ORIGINAL RESEARCH

CrossMark

Enhancing the Measurement of Composite Indicators of Corporate Social Performance

Juan Aparicio[1] · Magdalena Kapelko[2]

© The Author(s) 2019

Abstract
The aggregation of multiple metrics of corporate social performance (CSP) poses one of the major problems in CSP research. In this paper we propose a new method to compute a composite indicator of CSP from an efficiency perspective using data envelopment analysis. The new approach makes it possible to appropriately account for both positive and negative aspects of CSP, and also to measure CSP for firms for which some of the metrics are exhibiting zero values. Within the new method, we propose measuring global (meta-frontier) CSP inefficiency with regard to all firms in the sample and decomposing it into within-industry CSP inefficiency measured with regard to own industry group and industry gap CSP inefficiency. In this way, we measure inefficiencies due to shortcomings in managerial practices and due to the differences between industries in relation to CSP. Using a sample of United States firms in a variety of economic sectors from 2004 to 2015, as covered by the Kinder, Lydenberg and Domini database, we find considerable potential for improvement in CSP practices, with the most CSP-efficient firms representing mining and manufacturing sectors, and the most inefficient belonging to the construction sector. We also find that CSP inefficiency dropped in the period related to the recent financial crisis.

Keywords Corporate social responsibility · Data envelopment analysis · Directional distance function · Benefit of the doubt · Bad outputs

1 Introduction

Corporate social responsibility (CSR) is a concept that has attracted a great deal of attention in recent decades from practitioners and researchers across different disciplines. CSR consists of "discretionary responsibilities" of a business (Carroll 1979) and relates to firms' actions that promote social good beyond economic interest of the firm or shareholders and beyond legal requirements (McWilliams and Siegel 2001). In

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

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

2   Department of Logistics, Institute of Applied Mathematics, Wrocław University of Economics,
    Wrocław, Poland

general, CSR activities and outcomes are predictors of corporate social performance (CSP) (Waddock and Graves 1997), which consists of "a broad array of strategies and operating practices that a company develops in its effort to deal with and create relationships with its numerous stakeholders and the natural environment" (Surroca et al. 2010).

One of the major problems in the research on CSR and CSP lies in the construction of the appropriate overall measure of the complex CSP concept (Chen and Delmas 2011; Surroca et al. 2010; Waddock and Graves 1997). Because CSP encompasses a broad range of activities, multiple metrics are required in order to reflect its full scope. Researchers have used these metrics from a variety of sources, including questionnaires, annual reports, and dedicated databases created by professional companies such as Kinder, Lyndenberg, and Domini (KLD), Sustainalytics, or Thomson Reuters. However, the aggregation of these metrics is challenging. Most existing studies have used simple aggregates by summing up the scores on different metrics (dimensions of CSP) by assigning equal weights to these metrics (for example, Flammer 2015; Servaes and Tamayo 2013; Siegel and Vitaliano 2007) or introducing subjective weighting based on the views of a panel of experts (for example, Ruf et al. 1998; Waddock and Graves 1997).

Another stream of literature proposes using data envelopment analysis (DEA) (Banker et al. 1984; Charnes et al. 1978) to aggregate multiple CSP metrics and derive the composite measures of CSP from an efficiency perspective. DEA is based on mathematical programming and makes it possible to assess the relative efficiencies of decision making units (DMUs) with multiple inputs and multiple outputs. The main advantage of DEA is that it does not require a priori weights to aggregate different CSP metrics as weights are derived directly from the data, thus allowing for an objective weighting. The use of DEA to aggregate CSP metrics was initiated by Bendheim et al. (1998), who applied the so-called benefit of the doubt (BoD) DEA model or the model without inputs (Cherchye et al. 2004, 2007a, b; Lovell and Pastor 1997, 1999; Lovell 1995) to derive a composite indicator of CSP. The use of BoD in the CSP context was further extended in the studies by Belu and Manescu (2013) and Chen and Delmas (2011). However, the CSP research also used a standard DEA model considering both inputs and outputs, with examples being the studies by Lee and Saen (2012) and Belu (2009).

In the present study, we build on the aforementioned research that used DEA to assess CSP and propose a further extension of the methodology to aggregate CSP metrics in the DEA framework that makes it possible to tackle some of the issues pertaining to CSP measures. Firstly, CSP metrics usually contain both positive and negative aspects to represent strengths and concerns of CSP activities of firms. An example is the case of the KLD database. An example of a positive aspect of CSP includes investing in product and services that address the issues of resource conservation and climate change, while examples of a negative aspect could be paying a settlement, fine, or penalty for non-compliance with environmental regulations. In order to appropriately account for both positive and negative aspects of CSP, our method uses the approach developed by Chung et al. (1997) that measures efficiency asymmetrically modeling the two types of outputs; that is, increasing intended ("good") outputs while reducing undesirable ("bad") outputs. In our context, the proposed method increases good CSP practices and decreases negative CSP practices in order to measure the inefficiency of CSP. Chung et al.'s (1997) approach is based on the directional distance function (DDF) of Chambers et al. (1996, 1998) in the DEA framework, which considers both inputs and outputs in the model. However, as CSP metrics are usually output measures, we apply the DEA model without inputs—that is, the BoD model

in the directional distance function context as proposed by Fusco (2015)—and we further modify it to include "bad" outputs (CSP concerns).[1]

Secondly, it is often the case in practice that CSP metrics can present zero values for some firms. In general, traditional DEA models are designed for positive input or output values, which means they cannot be applied in this context. Only models that are characterized by the property of translation invariance (that is, those that are independent of affine translation of the input and output variables) are suitable in this case (Pastor and Aparicio 2015; Lovell and Pastor 1995). As a standard directional BoD model cannot deal with zero values for outputs, we propose a further extension of it by using the directional distance function model of Portela et al. (2004), which is translation-invariant following Aparicio et al. (2016).

Thirdly, the CSP metrics are created for a wide range of industries and it is important to be able to appropriately compare the CSP aggregated measures between different sectors. This is especially crucial when creating CSP measures from an efficiency perspective since not all inefficiency measures can be directly compared between them due to the possible existence of multiple different technologies. In order to account for this issue, we use the concept of a metafrontier (O'Donnell et al. 2008; Battese et al. 2004) and propose computing CSP inefficiency both with regard to a metafrontier composed of all observations in the sample and with regard to an industry-specific frontier, and finally calculating the difference and gap between both inefficiencies.[2]

The originality of this study consists of the development of a new approach to the measurement of CSP and aggregation of CSP metrics through DEA that makes it possible to appropriately account for both positive and negative aspects of CSP, as well as to measure CSP for firms that do not have some of the metrics presented (that is, exhibiting zero values). Overall, these are advantageous characteristics and the new approach can serve as a tool for academics and practitioners to compute CSP inefficiency for all firms covered by datasets. In addition, the proposed approach can be used to measure CSP inefficiency and compare it between firms representing different technologies as embedded in the industries. The resulting approach makes it possible to measure global CSP inefficiency and decompose it into within-industry CSP inefficiency and industry gap CSP inefficiency. This allows for a deeper understanding of differences in CSP inefficiency between industries. In particular, the approach makes it possible to identify the main source of CSP inefficiency in each sector, that is how much is caused by shortcomings in managerial practices and how much by the differences between industries in relation to CSP. This can provide further managerial and policy implications on how to improve CSP efficiency in different industries. The empirical application of the proposed approach uses the KLD dataset of CSP metrics for a large sample of United States firms in a variety of sectors from 2004 to 2015.

The remainder of the paper is structured as follows. In Sect. 2 we describe in details our proposal for the measurement of CSP and aggregation of CSP metrics. In Sect. 3 we explain the dataset of CSP metrics, and in Sect. 4 we present the results on CSP inefficiency using the proposed approach. Section 5 concludes, presenting the implications of results, limitations, and future research areas.

---

[1] Other extensions of BoD model are developed, for example, in Rogge (2018), Van Puyenbroeck and Rogge (2017), Vidoli et al. (2015) or Shen et al. (2013).

[2] The initial developments regarding these concepts date back to Charnes et al. (1981), who proposed distinguishing between managerial efficiency and program efficiency.

2 Measuring CSP—Our Proposal

In this section, we introduce the main notation and the methodology used in our data analysis. Our approach is based on aggregating CSP metrics through DEA in order to obtain an overall measurement of CSP.

DEA is a mathematical programming, non-parametric technique that is commonly used to measure the relative performance of a set of homogeneous processing units, which use several inputs to produce several outputs (Banker et al. 1984; Charnes et al. 1978). These operating units are usually called decision making units (DMUs) in recognition of their autonomy in setting their input and output levels. DEA allows one to determine a technical efficiency score for each assessed unit in the sample data by estimating a frontier that envelops the data from above.

DEA includes a battery of measures with plenty of possibilities for determining technical efficiency. Examples include Shephard's input and output distance functions (Shephard 1953), the Weighted Additive Model (Lovell and Pastor 1995), the Directional Distance Function (DDF) (Chambers et al. 1996, 1998), the Range-Adjusted Measure (Cooper et al. 1999), and the Enhanced Russell Graph/Slacks-Based Measure (Tone 2001; Pastor et al. 1999). The main difference among these is the way of implementing the notion of distance from each observation (DMU) to the estimated frontier of best practices. In particular, a measure that has attracted special attention in the literature is the DDF, which implies a directional vector that allows for a flexible choice of orientation in the output and input dimensions in order to project each DMU onto the frontier. Additionally, in a DEA framework, the DDF may be easily determined from a computational point of view since, in this case, researchers, analysts, practitioners, and scholars only need to resort to well-known and easy-to-use linear programming techniques.

DEA has played an important role recently in the context of building composite indicators. In particular, a related relevant approach based on DEA is the benefit of the doubt (BoD) model (see Cherchye et al. 2004, 2007a, b). This is a DEA model without inputs and where subindicators are all treated as outputs to generate an overall and objective aggregated indicator for each assessed DMU. In our framework, we collect information about a number of positive subindicators, or "good" outputs (CSP strengths), and negative subindicators, or "bad" outputs (CSP concerns) of the firms, rather than just positive subindicators, as is usually the case. Consequently, our approach for measuring CSP performance through the aggregation of CSP metrics must take this fact into account. The DDF has been one of the most used measures for dealing with undesirable ("bad") outputs in DEA since the publication of the paper by Chung et al. (1997). Those authors showed how to adapt the original DDF for working in the context of pollution, where firms produce intended ("good") outputs and unintended ("bad") outputs from a set of resources. The Chung et al. (1997) version of the DDF specifically seeks the largest feasible increase in desirable outputs compatible with a simultaneous reduction in undesirable outputs. This type of projection is what we are seeking for our application on CSP; that is, an increase in the positive aspects of CSP (strengths) and a decrease in negative aspects of CSP (concerns). In this way, it seems that a natural tool for building a composite indicator from CSP subindicators through a DEA-BoD model is the DDF, which also considers desirable and undesirable (output) metrics.

Another challenge that faces our approach is developing a composite indicator from CSP metrics that also allows us to appropriately account for firms with some metrics that may exhibit zero values, which is often the case of CSP datasets. Historically speaking,

Charnes et al. (1978) defined the first DEA model, the so-called CCR model, which is a constant returns to scale (CRS) radial measure, requiring strict positivity of all the input and output values. Ali and Seiford (1990) relaxed this requirement, but avoided the use of the CCR model. Specifically, they proposed considering an additive model, instead of a radial one, under variable returns to scale (VRS), showing that it is possible to deal with DMUs with all their inputs at level zero and/or all their outputs at level zero (something that happens for some firms in many of CSP datasets for CSP strengths or concerns). Ali and Seiford (1990) suggested considering a derived set of DMUs to be rated, where each derived unit is obtained by translating the original data. If the original DEA problem and the corresponding translated problem are equivalent—that is, if both problems rate each original unit and its corresponding translated unit in exactly the same way—then it is said that the corresponding model satisfies the translation invariance property. Another relevant contribution to the literature on DEA and translation-invariant models is that by Lovell and Pastor (1995), who concluded that the key to satisfying translation invariance is the VRS assumption. However, the same authors showed that the (radial) BCC model by Banker et al. (1984), which assumes VRS, is only partially translation-invariant: the input-oriented BCC model is invariant to output translations, whereas the output-oriented BCC model is invariant to input translations. Consequently, radial models, even under VRS, are not the solution in the context of working with zeros. More recently, Aparicio et al. (2016) analyzed the satisfaction of the translation invariance property in the case of the directional distance function, concluding that the DDF is translation-invariant if and only if the prefixed directional vector is also translation-invariant. This means that the most usual directional vector, which is that associated with the actual values of the evaluated DMU, is not translation-invariant and, therefore, neither its related DDF. Fortunately, Aparicio et al. (2016) also showed some examples of translation-invariant directional distance functions, such as the Range Directional Model (RDM) by Portela et al. (2004), which is additionally unit-invariant. In particular, the satisfaction of these good properties by the RDM are the reasons why we select this measure to be adapted to our context of the measurement of CSP where we jointly consider DEA, the BoD model, the directional distance function, the treatment of undesirable outputs, and the presence of zeros in the data.

Before introducing the specific mathematical model to be solved in this paper, it is worth mentioning that, as far as we are aware, there are no contributions in the literature that combine the DDF and the BoD and adapt them to deal with good and bad outputs and zeros. Nevertheless, a few recent papers have worked with a directional-benefit-of-doubt model. These are, for example, Fusco (2015), where a principal component analysis is applied to choose the directional vector, and Sahoo et al. (2017), who developed a composite indicator of research productivity of Indian management schools. However, neither of these papers focus on adapting the corresponding model to deal with bad outputs and the presence of zeros, as we do.

Additionally, and regarding the method of modeling bad outputs in DEA, it is worth mentioning that the main contributions draw on how the axioms underlying the production technology should reflect strong or weak disposability, and ultimately, whether they should be modeled as outputs or as inputs. Dakpo et al. (2016) is an updated revision of the most recent contributions on how to characterize undesirable production based on alternative approaches, which include the standard approach by Chung et al. (1997) based on the directional distance function; the materials balance principles requiring knowledge of the technical coefficients between desirable outputs, undesirable outputs and inputs (Hampf and Rødseth 2015); and the use of two sub-technologies (by-production), one generating the desirable outputs and a second generating the undesirable outputs (Førsund 2009;

Murty et al. 2012). We have adopted the Chung et al. approach because it is the most used model in the literature; while we recognize that other alternatives exist, comparing them is beyond the scope of this paper.

Let us now assume that we have observed M non-negative strengths (denoted as y) and I non-negative concerns (denoted as b) for each DMU in a sample of size K^h of units belonging to industry h, $h = 1, \ldots, H$. Mathematically, a particular observation j of firm j belonging to industry h is denoted as $\left(y_{j1}^h, \ldots, y_{jM}^h, b_{j1}^h, \ldots, b_{jI}^h\right)$.

Following this notation, the model to be solved in order to aggregate strengths and concerns in a final composite indicator when DMU_0 is evaluated with respect to its industry-specific frontier is as follows:

$$\overrightarrow{D}_0^h\left(y_0^h, b_0^h; g_0^y, g_0^b\right) = Max \ \beta \tag{1.1}$$
$$s.t.$$

$$\sum_{k=1}^{K^h} z_k y_{km}^h \geq y_{0m}^h + \beta g_{0m}^y, \quad m = 1, \ldots, M \tag{1.2}$$

$$\sum_{k=1}^{K^h} z_k b_{ki}^h = b_{0i}^h - \beta g_{0i}^b, \quad i = 1, \ldots, I \tag{1.3}$$

$$\sum_{k=1}^{K^h} z_k = 1, \tag{1.4}$$

$$z_n \geq 0, \quad\quad\quad k = 1, \ldots, K^h \tag{1.5}$$

$$(1)$$

where the directional vector $g = \left(g_0^y, g_0^b\right)$ is defined as $g_{0m}^y = \max\limits_{1 \leq h \leq H}\left\{\max\limits_{1 \leq k \leq K^h}\left\{y_{km}^h\right\}\right\} - y_{0m}^h$, $m = 1, \ldots, M$, and $g_{0i}^b = b_{0i}^h - \min\limits_{1 \leq h \leq H}\left\{\min\limits_{1 \leq k \leq K^h}\left\{b_{ki}^h\right\}\right\}$, $i = 1, \ldots, I$. This vector is, therefore, in line with the RDM by Portela et al. (2004), where an ideal point is defined from the data (with coordinates corresponding to the maximum value observed for outputs and the minimum value observed for inputs), which is used as a reference point in the projection of the assessed DMUs onto the DEA frontier. By analogy, our directional vector projects each unit towards an ideal point with coordinates corresponding to the maximum value observed for CSP strengths and the minimum value observed for CSP concerns for all the industries. Model (1) is an adaptation of Chung et al.'s (1997) directional distance function for dealing with good and bad outputs together, when inputs are not considered, so following the BoD approach, and using as a directional vector a variation of the directional vector used in the case of the RDM. Moreover, in order to estimate the inefficiency for all observations in the sample, model (1) must be used for each industry separately.

We also use the concept of a metafrontier by O'Donnell et al. (2008) to compare the performance of firms that may be classified into different groups, in our case in different industries. We next present how a metafrontier can be estimated using non-parametric techniques and show how to decompose the 'pooled' CSP inefficiency into within-industry CSP inefficiency and CSP industry gap.

In order to make inefficiency comparisons across groups of firms, we build a common frontier (the metafrontier), defined as the boundary that envelops all industry-specific frontiers. In this way, inefficiencies measured relative to the metafrontier can be decomposed into two components: one that measures the distance from a DMU to its industry-specific frontier (model (1) above) and one that measures the distance between the industry-specific

frontier and the metafrontier. This last term may be interpreted as representing the restrictive nature of the production environment. The initial idea of decomposing inefficiency into these two sources comes from Charnes et al. (1981), who distinguished between managerial efficiency measured in relation to firm's group-specific frontier and program efficiency that measures the difference between the group-specific frontier and the frontier composed of all observations. Using these ideas, in our context we can explore the inefficiencies due to shortcomings managerial practices and due to the differences between industries in relation to CSP.

Model (2) shows how to determine the distance between the DMU_0 belonging to industry h to the metafrontier, defined from all the observations for all the industries:

$$\overrightarrow{D}_0^{1,\ldots,H}\left(y_0^h, b_0^h; g_0^y, g_0^b\right) = Max \qquad \beta \qquad (2.1)$$

$$s.t.$$

$$\sum_{p=1}^{H} \sum_{k=1}^{K^p} z_k^p y_{km}^p \geq y_{0m}^h + \beta g_{0m}^y, \quad m = 1, \ldots, M \qquad (2.2)$$

$$\sum_{p=1}^{H} \sum_{k=1}^{K^p} z_k^p b_{ki}^p = b_{0i}^h - \beta g_{0i}^b, \quad i = 1, \ldots, I \qquad (2.3)$$

$$\sum_{p=1}^{H} \sum_{k=1}^{K^p} z_k^p = 1 \qquad (2.4)$$

$$z_k^p \geq 0, \qquad k = 1, \ldots, K^h, \; p = 1, \ldots, h, \ldots H \quad (2.5)$$

$$(2)$$

While h is fixed in $\overrightarrow{D}_0^{1,\ldots,H}(y_0^h, b_0^h; g_0^y, g_0^b)$, which denotes the distance for DMU_0 that belongs to a specific industry h to the metafrontier, superscript p, used in z_k^p, y_{km}^p and K^p in model (2), is variable ($p = 1, \ldots, h, \ldots H$). We use p in the constraints of model (2) so that y_{0m}^h and b_{ki}^h, where h is fixed, are not confused with y_{0m}^p and b_{ki}^p, with $p = 1, \ldots, h, \ldots H$.

Finally, the 'pooled' CSP inefficiency, $\overrightarrow{D}_0^{1,\ldots,H}(y_0^h, b_0^h; g_0^y, g_0^b)$, may be decomposed into 'within-industry' CSP inefficiency, $\overrightarrow{D}_0^h(y_0^h, b_0^h; g_0^y, g_0^b)$, plus a non-negative residual term, related to CSP 'industry gap':

$$\overrightarrow{D}_0^{1,\ldots,H}(y_0^h, b_0^h; g_0^y, g_0^b) = \overrightarrow{D}_0^h(y_0^h, b_0^h; g_0^y, g_0^b) + GAP \qquad (3)$$

It is possible to obtain the projection point corresponding to the resolution of model (1) when unit (y_0^h, b_0^h) is evaluated with respect to the frontier of industry h. In a DDF setting, it is well known that the projection point is $(y_0^{h*}, b_0^{h*}) :=$ $(y_0^h + \overrightarrow{D}_0^h(y_0^h, b_0^h; g_0^y, g_0^b)g_0^y, b_0^h - \overrightarrow{D}_0^h(y_0^h, b_0^h; g_0^y, g_0^b)g_0^b)$. It can then be proved that the CSP industry gap coincides with the directional distance function associated with point (y_0^{h*}, b_0^{h*}) when the metafrontier is used as a reference; that is, $\overrightarrow{D}_0^{1,\ldots,H}(y_0^{h*}, b_0^{h*}; g_0^y, g_0^b)$. In particular, the decomposition in (3) is well defined because the three directional distance functions utilize the same directional vector (g_0^y, g_0^b).

We illustrate the above ideas and notions through a simple numerical example with only one strength and one concern. Assume that we have observed two industries. In particular, we have observed a sample of four units (A, B, C, and D) of the first industry and a sample of three units (E, F and G) of the second one. The circles in Figs. 1 and 2 represent the DMUs of the first industry, whereas the triangles represent the DMUs of the second industry. On one hand, in Fig. 1, we show the estimation of the two industry-specific frontiers (the solid line for the first industry and the dash line for the second industry).
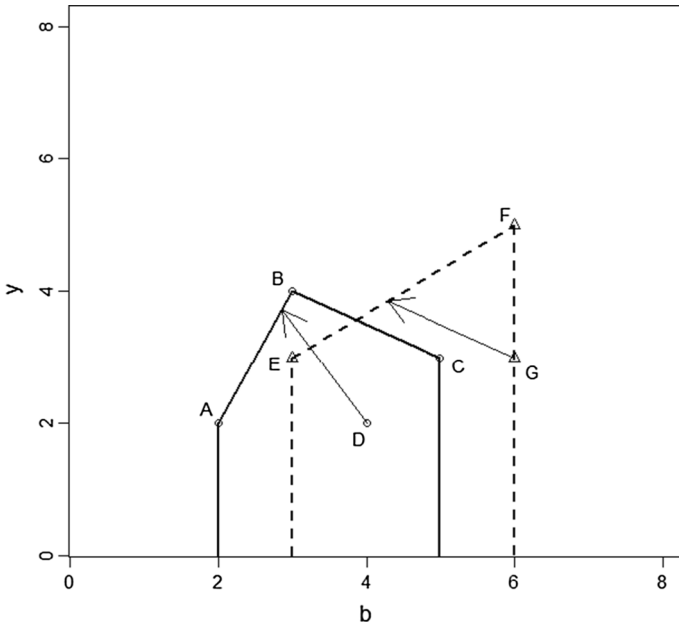
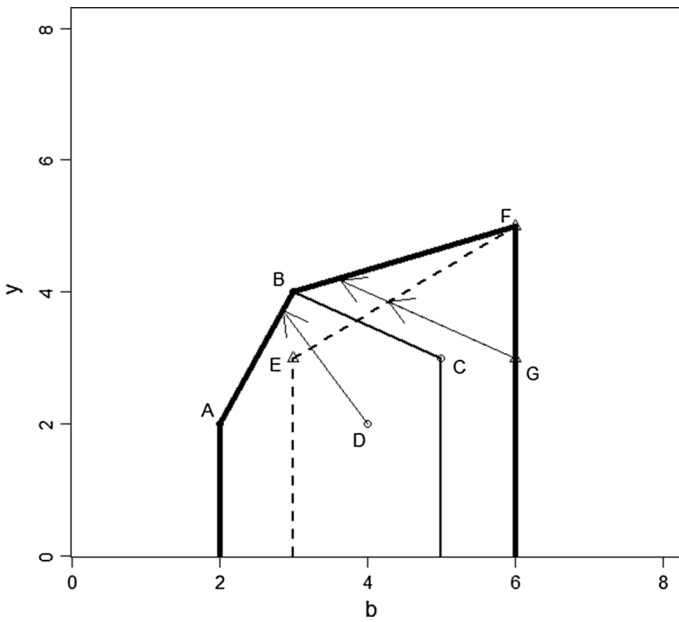**Fig. 1** Illustration of two industry-specific frontiers

Fig. 2 Illustration of a metafrontier

We also illustrate graphically how the directional distance function defined by model (1) works for two units. In the case of unit D, which belongs to the first industry, we show the directional vector that projects this DMU onto its industry-specific frontier. In the case of unit G, which belongs to the second industry, we also show its projection onto its specific frontier. On the other hand, in Fig. 2, we show the estimation of the corresponding meta-frontier, which envelops the two industry-specific frontiers. The directional distance function defined by model (2) coincides with that defined by model (1) in the case of evaluating DMU D, while yielding a greater value in the case of unit G. Thus, the industry gap equals zero in the case of unit D, but this is not true for G. In this last case, we illustrate in Fig. 2 the projection of unit G onto its industry-specific frontier and also the distance between this projection and the metafrontier, as a way of graphically illustrating the value of the industry gap measured at point G.

3 Dataset and Variables

In this research we use the KLD dataset as a source of information on CSP. KLD covers CSP activities of a large subset of publicly traded firms in the USA and has been used extensively in the scholarly literature (for example, in Flammer 2015; Servaes and Tamayo 2013; Chen and Delmas 2011; Lev et al. 2010; McWilliams and Siegel 2000). KLD assesses social performance based on multiple data sources, included surveys, firms' financial statements, annual and quarterly reports, articles in the popular press, and government reports. KLD has compiled information on CSP since 1991, but we have focused our analysis on the period 2004–2015, as this widens the coverage of firms in the database.

KLD rates firms along seven dimensions, including community, diversity, employee relations, human rights, product, environment, and corporate governance.[3] Following prior research (such as Lys et al. 2015) we further classify these dimensions into three major CSP categories: social (including community, diversity, employee relations, human rights, and product dimensions), environmental, and governance. Such aggregation also allows for a more viable analysis with DEA since it considerably reduces the problem of dimensionality, which is related to having many variables together with a limited number of observations.[4] For each dimension, KLD measures a number of positive aspects (strengths) and negative aspects (concerns), assigning a value of 1 when a firm presents a certain strength or concern and a value of 0 otherwise. The examples of community strengths include charitable giving and community engagement, while concerns include community impact controversies. The diversity dimension includes strengths such as board diversity and women/minority contracting, while workforce diversity controversies is an example of diversity concern. Employee relations dimension include indicators such as union relations strengths and concerns. Strengths connected with human rights are represented by indigenous

---

[3] KLD also contains the categories regarding firms' involvement in controversial activities, including alcohol, gambling, military contracting, nuclear power, and tobacco. The literature usually excludes these dimensions from the computation of CSP scores since they are inherently different to other areas covered by KLD (El Ghoul et al. 2011). Hence, for the same reason we do not consider these dimensions in our study.

[4] In general, the greater the number of variables, the less discerning the DEA analysis is (Jenkins and Anderson 2003). The literature indicates some empirical rules regarding the number of DMUs versus the number of inputs and outputs. See, for example, Cooper et al. (2007), Dyson et al. (2001) and Homburg (2001).

peoples relations and concerns relate to firms' operations in Burma and Sudan. Positive aspects of product dimension include quality, while negative aspects include product quality and safety controversies. Environmental strengths concern pollution prevention, while substantial emissions are categorized as concerns. Finally, corporate governance strength relates to reporting quality, while business ethics controversies is an example of governance concern. Because the number of categories in KLD has evolved over the years, it is not possible to directly compare strengths and concerns within each dimension across years. To account for this limitation, following prior research (for example, Servaes and Tamayo 2013; Deng et al. 2013), we have created adjusted measures of strengths and concerns by scaling the strength and concern scores for each firm year within each CSP dimension by the maximum number of items of the strength and concern scores of that dimension in each year.[5] We then took a sum of adjusted strengths and concerns for five dimensions—community, diversity, employee relations, human rights, and product—to obtain the measures of social strengths and social concerns, respectively (environmental and governance categories of CSP do not require any summation as they are not composed of any further dimensions).[6] Furthermore, to distinguish the industry to which each firm belongs, we match the data in KLD with information on industry (SIC codes) from Thomson Reuters Datastream. The final sample represents an unbalanced panel of 37,765 observations for 4824 firms.

Table 1 presents the sample distribution by the industry SIC code and year. The most heavily represented industries are finance (SIC code 37) and manufacturing (SIC code 38). Throughout the period 2004–2012, the sample size is relatively stable, but it decreased from 2013 due to the slight change in the coverage of US firms that occurred in the KLD dataset.

Table 2 reports descriptive statistics of adjusted CSP strengths and concerns for each of CSP category: social, environmental, and governance, and for each dimension of social CSP (community, diversity, employee relations, human rights and product). The mean values of CSP strengths for social and governance categories are smaller than CSP concerns for these categories; therefore, firms in the sample are, on average, irresponsible with regard to these aspects of CSP. Larger concerns than strengths for social category are due to the dimensions of diversity, employee relations and product. On the contrary, the mean values of CSP strengths with regard to environmental aspects of CSP are larger than their concerns. Nevertheless, there is also a considerable variation in the sample regarding these variables.

---

[5] Alternatively, one might consider scaling by the mean values of the strengths and concerns of each dimension in each year. However, such an approach would not be viable in the context of our data since for some categories and some years these mean values are equal to zero (that is, all firms present zero values for these specific scores). Nevertheless, even if the dataset is able to accommodate mean values, the usage of mean values would not change the inefficiency results because the DEA model we apply is characterized by the property of units invariance; that is, it is independent of the units in which the input and output variables are measured (Lovell and Pastor 1995).

[6] For example, suppose that in 2006 a company X has strengths in zero areas of community, one of diversity, one of employee relations, two of human rights, one of product, zero of environment, and one for governance, and the maximum number of strength indicators across these seven dimensions are 7, 8, 6, 3, 4, 6, and 5, respectively. The adjusted strength score for the company for community dimension is then 0/7, 1/8 for diversity, 1/6 for employee relations, 2/3 for human rights, 1/4 for product, 0/6 for environment, and 1/5 for governance. Thus, the adjusted strength score for the company for social category is equal to $0/7 + 1/8 + 1/6 + 2/3 + 1/4 = 1.21$, the score for environmental category is equal to $0/6 = 0$, and the score for the governance category is equal to $1/5 = 0.2$.

Table 1 Sample description—Industry representation by year

Year	CON	FIN	MAN	MIN	RTRADE	SERV	TRANS	WTRADE	All sectors
2004	27	1199	1047	90	178	427	234	62	3264
2005	27	1253	1003	106	178	420	240	66	3293
2006	30	1179	1000	117	173	398	242	69	2209
2007	30	1170	994	121	165	407	257	68	2159
2008	30	1172	1008	132	165	412	253	65	3237
2009	28	1117	1045	129	172	416	245	65	3217
2010	35	1202	1065	134	176	421	243	70	3346
2011	33	1164	1027	124	166	396	239	71	3220
2012	35	1202	1007	133	169	407	233	72	3258
2013	37	998	871	112	161	369	203	64	2815
2014	39	1017	885	115	154	374	209	65	2858
2015	36	1021	876	94	160	377	208	65	2837
TOTAL	387	13,694	11,828	1407	2017	4824	2806	802	37,765

CON = construction (SIC codes = 15–17); FIN = finance (SIC codes = 60–67); MAN = manufacturing (SIC codes = 20–39); MIN = mining (SIC codes = 10–14); RTRADE = retail trade (SIC codes = 52–59); SERV = services (SIC codes = 70–89); TRANS = transportation (SIC codes = 40–49); WTRADE = wholesale trade (SIC codes = 50–51)

Table 2 Descriptive statistics of CSP strengths and concerns, 2004–2015

Variable	Mean	SD	Coefficient of variation
CSP strengths-social	0.166	0.323	1.946
Community strengths	0.026	0.105	3.974
Diversity strengths	0.071	0.161	2.262
Employee relations strengths	0.042	0.100	2.352
Human rights strengths	0.008	0.072	9.347
Product strengths	0.018	0.069	3.792
CSP concerns-social	0.316	0.328	1.038
Community concerns	0.022	0.111	5.027
Diversity concerns	0.187	0.229	1.229
Employee relations concerns	0.058	0.117	2.026
Human rights concerns	0.007	0.045	6.286
Product concerns	0.043	0.123	2.888
CSP strengths-environmental	0.028	0.090	3.214
CSP concerns-environmental	0.018	0.069	3.833
CSP strengths-governance	0.042	0.115	2.738
CSP concerns-governance	0.068	0.118	1.735

4 Results

Our empirical analyses are divided into two parts. First, we look at the evolution over time (from 2004 till 2015) of each of CSP inefficiency indicators (pooled, within-industry, and industry gap) to assess how the inefficiency changes around the period related with recent

Table 3 Pooled CSP inefficiency, mean values reported

Year	CON	FIN	MAN	MIN	RTRADE	SERV	TRANS	WTRADE	All sectors
2004	0.373	0.302	0.316	0.317	0.342	0.313	0.307	0.328	0.312
2005	0.443	0.415	0.422	0.463	0.448	0.426	0.433	0.415	0.423
2006	0.572	0.527	0.511	0.540	0.550	0.536	0.516	0.538	0.525
2007	0.514	0.463	0.461	0.486	0.476	0.473	0.461	0.474	0.466
2008	0.495	0.437	0.429	0.442	0.448	0.443	0.424	0.437	0.436
2009	0.493	0.439	0.429	0.437	0.446	0.438	0.417	0.439	0.435
2010	0.623	0.578	0.556	0.579	0.537	0.578	0.541	0.570	0.566
2011	0.565	0.530	0.519	0.539	0.490	0.529	0.519	0.540	0.525
2012	0.570	0.564	0.557	0.518	0.582	0.584	0.564	0.582	0.564
2013	0.593	0.560	0.560	0.545	0.578	0.584	0.556	0.599	0.565
2014	0.515	0.494	0.507	0.432	0.515	0.521	0.501	0.534	0.502
2015	0.568	0.542	0.535	0.409	0.560	0.549	0.530	0.564	0.537

CON = construction; FIN = finance; MAN = manufacturing; MIN = mining; RTRADE = retail trade; SERV = services; TRANS = transportation; WTRADE = wholesale trade

Table 4 Within-industry CSP inefficiency, mean values reported

Year	CON	FIN	MAN	MIN	RTRADE	SERV	TRANS	WTRADE	All sectors
2004	0	0.062	0.287	0.048	0.173	0.090	0.163	0.058	0.150
2005	0.035	0.021	0.419	0.067	0.185	0.100	0.155	0.044	0.173
2006	0.034	0.070	0.501	0.062	0.153	0.145	0.123	0.039	0.221
2007	0.029	0.059	0.444	0.047	0.117	0.243	0.105	0.024	0.207
2008	0.019	0.102	0.415	0.065	0.148	0.201	0.113	0.026	0.211
2009	0.021	0.093	0.415	0.060	0.149	0.186	0.112	0.025	0.210
2010	0.005	0.269	0.543	0.083	0.244	0.318	0.190	0.027	0.340
2011	0.011	0.341	0.495	0.065	0.288	0.443	0.267	0.119	0.376
2012	0.067	0.404	0.485	0.202	0	0.328	0.247	0.086	0.369
2013	0.037	0.287	0.497	0.253	0	0.425	0.188	0	0.335
2014	0.033	0.199	0.490	0.190	0	0.331	0.201	0	0.289
2015	0.056	0.323	0.474	0.183	0.183	0.364	0.247	0.124	0.339

CON = construction; FIN = finance; MAN = manufacturing; MIN = mining; RTRADE = retail trade; SERV = services; TRANS = transportation; WTRADE = wholesale trade

financial crisis that, in the USA, began in 2007 and ended in 2009, as reported by the US National Bureau of Economic Research (2012). Then, in the second step, we analyze the mean values of each indicator for the whole period 2004–2015 to assess the differences between indicators as well as between industries.

We begin by looking at the results of evolution of CSP inefficiency over time for each sector. Tables 3, 4, and 5 summarize these results for CSP inefficiency estimated with regard to pooled frontier composed of firms in all industries, within-industry CSP inefficiency estimated with regard to industry-specific frontier, and the residual term of industry gap, respectively. The findings in Table 3 show that, for all industries, the CSP pooled

Table 5 CSP industry gap, mean values reported

Year	CON	FIN	MAN	MIN	RTRADE	SERV	TRANS	WTRADE	All sectors
2004	0.373	0.240	0.028	0.269	0.170	0.223	0.145	0.270	0.162
2005	0.407	0.394	0.003	0.396	0.264	0.326	0.278	0.372	0.250
2006	0.538	0.457	0.010	0.477	0.397	0.392	0.393	0.499	0.304
2007	0.486	0.404	0.017	0.439	0.359	0.230	0.356	0.450	0.259
2008	0.476	0.335	0.015	0.377	0.300	0.242	0.311	0.411	0.224
2009	0.472	0.346	0.014	0.377	0.299	0.252	0.305	0.414	0.224
2010	0.619	0.308	0.013	0.496	0.293	0.260	0.351	0.543	0.226
2011	0.554	0.189	0.024	0.473	0.202	0.086	0.251	0.421	0.149
2012	0.503	0.160	0.072	0.316	0.582	0.256	0.317	0.496	0.196
2013	0.556	0.273	0.063	0.291	0.578	0.160	0.368	0.599	0.229
2014	0.482	0.295	0.017	0.242	0.515	0.189	0.300	0.534	0.213
2015	0.512	0.219	0.061	0.227	0.560	0.185	0.282	0.441	0.198

CON = construction; FIN = finance; MAN = manufacturing; MIN = mining; RTRADE = retail trade; SERV = services; TRANS = transportation; WTRADE = wholesale trade

inefficiency (assessed with regard to year-specific frontier) increased steadily until 2006, but dropped consistently in the period related to recession; that is, 2007–2009. On average, across all industries CSP pooled inefficiency increased from 0.312 in 2004 to 0.525 in 2006, and then dropped to 0.435 in 2009. Such a result might indicate that the crisis actually improved CSP efficiency by forcing firms to increase CSP strengths and decrease CSP concerns. This result might be interpreted as follows: it is possible that firms in the sample used CSP activities and increased their efficiency during the crisis to protect themselves against the decreasing number of customers. Such an interpretation might further strengthen the benefits of CSP as a marketing tool (García de los Salmones et al. 2005) and positive consumer responses to CSP (Sen and Bhattacharya 2004). In addition, in the light of our findings, it is possible that in the period of recession firms in the sample tended to eliminate some inefficient CSP activities. Further, the findings in Table 3 indicate that pooled inefficiency increased considerably throughout all industries in the first year following from the crisis (2010) and then fluctuated in the years thereafter. On average, across all industries, CSP pooled inefficiency increased from 0.435 in 2009 to 0.566 in 2010. Hence, such a result might indicate that when economic conditions improved, firms in the sample worsened their CSP efficiency and possibly became less concerned about the efficiency of their CSP activities.

The sources of pooled CSP inefficiency can be explored further by looking at the results of within-industry CSP inefficiency (Table 4) or industry gaps in CSP inefficiency (Table 5). In particular, similar trends of CSP inefficiency evolution over time are found for the within-industry indicator as for the pooled indicator; that is, CSP within-industry inefficiency (assessed with regard to year-specific frontier) decreased in the crisis throughout all industries except for services (SERV), as shown in Table 4. Hence, this result might indicate that firms in the sample generally improved their efficient CSP performance in recession when assessed in comparison to firms belonging to the same industry. However, this finding was less pronounced than for pooled inefficiency since it was observed mainly in the first year of the crisis; that is, 2007. In average terms across all industries, within-industry CSP inefficiency increased from 0.150 in 2004 to 0.221 in 2006, and then dropped

to 0.207 in 2007. Also, as for pooled CSP inefficiency, within-industry CSP inefficiency generally increased in the first year after the crisis (with the exception of the construction industry, CON) and fluctuated in the years thereafter. In addition, for some industries and some years (for example, the construction sector in 2004) the results in Table 4 show the average values of within-industry CSP inefficiency being equal to 0, which requires further explanation. Such a finding occurs in situations when all firms in the sample have at least one concern (that is, at least one bad output) equal to 0. This stems from the optimization programs used to compute CSP inefficiency, which consider improvements in all CSP outputs simultaneously; this means that it is not possible to reduce the bad output with a zero, so generating a directional distance function value equals zero.[7] Finally, the results on the gaps in efficiency between the best practices of each industry and the pooled best practices (CSP industry gap) contained in Table 5 show that, similarly to the previous inefficiency indicators, the CSP industry gap decreased in the periods related to crisis (on average, across all industries, from 0.304 in 2006 till 0.224 in 2009), with the exception of the manufacturing industry (MAN), for which a slight increase was observed in these periods.[8]

We now turn to analyze the average CSP inefficiency indicators (pooled, within-industry, and industry gap) for the whole period 2004–2015, for each industry. These results are contained in Table 6, which also reports the findings of Simar and Zelenyuk's (2006) test (S-Z test) that allows statistical assessment of the differences between indicators for each industry as well as between industries for each indicator.

First of all, these results allow for the assessment of the differences in average CSP inefficiency between firms representing different industries (pooled CSP inefficiency). In particular, they show that there are small differences in the average CSP inefficiency between industries, as revealed by the similar average values of pooled inefficiency. However, in most cases, these small differences are statistically significant. The average pooled inefficiency ranges from 0.481 (mining) to 0.533 (construction), which indicates the considerable potential for improvement; that is, that firms can achieve more in their social, environmental, and governance CSP practices by increasing CSP strengths levels and decreasing the level of CSP concerns. The most CSP-efficient firms appear to belong to the mining and manufacturing sectors, as they produce more CSP strengths and fewer CSP concerns in comparison to other sectors. This finding could indicate that the sectors that tend to contribute most to environmental pollution and other harmful activities, such as mining and manufacturing, are more heavily scrutinized by public opinion and face more pressure, which results in them making better efforts regarding socially responsible practices. It is also possible that these sectors need to face stronger regulations towards CSP activities

---

[7] If a firm presents a value of zero for some concern (bad output)—that is, it presents the best possible performance in that dimension—then no other firm exists in the sample that strictly dominates its performance. In other words, one cannot find another DMU that dominates this firm in the sense of Pareto. Consequently, the "directional distance function" inefficiency is directly zero; that is, the evaluated firm is not inefficient. This is what happens in our approach. However, it is true that other types of technical inefficiency (slacks) can be present and the directional distance function approach is not able to consider them. This is a limitation of the directional distance function in general, which our approach naturally inherits.

[8] Nevertheless, the results on the evolution of inefficiency over time need to be interpreted with caution. This is because inefficiency is a relative measure estimated with regard to frontier of best practice firms, which makes it difficult to compare inefficiency over time estimated with regard to frontiers of different years. In order to provide more conclusive interpretations, one would need to (1) assume that technology does not change over time and estimate inefficiency for all years together or (2) assume that technology changes over time, estimate inefficiency for each year separately, but try to distinguish the absolute inefficiency change from the frontier shift. We thank an anonymous reviewer for these insights.

Table 6 Indicators of CSP inefficiency, mean values through the 2004–2015 period reported

Sector	Pooled CSP inefficiency	Within-industry CSP inefficiency	CSP industry gap	Differences between indicators at the critical 1% level (S–Z test)
CON	0.533	0.030	0.503	Yes
FIN	0.486	0.183	0.303	Yes
MAN	0.482	0.454	0.027	Yes
MIN	0.481	0.110	0.371	Yes
RTRADE	0.496	0.124	0.372	Yes
SERV	0.496	0.261	0.235	Yes
TRANS	0.479	0.174	0.305	Yes
WTRADE	0.504	0.048	0.455	Yes
All sectors	0.486	0.267	0.220	Yes
Differences between industries at the critical 1% level	Yes, except of between RTRADE and WTRADE, and SERV and WTRADE	Yes	Yes	

CON = construction; FIN = finance; MAN = manufacturing; MIN = mining; RTRADE = retail trade; SERV = services; TRANS = transportation; WTRADE = wholesale trade

than other sectors. On the other hand, construction firms appear to have the poorest results in CSP activities, as revealed by the largest values of CSP inefficiency. The results of previous research indicate that very few construction companies have positively embraced the ideas of sustainability and CSP, and have made slower progress towards CSP in comparison to other industries (Murray and Dainty 2008; Myers 2005).

The decomposition of pooled CSP inefficiency into inefficiency within each sector and the gap in inefficiency between the best practices of each industry and the overall best practices suggests that, on average across all industries, firms' management and skills regarding CSP is a slightly larger source of CSP pooled inefficiency than reasons embedded in industry (0.267 against 0.220 of inefficiency). Therefore, on average, shortcomings in managerial practices regarding CSP cause the most inefficiency. The smaller sources of this inefficiency are potential differences from industry characteristics as every industry has its specific combinations of CSP strengths and concerns arising from, for example, the industry's greater or lesser impact on the environment, the ease with which it conducts CSP activities, its greater or smaller scope for improving CSP, or its attitude towards marketing of CSP activities. Nevertheless, looking at specific sectors, we can see that the industries have different profiles in their CSP inefficiency. In particular, firms in the construction industry operate closest to their own industry profile compared to firms in other industries, as revealed by the lowest values of within-industry CSP inefficiency for these firms; that is, they obtain good results in management of their CSP activities. Hence, assessed in relation to own industry profile, construction companies produce CSP strengths and concerns under more stable conditions and are most homogenous in terms of their inefficiency in CSP. Taking high values of pooled inefficiency and low values of within-industry inefficiency for construction firms together results in the highest values of industry gap observed for this sector; that is, belonging to this specific industry creates large inefficiencies. On the contrary, the interesting case are firms in the manufacturing sector that have the highest values of within-industry CSP inefficiency compared to firms in other sectors, which means that they are more heterogeneous in terms of their inefficiency in CSP and are subject to highest fluctuations in CSP strengths and concerns. It also indicates shortcomings in managerial practices regarding CSP of these firms. However, the values of pooled inefficiency and within-industry inefficiency are comparable, which means that manufacturing firms have the lowest values of industry gap. Therefore, industry-related reasons consist of a relatively small portion of these firms CSP inefficiencies. Overall, we find that within-industry CSP inefficiency explains the largest part of pooled CSP inefficiency for firms in manufacturing and services sectors, while the CSP industry gap is much more relevant in explaining CSP pooled inefficiency for construction, finance, mining, retail trade, transportation, and wholesale trade sectors.[9]

---

[9] It would be interesting to compare the results obtained in the paper using the new approach to these applying the traditional methods. With this respect, the natural comparison would be with the standard 'radial' BoD model (Cherchye Lovell, et al. 2007a, b; Cherchye et al. 2004; Lovell and Pastor 1997, 1999; Lovell 1995). However, it is worth remembering that the traditional BoD model deals only with good outputs as well, as it cannot be computed for firms that present zero values for some indicators because it generally lacks the property of translation invariance (see Pastor and Aparicio 2015). Therefore, to be able to apply it with our dataset we needed to: (1) use only CSP strengths (positive CSP), hence reducing the number of outputs into three, and (2) remove all firms that present zero values for at least one of three outputs. After such modifications, the resulting dataset was only 1045 observations out of the total of 37,765 observations (in other words, approximately 97 percent of the sample size was removed). As a result, the majority of sectors were not sufficiently represented to be able to compute within-industry inefficiency and only pooled inefficiency was calculated with this reduced sample (these results on pooled inefficiency can be obtained from the authors upon request). Therefore, it is clear that the comparison efforts result in set of findings that are not comparable to these reported in the paper. These efforts also further emphasize the

5 Conclusions

This paper proposes a new approach to measure corporate social performance (CSP) through the definition of a composite indicator based on frontier techniques. In particular, a set of CSP metrics is aggregated into an overall measure by using data envelopment analysis. Our approach is based specifically on the benefit of the doubt model and was implemented through the directional distance function. Our approach is original in the sense that we introduce a model adapted for dealing with good outputs (CSP strengths), bad outputs (CSP concerns) and, at the same time, the presence of zeros in the CSP data. Another novel point of our contribution in the analysis of CSP is the determination of a metafrontier enveloping all the observations for all analyzed industries. This allows us to calculate several CSP measures: pooled CSP inefficiency, within-industry CSP inefficiency, and CSP industry gap. As a result, we are able to explore the sources of pooled inefficiency; that is, how much pooled inefficiency is due to management behavior with regard to CSP and how much is related to firms belonging to a particular industry.

The workability of the proposed approach is illustrated by means of empirical application that focused on CSP performance of US firms over the period 2004–2015 included in the Kinder, Lydenberg and Domini database. The results show that there is a considerable potential for improvement in the CSP performance of firms in the analyzed sample by increasing CSP strengths and reducing CSP concerns. The most CSP-efficient firms appear to belong to the mining and manufacturing sectors, while firms from the construction industry obtained the poorest results in terms of CSP efficiency. Moreover, we find that, on average across all industries, firms' managerial behavior towards CSP is a slightly larger source of CSP pooled inefficiency than differences related to industry. In addition, the findings of this study also indicate that sample firms actually improved their CSP efficiency in the period related to the recent financial crisis; that is 2007–2009. These results provide some managerial and policy implications. For example, the CSP efficiency results can help firm managers design operational strategies to improve CSP performance. From a policy perspective, the result of this study on the most inefficient CSP of construction firms might indicate that, for these firms to improve their social, environmental and governance performance, more investments in good CSP practices should be encouraged, for example through more favorable lending rates for these investments.

Future research could consider the extension of the model proposed in the paper towards the measurement of CSP inefficiency for each of the outputs separately. This will enable researchers to explore which of the dimensions of CSP (social, environmental, or governance) contributes most to the findings. As some of the CSP metrics are measured on the ordinal scale, another possible future research line could apply and extend DEA models designed for ordinal input and output variables, such as the one proposed by Cook and Zhu (2006). Finally, future research could undertake more in-depth analysis of the reasons for the differences in CSP inefficiency between industries.

Footnote 9 (continued)

advantageous characteristics of the developed approach in comparison to the traditional one and that using the traditional BoD model for CSP measurement can bias the findings when some important dimensions of CSP are not taken into account.

References

Ali, A. I., & Seiford, L. M. (1990). Translation invariance in data envelopment analysis. *Operations Research Letters, 9*(6), 403–405.

Aparicio, J., Pastor, J. T., & Vidal, F. (2016). The directional distance function and the translation invariance property. *Omega, 58,* 1–3.

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

Battese, G. E., Prasada Rao, D. S., & O'Donnell, Ch J. (2004). A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. *Journal of Productivity Analysis, 21,* 91–103.

Belu, C. (2009). Ranking corporations based on sustainable and socially responsible practices. A data envelopment analysis (DEA) approach. *Sustainable Development, 17,* 257–268.

Belu, C., & Manescu, C. (2013). Strategic corporate social responsibility and economic performance. *Applied Economics, 45,* 2751–2764.

Bendheim, C. L., Waddock, S. A., & Graves, S. B. (1998). Determining best practice in corporate-stakeholder relations using data envelopment analysis: An industry-level study. *Business and Society*, *37*(3), 306–338.

Carroll, A. B. (1979). A three-dimensional conceptual model of corporate social performance. *Academy of Management Review, 4*(1), 497–505.

Chambers, R. G., Chung, Y., & Färe, R. (1996). Benefit and distance functions. *Journal of Economic Theory, 70*(2), 407–419.

Chambers, R. G., Chung, Y., & Färe, R. (1998). Profit, directional distance functions, and Nerlovian efficiency. *Journal of Optimization Theory and Applications*, *98*, 351–364.

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

Charnes, A., Cooper, W. W., & Rhodes, E. (1981). Evaluating program and managerial efficiency: An application of data envelopment analysis to program follow through. *Management Science, 27*(6), 668–697.

Chen, Ch-M, & Delmas, M. (2011). Measuring corporate social performance: An efficiency perspective. *Production and Operations Management, 20*(6), 789–804.

Cherchye, L., Lovell, C. K., Moesen, W., & Van Puyenbroeck, T. (2007a). One market, one number? A composite indicator assessment of EU internal market dynamics. *European Economic Review, 51*(3), 749–779.

Cherchye, L., Moesen, W., Rogge, N., & Van Puyenbroeck, T. (2007b). An introduction to 'benefit of the doubt' composite indicators. *Social Indicators Research, 82,* 111–145.

Cherchye, L., Moesen, W., & Van Puyenbroeck, T. (2004). Legitimately diverse, yet comparable: on synthesizing social inclusion performance in the EU. *Journal of Common Market Studies, 42,* 919–955.

Chung, Y., Färe, R., & Grosskopf, S. (1997). Productivity and undesirable outputs: A directional distance function approach. *Journal of Environmental Management, 51,* 229–240.

Cook, W., & Zhu, J. (2006). Rank order data in DEA: A general framework. *European Journal of Operational Research, 174*(2), 1021–1038.

Cooper, W. W., Park, K. S., & Pastor, J. T. (1999). RAM: A range adjusted measure of inefficiency for use with additive models, and relations to other models and measures in DEA. *Journal of Productivity Analysis, 11*(1), 5–42.

Cooper, W. W., Seiford, L. M., & Tone, K. (2007). *Data envelopment analysis: A comprehensive text with models, applications, references and DEA-solver software* (2nd ed.). New York, NY: Springer.

Dakpo, K. H., Jeanneaux, P., & Latruffe, L. (2016). Modelling pollution-generating technologies in performance benchmarking: Recent developments, limits and future prospects in the nonparametric framework. *European Journal of Operational Research, 250,* 347–359.

Deng, X., Kang, J.-K., & Low, B. S. (2013). Corporate social responsibility and stakeholder value maximization: Evidence from mergers. *Journal of Financial Economics, 110,* 87–109.

Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., & Shale, E. A. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research, 132*(2), 245–259.

El Ghoul, S., Guedhami, O., Kwok, C., & Mishra, D. (2011). Does corporate social responsibility affect the cost of capital? *Journal of Banking & Finance, 35*(9), 2388–2406.

Flammer, C. (2015). Does corporate social responsibility lead to superior financial performance? A regression discontinuity approach. *Management Science, 61*(11), 2549–2568.

Førsund, F. R. (2009). Good modelling of bad outputs: Pollution and multiple-output production. *International Review of Environmental and Resource Economics, 3,* 1–38.

Fusco, E. (2015). Enhancing non-compensatory composite indicators: A directional proposal. *European Journal of Operational Research, 242,* 620–630.

García de los Salmones, M. M., Herrero Crespo, A., & Rodrígues del Bosque, I. (2005). Influence of corporate social responsibility in loyalty and valuation of services. *Journal of Business Ethics, 61,* 369–385.

Hampf, B., & Rødseth, K. L. (2015). Carbon dioxide emission standards for US power plants: An efficiency analysis perspective. *Energy Economics, 50,* 140–153.

Homburg, C. (2001). Using data envelopment analysis to benchmark activities. *International Journal of Production Economics, 73*(1), 51–58.

Jenkins, L., & Anderson, M. (2003). A multivariate statistical approach to reducing the number of variables in data envelopment analysis. *European Journal of Operational Research, 147*(1), 51–61.

Lee, K.-H., & Saen, R. F. (2012). Measuring corporate sustainability management: A data envelopment analysis approach. *International Journal of Production Economics, 140,* 219–226.

Lev, B., Petrovits, C., & Radhakrishnan, S. (2010). Is doing good good for you? How corporate charitable contributions enhance revenue growth. *Strategic Management Journal, 31,* 182–200.

Lovell, C. A. K. (1995). Measuring the macroeconomic performance of the Taiwanese economy. *International Journal of Production Economics, 39,* 165–178.

Lovell, C. A. K., & Pastor, J. T. (1995). Units invariant and translation invariant DEA models. *Operations Research Letters, 18,* 147–151.

Lovell, C. A. K., & Pastor, J. T. (1997). Target setting: An application to a bank branch network. *European Journal of Operational Research, 98,* 290–299.

Lovell, C. A. K., & Pastor, J. T. (1999). Radial DEA models without inputs or without outputs. *European Journal of Operational Research, 118,* 46–51.

Lys, T., Naughton, J. P., & Wang, C. (2015). Signaling through corporate accountability reporting. *Journal of Accounting and Economics, 60,* 56–72.

McWilliams, A., & Siegel, D. (2000). Corporate social responsibility and financial performance: Correlation or misspecification? *Strategic Management Journal, 21*(5), 603–609.

McWilliams, A., & Siegel, D. (2001). Corporate social responsibility: A theory of the firm perspective. *Academy of Management Review, 26*(1), 117–127.

Murray, M., & Dainty, A. (2008). *Corporate social responsibility in the construction industry*. London: Routledge.

Murty, S., Russell, R. R., & Levkoff, S. B. (2012). On modeling pollution-generating technologies. *Journal of Environmental Economics and Management, 64,* 117–135.

Myers, D. (2005). A review of construction companies' attitudes to sustainability. *Construction Management and Economics, 23,* 781–785.

National Bureau of Economic Research. (2012). US business cycle expansions and contractions. Available from: http://www.nber.org/cycles.html. Accessed 15 Feb 2018.

O'Donnell, Ch J, Prasada Rao, D. S., & Battese, G. E. (2008). Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. *Empirical Economics, 34,* 231–255.

Pastor, J. T., & Aparicio, J. (2015). Translation invariance in data envelopment analysis. In J. Zhu (Ed.), *Data envelopment analysis. A handbook of models and methods* (pp. 245–268). Boston, MA: Springer.

Pastor, J. T., Ruiz, J. L., & Sirvent, I. (1999). An enhanced DEA Russell graph efficiency measure. *European Journal of Operational Research, 115*(3), 596–607.

Portela, M. C. A., Thanassoulis, E., & Simpson, G. (2004). Negative data in DEA: A directional distance function approach applied to bank branches. *Journal of the Operational Research Society, 55,* 1111–1121.

Rogge, N. (2018). Composite indicators as generalized benefit-of-the-doubt weighted averages. *European Journal of Operational Research, 267,* 381–392.

Ruf, B., Muralidhar, K., & Paul, K. (1998). The development of a systematic, aggregate measure of corporate social performance. *Journal of Management, 24,* 119–133.

Sahoo, B. K., Singh, R., Mishra, B., & Sankaran, K. (2017). Research productivity in management schools of India during 1968–2015: A directional benefit-of-doubt model analysis. *Omega, 66,* 118–139.

Sen, S., & Bhattacharya, C. B. (2004). Doing better at doing good: When, why, and how consumers respond to corporate social initiatives. *California Management Review, 47*(1), 9–24.

Servaes, H., & Tamayo, A. (2013). The impact of corporate social responsibility on firm value: The role of customer awareness. *Management Science, 59*(5), 1045–1061.

Shen, Y., Hermans, E., Brijs, T., & Wets, G. (2013). Data envelopment analysis for composite indicators: A multiple layer model. *Social Indicators Research, 114,* 739–756.

Shephard, R. W. (1953). *Cost and production functions*. Princeton, NJ: Princeton University Press.

Siegel, D. S., & Vitaliano, D. F. (2007). An empirical analysis of the strategic use of corporate social responsibility. *Journal of Economics & Management Strategy, 16*(3), 773–792.

Simar, L., & Zelenyuk, V. (2006). On testing equality of distributions of technical efficiency scores. *Econometric Reviews, 25*(4), 497–522.

Surroca, J., Tribo, J. A., & Waddock, S. (2010). Corporate responsibility and financial performance: The role of intangible resources. *Strategic Management Journal, 31*(5), 463–490.

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

Van Puyenbroeck, T., & Rogge, N. (2017). Geometric mean quantity index numbers with Benefit-of-the-Doubt weights. *European Journal of Operational Research, 256,* 1004–1014.

Vidoli, F., Fusco, E., & Mazziotta, C. (2015). Non-compensability in composite indicators: A robust directional frontier method. *Social Indicators Research, 122,* 635–652.

Waddock, S. A., & Graves, S. B. (1997). The corporate social performance-financial performance link. *Strategic Management Journal, 18*(4), 303–319.