Author's personal copy

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.
© 2012 Elsevier B.V. All rights reserved.

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 © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.dss.2012.05.040

Author's personal copy

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]:

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

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:

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

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, $X_o^{GI}-S_o^{GI-}$, $X_o^{BI}+S_o^{BI-}$,

Author's personal copy

$Y_o^{GI}+S_o^{GI+}$, and $Y_o^{BI}-S_o^{BI+}$, the reference targets for integer factors, may or may not be equal to their projections on the efficient frontier, but must be dominated by their convex combinations of frontier DMUs.

As in the standard additive DEA model, however, efficiency scores obtained from model (2) can take values larger than one. To obtain an efficiency score between zero and one, we use a posteriori efficiency measure based upon the optimal solution to the additive model (2). Specifically, let (λ_j^*; S_o^{GR-*}, S_o^{BR-*}, S_o^{GR+*}, S_o^{BR+*}, S_o^{GI-*}, S_o^{BI-*}, S_o^{GI+*}, S_o^{BI+*}) be an optimal solution to model (2). We use the slacks-based measure (SBM) as the *additive efficiency* for DMU_o, which can then be computed through formula (3) [20,27]:

$$\rho_o^* = \frac{1-\frac{1}{m_{GR}+m_{GI}+s_{BR}+s_{BI}}\left(\sum\frac{s_{io}^{GR-*}}{x_{io}^{GR}}+\sum\frac{s_{io}^{GI-*}}{x_{io}^{GI}}+\sum\frac{s_{ro}^{BR+*}}{y_{ro}^{BR}}+\sum\frac{s_{ro}^{BI+*}}{y_{ro}^{BI}}\right)}{1+\frac{1}{s_{GR}+s_{GI}+m_{BR}+m_{BI}}\left(\sum\frac{s_{ro}^{GR+*}}{y_{ro}^{GR}}+\sum\frac{s_{ro}^{GI+*}}{y_{ro}^{GI}}+\sum\frac{s_{io}^{BR-*}}{x_{io}^{BR}}+\sum\frac{s_{io}^{BI-*}}{x_{io}^{BI}}\right)} \qquad (3)$$

In DEA literature, similar posterior efficiency indices have been developed in an attempt to incorporate slacks after a radial DEA score is obtained. Refer to [8,28] for examples.

Since we get $X_o^{GR}\geq S_o^{GR-*}$, $Y_o^{BR}\geq S_o^{BR+*}$, $X_o^{GI}\geq S_o^{GI-*}$, $Y_o^{BI}\geq S_o^{BI+*}$ according to the constraints from model (2), it is easily proved that $\left(\sum\frac{s_{io}^{GR-*}}{x_{io}^{GR}}+\sum\frac{s_{ro}^{BR+*}}{y_{ro}^{BR}}+\sum\frac{s_{io}^{GI-*}}{x_{io}^{GI}}+\sum\frac{s_{ro}^{BI+*}}{y_{ro}^{BI}}\right)$ is non-negative but no greater than $(m_{GR}+m_{GI}+s_{BR}+s_{BI})$. Thus the numerator of ρ_o^* in Eq. (3) ranges between zero and one, while its denominator is no less than one. Therefore, it is true that the value of ρ_o^* falls between zero and one ($0\leq\rho_o^*\leq 1$), and it is strictly decreasing with respect to input and output slacks. A larger efficiency value indicates that the DMU is closer to the efficient frontier and therefore has a better performance. Following *Definition* 1, DMU_o is efficient if and only if $\rho_o^*=1$, i.e., all the optimal slacks in model (2) are zero.

For an inefficient DMU_o with $\rho_o^*<1$, its performance can be improved to reach, or get as close as possible to, the efficient frontier by removing the excesses in desirable inputs and undesirable outputs, and by augmenting the shortfalls in desirable outputs and undesirable inputs through the following projections: $X_o^{GR*}=X_o^{GR}-S_o^{GR-*}$, $X_o^{BR*}=X_o^{BR}+S_o^{BR-*}$, $X_o^{GI*}=X_o^{GI}-S_o^{GI-*}$, $X_o^{BI*}=X_o^{BI}+S_o^{BI-*}$, $Y_o^{GR*}=Y_o^{GR}+S_o^{GR+*}$, $Y_o^{BR*}=Y_o^{BR}-S_o^{BR+*}$, $Y_o^{GI*}=Y_o^{GI}+S_o^{GI+*}$, $Y_o^{BI*}=Y_o^{BI}-S_o^{BI+*}$.

Note that in formulas (3) and (4), $R_i^{GR-}=\max_j\{x_{ij}^{GR}\}$, $R_i^{GI-}=\max_j\{x_{ij}^{GI}\}$, $R_i^{BR-}=\max_j\{x_{ij}^{BR}\}$, $R_i^{BI-}=\max_j\{x_{ij}^{BI}\}$, $R_r^{GR+}=\max_j\{y_{rj}^{GR}\}$, $R_r^{GI+}=\max_j\{y_{rj}^{GI}\}$, $R_r^{BR+}=\max_j\{y_{rj}^{BR}\}$ and $R_r^{BI+}=\max_j\{y_{rj}^{BI}\}$, can be used instead of $x_{io}^{GR}, x_{io}^{GI}, x_{io}^{BR}, x_{io}^{BI}, y_{ro}^{GR}, y_{ro}^{GI}, y_{ro}^{BR}$ and y_{ro}^{BI}, respectively. In that way, the positivity (non-zero) requirement on the concerning inputs and outputs can be dropped.

3. Additive super-efficiency model

In empirical applications of DEA, one may face the problem of having a high proportion of efficient DMUs in the results (see [1] for a related discussion). This can be fairly problematic as many decision-makings rely on a complete ranking of all DMUs. In this section we develop a super-efficiency model to discriminate DMUs that are efficient in model (2). The idea of super-efficiency was first introduced by Andersen and Petersen [2], which was further extended into the slacks-based measure by Tone [26].

Our super-efficiency measure is developed based upon the slacks-based measure, which considers input/output slacks. Suppose DMU_o is additively efficient. To obtain its super-efficiency score, we cannot simply modify additive model (2) by removing DMU_o from the reference set, as was done in the conventional super-efficiency model in [2]. If we do that, the resulting model may not have a feasible solution [12]. Therefore, for an additively efficient DMU_o, we propose the following unit-invariant super-efficiency model:

$$
\begin{aligned}
Min\ & \frac{1}{m_{GR}+m_{BR}+m_{GI}+m_{BI}+s_{GR}+s_{BR}+s_{GI}+s_{BI}} \\
& \times\left(\begin{array}{l}\sum\frac{t_{io}^{GR-}}{x_{io}^{GR}}+\sum\frac{t_{io}^{BR-}}{x_{io}^{BR}}+\sum\frac{t_{ro}^{GR+}}{y_{ro}^{GR}}+\sum\frac{t_{ro}^{BR+}}{y_{ro}^{BR}}\\ +\sum\frac{t_{io}^{GI-}}{x_{io}^{GI}}+\sum\frac{t_{io}^{BI-}}{x_{io}^{BI}}+\sum\frac{t_{ro}^{GI+}}{y_{ro}^{GI}}+\sum\frac{t_{ro}^{BI+}}{y_{ro}^{BI}}\end{array}\right) \\
s.t.\ & X_o^{GR}+T_o^{GR-}\geq\sum_{j=1,j\neq o}^{n}\lambda_j X_j^{GR}, \quad X_o^{BR}-T_o^{BR-}\leq\sum_{j=1,j\neq o}^{n}\lambda_j X_j^{BR} \\
& Y_o^{GR}-T_o^{GR+}\leq\sum_{j=1,j\neq o}^{n}\lambda_j Y_j^{GR}, \quad Y_o^{BR}+T_o^{BR+}\geq\sum_{j=1,j\neq o}^{n}\lambda_j Y_j^{BR} \\
& X_o^{GI}+T_o^{GI-}\geq\sum_{j=1,j\neq o}^{n}\lambda_j X_j^{GI}, \quad X_o^{BI}-T_o^{BI-}\leq\sum_{j=1,j\neq o}^{n}\lambda_j X_j^{BI} \\
& Y_o^{GI}-T_o^{GI+}\leq\sum_{j=1,j\neq o}^{n}\lambda_j Y_j^{GI}, \quad Y_o^{BI}+T_o^{BI+}\geq\sum_{j=1,j\neq o}^{n}\lambda_j Y_j^{BI} \\
& \sum_{j=1,j\neq o}^{n}\lambda_j=1,\ \lambda_j\geq 0,\ j=1,\ldots,n,j\neq o \\
& T_o^{BR-}\leq X_o^{BR},\ T_o^{GR+}\leq Y_o^{GR},\ T_o^{BI-}\leq X_o^{BI},\ T_o^{GI+}\leq Y_o^{GI} \\
& T_o^{GI-}\in Z_+^{m_{GI}},\ T_o^{BI-}\in Z_+^{m_{BI}},\ T_o^{GI+}\in Z_+^{s_{GI}},\ T_o^{BI+}\in Z_+^{s_{BI}} \\
& T_o^{GR-},\ T_o^{BR-},\ T_o^{GR+},\ T_o^{BR+}\geq 0
\end{aligned} \qquad (4)
$$

Here the variable vectors $T_o^{GR-}=(t_{io}^{GR-})_{m_{GR}\times 1}$, $T_o^{BR-}=(t_{io}^{BR-})_{m_{BR}\times 1}$, $T_o^{GI-}=(t_{io}^{GI-})_{m_{GI}\times 1}$, $T_o^{BI-}=(t_{io}^{BI-})_{m_{BI}\times 1}$ and $T_o^{GR+}=(t_{ro}^{GR+})_{s_{GR}\times 1}$, $T_o^{BR+}=(t_{ro}^{BR+})_{s_{BR}\times 1}$, $T_o^{GI+}=(t_{ro}^{GI+})_{s_{GI}\times 1}$, $T_o^{BI+}=(t_{ro}^{BI+})_{s_{BI}\times 1}$ represent the non-radial slack vectors of inputs and outputs with respect to DMU_o's super-efficiency.

In model (4), we remove DMU_o from the reference set of model (2). In addition, we modify the constraints and objective of model (2). In super-efficiency model (4), the reference set for the DMU under evaluation is constructed by all DMUs other than the evaluated DMU itself. The mathematical reason for choosing to solve a minimization problem instead of a maximization one is that the optimal value to the maximization problem is unbounded. In terms of efficiency measurement, we are trying to determine the minimal absolute changes in inputs and outputs that could render the super-efficient DMU on par with other DMUs on the efficient frontier. Solving a maximization

Table 1
Kaohsiung City Bus data from 1994 to 2009.

Year	Inputs			Outputs		
	Vehicles	Employees	Fuel consumption (liters)	Mileage (kilometers)	Passengers	Accidents (undesirable)
	Integer	Integer	Continuous	Continuous	Integer	Integer
1994	403	1120	5,331,608	18,363,085	31,277,605	154
1995	411	1125	5,883,209	19,695,025	30,840,680	127
1996	420	1112	5,940,485	19,028,804	30,586,165	152
1997	452	1121	6,626,290	20,103,803	33,175,128	140
1998	495	1126	8,154,603	21,189,140	36,092,750	194
1999	473	1146	8,502,509	21,136,856	37,271,623	175
2000	472	1087	9,031,125	21,770,235	38,828,087	161
2001	454	1132	8,788,094	20,801,147	39,107,870	142
2002	432	1108	9,157,097	21,698,043	38,143,325	180
2003	427	1045	8,868,206	21,051,481	33,723,130	171
2004	438	979	8,398,829	20,114,530	32,698,925	159
2005	450	923	9,083,108	22,023,502	32,641,359	232
2006	435	851	7,470,562	17,400,575	29,972,388	150
2007	424	869	7,188,360	17,101,044	28,763,740	147
2008	420	829	6,665,653	16,937,531	25,005,947	118
2009	474	879	6,489,033	17,681,062	22,596,922	123
Avg.	442.5	1028.25	7,598,673	19,755,992	32,545,353	157.81

Author's personal copy

Table 2
Efficiency and slack results for Kaohsiung City Bus from 1994 to 2009.

Year	Additive efficiency	Inputs			Outputs		
		Vehicles	Employees	Fuel (liters)	Mileage (km)	Passengers	Accidents (undesirable)
		s_1^{GI-*}	s_2^{GI-*}	s_1^{GR-*}	s_1^{GR+*}	s_1^{GI+*}	s_1^{BI+*}
1994	1	0	0	0	0	0	0
1995	1	0	0	0	0	0	0
1996	0.9397	8	0	31,134.336	526,403.019	6055	25
1997	1	0	0	0	0	0	0
1998	0.9132	40	28	5709.377	0	482,350	42
1999	0.9415	17	41	163,473.247	0	0	25
2000	1	0	0	0	0	0	0
2001	1	0	0	0	0	0	0
2002	1	0	0	0	0	0	0
2003	1	0	0	0	0	0	0
2004	1	0	0	0	0	0	0
2005	1	0	0	0	0	0	0
2006	1	0	0	0	0	0	0
2007	1	0	0	0	0	0	0
2008	1	0	0	0	0	0	0
2009	1	0	0	0	0	0	0

problem in this case obviously makes no sense, as we would project the super-efficient DMU to an infinitely inefficient point, which corresponds to an unbounded optimal value to the optimization problem. Different from model (2), in model (4) the desirable inputs and undesirable outputs are increased, while the desirable outputs and undesirable inputs are reduced, so that the evaluated DMU can be projected optimally close to the frontier constructed by the remaining DMUs. A set of constraints are imposed on slack vectors to ensure that they are not only non-negative themselves, but also make the reference targets after projection non-negative values.

Several studies [7,23] have pointed out that, standard radial super-efficiency models with VRS will become infeasible, if all possible convex combinations of any output r ($r=1, \ldots, s$) of the remaining DMUs are strictly smaller than the rth output of the evaluated DMU. However, our newly-developed additive super-efficiency model (4) overcomes this infeasibility issue, by using slacks to scale up the inputs (or undesirable outputs) and scale down the outputs (or undesirable inputs) of the DMU under evaluation. See also [13] for a relevant discussion on the feasibility issue.

Besides further differentiating efficient DMUs, additive super-efficiency models help in determining the maximum allowable increase in each desirable input or undesirable output, as well as the maximum allowable decrease in each desirable output or undesirable input, given that the efficient status of an efficient DMU remains unchanged. Specifically, denote ($\lambda_j^*, j \neq o$; T_{io}^{GI-*}, T_{io}^{BI-*}, T_{ro}^{GI+*}, T_{ro}^{BI+*}, T_{io}^{GR-*}, T_{io}^{BR-*}, T_{ro}^{GR+*}, T_{ro}^{BR+*}) as an optimal solution to model (4).

Following the definition proposed in [12], we define

$$\delta_o^* = \frac{\frac{1}{m_{GR}+m_{GI}+s_{BR}+s_{BI}}\left(\sum \frac{x_{io}^{GR}+t_{io}^{GR-*}}{x_{io}^{GR}} + \sum \frac{x_{io}^{GI}+t_{io}^{GI-*}}{x_{io}^{GI}} + \sum \frac{y_{ro}^{BR}+t_{ro}^{BR+*}}{y_{ro}^{BR}} + \sum \frac{y_{ro}^{BI}+t_{ro}^{BI+*}}{y_{ro}^{BI}}\right)}{\frac{1}{s_{GR}+s_{GI}+m_{BR}+m_{BI}}\left(\sum \frac{y_{ro}^{GR}-t_{ro}^{GR+*}}{y_{ro}^{GR}} + \sum \frac{y_{ro}^{GI}-t_{ro}^{GI+*}}{y_{ro}^{GI}} + \sum \frac{x_{io}^{BR}-t_{io}^{BR-*}}{x_{io}^{BR}} + \sum \frac{x_{io}^{BI}-t_{io}^{BI-*}}{x_{io}^{BI}}\right)} \tag{5}$$

as the *additive super-efficiency* for DMU_o. It is obvious that δ_o^* values no less than one ($\delta_o^* \geq 1$), and increases monotonically in both input and output slacks, thus a greater score represents a superior performance compared with other efficient DMUs.

4. Application to Kaohsiung Municipal Bus Company

In this section we apply the proposed additive integer DEA models to evaluate efficiency for the city bus company of Kaohsiung city in Taiwan from 1994 to 2009. Our data is compiled from the on-line version of Kaohsiung City Transportation Annual Report.[1] In this application we consider three input measures: the number of vehicles in operation, staff members, and the fuel consumption in liters. Among them, the number of operating vehicles and staff members are both restricted to non-negative integers. For outputs, we consider the mileage in kilometers, the number of passengers (integer), and finally the number of accidents, which we treat as an undesirable integer output. The data are tabulated in Table 1.

We should note that operating costs are an important input factor to consider, but we did not include the cost variable in our application for the following reasons. The operating costs for a bus company mainly consist of costs of labor, vehicle maintenance, fuel consumption, which have been captured by our current input variables. Other expenses, such as office rental costs and general overhead, can be considered as indirect fixed costs and are not directly related to the provision of bus services. On the side of outputs, the variables "mileage" and "the number of passengers" represent distinct constructs. "Mileage" from a social point of view represents the pervasiveness and availability of the bus services and is expected to be maximized. "The number of passengers", on the other hand, represents the effectiveness of the bus services provided. For example, a bus can travel extensively but only transports a few passengers due to an inappropriate design of the bus route and timing.

In this application, we view each observation from 1994 to 2009 as an independent decision-making unit (DMU). We first apply model (2) to the data set to obtain the additive efficiency score for each year. Table 2 presents the resultant efficiency scores and slacks.

Table 2 shows that only three (Year 1996, 1998 and 1999) out of the total sixteen DMUs are inefficient (with an efficiency score 0.9397, 0.9132 and 0.9415, respectively), whereas all the other thirteen DMUs are additively efficient. Therefore for this application, the results from the additive efficiency model (2) are not too helpful in that thirteen out of sixteen DMUs are considered commensurate in their performance. The reason that a high proportion of DMUs are efficient is probably due to the fact that the number of DMUs is not large compared with the total number of input and output variables [14] Therefore, we turn to the super-efficiency model (4) to further discriminate the performance of the thirteen efficient DMUs.

Table 3 reports the super-efficiency scores, the optimal slacks, and the resulting rankings for all DMUs. With the additive super-efficiency model, we are able to distinguish the performance of all thirteen

[1] Source: http://www.tbkc.gov.tw (in Chinese).

Author's personal copy

Table 3
Additive super-efficiency results for Kaohsiung City Bus from 1994 to 2009.

Year	Super-efficiency	Rank	Inputs			Outputs		
			Vehicles	Employees	Fuel (liters)	Mileage (km)	Passengers	Accidents (undesirable)
			t_1^{GI-*}	t_2^{GI-*}	t_1^{GR-*}	t_1^{GR+*}	t_1^{GI+*}	t_1^{BI+*}
1994	1.0392	5	8	5	551,600.9	0	436,925	0
1995	1.0688	1	8	0	0	1,051,503.198	0	18
1996	1	15	0	0	0	0	0	0
1997	1.0042	11	0	0	0	65,055.486	171,803	0
1998	1	16	0	0	0	0	0	0
1999	1	14	0	0	0	0	0	0
2000	1.0225	7	0	33	0	520,640.33	202,490	0
2001	1.0602	4	0	0	0	7481.684	4,038,563	3
2002	1.0273	6	22	0	0	613,739.967	0	0
2003	1.0085	10	2	0	0	304,497.038	0	0
2004	1.0027	13	0	0	0	108,635.158	0	0
2005	1.0619	3	0	53	0	1,973,630.172	0	0
2006	1.0129	8	0	44	0	0	0	0
2007	1.0029	12	5	0	0	0	0	0
2008	1.0668	2	35	68	0	0	0	12
2009	1.0108	9	0	0	0	378,895.058	0	0

efficient DMUs, and therefore now all sixteen DMUs can be completely ranked according to additive efficiency and super-efficiency scores.

The non-zero optimal slacks in Table 3 indicate the maximum allowable increases for the corresponding desirable inputs/undesirable outputs, as well as the maximum allowable decreases for the corresponding desirable outputs, such that these efficient DMUs can remain additively efficient. For example, the additive super-efficiency of Year 2006 is only associated with its number of employees (the only non-zero optimal slack for Year 2006). This indicates that Year 2006 can add 44 more staff members while still remain additively efficient. However, if the staff size for Year 2006 is increased by 45 or more, then it could become inefficient. In addition, Year 1995, which ranks 1st based on super-efficiency, excels in vehicles, travelling mileage, and accidents. In other words, Year 1995 can simultaneously add 8 more vehicles into operation, reduce the travelling mileage by 1051503.198 kilometers, and allow 18 more accidents, without altering its efficiency status.

It is worth noting that among the thirteen additively efficient DMUs, ten (with non-zero t_1^{GR-*} or t_1^{GR+*}) can afford either to increase the fuel consumption or to reduce the travelling mileage but still remain efficient. This implies that in most years during the 16-year period, Kaohsiung City Bus Company operates in a relatively fuel-efficient way. In addition, we note that Years 1995, 2001 and 2008 (with non-zero t_1^{BI+*}) take the lead in the number of accidents, indicating that in these three years, Kaohsiung City Bus operates in a comparatively safer manner and demonstrates a better accident-control capability.

Moreover, nine of the thirteen additively efficient DMUs (with non-zero t_1^{GI-*} or t_2^{GI-*}) are ahead of other years in utilizing operating vehicles and employees into generating outputs. This result implies that the operating scale for Kaohsiung City Bus in any of these nine years has the capacity to expand without lowering that year's overall performance.

5. Conclusion

In practice, it is not uncommon to see production outputs that are undesirable or can only take integer values. Standard DEA models are not able to handle these situations because the models assume that inputs and outputs are both continuous and desirable. Therefore the reference targets indentified through conventional DEA models are likely to be fractional and hence cannot be implemented in practice. This paper proposes additive efficiency and super-efficiency models to cope with integer constraints and undesirable factors, in order to help distinguish the performance of all DMUs.

We illustrate our models by an empirical efficiency evaluation of Kaohsiung City Bus operation spanning from the years 1994 to 2009. Each year's bus operation is viewed as a DMU, which utilizes inputs (both integer- and real-valued) to produce outputs (including an integer-valued undesirable output). All DMUs can be fully differentiated by additive super-efficiency scores. Also, we described some interesting managerial implications from the results obtained from the proposed models. As a final technical note, in this city bus application the number of passengers is modeled as an output measure, which may not be under the direct discretion of the bus company. In the presence of non-discretionary inputs and outputs, it is possible to develop similar additive efficiency and super-efficiency models as according to models (2) and (4) in this paper, simply by deleting the slack vectors from the corresponding constraints. However, if we take non-controllable/non-discretionary measures into consideration, the related super-efficiency model may become infeasible under the VRS assumption.

We also see a wide variety of potential application areas where models developed in this paper can be used. Take health-care providers for example. The numbers of full-time physicians and nurses are good examples of integer-valued inputs, while total operating expenses can be modeled as a regular input that take real values. For outputs, total operating revenue and the number of total medical cases are examples of real-valued and integer-valued desirable outputs, respectively. The numbers of post-surgery deaths and medical malpractices can be modeled as undesirable integer outputs.

Acknowledgments

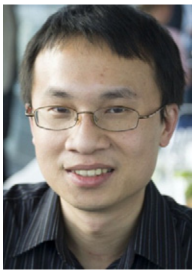
The authors are grateful to the helpful suggestions and comments made by two anonymous reviewers on an earlier version of the paper. Dr. Juan Du would like to thank the support by the National Natural Science Foundation of China (Grant No. 71101108) and China Postdoctoral Science Foundation (Grant No. 20110490696). Professor Huo would like to thank the support by the National Natural Science Foundation of China (Grant No. 70832005).

References

[1] N. Adler, L. Friedman, Z. Sinuany-Stern, Review of ranking methods in the data envelopment analysis context, European Journal of Operational Research 140 (2002) 249–265.
[2] P. Andersen, N.C. Petersen, A procedure for ranking efficient units in data envelopment analysis, Management Science 39 (1993) 1261–1264.
[3] R.D. Banker, A. Charnes, W.W. Cooper, Some models for estimating technical and scale inefficiencies in data envelopment analysis, Management Science 30 (1984) 1078–1092.

Author's personal copy

[4] S.A. Berg, F.R. Forsund, E.S. Jansen, Malmquist indices of productivity growth during the deregulation of Norwegian banking, 1980–89, The Scandinavian Journal of Economics 94 (1992) 211–228.
[5] A. Charnes, W.W. Cooper, E. Rhodes, Measuring the efficiency of decision making units, European Journal of Operational Research 2 (1978) 429–444.
[6] A. Charnes, W.W. Cooper, L.M. Seiford, J. Stutz, A multiplicative model for efficiency analysis, Socio-Economic Planning Sciences 16 (1982) 223–224.
[7] Y. Chen, Measuring super-efficiency in DEA in the presence of infeasibility, European Journal of Operational Research 161 (2005) 545–551.
[8] Y. Chen, H.D. Sherman, The benefits of non-radial vs. radial super-efficiency DEA: an application to burden-sharing amongst NATO member nations, Socio-Economic Planning Sciences 38 (2004) 307–320.
[9] C.-M. Chen, J. Zhu, Efficient resource allocation via efficiency bootstraps: an application to R&D project budgeting, Operations Research 59 (2011) 729–741.
[10] K.S. Chin, Y.M. Wang, G.K.K. Poon, J.S. Yang, Failure model and effects analysis by data envelopment analysis, Decision Support Systems 48 (2009) 246–256.
[11] Y.H. Chung, R. Färe, S. Grosskopf, Productivity and undesirable outputs: a directional distance function approach, Journal of Environmental Management 51 (1997) 229–240.
[12] J. Du, L. Liang, J. Zhu, A slacks-based measure of super-efficiency in data envelopment analysis: a comment, European Journal of Operational Research 204 (2010) 694–697.
[13] J. Du, C.-M. Chen, Y. Chen, W.D. Cook, J. Zhu, Additive super-efficiency in integer-valued data envelopment analysis, European Journal of Operational Research 218 (2012) 186–192.
[14] R.G. Dyson, R. Allen, A.S. Camanho, V.V. Podinovski, C.S. Sarrico, E.A. Shale, Pitfalls and protocols in DEA, European Journal of Operational Research 132 (2001) 245–259.
[15] R. Färe, S. Grosskopf, Modelling undesirable factors in efficiency evaluation: comments, European Journal of Operational Research 157 (2004) 242–245.
[16] R. Färe, S. Grosskopf, C.A.K. Lovell, C. Pasurka, Multilateral productivity comparisons when some outputs are undesirable: a nonparametric approach, The Review of Economics and Statistics 71 (1989) 90–98.
[17] C. Kao, S.-N. Hwang, Efficiency measurement for network systems: IT impact on firm performance, Decision Support Systems 48 (2010) 437–446.
[18] T. Kuosmanen, R. Kazemi Matin, Theory of integer-valued data envelopment analysis, European Journal of Operational Research 192 (2009) 658–667.
[19] W.B. Liu, J. Sharp, DEA Models Via Goal Programming, in: G. Westerman (Ed.), Data envelopment analysis in the public and private sector, Deutscher Universitats-Verlag, 1999.
[20] W.B. Liu, W. Meng, X.X. Li, D.Q. Zhang, DEA models with undesirable inputs and outputs, Annals of Operations Research 173 (2010) 177–194.
[21] S. Lozano, G. Villa, Data envelopment analysis of integer-valued inputs and outputs, Computers and Operations Research 33 (2006) 3004–3014.
[22] H. Scheel, Undesirable outputs in efficiency evaluation, European Journal of Operational Research 132 (2001) 400–410.
[23] L.M. Seiford, J. Zhu, Infeasibility of super-efficiency data envelopment analysis models, INFORS 37 (1999) 174–187.
[24] L.M. Seiford, J. Zhu, Modeling undesirable factors in efficiency evaluation, European Journal of Operational Research 142 (2002) 16–20.
[25] L.M. Seiford, J. Zhu, A response to comments on modeling undesirable factors in efficiency evaluation, European Journal of Operational Research 161 (2005) 579–581.
[26] K. Tone, A slacks-based measure of super-efficiency in data envelopment analysis, European Journal of Operational Research 143 (2002) 32–41.
[27] K. Tone, Dealing with undesirable outputs in DEA: a slacks-based measure (SBM) approach, Nippon Opereshonzu, Risachi Gakkai Shunki Kenkyu Happyokai Abusutorakutoshu (2004) 44–45.
[28] A.M. Torgersen, F.R. Forsund, S.A.C. Kittelsen, Slack-adjusted efficiency measures and ranking of effcient units, Journal of Productivity Analysis 7 (1996) 379–398.
[29] P. Zhou, B.W. Ang, K.L. Poh, Slacks-based efficiency measures for modeling environmental performance, Ecological Economics 60 (2006) 111–118.



Chien-Ming Chen is an assistant professor of operations management at the Nanyang Business School (NBS) of Nanyang Technological University in Singapore. Prior to his position at NBS, he was a postdoctoral scholar and lecturer at the UCLA Institute of the Environment and Sustainability. His main research interests include sustainability issues in operations and management, production economics, decision theories, and empirical research in operations management. His research work has been published in leading journals in operations management, such as Operations Research, Production and Operations Management, and European Journal of Operational Research.

Juan Du is a member at School of Economics and Management, Tongji University, Shanghai, China. She received her Ph.D. in Management Science and Engineering from University of Science and Technology of China (USTC) in late 2010. Her research interests focus on data envelopment analysis (DEA), decision analysis, multi-criteria decision modeling, and applications in health-care and banking. She has articles published in peer-reviewed journals, such as European Journal of Operational Research, Annals of Operations Research, OMEGA, and others.

Jiazhen Huo is Professor at School of Economics and Management, Tongji University, Shanghai, China. His research interests include logistics and supply chain management, management information system (MIS), enterprise operations management. He has published more than 30 articles in peer-reviewed academic journals, co-edited two books on supply chain management, and been approved six national software copyrights in China. He is Chair Professor of DHL, Germany.

Joe Zhu is Professor of Operations and Industrial Engineering, School of Business at Worcester Polytechnic Institute, Worcester, MA. His research interests include issues of productivity and benchmarking, and applications of Data Envelopment Analysis (DEA). He has published over 85 articles in peer-reviewed journals such as *Management Science, Operations Research, IIE Transactions, Naval Research Logistics, European Journal of Operational Research*, *Journal of Operational Research Society*, *Information Technology and Management Journal, Annals of Operations Research, Computer and Operations Research*, *OMEGA, International Journal of Production Economics, Socio-Economic Planning Sciences*, *Journal of Productivity Analysis*, *INFOR*, *Journal of Portfolio Management*, and others. He has published and co-edited seven books on performance evaluation and benchmarking using DEA. He is an Area Editor of OMEGA and an Associate Editor of INFOR. He is also a member of Computers & Operations Research Editorial Board. He is recognized as one of the top 10 authors in DEA with respect to h-index and g-index.