}\geq 0$, $Y_j^{GI}=(y_{rj}^{GI})_{s_{GI}\times 1}\geq 0$, $Y_j^{BR}=(y_{rj}^{BR})_{s_{BR}\times 1}\geq 0$, and $Y_j^{BI}=(y_{rj}^{BI})_{s_{BI}\times 1}\geq 0$. The corresponding production possibility set (PPS) with variable returns to scale (VRS) is defined as [20]:

$$P=\left\{\begin{array}{l}\left(X^{GR},X^{GI},X^{BR},X^{BI},Y^{GR},Y^{GI},Y^{BR},Y^{BI}\right)\;\middle|\;\begin{pmatrix}X^{GI}, & Y^{GI}\\ X^{BI}, & Y^{BI}\end{pmatrix}\in Z^{m_{GI}+m_{BI}+m_{GI}+m_{BI}};\\ \begin{pmatrix}X^{GR}\\ X^{GI}\end{pmatrix}\geq\sum_{j=1}^{n}\lambda_j\begin{pmatrix}X_j^{GR}\\ X_j^{GI}\end{pmatrix},\\ \begin{pmatrix}X^{BR}\\ X^{BI}\end{pmatrix}\leq\sum_{j=1}^{n}\lambda_j\begin{pmatrix}X_j^{BR}\\ X_j^{BI}\end{pmatrix};\;\begin{pmatrix}Y^{GR}\\ Y^{GI}\end{pmatrix}\leq\sum_{j=1}^{n}\lambda_j\begin{pmatrix}Y_j^{GR}\\ Y_j^{GI}\end{pmatrix},\;\begin{pmatrix}Y^{BR}\\ Y^{BI}\end{pmatrix}\geq\sum_{j=1}^{n}\lambda_j\begin{pmatrix}Y_j^{BR}\\ Y_j^{BI}\end{pmatrix};\\ \sum_{j=1}^{n}\lambda_j=1,\lambda_j\geq 0\end{array}\right\} \quad (1)$$

Note that if we drop the constraint $\sum_{j=1}^{n}\lambda_j=1$ in our models, we have models with constant returns to scale (CRS). Therefore, all subsequent discussions are similarly applicable to the CRS situation. Based upon the above PPS (1), we propose the following definition of an *efficient DMU.*

Definition 1. Efficient DMU

In the presence of integer and undesirable inputs and/or outputs, a $DMU_o(X_o^{GR}, X_o^{GI}, X_o^{BR}, X_o^{BI}, Y_o^{GR}, Y_o^{GI}, Y_o^{BR}, Y_o^{BI})$ is efficient if there does not exist a vector $(X^{GR}, X^{GI}, X^{BR}, X^{BI}, Y^{GR}, Y^{GI}, Y^{BR}, Y^{BI})\in P$, such that $X_o^{GR}\geq X^{GR}$, $X_o^{GI}\geq X^{GI}$, $X_o^{BR}\leq X^{BR}$, $X_o^{BI}\leq X^{BI}$, $Y_o^{GR}\leq Y^{GR}$, $Y_o^{GI}\leq Y^{GI}$, $Y_o^{BR}\geq Y^{BR}$, $Y_o^{BI}\geq Y^{BI}$ with at least one strict inequality.

In order to simultaneously handle the integrality and undesirable variables in one model, we modify the standard additive DEA model [6] based on PPS (1) as follows:

$$\begin{aligned}
&Max\ \frac{1}{m_{GR}+m_{BR}+m_{GI}+m_{BI}+s_{GR}+s_{BR}+s_{GI}+s_{BI}}\\
&\quad\times\left(\begin{array}{l}\sum\frac{s_{io}^{GR-}}{x_{io}^{GR}}+\sum\frac{s_{io}^{BR-}}{x_{io}^{BR}}+\sum\frac{s_{ro}^{GR+}}{y_{ro}^{GR}}+\sum\frac{s_{ro}^{BR+}}{y_{ro}^{BR}}\\ +\sum\frac{s_{io}^{GI-}}{x_{io}^{GI}}+\sum\frac{s_{io}^{BI-}}{x_{io}^{BI}}+\sum\frac{s_{ro}^{GI+}}{y_{ro}^{GI}}+\sum\frac{s_{ro}^{BI+}}{y_{ro}^{BI}}\end{array}\right)\\
&s.t.\ X_o^{GR}-S_o^{GR-}=\sum_{j=1}^{n}\lambda_jX_j^{GR},\quad X_o^{BR}+S_o^{BR-}=\sum_{j=1}^{n}\lambda_jX_j^{BR}\\
&\quad Y_o^{GR}+S_o^{GR+}=\sum_{j=1}^{n}\lambda_jY_j^{GR},\quad Y_o^{BR}-S_o^{BR+}=\sum_{j=1}^{n}\lambda_jY_j^{BR}\\
&\quad X_o^{GI}-S_o^{GI-}\geq\sum_{j=1}^{n}\lambda_jX_j^{GI},\quad X_o^{BI}+S_o^{BI-}\leq\sum_{j=1}^{n}\lambda_jX_j^{BI}\\
&\quad Y_o^{GI}+S_o^{GI+}\leq\sum_{j=1}^{n}\lambda_jY_j^{GI},\quad Y_o^{BI}-S_o^{BI+}\geq\sum_{j=1}^{n}\lambda_jY_j^{BI}\\
&\quad\sum_{j=1}^{n}\lambda_j=1,\ \lambda_j\geq 0,\ j=1,\ldots,n\\
&\quad S_o^{GI-}\in Z_+^{m_{GI}}, S_o^{BI-}\in Z_+^{m_{BI}}, S_o^{GI+}\in Z_+^{s_{GI}}, S_o^{BI+}\in Z_+^{s_{BI}}\\
&\quad S_o^{GR-}, S_o^{BR-}, S_o^{GR+}, S_o^{BR+}\geq 0
\end{aligned} \quad (2)$$

where variable vectors $S_o^{GR-}=(s_{io}^{GR-})_{m_{GR}\times 1}$, $S_o^{BR-}=(s_{io}^{BR-})_{m_{BR}\times 1}$, $S_o^{GI-}=(s_{io}^{GI-})_{m_{GI}\times 1}$, $S_o^{BI-}=(s_{io}^{BI-})_{m_{BI}\times 1}$ and $S_o^{GR+}=(s_{ro}^{GR+})_{s_{GR}\times 1}$, $S_o^{BR+}=(s_{ro}^{BR+})_{s_{BR}\times 1}$, $S_o^{GI+}=(s_{ro}^{GI+})_{s_{GI}\times 1}$, $S_o^{BI+}=(s_{ro}^{BI+})_{s_{BI}\times 1}$ represent the non-radial slack vectors of inputs and outputs for DMU_o. According to *Definition* 1, DMU_o is efficient if and only if the optimal objective for model (2) is zero. Note that model (2) is unit-invariant, which means that its optimal value does not depend on the units of measurement in input and output variables. Compared with classical radial DEA models such as CCR model [5] and BCC model [3], additive models compute efficiency scores based on input and output slacks, which provide a clearer view on which variables cause a specific DMU to be inefficient by a certain amount. With these slack results, directions for improvement are easily obtained for each input and output measure.

In model (2), slack variables S_o^{GR-}, S_o^{BR-}, S_o^{GR+}, S_o^{BR+}, S_o^{GI-}, S_o^{BI-}, S_o^{GI+}, S_o^{BI+} represent the absolute differences between the original input/output values and their respective reference points. It is worth noting that we use inequality in model (2) for integer-restricted inputs and outputs because the convex combinations of frontier DMUs are not necessarily integer-valued. Therefore, $X_o^{GI}-S_o^{GI-}$, $X_o^{BI}+S_o^{BI-}$,