A Well-Defined Composite Indicator: An Application to Corporate Social Responsibility

Juan Aparicio¹ · Magdalena Kapelko² · Juan F. Monge¹

Received: 31 January 2019 / Accepted: 4 June 2020 / Published online: 23 June 2020
© The Author(s) 2020

Abstract

This paper introduces a model to construct composite indicators for performance evaluation of decision making units, which is based upon the determination of the least distance from each assessed unit to a frontier estimated by data envelopment analysis. This generates less demanding targets from a benchmarking point of view. The model also makes it possible to account for the existence of slacks in all the considered dimensions (sub-indicators), playing with the notion of Pareto efficiency. Additionally, our approach satisfies units invariance, translation invariance and strong monotonicity and ensures that the weights used for the aggregation of the sub-indicators are always strictly positive. All previous approaches based on data envelopment analysis have failed to satisfy at least one of these properties. We also implement a new version of the Russell output measure of technical efficiency working with full-dimensional efficient facets. Finally, the new approach is illustrated by an application to the sphere of corporate social responsibility, showing the main empirical implications of the theoretical properties.

Keywords Data envelopment analysis · Benefit of the Doubt · Least distance · Benchmarking · Strong monotonicity

Mathematics Subject Classification 90B50 · 90C05 · 90C11 · 91B38

Communicated by Kaoru Tone.

✉ Magdalena Kapelko
magdalena.kapelko@ue.wroc.pl

Juan Aparicio
j.aparicio@umh.es

Juan F. Monge
monge@umh.es

¹ Center of Operations Research (CIO), University Miguel Hernandez of Elche (UMH), 03202 Elche, Alicante, Spain

² Department of Logistics, Wrocław University of Economics and Business, Wrocław, Poland

1 Introduction

Data envelopment analysis (DEA) is a methodology based on mathematical programming for the assessment of technical efficiency of a set of decision making units (DMUs), which allows a data-driven construction of a piece-wise frontier, enveloping the data cloud of observations and, at the same time, the determination of the distance from each DMU to this frontier. DEA provides both an efficiency measure (score) for each of the assessed DMUs and information on the targets that have been used in the efficiency assessment in the case of inefficient DMUs. The targets are the levels of operation of inputs and outputs that would make the corresponding inefficient DMU perform efficiently. In consequence, the targets can play an important role in practice, since they indicate keys for the inefficient DMUs to improve their performance.

Since its origins [1], DEA researchers have introduced many ways of implementing the idea of technical inefficiency as the distance from each evaluated unit to the estimated frontier. Each one of these ways is related to a different way of projecting the DMU onto the frontier. So, we have radial measures [1, 2], directional distance functions [3], the hyperbolic measure [4], weighted additive measures [5, 6], the Russell measures [4], the enhanced Russell graph or slacks-based measure (SBM) [7, 8], and so on. However, from a benchmarking viewpoint, it is worth mentioning that traditional DEA models can yield targets that are determined by the “furthest” efficient projection to the assessed DMU; see [9].¹ In contrast, measures, based on the notion of least distance, determine closer targets, suggesting directions of improvement for the inefficient units that may lead them to the frontier with less effort. In recent times, the calculation of the least distance to the frontier for determining the closest targets has been associated with the application of the principle of least action (PLA) in physics to the sphere of performance evaluation [10].

Additionally, DEA has been recently linked to the construction of composite indicators in social science through the notion of the Benefit-of-the-Doubt (BoD) model²; see [11–13]. This is a DEA model without real inputs [14], where sub-indicators are all treated as outputs to generate an overall and objective aggregated indicator for each assessed DMU through the determination of its efficiency. In that model, a unique ‘virtual’ input equal to one for all units is included in the model.

However, as far as we are aware, no previous contribution to the definition of composite indicators by BoD in DEA has applied the PLA, when determining the final aggregated indicator for each unit, neglecting the potentiality of the approach as a benchmarking tool for the evaluated DMUs. Previous research has used the variety of

¹ Radial measures, directional distance functions and the hyperbolic measure present this problem, when the assessed DMU is projected to the weakly efficient frontier. In the case of measures based on slacks, such as the weighted additive measure or the Slacks-Based Measure, the problem is more frequent since these measures maximize input and output slacks, generating the furthest targets.

² Standard DEA measures can be also understood as composite indicators since they allow the aggregation of information on several dimensions (inputs and outputs) in a final score. In this sense, it implies that standard DEA measures are able to composite indicators. However, DEA usually estimates the frontier in production contexts, where inputs are transformed into outputs. By contrast, in the traditional literature of composite indicators, which exists beyond DEA, each indicator could be seen as an output. Therefore, we are working in a non-usual context, without inputs and without the flavor of the traditional production framework.

measures within DEA to create BoD models. The usage of DEA to aggregate individual indicators was initiated by [14], who proposed a DEA radial model without inputs to evaluate macroeconomic performance based on four indicators. [11–13] popularized the usage of radial DEA models in the estimations of composite indicators by BoD. Recent contributions in this research area focus on the extensions of radial BoD toward the non-radial model, which takes into account the possible existence of slacks [15], the multiple-layer model that considers a hierarchical structure of indicators [16], the directional model to recognize the preference structure among indicators [17, 18], the directional model that allows dealing with undesirable output indicators [19], the robust and non-compensatory composite indicator based on the directional model [20], the composite indicator with imprecise data [21], the robust BoD model that considers external factors directly [22], the spatial directional robust BoD model [23], and a translation invariant directional distance function, which allows the computation of indicators for firms with zero values for outputs [24].

DEA is a nonparametric and deterministic technique for estimating technical efficiency. The literature contains an alternative parametric approach, called stochastic frontier analysis (SFA), which is grounded on statistical bases. An interesting characteristic of SFA is that the “goodness of fit” can be measured by statistical tests, while DEA lacks a goodness-of-fit tool. In this sense, as other authors have already pointed out [5, 7, 25, 26], the goodness of the measure of efficiency in DEA is checked by establishing a set of interesting properties, which the measure should satisfy. The above-cited authors have highlighted the following properties (adapted for output-oriented approaches): The measure should be greater or equal to one, with one signaling Pareto efficiency; units invariance; translation invariance; and strong monotonicity. Additionally, when DEA is used for aggregating sub-indicators, other properties can be also interesting. In particular, assigning zero weights to one or more sub-indicators (outputs) would imply that these performance criteria are (nearly) ignored in the construction of the composite performance score; see, for example, [23]. Finally, the application of the PLA should be considered essential, if the researcher needs to provide useful benchmarking information in the way of less demanding targets. Unfortunately, none of the existing measures in the literature satisfies all these properties at the same time.

In this paper, we fulfill this gap in the literature by introducing a BoD model in DEA for defining a composite indicator, which satisfies all of the above-mentioned properties. We have implemented a version of the Russell output measure, linked to the extension of full-dimensional efficient facets (FDEFs). In addition, and as a methodological contribution of our paper, we show—for the first time—how a Russell-type output measure, based upon the PLA and the extension of FDEFs in DEA, can be implemented in one step by a mixed integer linear program (MILP) model. Furthermore, for illustrative purposes, we apply the new approach to a dataset on the corporate social responsibility (CSR) activities of European food, beverage and tobacco manufacturing companies in 2017.

The paper unfolds as follows: In the following section, we briefly review the literature on DEA and the PLA. In Sect. 3, we introduce a new composite indicator. Section 4 includes an illustration of the new approach proposed in this paper, focused

Fig. 1 The standard DEA technology vs the extended DEA technology

on CSR data, showing the empirical implications of the theoretical properties. Finally, Sect. 5 concludes.

2 The Principle of Least Action in DEA

Regarding a brief revision of the literature linked to the PLA, Refs. [27, 28] were the precursors of this line of research in DEA. Frei and Harker [29] suggested resorting to the Euclidean distance to measure technical inefficiency. Cherchye and Van Puyenbroeck [30] defined a least-distance model in an input-oriented-space framework. González and Álvarez [31] maximized the input-oriented Russell efficiency measure, determining the closest efficient targets. Later, [32] established that the general aim in efficiency evaluation should be to globally minimize the input and output slacks, regardless of which DEA measure is finally applied. Aparicio et al. [9] identified the closest targets for a dataset of international airlines, applying a new version of the SBM. More recent approaches are [10, 33–36]. For more details, see [37].

DEA measures, based on the determination of the least distance, usually lack the important property of strong monotonicity; see [33, 38, 39]. Nevertheless, [34] proved that the output-oriented version of the Russell measure [4] is a well-defined efficiency measure, satisfying strong monotonicity on the strongly efficient frontier, if efficiency is evaluated with respect to an extended facet production possibility set based on FDEF, instead of standard DEA technology. Figure 1 illustrates the idea behind this method graphically. In this figure, two FDEFs appear, AB and BC, which are ‘extended’ by the dot lines generating a new production possibility set, which contains the original one. In this figure, A, B and C are the set of extreme efficient points (corners) of the output production set. Additionally, [35] showed the reason why strong monotonicity fails. The use of FDEFs avoids the problems associated with the dimensionality of the strongly efficient frontier and other related problems [40–43].

In the same line as [34], we will define in the next section a composite indicator, based upon the BoD model, which implements the determination of least distance, but ensuring strong monotonicity and the satisfaction of a set of key properties for effi-

ciency measurement. From a methodological point of view, we show how to implement the model of [34] for the first time, by resorting to mathematical programming in just one step. Instead, [34] needed to identify all the FDEFs in a first stage and then, determine the least distance from each DMU, which is a highly time-consuming step for big problems.

In what follows, we introduce some definitions that are needed to understand the approach, based upon the extension of the traditional DEA technology; see [34, 35]. Nevertheless, we first need to introduce some notation. Consider that we have observed n DMUs that use m inputs to produce s outputs. These are denoted by (X_j, Y_j) , $j = 1, \dots, n$. It is assumed $X_j = (x_{1j}, \dots, x_{mj}) \geq 0_m$, $j = 1, \dots, n$, and $Y_j = (y_{1j}, \dots, y_{sj}) \geq 0_s$, $j = 1, \dots, n$. The relative efficiency of each DMU₀ in the sample is assessed with reference to the so-called production possibility set $T := \{(X, Y) | X \text{ can produce } Y\}$, which can be empirically constructed from the n observations by assuming several postulates; see [2]. If, in particular, variable returns to scale is assumed, then T can be characterized as follows:

$$T = \left\{ (X, Y) \in R_+^{m+s} \left| \sum_{j=1}^n \lambda_j X_j \leq X, \sum_{j=1}^n \lambda_j Y_j \geq Y, \sum_{j=1}^n \lambda_j = 1, \lambda_j \geq 0, j = 1, \dots, n \right. \right\} \quad (1)$$

Hereinafter, we assume that each DMU is interested in maximizing outputs, while using no more than the observed amount of any input. This type of approach is called output oriented in the literature. 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, $P(X) := \{Y : (X, Y) \in T\}$. On the other hand, given X , the efficient frontier of $P(X)$, also called the weak efficient frontier, is defined as $\partial(P(X)) := \{Y \in P(X) : \hat{Y} > Y \Rightarrow \hat{Y} \notin P(X)\}$; see [28]. Following [44], in order to measure technical efficiency in the Pareto sense, it is necessary to isolate a certain subset of $\partial(P(X))$. We are referring to the strong efficient frontier, as defined below:

$$\partial^s(P(X)) := \left\{ Y \in P(X) : \hat{Y} \geq Y, \hat{Y} \neq Y \Rightarrow \hat{Y} \notin P(X) \right\} \quad (2)$$

Throughout the paper, with the aim of measuring technical efficiency, we will compare the actual performance of each DMU₀ with respect to the points belonging to the strong efficient frontier.

In this paper and under the BoD model, it is assumed that all DMUs use a unique input equal to one. Hence, hereinafter, we will use the notation $P(1)$ and $\partial^s(P(1))$ for denoting the output production set and its corresponding strong efficient frontier, respectively, both sets being common for all the units. Note, additionally, that due to the definition of $P(X)$ and T , $P(1) = \{Y \in R_+^s : \sum_{j=1}^n \lambda_j Y_j \geq Y, \sum_{j=1}^n \lambda_j = 1, \lambda_j \geq 0, \forall j\}$. Furthermore, the set of all the extreme efficient points of $P(1)$ will be denoted hereafter as E . In this way, the

polyhedron $P(1)$ can be then rewritten equivalently by substituting $j = 1, \dots, n$ by $j \in E$: $P(1) = \left\{ Y \in R_+^s : \sum_{j \in E} \lambda_j Y_j \geq Y, \sum_{j \in E} \lambda_j = 1, \lambda_j \geq 0, \forall j \right\}$.

To check Pareto efficiency, it is usual to resort to the weighted additive model; see [6]. In particular, an output vector $Y' \in R_+^s$ is assessed by (3) or by (4), being (4) the dual problem of (3).

$$\begin{aligned} & \text{Max} \sum_{r=1}^s w_r^+ s_r^+ \\ & \text{s.t.} \quad - \sum_{j=1}^n \lambda_j y_{rj} + s_r^+ \leq -y'_r, \quad s_r^+ \geq 0, \quad r = 1, \dots, s \end{aligned} \quad (3)$$

$$\begin{aligned} & \sum_{j=1}^n \lambda_j = 1, \quad \lambda_j \geq 0, \quad j = 1, \dots, n \\ & \text{Min} \quad - \sum_{r=1}^s u_r y'_r + \delta \\ & \text{s.t.} \quad - \sum_{r=1}^s u_r y_{rj} + \delta \geq 0, \quad j = 1, \dots, n \\ & \quad \quad u_r \geq w_r^+, \quad r = 1, \dots, s \end{aligned} \quad (4)$$

If $w_r^+ > 0$ for all $r = 1, \dots, s$, then it not difficult to prove that Y' is a Pareto-efficient point of $P(1)$, if and only if, the objective function of (3), or equivalently (4), equals zero.

Proposition 2.1 $Y' \in \partial^s(P(1))$, if and only if the optimal values of models (3) and (4) are equal to zero.

Proof See [6]. □

Now, before showing the version of the Russell output measure based on the PLA from [34], we turn to the definition of the traditional Russell output measure of technical efficiency [4, p. 149]: $\Gamma(Y_0) := \text{Max} \left\{ \frac{1}{s} \sum_{r=1}^s \phi_r : (\phi_1 y_{10}, \dots, \phi_s y_{s0}) \in P(1), (\phi_1, \dots, \phi_s) \geq 1_s \right\}$.

In the above model, ϕ_r evaluates the relative proportional expansion rate of output r , $r = 1, \dots, s$, whereas the objective function averages these proportional rates of output expansion. Also, in the above formulation, the constraints $\phi_r \geq 1$, $r = 1, \dots, s$ are the requirements for dominance. Additionally, by a change of variables, it is possible to prove that the traditional Russell output measure of technical efficiency is equivalent to the following formulation:

$$\Gamma(Y_0) = 1 + \text{Max} \left\{ \frac{1}{s} \sum_{r=1}^s \frac{s_r^+}{y_{r0}} : (Y_0 + s^+) \in \partial^s(P(1)), s^+ = (s_1^+, \dots, s_s^+) \geq 0_s \right\}, \quad (5)$$

where we use slacks and have changed $P(1)$ to $\partial^s(P(1))$. It is worth mentioning that the traditional Russell output measure determines the furthest targets from DMU_0 to the strong efficient frontier of $P(1)$, since the objective function maximizes the (weighted)

sum of the slacks and, at the optimum, $s_r^{+*} = \sum_{j=1}^n \lambda_j^* y_{rj} - y_{r0} = y_{r0}^* - y_{r0}$ for all $r = 1, \dots, s$.

Next, we show the formulation of the PLA version of the Russell output measure by [34]. In particular, “Max” is changed by “Min” in (5) in order to determine closest targets instead of furthest targets.

$$\Gamma^{\min}(Y_0) := 1 + \text{Min} \left\{ \frac{1}{s} \sum_{r=1}^s \frac{s_r^+}{y_{r0}} : (Y_0 + s^+) \in \partial^s(P(1)), s^+ = (s_1^+, \dots, s_s^+) \geq 0_s \right\} \quad (6)$$

Unfortunately, $\Gamma^{\min}(Y_0)$ lacks some properties. In particular, it does not fulfill neither translation invariance nor strong monotonicity. [34] studied the property of strong monotonicity in detail and its relationship with (6), showing that strong monotonicity fails. For this reason, these authors proposed a solution by identifying and extending FDEFs. In particular, they proved that the Russell output measure based on the PLA satisfies strong monotonicity, when we deal with an extended facet production possibility set instead of the usual DEA technology.

In what follows, we show the definition of the extended facet production possibility set. To this end, it is necessary to introduce some notation and definitions; see, for more details, [35, 42, 45]. We assume that $P(1) \subset R^s$ contains at least an interior point, and therefore, its dimension equals s .

Definition 2.1 An ‘efficient’ face F of $P(1)$ is a face of $P(1)$ such that it has at least one supporting hyperplane $\sum_{r=1}^s \mu_r y_r = \psi$, which presents nonnegative coefficients, that is, $\mu_r \geq 0, r = 1, \dots, s$.

Definition 2.2 A full-dimensional ‘efficient’ facet (FDEF) of $P(1)$ is an efficient face F of the polyhedron $\{Y \in R_+^s : \sum_{j=1}^n \lambda_j Y_j \geq Y, \sum_{j=1}^n \lambda_j = 1, \lambda_j \geq 0, \forall j\}$ such that a set of s affinely independent extreme efficient points belongs to F , and its corresponding supporting hyperplane $\sum_{r=1}^s \mu_r y_r = \psi$ presents strictly positive coefficients, that is, $\mu_r > 0, r = 1, \dots, s$.

The graphical idea behind the extension of FDEFs illustrated in Fig. 1 needs to be mathematically formalized, before being used in our problem. Let us introduce some notation. Let J_k be the subset of extreme efficient points that span the k th FDEF, $k = 1, \dots, K$. Finally, let $\sum_{r=1}^s \mu_r^k y_r = \psi^k$ be the mathematical expression of the supporting hyperplane³ associated with J_k . In this way, we can generate an empirical output production set, $P_{\text{EEF}(J_k)}(1)$, as the intersection between the half-space generated by the supporting hyperplane $\sum_{r=1}^s \mu_r^k y_r = \psi^k$ and the nonnegative orthant: $P_{\text{EEF}(J_k)}(x) = \{y \in R_+^s : \sum_{r=1}^s \mu_r^k y_r \leq \psi^k\}$. Using the last expression, the extended facet output production possibility set, denoted hereinafter as $P_{\text{EXFA}}(1)$, is defined as $P_{\text{EXFA}}(1) := \bigcap_{k=1}^K P_{\text{EEF}(J_k)}(1)$, which is equivalent to $P_{\text{EXFA}}(1) = \{y \in R_+^s : \sum_{r=1}^s \mu_r^k y_r \leq \psi^k, k = 1, \dots, K\}$.

³ If $\sum_{r=1}^s \mu_r^k y_r = \psi^k$ is one supporting hyperplane of some FDEF k , then $\sum_{r=1}^s (\varepsilon \mu_r^k) y_r = \varepsilon \psi^k$ for any $\varepsilon > 0$ is also a supporting hyperplane for the same FDEF.

Regarding the strongly efficient frontier of $P_{\text{EXFA}}(1)$, it is defined by analogy with $\partial^s(P(1))$ as $\partial^s(P_{\text{EXFA}}(1)) := \{y \in P_{\text{EXFA}}(1) : \hat{y} \geq y, \hat{y} \neq y \Rightarrow \hat{y} \notin P_{\text{EXFA}}(1)\}$.

We now turn to the definition of the Russell output measure based on both least distance and the extended facet production possibility set. In this sense, by analogy with (6), we directly derive this measure simply changing the reference set $\partial^s(P(1))$ by $\partial^s(P_{\text{EXFA}}(1))$.

$$\Gamma_{\text{EXFA}}^{\min}(Y_0) := 1 + \text{Min} \left\{ \frac{1}{s} \sum_{r=1}^s \frac{s_r^+}{y_{r0}} : (Y_0 + s^+) \in \partial^s(P_{\text{EXFA}}(1)), s^+ = (s_1^+, \dots, s_s^+) \geq 0_s \right\} \quad (7)$$

3 A New Aggregated Indicator Based on DEA and the Principle of Least Action

Next, we define the new composite indicator (NCI).

$$\text{NCI}_{\text{EXFA}}^{\min}(Y_0) := 1 + \text{Min} \left\{ \frac{1}{s} \sum_{r=1}^s \frac{s_r^+}{y_r^U - y_{r0}} : (Y_0 + s^+) \in \partial^s(P_{\text{EXFA}}(1)), s^+ = (s_1^+, \dots, s_s^+) \geq 0_s \right\} \quad (8)$$

In contrast to the original Russell output measure based on both PLA and the extended facet production possibility set (7), in (8), the denominator of the objective function has been transformed in order to ensure that translation invariance holds. Specifically, y_r^U corresponds to the maximum observed in the sample for output r , for all $r = 1, \dots, s$. Using $(y_r^U - y_{r0})$ as denominator, instead of y_{r0} , allows canceling any translation of the output data. Another possibility, which will allow us to define a translation invariance measure, would be directly using the ranges of each output in the denominators. However, as [46] pointed out, this last type of measure has a smaller discriminatory power with respect to the efficiency scores, which we would obtain for the set of DMUs. This is the reason why we prefer to resort to a unit-specific range as a denominator. Additionally, the NCI is well defined, if we adopt the next convention. If, for DMU_0 , output r' satisfies that $y_{r'0} = y_{r'}^U$, then there is no room for improvement according to data information, and, by convention, we will take $\frac{s_{r'}^{*+}}{y_{r'}^U - y_{r'0}} = 0$, where $*$ denotes optimality. This same idea was previously suggested in the literature by [46], for defining the bounded adjusted measure. Nevertheless, seeking simplicity, hereinafter we will describe what we have to do to evaluate a unit such that $y_{r0} \neq y_r^U$, for all $r = 1, \dots, s$.

Moreover, and from a computational viewpoint, the NCI is not easy to calculate. This difficulty results from the complexity of determining the least distance to the frontier of a convex set (a polyhedron in our context) from an interior point of this set, as this problem is equivalent to minimizing a convex function on the complement of a convex set. This is known in optimization theory as the reverse convex best approximation problem [47].

Additionally, (8) is not a directly implementable optimization program in usual optimizers. Hence, we want to repress (8) in a standard way. To do so, we first prove a result, which states that if you have explicitly determined all the FDEFs of your database, then we can get the optimal value of (8) by calculating K simple ratios.

Proposition 3.1 $\text{NCI}_{\text{EXFA}}^{\min}(Y_0) = 1 + \min_{1 \leq k \leq K} \left\{ \frac{\psi^k - \sum_{r=1}^s \mu_r^k y_{r0}}{\max\{s\mu_1^k(y_1^U - y_{10}), \dots, s\mu_s^k(y_s^U - y_{s0})\}} \right\}.$

Proof It is an adaptation, with other weights, of the proof of Proposition 5 in [34]. $\square$

Following Proposition 3.1, we would have to determine all the FDEFs before solving (8). In particular, if we resort, for example, to the Qhull software, then we have to carry out a multistage procedure as follows. First, using Qhull we would get a set of hyperplanes. Second, we should check both the slopes and the offset of each of them in order to discard those, which cannot define an efficient face. After applying this filter, we would get a group of candidates. Third, we should analyze the units, which span each hyperplane, in order to select those, that actually define a FDEF. In the fourth stage, for each inefficient unit and FDEF, a ratio—as in Proposition 3.1—should be solved. Finally, in the fifth stage, the minimum value of these ratios would be the desired optimal value of (8) for each assessed DMU.

However, in contrast to the multi-stage procedure described above, we now go on to show, how to determine the optimal value of model (8), invoking Proposition 3.1, by solving a MILP for each inefficient unit, something that is implementable in any optimizer. To get that result, we resort to bi-level programming, which has been previously used in DEA in [48, 49].

Next, we show the model that makes it possible to determine the new composite indicator. To do that, we next specify and explain each constraint in a separate way.

Proposition 3.1 states that the NCI can be computed through the ratios $\psi^k - \sum_{r=1}^s \mu_r^k y_{r0} / \max\{s\mu_1^k(y_1^U - y_{10}), \dots, s\mu_s^k(y_s^U - y_{s0})\}$, $k = 1, \dots, K$, once we have determined the supporting hyperplane $\sum_{r=1}^s \mu_r^k y_r = \psi^k$ associated with each FDEF of $P(1)$. Note that this fractional form can be linearized by defining $(\max\{s\mu_1^k(y_1^U - y_{10}), \dots, s\mu_s^k(y_s^U - y_{s0})\}) = 1$, which can be also modeled in a linear way through constraints:

$$\begin{aligned} s\mu_{r0}(y_r^U - y_{r0}) &\leq 1, \quad s\mu_{r0}(y_r^U - y_{r0}) \geq \delta_{r0}, \quad r = 1, \dots, s \\ \sum_{r=1}^s \delta_{r0} &\geq 1, \quad \delta_{r0} \in \{0, 1\}, \quad r = 1, \dots, s \end{aligned} \quad (9)$$

In this way, if we work with supporting hyperplanes $\sum_{r=1}^s \mu_r y_r = \psi$ that satisfy the above constraints, the NCI can be determined through the minimization of $(\psi_0 - \sum_{r=1}^s \mu_{r0} y_{r0})$ on the set of all the supporting hyperplanes of FDEFs of $P(1)$. Additionally, any feasible solution of the model, which makes it possible to determine the new composite indicator, must be related to a FDEF of $P(1)$ and, at the same time, any FDEF of $P(1)$ must be associated with a feasible solution of the new model to work properly. To obtain this, we need to incorporate the characteristics of a FDEF

to the constraints of the model. First, the supporting hyperplane, which we want to determine, linked to a FDEF, $\sum_{r=1}^s \mu_{r0} y_r = \psi_0$, should be a valid inequality of $P(1)$; that is,

$$\begin{aligned} \sum_{r=1}^s \mu_{r0} y_{rj} - \psi_0 + d_{j0} &= 0, \quad d_{j0} \geq 0, \quad j \in E \\ \mu_{r0} &\geq 0, \quad r = 1, \dots, s \end{aligned} \quad (10)$$

Additionally, this valid inequality makes it possible to define a face of $P(1)$: $F_0 := \{Y \in P(1) : \sum_{r=1}^s \mu_{r0} y_r = \psi_0\} \neq \emptyset$. In particular, (11) ensures that F_0 is not an empty set and there are s extreme efficient points, which belong to F_0 : $\tau_{j0} = 1$ implies $d_{j0} = 0$ and, therefore, $\sum_{r=1}^s \mu_{r0} y_{rj} = \psi_0$.

$$\begin{aligned} d_{j0} \tau_{j0} &= 0, \quad \tau_{j0} \in \{0, 1\}, \quad j \in E \\ \sum_{j \in E} \tau_{j0} &= s \end{aligned} \quad (11)$$

Let E_0 be the set of points such that $\tau_{j0} = 1$. F_0 is an efficient face of $P(1)$. To ensure that the dimension of this face is maximal, that is, $s - 1$, the s extreme efficient points must also be affinely independent points. By definition, the unique solution of $\sum_{j \in E_0} \alpha_j Y_j = 0$, $\sum_{j \in E_0} \alpha_j = 0$ must be $\alpha_j = 0$ for all $j \in E_0$. Being M a strictly positive large number, this is guaranteed by solving (12) and (13) for each $k \in E$ and forcing the satisfaction of (14) at optimum:

$$\begin{aligned} \min \quad & \alpha_{kk0}^{\min} \\ \text{s.t.} \quad & \sum_{j \in E} \alpha_{jk0}^{\min} y_{rj} = 0, \quad r = 1, \dots, s \\ & \sum_{j \in E} \alpha_{jk0}^{\min} = 0, \quad -M \tau_{j0} \leq \alpha_{jk0}^{\min} \leq M \tau_{j0}, \quad j \in E \end{aligned} \quad (12)$$

$$\begin{aligned} \min \quad & \alpha_{kk0}^{\max} \\ \text{s.t.} \quad & \sum_{j \in E} \alpha_{jk0}^{\max} y_{rj} = 0, \quad r = 1, \dots, s \\ & \sum_{j \in E} \alpha_{jk0}^{\max} = 0, \quad -M \tau_{j0} \leq \alpha_{jk0}^{\max} \leq M \tau_{j0}, \quad j \in E \end{aligned} \quad (13)$$

$$\alpha_{jk0}^{\min} = \alpha_{jk0}^{\max} = 0, \quad j, k \in E \quad (14)$$

Finally, since we are seeking supporting hyperplanes associated with FDEFs, we need to ensure that $\mu_{r0} > 0$, for all $r = 1, \dots, s$. This is obtained by checking the Pareto efficiency of the centroid of the s extreme efficient points in E_0 ; that is, the following expression:

$$\bar{y}_{r0} = \frac{1}{s} \sum_{j \in E} \tau_{j0} y_{rj}, \quad r = 1, \dots, s \quad (15)$$

To do that, we use program (16) and force the satisfaction of (17) at optimum.

$$\begin{aligned} \max \quad & \sum_{r=1}^s \frac{t_{r0}}{R_r} \\ \text{s.t.} \quad & \sum_{j \in E} \lambda_{j0} y_{rj} = \bar{y}_{r0} + t_{r0}, \quad r = 1, \dots, s \\ & \sum_{j \in E} \lambda_{j0} = 1, \quad t_{r0}, \lambda_{j0} \geq 0, \quad r = 1, \dots, s, j \in E \end{aligned} \quad (16)$$

where R_r is the output range, $r = 1, \dots, s$, which is assumed to be strictly positive.

$$\sum_{r=1}^s t_{r0} = 0. \quad (17)$$

By Lemma 3.2 and Proposition 2.1, the centroid is a point non-dominated in $P(1)$; consequently, $\mu_{r0} > 0$, for all $r = 1, \dots, s$. All this guarantees that $\sum_{r=1}^s \mu_{r0} y_r = \psi_0$ is a supporting hyperplane associated with a FDEF. Combining all the previous constraints, the program to be solved would be:

$$\begin{aligned} \min \quad & \psi_0 - \sum_{r=1}^s \mu_{r0} y_{r0} \\ \text{s.t.} \quad & (9) - (17) \end{aligned} \quad (18)$$

Now, we need to prove that the NCI coincides with one plus the optimal value of model (18). To do that, we first prove some lemmas, which we will use in the proof of the main result: Theorem 3.1.

Lemma 3.1 *Let E' be a set of s affinely independent extreme points. Then, the following linear system has a unique solution.*

$$\sum_{r=1}^s \mu_r y_{rj} - \psi = 0, \quad j \in E', \quad \sum_{r=1}^s \mu_r = 1. \quad (19)$$

Proof See “Appendix A.” □

Lemma 3.2 *Let E' be a set of s affinely independent extreme points on an efficient face F of $P(1)$. And let $\sum_{r=1}^s \mu'_r y_r = \psi'$ be a supporting hyperplane associated with F , with $\|\mu'\|_{\ell_1} = \sum_{r=1}^s \mu'_r = 1$. Then, $\mu'_r > 0$, $r = 1, \dots, s$, if and only if $\bar{Y}$, defined as $\bar{y}_r = \frac{1}{s} \sum_{j \in E'} y_{rj}$, $r = 1, \dots, s$, is a Pareto-efficient point in $P(1)$.*

Proof See “Appendix A.” □

Now, we are ready to prove the main result of this paper: the equivalence between (8) and (18).

Theorem 3.1 *One plus the optimal value of model (18) coincides with $\text{NCI}_{\text{EXFA}}^{\min}(Y_0)$.*

Proof See “Appendix A.” □

From a computational point of view, model (18) needs to be reformulated in order to be implemented in a standard optimizer. First, the multiple lower-level decision problems in (12), (13) and (16) should be treated in a way such that model (18) would be reformulated as a one-level problem. To do that, it is enough to apply the Karush–Kuhn–Tucker (KKT) optimality conditions to the lower-level problems [50]. Additionally, constraint $d_{j0}\tau_{j0} = 0$, $j \in E$, is not linear, but can be identified as a special ordered set (SOS) [51]. 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 CPLEX, also use SOS with continuous variables. In this way, we are able to rewrite model (18) as a mathematical problem, which can be implemented in standard optimizers, by transforming the bi-level program (18) in a standard one-level problem using KKT and resorting to SOS conditions.

Now, it is interesting to show the properties that the new Russell-type measure of technical efficiency satisfies, properties that will be inherited by the NCI composite indicator. Regarding this issue, [26] were the first who proposed a set of desirable properties that an ideal efficiency measure should meet. Later, [7] listed similar requirements and suggested a few others. In particular, the main properties for an output-oriented measure would be (P1) the measure should be greater or equal to one, where one represents full efficiency (Pareto–Koopmans efficiency); (P2) units invariance; (P3) translation invariance; and finally, (P4) strong monotonicity in outputs.

Theorem 3.2 *Let Y_0 be the output vector corresponding to DMU_0 . Then, the following is true for NCI:*

- [P1] (i) $NCI_{EXFA}^{\min}(Y_0) \geq 1$ and (ii) $NCI_{EXFA}^{\min}(Y_0) = 1$ if and only if $Y_0 \in \partial^s(P_{EXFA}(1))$.
- [P2] (iii) $NCI_{EXFA}^{\min}(Y_0)$ is units invariant.
- [P3] (iv) $NCI_{EXFA}^{\min}(Y_0)$ is translation invariant.
- [P4] (v) $NCI_{EXFA}^{\min}(Y_0) = 1$ is strongly monotonic in outputs.

Proof The proof is apparent from the definition of (8) and model (18). $\square$

Additionally, in the context of composite indicators, two supplementary properties could be of interest. The first one has been recently pointed out by [23] indicating that weights (shadow prices) associated with each sub-indicator (output) should be strictly positive for assuring that no dimension is omitted in the computation of the final composite indicator [P5]. Our approach satisfies this property as a consequence of determining FDEFs.⁴ Secondly, the new measure based on the least distance provides

⁴ As [52] pointed out ‘... Our approach makes it possible to select the weights associated with the facets of higher dimension that this unit generates and, in particular, it selects those weights associated with a FDEF if any. In this sense the weights provided by our procedure will have the maximum support from the production possibility set.’ This is one of the benefits of using this type of weight. Additionally, the sub-gradients at extreme efficient points in DEA models are not unique, and the number of supporting hyperplanes at this type of observations is infinite. In this sense, although our approach does not solve the problem of alternative solutions (weights), at least it drastically reduces this number, since our solutions are exclusively linked to FDEFs.

targets that can be achieved with the minimum effort [P6]. This fact is very important in practice since demanding targets could demotivate firms to achieve the efficient status by improving its outputs.

Furthermore, as we show in “Appendix B,” the new output-oriented measure $\text{NCI}_{\text{EXFA}}^{\min}(Y_0)$ is the only measure in the context of BoD models, among the most famous ones in the DEA literature that satisfies properties [P1]–[P6].

Finally, to get the optimal targets for each DMU_0 , we have to solve the linear system of Eqs. (20), where through the first equation we consider all the feasible projection points, which are on the optimal hyperplane determined by (20), and, through the second equation, we are sure that $Y_0^* = Y_0 + s_0^*$ is the point where $\text{NCI}_{\text{EXFA}}^{\min}(Y_0)$ is achieved:⁵

$$\begin{aligned} \sum_{r=1}^s \mu_{r0}^*(y_{r0} + s_{r0}) + \psi_0^* &= 0, \quad 1 + \frac{1}{s} \sum_{r=1}^s \frac{s_{r0}}{y_{r0} - y_{r0}} = \text{NCI}_{\text{EXFA}}^{\min}(Y_0), \\ s_{r0} &\geq 0, \quad r = 1, \dots, s \end{aligned} \quad (20)$$

4 Empirical Illustration

The new composite indicator is applied to the dataset on CSR activities of the main European companies in the food, beverage and tobacco manufacturing industry for the year 2017, obtained from Sustainalytics. The Sustainalytics dataset has largely been used in previous research; see, for example, [55–57]. The data in this dataset are supplied in the form of scores on over 70 indicators, and the raw scores range from 0 to 100, where 0 is the lowest rating. The dataset also provides information on the weighted scores, which reflect the importance of an indicator in a specific industry. In this study, we apply weighted scores of five indicators, which are classified by Sustainalytics as being specifically dedicated to the industry under study: fair trade products (output 1), water management programs (output 2), sustainable agriculture programs (output 3), organic products (output 4) and genetically modified organism (GMO) policy (output 5) (their values range from 0 to 4.14).⁶ The dataset comprises both owner firms and their subsidiaries, and to avoid repetition of scores, we restrict our analysis to owner firms. The resulting sample comprises 58 firms. With this sample size, we satisfy all popular rules of thumb regarding the number of DMUs versus the number of inputs and outputs in DEA; see, for example, [59–63]. We also pass the diagnostics of [64], which make it possible to determine whether dimensionality reduction is necessary.⁷

⁵ It is worth adding that, in the presence of multiple optimal solutions, the reference sets for inefficient units may not be unique. Recent literature deals with this issue and presents models to uniquely identify the reference sets; see [53, 54].

⁶ In fact, the dataset consists of a larger number of indicators specific to the industry under study (in total 15 indicators). However, we restricted our analysis to these five indicators due to two main reasons. The first reason was the limited availability of data for these other indicators. The second reason was related to the fact that these five indicators effectively summarize the main CSR challenges in the industry under study, associated with responsible sourcing, environmental issues such as water use, and the quality, healthiness and safety of food products [58].

⁷ In particular, the correlations between our outputs are relatively low, while eigensystem analysis yields that we cannot reduce our data to one principal component.

Overall, with this sample, we do not pretend to estimate the theoretical population frontier, for which statistical arguments would need to be applied; instead, we analyze the specific set of companies with the purpose of benchmarking. The sample contains the largest food, beverage and tobacco manufacturing firms by market capitalization in Europe, such as Danone, Nestlé and the Coca-Cola Company, which account for approximately 30% of the total revenues of this sector in Europe.

Table 1 summarizes the results for 25 selected DMUs: the value of our new composite indicator, and the DMUs actual values of CSR outputs together with the corresponding targets (in parentheses).⁸ The results for all DMUs can be obtained upon request. The analysis reveals seven DMUs (DMUs 23, 25, 38, 41, 44, 55 and 56) as Pareto-efficient and forming part of FDEFs. The remaining fifty-one DMUs are inefficient, that is, they could increase CSR outputs in order to reach an “ideal point” in terms of CSR activities, defined by the maximum values of each CSR output. For example, the target values indicate that DMU5 should exclusively increase its output 2 (that is, investment in water management programs) by 8.9%. Meanwhile, DMU6 should increase its output 4 (organic products) by 22.3% and output 5 (GMO policy) by 8.9%. The “ideal point” might not be feasible and not observed within the current values of DMUs outputs (as we work with FDEFs, which are not always observable), but still it is something that firms in the sample should seek to reach. For example, the value of the new indicator equal to 1.15 for DMU1 could be given the following interpretation: DMU1 could, on average, increase all of its CSR outputs by approximately 15% of the amount left to reach this “ideal point.” Examining the results of the table in depth, we now draw our attention to DMUs 19, 27 and 28, which are the most inefficient in terms of producing CSR activities (composite indicators equal to 1.25) and at the same time the targets obtained suggest making greater improvements in outputs. In particular, for these DMUs, the values of targets suggest that firms should increase their score in output 2 (that is, CSR projects related with water management programs) from 0 to 4.45. This is really a considerable improvement, which might not be possible for a DMU to reach in the short run and hence, gradual improvement in the target might be made year by year [66]. In fact, CSR in general is an asset, that is long-term, and firms, that have longer investor horizons in CSR, create larger shareholder values [67]. In particular, CSR activity of water management programs has a long-term character, which deals with water scarcity, water quality and climate change [68].

Furthermore, we wish to make some points regarding the ‘empirical’ importance of the properties that new model satisfies, compared to the models highlighted in “Appendix B.” In particular, beginning with the property of Pareto efficiency, we ran a traditional radial BoD model [11–14], adding small exemplary values to all output values in order to be able to undertake the analysis with the radial model with 0 values for some outputs; that is, two different constants $k = 1$ and $k = 4$ seeking robustness in our analysis. We found, for $k = 1$, that 34% of the DMUs (35 units)

⁸ The computation time of the new indicator was 65 s, which suggests that the problem is not time-consuming to solve. The calculations were conducted on a PC with a 2.5 GHz dual-core Intel Core i5 processor, 8 Gb of RAM and the operating system was OS X 10.9. The optimization software package CPLEX v12.5 (MIP solver) was used for the computations [65]. Our code for computations and data can be obtained upon request.

Table 1 PLA-based composite indicators for 25 selected DMUs, targets are reported in parentheses

DMU	Indicator	Targets				
		Output 1	Output 2	Output 3	Output 4	Output 5
1	1.15	0.00 (0.00)	1.79 (1.79)	2.24 (2.24)	0.00 (0.00)	0.28 (0.93)
2	1.23	0.00 (0.00)	0.90 (0.90)	0.00 (0.00)	0.00 (0.00)	0.00 (1.27)
3	1.22	0.00 (0.00)	0.00 (4.00)	0.75 (0.75)	0.00 (0.00)	0.20 (0.20)
4	1.07	0.00 (0.00)	0.90 (1.80)	0.00 (0.00)	1.60 (1.60)	1.00 (1.00)
5	1.02	0.00 (0.00)	1.79 (1.95)	0.56 (0.56)	1.79 (1.79)	0.84 (0.84)
6	1.04	0.00 (0.00)	0.45 (0.45)	0.56 (0.56)	1.79 (2.19)	1.12 (1.22)
12	1.23	0.00 (0.00)	0.90 (0.90)	0.00 (0.00)	0.00 (0.00)	0.00 (1.27)
13	1.20	0.00 (0.00)	0.90 (3.60)	0.75 (0.75)	1.60 (1.60)	0.00 (0.00)
18	1.18	0.00 (0.00)	1.80 (1.80)	0.75 (0.75)	0.00 (0.00)	0.00 (1.03)
19	1.25	0.00 (0.00)	0.00 (4.45)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
23	1.00	0.00 (0.00)	1.79 (1.79)	2.24 (2.24)	1.79 (1.79)	0.84 (0.84)
25	1.00	1.66 (1.66)	0.00 (0.00)	1.50 (1.50)	4.00 (4.00)	1.00 (1.00)
27	1.25	0.00 (0.00)	0.00 (4.45)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
28	1.25	0.00 (0.00)	0.00 (4.45)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
30	1.21	0.00 (0.00)	0.90 (3.67)	0.00 (0.00)	1.60 (1.60)	0.00 (0.00)
35	1.07	0.00 (0.00)	0.43 (1.53)	1.07 (1.07)	0.00 (0.00)	1.07 (1.07)
38	1.00	4.14 (4.14)	0.00 (0.00)	3.00 (3.00)	0.00 (0.00)	0.00 (0.00)
41	1.00	0.00 (0.00)	0.90 (0.90)	2.24 (2.24)	0.00 (0.00)	1.12 (1.12)
42	1.16	0.00 (0.00)	1.70 (1.70)	2.13 (2.13)	0.00 (0.00)	0.21 (0.96)
44	1.00	1.34 (1.34)	1.63 (1.63)	2.04 (2.04)	1.63 (1.63)	0.76 (0.76)
48	1.02	0.00 (0.00)	0.85 (1.19)	2.13 (2.13)	0.00 (0.00)	1.07 (1.07)
55	1.00	0.00 (0.00)	1.80 (1.80)	0.00 (0.00)	1.60 (1.60)	1.00 (1.00)
56	1.00	0.00 (0.00)	3.60 (3.60)	0.75 (0.75)	1.60 (1.60)	0.00 (0.00)
57	1.14	1.34 (1.34)	0.43 (2.67)	0.53 (0.53)	1.70 (1.70)	0.27 (0.27)
58	1.17	1.34 (1.34)	0.81 (3.16)	1.02 (1.02)	1.63 (1.63)	0.00 (0.00)

are not Pareto-efficient and, for $k = 4$, we observed that this is the case for 50% of the DMUs (29 units). In particular, for $k = 1$, DMU2 is radially inefficient (score = 1.39). Its corresponding output projection is (1.39, 2.64, 1.39, 1.39, 1.39), which is dominated in the sense of Pareto by a convex combination of DMU44 (0.995) and DMU56 (0.05). Therefore, in practice, if we use the radial model (equivalently the DDF BoD model with $g = Y_0$), then the approach will provide targets that could be improved. This result is in contrast to the new model, for which all DMUs satisfy the property of Pareto-efficiency. Regarding units invariance, there is nothing to say, since all the measures satisfy this property. Furthermore, we analyzed the property of translation invariance; that is, invariance of efficiency scores with respect to adding a strictly positive constant to any considered input and/or output variable. To show its importance, we ran an output-oriented SBM model, which coincides with Russell model, with $k = 1$ and $k = 4$ added to output variables. Otherwise, these measures could

not be correctly determined since our database contains zeros (the same happens with respect to the radial and DDF BoD models). None of these measures satisfy translation invariance. Let us show an example of the implications of this drawback in practice. For DMU2, for example, the score for the SBM indicator changed from 2.37, for $k = 1$, to 1.34, for $k = 4$, something that would not occur with the new model. This finding has important practical implications, because, when we use the model that is not translation-invariant, it is not clear which values of efficiency should be reported, since the selection of constant k is completely arbitrary, and the resulting score can be very different depending on the selected constant. Moreover, the importance of the property of strong monotonicity can be assessed by looking at DMUs 1 and 23, in which DMU23 dominates DMU1, as all its outputs are larger or equal to outputs of DMU1. Therefore, the performance of DMU23 is clearly better than that of DMU1, regardless of which measure is selected. However, for these DMUs, the radial BoD model (and also DDF BoD model with $g = Y_0$) with, for example, $k = 1$ added presents scores equal to 1 for both DMUs, which can be understood as an inconsistency with respect to the observed output values; consequently, strong monotonicity is not satisfied. For the new model, DMU's 23 score was 1 and for DMU1 it was 1.15; hence strong monotonicity is obeyed. Therefore, new model gives consistent results, which is important in practice, since otherwise the practitioner would see that DMU23 is clearly better than DMU1. However, the measure did not tell the same story. Regarding the importance of the property of the strict positivity of the weights (shadow prices), using the radial BoD model (and also DDF BoD model) with $k = 1$ added, all but one of the DMUs had at least one weight equal to 0, while, for $k = 4$, this was the case for all DMUs except four. This is in contrast to the new model, in which all weights are strictly positive, which ensures that no dimension (information) is omitted in the computation of the composite indicator. This demonstrates the advantage of our model with regard to weights compared to the traditional radial and DDF BoD models in practice: Certain traditional measures can neglect some information (output dimensions), when the composite indicator is built. Finally, regarding the property of least distance, the BoD model was run based on the well-known range-adjusted measure (RAM) [25]; for example, for DMU1, we obtained the targets such that for output 4 were considerably larger than for the new model (1.79 compared to 0) and for output 5 were negligibly smaller (0.84 compared to 0.93), without changes in the other dimensions. Therefore, the new model provided smaller targets overall than those based on the RAM, which signifies smaller improvements for DMUs to become efficient. This fact is important in practice since demanding targets could demotivate firms to achieve the efficient status by improving its outputs in the short run. “Appendix C” summarizes the results of radial BoD model for $k = 1$ added to outputs equal to zero, for the same 25 selected DMUs. Due to space limitations, we do not report the detailed results for all other models, which prove the empirical importance of the properties that our new model satisfies. These results can be obtained from the authors upon request.

5 Conclusions

In this paper, for the first time, a composite indicator based on the PLA in DEA has been introduced to the literature that, additionally, satisfies a set of interesting properties. Indeed, it satisfies more properties than previous DEA approaches. This development is important for several reasons. First, the traditional DEA models yield targets that are related in general to the furthest projection to the assessed DMU onto the efficient frontier, making this benchmarking evaluation not credible. In contrast, finding the closest targets allows determining directions of improvement associated with less effort as opposed to the traditional approach. Second, our approach satisfies several interesting properties from an economic and mathematical point of view. Our indicator is units invariant, translation invariant, strong monotonic and, additionally, takes into account all sources of inefficiency, since it evaluates Pareto–Koopmans efficiency instead of the Debreu–Farrell notion of technical efficiency. Third, our methodology guarantees that the weights used for aggregating all sub-indicators are always strictly positive, something that contrasts with traditional approaches based upon radial models.

In order to ensure the satisfaction of all the properties, we resorted to a solution, which is based upon extending the FDEFs of the original DEA polyhedral technology, and introduced, for calculating the composite indicator, a new bi-level linear program, that avoids determining explicitly all the FDEFs and does its job in only one step. This feature contrasts to previous existing methods, which needed a multistage procedure.

To show the workability of the developed model and the empirical implications of the theoretical properties, the new model was applied using a recent dataset consisting of CSR activities for the main food, beverage and tobacco manufacturing firms in Europe. The results of this study are important for efficiency researchers and practitioners, as the new model with important properties is put forward. However, it could also be of interest to CSR analysts on how firms could improve CSR performance with minimal effort. Future research efforts could be dedicated toward the analysis of computational complexity of the developed estimator, given it is a NP-hard problem.

Acknowledgements The authors are grateful to the associate editor and three anonymous reviewers for their comments. We also wish to thank the National Science Centre in Poland (Grant Number 2016/23/B/HS4/03398) and the Spanish Ministry for Economy and Competitiveness (*Ministerio de Economía, Industria y Competitividad*), the State Research Agency (*Agencia Estatal de Investigación*) and the European Regional Development Fund (*Fondo Europeo de Desarrollo Regional*) under Grant PID2019-105952GB-I00 (AEI/FEDER, UE) for providing financial support for this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

A: Proofs of Lemmas and Theorems

Proof of Lemma 3.1 Without loss of generality, let us assume that the s extreme points in E' coincide with the first s observed DMUs. We can write (19) as a matrix:

$$\underbrace{\begin{pmatrix} y_{11} & y_{21} & \cdots & y_{s1} & -1 \\ y_{12} & y_{22} & \cdots & y_{s2} & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{1s} & y_{2s} & \cdots & y_{ss} & -1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix}}_A \underbrace{\begin{pmatrix} \mu \\ \psi \end{pmatrix}}_b = \underbrace{\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}}_b \quad (21)$$

We want to prove that $\text{rank}(A) = \text{rank}(A|b) = s + 1$ [the number of variables in (21)], implying that the system (21) has a unique solution. We will work with the rows of matrix A . The vectors $\begin{pmatrix} Y_1 \\ -1 \end{pmatrix}, \dots, \begin{pmatrix} Y_s \\ -1 \end{pmatrix}$ are linearly independent, because $Y_1, \dots, Y_s$ are affinely independent [45, Proposition 1.3 on page 84]. In this way, $\text{rank}\left(\begin{pmatrix} Y_1 \\ -1 \end{pmatrix}, \dots, \begin{pmatrix} Y_s \\ -1 \end{pmatrix}\right) = s$. Additionally, $\text{rank}\left(\begin{pmatrix} Y_1 \\ -1 \end{pmatrix}, \dots, \begin{pmatrix} Y_s \\ -1 \end{pmatrix}\right) = \text{rank}\left(\begin{pmatrix} Y_1 \\ -1 \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} Y_s \\ -1 \\ 0 \end{pmatrix}\right)$. This means that the system of equations $\sum_{r=1}^s \mu_r y_{rj} - \psi = 0, j = 1, \dots, s$, has infinite solutions, since the number of variables is $s + 1$. Additionally, there is at least one solution for this system with $\mu \neq 0_s$ since, otherwise, $\psi = 0$ and the unique solution of the system would be $(\mu, \psi) = (0_s, 0)$. This implies that given any solution of $\sum_{r=1}^s \mu_r y_{rj} - \psi = 0, j = 1, \dots, s$, (μ', ψ') , with $\mu' \neq 0_s$, we have that $\left(\frac{\mu'}{\sum_{r=1}^s \mu'_r}, \frac{\psi'}{\sum_{r=1}^s \mu'_r}\right)$ is a solution of system (19). Hence, rank

$(A) = \text{rank}(A|b)$. Now, we consider rows of $(A|b)$. The system $\alpha_1 \begin{pmatrix} Y_1 \\ -1 \\ 0 \end{pmatrix} + \dots + \alpha_s$

$\begin{pmatrix} Y_s \\ -1 \\ 0 \end{pmatrix} + \alpha_{s+1} \begin{pmatrix} 1_s \\ 0 \\ 1 \end{pmatrix} = 0$, with $\alpha_l \in R, l = 1, \dots, s + 1$, is such that $\alpha_{s+1} = 0$ [by

the last row]. So, we need to know if the system $\alpha_1 \begin{pmatrix} Y_1 \\ -1 \\ 0 \end{pmatrix} + \dots + \alpha_s \begin{pmatrix} Y_s \\ -1 \\ 0 \end{pmatrix} = 0$ gives the unique solution $\alpha_l = 0, l = 1, \dots, s$. If this system gives a solution $\hat{\alpha} \neq 0_s$, then $\hat{\alpha}$ is also a solution of $\alpha_1 \begin{pmatrix} Y_1 \\ -1 \end{pmatrix} + \dots + \alpha_s \begin{pmatrix} Y_s \\ -1 \end{pmatrix} = 0$, which is not possible, since the vectors $\begin{pmatrix} Y_1 \\ -1 \end{pmatrix}, \dots, \begin{pmatrix} Y_s \\ -1 \end{pmatrix}$ are linearly independent, as we saw

above. Consequently, the system $\alpha_1 \begin{pmatrix} Y_1 \\ -1 \\ 0 \end{pmatrix} + \dots + \alpha_s \begin{pmatrix} Y_s \\ -1 \\ 0 \end{pmatrix} = 0$ gives the unique solution $\alpha_l = 0$, $l = 1, \dots, s$, and, definitely, the vectors $\begin{pmatrix} Y_1 \\ -1 \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} Y_s \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1_s \\ 0 \\ 1 \end{pmatrix}$ are linearly independent. Hence, $\text{rank}(A|b) = s + 1$. Therefore, we have that $\text{rank}(A) = \text{rank}(A|b) = s + 1$, which coincides with the number of variables of the system (21). This implies that system (21) and, consequently, system (19) gives a unique solution. $\square$

Proof of Lemma 3.2 (i) Let us assume that $\mu'_r > 0$, $r = 1, \dots, s$. Then, $(\hat{\mu}, \hat{\psi})$ with $\hat{\mu}_r = \frac{\mu'_r}{\min\{\mu'_1, \dots, \mu'_s\}}$ and $\hat{\psi} = \frac{\psi'}{\min\{\mu'_1, \dots, \mu'_s\}}$ is a feasible solution of (4), when point $\bar{Y}$ is evaluated and weights are fixed as $w_r^+ = 1$, $r = 1, \dots, s$, as we show next. Additionally, and with respect to the objective function of model (4) at

$$(\hat{\mu}, \hat{\psi}), \text{ we have } -\sum_{r=1}^s \hat{\mu}_r \bar{y}_r + \hat{\psi} = \frac{-\sum_{r=1}^s \mu'_r \bar{y}_r + \psi'}{\min\{\mu'_1, \dots, \mu'_s\}} = \frac{-\sum_{r=1}^s \mu'_r \left(\frac{1}{s} \sum_{j \in E'} y_{rj}\right) + \psi'}{\min\{\mu'_1, \dots, \mu'_s\}} = \frac{\frac{1}{s} \sum_{j \in E'} \left(-\sum_{r=1}^s \mu'_r y_{rj} + \psi'\right)}{\min\{\mu'_1, \dots, \mu'_s\}} = 0, \text{ since, by hypothesis, } \sum_{r=1}^s \mu'_r y_{rj} = \psi' \text{ for all } j \in E'.$$

But this is the best result for the objective function in (4), which is possible to obtain due to the first set of constraints in (4). Therefore, the optimal value of the weighted additive model for $\bar{Y}$ is zero and, consequently, $\bar{Y}$ is a Pareto-efficient point in $P(1)$ by Proposition 2.1. (ii) Let us now assume that $\bar{Y}$ is a Pareto-efficient point in $P(1)$. Then, the optimal value of model (4), when $\bar{Y}$ is evaluated with weights $w_r^+ = 1$, $r = 1, \dots, s$, equals zero (Proposition 2.1) and there is an optimal solution (μ^*, ψ^*) such that $\mu_r^* \geq 1 > 0$, $r = 1, \dots, s$. Moreover, this vector defines a hyperplane, $\sum_{r=1}^s \mu_r^* y_r = \psi^*$, which passes on $\bar{Y}$ since the objective function in (4) at (μ^*, ψ^*) is equal to zero, $-\sum_{r=1}^s \mu_r^* \bar{y}_r + \psi^* = 0$, which is equivalent to $\sum_{r=1}^s \mu_r^* \bar{y}_r = \psi^*$. By the first set of constraints in (4), we have that $-\sum_{r=1}^s \mu_r^* y_{rj} + \psi^* \geq 0$, for all $j = 1, \dots, n$. This implies that $-\sum_{r=1}^s \mu_r^* y_{rj} + \psi^* \geq 0$, for all $j \in E'$, which is mathematically equivalent to $\sum_{r=1}^s \mu_r^* y_{rj} \leq \psi^*$. Let us suppose that $\exists j' \in E'$ such that

$$\sum_{r=1}^s \mu_r^* y_{rj'} < \psi^*. \text{ Then, } \sum_{r=1}^s \mu_r^* \bar{y}_r = \sum_{r=1}^s \mu_r^* \left(\frac{1}{s} \sum_{j \in E'} y_{rj}\right) = \frac{1}{s} \sum_{j \in E'} \left(\sum_{r=1}^s \mu_r^* y_{rj}\right) = \frac{1}{s} \left[\underbrace{\sum_{r=1}^s \mu_r^* y_{rj'}}_{< \psi^*} + \sum_{\substack{j \in E' \\ j \neq j'}} \underbrace{\left(\sum_{r=1}^s \mu_r^* y_{rj}\right)}_{\leq \psi^*} \right] < \psi^*. \text{ However, this is a contradiction with the}$$

fact that $\sum_{r=1}^s \mu_r^* \bar{y}_r = \psi^*$. Hence, $\sum_{r=1}^s \mu_r^* y_{rj} = \psi^*$ for all $j \in E'$. This implies that $(\tilde{\mu}^*, \tilde{\psi}^*)$ with $\tilde{\mu}_r^* = \frac{\mu_r^*}{\sum_{r=1}^s \mu_r^*} > 0$ and $\tilde{\psi}^* = \frac{\psi^*}{\sum_{r=1}^s \mu_r^*}$ defines a hyperplane such that

$\sum_{r=1}^s \tilde{\mu}_r^* y_{rj} = \tilde{\psi}^*$, for all $j \in E'$. But then, by hypothesis and Lemma 3.1, $(\tilde{\mu}^*, \tilde{\psi}^*)$ is the unique solution of the system (19). Consequently, $(\tilde{\mu}^*, \tilde{\psi}^*) = (\mu', \psi')$ and $\mu'_r > 0$, for all $r = 1, \dots, s$. $\square$

Proof of Theorem 3.1 First, we are going to prove that given any feasible solution of model (18), then it is associated with a FDEF and the other way round. On the one hand, if $\hat{z}_0$ ⁹ is a feasible solution of (18), then the hyperplane $\sum_{r=1}^s \hat{\mu}_{r0} y_r = \hat{\psi}_0$ is such that (i) $\hat{\mu}_{r0} \geq 0$, $r = 1, \dots, s$ by (10); (ii) $\sum_{r=1}^s \hat{\mu}_{r0} y_r \leq \hat{\psi}_0$ is a valid inequality of $P(1)$ since, by (10), $\sum_{r=1}^s \hat{\mu}_{r0} y_{rj} \leq \hat{\psi}_0$, $\forall j \in E$, which means that $\sum_{r=1}^s \hat{\mu}_{r0} y'_r \leq \hat{\psi}_0$, for all $Y' = \sum_{j \in E} \lambda_j Y_j$, $\sum_{j \in E} \lambda_j = 1$, $\lambda_j \geq 0$, $j \in E \Leftrightarrow \sum_{r=1}^s \hat{\mu}_{r0} y'_r \leq \hat{\psi}_0$, for all $Y' \in P(1)$; (iii) let $\hat{F} = \{Y' \in P(1) : \sum_{r=1}^s \hat{\mu}_{r0} y'_r = \hat{\psi}_0\}$. By (11), there is a subset of E , denoted as $\hat{E}$, such that $\hat{\tau}_{j0} = 1$ and $\hat{d}_{j0} = 0$, which implies that $\sum_{r=1}^s \hat{\mu}_{r0} y_{rj} = \hat{\psi}_0$, for all $j \in \hat{E}$, by (12.1). So, $\hat{F} \neq \emptyset$ and $\hat{F}$ is an efficient face of $P(1)$. Now, it is possible to prove that $\hat{F}$ is a FDEF of $P(1)$ by showing that (a) the set of s extreme efficient points $\{Y_j, j \in \hat{E}\}$ is affinely independent and (b) $\hat{\mu}_{r0} > 0$, for all $r = 1, \dots, s$. On the other hand, if $F' = \{\tilde{Y} \in P(1) : \sum_{r=1}^s \tilde{\mu}_{r0} \tilde{y}_r = \tilde{\psi}_0\} \neq \emptyset$ is a FDEF of $P(1)$, we have to prove that we are able to determine a feasible solution z'_0 of model (18). Let $\sum_{r=1}^s \tilde{\mu}_{r0} y_r = \tilde{\psi}_0$ be a supporting hyperplane associated with F' such that $\tilde{\mu}_r > 0$, $r = 1, \dots, s$. Moreover, there are s affinely independent extreme points, that belong to F' . Let us denote this set of point by E' . Then, we define $\mu'_{r0} = \gamma \tilde{\mu}_{r0} > 0$ and $\psi'_0 = \gamma \tilde{\psi}_0$ with $\gamma = (\max\{s \tilde{\mu}_{10}(y_1^U - y_{10}), \dots, s \tilde{\mu}_{s0}(y_s^U - y_{s0})\})^{-1} > 0$. Additionally, we define $d'_{j0} = \psi'_0 - \sum_{r=1}^s \mu'_{r0} y_{rj}$, $j \in E$, $\delta'_{r0} = \begin{cases} 1, & s \tilde{\mu}_{r0}(y_r^U - y_{r0}) = \gamma^{-1}, r = 1, \dots, s, \\ 0, & \text{otherwise} \end{cases}$, $\tau'_{j0} = \begin{cases} 1, & j \in E' \\ 0, & \text{otherwise} \end{cases}$, $j \in E$, $t'_{r0} = 0$, $r = 1, \dots, s$, $\lambda'_{j0} = \frac{1}{s} \tau'_{j0}$, $j \in E$, and $\alpha'_{jk0} = \alpha'_{jk0}^{\max} = 0$, $j \in E$, $k \in E$. This determines a feasible solution of model (18). Therefore, invoking Proposition 3.1, we have that $\text{NCI}_{\text{EXFA}}^{\min}(Y_0) = 1 + \psi_0^* - \sum_{r=1}^s \mu_{r0}^* y_{r0}$. $\square$

⁹ This is the way of representing a vector of all decision variables of problem (19) with a zero as subscript and a hat over them.

B: Comparison of DEA Measures in the Context of BoD Models Regarding the Satisfaction of the Properties [P1]–[P6]

Measure and Reference	Pareto efficiency [P1]	Units invariance [P2]	Translation invariance [P3]	Strong monotonicity [P4]	Weights > 0 [P5]	Closest targets [P6]
Radial model [1, 2]	Not always	Yes	Not always	Not always	Not always	Not always
Directional distance function [3]	Not always	Yes ^a	Yes ^c	Not always	Not always	Not always
Weighted additive [5, 6]	Yes	Yes ^b	Yes ^d	Yes	Yes	Not always
Russell-oriented measures [4]	Yes	Yes	Not always	Yes	Yes	Not always
Slacks-Based Measure [7, 8]	Yes	Yes	Not always	Yes	Yes	Not always
The new approach	Yes	Yes	Yes	Yes	Yes	Yes

^aOnly if the selected directional vector is units-invariant

^bOnly if the selected weights are units-invariant

^cOnly if the selected directional vector is translation-invariant

^dOnly if the selected weights are translation-invariant

C: Composite Indicators Based on the Radial BoD Model for 25 Selected DMUs, with $k = 1$ Added, Targets Are Reported in Parentheses

DMU	Indicator	Targets				
		Output 1	Output 2	Output 3	Output 4	Output 5
1	1.00	0.00 (0.00)	1.79 (1.79)	2.24 (2.24)	0.00 (0.00)	0.28 (0.28)
2	1.39	1.00 (1.39)	1.90 (2.64)	1.00 (1.39)	1.00 (1.39)	1.00 (1.39)
3	1.65	1.00 (1.65)	1.00 (1.65)	1.75 (2.88)	1.00 (1.65)	1.20 (1.98)
4	1.00	0.00 (0.00)	0.90 (0.90)	0.00 (0.00)	1.60 (1.60)	1.00 (1.00)
5	1.00	0.00 (0.00)	1.79 (1.79)	0.56 (0.56)	1.79 (1.79)	0.84 (0.84)
6	1.00	0.00 (0.00)	0.45 (0.45)	0.56 (0.56)	1.79 (1.79)	1.12 (1.12)
12	1.39	1.00 (1.39)	1.90 (2.64)	1.00 (1.39)	1.00 (1.39)	1.00 (1.39)
13	1.36	1.00 (1.36)	1.90 (2.59)	1.75 (2.39)	2.60 (3.54)	1.00 (1.36)
18	1.00	0.00 (0.00)	1.80 (1.80)	0.75 (0.75)	0.00 (0.00)	0.00 (0.00)
19	1.67	1.00 (1.67)	1.00 (1.67)	1.00 (1.67)	1.00 (1.67)	1.00 (1.67)
23	1.00	0.00 (0.00)	1.79 (1.79)	2.24 (2.24)	1.79 (1.79)	0.84 (0.84)
25	1.00	1.66 (1.66)	0.00 (0.00)	1.50 (1.50)	4.00 (4.00)	1.00 (1.00)
27	1.67	1.00 (1.67)	1.00 (1.67)	1.00 (1.67)	1.00 (1.67)	1.00 (1.67)
28	1.67	1.00 (1.67)	1.00 (1.67)	1.00 (1.67)	1.00 (1.67)	1.00 (1.67)
30	1.33	1.00 (1.34)	1.90 (2.54)	1.00 (1.34)	2.60 (3.47)	1.00 (1.34)
35	1.00	0.00 (0.00)	0.43 (0.43)	1.07 (1.07)	0.00 (0.00)	1.07 (1.07)
38	1.00	4.14 (4.14)	0.00 (0.00)	3.00 (3.00)	0.00 (0.00)	0.00 (0.00)
41	1.00	0.00 (0.00)	0.90 (0.90)	2.24 (2.24)	0.00 (0.00)	1.12 (1.12)
42	1.00	0.00 (0.00)	1.70 (1.70)	2.13 (2.13)	0.00 (0.00)	0.21 (0.21)
44	1.00	1.34 (1.34)	1.63 (1.63)	2.04 (2.04)	1.63 (1.63)	0.76 (0.76)
48	1.02	1.00 (1.02)	1.85 (1.89)	3.13 (3.20)	1.00 (1.02)	2.07 (2.11)
55	1.00	0.00 (0.00)	1.80 (1.80)	0.00 (0.00)	1.60 (1.60)	1.00 (1.00)
56	1.00	0.00 (0.00)	3.60 (3.60)	0.75 (0.75)	1.60 (1.60)	0.00 (0.00)
57	1.13	2.34 (2.66)	1.43 (1.61)	1.53 (1.74)	2.70 (3.06)	1.27 (1.44)
58	1.11	2.34 (2.61)	1.81 (2.02)	2.02 (2.25)	2.63 (2.92)	1.00 (1.11)

References

1. Charnes, A., Cooper, W.W., Rhodes, E.: Measuring the efficiency of decision making units. *Eur. J. Oper. Res.* **2**(6), 429–444 (1978)
2. Banker, R.D., Charnes, A., Cooper, W.W.: Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis. *Manag. Sci.* **30**, 1078–1092 (1984)
3. Chambers, R.G., Chung, Y., Färe, R.: Profit, directional distance functions, and Nerlovian efficiency. *J. Optim. Theory Appl.* **98**(2), 351–364 (1998)
4. Färe, R., Grosskopf, S., Lovell, C.A.K.: *The Measurement of Efficiency of Production*. Kluwer Academic Publishers, Boston (1985)
5. Lovell, C.A.K., Pastor, J.T.: Units invariant and translation invariant DEA models. *Oper. Res. Lett.* **18**(3), 147–151 (1995)

6. Aparicio, J., Pastor, J.T., Vidal, F.: The weighted additive distance function. *Eur. J. Oper. Res.* **254**(1), 338–346 (2016)
7. Pastor, J.T., Ruiz, J.L., Sirvent, I.: An enhanced DEA Russell graph efficiency measure. *Eur. J. Oper. Res.* **115**, 596–607 (1999)
8. Tone, K.: A slacks-based measure of efficiency in data envelopment analysis. *Eur. J. Oper. Res.* **130**(3), 498–509 (2001)
9. Aparicio, J., Ruiz, J.L., Sirvent, I.: Closest targets and minimum distance to the Pareto-efficient frontier in DEA. *J. Prod. Anal.* **28**, 209–218 (2007)
10. Aparicio, J., Mahlberg, B., Pastor, J.T., Sahoo, B.K.: Decomposing technical inefficiency using the principle of least action. *Eur. J. Oper. Res.* **239**(3), 776–785 (2014)
11. Cherchye, L., Lovell, C.K., Moesen, W., Van Puyenbroeck, T.: One market, one number? A composite indicator assessment of EU internal market dynamics. *Eur. Econ. Rev.* **51**(3), 749–779 (2007)
12. Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T.: An introduction to ‘benefit of the doubt’ composite indicators. *Soc. Indic. Res.* **82**, 111–145 (2007)
13. Cherchye, L., Moesen, W., Van Puyenbroeck, T.: Legitimately diverse, yet comparable: on synthesizing social inclusion performance in the EU. *J. Common Mark. Stud.* **42**, 919–955 (2004)
14. Lovell, C.A.K.: Measuring the macroeconomic performance of the Taiwanese economy. *Int. J. Prod. Econ.* **39**, 165–178 (1995)
15. Sahoo, B.K., Acharya, D.: Constructing macroeconomic performance index of Indian states using DEA. *J. Econ. Stud.* **39**(1), 63–83 (2012)
16. Shen, Y., Hermans, E., Brijs, T., Wets, G.: Data Envelopment Analysis for composite indicators: A multiple layer model. *Soc. Indic. Res.* **114**, 739–756 (2013)
17. Fusco, E.: Enhancing non-compensatory composite indicators: a directional proposal. *Eur. J. Oper. Res.* **242**, 620–630 (2015)
18. Sahoo, B.K., Singh, R., Mishra, B., Sankaran, K.: Research productivity in management schools of India during 1968–2015: a directional Benefit-of Doubt model analysis. *Omega* **66**, 118–139 (2017)
19. Zanella, A., Camanho, A.S., Dias, T.G.: Undesirable outputs and weighting schemes in composite indicators based on data envelopment analysis. *Eur. J. Oper. Res.* **245**, 517–530 (2015)
20. Vidoli, F., Fusco, E., Mazziotta, C.: Non-compensability in composite indicators: a robust directional frontier method. *Soc. Indic. Res.* **122**, 635–652 (2015)
21. Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T.: Constructing composite indicators with imprecise data: a proposal. *Expert Syst. Appl.* **38**, 10940–10949 (2011)
22. Fusco, E., Vidoli, F., Sahoo, B.K.: Spatial heterogeneity in composite indicator: a methodological proposal. *Omega* **77**, 1–14 (2018)
23. Fusco, E., Vidoli, F., Rogge, N.: Spatial directional robust Benefit of the Doubt approach in presence of undesirable output: an application to Italian waste sector. *Omega* **94**, 102053 (2020)
24. Aparicio, J., Kapelko, M.: Enhancing the measurement of composite indicators of corporate social performance. *Soc. Indic. Res.* **114**(2), 807–826 (2019)
25. Cooper, W.W., Park, K.S., Pastor, J.: RAM: a range adjusted measure of inefficiency for use with additive models, and relations to other models and measures in DEA. *J. Prod. Anal.* **11**, 5–42 (1999)
26. Färe, R., Lovell, C.A.K.: Measuring the technical efficiency of production. *J. Econ. Theory* **19**, 150–162 (1978)
27. Coelli, T.: A multi-stage methodology for the solution of orientated DEA models. *Oper. Res. Lett.* **23**(3–5), 143–149 (1998)
28. Briec, W.: Hölder distance functions and measurement of technical efficiency. *J. Prod. Anal.* **11**, 111–131 (1998)
29. Frei, F.X., Harker, P.T.: Projections onto efficient frontiers: theoretical and computational extensions to DEA. *J. Prod. Anal.* **11**(3), 275–300 (1999)
30. Cherchye, L., Van Puyenbroeck, T.: A comment on multi-stage DEA methodology. *Oper. Res. Lett.* **28**(2), 93–98 (2001)
31. González, E., Álvarez, A.: From efficiency measurement to efficiency improvement: the choice of a relevant benchmark. *Eur. J. Oper. Res.* **133**(3), 512–520 (2001)
32. Portela, M.C.A.S., Borges, P.C., Thanassoulis, E.: Finding closest targets in non-oriented DEA models: the case of convex and non-convex technologies. *J. Prod. Anal.* **19**, 251–269 (2003)
33. Pastor, J.T., Aparicio, J.: The relevance of DEA benchmarking information and the least-distance measure: comment. *Math. Comput. Model.* **52**(1–2), 397–399 (2010)

34. Aparicio, J., Pastor, J.T.: A well-defined efficiency measure for dealing with closest targets in DEA. *Appl. Math. Comput.* **219**(17), 9142–9154 (2013)
35. Aparicio, J., Pastor, J.T.: Closest targets and strong monotonicity on the strongly efficient frontier in DEA. *Omega* **44**, 51–57 (2014)
36. Fukuyama, H., Leth Hougaard, J., Sekitani, K., Shi, J.: Efficiency measurement with a non-convex free disposal hull technology. *J. Oper. Res. Soc.* **67**(1), 9–19 (2016)
37. Aparicio, J.: A survey on measuring efficiency through the determination of the least distance in data envelopment analysis. *J. Cent. Cathedra* **9**(2), 143–167 (2016)
38. Ando, K., Kai, A., Maeda, Y., Sekitani, K.: Least distance based inefficiency measures on the Pareto-efficient frontier in DEA. *J. Oper. Res. Soc. Jpn.* **55**(1), 73–91 (2012)
39. Fukuyama, H., Maeda, Y., Sekitani, K., Shi, J.: Input–output substitutability and strongly monotonic p-norm least distance DEA measures. *Eur. J. Oper. Res.* **237**(3), 997–1007 (2014)
40. Bessent, A., Bessent, W., Elam, J., Clark, T.: Efficiency frontier determination by constrained facet analysis. *Oper. Res.* **36**, 785–796 (1988)
41. Green, R.H., Doyle, J.R., Cook, W.D.: Efficiency bounds in Data Envelopment Analysis. *Eur. J. Oper. Res.* **89**, 482–490 (1996)
42. Olesen, O.B., Petersen, N.C.: Indicators of ill-conditioned data sets and model misspecification in Data Envelopment Analysis: an extended facet approach. *Manag. Sci.* **42**, 205–219 (1996)
43. Räty, T.: Efficient facet based efficiency index: a variable returns to scale specification. *J. Prod. Anal.* **17**, 65–82 (2002)
44. Koopmans, T.C.: Analysis of production as an efficient combination of activities. In: Koopmans, T.C. (ed.) *Activity Analysis of Production and Allocation*. Wiley, New York (1951)
45. Nemhauser, G.L., Wolsey, L.A.: *Integer and Combinatorial Optimization*. Wiley, New York (1999)
46. Cooper, W.W., Pastor, J.T., Borras, F., Aparicio, J., Pastor, D.: BAM: a bounded adjusted measure of efficiency for use with bounded additive models. *J. Prod. Anal.* **35**(2), 85–94 (2011)
47. Singer, I.: *Duality for Nonconvex Approximation and Optimization*. Springer, Boston, MA (2007)
48. Wu, D.D.: Bilevel programming Data Envelopment Analysis with constrained resource. *Eur. J. Oper. Res.* **207**, 856–864 (2010)
49. Whittaker, G., Färe, R., Grosskopf, S., Barnhart, B., Bostian, M., Mueller-Warrant, G., Griffith, S.: Spatial targeting of agri-environmental policy using bilevel evolutionary optimization. *Omega* **66**, 15–27 (2017)
50. Shi, C., Lu, J., Zhang, G.: An extended Kuhn-Tucker approach for linear bilevel programming. *Appl. Math. Comput.* **162**(1), 51–63 (2005)
51. Beale, E.M.L., Tomlin, J.A.: Special facilities in a general mathematical programming system for non-convex problems using ordered sets of variables. In: Lawrence, J. (ed.) *Proceedings of the Fifth International Conference on Operational Research*, pp. 447–454. Tavistock Publications, London (1970)
52. Cooper, W.W., Ruiz, J.L., Sirvent, I.: Choosing weights from alternative optimal solutions of dual multiplier models in DEA. *Eur. J. Oper. Res.* **180**(1), 443–458 (2007)
53. Mehdiloozad, M., Mirdehghan, S.M., Sahoo, B.K., Roshdi, I.: On the identification of the global reference set in Data Envelopment Analysis. *Eur. J. Oper. Res.* **245**(3), 779–788 (2015)
54. Mehdiloozad, M., Zhu, J., Sahoo, B.K.: Identification of congestion in Data Envelopment Analysis under the occurrence of multiple projections: a reliable method capable of dealing with negative data. *Eur. J. Oper. Res.* **265**(2), 644–654 (2018)
55. Engida, T.G., Rao, X., Berentsen, P.B.M., Oude Lansink, A.: Measuring corporate sustainability performance—the case of European food and beverage companies. *J. Clean. Prod.* **195**, 734–743 (2018)
56. Auer, B.R.: Do socially responsible investment policies add or destroy European stock portfolio value? *J. Bus. Ethics* **135**, 381–397 (2016)
57. Surroca, J., Tribó, J.A., Waddock, S.: Corporate responsibility and financial performance: the role of intangible resources. *Strateg. Manag. J.* **31**, 463–490 (2010)
58. Maloni, M.J., Brown, M.E.: Corporate social responsibility in the supply chain: an application in the food industry. *J. Bus. Ethics* **68**, 35–52 (2006)
59. Golany, B., Roll, Y.: An application procedure for DEA. *Omega* **17**, 237–250 (1989)
60. Nunamaker, T.R.: Using data envelopment analysis to measure the efficiency of non-profit organizations: a critical evaluation. *Manag. Decis. Econ.* **6**, 50–58 (1985)
61. Raab, R.L., Lichty, R.W.: Identifying subareas that comprise a greater metropolitan area: the criterion of county relative efficiency. *J. Reg. Sci.* **42**, 579–594 (2002)

62. Dyson, R.G., Allen, R., Camanho, A.S., Podinovski, V.V., Sarrico, C.S., Shale, E.A.: Pitfalls and protocols in DEA. *Eur. J. Oper. Res.* **132**, 245–259 (2001)
63. Cooper, W.W., Seiford, L.M., Tone, K.: *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. Springer, New York (2007)
64. Wilson, P.: Dimension reduction in nonparametric models of production. *Eur. J. Oper. Res.* **267**, 349–367 (2018)
65. CPLEX Optimizer User Manual (2013). IBM ILOG. <https://www.ibm.com/products/ilog-cplex-optimization-studio>. Accessed 30 May 2020
66. Lozano, S., Villa, G.: Gradual technical and scale efficiency improvement in DEA. *Ann. Oper. Res.* **173**, 123–136 (2010)
67. Nguyen, P.-A., Kecskés, A., Mansi, S.: Does corporate social responsibility create shareholder value? The importance of long-term investors. *J. Bank. Finance* **112**, 105217 (2020)
68. Gleick, P.H., Allen, L., Christian-Smith, J., Cohen, M.J., Cooley, H., Heberger, M., Morrison, J., Palaniappan, M., Schulte, P.: *The World's Water Volume 7: The Biennial Report on Freshwater Resources*. Island Press, Washington (2011)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.