Omega 71 (2017) 1–10

Contents lists available at ScienceDirect

Omega

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

The determination of the least distance to the strongly efficient frontier in Data Envelopment Analysis oriented models: Modelling and computational aspects ☆

Juan Aparicio [a,*], Jose M. Cordero [b], Jesus T. Pastor [a]

[a] Center of Operations Research (CIO). University Miguel Hernandez of Elche (UMH), 03202 Elche (Alicante), Spain
[b] Department of Economics, University of Extremadura (UEX), 06006 Badajoz, Spain

ARTICLE INFO

Article history:
Received 5 May 2016
Accepted 19 September 2016
Available online 23 September 2016

Keywords:
Data Envelopment Analysis
least distance
oriented models

ABSTRACT

Determining the least distance to the efficient frontier for estimating technical inefficiency, with the consequent determination of closest targets, has been one of the relevant issues in recent Data Envelopment Analysis literature. This new paradigm contrasts with traditional approaches, which yield furthest targets. In this respect, some techniques have been proposed in order to implement the new paradigm. A group of these techniques is based on identifying all the efficient faces of the polyhedral production possibility set and, therefore, is associated with the resolution of a NP-hard problem. In contrast, a second group proposes different models and particular algorithms to solve the problem avoiding the explicit identification of all these faces. These techniques have been applied more or less successfully. Nonetheless, the new paradigm is still unsatisfactory and incomplete to a certain extent. One of these challenges is that related to measuring technical inefficiency in the context of oriented models, i.e., models that aim at changing inputs or outputs but not both. In this paper, we show that existing specific techniques for determining the least distance without identifying explicitly the frontier structure for graph measures, which change inputs and outputs at the same time, do not work for oriented models. Consequently, a new methodology for satisfactorily implementing these situations is proposed. Finally, the new approach is empirically checked by using a recent PISA database consisting of 902 schools.

1. Introduction

Data Envelopment Analysis (DEA) is a non-parametric methodology for estimating technical efficiency of a set of Decision Making Units (DMUs) from a dataset of inputs and outputs. This methodology is fundamentally based on Mathematical Programming and allows a piece-wise linear production frontier enveloping the input–output observations to be determined. Moreover, and as a byproduct of the estimation process, a projection on the frontier and a value of technical inefficiency for each DMU are determined through the calculation of a measure with sense of distance from each unit to the frontier.

A DMU is considered to be technically inefficient if it is possible to expand its output bundle without requiring any increase in its inputs and/or to contract its input bundle without requiring a reduction in its outputs. The potential for augmenting the output bundle reflects output-oriented inefficiency, while potential reduction in inputs means input-oriented inefficiency. In most empirical applications, technical efficiency is measured either in input- or in output-orientation. The selection between one of the two depends on the situation being considered. Additionally, when there is no particular reason to select either the input or output orientation, it is desirable to resort to a technical efficiency measure that includes both input-saving and output-expanding components. These last measures are usually known as graph or non-oriented in contrast to the oriented ones.

Measures in DEA may also be categorized into two groups. The first one yields projection points on the frontier of the technology without considering whether these are dominated in the sense of Pareto or not. In contrast, the second group ensures that the projection points will be non-dominated, following Koopmans' definition of technical efficiency [35]. While for measures belonging to the first category we deal with the concept of weakly efficient frontier, in the second case, the main character is the strongly efficient frontier, which represents a subset of the weakly efficient frontier.

In the case of the graph measures, we note that nowadays there are two clearly different paradigms for estimating technical inefficiency in DEA. On the one hand, we have the traditional measures, which are associated with the determination of demanding

☆ This manuscript was processed by Associate Editor Podinovski.

* Corresponding author. Fax: +34 966658715.
E-mail address: j.aparicio@umh.es (J. Aparicio).

http://dx.doi.org/10.1016/j.omega.2016.09.008
0305-0483/

targets. The targets are in particular the coordinates of the projection point on the frontier and thus represent levels of operation of inputs and outputs that would make the corresponding inefficient DMU perform efficiently. This first philosophy is followed by, for example, the Weighted Additive Models [36], the Range-Adjusted Measure [22] and the Enhanced Russell Graph/Slacks-Based Measure [40,54], where the total technical effort required by a DMU to become technically efficient is maximized instead of minimized, thereby generating the furthest projection points on the frontier. On the other hand, other proposals have suggested determining the closest efficient targets instead, minimizing in some sense the slacks in the corresponding mathematical programming model. The argument behind this idea is that closer targets suggest directions of improvement for the inputs and outputs of the inefficient DMUs that can lead them to efficiency with less technical effort. Regarding this second and more recent approach, all began with Briec's (1998) paper, where the Hölder distance functions were defined in order to determine the least distance from each DMU to the frontier of the production possibility set. This paper gives the go-ahead for the publication of a sequel of related works: Briec and Lemaire [14], Briec and Lesourd [15] and Briec and Leleu [16]. In the same line, Frei and Harker [27] suggested determining projection points by minimizing the Euclidean distance to the strongly efficient frontier. Later, Portela et al. [43] introduced the notion of similarity in DEA as closeness between the values of inputs and outputs of the evaluated DMU and the targets, and proposed determining projection points as similar as possible to the assessed DMU. Additionally, Lozano and Villa [37] introduced a method that determines a sequence of targets to be achieved in successive steps, which converge on the strongly efficient frontier. Aparicio et al. [5] determined closest targets for a set of international airlines applying a new version of the Enhanced Russell Graph/Slacks-Based Measure. More recently, Baek and Lee [9], Amirteimoori and Kordrostami [2] and Aparicio and Pastor [7] have focused on the determination of a weighted Euclidean distance to the strongly efficient frontier, whereas Pastor and Aparicio [41], Ando et al. [3], Aparicio and Pastor [6], Aparicio and Pastor [8], Fukuyama et al. [28,29] and Fukuyama et al. [30] are methodological papers focused on checking the fulfillment of suitable properties by the measures based on the new paradigm.

We need to highlight that, in practice, implementing the approach based on the determination of closest targets is not so easy from a computational point of view. This difficulty is consequence of the complexity of determining the least distance to the frontier of a DEA technology from an interior point (inefficient DMU). This problem is reduced by minimizing a convex function on the complement of a convex set (also called reverse convex set) and it is computationally hard (see, for example, [12]). Hence, nowadays there are mainly two paths for dealing with this problem in the literature. The first one is based on identifying in a first stage all the efficient faces of the efficient frontier in order to determine the minimum distance to the frontier as the minimum of the distances to each of the faces in a multi-stage process. Obviously, this path is related to the resolution of a combinatorial NP-hard (Non-deterministic Polynomial-time hard) problem. The second path corresponds to the approach proposed by Aparicio et al. [5], where Mixed Integer Linear Programming (MILP) is used to determine closest targets without calculating explicitly all the efficient faces. As additional advantages, the Aparicio et al. method allows the least distance to be calculated in one step and the code of standard optimizers to be utilized. The Aparicio et al. method has already been used in the literature for estimating technical inefficiency under carbon emissions for a sample of 20 APEC (Asia-Pacific Economic Cooperation) economies in Wu et al. [57], for benchmarking units in the evaluation of the educational performance of Spanish universities [48], for ranking units through a common set of weights in Ruiz and Sirvent [47] and for determining overall inefficiency and its decomposition in Ruiz and Sirvent [46], avoiding in these all cases determining explicitly all the efficient faces of the piece-wise linear frontier of DEA.

Although the new paradigm has already matured as a trend in the DEA literature, it is still unsatisfactory and incomplete to a certain extent. One of the principle challenges is that related to measure technical inefficiency in the context of oriented models, i.e., models that aim at changing inputs or outputs but not both. Most methodological and empirical papers dealing with least distance and closest targets implement graph measures, seeking potential changes in inputs and outputs at the same time. However, sometimes practitioners work with contexts where only oriented models make sense. The empirical application that we will use at the end of the paper to illustrate the new methodology can serve as example. In this empirical application, the objective is to analyse the efficiency of a set of schools with inputs like the average of the socio-economic status of students in the school, the availability of material resources, the human resources employed by schools, and outputs like the averaged test scores achieved by students belonging to the same school in reading and maths. In this framework, the usual approach assumes that it is not possible or not desirable to change the inputs, at least in the short run, and that the model utilized must be always output-oriented (see [1] and De Witte and Lopez-Torres, 2015, to name just a few).

Unfortunately, there are very few contributions that mix oriented measures and least distance. As far as we are aware, only three papers have dealt with this issue. The first one was Coelli [20] who, in the context of radial measures, suggested a multi-stage method for determining closest targets instead of furthest targets in the well-known second phase of radial models. Later, Cherchye and Van Puyenbroeck [19] defined the deviation between mixes in an input-oriented setting as the angle between the input vector of the assessed DMU and its projection, maximizing the corresponding cosine in order to find the closest targets on the strongly efficient frontier. In the same year, Gonzalez and Alvarez [31] defined a new version of the traditional Russell input efficiency measure [26] based on the minimization of the sum of input contractions required to reach the efficient subset of the production frontier instead of the usual maximization criterion.

Regarding limitations of these three last mentioned papers, it is worth mentioning that Coelli [20] was only created for dealing with the second stage of the radial model and, therefore, the corresponding projection conserves the (input or output) mix in the movements towards the boundary of the production possibility set. However, a well-known drawback of radial measures is the arbitrariness in imposing targets preserving the mix within inputs or within outputs, when the firm's very reason to change its input/output levels might often be the desire to change that mix [17]. As for the contribution of Cherchye and Van Puyenbroeck [19], these authors resorted to the 'combinatorial' methodology associated with the determination of all the faces of the polyhedral DEA technology, which is linked to a NP-hard problem. Finally, the approach by Gonzalez and Alvarez [31] applies an ad-hoc method, defined for a new version of the Russell input efficiency measure, which should generate the closest targets on the strongly efficient frontier. However, we will show in this paper that it is not always true. Consequently, regarding oriented models in the new paradigm, no existing method is sufficiently flexible or interesting from a computational point of view when it comes to tackling the implementation of the problem.

Apart from these methods in the oriented setting, the approach introduced by Aparicio et al. [5], originally defined for graph-type measures, could *a priori* be utilized for oriented models, at least that is what it may seem. However, we will also show that this technique, which works correctly in the case of non-oriented measures, cannot be successfully applied in the case of input or output oriented models.

In view of the preceding discussion, it seems necessary to propose a new and valid solution for determining least distance and closest targets in the context of oriented models. In particular, we will focus our contribution on the identification of the Pareto-efficient projection point that dominates the evaluated unit and, at the same time, produces the least corresponding distance. All these analyses will be carried out in an oriented setting. To do that, we will introduce a Bilevel Linear Programming (BLP) model that will allow us to calculate both the desired closest targets and the minimum distance to the strongly efficient frontier.

The remainder of the paper is organized as follows: In Section 2, we introduce the necessary notation and background. Moreover, we particularly show that neither the approach by Gonzalez and Alvarez [31] performs correctly nor does the methodology proposed for non-oriented contexts by Aparicio et al. [5] work in the oriented setting except for limited cases. Subsequently, in Section 3, we introduce a new methodology, based on Bilevel Linear Programming, in order to be able to determine the Pareto-efficient closest targets and least distance for oriented models in DEA. An empirical illustration of the introduced methodology based on recent PISA data is carried out in Section 4. In Section 5, we present the conclusions.

2. Notation, background and analysis of the literature

In this section, we review the literature on least distance and closest targets in Data Envelopment Analysis, showing some unknown results and limitations of existing approaches in the oriented framework. Nevertheless, before doing that we need to introduce some notation and notions.

Working in the usual DEA context, let us consider n decision making units (DMUs) to be evaluated. DMU_j consumes $x_j = (x_{1j}, \ldots, x_{mj}) \in R^m_+$, $x_j \neq 0_m$, amounts of inputs for the production of $y_j = \left(y_{1j}, \ldots, y_{sj}\right) \in R^s_+$, $y_j \neq 0_s$, amounts of outputs. The relative efficiency of each DMU in the sample is assessed with reference to the so-called production possibility set, which can be non-parametrically constructed from the observations by assuming certain postulates (see [10]). In this way, the production possibility set in DEA, T, can then be mathematically characterized under Constant Returns to Scale (CRS) and Variable Returns to Scale (VRS) as follows:

$$T_{CRS} = \left\{ (x, y) \in R^m_+ \times R^s_+ : x \geq \sum_{j=1}^{n} \lambda_j x_j, y \leq \sum_{j=1}^{n} \lambda_j y_j, \lambda_j \geq 0, j = 1, \ldots, n \right\}. \quad (1)$$

$$T_{VRS} = \left\{ (x, y) \in R^m_+ \times R^s_+ : x \geq \sum_{j=1}^{n} \lambda_j x_j, y \leq \sum_{j=1}^{n} \lambda_j y_j, \sum_{j=1}^{n} \lambda_j = 1, \lambda_j \geq 0, j = 1, \ldots, n \right\}. \quad (2)$$

Below we introduce additional notions related to the production possibility set regardless of the assumed returns to scale, using T instead of T_{CRS} or T_{VRS}. Nevertheless, these notions are also applicable to T_{CRS} or T_{VRS} simply by incorporating the corresponding subscript.

In the production literature, we can find the concept of frontier linked to the notion of technology. Specifically, the weakly efficient frontier of T is defined as $\partial^w(T) := \{(x, y) \in T : \hat{x} < x,\ \hat{y} > y \Rightarrow (\hat{x}, \hat{y}) \notin T\}$. Following Koopmans [35], in order to measure technical efficiency in the Pareto sense, isolating a certain subset of $\partial^w(T)$ is necessary. We are referring to the strongly efficient frontier of T, defined as $\partial^s(T) := \{(x, y) \in T : \hat{x} \leq x, \hat{y} \geq y,\ (\hat{x}, \hat{y}) \neq (x, y) \Rightarrow (\hat{x}, \hat{y}) \notin T\}$. In words, $\partial^s(T)$ is the set of all the Pareto-Koopmans efficient points of T.

Regarding the oriented framework, the two usual approaches are linked to the input and output orientations. Seeking simplicity, hereafter, we will focus our analysis on the output-oriented approach. Nevertheless, a similar analysis could be performed in the case of input orientation. In this way, output-oriented models assume that each DMU is interested in maximizing outputs while using no more than the observed amount of any input. In order to implement this approach, it is useful to introduce the output production set. In this sense, for each input vector, x, let $P(x)$ be the set of feasible (producible) outputs. Formally, $P(x) = \{y : (x, y) \in T\}$. Regarding the strongly efficient frontier of $P(x)$, it is defined as the set of all the Pareto-Koopmans points of $P(x)$, i.e. $\partial^s(P(x)) := \{y \in P(x) : \hat{y} \geq y, \hat{y} \neq y \Rightarrow \hat{y} \notin P(x)\}$, and it is a subset of the weakly efficient frontier of $P(x)$, denoted and defined as $\partial^w(P(x)) := \{y \in P(x) : \hat{y} > y \Rightarrow \hat{y} \notin P(x)\}$.

As in the graph case, and since the definition of $P(x)$ depends on T, we consider two returns to scale for the oriented framework throughout the paper, CRS and VRS and, consequently, we will utilize the following notation where appropriate: $P_{CRS}(x)$, $P_{VRS}(x)$, $\partial^s(P_{CRS}(x))$, $\partial^s(P_{VRS}(x))$, $\partial^w(P_{CRS}(x))$ and $\partial^w(P_{VRS}(x))$.

In order to measure technical inefficiency, there are a lot of models in DEA (see [23]). One of them is the well-known weighted additive model [36], which in the context of determining the graph inefficiency of DMU_0 with data (x_0, y_0) can be formulated under Variable Returns to Scale as follows[1]:

$$\begin{aligned} WA^{\max}(x_0, y_0; w^-, w^+) := Max & \sum_{i=1}^{m} w_i^- s_{i0}^- + \sum_{r=1}^{s} w_r^+ s_{r0}^+ \\ s.t. & \\ & \sum_{j \in E_{VRS}} \lambda_{j0} x_{ij} = x_{i0} - s_{i0}^-, i = 1, \ldots, m \quad (3.1) \\ & \sum_{j \in E_{VRS}} \lambda_{j0} y_{rj} = y_{r0} + s_{r0}^+, r = 1, \ldots, s \quad (3.2), \\ & \sum_{j \in E_{VRS}} \lambda_{j0} = 1, \quad (3.3) \\ & s_{i0}^- \geq 0, i = 1, \ldots, m \quad (3.4) \\ & s_{r0}^+ \geq 0, r = 1, \ldots, s \quad (3.5) \\ & \lambda_{j0} \geq 0, j \in E_{VRS} \quad (3.6), \end{aligned} \quad (3)$$

where $w^- = (w_1^-, \ldots, w_m^-) \in R^m_{++}$ and $w^+ = (w_1^+, \ldots, w_s^+) \in R^s_{++}$ are weights representing the relative importance of unit inputs and unit outputs, and E_{VRS} denotes the set of extreme efficient DMUs in the case of assuming VRS.[2]

The linear dual of model (3) can be written as follows:

$$\begin{aligned} WA^{\max}(x_0, y_0; w^-, w^+) = Min & \sum_{i=1}^{m} v_{i0} x_{i0} - \sum_{r=1}^{s} u_{r0} y_{r0} + \alpha \\ s.t. & \\ & -\sum_{i=1}^{m} v_{i0} x_{ij} + \sum_{r=1}^{s} u_{r0} y_{rj} - \alpha \leq 0, j \in E_{VRS} \quad (4.1) \\ & v_{i0} \geq w_i^-, i = 1, \ldots, m \quad (4.2) \\ & u_{r0} \geq w_r^+, r = 1, \ldots, s \quad (4.3) \end{aligned} \quad (4)$$

Model (3) 'maximizes' a weighted ℓ_1 distance from the DMU_0 to the frontier of the production possibility set, thereby increasing outputs and reducing inputs at the same time. Let $(s_0^{-*}, s_0^{+*}, \lambda_0^*)$ be an optimal solution of model (3), then (x_0^*, y_0^*), defined as $x_{i0}^* = \sum_{j \in E_{VRS}} \lambda_{j0}^* x_{ij}$, $\forall i$, and $y_{r0}^* = \sum_{j \in E_{VRS}} \lambda_{j0}^* y_{rj}$, $\forall r$, denotes the projection point associated with the assessed DMU (x_0, y_0). In this way, the targets are the coordinates of the projection point (x_0^*, y_0^*) and represent levels of operation of inputs and outputs which would make the evaluated unit, if it were technically inefficient, perform efficiently. In the case of the traditional weighted additive model,

[1] Note that (3) comes down to the maximization of a convex function on a convex set. This is a relevant issue in optimization theory that have been extensively studied in previous literature (for instance, see [34,42] and [53]).

[2] The extreme efficient units are the DMUs spanning the efficient faces of the frontier that cannot be expressed as a linear combination of the other DMUs. For a formal definition, see Charnes et al. [18]. In the same way, E_{CRS} denotes the set of extreme efficient DMUs under CRS.

it yields targets that are determined by the 'furthest' efficient projection to the assessed DMU. Additionally, it is well-known that the projection points generated by the weighted additive model are always located onto the strongly efficient frontier $\partial^s(T)$.

In contrast to models that determine the furthest targets, there is a stream of the literature in DEA that defends the opposite, i.e. the projected points on the efficient frontier obtained as such are not a suitable representative projection for the assessed DMU. The research line devoted to determining the closest efficient targets and the least distance to the efficient frontier arose from this philosophy, which was briefly revised in the Introduction. However, the implementation of this approach is not as easy as replacing 'Max' by 'Min' in model (3). As we mentioned in the Introduction, the determination of the least distance and closest targets is a hard task from a computational point of view. This difficulty is consequence of the complexity of determining the least distance to the frontier of a DEA technology from an interior point, since this problem is equivalent to minimizing a convex function on the complement of a convex set.

Nowadays, there are principally two paths for determining closest targets in the DEA literature. The first one is based on identifying all the faces of the efficient frontier of the polyhedral DEA technology in a first stage, determining the minimum distance as the minimum of the distances to each of the faces in a multi-stage process. In this way, this first path is related to a combinatorial NP-hard problem and will not be explored in this paper. The second path corresponds to the approach proposed by Aparicio et al. [5], where the strongly efficient frontier is characterized by linear constraints and binary variables, which consequently allows the closest targets to be determined without calculating explicitly all the efficient faces by resorting to Mixed Integer Linear Programming. Next, we show the main result of Aparicio et al. [5].[3]

Theorem 1. *[5].*

Let $S(x_0, y_0; T_{VRS}) := \partial^s(T_{VRS}) \cap \{(x,y) \in R^m_+ \times R^s_+ : x_i \le x_{i0}, \forall i, y_r \ge y_{r0}, \forall r\}$ be the set of strongly efficient points in T_{VRS} dominating (x_0, y_0) in the sense of Pareto. Then, $(x,y) \in S(x_0, y_0; T_{VRS})$ if and only if $\exists \lambda, s^-, s^+, v, u, \alpha, d, b$ such that $x_i = \sum_{j \in E_{VRS}} \lambda_j x_{ij}, \ \forall i$, $y_r = \sum_{j \in E_{VRS}} \lambda_j y_{rj}, \ \forall r$ and

$$
\begin{array}{lll}
\sum_{j \in E_{VRS}} \lambda_j x_{ij} = x_{i0} - s_i^-, & i = 1,\dots,m & (5.1)\\
\sum_{j \in E_{VRS}} \lambda_j y_{rj} = y_{r0} + s_r^+, & r = 1,\dots,s & (5.2)\\
\sum_{j \in E_{VRS}} \lambda_j = 1, & & (5.3)\\
-\sum_{i=1}^{m} v_i x_{ij} + \sum_{r=1}^{s} u_r y_{rj} - \alpha + d_j = 0, & j \in E_{VRS} & (5.4)\\
v_i \ge w_i^-, & i = 1,\dots,m & (5.5)\\
u_r \ge w_r^+, & r = 1,\dots,s & (5.6)\\
s_i^- \ge 0, & i = 1,\dots,m & (5.7)\\
s_r^+ \ge 0, & r = 1,\dots,s & (5.8)\\
\lambda_j \ge 0, & j \in E_{VRS} & (5.9)\\
d_j \le M b_j, & j \in E_{VRS} & (5.10)\\
\lambda_j \le 1 - b_j, & j \in E_{VRS} & (5.11)\\
d_j \ge 0, & j \in E_{VRS} & (5.12)\\
b_j \in \{0,1\}, & j \in E_{VRS} & (5.13)
\end{array}
\qquad (5)
$$

where M is a sufficiently big positive number.

Note that their result combines the constraints of programs (3) and (4) in (5). Indeed, (5.1)–(5.9) coincide with (3.1)–(3.6) and (4.1)–(4.3). Whereas the new constraints, (5.10)–(5.13), are the key to suitably mixing all the aforementioned restrictions, resorting to a set of $card(E_{VRS})$ binary variables b_j.

Invoking Theorem 1, we may formulate a new version of the weighted additive model, which seeks to determine the least distance and closest targets, based on Mixed Integer Linear Programming. In its compact format, it would be expressed as:

$$
WA^{min}_{VRS}(x_0, y_0; w^-, w^+) := Min\left\{ \sum_{i=1}^{m} w_i^- s_{i0}^- + \sum_{r=1}^{s} w_r^+ s_{r0}^+ : (x_0 - s_0^-, y_0 + s_0^+) \in S(x_0, y_0; T_{VRS}) \right\} \qquad (6)
$$

We now turn to the output-oriented framework and try to show what happens when Aparicio et al.'s result is applied. In this sense, the first approach to the problem is to adapt (6) for working in the oriented context through the following model:

$$
WA^{min}_{VRS,O}(x_0, y_0; w^+) := Min\left\{ \sum_{r=1}^{s} w_r^+ s_{r0}^+ : y_0 + s_0^+ \in S(y_0; P_{VRS}(x_0)) \right\}, \qquad (7)
$$

where $S(y_0; P_{VRS}(x_0)) = \partial^s(P_{VRS}(x_0)) \cap \{y \in R^s_+ : y_r \ge y_{r0}, \forall r\}$. In words, $S(y_0; P_{VRS}(x_0))$ denotes the set of strongly efficient points in $P_{VRS}(x_0)$ dominating y_0 in the sense of Pareto.

The first question that arises when one wants to apply Theorem 1 in the oriented context is: $S(y_0; P_{VRS}(x_0)) = \{y \in R^s_+ : (x_0, y) \in S(x_0, y_0; T_{VRS})\}$? If the answer were yes, then (7) could be easily expressed as a MILP model through the theorem. Unfortunately, the answer to the question is negative in general as we next illustrate by means of a simple counterexample.

Counterexample 1. Let A=(1,1) and B=(2,0.5) be two DMUs that consume one input to produce one output under Variable Returns to Scale. For this example, $\partial^s(T_{VRS}) = E_{VRS} = \{A\}$. We now want to evaluate the performance of DMU B through the output-oriented approach. This means that $(x_0, y_0) = (2, 0.5)$. In this way, we have that $y = 1 \in S(0.5; P_{VRS}(2))$ since, in this example, $\partial^s(P_{VRS}(2)) = \{1\}$. However, $(2,1) \notin \partial^s(T_{VRS})$ and, therefore, $(2,1) \notin S(2, 0.5; T_{VRS})$.

Our second approach to the problem is to derive a result similar to Theorem 1 for the oriented case by analogy with the steps followed for the graph framework. The idea is to introduce both the primal and dual of the output-oriented weighted additive model (see, for example, [32], or [44]) and to combine the corresponding constraints. Under Variable Returns to Scale, these models are the following:

$$
\begin{array}{lll}
WA^{max}_{VRS,O}(x_0, y_0; w^+) := Max \sum_{r=1}^{s} w_r^+ s_{r0}^+ & & \\
s.t. & & \\
\quad \sum_{j \in E_{VRS}} \lambda_{j0} x_{ij} \le x_{i0}, i = 1,\dots,m & & (8.1)\\
\quad \sum_{j \in E_{VRS}} \lambda_{j0} y_{rj} = y_{r0} + s_{r0}^+, r = 1,\dots,s & & (8.2)\\
\quad \sum_{j \in E_{VRS}} \lambda_{j0} = 1, & & (8.3)\\
\quad s_{r0}^+ \ge 0, \quad r = 1,\dots,s & & (8.4)\\
\quad \lambda_{j0} \ge 0, \quad j \in E_{VRS} & & (8.5),
\end{array}
\qquad (8)
$$

$$
\begin{array}{ll}
WA^{max}_{VRS,O}(x_0, y_0; w^+) = Min \sum_{i=1}^{m} v_{i0} x_{i0} - \sum_{r=1}^{s} u_{r0} y_{r0} + \alpha & \\
s.t. & \\
\quad -\sum_{i=1}^{m} v_{i0} x_{ij} + \sum_{r=1}^{s} u_{r0} y_{rj} - \alpha \le 0, j \in E_{VRS} & (9.1)
\end{array}
$$

[3] In Aparicio et al. [5] $w_i^- = 1, \ \forall i$ and $w_r^+ = 1, \ \forall r$ and the assumed returns to scale was CRS. Nevertheless, the adaptation of their result to our context is trivial.

$$v_{i0} \geq 0, i = 1, \ldots, m \tag{9.2}$$

$$u_{r0} \geq w_r^+, r = 1, \ldots, s \tag{9.3}$$

(9)

If the non-oriented and the oriented models are compared, then we observe that the input slacks of model (3) are missing in model (8), the equality constraint (3.1) has been transformed into an inequality and, finally, (4.2) has been converted into $v_{i0} \geq 0, \forall i$.

Now, by mixing the constraints of (8) and (9) and adding (5.10)–(5.13), we get the definition of the following set:

$$\hat{S}(y_0; P_{VRS}(x_0)) = \left\{ y \in R^s_+ : \begin{array}{lll} y_r = \sum_{j \in E_{VRS}} \lambda_j y_{rj}, & r = 1, \ldots, s & (10.1) \\ \sum_{j \in E_{VRS}} \lambda_j x_{ij} \leq x_{i0}, & i = 1, \ldots, m & (10.2) \\ \sum_{j \in E_{VRS}} \lambda_j y_{rj} = y_{r0} + s_r^+, & r = 1, \ldots, s & (10.3) \\ \sum_{j \in E_{VRS}} \lambda_j = 1, & & (10.4) \\ -\sum_{i=1}^{m} v_i x_{ij} + \sum_{r=1}^{s} u_r y_{rj} - \alpha + d_j = 0, & j \in E_{VRS} & (10.5) \\ v_i \geq 0, & i = 1, \ldots, m & (10.6) \\ u_r \geq w_r^+, & r = 1, \ldots, s & (10.7) \\ s_r^+ \geq 0, & r = 1, \ldots, s & (10.8) \\ \lambda_j \geq 0, & j \in E_{VRS} & (10.9) \\ d_j \leq M b_j, & j \in E_{VRS} & (10.10) \\ \lambda_j \leq 1 - b_j, & j \in E_{VRS} & (10.11) \\ d_j \geq 0, & j \in E_{VRS} & (10.12) \\ b_j \in \{0, 1\}, & j \in E_{VRS} & (10.13) \end{array} \right\} \tag{10}$$

$$\hat{S}^{=}(y_0; P_{CRS}(x_0)) := \left\{ y \in R^s_+ : \begin{array}{lll} y_r = \sum_{j \in E_{CRS}} \lambda_j y_{rj}, & r = 1, \ldots, s & (11.1) \\ \sum_{j \in E_{CRS}} \lambda_j x_j = x_0, & & (11.2) \\ \sum_{j \in E_{CRS}} \lambda_j y_{rj} = y_{r0} + s_r^+, & r = 1, \ldots, s & (11.3) \\ -v x_j + \sum_{r=1}^{s} u_r y_{rj} + d_j = 0, & j \in E_{CRS} & (11.4) \\ v \geq 0, & & (11.5) \\ u_r \geq w_r^+, & r = 1, \ldots, s & (11.6) \\ s_r^+ \geq 0, & r = 1, \ldots, s & (11.7) \\ \lambda_j \geq 0, & j \in E_{CRS} & (11.8) \\ d_j \leq M b_j, & j \in E_{CRS} & (11.9) \\ \lambda_j \leq M(1 - b_j), & j \in E_{CRS} & (11.10) \\ d_j \geq 0, & j \in E_{CRS} & (11.11) \\ b_j \in \{0, 1\}, & j \in E_{CRS} & (11.12) \end{array} \right\} \tag{11}$$

Then the question is: $S(y_0; P_{VRS}(x_0)) = \hat{S}(y_0; P_{VRS}(x_0))$? Unfortunately, as we show below through a counterexample, the answer is once again negative in general.

Counterexample 2. Let A=(5;1,7), B=(1;5,1), C=(4;4,5) and D=(4;3,4) be four DMUs that consume one input to produce two outputs under Variable Returns to Scale. For this example, $E_{VRS} = \{A, B, C\}$. Let $\lambda_A = \lambda_B = 1/2$, $s_1^+ = s_2^+ = 0$, $v = 5$, $u_1 = 1$, $u_2 = 4$, $\alpha = 4$, $d_A = d_B = d_C = 0$ and $b_A = b_B = b_C = 0$. Then, $(y_1, y_2) = \left(\sum_{j \in E_{VRS}} \lambda_j y_{rj}, \sum_{j \in E_{VRS}} \lambda_j y_{rj} \right) = (3, 4) \in \hat{S}((3, 4); P_{VRS}(4))$. However, since C=(4;4,5) has been observed, $(y_1, y_2) = (4, 5) \in P_{VRS}(4)$. In this way, $(4, 5)$ dominates $(3, 4)$ in the sense of Pareto. Therefore, $(3, 4) \notin S((3, 4); P_{VRS}(4))$ and $S((3, 4); P_{VRS}(4)) \neq \hat{S}((3, 4); P_{VRS}(4))$ in this example.

In the case of assuming Constant Returns to Scale, numerical examples on $S(y_0; P_{CRS}(x_0)) \neq \{y \in R^s_+ : (x_0, y) \in S(x_0, y_0; T_{CRS})\}$ and $S(y_0; P_{CRS}(x_0)) \neq \hat{S}(y_0; P_{CRS}(x_0))$ can be also provided. Nevertheless, seeking simplicity, we do not show them explicitly in this paper. Anyway, it is worth mentioning that the conical nature of the frontier of the CRS technology allows to prove that $S(y_0; P_{CRS}(x_0))$ coincides with a similar set to $\hat{S}(y_0; P_{CRS}(x_0))$ in a very restrictive context: when the production possibility set is generated from an only input, i.e. $m = 1$. The next proposition establishes this result. Nonetheless, we need first to introduce three related lemmas.

Lemma 1. *[22]. $y \in \partial^s(P_{CRS}(x_0))$ if and only if $WA^{\max}_{CRS,O}(x_0, y; w^+) = 0$.*

Lemma 2. *[22]. Let (s^{+*}, λ^*) be an optimal solution of (8) under CRS.[4] Then, $y_0^* := \left(\sum_{j \in E_{CRS}} \lambda_j^* y_{1j}, \ldots, \sum_{j \in E_{CRS}} \lambda_j^* y_{sj} \right) \in \partial^s(P_{CRS}(x_0))$.*

Lemma 3. *Let $m = 1$ and (s^{+*}, λ^*) be an optimal solution of (8) under CRS. Then, $\sum_{j \in E_{CRS}} \lambda_{j0}^* x_j = x_0$.*

Proof. See Appendix.

In order to prove the desired result, we need to adapt expression (10) to Constant Returns to Scale, deleting (10.4) and α, adding M to (10.10) and considering $m = 1$. Additionally, it is necessary to slightly modify (10.2), transforming the inequality into an equality:

Proposition 1. Let $m = 1$. Then $S(y_0; P_{CRS}(x_0)) = \hat{S}^{=}(y_0; P_{CRS}(x_0))$.

Proof. See Appendix.

From all the above discussion, we conclude that the approach based on the Aparicio et al. theorem does not work in general terms for the oriented framework, except for the restrictive case of assuming Constant Returns to Scale and the existence of a unique input.

In the context of measuring technical efficiency through an input-oriented model, Gonzalez and Alvarez [31] suggested minimizing the sum of all the input-specific contractions in order to reach the strongly efficient frontier, a proposal that is mathematically equivalent to maximizing the well-known Russell input measure, instead of

[4] This means that in (8) the constraint (8.3) is omitted.

minimizing it as usual. This approach is, therefore, related to the determination of closest targets under the Pareto-Koopmans criterion of technical efficiency. To implement their new model, Gonzalez and Alvarez introduced a multistage process based on the solution, in the first stage, of m linear models, each of them providing the k th input-specific contraction. In the second stage, the desired value of the new version of the Russell input measure is obtained as the minimum of all the input-specific contraction determined previously (see [31], Proposition 1, p. 517). Unfortunately, this algorithm does not always lead to the correct solution.

To illustrate the above claim, we first need to introduce the linear model that is solved for each input k in the first stage of the Gonzalez and Alvarez process:

$$
\begin{aligned}
& Min \quad M\theta_k^k + \sum_{i \neq k}^{m} \theta_i^k \\
& s.t. \\
& \qquad \sum_{j=1}^{n} \lambda_j^k x_{ij} \le \theta_i^k x_{i0}, \quad i = 1, \dots, m \\
& \qquad \sum_{j=1}^{n} \lambda_j^k y_{rj} \ge y_{r0}, \qquad r = 1, \dots, s \\
& \qquad \sum_{j=1}^{n} \lambda_j^k = 1, \\
& \qquad \lambda_j^k \ge 0, \qquad\qquad j = 1, \dots, n \\
& \qquad \theta_i^k \le 1, \qquad\qquad i \neq k
\end{aligned}
\tag{12}
$$

where M is a sufficiently big positive number.

Then, from an optimal solution of (12), $\left(\lambda^{k*}, \theta_1^{k*}, \dots, \theta_m^{k*}\right)$, the input-specific contraction for input k can be computed as $C(x_0, y_0)_k = \sum_{i=1}^{m} \left(1 - \theta_i^{k*}\right)$. Finally, the input-oriented Russell measure associated with the least distance and closest targets is determined following the Gonzalez and Alvarez approach as $C(x_0, y_0) = \min\{C(x_0, y_0)_k : k = 1, \dots, m\}$. Next we show that this is not always true through a numerical example.[5]

Counterexample 3. Let us assume that we have observed five DMUs that produce one output from the consumption of three inputs (see Table 1). Considering $M = 100,000$, and the evaluation of the performance of unit E, we obtain, applying (12), that $C(x_E, y_E)_1 = 2/3$, $C(x_E, y_E)_2 = 2/3$ and $C(x_E, y_E)_3 = 0.43$. In this way, we conclude that $C(x_E, y_E) = 0.43$. However, unit C produces the same quantity of output and dominates unit E in the sense of Pareto. If we use the inputs of unit C for evaluating unit E, we get $\tilde{\theta}_1 = \frac{2.8}{3}$, $\tilde{\theta}_2 = \frac{2.2}{3}$ and $\tilde{\theta}_3 = \frac{4.6}{5}$. This leads to the following sum of input contractions: $\sum_{i=1}^{3} \left(1 - \tilde{\theta}_i\right) = 0.41 < 0.43$. Consequently, the smallest contraction to the strongly efficient frontier does not always coincide with the smallest $C(x_0, y_0)_k$, $k = 1, \dots, m$.

In summary, as we are aware, none of the existing approaches allows the determination of the closest Pareto-efficient targets in the oriented framework to be dealt with in a suitable way. In the next section, we will propose a solution to this problem. In particular, we will introduce a new methodology based on Bilevel Linear Programming.

[5] If we focus our attention on the weakly efficient frontier instead of the strongly efficient one, then it is possible to show that the Gonzalez and Alvarez [31] approach also presents some drawbacks using the same counterexample. Following Briec [13], the correct way of calculating the smallest contraction to the weakly efficient frontier would consist in solving m models like (12) where the objective function is substituted by "Min θ_k^k".

Table 1
Data for Counterexample 3.

DMU	Input 1	Input 2	Input 3	Output
A	3	1	5	1
B	1	3	5	1
C	2.8	2.2	4.6	1
D	2.2	2.8	4.5	1
E	3	3	5	1

3. A solution based on Bilevel Linear Programming

In this section, we first briefly review the mainly notions related to Bilevel Programming in order to introduce, in the second part of the section, an approach on these grounds to determine the closest targets and the least distance through oriented models in DEA.

A Bilevel Programming model refers to a mathematical programming problem where one of the constraints is an optimization problem. This theory has been successfully applied to model different real situations with a common feature: the existence of a hierarchical structure (see [55]). A Bilevel Programming problem where both the objective functions and the constraints are linear is called a Bilevel Linear Programming problem. Denote by $z \in Z \subset R^p$ and $t \in T \subset R^q$ the decision variables corresponding to the first and second level, respectively. The general formulation of a Bilevel Linear Programming (BLP) problem is as follows:

$$
\begin{aligned}
& \underset{z,t}{Min} \quad c_1 z + d_1 t \\
& \quad s.t. \\
& \qquad A_1 z + B_1 t \le b_1, \\
& \qquad \underset{t}{Min} \quad c_2 z + d_2 t \\
& \qquad s.t. \\
& \qquad\qquad A_2 z + B_2 t \le b_2, \\
& \qquad z \ge 0, \ t \ge 0
\end{aligned}
\tag{13}
$$

Program (13) consists of two subproblems. On the one hand, the higher level decision problem and, the other hand, the lower level decision problem, which appears as a constraint in (13). Both problems are connected in a way that the higher problem sets parameters influencing the lower level problem and the higher problem, in turn, is affected by the outcome of the lower level problem.

It is known that even for the Bilevel Programming problem where all the functions are linear, like in (13), the model to be solved is non-convex and NP-hard [21]. This complexity is the reason why many different techniques have been proposed in the literature to study the computational aspects of Bilevel Programming problems. The formulation of optimality conditions for this type of problems usually starts with a suitable reformulation of the problem as a one-level model. First conditions are based on replacing the lower level problem by an implicitly determined function [45]. A second possibility is to transform the original problem into a single optimization problem by applying the well-known Karush–Kuhn–Tucker (KKT) optimality conditions of the lower level problem [50]. Other alternative options might be the application of specific branch and bound algorithms [52], the Kth-best approach [51] by enumerating the extreme points of the feasible region when all the constraints are linear, using heuristic algorithms [4] or even neural network techniques [33].

Regarding the solutions of a BLP problem, $(z^*, t^*) \ge 0$ is a feasible solution of (13) if t^* is an optimal solution of the lower level program with $z = z^*$ and, at the same time, $A_1 z^* + B_1 t^* \le b_1$. In this way, (z^*, t^*) is an optimal solution if additionally$c_1 z^* + d_1 t^* \le c_1 z + d_1 t$ for all feasible solution (z, t) of (13), being $c_1 z^* + d_1 t^*$ the corresponding optimal value of the BLP problem.

Now we are ready to introduce the model that permits the closest Pareto-efficient targets in the output-oriented case to be determined.

The input-oriented case could be derived by analogy. The key idea is to exploit the hierarchical structure of the BLP problems, using the measure that needs to be determined as the higher level problem and the lower level problem being the output-oriented weighted additive model that, by Lemma 1, is able to characterize the belonging to the strongly efficient frontier in the oriented case by its optimal value. Let us assume that we are interested in determining the Russell output measure under the least distance criterion. In this case, the model to be solved is the following:

$$
\begin{array}{lll}
Min \ \frac{1}{s}\sum_{r=1}^{s}\phi_r & & (14.1)\\
s.t. & & \\
\quad \sum_{j\in E_{VRS}} \lambda_j x_{ij} \le x_{i0}, & i=1,\dots,m & (14.2)\\
\quad \sum_{j\in E_{VRS}} \lambda_j y_{rj} \ge \phi_r y_{r0}, & r=1,\dots,s & (14.3)\\
\quad \sum_{j\in E_{VRS}} \lambda_j = 1, & & (14.4)\\
\quad \sum_{r=1}^{s} w_r^+ s_r^+ = 0, & & (14.5)\\
\quad Max \ \sum_{r=1}^{s} w_r^+ s_r^+ & & (14.6)\\
\quad s.t. & & \\
\qquad \sum_{j\in E_{VRS}} \gamma_j x_{ij} \le x_{i0}, & i=1,\dots,m & (14.7)\\
\qquad \sum_{j\in E_{VRS}} \gamma_j y_{rj} = \phi_r y_{r0} + s_r^+, & r=1,\dots,s & (14.8)\\
\qquad \sum_{j\in E_{VRS}} \gamma_j = 1, & & (14.9)\\
\quad \phi_r \ge 1,\ \lambda_j,\ s_r^+,\ \gamma_j \ge 0, & \forall r,j & (14.10)
\end{array}
\qquad (14)
$$

In (14) the higher level problem coincides with the Russell output measure except for the fact that the objective function is minimized instead of maximized as happens with the traditional definition of the Russell output measure ([26], p. 149), while the lower level problem matches (8) when the evaluated output vector is $(\phi_1 y_{10}, \dots, \phi_s y_{s0})$.

The next proposition states that the optimal value of (14) equals the Russell output measure under the philosophy of seeking the least distance to the strongly efficient frontier of the production possibility set.

Proposition 2. Let $(\phi^*, \lambda^*, s^{+*}, \gamma^*)$ be an optimal solution of (14). Then, $\frac{1}{s}\sum_{r=1}^{s}\phi_r^* = Min\left\{\frac{1}{s}\sum_{r=1}^{s}\phi_r : (\phi_1 y_{10}, \dots, \phi_s y_{s0}) \in S(y_0; P_{VRS}(x_0))\right\}$.

Proof. See Appendix.

Regarding the weights w_r^+ in (14), we will assume from now on that $w_r^+ = 1$, $r = 1, \dots, m$.

As for the implementation of the BLP problem, we will particularly use the KKT optimality conditions in order to solve (14). Accordingly, (14.6)–(14.9) must be substituted by (15.1)–(15.9).

$$
\begin{array}{lll}
\sum_{j\in E_{VRS}} \gamma_j x_{ij} + l_i = x_{i0}, & i=1,\dots,m & (15.1)\\
\sum_{j\in E_{VRS}} \gamma_j y_{rj} = \phi_r y_{r0} + s_r^+, & r=1,\dots,s & (15.2)\\
\sum_{j\in E_{VRS}} \gamma_j = 1, & & (15.3)\\
-\sum_{i=1}^{m}\eta_i x_{ij} + \sum_{r=1}^{s}\mu_r y_{rj} + \psi_j + \tau_j = 0, & j\in E_{VRS} & (15.4)\\
-\eta_i + e_i = 0, & i=1,\dots,m & (15.5)\\
\mu_r \ge 1, & r=1,\dots,s & (15.6)\\
\gamma_j \tau_j = 0, & j\in E_{VRS} & (15.7)\\
l_i e_i = 0, & i=1,\dots,m & (15.8)\\
l_i, \eta_i, \mu_r, \tau_j, e_i \ge 0, & \forall i,j & (15.9)
\end{array}
\qquad (15)
$$

Constraints (15.7)–(15.8) are not linear. Nevertheless, restrictions of this nature are not difficult to be implemented by means of a Special Ordered Set (SOS)[6] [11].

Finally, before applying model (14) to a real database in the next section, it is important to clarify some questions regarding the general setting when we are interested in computing a technical efficiency measure that includes both input-saving and output-expansion (graph measures). As we pointed out in Section 2, the approach by Aparicio et al. [5] works well in the general framework, but it fails in the case of dealing with oriented models. However, the new methodology described in this paper is valid in any context. Therefore, if we want to evaluate the technical inefficiency of a set of DMUs using the weighted additive model (3), but determining the closest targets instead of the usual furthest targets that are generated by this model, we should solve the following model:

$$
\begin{array}{lll}
Min \ \sum_{i=1}^{m} w_i^- s_{i0}^- + \sum_{r=1}^{s} w_r^+ s_{r0}^+ & & (16.1)\\
s.t. & & \\
\quad \sum_{j\in E_{VRS}} \lambda_{j0} x_{ij} = x_{i0} - s_{i0}^-, & i=1,\dots,m & (16.2)\\
\quad \sum_{j\in E_{VRS}} \lambda_{j0} y_{rj} = y_{r0} + s_{r0}^+, & r=1,\dots,s & (16.3)\\
\quad \sum_{j\in E_{VRS}} \lambda_{j0} = 1, & & (16.4)\\
\quad \sum_{i=1}^{m}\tau_{i0}^- + \sum_{r=1}^{s}\tau_{r0}^+ = 0, & & (16.5)\\
\quad Max \ \sum_{i=1}^{m}\tau_{i0}^- + \sum_{r=1}^{s}\tau_{r0}^+ & & (16.6)\\
\quad s.t. & & \\
\qquad \sum_{j\in E_{VRS}} \gamma_{j0} x_{ij} = x_{i0} - s_{i0}^- - \tau_{i0}^-, & i=1,\dots,m & (16.7)\\
\qquad \sum_{j\in E_{VRS}} \gamma_{j0} y_{rj} = y_{r0} + s_{r0}^+ + \tau_{r0}^+, & r=1,\dots,s & (16.8)\\
\qquad \sum_{j\in E_{VRS}} \gamma_{j0} = 1, & & (16.9)\\
\quad s_{i0}^-, s_{r0}^+, \lambda_{j0}, \tau_{i0}^-,\ \tau_{r0}^+,\ \gamma_{j0} \ge 0, & \forall r,j & (16.10)
\end{array}
\qquad (16)
$$

4. Empirical illustration: efficiency of schools using PISA data

This section includes an empirical illustration with real data applying the methodology proposed in this paper. In particular, we use Spanish data from the PISA (Programme for International Student Assessment) 2012 survey, where data from student and school questionnaires (with students and school level information, respectively) were merged. This dataset provides results on the performance of 15 year-old students in different competences as well as other factors potentially related to those results such as variables representing student background, school environment or educational provision.

Following the well-established literature on school efficiency (e.g. [1,24,25,49]), we select the results from a standardized test as educational outputs and three usual inputs in education production functions such as the students (raw material), infrastructures (school resources) and teachers (human capital). Table 2 reports the descriptive statistics for these five variables considering the total number of schools (902) included in the sample. A detailed explanation of the specific indicators considered in the empirical analysis is provided below.

[6] SOS is a way to specify that a pair of variables cannot take strictly positive values at the same time and is a technique related to using special branching strategies. Traditionally, SOS was used with discrete and integer variables, but modern optimizers, like for example CPLEX, use also SOS with continuous variables.

Table 2
Descriptive statistics.

Variables	Type of variable	Mean	SD	Min	Max
ESCS	Input	4.396	0.538	2.190	5.970
SCMATEDU	Input	3.683	0.891	0.008	5.576
TEACHERS	Input	10.14	6.309	0.719	90.01
PVMATH	Output	490.9	46.17	130.1	609.6
PVREAD	Output	490.4	46.66	297.7	626.6

Table 3
Results of model (14) for 902 schools.

Model (14)	Mean	SD	Min	Max
Score of efficiency	1.122	0.084	1.000	1.672
Time of execution (s)	10.782	4.854	1	42

– As a proxy for the quality of students in the school, we use the average of the socio-economic status of students in the school, represented by the ESCS index, which provides a measure of family background that includes the highest levels of parents´ occupation, educational resources and cultural possessions at home. Since the original values of this variable presented positive and negative values, all of them were rescaled to show positive values.
– As a proxy for the availability of material resources, we use an index created by PISA analysts (SCMATEDU) from the responses given by school principals regarding several educational resources such as computers, educational software, calculators, books, audio visual resources or laboratory equipment. In this case, we have also rescaled the original values to assure that all values are positive.
– The inverse of the student-teacher ratio, i.e., the number of teachers per (hundred) students (TEACHERS), as a proxy for human resources employed by schools.
– The output variables are represented by the averaged test scores achieved by students belonging to the same school in reading and maths. Regarding this point, it is worth noting that PISA reports five plausible values randomly drawn from the estimated distribution of results for each student according to their answers to the questions in the test (see [39] for details). Those plausible values can be interpreted as a measure of their performance in order to approximate the real distribution of the latent variable being measured (cognitive skills) [38,56].

Table 3 shows a summary of the results obtained with the approach proposed in this paper, model (14). The mean of the technical efficiency of the Spanish sample is 1.122 (1.135 in reading and 1.109 in maths), which means that, on average, the schools could increase their outputs levels by 12%, needing a greater effort in the reading dimension, without changing their resources. With regard to the resolution of the 902 optimization programs, we used CPLEX to solve the different problems and code in C on a CPU AMD Phenom II X6 1075 T (hexa-core) with 3 GHz and 16 RAM GB. In this respect, the average time of execution was 10.782 seconds, i.e., a total amount of around 3 hours.

5. Conclusions

In this paper, we have shown that all the existing approaches to determine the closest Pareto-efficient targets in DEA present some weaknesses when they are applied or adapted to the oriented framework, when the interest of the firm/organization is to expand its output bundle without requiring any increase in its inputs or to contract its input bundle without requiring a reduction in its outputs.

To deal with this problem in a suitable way, a new methodology based upon Bilevel Linear Programming was introduced to determine the desired targets in the case of using a new version of the Russell oriented measure. Its implementation is grounded on the application of the Karush-Kuhn-Tucker (KKT) optimality conditions to the lower level problem and Special Ordered Sets (SOS).

Finally, the new approach was illustrated through an empirical analysis using data on the 902 Spanish schools participating in PISA 2012. The results show that there is room for improvement, especially in reading (one of the outputs selected). Likewise, the computation time is relatively low considering the size of the available sample.

Acknowledgements

We thank two anonymous referees for providing constructive comments and help in improving the contents and presentation of this paper. Additionally, Juan Aparicio and Jesus T. Pastor would like to express their gratitude to the Spanish Ministry for Economy and Competitiveness for supporting this research through grant MTM2013-43903-P and Jose M. Cordero also acknowledges the support from the same institution through grant ECO2014-53702-P.

Appendix

Proof of Lemma 3. Let us assume that $\sum_{j \in E_{CRS}} \lambda_{j0}^* x_j < x_0$. Let us define $x_0^* = \sum_{j \in E_{CRS}} \lambda_{j0}^* x_j$. In this way, by (1), $(x_0^*, y_0^*) \in T_{CRS}$. Moreover, note that $\exists j \in E_{CRS}$ such that $\lambda_{j0}^* > 0$, since otherwise constraint (8.2) would be violated for $y_0 \in R_+^s$, $y_0 \neq 0_s$. Consequently, $x_0^* = \sum_{j \in E_{CRS}} \lambda_{j0}^* x_j > 0$ since $x_j \in R_+^m$, $x_j \neq 0_m$, and $m = 1$ for all $j = 1, ..., n$. By the definition of x_0^*, $x_0^* < x_0 \Leftrightarrow \frac{x_0}{x_0^*} > 1$. This implies that, under CRS, $\frac{x_0}{x_0^*}(x_0^*, y_0^*) = \left(x_0, \frac{x_0}{x_0^*} y_0^*\right) \in T_{CRS}$. Therefore, $\frac{x_0}{x_0^*} y_0^* \in P_{CRS}(x_0)$. Additionally, $y_0^* \neq 0_s$ since, by definition, $y_0^* = \left(\sum_{j \in E_{CRS}} \lambda_j^* y_{1j}, \ldots, \sum_{j \in E_{CRS}} \lambda_j^* y_{sj}\right)$, which must be equal or greater than y_0 by (8.2) under CRS, which satisfies $y_0 \in R_+^s$, $y_0 \neq 0_s$. In this way, $\frac{x_0}{x_0^*} y_{r0}^* \geq y_{r0}^*$, $\forall r = 1, \ldots, s$, with some $r' = 1, \ldots, s$ satisfying $\frac{x_0}{x_0^*} y_{r0}^* > y_{r0}^*$. Thus, $y_0^* \notin \partial^s(P_{CRS}(x_0))$, which is a contradiction to Lemma 2. So, by (8.1), $\sum_{j \in E_{CRS}} \lambda_{j0}^* x_j = x_0$. ■

Proof of Proposition 1. (i) Let $\left(\overline{\lambda}, \overline{s}^+, \overline{v}, \overline{u}, \overline{d}, \overline{b}\right)$ be a vector that satisfies constraints 11.2-11.12. From this solution, it is possible to generate $\overline{y} \in \hat{S}^{=}(y_0; P_{CRS}(x_0))$ as $\overline{y}_r = \sum_{j \in E_{CRS}} \overline{\lambda}_j y_{rj}$, $r = 1, \ldots, s$. Let us note that, by (1), $(\overline{x}, \overline{y}) \in T_{CRS}$, with $\overline{x} := \sum_{j \in E_{CRS}} \overline{\lambda}_j x_j = x_0$, and, consequently, $\overline{y} \in P_{CRS}(\overline{x}) = P_{CRS}(x_0)$. Let us now assume that $\exists y' \in P_{CRS}(x_0)$ such that $y' \geq \overline{y}$ with $y' \neq \overline{y}$. We want to arrive at a contradiction. By (11.4) and (11.11), we have that $-\overline{v} x_j + \sum_{r=1}^{s} \overline{u}_r y_{rj} \leq 0$ for all $j \in E_{CRS}$. And, by the definition of T_{CRS} in (1), we know that $-\overline{v} x + \sum_{r=1}^{s} \overline{u}_r y_r \leq 0$ for all $(x, y) \in T_{CRS}$. This implies that $-\overline{v} x_0 + \sum_{r=1}^{s} \overline{u}_r y_r \leq 0$ for all $(x_0, y) \in T_{CRS}$. In this way, $-\overline{v} x_0 + \sum_{r=1}^{s} \overline{u}_r y'_r \leq 0$ since

$(x_0, y') \in T_{CRS}$ because $y' \in P_{CRS}(x_0)$. Additionally, $-\overline{v}x_0 + \sum_{r=1}^{s} \overline{u}_r \overline{y}_r = -\overline{vx} + \sum_{r=1}^{s} \overline{u}_r \overline{y}_r = -\overline{v}\left(\sum_{j \in E_{CRS}} \overline{\lambda}_j x_j\right) + \sum_{r=1}^{s} \overline{u}_r \left(\sum_{j \in E_{CRS}} \overline{\lambda}_j y_{rj}\right) = \sum_{j \in E_{CRS}} \overline{\lambda}_j \left(-\overline{v}x_j + \sum_{r=1}^{s} \overline{u}_r y_{rj}\right) = -\sum_{j \in E_{CRS}} \overline{\lambda}_j \overline{d}_j$ by (11.4). Now, by (11.8)-(11.12), $\overline{\lambda}_j \overline{d}_j = 0$, $\forall j \in E_{CRS}$, and, therefore, $-\overline{v}x_0 + \sum_{r=1}^{s} \overline{u}_r \overline{y}_r = 0$. Finally, we have that $-\overline{v}x_0 + \sum_{r=1}^{s} \overline{u}_r y'_r > -\overline{v}x_0 + \sum_{r=1}^{s} \overline{u}_r \overline{y}_r = 0$ since $\overline{u}_r > 0$ (11.6) and $y' \geq \overline{y}$ with $y' \neq \overline{y}$. Then, we proved that, on the one hand, $-\overline{v} x_0 + \sum_{r=1}^{s} \overline{u}_r y'_r \leq 0$ and, on the other hand, $-\overline{v}x_0 + \sum_{r=1}^{s} \overline{u}_r y'_r > 0$. This is the contradiction that we were seeking. Therefore, $\overline{y} \in \partial^s(P_{CRS}(x_0))$. Additionally, by (11.3) and (11.7), $\overline{y} \geq y_0$, which implies that $\overline{y} \in S(y_0; P_{CRS}(x_0))$.

(ii) Let $\overline{y} \in S(y_0; P_{CRS}(x_0)) = \partial^s(P_{CRS}(x_0)) \cap \{y \in R^s_+ : y_r \geq y_{r0}, \forall r\}$. We have that $(x_0, y_0) \in T_{CRS}$, which implies that $y_0 \in P_{CRS}(x_0)$. Under our assumptions, $x_0 \in R^m_+$, $x_0 \neq 0_m$, and $y_0 \in R^s_+$, $y_0 \neq 0_s$. Therefore, $\overline{y} \neq 0_s$ since otherwise $\overline{y}$ is dominated by y_0, what is a contradiction with $\overline{y} \in \partial^s(P_{CRS}(x_0))$. Now, we can solve model (8) under CRS for $(x_0, y_0) = (x_0, \overline{y})$, obtaining, by Lemma 1, that $WA^{\max}_{CRS,O}(x_0, \overline{y}; w^+) = 0$. Therefore, an optimal solution $\left(\overline{s}^+, \overline{\lambda}\right)$ with $\overline{s}^+ = 0_s$ for (8) under CRS exists that satisfies, by (8.2), the constraint $\sum_{j \in E_{CRS}} \overline{\lambda}_j y_{rj} = \overline{y}_r$, $\forall r$ [(11.1)]. By Lemma 3, $\sum_{j \in E_{CRS}} \overline{\lambda}_j x_j = x_0$ [(11.2)]. Moreover, $\overline{\lambda}_j \geq 0$, $j \in E_{CRS}$ [(11.8)] and $\tilde{s}^+_r := \overline{y}_r - y_{r0} \geq 0$, $r = 1, \ldots, s$, [(11.7)] allow the satisfaction of $\sum_{j \in E_{CRS}} \overline{\lambda}_j y_{rj} = y_{r0} + \tilde{s}^+_r$, $r = 1, \ldots, s$ [(11.3)]. By duality in Linear Programming, there exists an optimal solution for (9) under CRS[7] $(\overline{v}, \overline{u})$ such that $-\overline{v}x_{ij} + \sum_{r=1}^{s} \overline{u}_r y_{rj} \leq 0$, $j \in E_{CRS}$, $\overline{v} \geq 0$ [(11.5)] and $\overline{u}_r \geq w^+_r$, $r = 1, \ldots, s$ [(11.6)]. Defining $\overline{d}_j := \overline{v}x_{ij} - \sum_{r=1}^{s} \overline{u}_r y_{rj}$, $j \in E_{CRS}$, (11.4) and (11.11) are satisfied. Finally, by the Complementarity Theorem in Linear Programming, we have that $\overline{\lambda}_j \overline{d}_j = 0$, $j \in E_{CRS}$. In this way, defining $M := \max_{j \in E_{CRS}} \left\{\overline{d}_j, \overline{\lambda}_j\right\}$ and $\overline{b}_j := 0$ if $\overline{\lambda}_j > 0$, and $\overline{b}_j := 1$ otherwise, $j \in E_{CRS}$, constraints (11.9), (11.10) and (11.12) hold. Therefore, all the constraints in (11) are satisfied, which means that $\overline{y} \in \hat{S}^{=}(y_0; P_{CRS}(x_0))$. ∎

Proof of Proposition 2. On the one hand, $(\phi^*_1 y_{10}, \ldots, \phi^*_s y_{s0}) \in \{y \in R^s_+ : y_r \geq y_{r0}, \forall r\}$ by (14.10). Regarding if $(\phi^*_1 y_{10}, \ldots, \phi^*_s y_{s0}) \in \partial^s(P_{VRS}(x_0))$ or not, we have that $(\phi^*_1 y_{10}, \ldots, \phi^*_s y_{s0}) \in P_{VRS}(x_0)$ by (14.2)-(14.4). Moreover, (s^{+*}, γ^*) is an optimal solution of the lower level problem with $\phi_r = \phi^*_r$, $r = 1, \ldots, s$, since $(\phi^*, \lambda^*, s^{+*}, \gamma^*)$ is a feasible solution of the BLP problem. Additionally, $\sum_{r=1}^{s} w^+_r s^{+*}_r = 0$ by (14.5), which implies that $WA^{\max}_{VRS,O}(x_0, \phi^* y_0; w^+) = 0$, where $\phi^* y_0 = (\phi^*_1 y_{10}, \ldots, \phi^*_s y_{s0})$. In this way, invoking Lemma 1, we have that $(\phi^*_1 y_{10}, \ldots, \phi^*_s y_{s0}) \in \partial^s(P_{VRS}(x_0))$. Consequently, $(\phi^*_1 y_{10}, \ldots, \phi^*_s y_{s0}) \in S(y_0; P_{VRS}(x_0))$. On the other hand, let us suppose that $\exists \left(\overline{\phi}_1, \ldots, \overline{\phi}_s\right)$ with $\left(\overline{\phi}_1 y_{10}, \ldots, \overline{\phi}_s y_{s0}\right) \in S(y_0; P_{VRS}(x_0))$ such that $\frac{1}{s}\sum_{r=1}^{s} \overline{\phi}_r < \frac{1}{s}\sum_{r=1}^{s} \phi^*_r$. Then, $\left(\overline{\phi}_1 y_{10}, \ldots, \overline{\phi}_s y_{s0}\right) \in P_{VRS}(x_0)$, which implies by (2) that $\exists \overline{\lambda}_j \geq 0$, $j \in E_{VRS}$, satisfying $\sum_{j \in E_{VRS}} \overline{\lambda}_j x_{ij} \leq x_{i0}$, $i = 1, \ldots, m$, $\sum_{j \in E_{VRS}} \overline{\lambda}_j y_{rj} \geq \overline{\phi}_r y_{r0}$, $r = 1, \ldots, s$ and $\sum_{j \in E_{VRS}} \overline{\lambda}_j = 1$. Let also $\left(\overline{\overline{s}}^+, \overline{\overline{\lambda}}\right)$ be an optimal solution of model (8) for evaluating $\left(x_0, \overline{\phi} y_0\right)$, where $\overline{\phi} y_0 = \left(\overline{\phi}_1 y_{10}, \ldots, \overline{\phi}_s y_{s0}\right)$. Observe that $\overline{\overline{s}}^+_r = 0$, $r = 1, \ldots, s$, thanks to Lemma 1 since $\left(\overline{\phi}_1 y_{10}, \ldots, \overline{\phi}_s y_{s0}\right) \in \partial^s(P_{VRS}(x_0))$ by hypothesis. Now, considering $\left(\overline{\phi}, \overline{\lambda}, \overline{s}^+, \overline{\gamma}\right)$ with $\overline{\gamma}_j := \overline{\overline{\lambda}}_j$, $j \in E_{VRS}$, we get that $\left(\overline{\phi}, \overline{\lambda}, \overline{s}^+, \overline{\gamma}\right)$ is a feasible solution of (14). But then $\frac{1}{s}\sum_{r=1}^{s} \overline{\phi}_r < \frac{1}{s}\sum_{r=1}^{s} \phi^*_r$ would contradict the fact that $(\phi^*, \lambda^*, s^{+*}, \gamma^*)$ is an optimal solution of (14). Consequently, $\frac{1}{s}\sum_{r=1}^{s} \phi^*_r = Min\left\{\frac{1}{s}\sum_{r=1}^{s} \phi_r : (\phi_1 y_{10}, \ldots, \phi_s y_{s0}) \in S(y_0; P_{VRS}(x_0))\right\}$.

[7] In model (9) the decision variable α disappears.

References

[1] Agasisti T, Zoido P. The efficiency of secondary schools in an international perspective: preliminary results from PISA 2012. Paris: OECD Publishing; 2015.
[2] Amirteimoori A, Kordrostami S. A Euclidean distance-based measure of efficiency in data envelopment analysis. Optimization 2010;59:985–96.
[3] Ando K, Kai A, Maeda Y, Sekitani K. Least distance based inefficiency measures on the Pareto-efficient frontier in DEA. Journal of the Operations Research Society of Japan 2012;55:73–91.
[4] Angelo JS, Barbosa HJC. A study on the use of heuristics to solve a bilevel programming problem. International Transactions in Operational Research 2015;22:861–82.
[5] Aparicio J, Ruiz JL, Sirvent I. Closest targets and minimum distance to the Pareto-efficient frontier in DEA. Journal of Productivity Analysis 2007;28:209–18.
[6] Aparicio J, Pastor JT. A well-defined efficiency measure for dealing with closest targets in DEA. Applied Mathematics and Computation 2013;219:9142–54.
[7] Aparicio J, Pastor JT. On how to properly calculate the Euclidean distance-based measure in DEA. Optimization 2014;63(3):421–32.
[8] Aparicio J, Pastor JT. Closest targets and strong monotonicity on the strongly efficient frontier in DEA. Omega 2014;44:51–7.
[9] Baek C, Lee J. The relevance of DEA benchmarking information and the least-distance measure. Mathematical and Computer Modelling 2009;49:265–75.
[10] Banker RD, Charnes A, Cooper WW. Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis. Management Science 1984;30:1078–92.
[11] Beale, E.M.L. and Tomlin, J.A. (1970). Special facilities in a general mathematical programming system for non-convex problems using ordered sets of variables. In: Proceedings of the fifth international conference on operational research, edited by Lawrence, J, London: Tavistock Publications, p. 447–54.
[12] Briec W. Minimum distance to the complement of a convex set: duality result. Journal of Optimization Theory and Applications 1997;93:301–19.
[13] Briec W. Hölder distance function and measurement of technical efficiency. Journal of Productivity Analysis 1998;11:111–31.
[14] Briec W, Lemaire B. Technical efficiency and distance to a reverse convex set. European Journal of Operational Research 1999;114:178–87.
[15] Briec W, Lesourd JB. Metric distance function and profit: some duality results. Journal of Optimization Theory and Applications 1999;101(1):15–33.
[16] Briec W, Leleu H. Dual representations of non-parametric technologies and measurement of technical efficiency. J Product Anal 2003;20:71–96.
[17] Chambers RG, Mitchell T. Homotheticity and non-radial changes. J Product Anal 2001;15 31–19.
[18] Charnes A, Cooper WW, Thrall RM. A structure for classifying and characterizing efficiency and inefficiency in data envelopment analysis. Journal of Productivity Analysis 1991;2:197–237.
[19] Cherchye L, Van Puyenbroeck T. A comment on multi-stage DEA methodology. Operations Research Letters 2001;28:93–8.
[20] Coelli T. A multi-stage methodology for the solution of orientated DEA models. Operations Research Letters 1998;23:143–9.
[21] Colson B, Marcotte P, Savard G. An overview of bilevel optimization. Annals of Operations Research 2007;153:235–56.
[22] Cooper WW, Park KS, Pastor JT. RAM: a range adjusted measure of inefficiency for use with additive models, and relations to others models and measures in DEA. Journal of Productivity Analysis 1999;11:5–42.
[23] Cooper WW, Seiford LM, Tone K. Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software. Springer, New York; 2007.
[24] Crespo-Cebada E, Pedraja-Chaparro F, Santín D. Does school ownership matter? An unbiased efficiency comparison for regions of Spain Journal of Productivity Analysis 2014;41(1):153–72.
[25] De Witte K, López-Torres L. Efficiency in education. A review of literature and a way forward. Journal of Operational Research Society 2015 forthcoming.

[26] Färe R, Grosskopf S, Lovell CAK. The measurement of efficiency of production. Kluwer Nijhof Publishing, Boston; 1985.
[27] Frei FX, Harker PT. Projections onto efficient frontiers: theoretical and computational extensions to DEA. Journal of Productivity Analysis 1999;11:275–300.
[28] Fukuyama H, Maeda Y, Sekitani K, Shi J. Input-output substitutability and strongly monotonic p-norm least-distance DEA measures. European Journal of Operational Research 2014;237:997–1007.
[29] Fukuyama H, Masaki H, Sekitani K, Shi J. Distance optimization approach to ratio-form efficiency measures in data envelopment analysis. Journal of Productivity Analysis 2014;42:175–86.
[30] Fukuyama H, Hougaard JL, Sekitani K, Shi J. Efficiency measurement with a nonconvex free disposal hull technology. Journal of the Operational Research Society 2016;67:9–19.
[31] Gonzalez E, Alvarez A. From efficiency measurement to efficiency improvement: the choice of a relevant benchmark. European Journal of Operational Research 2001;133:512–20.
[32] Grifell-Tatje E, Lovell CAK, Pastor JT. A Quasi-Malmquist productivity index. Journal of Productivity Analysis 1998;10:7–20.
[33] He X, Li C, Huang T, Li C, Huang J. A recurrent neural network for solving Bilevel Linear Programming problem. IEEE Transactions on Neural Networks and Learning Systems 2014;25(4):824–30.
[34] Hiriart-Urruty J-B. Conditions for global optimality 2. Journal of Global Optimization 1998;13(4):349–67.
[35] Koopmans TC. Analysis of production as an efficient combination of activities. In: Koopmans TC, editor. Activity Analysis of Production and Allocation. New York: John Wiley; 1951.
[36] Lovell CAK, Pastor JT. Units invariant and translation invariant DEA models. Operations Research Letters 1995;18:147–51.
[37] Lozano S, Villa G. Determining a sequence of targets in DEA. Journal of Operational Research Society 2005;56:1439–47.
[38] Mislevy RJ, Beaton AE, Kaplan B, Sheehan KM. Estimating population characteristics from sparse matrix samples of item responses. Journal of Educational Measurement 1992;29(2):133–61.
[39] OECD. PISA 2009 Technical Report. PISA, Paris: OECD Publishing; 2012.
[40] Pastor JT, Ruiz JL, Sirvent I. An enhanced DEA Russell Graph efficiency measure. European Journal of Operational Research 1999;115:596–607.
[41] Pastor JT, Aparicio J. The relevance of DEA benchmarking information and the Least-Distance Measure: comment. Mathematical and Computer Modelling 2010;52:397–9.
[42] Penot J-P. Duality for anticonvex programs. Journal of Global Optimization 2001;19(2):163–82.
[43] Portela MCAS, Castro P, Thanassoulis E. Finding closest targets in non-oriented DEA models: the case of convex and non-convex technologies. J Product Anal 2003;19:251–69.
[44] Prieto AM, Zofio JL. Evaluating effectiveness in public provision of infrastructure and equipment: the case of Spanish municipalities. Journal of Productivity Analysis 2001;15:41–58.
[45] Ralph D, Dempe S. Directional derivatives of the solution of a parametric nonlinear program. Mathematical Programming 1995;70:159–72.
[46] Ruiz JL, Sirvent I. A DEA approach to derive individual lower and upper bounds for the technical and allocative components of the overall profit efficiency. Journal of the Operational Research Society 2010;62:1907–16.
[47] Ruiz JL, Sirvent I. Common benchmarking and ranking of units with DEA. Omega 2016;65:1–9.
[48] Ruiz JL, Segura JV, Sirvent I. Benchmarking and target setting with expert preferences: an application to the evaluation of educational performance of Spanish universities. European Journal of Operational Research 2015;242:594–605.
[49] Santín D, Sicilia G. Measuring the efficiency of public schools in Uruguay: main drivers and policy implications. Latin American Economic Review 2015;24 (1):1–28.
[50] Shi C, Lu J, Zhang G. An extended Kuhn–Tucker approach for linear bilevel programming. Applied Mathematics and Computation 2005;162:51–63.
[51] Shi C, Lu J, Zhang G. An extended Kth-best approach for linear bilevel programming. Applied Mathematics and Computation 2005;164:843–55.
[52] Shi C, Zhang G, Lu J, Zhou H. An extended branch and bound algorithm for linear bilevel programming. Applied Mathematics and Computation 2006;180:529–37.
[53] Singer I. Duality for nonconvex approximation and optimization. New York: Springer; 2006.
[54] Tone K. A slacks-based measure of efficiency in data envelopment analysis. European Journal of Operational Research 2001;130:498–509.
[55] Wu DD. Bilevel programming Data Envelopment Analysis with constrained resource. European Journal of Operational Research 2010;207:856–64.
[56] Wu M. The role of plausible values in large-scale surveys. Studies in Educational Evaluation 2005;31(2):114–28.
[57] J. Wu, J.-F. Chu, L.Liang, Target setting and allocation of carbon emissions abatement based on DEA and closest target: an application to 20 APEC economies, Nature Hazards 2015, http://dx.doi.org/10.1007/s11069-015-1973-6, in press.