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

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 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.

© 2016 Elsevier Ltd. All rights reserved.

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 efficiency of 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
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 slack 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 Brie’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: Brie and Lemaire [14], Brie and Lesourd [15] and Brie 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 [28] 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\(_i\) consumes \( x_j = (x_{j1}, \ldots, x_{jm}) \in \mathbb{R}^m_+ \), \( x_j \neq 0_m \), amounts of inputs for the production of \( y_j = (y_{j1}, \ldots, y_{jm}) \in \mathbb{R}^m_+ \), \( y_j \neq 0_m \), 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:

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

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

Below we introduce additional notions related to the production possibility set regardless of the assumed returns to scale, using \( T \) instead of \( T_{\text{CRS}} \) or \( T_{\text{VRS}} \). Nevertheless, these notions are also applicable to the \( T_{\text{CRS}} \) or \( T_{\text{VRS}} \) 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 \( \partial^w(T) = \left\{ (x, y) \in T : \exists \bar{x} < x, \bar{y} > y \Rightarrow (\bar{x}, \bar{y}) \notin T \right\} \). 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 \( \partial^s(T) = \left\{ (x, y) \in T : \bar{x} \leq x, \bar{y} \geq y \Rightarrow (\bar{x}, \bar{y}) \notin T \right\} \). 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. \( \partial^s(P(x)) = \{ y \in P(x) : y \geq \bar{y}, y \neq \bar{y} = \gamma \notin P(x) \} \), and it is a subset of the weakly efficient frontier of \( P(x) \), denoted and defined as \( \partial^w(P(x)) = \{ y \in P(x) : y > y \Rightarrow y \notin P(x) \} \).

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: \( P_{\text{CRS}}(x), P_{\text{VRS}}(x), \partial_{\text{CRS}}^s(P(x)), \partial_{\text{VRS}}^s(P(x)), \partial_{\text{CRS}}^w(P(x)) \) and \( \partial_{\text{VRS}}^w(P(x)) \).

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\(_j\) with data \((x_{0j}, y_{0j})\) can be formulated under Variable Returns to Scale as follows\(^\dagger\):

\[
W_{\text{A}}^{\text{max}}(x_{0j}, y_{0j}; w^-, w^+) = \max \left\{ \sum_{r=1}^m w_r^+ s_r^+ + \sum_{r=1}^s w_r^- s_r^- \right\} \\
\text{s.t.} \\
\sum_{j=1}^m \lambda_j x_{0j} = x_{0j} - s_0^-, i = 1, \ldots, m \quad (3.1) \\
\sum_{j=1}^m \lambda_j y_{0j} = y_{0j} + s_0^+, r = 1, \ldots, s \quad (3.2) \\
\sum_{j=1}^m \lambda_j = 1 \quad (3.3) \\
s_0^- \geq 0, i = 1, \ldots, m \quad (3.4) \\
s_0^+ \geq 0, r = 1, \ldots, s \quad (3.5) \\
\lambda_j \geq 0, j \in E_{\text{CRS}} \quad (3.6)
\]

where \( w^- = (w_1^-, \ldots, w_m^-, w_s^-) \in \mathbb{R}^{m+s}_- \) and \( w^+ = (w_1^+, \ldots, w_s^+) \in \mathbb{R}^s_+ \) are weights representing the relative importance of unit inputs and unit outputs, and \( E_{\text{CRS}} \) denotes the set of extreme efficient DMUs in the case of assuming VRS.\(^\dagger\)

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

\[
W_{\text{A}}^{\text{min}}(x_{0j}, y_{0j}; w^-, w^+) = \min \left\{ \sum_{i=1}^m v_i x_{0i} - \sum_{r=1}^s u_r y_{0r} + \alpha \right\} \\
\text{s.t.} \\
- \sum_{i=1}^m v_i x_{0i} + \sum_{r=1}^s u_r y_{0r} - \alpha \leq 0, j \in E_{\text{CRS}} \quad (4.1) \\
v_i \geq w_i^-, i = 1, \ldots, m \quad (4.2) \\
u_r \geq w_r^+, r = 1, \ldots, s \quad (4.3) \\
v_0 \geq w_i^-, r = 1, \ldots, s 
\]

Model (3) ‘maximizes’ a weighted \( \varepsilon \) distance from the DMU\(_0\) to the frontier of the production possibility set, thereby increasing outputs and reducing inputs at the same time. Let \((x_{0j}^+, y_{0j}^+, x_{0j}^0, y_{0j}^0)\) be an optimal solution of model (3), then \((x_{0j}^+, y_{0j}^+)\) defines as \( x_{0j}^0 = \sum_{j \in E_{\text{CRS}}} \lambda_j x_{0j}, i, \) and \( y_{0j}^0 = \sum_{j \in E_{\text{CRS}}} \lambda_j y_{0j}, r, \) \( vr \), denotes the projection point associated with the assessed DMU \((x_{0j}, y_{0j})\). In this way, the targets are the coordinates of the projection point \((x_{0j}^+, y_{0j}^+)\) 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,

\(^\dagger\) 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]).

\(^\dagger\) 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, \( E_{\text{CRS}} \) 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 $\hat{\sigma}(T)$.

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].

**Theorem 1.** [5].

Let $S(x_0, y_0; T_{VRS}) : = \hat{\sigma}(T_{VRS}) \cap \{(x, y) \in \mathbb{R}^m_+ \times \mathbb{R}^s_+ : x_i \leq x_0, \forall i, y_r \geq y_0, \forall r\}$ 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 $\exists \lambda, s, \alpha, \beta, v, u, \sigma, d$ such that $x_i = \sum_{j \in EVRS} \lambda_j x_j, \forall i,$ $y_r = \sum_{j \in EVRS} \lambda_j y_{jr}, \forall r$ and

$$\sum_{j \in EVRS} \lambda_j x_0 = x_0 - s_i^- , \quad i = 1, \ldots, m \quad (5.1)$$

$$\sum_{j \in EVRS} \lambda_j y_{r} = y_{r0} + s_r^+ , \quad r = 1, \ldots, s \quad (5.2)$$

$$\sum_{j \in EVRS} \lambda_j = 1 , \quad (5.3)$$

$$- \sum_{i = 1}^m v_i x_i + \sum_{r = 1}^s u_r y_r - \alpha + d_j = 0 , \quad j \in EVRS \quad (5.4)$$

$$v_i \geq w_i^- , \quad i = 1, \ldots, m \quad (5.5)$$

$$u_r \geq w_r^+ , \quad r = 1, \ldots, s \quad (5.6)$$

$$s_i^- \geq 0 , \quad i = 1, \ldots, m \quad (5.7)$$

$$s_r^+ \geq 0 , \quad r = 1, \ldots, s \quad (5.8)$$

$$\lambda_j \geq 0 , \quad j \in EVRS \quad (5.9)$$

$$d_j \leq M \alpha_j , \quad j \in EVRS \quad (5.10)$$

$$\lambda_j \leq 1 - b_j , \quad j \in EVRS \quad (5.11)$$

$$b_j \geq 0 , \quad j \in EVRS \quad (5.12)$$

$$b_j \in \{0, 1\} , \quad j \in EVRS \quad (5.13)$$

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 $\text{card}(EVRS)$ 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:

$$W_{VRS}^\text{min}(x_0, y_0; \{w^-, w^+\}) := \text{Min} \left\{ \sum_{i = 1}^m w_i^- s_i^- + \sum_{r = 1}^s w_r^+ s_r^+ : (x_0 - s_i^- , y_0 + s_r^+) \in S(x_0, y_0; T_{VRS}) \right\} \quad (6)$$

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:

$$W_{VRS,0}^\text{max}(x_0, y_0; \{w^+, w^+\}) := \text{Max} \left\{ \sum_{i = 1}^m w_i^+ s_i^+ : y_0 + s_i^+ \in S(y_0; P_{VRS}(x_0)) \right\} \quad (7)$$

where $S(y_0; P_{VRS}(x_0)) = \hat{\sigma}(P_{VRS}(x_0)) \cap \{y \in \mathbb{R}^s_+ : y_r \geq y_{r0}, \forall r\}$. 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: $S(y_0; P_{VRS}(x_0)) = \{y \in \mathbb{R}^s_+ : \forall y_r \geq y_{r0}, \forall r\}$. In this example, $\hat{\sigma}(P_{VRS}(2)) = (1)$. However, $(2, 1) \notin \hat{\sigma}(T_{VRS})$ 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:

$$W_{VRS,0}^\text{max}(x_0, y_0; \{w^+, w^+\}) = \text{Max} \sum_{i = 1}^m w_i^+ s_i^+ \quad (8.1)$$

$$\text{s.t.} \quad \sum_{j \in EVRS} \lambda_j x_0 \leq x_0 , i = 1, \ldots, m \quad (8.2)$$

$$\sum_{j \in EVRS} \lambda_j y_{rj} = y_{r0} + s_r^+ , r = 1, \ldots, s \quad (8.3)$$

$$\lambda_j \geq 0 , \quad j \in EVRS \quad (8.4)$$

$$\lambda_j \leq 1 , \quad j \in EVRS \quad (8.5)$$

$$b_j \geq 0 , \quad j \in EVRS \quad (8.6)$$

$$b_j \in \{0, 1\} , \quad j \in EVRS \quad (8.7)$$

where $M$ is a sufficiently big positive number.

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.
\[ v_0 \geq 0, i = 1, \ldots, m \]  
\[ u_0 \geq w_j^r, r = 1, \ldots, s \]  
\[ C \alpha \]  
\[ (9.2) \]  
\[ (9.3) \]  

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 \( v_0 \geq 0, \forall i \).

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

\[ \tilde{S}(y_0; PVRS(x_0)) = \left\{ y \in R^s_+ : \begin{cases} y_r = \sum_{j \in ECRS} \lambda_j y_{j,r}, & r = 1, \ldots, s, \\ \sum_{j \in ECRS} \lambda_j y_{j} = y_0, \\ \sum_{j \in ECRS} \lambda_j y_{j} = y_0 + s^+, & r = 1, \ldots, s, \\ -\sum_{r=1}^n v_r y_{r} + \sum_{r=1}^s u_r y_{r} - \alpha + d_0 = 0, & j \in ECRS \end{cases} \right\} \]  
\[ (10) \]

Then the question is: \( S(y_0; PVRS(x_0)) = \tilde{S}(y_0; PVRS(x_0)) \) ? 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, \( EVRS = \{A, B, C\} \). Let \( \lambda_1 = A_{1,1} = 1/2, s_1^+ = s_2^+ = 0, v = 5, u_1 = 1, u_2 = 4, \alpha = 4, \) \( d_1 = d_2 = d_0 = 0 \) and \( b_1 = b_2 = b_0 = 0 \). Then, \( \left( y_1, y_2 \right) = \left( \sum_{j \in ECRS} \lambda_j y_{j,1}, \sum_{j \in ECRS} \lambda_j y_{j,2} \right) = (3.4) \in \tilde{S}(3.4; PVRS(4)) \). However, since \( C = (4; 4.5) \) has been observed, \( \left( y_1, y_2 \right) = (4.5) \in PVRS(4) \). In this way, (4.5) dominates (3.4) in the sense of Pareto. Therefore, \( (3.4) \notin S \) \((3.4; PVRS(4)) \) and \( S((3.4); PVRS(4)) \) is in this example.

In the case of assuming Constant Returns to Scale, numerical examples on \( S(y_0; PCRS(x_0)) \) \( \neq \{ y \in R^s_+ : (x_0, y) \in S(x_0,y_0; TCRS) \} \) and \( S(y_0; PCRS(x_0)) \) 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 \( S(y_0; PCRS(x_0)) \) coincides with a similar set to \( \tilde{S}(y_0; PVRS(x_0)) \) 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]. \( y \in \partial(PVRS(x_0)) \) if and only if \( W_{Max}^{TCS}(x_0, y; w^+) = 0 \).

**Lemma 2.** [22]. Let \( (s^+, \lambda^*) \) be an optimal solution of (8) under CRS. Then, \( y_0^* = \left( \sum_{j \in ECRS} \lambda_j^* y_{j,1}, \ldots, \sum_{j \in ECRS} \lambda_j^* y_{j,s} \right) \in \partial(PVRS(x_0)) \).

**Lemma 3.** Let \( m = 1 \) and \( (s^*, \lambda^*) \) be an optimal solution of (8) under CRS. Then, \( \sum_{j \in ECRS} \lambda_j^* x_j = x_0 \).

Proof. See Appendix.

In order to prove the desired result, we need to adapt expression (10) to Constant Returns to Scale, deleting (10.2) and \( \alpha \), 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 \( S(y_0; PCRS(x_0)) = \tilde{S}(y_0; PCRS(x_0)) \).

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:

$$
\begin{align*}
\text{Min} & \quad \theta_k^k \sum_{i=1}^{n} \phi_k^i \\
\text{s.t.} & \quad \sum_{j=1}^{n} \theta_k^i x_{ij} \leq \theta_k^k x_{i0}, \quad i = 1, \ldots, m \\
& \quad \sum_{j=1}^{n} \theta_k^i y_{ij} \geq y_{i0}, \quad r = 1, \ldots, s \\
& \quad \sum_{j=1}^{n} \theta_k^i = 1, \\
& \quad \theta_k^i \geq 0, \quad j = 1, \ldots, n \\
& \quad \theta_k^k \leq 1, \quad i \neq k
\end{align*}
$$

where M is a sufficiently big positive number.

Then, from an optimal solution of (12), \( \left( x^{\theta_k}, \theta_1^k, \ldots, \theta_m^k \right) \), the
input-specific contraction for input k can be computed as

$$
C(x_0, y_0)_k = \sum_{i=1}^{m} (1 - \theta_k^i).
$$

Finally, the input-oriented Russell measure associated with the least distance and closest targets is
determined following the Gonzalez and Alvarez approach as

$$
C(x_0, y_0) = \min \{ C(x_0, y_0)_k : k = 1, \ldots, m \}.
$$

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

$$
C(x_0, y_0)_E = 2,3, \quad C(x_0, y_0)_D = 2,3 \quad \text{and} \quad C(x_0, y_0)_E = 0.43.
$$

In this way, we conclude that \( C(x_0, y_0)_1 = 0.43 \). 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

$$
C(x_0, y_0)_C = \frac{2}{3}, \quad C(x_0, y_0)_E = \frac{4}{3} \quad \text{and} \quad C(x_0, y_0)_E = \frac{4}{3}.
$$

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 \( C(x_0, y_0)_k, \quad k = 1, \ldots, m \).

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 par-
cular, 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
approach also presents some drawbacks using the same counterexample. Following
Brice [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 \( \text{Min} \ \theta_k^k \).

### Table 1: Data for Counterexample 3.

<table>
<thead>
<tr>
<th>DMU</th>
<th>Input 1</th>
<th>Input 2</th>
<th>Input 3</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>2.8</td>
<td>2.2</td>
<td>4.6</td>
<td>1</td>
</tr>
<tr>
<td>D</td>
<td>2.2</td>
<td>2.8</td>
<td>4.5</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

### 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 pro-
gramming problem where one of the constraints is an optimization
problem. This theory has been successfully applied to model different
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 \( z \in R^q \)
and \( t \in C \subseteq R^l \) the decision variables corresponding to the first and
second level, respectively. The general formulation of a Bilevel
Linear Programming (BLP) problem is as follows:

$$
\begin{align*}
\text{Min} & \quad c_1 z + d_1 t \\
\text{s.t.} & \quad A_1 z + B_1 t \leq b_1, \\
& \quad M \text{Min} \quad c_2 z + d_2 t \\
\text{s.t.} & \quad A_2 z + B_2 t \leq b_2, \\
& \quad z \geq 0, \quad t \geq 0
\end{align*}
$$

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
where all the functions are linear, like in (13), the model to be
described 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, \( (z^*, t^*) \geq 0 \) is a feas-
sible solution of (13) if \( t^* \) is an optimal solution of the lower level
program with \( z = z^* \) and, at the same time, \( A_1 z^* + B_1 t^* \leq b_1 \). In this
way, \( (z^*, t^*) \) is an optimal solution if additionally \( c_1 z^* + d_1 t^* \leq c_1 z +
\quad d_1 t \) 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:

\[
\begin{align*}
\text{Min} & \quad \frac{1}{2} \sum_{r=1}^{s} \phi_r, \\
\text{s.t.} & \quad \sum_{j \in E_{\text{VRS}}} \lambda_j x_j \leq x_{0}, & i = 1, \ldots, m \quad (14.2) \\
& \quad \sum_{j \in E_{\text{VRS}}} \lambda_j y_{ij} \geq \phi_j y_{0}, & r = 1, \ldots, s \quad (14.3) \\
& \quad \sum_{j \in E_{\text{VRS}}} \lambda_j = 1, \quad (14.4) \\
& \quad \sum_{r=1}^{s} w^+_r = 0, \quad (14.5) \\
& \quad \text{Max} \quad \sum_{r=1}^{s} w^+_r \gamma^+_r, \\
\text{s.t.} & \quad \sum_{j \in E_{\text{VRS}}} \gamma_j x_j \leq x_{0}, & i = 1, \ldots, m \quad (14.7) \\
& \quad \sum_{j \in E_{\text{VRS}}} \gamma_j y_{ij} = \phi_j y_{0} + s^+_j + \tau_j, & r = 1, \ldots, s \quad (14.8) \\
& \quad \sum_{j \in E_{\text{VRS}}} \gamma_j = 1, \quad (14.9) \\
\phi_j \geq 1, \quad \lambda_j, \gamma^+_r, \gamma_j \geq 0, \quad \forall r, j \quad (14.10)
\end{align*}
\]

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], [149], while the lower level problem matches (8) when the evaluated output vector is \((\phi_1 y_{0}, \ldots, \phi_j y_{0})\).

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, \(\frac{1}{2} \sum_{r=1}^{s} \phi_r^* = \text{Min} \left\{ \frac{1}{2} \sum_{r=1}^{s} \phi_r : (\phi_1 y_{10}, \ldots, \phi_j y_{0}) \in S(y_{0}; P_{\text{VRS}}(x_{0})) \right\} \).

Proof. See Appendix.

Regarding the weights \(w^+_r\) in (14), we will assume from now on that \(w^+_r = 1, \quad r = 1, \ldots, m\).

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).

\[
\begin{align*}
& \sum_{j \in E_{\text{VRS}}} \gamma_j x_j + l_i = x_{0}, & i = 1, \ldots, m \quad (15.1) \\
& \sum_{j \in E_{\text{VRS}}} \gamma_j y_{ij} = \phi_j y_{0} + s^+_j + \tau_j, & r = 1, \ldots, s \quad (15.2) \\
& \sum_{j \in E_{\text{VRS}}} \gamma_j = 1, \quad (15.3) \\
& \sum_{i=1}^{m} \eta_i x_i + \sum_{t=1}^{s} \mu_t y_{ij} + \psi_j + \tau_j = 0, & j \in E_{\text{VRS}} \quad (15.4) \\
& \eta_i + \epsilon_i = 0, & i = 1, \ldots, m \quad (15.5) \\
& \mu_t \geq 1, & r = 1, \ldots, s \quad (15.6) \\
& \gamma_j \tau_j = 0, & j \in E_{\text{VRS}} \quad (15.7) \\
& \lambda_i \epsilon_i = 0, & i = 1, \ldots, m \quad (15.8) \\
& \lambda_i, \eta_i, \mu_r, \tau_j, \epsilon_i \geq 0, & \forall i, j \quad (15.9)
\end{align*}
\]

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:

\[
\begin{align*}
\text{Min} & \quad \sum_{i=1}^{m} w^-_i s^-_i + \sum_{r=1}^{s} w^+_r \tau^+_r, \\
\text{s.t.} & \quad \sum_{j \in E_{\text{VRS}}} \lambda_j x_j = x_{0} - s^-_i, & i = 1, \ldots, m \quad (16.2) \\
& \quad \sum_{j \in E_{\text{VRS}}} \lambda_j y_{ij} = y_{0} + s^+_r, & r = 1, \ldots, s \quad (16.3) \\
& \quad \sum_{j \in E_{\text{VRS}}} \lambda_j = 1, \quad (16.4) \\
& \quad \sum_{i=1}^{m} \tau^-_i + \sum_{r=1}^{s} \tau^+_r = 0, \quad (16.5) \\
& \quad \text{Max} \quad \sum_{i=1}^{m} \tau^-_i + \sum_{r=1}^{s} \tau^+_r, \\
\text{s.t.} & \quad \sum_{j \in E_{\text{VRS}}} \gamma_j x_j = x_{0} - s^-_i - \tau^-_i, & i = 1, \ldots, m \quad (16.7) \\
& \quad \sum_{j \in E_{\text{VRS}}} \gamma_j y_{ij} = y_{0} + s^+_r + \tau^+_r, & r = 1, \ldots, s \quad (16.8) \\
& \quad \sum_{j \in E_{\text{VRS}}} \gamma_j = 1, \quad (16.9) \\
& \quad s^-_i, s^+_r, \tau^-_i, \tau^+_r, \gamma_j \geq 0, \quad \forall r, j \quad (16.10)
\end{align*}
\]

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.
As a proxy for the quality of students in the school, we use 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.

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 reports

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 \( \sum_{j \in E_{RS}} A_j y_j < x_0 \). Let us define \( x'_0 = \sum_{j \in E_{RS}} A_j y_j \). In this way, by (1), \((x'_0, y'_0) \in T_{CRS} \). Moreover, note that \( j \in E_{CRS} \) such that \( A_j y_j > 0 \), since otherwise constraint (8.2) would be violated for \( y_0 \in R^s_+ \) with \( c_0 \neq 0 \). Consequently, \( x'_0 = \sum_{j \in E_{CRS}} A_j y_j > 0 \) since \( x_j \in R^n_+ \), \( x_j \neq 0 \), and \( m = 1 \) for all \( j = 1, \ldots, n \). By the definition of \( x'_0, x'_0 < x_0 \). This implies that, under CRS, \( \frac{n}{x_0} A_j y_j < x_0 \in E_{CRS} \). Therefore, \( \sum_{j \in E_{RS}} A_j y_j = x_0 \in E_{CRS} \). Additionally, \( y'_0 \neq 0 \), since, by definition, \( y'_0 = \left( \sum_{j \in E_{RS}} A_j y_j, \ldots, \sum_{j \in E_{RS}} A_j y_j \right) \), which must be equal or greater than \( y_0 \) by (8.2) under CRS, which satisfies \( y_0 \in R^n_+ \). Therefore, \( \sum_{j \in E_{RS}} A_j y_j = x_0 \).

#### Proof of Proposition 1

(i) Let \((\overline{x}, \overline{y}, \overline{r}, \overline{P}, \overline{F})\) be a vector that satisfies constraints 11.2-11.12. From this solution, it is possible to generate \( \overline{y} \in S \) (y:PCRS(x0)) as \( \overline{y} = \sum_{j \in E_{RS}} A_j y_j, r = 1, \ldots, s \). Let us note that, by (1), \((\overline{x}, \overline{y}) \in T_{CRS} \), with \( \overline{x} : = \sum_{j \in E_{RS}} A_j x_j = x_0 \), and, consequently, \( \overline{y} \in P_{CRS}(\overline{x}) = P_{CRS}(x_0) \). Let us now assume that \( \overline{y} \in P_{CRS}(x_0) \) such that \( \overline{y} \succeq \overline{y} \notin \overline{y} \). We want to arrive at a contradiction. By (11.4) and (11.11), we have that \(-\overline{y} x + \sum_{r = 1}^s \pi_r y_r \leq 0 \) for all \((x, y) \in T_{CRS} \). This implies that \(-\overline{y} x_0 + \sum_{r = 1}^s \pi_r y_r \leq 0 \) for all \((x_0, y) \in T_{CRS} \). In this way, \(-\overline{y} x_0 + \sum_{r = 1}^s \pi_r y'_r \leq 0 \) since

### Table 2

Descriptive statistics.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Type of variable</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESCS</td>
<td>Input</td>
<td>4.396</td>
<td>0.538</td>
<td>2.190</td>
<td>5.970</td>
</tr>
<tr>
<td>SCMATEDU</td>
<td>Input</td>
<td>3.683</td>
<td>0.891</td>
<td>0.008</td>
<td>5.576</td>
</tr>
<tr>
<td>TEACHERS</td>
<td>Input</td>
<td>10.14</td>
<td>6.309</td>
<td>0.719</td>
<td>90.01</td>
</tr>
<tr>
<td>PVMATH</td>
<td>Output</td>
<td>490.9</td>
<td>46.17</td>
<td>130.1</td>
<td>609.6</td>
</tr>
<tr>
<td>PVREAD</td>
<td>Output</td>
<td>490.4</td>
<td>46.56</td>
<td>297.7</td>
<td>626.6</td>
</tr>
</tbody>
</table>

### Table 3

Results of model (14) for 902 schools.

<table>
<thead>
<tr>
<th>Model (14)</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score of efficiency</td>
<td>1.122</td>
<td>0.084</td>
<td>1.000</td>
<td>1.672</td>
</tr>
<tr>
<td>Time of execution (s)</td>
<td>10.782</td>
<td>4.854</td>
<td>1</td>
<td>42</td>
</tr>
</tbody>
</table>

---

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.
\((x_0, y') \in T_{CRS}\) because \(y' \in P_{CRS}(x_0)\). Additionally, 
\[-\nabla x_0 + \sum_{r = 1}^{s} \mathbf{p}_r y_r = -\nabla x_0 + \sum_{r = 1}^{s} \mathbf{p}_r y_r = -\nabla x_0 + \sum_{r = 1}^{s} \mathbf{p}_r y_r\]
implies by (2) that \(\exists \lambda_j \geq 0, j \in E_{VRS}\), satisfying \(\sum_{j \in E_{VRS}} \lambda_j y_j \leq x_0\), 
\(i = 1, \ldots, m, \sum_{j \in E_{CRS}} \lambda_j y_j \geq \bar{y}_j y_{i0}, r = 1, \ldots, s\) and \(\sum_{j \in E_{CRS}} \lambda_j = 1\). Let also 
\((s^+, \bar{y}^+)'\) be an optimal solution of model (8) for evaluating 
\((x_0, \bar{y}_{i0})\), where \(\bar{y}_{i0} = (\bar{y}_{i1}, \ldots, \bar{y}_{is})\). Observe that \(\bar{y}^+_j = 0, \quad r = 1, \ldots, s\), thanks to Lemma 1 since 
\((\bar{y}_{i1}, \ldots, \bar{y}_{is}) \in \mathcal{P}(P_{CRS}(x_0))\) by hypothesis. Now, considering 
\((\bar{y}_{i1}, \ldots, \bar{y}_{is}, \bar{y}^+)\) with \(\bar{y}^+_j = 0\), \(j \in E_{CRS}\), we get that 
\((\bar{y}_{i1}, \bar{y}^+)\) is a feasible solution of 
(14). But then \(\frac{1}{r} \sum_{j = 1}^{r} \phi_j \leq \sum_{r = 1}^{s} \phi_r\) would contradict the fact that 
\((\phi^*, \bar{y}^+, \bar{y}'^+)\) is an optimal solution of (14). Consequently, 
\(\frac{1}{r} \sum_{j = 1}^{r} \phi_j = \min \left\{ \frac{1}{r} \sum_{j = 1}^{r} \phi_j : (\phi_{i1}, \ldots, \phi_{is}) \in S(y_0; P_{CRS}(x_0)) \right\}\)

\[\text{References}\]