ORIGINAL PAPER

Selective proportionality and integer-valued data in DEA: an application to performance evaluation of high schools

Zohreh Moghaddas[1] · Alireza Amirteimoori[2] · Reza Kazemi Matin[3]

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Conventional data envelopment analysis (DEA) models are often extended for constant or variable returns to scale assumptions based on the under-investigated technology. It is assumed that all inputs and outputs are real-valued data. However, in many practical applications, proportionality or convexity axioms require to be modified. This study attempts to further expand upon the hybrid returns to scale DEA models in the presence of integer-valued input and output data. We refine the previous axioms to introduce a new minimal extrapolation technology set. Moreover, we formulate a couple of mixed-integer linear programming models for efficiency evaluation and target setting. An empirical application on 30 high schools in Iran is provided to validate the proposed approach. The data analysis, including efficiency evaluations along with providing benchmark units, is also performed.

✉ Reza Kazemi Matin
rkmatin@kiau.ac.ir

Zohreh Moghaddas
moghaddas.sahar@gmail.com

Alireza Amirteimoori
ateimoori@iaurasht.ac.ir

[1] Department of Mathematics, Qazvin Branch, Islamic Azad University, Qazvin, Iran

[2] Department of Mathematics, Rasht Branch, Islamic Azad University, Rasht, Iran

[3] Department of Mathematics, Karaj Branch, Islamic Azad University, Karaj, Iran

1 Introduction

Data envelopment analysis (DEA) is a non-parametric mathematical programming technique originally introduced by Charnes et al. (1979) and extended by Banker et al. (1984). This data-oriented technique is utilized for performance assessment of a set of homogeneous decision making units (DMUs) which consume multiple inputs to produce multiple outputs. Conventional DEA models are often built on some basic information about returns to scale assumptions such as constant or variable return to scale for the underlying production set (Cooper et al. 2006).

It is realized that input and output data in real applications of performance evaluation of production systems are sometimes integer-valued. In this sense, to conduct in-depth performance analysis, it is important to modify the traditional DEA models to take into discrete types of input/output data. Neglecting the integrality assumption of inputs and/or outputs may lead to the overestimation of results in the efficiency evaluation of production units. At the first sight, it seems that rounding to the nearest integer target is a rational procedure to get an optimal target. However, this method may lead to inefficient or even infeasible target points and misleading efficiency evaluations. Handling integer data in a modified DEA model was first studied in the DEA framework by Lozano and Villa (2006) and then extended and improved by Kuosmanen and Kazemi Matin (2009), Kazemi Matin and Kuosmanen (2009) , Kuosmanen et al. (2015).

As Podinovski (2004) discussed, in many real DEA applications, the full proportionality between all inputs and outputs also cannot be assumed. This means that it is possible to have a subset of inputs that are not proportional to a subset of outputs in the evaluation.

These issues show that there is a need to modify the existing approaches in such a way that they can be applied for the cases that selective proportionality and discrete data coexist. This work extends the hybrid returns to scale (HRS) models in DEA for the case of integer-valued input and output data by taking an axiomatic approach. An associated minimal extrapolation technology set is then introduced as well as some mixed integer linear programming (MILP) models that are suggested for efficiency evaluation in the resulting technology. To demonstrate the applicability and capability of the new approach, two examples are provided.

This paper is organized as follows: A literature review on integer DEA and HRS models is given in Sect. 2. Section 3 briefly presents the concept of selective proportionality in DEA frameworks. The HRS technology is extended in Sect. 4 and a new Farrell-type efficiency measure is presented for the treatment of integer-valued input and output. Section 5 discusses a descriptive application of performance evaluation of 30 high schools in Iran. The paper concludes in Sect. 6.

2 Literature review

Different assumptions about DEA axioms and various types of data have been widely studied in DEA literature for better analyzing the performance of production units. The first two basic DEA models, i.e., CCR (Charnes et al. 1978) and

BCC (Banker et al. 1984) models, have been developed based on constant returns to scale (CRS) and variable returns to scale (VRS) technologies, respectively. Thenceforth, a wide range of DEA models has been introduced to deal with the performance evaluation problem with different assumptions. These models possess their advantages and limitations. Podinovski (2004) highlighted that the VRS technology is conservative and may overestimate the true efficiency scores in many cases. On the other hand, the underlying full proportionality assumption in CRS technology does not hold in all cases. Consequently, the author introduced a novel hybrid returns to scale (HRS) technology to make a bridge between constant and variable returns to scale technologies. This type of hybrid scenario is also used in other studies in DEA literature; for example, in efficiency measurement with mixed orientation in Wu and O'Brien (2010), where a mixed radial DEA model is suggested and then it is verified that the obtained results lie within solutions of input and output orientations.

In the HRS models, it is possible to build the technology set based on a selective proportionality axiom. This can be done by a modification on full proportionality axiom in the CRS technology. The HRS production technology is a convex polyhedral set where a subset of inputs and outputs operates under the CRS assumption while other inputs and outputs activate under the VRS assumption. The most important feature of HRS technology is that it simultaneously covers both quantity and quality of data in a unified production set which often appears in technologies with different types of returns to scale (RTS) assumptions. Several extensions and applications of HRS technologies have been reported in the recent DEA literature. Podinovski (2009) developed a production possibility set that exhibits both full and selective proportionality in a unified set. Cook and Zhu (2011) introduced a new multiple variable proportionality (MVP) concept as an extension of HRS technology. The MVP shows how output bundles are separated into distinct subsets with different types of RTS. Kazemi Matin and Emrouznejad (2011) utilized a variant of the HRS assumption and developed a bounded technology set by taking into account integer-valued data. Alirezaee and Boloori (2012) proposed a proportional model of trade-offs in the DEA framework by modifying the proportionality axiom and by considering some input–output replacement. Huang et al. (2012) considered a hybrid DEA model and evaluated proportionate and non-proportionate inputs with radial and non-radial measures in an application to Taiwanese international tourist hotels. The authors utilized the HRS to evaluate the impact of marketing expenses on subjects. Podinovski et al. (2014) reported another application of HRS for performance assessment of secondary schools. Recently, Afsharian et al. (2015) proposed a modification of free disposal hull (FDH) technology by incorporating the selective proportionality axiom. The HRS model incorporating production trade-offs was developed by Podinovski et al. (2017) with an application to the efficiency assessment of public universities in Malaysia.

More recently, Ferreira et al. (2018) and Ferreira and Marques (2020) proposed an extension of the traditional order-α method to estimate an empirical convex α-level. They proposed a step forward on the order-α robust nonparametric method for technical efficiency assessment. As Ferreira and Marques (2020) stated, the order-α robust nonparametric method is computationally complex and expensive. In

other words, in the DEA framework with at most 500 DMUs, we prefer to solve a MILP model instead of the order-α method.

To the best of our knowledge, no research has directly considered integer-valued data in HRS models. This paper is the first attempt that takes the assumptions of the integrality of the data and selective proportionality of inputs and outputs into consideration.

3 Selective proportionality in DEA

It is initially assumed that a set of n decision making unit (DMUs) exists where $DMU_j, (j \in \{1, ..., n\})$. consumes m inputs $(x_j = (x_{ij}), i \in \{1, ..., m\})$ to produce s outputs $(y_j = (y_{rj}), r \in \{1, ..., s\})$. Basic DEA models are built upon a technology set that satisfies a set of underlying axioms of disposability and returns to scale properties. A common example could be introducing the following basic axioms for the CCR production set T_c:

$\dot{A}$1. *Observations inclusion*: For all j: $(x_j, y_j) \in T_c$.

$\dot{A}$2. *Convexity*: T_c is a convex set.

$\dot{A}$3. *Constant returns to scale*: For all $(x, y) \in T_c$ and $\alpha \in R_+$: $(\alpha x_j, \alpha y_j) \in T_c$.

$\dot{A}$4. *Free disposability*: If $(x, y) \in T_c$, $x \le \bar{x}$ and $0_s \le \bar{y} \le y 0_s \mathbb{R}^s 0_s = (0, ..., 0) \in \mathbb{R}^s$, then $(x, y) \in T_c$.

$\dot{A}$5. *Closedness*: T_c is a closed set.

$\dot{A}$6. *Minimum extrapolation*: T_c is the intersection of all sets that satisfy axioms $\dot{A}1 - \dot{A}5$.

Regarding the axioms $\dot{A}1 - \dot{A}6$, the associated production possibility set (PPS) could be stated as follows:

$$T_c = \left\{ (x, y) \in R_+^{m+s} \middle| x \ge \sum_{j=1}^n \lambda_j x_j, y \le \sum_{j=1}^n \lambda_j y_j, \lambda_j \ge 0 \ \forall j \right\} \tag{1}$$

For deriving the VRS technology set, axiom $\dot{A}$3 from the CRS technology is ignored and the associated minimal extrapolation technology set is constructed by Banker et al. (1984) as follows:

$$T_v = \left\{ (x, y) \in R_+^{m+s} \middle| x \ge \sum_{j=1}^n \lambda_j x_j, y \le \sum_{j=1}^n \lambda_j y_j, \sum_{j=1}^n \lambda_j = 1, \lambda_j \ge 0 \ \forall j \right\} \tag{2}$$

The DEA CRS form of technology assumes full proportionality among all inputs and outputs while VRS refers to no proportionality. In most practical situations, however, only some of the inputs and the outputs may be assumed to be proportional to each other. The remaining variables are not part of the proportion.

Podinovski (2004) addressed similar circumstances and introduced a hybrid model that combines CRS and VRS assumptions. In this model, the CRS assumption is considered for selected sets of inputs and outputs while the VRS assumption holds for the other data. Podinovski (2004) claimed that the discriminating power

of an HRS model is better than either VRS or CRS formulations in most practical situations.

In what follows, we partition the set of inputs and outputs into two groups: those inputs and outputs that are proportional to each other and those that are not. We use the superscript "P" and "NP" to show proportional and non-proportional input–output variables, respectively. To this end, consider the following partitions according to the selective proportionality:

$$I = I^P \cup I^{NP}, O = O^P \cup O^{NP}$$

It is noted that O^P is assumed to be proportional to I^P while I^{NP} and O^{NP} are not. Subsets "P" and "NP" of inputs and outputs are non-empty, mutually exclusive and collectively exhanustive. Thus, each production plan can be stated as below

$$(\boldsymbol{x}, \boldsymbol{y}) = \left(\boldsymbol{x}^P, \boldsymbol{x}^{NP}, \boldsymbol{y}^P, \boldsymbol{y}^{NP}\right)$$

where $\boldsymbol{x}^P \in \mathbb{R}^{|I^P|}$, $\boldsymbol{y}^P \in \mathbb{R}^{|O^P|}$, $\boldsymbol{x}^{NP} \in \mathbb{R}^{m-|I^P|}$, and $\boldsymbol{y}^{NP} \in \mathbb{R}^{s-|O^P|}$.

Podinovski ([2004](#)) defined the HRS technology based on the following axioms:

A1. *Observations inclusion:* For all j: $\left(\boldsymbol{x}_j, \boldsymbol{y}_j\right) \in T_{HRS}$.

A2. *Convexity:* T_{HRS} is a convex set.

A3. *Free disposability:* If $(\boldsymbol{x}, \boldsymbol{y}) \in T_{HRS}$, $\boldsymbol{x} \leq \bar{\boldsymbol{x}}$ and $0_s \leq \bar{\boldsymbol{y}} \leq \boldsymbol{y}$ then $(\boldsymbol{x}, \boldsymbol{y}) \in T_{HRS}$.

A4. *Selective proportionality*: Let $(\boldsymbol{x}, \boldsymbol{y}) \in T_{HRS}$, then $(\alpha\boldsymbol{x}, \alpha\boldsymbol{y})$ is defined in expansion and contraction scenarios as follows:

Expansion scenario
$$\left(\alpha\boldsymbol{x}^P, \boldsymbol{x}^{NP}, \alpha\boldsymbol{y}^P, \boldsymbol{y}^{NP}\right) \in T_{HRS}, \forall \alpha(\alpha > 1) \tag{3}$$

Podinovski ([2004](#)) proved that under the convexity assumption the contraction scenario can be stated as follows:

Contraction scenario
$$\left(\alpha\boldsymbol{x}^P, \boldsymbol{x}^{NP}, \alpha\boldsymbol{y}^P, \alpha\boldsymbol{y}^{NP}\right) \in T_{HRS}, \forall \alpha(0 \leq \alpha < 1) \tag{4}$$

A5. Closedness: T_{HRS} is a closed set.

A6. Minimum extrapolation: T_{HRS} is the minimal set that satisfies axioms $A1 - A5$.

Regarding the axioms $A1 - A6$, it is shown that the associated production possibility set (PPS) can be stated as follows:

$$T_{HRS} = \left\{ \left(\boldsymbol{x}^P, \boldsymbol{x}^{NP}, \boldsymbol{y}^P, \boldsymbol{y}^{NP}\right) \in R_+^{m+s} \middle| \begin{array}{l} \boldsymbol{x}^P \geq \sum_{j=1}^{n} \lambda_j \left(\alpha_j \beta_j \boldsymbol{x}_j^P\right), \boldsymbol{y}^P \leq \sum_{j=1}^{n} \lambda_j \left(\alpha_j \beta_j \boldsymbol{y}_j^P\right) \\ \boldsymbol{x}^{NP} \geq \sum_{j=1}^{n} \lambda_j \boldsymbol{x}_j^{NP}, \boldsymbol{y}^{NP} \leq \sum_{j=1}^{n} \lambda_j \left(\alpha_j \boldsymbol{y}_j^{NP}\right), \\ \sum_{J=1}^{N} \lambda_j = 1, \forall j, \lambda_j \geq 0, \alpha_j \leq 1, \beta_j \geq 1 \end{array} \right\} \tag{5}$$

In the presentation of the production set T_{HRS} in (5), $\alpha_j \leq 1$ and $\beta_j \geq 1$ are respectively expansion and contraction factors.

Different assumptions in DEA axioms have now been considered for constructing a variety of technologies for better formulating and analyzing practical production systems. Podinovski (2009) considered full and selective proportionality simultaneously and applied the proposed models to a practical example. As a result, he noted that there is a notable difference in the discrimination powers of CRS and HRS models.

Based on the above discussions and in order to produce more reliable results in performance assessments of production units, the common DEA models need to be modified in order to address applications where selective proportionality in the presence of integer-valued input and output data. We employ an axiomatic approach aims at addressing the integer-valued data in the HRS models.

4 Selective proportionality and integer-valued data

For simplicity of presentation and without loss of generality, it has been assumed that all input and output data can only take integer values. This assumption would be relaxed later. Consider the following axiom defined in integer environment by Kuosmanen and Kazemi Matin (2009):

Integrality : If $(x, y) \in T$ then $(x, y) \in Z^{m+s}$

Obviously, the integrality assumption is in contradiction with standard DEA assumptions such as free disposability and convexity assumptions. This indicates that traditional axioms need to be modified to meet new conditions. To do this, they introduced the following axioms for contraction and expansion scenarios:

Natural divisibility: If $(x, y) \in T$ and $\exists \lambda \in [0, 1)$ such that $(\lambda x, \lambda y) \in Z^{m+s}$, then $(\lambda x, \lambda y) \in T$.

Natural augmentability: If $(x, y) \in T$ and $\exists \lambda > 1$ such that $(\lambda x, \lambda y) \in Z^{m+s}$, then $(\lambda x, \lambda y) \in T$.

These axioms are integer-restricted counterparts of non-increasing and non-decreasing RTS axioms of conventional DEA models. Based on our earlier discussion on selective proportionality, we propose the following set of axioms for taking into account both integrality and proportionality relations in a unified technology set, $\hat{T}$.

B1. *Observations inclusion*: For all j: $(x_j, y_j) \in \hat{T}$.

B2. *Natural convexity*: If $(x, y), (\bar{x}, \bar{y}) \in \hat{T}$, $\lambda \in (0, 1)$ and $\lambda(x, y) + (1 - \lambda)(\bar{x}, \bar{y}) \in Z_+^{m+s}$, then $(\lambda x + (1 - \lambda)\bar{x}, \lambda y + (1 - \lambda)\bar{y}) \in \hat{T}$.

B3. *Natural disposability*: If $(x, y) \in \hat{T}$, $(u, v) \in Z^{m+s}$ and $0_s \leq v \leq y$, then $(x + u, y - v) \in \hat{T}$.

B4. *Natural selective proportionality*:

a. If $(x, y) \in \hat{T} \& \exists \alpha > 1$, while considering the definition of expansion scenario for $(\alpha x, \alpha y)$ in (3), with $(\alpha x, \alpha y) \in Z_+^{m+s}$ then $(\alpha x, \alpha y) \in \hat{T}$.

b. If $(x, y) \in \hat{T} \& \exists 0 \leq \alpha < 1$, while considering the definition of contraction scenario for $(\alpha x, \alpha y)$ in (4), with $(\alpha x, \alpha y) \in Z_+^{m+s}$ then $(\alpha x, \alpha y) \in \hat{T}$.

B5. *Minimum extrapolation*: $\hat{T}$ is the minimal set that satisfies axioms **B**1 − **B**5.

The third axiom **B**3 states that if we can produce a certain quantity of outputs with a given quantity of input (all with integer values), then we can also produce less integer-valued outputs with more integer-valued inputs. The notion of natural disposability (**B**3) can be interpreted as the integer-restricted counterpart of free disposability. Standard convexity is also restricted with integer-valued data and written by (**B**2) as natural convexity. Both parts of (**B**4) can be interpreted as discrete variants of RTS axioms since they limit continuous re-scaling of input and output data.

The following theorem proves that under our adapted set of axioms, reference technology $T_{HRS}^{IDEA} = T_{HRS} \cap Z_+^{m+s}$ is the smallest set among all possible technologies that satisfy the mentioned axioms.

Theorem 1 *Production set T_{HRS}^{IDEA} is the intersection of all sets that satisfy the axioms of feasibility (**B**1), natural convexity (**B**2), natural disposability (**B**3), and natural selective proportionality (**B**4).*

Proof See Appendix 1. Referencing Theorem 1, T_{HRS}^{IDEA} is the minimum extrapolation technology set under the refined set of axioms, **B**1 − **B**4. Now consider the following PPS of the HRS technology that is consist of all integer-valued input and output data:

$$T_{HRS}^{IDEA} = \left\{ (x,y) \in Z_+^{m+s} \middle| \begin{array}{l} x^P \geq \sum_{j=1}^{n} \lambda_j \left(\alpha_j \beta_j x_j^P \right) \ , y^P \leq \sum_{j=1}^{n} \lambda_j \left(\alpha_j \beta_j y_j^P \right) \\ x^{NP} \geq \sum_{j=1}^{n} \lambda_j x_j^{NP} \ , y^{NP} \leq \sum_{j=1}^{n} \lambda_j \left(\alpha_j y_j^{NP} \right) \\ \sum_{J=1}^{N} \lambda_j = 1 \ , \forall j, \lambda_j \geq 0, \alpha_j \leq 1, \beta_j \geq 1 \end{array} \right\} \qquad (6)$$

To avoid non-linear terms in (6) and following Podinovski (2004), we use some variable substitutions as follows:

Let, $\lambda_j + \mu_j - v_j = \lambda_j \alpha_j \beta_j$ and $\lambda_j - v_j = \lambda_j \alpha_j$, $j = 1, \ldots, n$. Accordingly, we have; $\mu_j = \lambda_j \alpha_j (\beta_j - 1), v_j = \lambda_j (1 - \alpha_j), j = 1, \ldots, n$. This means that the technology set T_{HRS}^{IDEA} can be equivalently state as:

$$T_{HRS}^{IDEA} = \left\{ (x,y) \in Z_+^{m+s} \middle| \begin{array}{l} x^P \geq \sum_{j=1}^{n} \left(\lambda_j + \mu_j - v_j \right) x_j^P \ , y^P \leq \sum_{j=1}^{n} \left(\lambda_j + \mu_j - v_j \right) y_j^P \ , \\ x^{NP} \geq \sum_{j=1}^{n} \lambda_j x_j^{NP}, \ , y^{NP} \leq \sum_{j=1}^{n} \left(\lambda_j - v_j \right) y_j^{NP} \\ \sum_{j=1}^{n} \lambda_j = 1 \ , \forall j, \lambda_j - v_j \geq 0, \mu_j \geq 0, v_j \geq 0 \end{array} \right\}$$

$$(7)$$

T_{HRS}^{IDEA} is linear in terms of the unknown variables λ_j, μ_j and v_j. The right-hand side of the constraints of T_{HRS}^{IDEA} are free of scaling variables.

Having access to the structure of T_{HRS}^{IDEA}, now we can proceed to evaluate the performance measurement of any production units in the new technology set.

5 Efficiency assessment

Analysis step is now performed by constructing the true production set based on the refined axioms by computing efficiency measures for the observed units. To do this, we adapt Farrell (1957)'s input radial measure in the new discrete production set T_{HRS}^{IDEA} to suggest the following radial measure for evaluating DMU_k:

$$\text{Eff}_{HRS}^{IDEA}(x_k, y_k) = \min\{\theta | \exists (x, y) \in T_{HRS}^{IDEA}; \theta x_k \geq x, y_k \leq y\} \tag{8}$$

Adapting this radial measure with the linear version of the HRS technology, introduced by Podinovski (2004), we provide the following input-oriented MILP model for measuring efficiency score of DMU_k:

$$
\begin{aligned}
&\min_{\lambda, \mu, v, \breve{x}} \theta \\
&s.t. \\
&\sum_{j=1}^{n} (\lambda_j + \mu_j - v_j) x_{i_1 j}^{IP} \leq \breve{x}_{i_1 k}^{IP} \quad i_1 = 1, ..., m_1 \\
&\breve{x}_{i_1 k}^{IP} \leq \theta x_{i_1 k}^{IP} \quad i_1 = 1, ..., m_1 \\
&\sum_{j=1}^{n} \lambda_j x_{i_2 j}^{INP} \leq \breve{x}_{i_2 k}^{INP} \quad i_2 = 1, ..., m_2 \\
&\breve{x}_{i_2 k}^{INP} \leq \theta x_{i_2 k}^{INP} \quad i_2 = 1, ..., m_2 \\
&\sum_{j=1}^{n} (\lambda_j + \mu_j - v_j) y_{r_1 j}^{IP} \geq y_{r_1 k}^{IP} \quad r_1 = 1, ..., s_1 \\
&\sum_{j=1}^{n} (\lambda_j - v_j) y_{r_2 j}^{INP} \geq y_{r_2 k}^{INP} \quad r_2 = 1, ..., s_2 \\
&\sum_{j=1}^{n} \lambda_j = 1 \\
&\breve{x}_{i_1 k}^{IP} \in Z_+ \quad i_1 = 1, ..., m_1 \\
&\breve{x}_{i_2 k}^{NIP} \in Z_+ \quad i_2 = 1, ..., m_2 \\
&\lambda_j - v_j \geq 0, \mu_j \geq 0, \ v_j \geq 0 \ j = 1, ..., n
\end{aligned} \tag{9}
$$

The proposed model requires the outmost input radial reduction to reach an integer-valued point in the corresponding PPS that dominates DMU_k. This approach is different from the first proposed model by Lozano and Villa (2006). For dealing with Pareto-efficiency, the second phase optimization model is needed to obtain non-dominated target points in the corresponding PPS. A modified second phase model is suggested in Jie et al. (2015).

In the presence of proportional inputs/outputs and integer-valued inputs, this model can be formulated as follows:

$$
\rho^* = \max \sum_{i_1=1}^{m_1} s_{i_1 k}^{-IP} + \sum_{r_2=1}^{s_2} s_{r_1 k}^{+IP} + \sum_{i_2=1}^{m_2} s_{i_2 k}^{-INP} + \sum_{r_2=1}^{s_2} s_{r_2 k}^{+INP}
$$

$$
+ \sum_{i_1=1}^{m_1} \hat{s}_{i_1 k}^{-IP} + \sum_{i_2=1}^{m_2} \hat{s}_{i_2 k}^{-INP}
$$

$s.t.$

$$
\sum_{j=1}^{n} \left(\lambda_j + \mu_j - \nu_j \right) x_{i_1 j}^{IP} - \hat{s}_{i_1 k}^{-IP} = \widetilde{x}_{i_1 k}^{IP} \quad i_1 = 1, ..., m_1
$$

$$
\sum_{j=1}^{n} \left(\lambda_j + \mu_j - \nu_j \right) y_{r_1 j}^{IP} - s_{r_1 k}^{+IP} = y_{r_1 k}^{IP} \quad r_1 = 1, ..., s_1
$$

$$
\sum_{j=1}^{n} \lambda_j x_{i_2 j}^{INP} + \hat{s}_{i_2 k}^{-INP} = \widetilde{x}_{i_2 k}^{INP} \quad i_2 = 1, ..., m_2
$$

$$
\sum_{j=1}^{n} \left(\lambda_j - \nu_j \right) y_{r_2 j}^{INP} - s_{r_2 k}^{+INP} = y_{r_2 k}^{INP} \quad r_2 = 1, ..., s_2 \qquad (10)
$$

$$
\theta^* x_{i_1 k}^{IP} - s_{i_1 k}^{-IP} \geq \widetilde{x}_{i_1 k}^{IP} \quad i_1 = 1, ..., m_1
$$

$$
\theta^* x_{i_2 k}^{INP} - s_{i_2 k}^{-INP} \geq \widetilde{x}_{i_2 k}^{INP} \quad i_2 = 1, ..., m_2
$$

$$
\sum_{j=1}^{n} \lambda_j = 1
$$

$$
\lambda_j - \nu_j \geq 0, \lambda_j \geq 0, \mu_j \geq 0, \nu_j \geq 0 \quad j = 1, ..., n
$$

$$
\widetilde{x}_{i_1 k}^{IP} \in Z_+, \widetilde{x}_{i_2 k}^{NIP} \in Z_+ \quad i_1 = 1, ..., m_1, i_2 = 1, ..., m_2
$$

$$
s^{-IP} \geq 0, s^{+IP} \geq 0, s^{-INP} \geq 0, s^{+INP} \geq 0
$$

$$
\hat{s}_{i_1 k}^{-IP} \geq 0, \hat{s}_{i_2 k}^{-INP} \geq 0 \quad i_1 = 1, ..., m_1, i_2 = 1, ..., m_2
$$

Model (10) is solved to obtain benchmark units. Having access to the optimal solution of this model, the benchmark units can be obtained from the following formula:

$$
\left(x_{i_1 k}^{IP} - s_{i_1 k}^{-IP*}, y_{r_1 k}^{IP} + s_{r_1 k}^{+IP*}, x_{i_2 k}^{INP} - s_{i_2 k}^{-INP*}, y_{r_2 k}^{INP} + s_{r_2 k}^{+INP*} \right) \qquad (11)
$$

6 Illustrative examples

In this section, we apply our provided approach for illustration and comparison purposes. First, we present a simple example to compare the results of the traditional and the new models in the efficiency evaluation of the observed units. Then we apply our provided approach in the performance assessment of 30 high schools in Iran.

6.1 A simple example

To illustrate the applicability of the proposed approach, we take data for 10 hypothetical DMUs, containing single input i_1 and two outputs o_1 and o_2, where all the data is assume to be integer. It is also assumed that i_1 and o_1 are proportional. The input/output data set are listed in Table 1.

Let DMU_5 be the unit under evaluation. When the input and the first output are proportional, the proposed model by Podinovski (2004) may be stated as follows:

$$
\min \theta
$$
s.t.
$$
33(\lambda_1 + \mu_1 - \nu_1) + 39(\lambda_2 + \mu_2 - \nu_2) + 35(\lambda_3 + \mu_3 - \nu_3) +
$$
$$
32(\lambda_4 + \mu_4 - \nu_4) + 36(\lambda_5 + \mu_5 - \nu_5) + 35(\lambda_6 + \mu_6 - \nu_6) +
$$
$$
34(\lambda_7 + \mu_7 - \nu_7) + 40(\lambda_8 + \mu_8 - \nu_8) + 38(\lambda_9 + \mu_9 - \nu_9) +
$$
$$
39(\lambda_{10} + \mu_{10} - \nu_{10}) \le 36\theta
$$
$$
22(\lambda_1 + \mu_1 - \nu_1) + 54(\lambda_2 + \mu_2 - \nu_2) + 76(\lambda_3 + \mu_3 - \nu_3) +
$$
$$
84(\lambda_4 + \mu_4 - \nu_4) + 87(\lambda_5 + \mu_5 - \nu_5) + 43(\lambda_6 + \mu_6 - \nu_6) +
$$
$$
33(\lambda_7 + \mu_7 - \nu_7) + 45(\lambda_8 + \mu_8 - \nu_8) + 23(\lambda_9 + \mu_9 - \nu_9) + \qquad (12)
$$
$$
54(\lambda_{10} + \mu_{10} - \nu_{10}) \ge 87
$$
$$
43(\lambda_1 - \nu_1) + 51(\lambda_2 - \nu_2) + 176(\lambda_3 - \nu_3) + 34(\lambda_4 - \nu_4) +
$$
$$
80(\lambda_5 - \nu_5) + 5(\lambda_6 - \nu_6) + 136(\lambda_7 - \nu_7) + 56(\lambda_8 - \nu_8) +
$$
$$
67(\lambda_9 - \nu_9) + 68(\lambda_{10} - \nu_{10}) \ge 80
$$
$$
\sum_{j=1}^{10} \lambda_j = 1
$$
$$
\lambda_j - \nu_j \ge 0 \quad j = 1, ..., 10
$$
$$
\nu_j \ge 0 \quad j = 1, ..., 10
$$

By computing the above linear programming model the optimal (non-zero) results of assessing DMU_5 is $\theta^* = 0.97$, $\lambda_3^* = 0.32$, $\lambda_4^* = 0.68$, $\mu_4^* = 0.07$.

Now consider the new model presented for dealing with both proportionality and integer-valued data:

Table 1 Input / output data

DMU	i_1	o_1	o_2	DMU	i_1	o_1	o_2
1	33	22	43	6	35	43	5
2	39	54	51	7	34	33	136
3	35	76	176	8	40	45	56
4	32	84	34	9	38	23	67
5	36	87	80	10	39	54	68

$$\min \theta$$

s.t.

$$33(\lambda_1 + \mu_1 - \nu_1) + 39(\lambda_2 + \mu_2 - \nu_2) + 35(\lambda_3 + \mu_3 - \nu_3)+$$
$$32(\lambda_4 + \mu_4 - \nu_4) + 36(\lambda_5 + \mu_5 - \nu_5) + 35(\lambda_6 + \mu_6 - \nu_6)+$$
$$34(\lambda_7 + \mu_7 - \nu_7) + 40(\lambda_8 + \mu_8 - \nu_8) + 38(\lambda_9 + \mu_9 - \nu_9)+$$
$$39(\lambda_{10} + \mu_{10} - \nu_{10}) \leq \tilde{x}$$
$$22(\lambda_1 + \mu_1 - \nu_1) + 54(\lambda_2 + \mu_2 - \nu_2) + 76(\lambda_3 + \mu_3 - \nu_3)+$$
$$84(\lambda_4 + \mu_4 - \nu_4) + 87(\lambda_5 + \mu_5 - \nu_5) + 43(\lambda_6 + \mu_6 - \nu_6)+$$
$$33(\lambda_7 + \mu_7 - \nu_7) + 45(\lambda_8 + \mu_8 - \nu_8) + 23(\lambda_9 + \mu_9 - \nu_9)+$$
$$54(\lambda_{10} + \mu_{10} - \nu_{10}) \geq 87 \tag{13}$$
$$43(\lambda_1 - \nu_1) + 51(\lambda_2 - \nu_2) + 176(\lambda_3 - \nu_3) + 34(\lambda_4 - \nu_4)+$$
$$80(\lambda_5 - \nu_5) + 5(\lambda_6 - \nu_6) + 136(\lambda_7 - \nu_7) + 56(\lambda_8 - \nu_8)+$$
$$67(\lambda_9 - \nu_9) + 68(\lambda_{10} - \nu_{10}) \geq 80$$
$$36\theta \geq \tilde{x}$$
$$\tilde{x} \in Z_+$$
$$\sum_{j=1}^{10} \lambda_j = 1$$
$$\lambda_j - \nu_j \geq 0 \quad j = 1, ..., 10$$
$$\nu_j \geq 0 \quad j = 1, ..., 10$$

Solving the above MILP model for evaluating DMU_5 leads to $\theta^* = 1, \lambda_3^* = 1, \mu_4^* = 0.61, \nu_3^* = 0.53$. Comparing the solutions indicaates different results for DMU_5 in the two models. In Model (12), unit is classified as inefficient while in (13) it is efficient. This is a notable difference. The new MILP model is designed to compare the unit under evaluation with the integer-valued production possibilities. With regards to the integrality assumption, DMU_5 is compared with the continuous part of production set in (12) which is not necessarily feasible. This gives evidence to the fact that in applications where integer-valued data exist, applying appropriate production set is necessary and the new model provides more reliable results.

DMU	R-EFFI	I-EFFI	FDH	DMU	R-EFFI	I-EFFI	FDH
1	0.297	0.303	1	6	0.486	0.486	0.914
2	0.563	0.564	0.897	7	0.795	0.826	1
3	1	1	1	8	0.468	0.475	0.875
4	1	1	1	9	0.358	0.368	0.895
5	0.974	1	1	10	0.576	0.590	0.897

Table 2 Efficiency scores of Model (8) and (9)

Computed scores for the rest of the units are reported in Table 2, where R-EFFI and I-EFFI stand for efficiencies with real and integer-valued data, respectively. Comparing the columns under R-EFFI and I-EFFI, we can find that seven out of units have different efficiency scores in these two approaches. Moreover, as the Table shows, although, DMU_5 is efficient in R-EFFI, it is inefficient in I–EFFI. This shows that the classification of efficient and inefficient units is quite different in these approaches. The new method brings more discrimination power into evaluation.

Finally, the FDH model is run for the data set by ignoring the selective proportionality and integrality assumptions. The results are reported in the last column of Table 2. As the results show, five out of ten units prevail as efficient. This indicates that the FDH model overestimates the efficiency of the units.

6.2 Empirical example

Performance analysis and evaluation in educational parts such as schools and universities have been widely studied by researchers from different perspectives. Kadoić et. al. (2018) in their research considered an approach that includes problem identification, objectives of a solution, design and development, demonstration of the artifact, evaluation, and dissemination for strategic decision-making in higher education. They have used a design science research process and provided new decision-making with two components that are based on the analytic network process and social network analysis. Begičević et. al. (2010) considered two plans in higher education institutions. First, the authors have made a plan ready for those activities yield in the execution of a portfolio of projects at the institutional level. Second, those scenarios are important for decision on whether to start a new project application. In this scenario, it is also dealt with that which project to choose, in case of several project ideas and limited resources. Jablonsky (2016), presented a new method, for overcoming the disadvantage of existing models, that considered efficiency assessment of decision-making units within the whole production chain. Cordero et. al. (2016) presented a new method for educational performance assessment in Spanish. They have used the non-parametric free disposal hull model and provided a decomposition based on overall inefficiency between different components while considering the differences between public and state-subsidized private schools. Čampelj et. al. (2018), provided a multi-attribute modeling approach for assessing the implementation of Information and Communication Technologies in schools. They claimed that the key feature of their study is utilizing qualitative value scales for attributes that do not have exact values. Cherchye et al. (2019) presented a unified method for assessing the productivity of secondary schools in the DEA framework.

All of the previously studied DEA-based research, considered the constant or variable returns to scale models, and issues such as integrality of the data and selective proportionality are not considered. However, two important points must be considered in performance analysis in educational parts: factors such as the number of students and number of teachers are integer and real values are not allowed to these factors. Moreover, it is possible to have a subset of inputs such as sport training camps that are not proportional with a subset of outputs such as the number of students who won in sports competitions. These issues show that there is a need to modify the existing approaches in such a way that they can be applicable in these hybrid cases.

In this section, an empirical application on performance evaluation of thirty high schools in Iran is presented. According to the Iranian constitution, education for all Iranian children and adolescents is free of charge, and the government has to provide education for all through the Ministry of Education. Iran's educational system had elementary, secondary, and pre-university courses. In the final years of the eighties, the educational system changed to 3–3–3–3. This system has generally changed to six years of elementary and secondary education. The first people trained in this system were born in the second half of the 79th and the first half of the 80 s who had the elementary sixth grade. They graduated at the end of the first high school, and this year they study at eleventh grade. Practicing and technical students require skills in vocational schools and institutes of labor and knowledge. Graduates from the pre-university period entered the university if they succeeded in the national examinations (entrance examination), which had been removed by changing the system and replacing it with the sixth elementary element. Student Olympiads are a test that is held annually at high school students' level. The purpose of this test is to boost talented Iranian student and selecting a multiplayer team to take part in international scientific Olympiads. At present, the student Olympiads in the country are held in eight literary: mathematical, biology, rehabilitation stem and medical stem cells, chemistry, physics, computers and astronomy, astrophysics and geography, among which the Stem Cell and Medical Stem Cell Olympiad was rebuilt. Other Olympiads are officially held and include exam facilities.[1]

Based on the foregoing discussion, the performance of the Iranian educational system is an important issue that needs to be analyzed in dept. This section applies our proposed model to data of thirty high schools in Tehran. There are two main education types in Iran[2]: K-12 and higher educations. K-12 indicates the sum of kindergarten, primary and secondary education in the country which takes 12 years. K-12 education in Iran is under the supervision of the ministry of education[3] and higher education is supervised by the ministry of science, research, and technology[4] and the ministry of health and medical education*.[5]

We investigated the performance of high school education systems in Tehran by comparing and contrasting our new approach by proportional DEA models

---

[1] http://olympiad.sanjesh.org/En/Default.aspx.

[2] https://en.wikipedia.org/wiki/Education_in_Iran.

[3] http://medu.ir/fa/.

[4] https://www.msrt.ir/en.

[5] * http://www.behdasht.gov.ir/page/en.

Table 3 Inputs and outputs of high schools

Inputs		Outputs	
Description	Notation	Description	Notation
Students	I_1	Graduate students	O_1
Expert teaching staff	I_2	Students passed central exams of universities and Olympiad	O_2
Sport training camps	I_3	Students won sport medals	O_3

with integer-valued input/output data and the tradidional FDH model. Moreover, we highlighted the efficiency scores and also target points with different approaches. Table 3 summarises the selected inputs and outputs for 30 high schools in Tehran.

I_1 indicates the number of students in their third year of high school. This quantity shows the population density of each high school that obviously takes integer values. I_2 designates the number of specialized teachers for classes held in mathematics and physics. A specialized teacher needs to be employed for each specific course work that can teach and mentor students in their schooling. These teachers may also assist students in university entrance exams and the Olympiad selection process. I_3 specifies the number of sports training camps held in each high school to support those interested in sports competitions and activities. O_1 shows the number of students who passed all course work and graduated in their third year-inherently. O_2 is the number of students who succeded in university entrance exams plus the number of students who entered to the Olympiad competitions. O_3 indicates the number of students who won in sports competetions.

The important issue here is the possibility of having a subset of inputs that is proportional with a subset of outputs. This would point out that alterations in the number of teachers and the number of students who are able to pass all the courses and graduate from their current level have the same portion as the number of students who pass university central exams and Olympics. Thus, these subsets of inputs and outputs are considered to be proportional to each other since factors such as 'the availability of sports facilities' are important. Therefore, I_3 and O_3 are not considered to be proportional to each other, and they are treated as non-proportional input and output.

Table 4 exhibits the integer-valued inputs and outputs data for 30 high schools in Tehran.

Table 5 provides the results of efficiency analysis while applying real and integer-valued models for evaluating efficiency scores. The results point out that the efficiency scores obtained from the new integer-valued model (I-EFFI) are always greater than or equal to those obtained from the traditional real valued model (R-EFFI). The maximum difference is related to DMU_3 where its I-EFF and R-EFF scores is 0.7632 and 0.7223, respectively. There are 13 efficient DMUs by our new integer-valued approach meanwhile this number for the R-EFFI approach is reduced to 9. To be more specific, DMUs 19, 20, 23, and 25 are in the efficient set in our approach while they are in the inefficient set in the real-valued approach.

Table 4 Inputs / outputs of 30 high schools

DMU	I_1	I_2	I_3	O_1	O_2	O_3	DMU	I_1	I_2	I_3	O_1	O_2	O_3
1	113	10	14	88	51	23	16	110	10	19	94	75	7
2	133	12	20	112	75	13	17	95	7	18	89	71	5
3	114	9	17	68	61	12	18	136	8	17	112	93	12
4	113	11	18	95	76	2	19	88	7	14	81	66	5
5	145	11	21	130	107	35	20	99	7	17	89	72	10
6	118	9	15	104	57	33	21	107	7	19	94	68	22
7	91	7	16	83	69	7	22	128	12	22	98	75	3
8	83	8	14	71	39	4	23	110	7	14	90	75	22
9	124	10	11	101	72	10	24	95	8	18	82	61	3
10	94	9	19	73	74	1	25	137	9	13	103	94	5
11	111	11	18	84	60	8	26	104	6	19	78	62	93
12	125	12	15	106	90	10	27	126	9	21	104	85	26
13	96	10	17	76	61	3	28	144	12	16	106	84	3
14	140	9	18	106	82	3	29	96	9	15	89	71	2
15	119	11	22	104	86	9	30	131	11	19	106	85	7

Taking integer-valued data and attempting to calculate non-dominated target points affect the analysis so that feasible solutions may not be achievable in some cases. This is not the case in the presented approach here because it is attempted to find suitable points within the corresponding PPS and not merely on the efficient frontier. These two approaches differ from each other as some units are efficient considering integer-valued data since they are evaluated as inefficient units when a model is not obliged to search for the integer-valued benchmark. Efficiency scores of these units exhibit a decline when respectively considered having integer and real data like DMUs 19, 20, 23, and 25 that cover 13.3% of the results.

DMUs do exist that are inefficient in both analyzes but with different scores. These units did not manage to efficiently use inputs to produce outputs. As an important matter, these units should imitate the corresponding target units out of the set of DMUs under evaluation to make changes in their strategy in order to reach an efficient state. These units constitute 56.7% of all results. Around 30% of the high schools remain efficient in both analyzes.

Restricting the analysis to integer-valued data does not have an effect on the operation of the units. As it may be observed, efficiency classification taking integer-valued data into consideration differs from real-valued data in the analysis. The average of efficiencies in real and integer analyses are quite close, however, 0.93338 and 0.940167 respectively. It is important to mention that these results are different in the interpretation and classification of efficient and inefficient units. Some units perform efficiently with real-valued data but not with integer-valued data are considered in the analysis. Thus, relevant managers would change their strategies. For inefficient units, target units need to be introduced to be followed and imitated (Table 5).

The last notable point is that the conventional proportional model may fail to produce target points in the presence of integer-valued data. Table summarizes the

Table 5 Efficiencies obtained by Model (9) and Podinovski et al. (2014)

DMU	R-EFFI	I-EFFI	FDH	DMU	R-EFFI	I-EFFI	FDH
1	0.9103	0.9286	1.0000	16	0.9171	0.9273	1.0000
2	0.9121	0.9173	1.0000	17	1.0000	1.0000	1.0000
3	0.7223	0.7632	0.9649	18	1.0000	1.0000	1.0000
4	0.899	0.9027	1.0000	19	0.998	1.0000	1.0000
5	1.0000	1.0000	1.0000	20	0.9961	1.0000	1.0000
6	1.0000	1.0000	1.0000	21	1.0000	1.0000	1.0000
7	1.0000	1.0000	1.0000	22	0.8172	0.8203	0.9844
8	0.9203	0.9277	1.0000	23	0.9914	1.0000	1.0000
9	1.0000	1.0000	1.0000	24	0.9213	0.9263	0.9579
10	1.0000	1.0000	1.0000	25	0.9643	1.0000	1.0000
11	0.8185	0.8288	0.9444	26	1.0000	1.0000	1.0000
12	0.9529	0.9600	1.0000	27	0.9348	0.9683	1.0000
13	0.8485	0.8542	0.9167	28	0.7886	0.7917	1.0000
14	0.8729	0.8889	0.9714	29	0.9896	0.9896	1.0000
15	0.9569	0.9664	1.0000	30	0.8693	0.8779	1.0000

achieved targets of real-valued data of Podinovski (2004) and the integer-valued data of Models (9) and (10). Some targets are identical by all three models and some others differ which have been highlighted. In the highlighted cells of Table 6, the measured targets by the first phase Model (9) and the second phase Model (10) are presented within a pair of brackets and parentheses, respectively.

As already indicated, initial input data have integer values. Therefore, it is rational that corresponding targets must also have integer values. This is not a given in the conventional HRS model. Comparing the two analyzes with real- and real-valued data of Model (9), DMUs may have similar or dissimilar efficiency scores. Some units such as DMU_3 and DMU_{27} have different efficiency scores and unit DMU_{29} has scored the same in both analyzes. In performance analysis, target setting (benchmarking) is also of great importance. It assists decision-makers in better performance analysis of inefficient units and suggesting improvement schemes. Targets of efficient DMUs are not scrutinized since they perform efficiently, and there is no need to introduce specific targets from among the sets of DMUs. A notable difference could be observed in target setting for inefficient units. For this group, the computed targets of 70% of the units in the HRS model are not applicable when considering integer-valued inputs. In accordance with obtained results for target units, it is observed that these results are not just obtained by rounding (up or down) the data values. Rounding the obtained results is not always an acceptable solution since in some cases it may lead to infeasible points. This rounding up may be verified in units such as DMUs 1, 3, 25, and 15.

Consider relations in (11) for introducing non-dominated targets. The model presented by Kuosmanen and Kazemi Matin (2009) is used to find the efficient integer-valued points of the PPS to introduce target (benchmark) units. After solving Model

Table 6 Real-valued and integer-valued targets

DMUs	I_1	I_2	I_3	DMUs	I_1	I_2	I_3
1	104.929 [104] (104)	9.286 [9] (8)	13	16	101	9.182 [9] (9)	17.445 [17](17)
2	122	11.008 [10] (10)	18.346 [18] (18)	17	95	7	18
3	87	6.868 [6] (6)	12.974 [12] (12)	18	136	8	17
4	102	9.929 [9] (8)	16.248 [16] (12)	19	88	7	14
5	145	11	21	20	99	7	17
6	118	9	15	21	107	7	19
7	91	7	16	22	105	9.844 [9] (8)	18.049 [18] (13)
8	77	7.422 [7] (7)	12.988 [12] (12)	23	110	7	14
9	124	10	11	24	88	7.411 [7] (7)	16.674 [16] (13)
10	94	9	19	25	137	9	13
11	91	9.018 [9] (9)	14.757 [14] (14)	26	104	6	19
12	120	11.52 [11] (10)	14.4 [14] (14)	27	122	8	16
13	82	8.542 [8] (7)	14.521 [14] (13)	28	114	9	12
14	124.444 [124] (124)	8	16	**29**	95	8	14
15	114	10.538 [10] (9)	21.076 [21] (16)	30	114	9	16

(9) the second phase Model (10) is used for calculating the benchmarks with integer-valued components.

Optimal slacks obtained from solving Model (10) are listed in Table 7 for DMUs with at least one non-zero slack.

Moreover, Table 8 demonstrates the obtained input targets acquired from Podinovski et al. (2014) model when all data is assumed to be real. As can be easily verified, some of these points dominate acceptable integer-valued targets. It is rational and logical that targets not having integer data cannot be considered as acceptable targets when integer data exist in the analysis.

Consider Table 8 and Model (9) for assessing DMU_1. The observed input vector for this DMU is $(113, 10, 14)$ and the obtained results show that input radial real-valued target points for this unit to be $(104.9286, 9.285714, 13)$ and the resulting suitable integer-valued target point as $(104, 9, 13)$. It may be noted for further discussion of results that input improvements are $(0.9285682, 1.285714, 0)$ after non-radial improvement and solving the second phase model.

Table 7 Optimal slacks obtained from Model (10)

DMUs	I_1	I_2	I_3	o_1	o_2	o_3	Sum
1	0.929	1.286	0	4.574	22.795	0.232	29.816
2	0	0	0	0	0	0	0
3	0	0.868	0.974	8.67	0.229	0	10.744
4	0	1.929	4.248	0	0.098	1.358	7.633
12	0	1.52	0.4	2.665	0	0.973	5.558
13	0	1.542	1.521	0	2.344	5.498	5.498
14	0.4444	0	0	0	0.589	28.5	29.533
15	0	1.5378	5.0756	0	0	0.295	6.908
19	0	0	0	0	0	1.275	1.275
20	0	0	3	929	0	0.119	4.048
22	0	1.844	5.047	0	3.387	0.185	10.463
23	0	0	0	2.678	0	6.206	8.885
24	0	0.411	3.674	0	4.665	0.5367	9.286
25	3	0	0	14.188	0.258	18.25	35.695
27	0	0.714	4.333	0.614	0	0	5.662
28	0	0.5	0.667	0	0.849	1.45	3.466

Applying these non-radial changes to the real-valued input radial target, the integer-valued point $(104, 8, 13)$ is obtained which is a better estimate than the integer-valued target point obtained from the first phase model. Additionally, it can be observed that by non-radial improvements, $(0, 1.752632, 0)$, the point $(104, 7.247368, 13)$ is obtained from this model. This point is non-dominated since it is obtained from the second phase model of HRS Podinovski et al. (2014), but it does not have integer values. Thus, it cannot be an acceptable target unit for the DMU under evaluation. Although, it dominates the integer input target and is located on the frontier. Consequently, $(104, 8, 13)$ is the best integer-valued input target point. Model (10), the second phase model, maximizes the non-radial improvement and returns the optimal output slacks as $(4.574, 22.795, 0.232)$. For DMU_2, the initial input data is $(113, 12, 20)$ and the real-valued input target point is $(122, 11.007, 18.346)$ when utilizing the first phase Model (9). Imposing limitations, the integer-valued input target point is $(122, 10, 18)$ and considering Model (10), the second phase model, all the input and output slacks are equal to zero.

In evaluating DMU_4 the input vector is $(113, 11, 18)$. Solving the first phase Model (9) leads to the real-valued input radial target $(102, 9.929, 16.248)$. The nearest dominated point is $(102, 9, 16)$. Input and output slacks are $(1.929, 0, 4.248)$ and $(0, 0.098, 1.358)$, respectively for non-radial improvements reaching the non-dominated point. Thus, input integer-valued target is $(102, 8, 12)$. If it is desired to find a non-dominated point, Model (10) produces slacks of $(0, 1.454, 4.045)$ which lead to point $(102, 7.546, 11.955)$. This point dominates the integer-valued point obtained from the second phase model. However, it is not considered suitable as a target point due to it containment of real-valued input.

Table 8	Input targets obtained from Podinovski et al. (2014)

DMU	I_1	I_2	I_3	DMU	I_1	I_2	I_3
1	104	7.247	13	16	102	9	17
2	122	10	18	17	95	7	18
3	86.363	6	12	18	136	8	17
4	102	7.546	11.955	19	88	6.627	13.898
5	145	11	21	20	99	7	13.834
6	118	9	15	21	107	7	19
7	91	7	15	22	105	7.757	12.74
8	77	7	12	23	110	7	14
9	124	10	11	24	88	6.508	12.249
10	94	9	19	25	133.404	9	13
11	91	9	14	26	104	6	19
12	120	9.171	14	27	122	8	16.49
13	82	6.086	12.405	28	114	8.403	12
14	124	8	16	29	95	8	14
15	114	8.717	16.546	30	114	9	16

It should now be clear that solving Model (10) is necessary since the second phase model can find better target points when assessing most DMUs with integer values. Some of the DMUs are efficient while being assessed by Model (10) and there is no need for finding corresponding target units for them as they perform efficiently. In obtained results for some inefficient DMUs, it is only possible to find radial integer-valued input targets, and it is not possible to find improvements in inputs by non-radial changes. This indicates that the optimal value of the objective function of the second phase model for finding an integer-valued target point is zero. Finding reference points is of great importance in the efficiency evaluation of DMUs and target setting.

Having this information, managers may establish a role model for their systems. By manipulating them, managers can reach an efficient states. As Podinovski (2004) stated, for any j where $\lambda_j^* - v_j^* > 0$, the corresponding DMU_j is efficient. Therefore, referenced DMUs by this formula are detected from the analysis as presented in Table. As indicated in Table 9 there are 10 efficient units: DMUs 5, 6, 7, 9, 10, 17, 18, 19, 21, and 26. In accordance with the basic definition of the reference set, the mentioned efficient DMUs have a positive value of $\lambda^* - v^*$.

Models (9) and (10) show radial and non-radial changes in the inputs. With these changes, it is possible to reach points with integer values that dominate the DMUs under assessment. As it can be seen, based on the optimal solution of the first and the second phase models listed in Table 8, not all these points have integer values. Thus, although they dominate integer target points obtained from our proposed model, they cannot be considered as acceptable target units. For more substantial analysis of the results, norm L_1 of three input elements to their corresponding target points is calculated. Units with the smallest norm L_1 values are DMUs 20, 25, and 29. Corresponding norm L_1 value for all of these three units is 3 and the corresponding efficiency scores are 0.9961, 0.9643, and 0.99896 respectively. It may be stated that these units perform efficiently. The high efficiency scores indicate that

Table 9 Reference set

DMU	$(\lambda_j^* - v_j^* > 0)$		DMU	$(\lambda_j^* - v_j^* > 0)$	
1	$DMU_{26}(0.25)$		16	$DMU_5(0.097)$	$DMU_{17}(0.718)$
2	$DMU_5(0.287)$	$DMU_{17}(0.590)$	17	$DMU_{17}(1)$	
3	$DMU_9(0.0375)$	$DMU_{26}(0.125)$	18	$DMU_{18}(1)$	
4	$DMU_5(0.0955)$	$DMU_{17}(0.006)$	19	$DMU_{19}(1)$	
5	$DMU_5(1)$		20	$DMU_{17}(0.266)$	$DMU_{26}(0.084)$
6	$DMU_6(1)$		21	$DMU_{21}(1)$	
7	$DMU_7(1)$		22	$DMU_5(0.063)$	$DMU_{17}(0.196)$
8	$DMU_5(0.0895)$	$DMU_6(0.0263)$	23	$DMU_{18}(0.115)$	$DMU_{26}(0.288)$
9	$DMU_9(1)$		24	$DMU_5(0.076)$	$DMU_{17}(0.178)$
10	$DMU_{10}(1)$		25	$DMU_{26}(1)$	
11	$DMU_5(0.210)$		26	$DMU_5(0.099)$	$DMU_{17}(0.054)$
12	$DMU_5(0.294)$	$DMU_{26}(0.007)$	27	$DMU_5(0.089)$	$DMU_{17}(0.015)$
13	$DMU_5(0.140)$	$DMU_{17}(0.085)$	28	$DMU_5(0.082)$	$DMU_{17}(0.026)$
14	$DMU_{21}(0.375)$	$DMU_{26}(0.25)$	29	$DMU_5(0.136)$	$DMU_{17}(0.518)$
15	$DMU_5(0.109)$	$DMU_7(0.781)$	30	$DMU_5(0.121)$	$DMU_{17}(0.531)$

it is rational for the corresponding norm L_1 to be small. DMUs 11, 13, and 28 have larger norm L_1 values of 35, 26, 21, and 37 respectively; their individual efficiency scores are 0.7223, 0.8181, 0.8485, and 0.7883. High norm L_1 values are relatively predicted by the efficiency scores, as mentioned above. This study compares variations of three inputs with their target points. Considering the norm L_1 values, it may be said that x_1 has more reduction after x_3 and then x_2.

Although the CRS and VRS technologies respectively imply safer assumptions and better discrimination, HRS technology is a more powerful tool for efficiency discrimination. Therefore, in order to increase the discriminatory power of DEA models, it may be more useful to consider HRS over conventional technologies. This may be particularly useful when inputs and outputs do not necessarily exhibit the same proportionality.

The VRS technology is more limited than the HRS technology because the former technology is a subset of the latter one. Comparing average efficiency in HRS and VRS technologies, the former is 0.95 and the latter is 0.93. Note that that HRS is not a subset of CRS technology. However, there are sets in the sample that exhibit differing efficiencies that may be lower in CRS or HRS models. Note that simple rounding methods for obtaining integer-valued points may lead to dominated or even infeasible points. Thus, utilizing HRS analysis is more discriminating.

We have also applied the FDH model to this data set. The results of efficiency scores are listed in the last column of Table 5. The target points corresponding to inefficient units are given in Table 10. All the computed efficiency scores are estimated tobe greater than 0.9000 and 24 units were classified as efficient, while only 6 units prevailed as inefficient. This clearly shows that the efficiency scores are overestimated.

Table 10 Input–Output targets obtained from FDH model

DMU	I_1	I_2	I_3	o_1	o_2	o_3	DMU	I_1	I_2	I_3	o_1	o_2	o_3
1	113	10	14	88	51	23	16	110	10	19	94	75	7
2	133	12	20	112	75	13	17	95	7	18	89	71	5
3	110	7	14	90	75	22	18	136	8	17	112	93	12
4	113	11	18	95	76	2	19	88	7	14	81	66	5
5	145	11	21	130	107	35	20	99	7	17	89	72	10
6	118	9	15	104	57	33	21	107	7	19	94	68	22
7	91	7	16	83	69	7	22	126	9	21	104	85	26
8	83	8	14	71	39	4	23	110	7	14	90	75	22
9	124	10	11	101	72	10	24	91	7	16	83	69	7
10	94	9	19	73	74	1	25	137	9	13	103	94	5
11	99	7	17	89	72	10	26	104	6	19	78	62	93
12	125	12	15	106	90	10	27	126	9	21	104	85	26
13	88	7	14	81	66	5	28	125	12	15	106	90	10
14	136	8	17	112	93	12	29	96	9	15	89	71	2
15	119	11	22	104	86	9	30	131	11	19	106	85	7

As a final word, it can strongly be stated that the proposed model is more useful than the prior arts in the presence of selective proportionality and integer-valued data.

7 Conclusion

In real applications of DEA, we sometimes face cases in which some input/output variables are integer-valued and simultaneously, a subset of input variables is not proportional with a subset of output variables. Ignoring this important issue leads to a misestimating of efficiency analysis. This paper introduces an axiomatic foundation to construct a production technology set with selective proportionality assumption in the presence of integer-valued data. A novel MILP model is then provided to compute modified Farrell measure with respect to both HRS and integrality assumptions. A numerical example of an actual data set for 30 high schools in Tehran is utilized to demonstrate the proposed approach. This work demonstrates that efficiency obtained from the proposed model is more reliable than those obtained from conventional HRS models where integrality assumption is completely overlooked. Insignificant but notable differences are reported in setting targets for both conventional HRS and the proposed approach. It is illustrated that the proposed model is useful for benchmarking.

Two potential research issues need worth mentioning. First, one can explore from our proposed model to incorporate selective convexity and provide Pareto-efficient targets. Second, one could think of extending the approach to study the nature of returns to scale and determining an optimal scale size for the observations. These may consider as interesting challenges in expanding our knowledge on the subject.

Appendix 1

Proof of Theorem 1 The following items need to verify:

a. T_{HRS}^{IDEA} satisfies the mentioned axioms.
b. If $\hat{T}$ is an arbitrary technology set that satisfies the same axioms, $\boldsymbol{B}1 - \boldsymbol{B}4$, then it contains T_{HRS}^{IDEA}. The first step is straightforward and a trivial verification points that T_{HRS}^{IDEA}

   involves the observed DMUs ($\boldsymbol{B}1$), satisfies natural convexity ($\boldsymbol{B}2$), natural disposability ($\boldsymbol{B}3$), and natural selective proportionality ($\boldsymbol{B}4$). Regarding the second step, consider integer-valued data and assume that $\hat{T} \subseteq Z_+^{m+s}$ be any arbitrary technology set that satisfies $\boldsymbol{B}1 - \boldsymbol{B}4$. Let $T = \text{conv}(\hat{T}) \subseteq R_+^{m+s}$ be the convex hull of $\hat{T}$. According to Kuosmanen, Kazemi Matin (2009), T is still contains the observations and satisfies the continues versions of $\boldsymbol{B}2 - \boldsymbol{B}4$. As it is shown in Podinovski et al. (2014), the production set T is the smallest set that satisfies the mentioned axioms $\boldsymbol{B}1 - \boldsymbol{B}4$. So, we have $T_{HRS} \subseteq T$. Now, by restricting the vectors to the integer valued points, we obtain

$$T_{HRS} \cap Z_+^{m+s} \subseteq T \cap Z_+^{m+s} \tag{14}$$

   The left and right sides of (6) can be considered as T_{HRS}^{IDEA} and $\hat{T}$, respectively. This concludes that $T_{HRS}^{IDEA} \subseteq \hat{T}$, which completes the proof.

Appendix 2

Generalized model: In Sect. 4, we assumed that all input and output data can only take integer values. Now, we extend the proposed model to the case in which some inputs and outputs are real-valued. Suppose the input–output vector (x, y) is partitioned as $(x^{IP}, x^{INP}, x^P, x^{NP}, y^{IP}, y^{INP}, y^P, y^{NP})$ in which:

x^{IP} are integer $-$ valued proportional input $m_1 -$ vector,

x^{INP} are integer $-$ valued nonproportional input $m_2 -$ vector,

x^P are real $-$ valued proportional input $m_3 -$ vector,

x^{NP} are real $-$ valued nonproportional input $m_4 -$ vector,

y^{IP} are integer $-$ valued proportional output $s_1 -$ vector,

y^{INP} are integer $-$ valued nonproportional output $s_2 -$ vecto*r*,

y^P are real $-$ valued proportional output s_3 $-$ vector,

y^{NP} are real $-$ valued nonproportional output s_4 $-$ vector,

The following input-oriented MILP model is proposed to analyze the relative efficiency of DMU_k:

$$\min_{\lambda,\mu,\nu,\breve{x}} \theta$$

s.t.

$$\sum_{j=1}^{n} \left(\lambda_j + \mu_j - \nu_j\right)x_{i_1j}^{IP} \leq \breve{x}_{i_1k}^{IP} \quad i_1 = 1,...,m_1$$

$$\breve{x}_{i_1k}^{IP} \leq \theta x_{i_1k}^{IP} \quad i_1 = 1,...,m_1$$

$$\sum_{j=1}^{n} \lambda_j x_{i_2j}^{INP} \leq \breve{x}_{i_2k}^{INP} \quad i_2 = 1,...,m_2$$

$$\breve{x}_{i_2k}^{INP} \leq \theta x_{i_2k}^{INP} \quad i_2 = 1,...,m_2$$

$$\sum_{j=1}^{n} \left(\lambda_j + \mu_j - \nu_j\right)x_{i_1j}^{P} \leq \theta x_{i_1k}^{P} \quad i_3 = 1,...,m_3$$

$$\sum_{j=1}^{n} \lambda_j x_{i_2j}^{NP} \leq \theta x_{i_2k}^{NP} \quad i_4 = 1,...,m_4$$

$$\sum_{j=1}^{n} \left(\lambda_j + \mu_j - \nu_j\right)y_{r_1j}^{IP} \geq y_{r_1k}^{IP} \quad r_1 = 1,...,s_1$$

$$\sum_{j=1}^{n} \left(\lambda_j - \nu_j\right)y_{r_2j}^{INP} \geq y_{r_2k}^{INP} \quad r_2 = 1,...,s_2$$

$$\sum_{j=1}^{n} \left(\lambda_j + \mu_j - \nu_j\right)y_{r_1j}^{P} \geq y_{r_1k}^{P} \quad r_3 = 1,...,s_3$$

$$\sum_{j=1}^{n} \left(\lambda_j - \nu_j\right)y_{r_2j}^{NP} \geq y_{r_2k}^{NP} \quad r_4 = 1,...,s_4$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

$$\breve{x}_{i_1k}^{IP} \in Z_+ \quad i_1 = 1,...,m_1$$

$$\breve{x}_{i_2k}^{NIP} \in Z_+ \quad i_2 = 1,...,m_2$$

$$\lambda_j - \nu_j \geq 0, \mu_j \geq 0, \nu_j \geq 0 \quad j = 1,...,n$$

It should be noted that $m_1 + m_2 + m_3 + m_4 = m$ and $s_1 + s_2 + s_3 + s_4 = s$ and m and s are respectively the total number of inputs and outputs.

References

Afsharian M, Ahn H, Alirezaee M (2015) Developing selective proportionality on the FDH models: new insight on the proportionality axiom. Int J Inf Decis Sci 7(2):99

Alirezaee MR, Boloori F (2012) Proportional production trade-offs in DEA. Asia-Pac J Oper Res 29(06):1250035

Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manage Sci 30(9):1078–1092

Begičević N, Divjak B, Hunjak T (2010) Decision-making on prioritization of projects in higher education institutions using the analytic network process approach. CEJOR 18(3):341–364

Čampelj B, Karnet I, Brodnik A, Jereb E, Rajkovič U (2018) A multi-attribute modelling approach to evaluate the efficient implementation of ICT in schools. Centr Eur J Oper Res 27(3):851–62

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2:429–444

Charnes A, Cooper WW, Rhodes E (1979) Measuring the efficiency of decision-making units. Eur J Oper Res 3(4):339

Cherchye L, De Witte K, Perelman S (2019) A unified productivity-performance approach applied to secondary schools. J Oper Res Soc 70(9):1522–37

Cook WD, Zhu J (2011) Multiple variable proportionality in data envelopment analysis. Oper Res 59(4):1024–1032

Cooper WW, Seiford LM, Tone, K (2006) Introduction to data envelopment analysis and its uses: With DEA-solver software and references. Introduction to data envelopment analysis and its uses: with DEA-solver software and references

Cordero JM, Prior D, Simancas R (2016) A comparison of public and private schools in Spain using robust nonparametric frontier methods. CEJOR 24(3):659–680

Farrell MJ (1957) The measurement of productive efficiency. J R Stat Soc Ser A Gener 120(3):253–290

Ferreira D, Marques R (2020) A step forward on order-a robust nonparametric method: inclusion of weight restrictions, convexity, and non-variable returns to scale. Oper Res Int J 20:1011–1046

Ferreira D, Nunes A, Marques R (2018) Economies of scope in the health sector: the case of Portuguese hospitals. Eur J Oper Res 266(2):716–735

Huang C-W, Chiu Y-H, Ting C-T, Lin C-H (2012) Applying a hybrid DEA model to evaluate the influence of marketing activities to operational efficiency on Taiwan's international tourist hotels. J Oper Res Soc 63(4):549–560

Jablonsky J (2016) Efficiency analysis in multi-period systems: an application to performance evaluation in Czech higher education. CEJOR 24(2):283–296

Jie T, Yan Q, Xu W (2015) A technical note on "A note on integer-valued radial model in DEA." Comput Ind Eng 87:308–310

Kadoić N, Ređep NB, Divjak B (2018) A new method for strategic decision-making in higher education. CEJOR 26(3):611–628

Kazemi Matin R, Emrouznejad A (2011) An integer-valued data envelopment analysis model with bounded outputs. Int Trans Oper Res 18(6):741–749

Kazemi Matin R, Kuosmanen T (2009) Theory of integer-valued data envelopment analysis under alternative returns to scale axioms. Omega 37(5):988–995

Kuosmanen T, Kazemi Matin R (2009) Theory of integer-valued data envelopment analysis. Eur J Oper Res 192(2):658–667

Kuosmanen T, Keshvari A, Kazemi Matin R (2015) Discrete and integer-valued inputs and outputs in data envelopment analysis. In: Zhu J (ed) Handbook on data envelopment analysis. Springer, Berlin

Lozano S, Villa G (2006) Data envelopment analysis of integer-valued inputs and outputs. Comput Oper Res 33(10):3004–3014

Podinovski VV (2004) Bridging the gap between the constant and variable returns-to-scale models: selective proportionality in data envelopment analysis. J Oper Res Soc 55(3):265–276

Podinovski VV (2009) Production technologies based on combined proportionality assumptions. J Prod Anal 32(1):21–26

Podinovski VV, Wan Husain WR (2017) The hybrid returns-to-scale model and its extension by production trade-offs: an application to the efficiency assessment of public universities in Malaysia. Ann Oper Res 250(1):65–84

Podinovski VV, Ismail I, Bouzdine-Chameeva T, Zhang W (2014) Combining the assumptions of variable and constant returns to scale in the efficiency evaluation of secondary schools. Eur J Oper Res 239(2):504–513

Wu L, O'Brien GC (2010) Radial data envelopment analysis models with mixed orientation of input and output. Int Trans Oper Res 17(2):287–302