Vol. 32, No. 3 (2022) | DOI: 10.37190/ord220306

OPEN ACCESS

Operations Research and Decisions

www.ord.pwr.edu.pl

OPERATIONS RESEARCH AND DECISIONS QUARTERLY

A slacks-based nonlinear DEA model with integer data: an application to departments of the Islamic Azad University, Karaj Branch in Iran

Zahra Moazenzadeh[1] Saber Saati*[1] Reza Farzipoor Saen[2] Reza Kazemi Matin[3] Sevan Sohraee[1]

[1] *Department of Mathematics, North Tehran Branch, Islamic Azad University, Tehran, Iran*
[2] *Department of Operations Management & Business Statistics, College of Economics and Political Science, Sultan Qaboos University, Muscat, Oman*
[3] *Department of Mathematics, Islamic Azad University, Karaj, Iran*
**Corresponding author: s_saatim@iau-tnb.ac.ir*

Abstract

Although Data Envelopment Analysis (DEA) assumes that inputs and outputs take non-negative real values, in some real-world cases, data are integer-valued. In some situations, rounding a fractional value to the closest integer can lead to a misleading evaluation of efficiency and in some cases may lead to an infeasible projection point. To date, various radial and non-radial models have been presented. This paper proposes a slacks-based non-linear model that guarantees an integer-valued reference point for all integer targets. Also, the reference point of each target is feasible under the proposed model. The lack of a need to round answers to the closest whole value is an advantage of this method. In addition, the results of this model are compared with other models. An example is used to clarify the suggested method.

Keywords: *data envelopment analysis, non-radial model, integer-valued data, slack based measure*

1. Introduction

Data Envelopment Analysis (DEA) is a mathematical approach in operations research to assess the performance of a set of homogenous decision making units (DMUs). DEA is a non-parametric method in which only the amounts of inputs consumed and the amounts of outputs produced are used for the relative evaluation of DMUs.

Two basic DEA models are CCR (Charnes et al. [3]) and BCC (Banker et al. [2]), which rely on the Constant Return to Scale (CRS) and Variable Return to Scale (VRS) assumptions regarding the reference technology, respectively. CRS reflects the fact that output should change in proportion to the inputs (e.g doubling all the inputs will double the outputs). VRS reflects the fact that production technology may

Received 31 January 2022, accepted 22 July 2022, published online 22 November 2022
ISSN 2391-6060 (Online)/

exhibit increasing, constant, or decreasing returns to scale. The CCR model is based on the earlier work of Farrell [8].

Radial and non-radial models are two types of models in DEA. The first DEA models were radial and represented by the CCR model. Radial models deal with proportional changes in inputs or outputs. As such, the CCR score reflects the proportional maximum input (output) reduction (expansion) rate that is common to all inputs (outputs). Radial models have two drawbacks. The first is that, in the real world, not all inputs (outputs) are proportionate. Second, slacks are not taken into account in reporting the efficiency score. Radial approaches may mislead a decision maker when we use the efficiency score as the only index for evaluating the performance of DMUs, if non-radial slacks play a vital role in evaluating efficiency.

Non-radial models include the SBM model (slacks-based measure) presented by Tone [19]. The advantages of the non-radial SBM model over radial models are as follows. The non-radial SBM approach discards the assumption of proportionate changes in the data and deals directly with slacks. Furthermore, the SBM model normally computes efficiency measures using a one-stage method. The following conditions should be satisfied by SBM methods.

1. The measure should be monotonically decreasing in each slack in the data.
2. The measure should be invariant to the units used to measure inputs and outputs.

Tone's SBM model [19] assumes that inputs and outputs have real values, which is not the case in many real situations. It is common in DEA for some inputs and/or outputs to be integer. Therefore, Lozano and Villa [16] proposed a Mixed-Integer Linear Programming (MILP) model for DEA, which is radial. Lozano and Villa's model [16] defines an integer-valued measure. This model had two drawbacks. First, the theoretical foundation of Lozano and Villa's model [16] is ambiguous. Assuming integer-valued inputs and outputs immediately violates the standard convexity, free disposability, and returns to scale properties of DEA. Thus, Lozano and Villa's model is not consistent with the minimum extrapolation principle (Banker et al. [2]), which is the foundation of all DEA models. Second, Lozano and Villa's [16] MILP formulation for computing efficiency scores can lead to the overestimation of efficiency.

Later, Kuosmanen and Matin [15] revised the Farrell input efficiency measure based on integer-valued measures and introduced a MILP problem. Du et al. [7] introduced an additive integer-valued non-oriented DEA model that provides an integrated efficiency score between zero and one. Jie et al. [11] improved Kuosmanen and Matin's model [15] and pointed out that the proposed model solves problems appropriately. Based on such radial models, the efficiency scores are computed using a two-stage method, where the radial input efficiency component θ is minimized in the first stage and the non-radial slacks are maximized in the second stage.

As mentioned by Kuosmanen and Matin [15], rounding a fractional measure to the nearest integer can sometimes lead to a misleading evaluation of efficiency and in some cases may lead to a projection point that is not feasible. Instead of rounding a continuous DEA projection, one possibility would be the Free Disposal Hull (FDH) approach, proposed by Tulkens [20].

Traditional DEA models assume that the levels of all inputs and outputs have real values. However, the levels of some inputs and outputs have integer values in real-world cases. For example, when the efficiency of a university is analyzed, inputs such as the number of students and outputs like the number of articles published have integer values. When one uses sequential or explicit data, it is necessary for DEA to take integer data into account. Such cases have been considered in Banker and Morey [1], Kamakura [12], Rousseau and Semple [17]. This paper attempts to provide models for evaluating the relative efficiency of DMUs, where some data are integer-valued.

Various models have been presented in the field of integer-valued DEA: see, for instance, Chen et al. [5], Jahanshahloo and Piri [10], Khezrimotlagh et al. [13], Hussain et al. [9], Miranda et al. [6], Taleb et al. [18]. Kordrostami et al. [14], as well as Chen et al. [4], proposed approaches based on DEA to assess the relative efficiency of DMUs in the presence of both flexible and integer-valued measures.

The radial and non-radial mixed-integer linear programming (MILP) methods proposed by Lozano and Villa [16], Kuosmanen and Kazemi Matin [15], and Du et al. [7] cannot benchmark efficient units and only define efficiency scores with respect to the set of DMUs conisdered. They cannot describe all of the types of inefficiencies that methods such as slacks-based measures (SBM) can detect.

In this paper, we concentrate on extending a new integer SBM (ISBM) DEA procedure which can estimate the efficiency scores of entities and guarantee integer-valued projection points for integer valued variables. This is a valuable contribution of this paper, which means that we do not need to round off an optimal answer to the nearest integer.

In Khezrimotlagh et, al.'s [13] model, the authors suggested an ISBM model without clarification of when and why to use it. The model proposed in this paper has some advantages over Khezrimotlagh et al.'s [13] model. First, integer-valued inputs and/or outputs have integer-valued projection points. More importantly, the model proposed here is compared to other models, and various theoretical results are presented. Our model does not have the limitations of radial models such as CRS, Lozano and Villa [16] and Kuosmanen and Kazemi Matin [15].

The remainder of the paper is organized as follows. Section 2 gives an overview of mixed integer-valued DEA, as well as the integer radial models introduced by Kuosmanen and Matin [15] and Lozano and Villa [16]. A slacks-based measure dealing with hybrid integer inputs and outputs is proposed in Section 3. Section 4 illustrates the applicability and usefulness of the proposed model using a case study. The conclusions of the study are presented in Section 5. The proofs of theorems are provided in Appendix A.

2. Preliminaries

In this section, radial integer-valued DEA approaches are first presented. Then, the radial models suggested by Lozano and Villa [16], Kuosmanen and Matin [15] are described.

Suppose x_{ij}, $i = 1, \ldots, m$, and y_{rj}, $r = 1, \ldots, s$, are the amounts of the ith input and rth output of DMU$_j$, $j = 1, \ldots n$, respectively. Suppose x_j and y_j correspond to the column vector of DMU$_j$ and $x = (\bar{x}_1, \ldots, \bar{x}_n)$ and $y = (\bar{y}_1, \ldots, \bar{y}_n)$ are data matrices. The production technology, mentioned in Lozano and Villa [16], corresponds to CRS and VRS as follows:

$$T_{\text{CRS}} = \{(\bar{x}, \bar{y}) : \exists(\lambda_1, \ldots, \lambda_n) \;\; \lambda_j \geq 0 \;\; \forall_j, \;\; \bar{x} \geq \lambda x, \; \bar{y} \leq \lambda y\}$$

$$T_{\text{VRS}} = \{(\bar{x}, \bar{y}) : \exists(\lambda_1, \ldots, \lambda_n) \;\; \lambda_j \geq 0 \;\; \forall_j, \;\; \sum_j \lambda_j = 1, \; \bar{x} \geq \lambda x, \;\; \bar{y} \leq \lambda y\}$$

If all the inputs and outputs are integers, then according to Lozano and Villa [16]

$$T_{\text{IDEA}} = \{(x, y) \in Z_+^{m+s} : x \geq \sum_{j=1}^{n} \lambda_j X_j, \;\; y \leq \sum_{j=1}^{n} \lambda_j Y_j, \;\; \lambda_j \geq 0, \;\; \forall_j\}$$

If some of the inputs and outputs are integers, suppose $I = \{1, \ldots, m\}$ is the set of inputs and $O = \{1, \ldots, s\}$ is the set of outputs. The sets of integer and real inputs are I^I and I^{NI}, respectively, and the sets of integer and real outputs are denoted by O^I, O^{NI}, respectively. Therefore,

- $I^I \bigcup I^{NI} = I$, $\left|I^I\right| = p \leq m$, where $\left|I^I\right|$ is the cardinality of the set of integer inputs,
- $O^I \bigcup O^{NI} = O$, $\left|O^I\right| = q \leq s$, where $\left|O^I\right|$ is the cardinality of the set of integer outputs.

Each production point is denoted by an (x, y) vector, such that

$$x = \binom{x^I}{x^{NI}}, y = \binom{y^I}{y^{NI}}$$

Based on this compound set, the following production technology set (PTS) is proposed:

$$T_{\text{HIDEA}} = \left\{ \begin{pmatrix} x^I & y^I \\ x^{NI} & y^{NI} \end{pmatrix} : (x^I, y^I) \in Z_+^{p+q}, \right.$$
$$\left. \begin{pmatrix} x^I \\ x^{NI} \end{pmatrix} \geq \sum_{j=1}^{n} \lambda_j \begin{pmatrix} X_j^I \\ X_j^{NI} \end{pmatrix}, \begin{pmatrix} y^I \\ y^{NI} \end{pmatrix} \leq \sum_{j=1}^{n} \lambda_j \begin{pmatrix} Y_j^I \\ Y_j^{NI} \end{pmatrix}, \lambda_j \geq 0, \forall j \right\}$$

Lozano and Villa [16] introduced a radial model of hybrid integer-valued DEA for evaluating the efficiency of DMUs when some of the inputs and outputs can only take integer values. Their model can be presented as (1).

$$\begin{array}{lll} \min & \theta_o - \varepsilon(\sum_i s_i^- + \sum_r s_r^+) & \\ \text{s.t.} & \sum_j \lambda_j x_{ij} = x_i & \forall i \\ & x_i = \theta_o x_{io} - s_i^- & \forall i \\ & \sum_j \lambda_j y_{rj} = y_r & \forall r \\ & y_r = y_{ro} + s_r^+ & \forall r \\ & \lambda_j \geq 0 & \forall j \\ & s_i^-, x_i \geq 0 & \forall i \\ & s_r^+, y_r \geq 0 & \forall r \\ & \theta \text{ free} & \\ & x_i \in Z & \forall i \in I^I \\ & y_r \in Z & \forall r \in O^I \end{array} \tag{1}$$

where ε is a very small positive value and non-Archimedean number and the variables s_i^- and s_r^+ are slacks. This model has $2(m+s)$ constraints, because $\sum_{j=1}^n \lambda_j x_{ij} = x_i, x_i = \theta x_{io} - s_i^-, \forall i$, and $\sum_{j=1}^n \lambda_j y_{rj} = y_r$, $y_r = y_{ro} + s_r^+$, $\forall r$, are constraints. The number of variables is $n + 2(m+s) + 1$, because θ, λ_j, $\forall j$, (s_i^-, x_i), $\forall i$, (s_r^+, y_r), $\forall r$, are variables. This is a radial model. The CRS integer-efficiency score θ_o^* is the optimal solution for DMU$_o$.

The integer input-oriented model concurrently dealing with real and integer inputs and outputs introduced in Kuosmanen and Matin [15] can be presented as follows:

$$\begin{array}{rlll} Eff(x_o, y_o) = & \min & \theta - \varepsilon(\sum_{r=1}^{s} s_r^+ + \sum_{i=1}^{m} s_i^- + \sum_{i=1}^{p} s_i^I) & \\ & \text{s.t.} & y_{ro} + s_r^+ = \sum_{j=1}^{n} \lambda_j y_{rj} & r \in O \\ & & \theta x_{io} - s_i^- = \sum_{j=1}^{n} \lambda_j x_{ij} & i \in I^{NI} \\ & & \tilde{x}_i - s_i^- = \sum_{j=1}^{n} \lambda_j x_{ij} & i \in I^I \\ & & \theta x_{io} - s_i^I = \tilde{x}_i & i \in I^I \\ & & \tilde{x}_i \in Z_+ & i \in I^I \\ & & \lambda_j \geq 0 & \forall j \\ & & s_r^+ \geq 0, s_i^- \geq 0, s_j^I \geq 0 & r \in O, i \in I, j \in I^I \end{array} \tag{2}$$

The integer-valued projection points of the integer-valued inputs are denoted $\tilde{x}$, where $\tilde{x} \in Z_+$.

Model (1) assumes that the levels of all inputs and outputs are integer. In contrast, model (2) can deal with hybrid cases in which the levels of inputs and outputs can be real or integer (see Kuosmanen and

Matin [15]). In model (2), the vector $(\sum \lambda_j x_j^I, \sum \lambda_j y_j^I)$ is dominated by an integer reference point, i.e., $(\tilde{x}, y_o^I) \in Z_+^{p+q}$. However, in model (2), it is necessary that $(\sum \lambda_j x_j^I, \sum \lambda_j y_j^I)$ is integer valued.

3. An integer-valued slacks-based nonlinear model (ISBMNL)

This section considers a proposed integer-valued DEA model and some theoretical results based on Tone's [19] SBM model and Kuosmanen and Matin [15] for evaluating the efficiency of DMUs when the levels of some of their inputs and outputs are integer values.

$$
\begin{aligned}
\min \quad & \rho = \frac{1-(\frac{1}{m})(\sum_{i\in I^{NI}} \frac{s^-_{i_{NI}o}}{x_{i_{NI}o}} + \sum_{i\in I^{I}} \frac{s^-_{i_{I}o}}{x_{i_{I}o}})}{1+(\frac{1}{s})(\sum_{r\in O^{NI}} \frac{s^+_{r_{NI}o}}{y_{r_{NI}o}} + \sum_{r\in O^{I}} \frac{s^+_{r_{I}o}}{y_{r_{I}o}})} \\
\text{s.t.} \quad & x_{i_{NI}O} - s^-_{i_{NI}o} = \sum_{j=1}^{n} x_{i_{NI}j}\lambda_j \qquad i_{NI} \in I^{NI} \\
& x_{i_{I}o} - s^-_{i_{I}o} \geq \sum_{j=1}^{n} x_{i_{I}j}\lambda_j \qquad i_{I} \in I^{I} \\
& y_{r_{NI}o} + s^+_{r_{NI}o} = \sum_{j=1}^{n} y_{r_{NI}j}\lambda_j \qquad r_{NI} \in O^{NI} \\
& y_{r_{I}o} + s^+_{r_{I}o} \leq \sum_{j=1}^{n} y_{r_{I}j}\lambda_j \qquad r_{I} \in O^{I} \\
& \lambda_j \geq 0 \qquad j = 1,\ldots,n \\
& x_{i_{I}o}, y_{r_{I}o} \in Z_+^{m+s}, x_{i_{NI}o}, y_{r_{NI}o} \geq 0 \\
& s^-_{i_{I}o}, s^+_{r_{I}o} \in Z_+^{m+s}, s^-_{i_{NI}o}, s^+_{r_{NI}o} \geq 0 \\
& i_{NI} \in I^{NI}, i_I \in I^I, r_{NI} \in O^{NI}, r_I \in O^I
\end{aligned}
\tag{3}
$$

where s_i^- and s_r^+ represent the slack variables. These slacks are the actual amounts by which the level of inputs should be reduced and level of outputs should be increased to attain the best feasible solution; $x_i \in z^+$, $y_r \in z^+$ are integer-valued projection points for the integer levels of an input and output, respectively.

According to this model, $\lambda_j^*, j = 1,\ldots,n$, $s_{io}^{-*}, \forall i$, $s_{ro}^{+*}, \forall r$, is an optimal solution for DMU$_o$. Compared with model (1), not all the constraints in this model are expressed as equalities. Similar to model (1), there are $2(m+s)$ constraints and $n+2(m+s)+1$ variables. This model is based on slacks and is non-linear. The constraints in this model imply that $x_{io} - s_i^- \geq \sum_j \lambda_j x_{ij} \geq 0, \forall i$, implying that the efficiency score ρ^* lies between zero and one. DMU$_o$ is efficient if and only if $\rho^* = 1$, which indicates that the slack variables are equal to zero. Otherwise, DMU$_o$ is inefficient.

Theorem 1. The efficiency score of each unit according to model (3) is not higher than the relative efficiency of the same DMU according to the model presented by Kuosmanen and Matin [15] (the proof is given in Appendix A).

Theorem 2. The efficiency score of each unit according to model (3) does not exceed the relative efficiency of the same unit according to FDH (the proof is included in Appendix A).

4. Numerical example

Our proposed model is applied to real-world data from 42 university departments of the Islamic Azad University, Karaj Branch (IAUK). The dataset is taken from [15]. For comparison, five alternative models are considered. The results are shown in Table 1.

The input variables are as follows: x_1 is the number of postgraduate students, x_2 and x_3 are the number of bachelor students and master students, respectively. The output variables are the number of graduates y_1, scholarships y_2, research products y_3, and the manager satisfaction level y_4. All of the variables are integer and y_4 is a variable with an ordinal scale. The efficiency scores are presented in the second to seventh column of Table 1. Model (3) is non-radial and based on slacks.

It is noteworthy that based on radial models, θ is minimzed in the first stage and the non-radial slack (s) is maximized in the next stage. The relative efficiency according to model (3) is placed in the ninth column. The radial input efficiency scores according to model (2), model (1), common CRS model, and FDH model are presented in the tenth, eleventh, twelfth, and thirteenth columns of Table 1, respectively. We used the GAMS program to calculate the efficiencies.

The efficiency scores according to the FDH model are higher than according to other models. This is due to the principles of natural divisibility and additivity that increase the power of other models to discriminate. As shown in Table 1, the efficiency scores according to model (3) do not exceed the relative efficiency according to the other models.

Radial models compared in terms of performance in Kuosmanen, Kazemi Matin [15]. The efficiency scores of DMUs according to model (3) are not higher than those according to radial models.

In general, the efficiency scores for each unit according to our nonlinear integer (SBM) DEA model (model (3)) cannot exceed those according to Kuosmanen and Matin's [15] model. The efficiency scores according to model (3) and Kuosmanen and Matin's [15] model are presented in the ninth and tenth columns of Table 1, respectively. The results in these two columns show that the efficiency score of each unit according to model (3) is not higher than the relative efficiency of the same DMU according to the model developed by Kuosmanen and Matin [15].

Table 2 presents the reference points for the inputs and outputs (i.e., x_i^* and y_j^*) according to model (3). The interpretation of model (3) can be illustrated by considering a particular department, say DMU 3. The integer DEA relative efficiency of this unit was 0.266. The weights for DMU 3 are $\lambda_7^* = 0.413$, $\lambda_{14}^* = 0.788$ and $\lambda_{17}^* = 1.693$ (the weights corresponding to other units are equal to zero). Our integer-valued reference input and output targets are $x_2^* = (0, 281, 52)$ and $y_2^* = (326, 11, 0, 10)$ respectively (see Table 2).

The integer-valued input targets (i.e., $\tilde{x}$), as illustrated in Kuosmanen and Matin [15], for 20 DMUs, rounding the DEA CRS benchmark to the nearest integer-valued target or rounding upward, were projected to a different target point. For instance, consider DMU 13, the efficiency score of this DMU according to Kuosmanen and Matin's 6 model was 0.883, obtained using the weights,$\lambda_{16}^* = 0.7376378$, $\lambda_{17}^* = 0.510961$ and $\lambda_{18}^* = 0.3827820$ (the weights corresponding to the other DMUs are equal to zero). So $\sum_{j=1}^{42} \lambda_j x_j = (0.871, 809, 0)$ dominates$\tilde{x} = (0, 872, 0)$. Lozano and Villa's (2006) MILP formulation yields an input target (0,874,0) by rounding to the nearest integer-valued target. Note that our proposed model yields integer targets for integer-valued data.

5. Conclusions

In traditional DEA, all of the data are assumed to be real. However, in many situations, it is necessary to work with integer variables. The reference levels calculated using the classical DEA model are usually fractional values, which may be invalid. Thus, rounding down fractional values of the input targets and/or rounding up the output targets may lead to a production profile that is infeasible. In this paper, a slack-based model (i.e., model (3)) has been proposed that guarantees an integer-valued projection point for integer-valued variables. The proposed model (3) does not require the convex combination $(\sum \lambda_j x_j^I, \sum \lambda_j y_j^I)$ to be integer (see Du et al. [7]).

This model was illustrated using an example. The efficiency of 42 units was evaluated, based on the same data as for the numerical example of Kuosmanen and Matin [15]. Using this example, the results for model (3) were compared with those of Lozano and Villa [16], Kuosmanen and Matin [15], CRS,

Table 1. Data and efficiency scores

DMU	x_1	x_2	x_3	y_1	y_2	y_3	y_4	ISBNLM	Integer DEA	Lozano and Villa	DEA CRS	FDH
1	0	261	0	225	1	1	3	0.309	0.881	0.881	0.880	1.000
2	0	170	56	213	2	0	3	0.408	0.964	0.977	0.956	1.000
3	0	281	70	326	2	0	3	0.266	0.943	0.947	0.940	1.000
4	0	138	33	159	1	0	2	0.316	0.942	0.964	0.941	1.000
5	164	0	0	52	1	0	3	1.000	1.000	1.000	1.000	1.000
6	291	815	0	1014	2	2	2	0.131	0.918	0.918	0.917	1.000
7	0	0	61	50	0	0	4	1.000	1.000	1.000	1.000	1.000
8	113	95	0	73	0	0	2	0.409	0.495	0.513	0.487	1.000
9	0	727	0	675	3	0	3	0.176	0.928	0.931	0.928	1.000
10	0	773	0	697	2	0	3	0.127	0.902	0.904	0.902	1.000
11	0	0	66	46	0	0	3	0.756	0.758	0.758	0.758	0.924
12	346	197	0	132	0	0	1	0.142	0.266	0.274	0.264	0.629
13	0	988	0	812	8	10	2	0.347	0883	0.885	0.882	1.000
14	0	0	34	32	0	0	2	1.000	1.000	1.000	1.000	1.000
15	0	795	0	601	6	2	2	0.255	0.758	0.764	0.758	1.000
16	0	672	0	591	6	12	2	1.000	1.000	1.000	1.000	1.000
17	0	166	0	166	7	0	4	1.000	1.000	1.000	1.000	1.000
18	0	761	0	761	0	3	2	1.000	1.000	1.000	1.000	1.000
19	193	124	0	293	0	0	3	1	1.000	1.000	1.000	1.000
20	484	0	0	361	0	0	1	0.283	0.893	0.998	0.892	1.000
21	0	517	0	434	0	4	2	0.478	0.880	0.880	0.879	1.000
22	0	584	0	492	1	4	2	0.168	0.875	0.875	0.874	1.000
23	0	682	0	565	2	3	2	0.198	0.840	0.841	0.840	0.985
24	0	565	0	423	1	2	2	0.143	0.758	0.758	0.756	1.000
25	0	603	0	433	1	3	2	0.145	0.740	0.740	0.738	0.969
26	0	373	0	332	1	1	1	0.172	0.890	0.895	0.890	1.000
27	0	347	0	328	2	3	3	0.585	0.997	0.997	0.996	1.000
28	0	0	70	51	0	3	4	1.000	1.000	1.000	1.000	1.000
29	0	328	0	170	0	1	3	0.412	0.540	0.543	0.539	0.796
30	0	267	0	123	0	0	3	0.463	0.468	0.498	0.466	0.622
31	262	0	0	219	3	0	3	1.000	1.000	1.000	1.000	1.000
32	0	1023	0	794	2	0	4	0.103	0.776	0.780	0.776	1.000
33	366	995	0	1111	2	2	3	0.096	0.817	0.819	0.816	1.000
34	0	266	15	238	3	4	3	0.706	0.951	0.955	0.949	1.000
35	172	375	0	547	4	3	3	1.000	1.000	1.000	1.000	1.000
36	0	460	0	385	4	8	3	1.000	1.000	1.000	1.000	1.000
37	223	0	535	232	14	6	4	1.000	1.000	1.000	1.000	1.000
38	0	1202	58	1158	12	0	3	0.101	0.923	0.924	0.922	1.000
39	0	1025	61	394	4	1	3	0.082	0.365	0.367	0.364	1.000
40	0	0	69	50	0	2	4	0.971	0.971	0.986	0.971	1.000
41	314	0	0	204	0	0	1	0.380	0.780	0.834	0.777	0.834
42	371	0	0	226	0	0	1	0.320	0.730	0.863	0.729	1.000

and FDH. Unlike model (3), the approaches proposed by Lozano and Villa [16], Kousmanen and Kazemi Matin [15], and CRS are radial.

The results presented in Kousmanen and Matin [15] present the integer-valued input targets for their model (IDEA), Lozano and Villa's [16] integer DEA model, and the conventional DEA CRS model. For 2 departments, rounding the DEA CRS benchmark to the nearest integer results in a different target point from the one derived from our integer DEA model. The result was somewhat better when the DEA CRS benchmarks were rounded upward: 8 departments were projected to a different target point. Rounding the conventional DEA benchmarks can give over-optimistic (or pessimistic) performance goals, especially for small departments.

Table 2. Integer-valued input and output targets

DMU	x_1^*	x_2^*	x_3^*	y_1^*	y_2^*	y_3^*	y_4^*
1	0	258	0	251	9	1	5
2	0	170	48	213	7	0	7
3	0	281	52	326	11	0	10
4	0	138	26	159	5	0	5
5	164	0	0	52	1	0	3
6	253	815	0	1014	33	2	20
7	0	0	61	50	0	0	4
8	0	83	0	83	0	0	2
9	0	712	0	712	30	0	17
10	0	712	0	712	30	0	17
11	0	0	51	48	0	0	3
12	346	28	0	138	0	0	7
13	0	959	0	865	21	10	13
14	0	0	34	32	0	0	2
15	0	790	0	771	29	2	17
16	0	672	0	591	6	12	2
17	0	166	0	166	7	0	4
18	0	761	0	761	0	3	2
19	193	124	0	293	0	0	3
20	479	0	0	361	0	0	6
21	0	500	0	462	0	4	8
22	0	580	0	553	17	4	9
23	0	680	0	657	23	3	13
24	0	563	0	549	20	2	11
25	0	587	0	560	19	3	11
26	0	353	0	346	13	1	7
27	0	347	0	328	7	3	4
28	0	0	70	51	0	3	4
29	0	293	0	255	0	4	3
30	0	249	0	249	0	0	6
31	262	0	0	219	3	0	3
32	0	830	0	830	35	0	20
33	155	995	0	1111	40	2	23
34	0	265	8	238	4	4	3
35	172	375	0	547	4	3	3
36	0	460	0	385	4	8	3
37	223	0	535	232	14	6	4
38	0	1162	0	1162	49	0	28
39	0	920	0	770	8	16	6
40	0	0	67	50	0	2	4
41	298	0	0	204	0	0	4
42	35	0	0	227	0	0	5

The model proposed in this paper differs from Khezrimotlagh et al.'s method [13] and Kordrostami et al.'s model [14]. Khezrimotlagh et al. [13] extended the SBM-based model (ISBM). Since an additive model does not give a relative measure of efficiency for a DMU, the SBM score can be selected to measure the relative efficiency of a unit. Also, Khezrimotlagh et al. [13] state theorems without proving them. The method introduced in this paper can ascribe an efficiency score to a set of DMUs. In addition, some theoretical results have been proved. Moreover, by modifying the proposed approach, the relative efficiency of units can be evaluated in situations where integer-valued data and flexible measures are present. The projection points of real-valued data do not need to lie on the frontier.

Finally, if we add the constraint $\sum_{j=1}^{n} \lambda_j = 1$ to our model, then we have a model with VRS. Thus, all of the discussions are applicable to the VRS case.

References

[1] Banker, R. D., and Morey, R. C. The use of categorical variables in data envelopment analysis. *Management Science* 32, 12 (1986), 1613–1627.

[2] Banker, R. J., Charnes, A., and Cooper, W. W. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science* 30, 9 (1984), 1078–1092.

[3] Charnes, A., Cooper, W. W., and Rhodes, E. Measuring the efficiency of decision making units. *European Journal of Operational Research* 2, 6 (1978), 429–444.

[4] Chen, C., Ren, J., Tang, L., and Liu, H. Additive integer-valued data envelopment analysis with missing data: A multi-criteria evaluation approach. *PloS One* 15, 6 (2020), 1–20.

[5] Chen, C.-M., Du, J., Huo, J., and Zhu, J. Undesirable factors in integer-valued DEA: Evaluating the operational efficiencies of city bus systems considering safety records. *Decision Support Systems* 54, 1 (2012), 330–335.

[6] de Carvalho Miranda, R., Montevechi, J. A. B., da Silva, A. F., and Marins, F. A. S. Increasing the efficiency in integer simulation optimization: Reducing the search space through data envelopment analysis and orthogonal arrays. *European Journal of Operational Research* 262, 2 (2017), 673–681.

[7] Du, J., Chen, C.-M., Chen, Y., Cook, W. D., and Zhu, J. Additive super-efficiency in integer-valued data envelopment analysis. *European Journal of Operational Research* 218, 1 (2012), 186–192.

[8] Farrell, M. J. The measurement of productive efficiency. *Journal of the Royal Statistical Society: Series A* 120, 3 (1957), 253–281.

[9] Hussain, M. T., Ramli, R., and Khalid, R. A hybrid integer data envelopment analysis based on alternative approach of super slack based measure for measuring super efficiency and inefficiency of decision making units. *Far East Journal of Mathematical Sciences (FJMS)* 100, 1 (2016), 147–170.

[10] Jahanshahloo, G. R., and Piri, M. Data envelopment analysis (DEA) with integer and negative inputs and outputs. *Journal of Data Envelopment Analysis and Decision Science* 2013 (2013), 1–15.

[11] Jie, T., Yan, Q., and Xu, W. A technical note on "A note on integer-valued radial model in DEA". *Computers and Industrial Engineering* 87 (2015), 308–310.

[12] Kamakura, W. A. A note on the use of categorical variables in data envelopment analysis. *Management Science* 34, 10 (1988), 1273–1276.

[13] Khezrimotlagh, D., Salleh, S., and Mohsenpour, Z. A new robust mixed integer-valued model in DEA. *Applied Mathematical Modelling* 37, 24 (2013), 9885–9897.

[14] Kordrostami, S., Amirteimoori, A., and Noveiri, M. J. S. Inputs and outputs classification in integer-valued data envelopment analysis. *Measurement* 139 (2019), 317–325.

[15] Kousmanen, T., and Matin, R. K. Theory of integer-valued data envelopment analysis. *European Journal of Operational Research* 192, 2 (2009), 658–667.

[16] Lozano, S., and Villa, G. Data envelopment analysis of integer-valued inputs and outputs. *Computers and Operations Research* 33, 10 (2006), 3004–3014.

[17] Rousseau, J. J., and Semple, J. H. Categorical outputs in data envelopment analysis. *Management Science* 39, 3 (1993), 384–386.

[18] Taleb, M., Khalid, R., and Ramli, R. Estimating the return to scale of an integer-valued data envelopment analysis model: efficiency assessment of a higher education institution. *Arab Journal of Basic and Applied Sciences* 26, 1 (2019), 144–152.

[19] Tone, K. A slack-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research* 130, 3 (2001), 498–509.

[20] Tulkens, H. On FDH efficiency analysis: some methodological issues and applications to retail banking, courts, and urban transit. *Journal of Productivity Analysis* 4 (1993), 183–210.

A. Proofs of theorems

Proof of Theorem 1. It was shown in the proof of Kuosmanen and Matin's [15][Theorem 3] that model (2) is equivalent to the following model:

$$
\begin{array}{lll}
\min & \theta & \\
\text{s.t.} & y_{ro} \le \sum_{j=1}^{n} \lambda_j y_{rj} & r \in O \\
& \theta x_{io} \ge \sum_{j=1}^{n} \lambda_j x_{ij} & i \in I^{NI} \\
& \theta x_{io} \ge \tilde{x}_i \ge \sum_{j=1}^{n} \lambda_j x_{ij} & i \in I^{I} \\
& \tilde{x}_i \in Z_+ & i \in I^{I} \\
& \lambda_j \ge 0 & j \in J
\end{array}
\tag{4}
$$

Model (3) is also equivalent to the following model:

$$
\begin{array}{lll}
\min & \rho = \dfrac{1 - \frac{1}{m}(\sum_{i \in I} \frac{s_i^-}{x_{io}})}{1 + \frac{1}{s}(\sum_{r \in O} \frac{s_r^+}{y_{ro}})} & \\
\text{s.t.} & \sum_j \lambda_j x_{ij} \le x_i & \forall i \\
& x_i = x_{io} - s_i^- & \forall i \\
& \sum_j \lambda_j y_{rj} \ge y_r & \forall r \\
& y_r = y_{ro} + s_r^+ & \forall r \\
& \lambda_j \ge 0 & \forall j \\
& s_i^-, x_i \ge 0 & \forall i \\
& s_r^+, y_r \ge 0 & \forall r \\
& x_i \in Z & \forall i \in I^I \\
& y_r \in Z & \forall r \in O^I
\end{array}
\tag{5}
$$

To prove Theorem 1, it is first shown that the constraints in model (5) and model (4) are equivalent. Then, it is proved that the results according to model (5) are not exceeded by those according to model (4).

To show that the constraints in model (5) and model (4) are equivalent, the constraints in model (4) are shown to follow from the constraints of model (5) and vice versa.

From the relationships $\sum_j \lambda_j y_{rj} \ge y_r$, $y_r = y_{ro} + s_r^+$, $r \in O^{NI}$, $s_r^+ \ge 0$, we can conclude that $\sum_j \lambda_j y_{rj} \ge y_{ro} + s_r^+$. Therefore, we have

$$\sum_j \lambda_j y_{rj} \ge y_{ro} \quad r \in O^{NI} \tag{6}$$

Also, by considering the relationships $\sum_j \lambda_j y_{rj} \ge y_r \quad r \in O^I$, $y_r = y_{ro} + s_r^+$, $y_r \ge 0$, $y_r \in Z$, we have $\sum_j \lambda_j y_{rj} \ge y_{ro} + s_r^+$. Therefore,

$$\sum_j \lambda_j y_{rj} \ge y_{ro} \quad r \in O^I \tag{7}$$

Inequalities (6) and (7) indicate that

$$\sum_j \lambda_j y_{rj} \ge y_{ro} \quad r \in O \tag{8}$$

Consider the following relationships:

$$\sum_j \lambda_j x_{ij} \le x_i \quad i \in I^I \tag{9}$$

$$x_i = x_{io} - s_i^- \tag{10}$$

$$s_i^- \ge 0, \; x_i \in Z_+ \tag{11}$$

From (9), (10), (11), we deduce that $\sum_j \lambda_j x_{ij} \le x_{io} - s_i^- \le \theta x_{io}$ and hence $\theta \ge \frac{x_{io} - s_i^-}{x_{io}}$. Let $\tilde{x}_i = x_{io} - s_i^-$. It follows that

$$\sum_j \lambda_j x_{ij} \le \tilde{x}_i \le \theta x_{io} \tag{12}$$

By considering $\sum_j \lambda_j x_{ij} \le x_i$, $i \in I^{NI}$, $x_i = x_{io} - s_i^-$, $s_i^- \ge 0$, $x_i \ge 0$, we conclude that $\sum_j \lambda_j x_{ij} \le x_{io} - s_i^- \le \theta x_{io}$ and hence $\theta \ge \frac{x_{io} - s_i^-}{x_{io}}$. It follows that

$$\sum_j \lambda_j x_{ij} \le \theta x_{io} \quad i \in I^{NI} \tag{13}$$

We showed above that the constraints in model (4) can be derived from the constraints in model (5). We have $\sum_j \lambda_j x_{ij} \le \theta x_{io}$, $i \in I^{NI}$, $\theta \le 1$, so $\theta x_{io} \le x_{io}$, $i \in I^{NI}$, and therefore

$$\sum_j \lambda_j x_{ij} \le x_{io} \quad i \in I^{NI} \tag{14}$$

There exists an $s_i^{o-} \ge 0$ such that $\sum_j \lambda_j x_{ij} \le x_{io} - s_i^{o-}$, so we can define $x_i = x_{ip} - s_i^{o-}$.

$$\begin{aligned} &\textstyle\sum_j \lambda_j x_{ij} \le x_i \\ &x_i = x_{io} - s_i^{o^-} \\ &x_i \ge 0 \end{aligned} \tag{15}$$

The relationships $\sum_j \lambda_j x_{ij} \le \tilde{x}_i \le \theta x_{io}$, $i \in I^I$, $\tilde{x}_i \in Z_+$, $\theta \le 1$, indicate that $\tilde{x}_i \le x_{ip}$. So, there exists $s_i^{\Delta-} \in Z_+$ such that $\tilde{x}_i = x_{ip} - s_i^{\Delta-}$, $\tilde{x}_i \in Z_+$ and it follows that

$$\begin{aligned} &\textstyle\sum_j \lambda_j x_{ij} \le \tilde{x}_i \\ &\tilde{x}_i = x_{io} - s_i^{\Delta^-} \\ &\tilde{x}_i \in Z_+ \end{aligned} \tag{16}$$

We have $\sum_j \lambda_j y_{rj} \ge y_{ro}$, $r \in O^I$, so there exists $s_r^{o+} \ge 0$ such that $\sum_j \lambda_j y_{rj} \ge y_{ro} + s_r^{o+}$. Now, $y_r = y_{ro} + s_r^{o+}$ and it follows that

$$\begin{aligned} &\textstyle\sum_j \lambda_j y_{rj} \ge y_r \\ &y_r = y_{ro} + s_r^{o^+} \\ &y_r \ge 0 \end{aligned} \tag{17}$$

From $\sum_j \lambda_j y_{rj} \ge y_{ro}$, $r \in O^I$, we conclude that there exists $s_r^{\Delta+} \in Z_+$, such that $\sum_j \lambda_j y_{rj} \ge y_{ro} + s_r^{\Delta+}$. Set $y_r = y_{ro} + s_r^{\Delta+}$. It follows that

$$\begin{aligned} &\textstyle\sum_j \lambda_j y_{rj} \ge y_r \\ &y_r = y_{ro} + s_r^{\Delta^+} \\ &y_r \in Z_+ \end{aligned} \tag{18}$$

The relationships (15), (16), (17), and (18) indicate that we have derived the constraints in model (5) from the constraints in model (4). Thus, the constraints in model (5) and model (4) are equivalent.

The constraints in model (4) can be written in the form of equalities in the form

$$\begin{array}{lll} \min & \theta & \\ \text{s.t.} & \sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = y_{ro} & r \in O \\ & \sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = \theta x_{io} & i \in I^{NI} \\ & \sum_{j=1}^{n} \lambda_j x_{ij} + \bar{s}_i^- = \tilde{x}_i & i \in I^I \\ & \tilde{x}_i + \hat{s}_i^- = \theta x_{io} & i \in I^I \\ & \tilde{x}_i \in Z_+ & i \in I^I \\ & \lambda_j \ge 0 & j \in J \\ & s_r^+ \ge 0 & \forall r \\ & s_i^-, \bar{s}_i^-, \hat{s}_i^- \ge 0 & \forall i \\ & \theta \text{ free} & \end{array} \tag{19}$$

Suppose that $(\theta^*, \lambda^*, \tilde{x}_i^*, s_i^{*-}, s_r^{*+}, \bar{s}_i^{*-}, \hat{s}_i^{*-})$ is an optimal solution of model (19). $\sum_{j=1}^{n} \lambda_j^* x_{ij} = \theta^* x_{io} - s_i^{*-}$, so $\lambda^* X = \theta^* X_o - s^{*-}$ and thus $\lambda^* X + s^{*-} - \theta^* X_o = 0$. By adding X_o to both sides of this equation, we obtain $\lambda^* X + s^{*-} - \theta^* X_o + X_o = X_o$, so $X_o = \lambda^* X + s^{*-} + (1 - \theta^*) X_o$.

On the other hand, we have $\sum \lambda_j^* y_{rj} - s_r^+ = y_{ro}$. So, $\lambda^* Y \geq Y_o$ and, therefore $\lambda^* Y - s^{*+} = Y_o$. Now, by defining $\hat{s}^+ = s^{*+}$, $\hat{s}^- = s^{*-} + (1-\theta^*)X_o$, $\bar{\lambda} = \lambda^*$. A feasible solution of model (5), denoted $(\bar{\lambda}, \hat{s}^-, \hat{s}^+)$, is given by:

$$\rho = \frac{1 - \frac{1}{m}(\sum \frac{\hat{s}_i^-}{x_{io}})}{1 + \frac{1}{s}(\sum \frac{\hat{s}_r^+}{y_{ro}})} = \frac{1 - \frac{1}{m}(\sum_{i=1}^m \frac{s_i^{*-} + (1-\theta^*)x_0}{x_{io}})}{1 + \frac{1}{s}(\sum \frac{s_r^{*+}}{y_{ro}})}$$

So, $\rho = \frac{1 - \frac{1}{m}(\sum \frac{s_i^{*-}}{x_{io}}) - \frac{1}{m}\sum_{i=1}^m (1-\theta^*)}{1 + \frac{1}{s}(\sum \frac{s_r^{*+}}{y_{ro}})} = \frac{1 - \frac{1}{m}(\sum \frac{s_i^{*-}}{x_{io}}) - \frac{1}{m}(m)((1-\theta^*)}{1 + \frac{1}{s}(\sum \frac{s_r^{*+}}{y_{ro}})}$. Therefore, $\rho = \frac{\theta^* - \frac{1}{m}(\sum \frac{s_i^{*-}}{x_{io}})}{1 + \frac{1}{s}(\sum \frac{s_r^{*+}}{y_{ro}})}$. Since the denominator is strictly greater than one, $\rho \leq \theta^*$. By definition, $\rho^* \leq \rho$, and finally $\rho^* \leq \theta^*$. Hence, we have showed that the efficiency scores according to model (5) do not exceed those according to model (4).

It was indicated in the proof of [15][Theorem 3] by Kuosmanen and Matin that the efficiency scores according to model [15] are equivalent to the efficiency scores according to model (4). Therefore, the efficiency scores according to model (5) do not exceed the efficiency scores according to Kuosmanen and Matin's model [15]. □

Proof of Theorem 2. Kuosmanen and Matin [15] showed that the efficiency score of a DMU according to model (2) does not exceed the efficiency score of the corresponding DMU based on FDH. In Theorem 1, we proved that the efficiency score for a unit according to model (3) is not lower than the efficiency score of the corresponding unit according to model (2). Thus, the efficiency score of a DMU according to model (3) does not exceed the efficiency score of the corresponding unit according to FDH. □