Measuring Corporate Social Performance:

An Efficiency Perspective

Chien-Ming Chen,
Post Doctoral Fellow
UCLA Institute of the Environment,
University of California, Los Angeles
La Kretz Hall, Suite 300 • Box 951496 • Los Angeles, CA 90095-1496
cmchen@ioe.ucla.edu

Magali Delmas
Professor of Management
UCLA Institute of the Environment, and Anderson School of Management
University of California, Los Angeles
La Kretz Hall, Suite 300 • Box 951496 • Los Angeles, CA 90095-1496
delmas@ioe.ucla.edu

June 2010

Forthcoming *Production and Operations Management*

Abstract

Aggregation of corporate social performance (CSP) metrics poses a major challenge to researchers and practitioners. This study provides a critical evaluation of current aggregation approaches and proposes a new methodology based on Data Envelopment Analysis (DEA) to compute a CSP index. DEA is independent of subjective weight specifications and provides an efficiency index to benchmark the CSP of firms. Using CSP data from 2,190 firms in three major industries from the Kinder, Lydenberg, and Domini Inc. database in 2007, our study presents the first application of the DEA model for CSP and ordinal data and opens up a new path for future empirical CSP research.

Keywords: Corporate Social Performance; KLD; Efficiency; Data Envelopment Analysis.

1 INTRODUCTION

Stakeholders are becoming more and more concerned about the corporate social performance (CSP) of firms' operations. CSP can be defined as "a construct that emphasizes a company's responsibilities to multiple stakeholders, such as employees and the community at large, in addition to its traditional responsibilities to economic shareholders" (Turban and Greening 1996, p.658). For example, investors are increasingly using socially responsible investing (SRI) screens to select or avoid investing in firms according to their environmental and social preferences (Chatterji et al. 2009). Similarly, a growing number of consumers purchase eco-labeled products that signal a lower environmental and social impact of corporate operations (Loureiro and Lotade 2005). Some corporations are also developing socially responsible purchasing practices to promote more sustainable supply chains (e.g., Drumwright 1994, Bowen et al. 2001, Srivastava 2007, Carter 2008, Seuring and Müller 2008). However, measuring CSP has proven to be a daunting task because it represents a broad range of economic, social, and environmental impacts caused by business operations and thus requires multiple metrics to fully cover its scope (Gond and Crane 2009, Rowley and Berman 2000).

As a result, researchers often need aggregate CSP measures to assess the overall corporate social performance of firms. Most empirical studies on CSP employ simple linear aggregations, weighted or non-weighted, to derive a composite CSP score from a selection of CSP metrics. These types of approaches would seem appropriate when the weights are exogenously given. For example, NGOs may have a specific weighting scheme based on the priorities of their members. However, for managers who face a variety of stakeholder pressures, the choice of weights is more ambiguous. Specifically, one primary stakeholder group (e.g., customers) may very well hold opinions that conflict with those of another primary or secondary group (e.g., employees) about the same corporate social policy of a firm (Clarkson 1995). In addition, because stakeholder characteristics and preferences can shift dramatically under different contexts and times (Griffin 2000), prioritizing CSP categories can turn into a formidable task.

Furthermore, CSP assessment contains both negative and positive metrics to represent strengths and concerns regarding CSP practices. For example, generously giving to charities in the community is often perceived as a positive practice, whereas investments that would lead to controversies might be considered detrimental to CSP. Similarly, the use of clean energy is often considered a positive practice, whereas making profits from fossil fuel products might be considered negative because of the impact on climate change. When stakeholders want to balance concerns over strengths, they also face the challenge of assessing the respective importance of different CSP categories.

Considering the multiple dimensions of the CSP construct, we argue the existing CSP aggregation methodologies fail to provide an effective measure of CSP. We show the scores resulting from these aggregation methodologies differ in terms of their median and variance and are sensitive to changes in aggregation weights. This sensitivity can be fairly problematic. Since expressing CSP through an aggregate measure is necessary for most analyses, we propose an alternative methodology to calculate a CSP index. Our methodology is based on Data Envelopment Analysis (DEA), a mathematical programming method for evaluating the relative efficiencies of firms (Charnes et al. 1978, Cook and Zhu 2006) that does not require a priori weights to aggregate different CSP dimensions.

DEA computes an efficient frontier that represents the best performers in a peer group. The DEA CSP score represents the distance of a firm to the efficient frontier and the extent to which a firm can reduce its current concerns, given its strengths relative to those of the best performers. We argue that DEA has several advantages in addressing the challenges of assessing CSP. First, DEA produces a ratio index that incorporates both good and bad CSP metrics. Second, DEA does not require an a priori weight specification for different CSP criteria. Third, the DEA score represents the distance to the efficient frontier and is easy to interpret. These features help compare firms' CSP both within and across industries.

To meet the ordinal nature of the CSP data, we use the DEA model for rank order data (Cook and Zhu 2006) and present the first large-scale empirical application of DEA to ordinal data. Our model is inspired by the study conducted by Benheim et al. (1998), who used DEA to assess best management practices regarding stakeholder relations. However, their study did not consider the trade-offs between strengths and concerns. Our model can be contrasted with previous eco-efficiency studies based on DEA (e.g., Dyckhoff and Allen 2001, Färe et al. 2006, Kuosmanen and Kortelainen 2007), which draw on concrete quantities of environmental data such as total CO_2 and SO_2 emissions.

In this paper, we focus on the Kinder, Lydenberg, and Domini Inc. (KLD) database, currently the most widely used and comprehensive information source for CSP research (Waddock 2003). KLD publishes the CSP ratings of major publicly traded companies in the United States, and the data cover areas of *environmental performance*, *social contribution*, *corporate governance*, and *controversial business involvement*. Our empirical analysis shows that DEA is more robust than the existing CSP aggregation methodologies, whose aggregation results are sensitive to weight changes. Our analysis also highlights the ease of interpretation of the DEA score for benchmarking purposes.

In the next section, we review the empirical CSP literature and outline the advantages of DEA and its formulations. In section 3, we compute the DEA efficiency score for CSP using KLD data and compare our results with those using prior existing aggregation methodologies. In section 4, we summarize our findings and suggest directions for future CSP research based on the DEA methodology.

2 LITERATURE REVIEW: MEASURING CSP

Because the full spectrum of CSP is broad, generating a proxy that can reflect its full scope is challenging. Although measures that represent a firm's financial performance are clearly defined and readily available (like Return on Assets, Return on Investment, etc.), the CSP counterparts are not. Because of the qualitative nature of CSP, the assessment of CSP relies mostly on "soft" measures related to management practices, rather than the "harder" measures (e.g., tons of CO_2 emission or of toxic releases). Common CSP measures include, for example, labor right protection and the transparency of social and environmental performance reporting. Several authors have described the challenges associated with measuring CSP (Carroll 1999, Graves and Waddock 1994, Wokutch and McKinney 1991).

The multi-dimensionality of the CSP construct is the primary difficulty in measuring CSP. As Hirsch and Levin (1999, p.200) note, CSP is "a broad concept or idea used loosely to encompass and account for a broad set of diverse phenomena." Rowley and Berman (2000) criticize for two main reasons studies that proxy CSP using a single-dimensional measure: the one-dimensional measure cannot represent the full breadth of CSP construct (i.e., the validity problem), and it makes comparing and unifying different studies extremely difficult.

Recent studies attempt to grapple with this issue by using simple linear aggregation of CSP data to create an aggregate CSP score for either a specific subset of CSP criteria or the entire CSP construct. In spite of their ease of implementation, these aggregation approaches have suffered from several major drawbacks. They often lack general applicability and are difficult to interpret in different contexts (Berman 1999). We next introduce these approaches and describe their limitations.

2.1 Linear aggregation methods

Academic researchers have measured corporate social performance using survey questionnaires, content analyses of annual reports, expert evaluations, and regulatory compliance data (Aupperle 1991, Bowman and Haire 1975, Wolfe 1991, Zahra et al. 1993). More recently, several for-profit organizations have taken up the task of measuring CSP. These include the SAM Group Inc. (SAM), the Riskmetrics Group, and KLD. SAM, for example, gathers CSP information such as board structure, ability to manage risk, and environmental reporting system (http://www.sam-group.com). The Riskmetrics Group evaluates corporate governance, employee and stakeholder management, and corporate environmental performance (http://www.riskmetrics.com). KLD ratings include the following categories: employee relations, diversity, community relationships, human rights, the environment, governance, and controversial issues (http://www.kld.com) for the period 1991 to 2007. Up until now, the KLD database has been the most commonly used database for assessing CSP (Graves and Waddock 1994, Turban and Greening 1996, Waddock 2003). A search for "Corporate Social Responsibility" and "KLD" in Google Scholar in May 2009 produced over 700 hits.

2.1.1 CSP studies using linear aggregation approaches

KLD has generated a flourishing literature on CSP in prominent academic management journals. In Table 1, we tally these papers by journals. The results show that the *Journal of Business Ethics*, *Business and Society*, and the *Academy of Management Journal* have published the largest number of KLD-CSP articles.

***

[Insert Table 1 here]

***

The literature uses two main types of aggregation methodologies. The first consists of assigning equal weight to all categories (community relationships, environmental performance, human rights, etc.). For example, Hillman and Keim (2001) use the equal weights aggregation method because the literature "has yet to identify a ranking of importance [of different CSP categories] for various stakeholder groups and issues" (Hillman and Keim 2001, p. 131). By assigning equal weights, however, the researcher assumes all criteria are of the same or at least similar importance. As Bird et al. (2007) argue, this assumption is invalid in most cases.

The second methodology is to gather information on stakeholder preferences in order to assign weights to specific CSP categories. Using this method, Ruf et al. (1998) generated weights for the different KLD dimensions through a survey of preference of 101 public officers, executives of non-profit organizations, and managerial accountants. The respondents considered product/liability issues to have the highest weight (23%), followed by employee relations (18%), women/minority (15%), environmental (14%), and community relations (12%). The three social dimensions considered least important were nuclear power (7%), military (5%), and South Africa (5%) (Ruf et al. 1998). Similarly, Waddock and Graves (1997b) developed a weighting scheme based on the opinion of three experts from the Social Issues in Management division of the Academy of Management who had been active in the social issues arena for more than 15 years. Employee relations were found to be the most important category (17%), followed by product/liability issues (15%) and community relations (15%), and then the environment (14%). Other social issues considered include the treatment of women and minorities (13.6%), nuclear power (8.9%), military contracts (8.6%), and South Africa (7.6%).

An analysis of the publication counts by aggregation types of the journals listed in Table 1 shows that, of the 43 publications, 26 used equal weights and 9 used unequal weights, whereas the other 8 studies did not use aggregation. We will argue below that these aggregation methods can lead to non-robust results.

2.1.2 Limitations of the aggregation approaches

The first question with these aggregation approaches is whether these weights are justifiable. The answer from the literature, however, tends to be unfavorable. Research on stakeholder management and social participation has pointed out that no universally agreed-upon weights or prioritization of social or environmental issues can exist for different stakeholder groups in different situations, since stakeholder attributes (e.g., stakeholder composition, perceptions, and preferences) are dynamic and could change over time (Mitchell et al. 1997, Hillman and Keim 2001, Bird et al. 2007). Even for a specific stakeholder group, the current weight elicitation encounters great difficulties when evaluating less tangible goods such as clean air and noise (Freeman 2003, Kuosmanen and Kortelainen 2007). Chatterji and Levine (2006) note that even major social investment indexes (SRI) weight the non-financial CSP indicators of listed companies quite differently, which makes comparing the CSP of firms difficult. Delquié (1997) further shows that biases can arise in the elicitation of weights.

Rowley and Berman (2000) further highlight several concerns for the simple weight-aggregation approach the CSP-Corporate Financial Performance literature uses: the aggregate score lacks a simple interpretation; the weights are not representative of the trade-off between CSP criteria; and when a different data source is used (e.g., a new database with or without the addition or removal of the original CSP criteria), the weights and aggregate scores could lose their applicability and comparability.

2.2 Strengths versus concerns

As noted earlier, the CSP construct consists of both positive and negative firm behavior. The KLD database, which contains both "strength" and "concern" measures for each CSP issue, reflects this trait. Many empirical researches conduct simple aggregation of the measures (e.g., strength scores minus concern scores) to create a CSP-item score. Yet Mattingly and Berman (2006) have found through a factor analysis that the "strength" and "concern" measures in KLD data represent four distinct constructs, and thus they should not be combined without "carefully clarify[ing] the social construct that we intend to measure" (Mattingly and Berman 2006, p.41).

Previous research has found that firms with high scores on their strengths also tend to have high scores on their concerns, as indicated by the positive correlation between the KLD strengths and concerns (Delmas and Doctori-Blass 2010, Mattingly and Berman 2006). Simple aggregation methods (subtraction of strengths from concerns), however, consider that firms with high scores on both strengths and concerns are similar to firms with low scores on both strengths and concerns.

In spite of the extensive discussion of the aggregation issues we just described, the CSP literature has yet to provide empirical researchers with a general methodology to tackle all of these criticisms. In this paper, we propose a weight-free evaluation approach, called Data Envelopment Analysis (DEA), to evaluate CSP from an efficiency perspective. In the following section, we introduce the DEA approach.

3 DATA ENVELOPMENT ANALYSIS

Our methodology is based on DEA, which is a mathematical programming method the operations research and management literature has used extensively to evaluate firms' efficiency (Charnes et al. 1978, Cooper et al. 2006). In the DEA methodology, efficient firms are those that use minimal inputs to produce maximum outputs. DEA evaluates a firm's multi-factor performance by a composite efficiency index with a value between zero and one, with "one" representing the efficient firms. It does so without the need for explicit weight specifications for inputs and outputs. These weights are generated automatically through an optimization procedure, such that the evaluated firm will be assigned a set of "optimal weights" that maximizes the firm's efficiency relative to the other firms in the sample. Each firm will therefore receive its most favorable weights, and the influence of subjective weightings can be eliminated. In this paper, we consider CSP concerns as inputs (i.e., factors to be minimized) and CSP strengths as outputs (i.e., factors to be maximized). Thus the DEA score can account for the trade-off between positive and negative CSP indicators.

Benheim et al. (1998) utilize the conventional DEA model to identify firms' best practices regarding their stakeholder relationship management. In their analysis, they select output variables as the aggregated scores of five CSP categories, whereas dummy variables represent input variables (i.e., all firms have the same input value). Although the authors use CSP categories as outputs, their study does not differentiate between strengths and concerns. In contrast, our approach uses CSP concerns as inputs and CSP strengths as outputs.

We use an input-oriented DEA model, where the objective is to minimize CSP concerns (the inputs) given current CSP strengths (the outputs). Figure 1 illustrates the fundamental mechanism of the DEA model. In the figure, we consider one CSP concern and one CSP strength. First, DEA constructs the efficient frontier as a piecewise linear function that envelops the observed sample. Subsequently, each firm is benchmarked against its unique target located on the frontier. The DEA score represents the distance between the firm and the efficiency target (e.g., the length from *O* to *i** divided by the length from *O* to *i*). Firms on the frontier are identified as efficient and hence their DEA values are equal to one, whereas firms with a score lower than one are considered inefficient (i.e., they should further reduce their concern levels).

***

[Insert Figure 1 here]

***

However, because we determine the efficiency frontier based on observed data, small or unrepresentative samples can often result in underestimation of the efficient frontier (i.e., closer to the observed firms), which may in turn reduce the variation of efficiency scores among firms; see Podinovski and Thanassoulis (2007) for further discussion and possible remedies for the problem.

3.1.1 DEA formulations

Consider *n* firms under evaluation. In evaluating a firm, we consider *s* desirable criteria and *m* undesirable criteria; accordingly, we denote the observed performance of firm *j* as y_{j1} to y_{js} and x_{j1} to x_{jm}, respectively. Note that we assume these variables have a ratio-scale measure. In general, we would consider a firm superior if it has higher desirable values than other firms, keeping the level of undesirable criteria constant; or vice versa. As such, we can construct the performance index as

$$I_j = \sum_{r=1}^{s} u_r y_{jr} \Big/ \sum_{i=1}^{m} v_i x_{ji} \text{, for } j = 1, \ldots, n, \tag{1}$$

where the u_r and v_i in the formula are the weights attached to the r*th* desirable and the i*th* undesirable criterion, respectively.

As in the classical productivity efficiency index, the composite CSP in this formulation is represented as a ratio between the aggregated good and bad. The weight parameters are assumed to be known and are supposed to reflect the relative importance among different criteria. A higher index score then indicates better CSP, and a lower score indicates worse CSP.

As noted earlier, however, meaningful weights or rankings for CSP criteria are difficult to assess, especially for different stakeholder groups. DEA can be helpful in addressing this problem. Instead of assigning fixed weights, DEA allows weights to be variable, and the following optimization problem determines the weights (for firm 1):

$$\text{Max} \ \sum_{r=1}^{s} u_r y_{1r} \Big/ \sum_{i=1}^{m} v_i x_{1i}$$

$$\text{subject to} \ \sum_{r=1}^{s} u_r y_{jr} \Big/ \sum_{i=1}^{m} v_i x_{ji} \le 1, \ \text{for} \ j = 1,\dots,n.,$$

$$u_r \ge 0 \ \text{for} \ r = 1,\dots,s; \ v_i \ge 0 \ \text{for} \ i = 1,\dots,m. \qquad (2)$$

Model (2) is commonly called *the DEA multiplier model* in the literature. The model will select weights that maximize the efficiency of the evaluated firm. The first set of constraints standardizes the evaluation results such that the efficiency scores of all firms should not exceed one. The second set of constraints guarantees the weights are non-negative. We then solve the problem for each evaluated firm by replacing the parameter values in the objective function.

For computational convenience, we can reformulate the problem as an equivalent linear programming problem by using the Charnes-Cooper transformation for fractional linear problem; namely, we replace the objective function with

$$\text{Max} \ \sum_{r=1}^{s} u_r y_{1r}, \qquad (3)$$

and add a linearizing constraint

$$\sum_{i=1}^{m} v_i x_{1i} = 1. \qquad (4)$$

The objective value (3) can be interpreted as the distance between the focal firm's CSP and the best CSP performer in the sample. In this case, we can define the *CSP efficiency* as the extent to which a firm can reduce its current concerns, given its strengths; for example, a score of 0.9 means the firm can decrease its overall CSP concerns by 10% relative to the best practice (i.e., CSP-efficient firms). Without the exogenous influence of weights, DEA scores can capture the true underlying difference in CSP (Charnes et al. 2006).

The dual linear programming problem of the multiplier model (2) is generally referred to as the *envelopment model*:

$$\text{Min } \theta_1$$

$$\text{subject to} \qquad \sum_{j=1}^{n} \lambda_j x_{ji} \le \theta_1 x_{1i}, \text{ for } i = 1, \ldots, m,$$

$$\sum_{j=1}^{n} \lambda_j y_{jr} \ge y_{1r}, \text{ for } r = 1, \ldots, s,$$

$$\lambda_j \ge 0, \text{ for } j = 1, \ldots, n. \qquad (5)$$

The envelopment model (5) minimizes the contraction ratio θ_1 such that the evaluated firm can become CSP efficient after contraction (i.e., finding its benchmark on the frontier). This interpretation corresponds to the graphical illustration in Figure 1.

Traditional DEA models, however, assume the input and output variables are in the ratio or interval scale. They are therefore not appropriate for the KLD data. Cook and Zhu (2006) developed the extension for the ordinal input and output variables, that is, variables measured in Likert scale. This feature is necessary for our DEA application to KLD data. The strategy literature that employs KLD data has ignored the issue of measurement scale. Specifically, we know a firm that scores three points in the environmental strength category is not "three times better" than a firm that scores one point in the same category—these scores can only be appropriately compared in the ordinal scale. The Cook and Zhu model is developed based on the extension of the classical DEA model, but it involves a higher level of mathematical detail; see Zhu (2003) and Cook and Zhu (2006).

3.1.2 Identifying benchmark targets using DEA

Scores from the DEA model can be interpreted as the reduction ratio of concern levels necessary for the firm to become CSP efficient, because our DEA model is input-oriented. We therefore seek to reduce concerns, given the firm's current strengths. We provide an example in Table 2. The KLD scores are presented in the first column. The benchmark KLD scores are presented in the second column, and we obtain them by multiplying the KLD score by the DEA score (here 0.9475). One problem is that the benchmark efficiency score could be fractional. For example, in Table 2, the target of the Community concern is 3.8. One way to interpret the score in terms of changes for CSP variables is to round the fractional number down to its nearest integer. Column 3 in Table 2 presents the rounded benchmarks for all concerns. If we wanted to find out how much a firm needs to increase its strengths to reach efficiency, we would need to compute an output-oriented DEA model.

***

[Insert Table 2 about here]

***

Several studies have used DEA or its variants to assess the environmental performance of firms (see Zhou et al. [2009] for a survey). What is similar in evaluating environmental performance and CSP is that we need to consider both desirable and undesirable performance, for example, electricity generated and the greenhouse gas emission from a utility generation plant. Compared with CSP, the environmental data these studies use are more directly observable, such as energy consumption, pollutants emitted, and some financial or economic measures. Our paper is thus distinct from previous DEA studies of environmental issues in that we need to particularly account for the qualitative and ordinal nature of CSP data, namely, that CSP variables are related to management practices rather than actual performance outputs. However, we should note that the KLD database also considers some quantitative data in its qualitative assessment of firm performance (e.g., information from the toxic release inventory by the Environmental Production Agency [EPA]).

4 EMPIRICAL ANALYSIS

In this section, we present the KLD data and the aggregation methodologies used in the literature. We then compare the composite CSP scores we obtained from the DEA model with those from the weight aggregation methods the literature reports.

4.1 CSP and KLD

Several studies in the literature have criticized the measures used in empirical studies (including the KLD database) as not fully grounded in the theoretical development of CSP, and using a fixed set of measures presupposes a "one-size-fits-all" property of CSP for different industries (see, e.g., Rowley and Berman 2000, Mattingly and Berman 2006, Gond and Crane 2009). Yet in view of the vague boundary and complexity of CSP, the KLD database has been deemed "the de facto research standard at this moment" and "the best currently available to scholars'' (Waddock 2003, pp.369 and 371). Its 2007 version includes the CSP assessment of the 3,000 largest U.S. publicly traded companies over 21 CSP issues, which can be classified into four major CSP dimensions: environmental, social, governance, and controversial business involvement ratings.

KLD evaluates each of the CSP issues by a number of *concern* and *strength* variables, and these variables are coded as binary variables. For example, in the climate change issue, "taking significant measure to reduce emissions by using clean energy" is considered a strength, whereas "deriving significant profits from the sale of fossil fuels and their derivatives" is regarded as a concern. A team of experts from KLD investigates, using a variety of data sources, how a firm can score on individual KLD strength and concern variables. Their sources include direct communication with the company managers, public documents, government and NGO information, and media reports (see http://www.kld.com for a complete description).

As do most studies in the literature, we exclude in our analysis the nine CSP issues under the "controversial business involvement" umbrella, as no theory or evidence yet supports their roles in the CSP research (Turban and Greening 1996, Berman et al. 1999). The nine CSP issues include Abortion, Adult Entertainment, Alcohol, Contraceptives, Firearms, Gambling, Military, Nuclear Power, and Tobacco. In what follows, we will describe the main variables included in the database and how the literature has aggregated the KLD data. Figure 2 illustrates the structure of the KLD database, which includes three main categories: environmental performance, social ratings, and governance ratings. Within each of these main categories, several issues are considered, such as climate change and operations and management within the environmental performance category. A number of concern and strength binary variables (i.e., the variable is equal to one if the firm meets the criteria of the concern or strength variable and equal to zero otherwise) subsequently represent each issue. For example, in determining the *Climate Change* concern item for the environmental performance category, the team of KLD experts will use various data sources to assess whether "*The company derives substantial revenues, directly or indirectly, from the sale of coal or oil and its derivative fuel products*." If the KLD team concludes the evaluated firm satisfies the above description, this firm's climate concern score is one; otherwise, the score will be zero. We provide a partial list of concern and strength items in Figure 2; see http://www.kld.com for the full list of these items and their definitions.

***

[Insert Figure 2 about here]

***

4.2 Data and methods

We utilize the 2007 KLD data, which contain the CSP ratings of around 3,000 of the largest publicly traded firms in the United States. Table 3 gives the descriptive statistics at the issue level (e.g., climate change, diversity, human rights, and so on), which we use as the basis for subsequent calculation.

***

[Insert Table 3 about here]

***

The empirical and conceptual CSP literatures have both reiterated the substantial influence of industrial effects on the analysis of CSP (Waddock and Graves 1997b, McWilliams and Siegel 2000, McWilliams and Siegel 2001, Griffin 2000). To take into account the industry effect, we classified the 2007 sample according to the first two digits of the SIC code (see Table 4). Our analysis focuses on the three largest industries in the 2007 sample: *Manufacturing*, *Finance*, and *Service* industries.

***

[Insert Table 4 about here]

***

In the subsequent analysis, we use the three most widely used CSP weighting schemes. These are shown in Table 5 and include (i) equal weights, (ii) weights derived from expert opinions

(Waddock and Graves 1997b), and (iii) weights derived from survey of public affairs officers, executives of non-profit organizations, and managerial accountants (Ruf et al. 1998).

However, as the KLD database has updated the evaluated CSP items and the number of strength and concern variables over the years, the weights Ruf et al. (1998) and Waddock and Graves (1997b) developed no longer match the 2007 version of the KLD database.[1] In order to compare our results from DEA with those of these previous methods, we choose to remove the *corporate ratings* category from the sample since neither Ruf et al. nor Waddock and Graves provided a weight for this CSP category. For the same reason, we also combine the four issues of the category *Environmental performance* (see Figure 2).

***

[Insert Table 5 about here]

***

With these weights, we calculate the KLD-CSP score according to the formula

$$\text{KLD\_CSP score}_j = \sum_{i=1}^{m} \rho_i (y_{ji} - x_{ji}) \,. \qquad (5)$$

y_{ji} and x_{ji} denote firm *j*'s number of strengths and concerns in CSP category *i*, respectively; ρ_i is the weight for category *i*. In the formula, we first calculate the category score by subtracting *concerns* from *strengths*. Then we can obtain the KLD-CSP score simply as the weighted sum of the category scores. Note that the fixed-weight approach only applies

---

[1] In 1996, KLD removed the Property, Plant, Equipment item from the environmental performance category, and in 1999, it added a Climate Change item to the environmental performance category. In 2005, KLD added the governance rating category. In 2006, it added a Management Systems Strength item to the environmental performance category.

weighting to CSP categories; the concern and strength variables are merged by subtraction. This research design implies that any *concern* matters as much as any *strength* in a category.

By contrast, we do not need to impose such an assumption in the DEA model. In the empirical application, we consider separately the strength and concern levels of the six CSP categories that correspond to the W1 to W6 categories in Table 5 (so in total the DEA model uses 12 variables). We follow the original measurement scale of KLD data; that is, different CSP levels are only compared in an ordinal fashion, and the relative weights for different categories are determined by the DEA program as described. We conduct the DEA analysis independently for the three industries considered.

As we noted earlier, the KLD database consists of data that only have meaning in the ordinal scale. To maintain research validity, scientific analysis should therefore be carried out in concordance with the measurement scale of observations. We therefore adopt the DEA model Cook and Zhu (2006) developed for ordinal data.

4.3 Results

We apply both the DEA method and the three weighting schemes (i.e., equal weights and those developed by Ruf et al. [1998] and by Waddock and Graves [1997b]) to the KLD 2007 dataset. Table 6 shows the descriptive statistics of the scores from the different approaches. We obtained the scores by applying the weights from Table 5 to Eq. (5).

***

[Insert Table 6 about here]

***

Figure 3 provides the distribution plots of KLD scores. These figures in general appear to be bell-shaped, although none of them can pass the Shapiro-Wilk test for normality at the 1-percent significance level.

***

[Insert Figure 3 about here]

***

From the DEA scores we can obtain additional insights. For all three industries, only a small proportion of the firms are CSP-efficient (manufacturing: 3.17%; finance: 4.08%; service: 2.63%). The average DEA score of all inefficient firms from the three industries is approximately 0.976; this finding indicates that on average the inefficient firms from our sample can reduce their CSP concern levels by 2.4%, given their current level of CSP strengths. Obtaining high average efficiency in empirical DEA applications is not uncommon. For instance, Majumdar and Marcus (2001) report an average efficiency score of .78 with a standard deviation of .24. Similarly, Goto and Tsutsui (1998) report an average efficiency score of .90 for U.S. utilities for 1984–93. The low average inefficiency in our current study is also due to the low level of variations in the KLD data. Table 7 contains the percentages of the firms with the highest strength scores (e.g., 5 out of 5 possible points) and the lowest concern scores (e.g., 0 out of 7 possible points) in the six KLD concern and strength categories. For most categories, we observe a high proportion of firms that are ranked best. This finding suggests a majority of firms in the KLD sample are located in the vicinity of the efficient frontier and explains why we have a high average efficiency score.

***

[Insert Table 7 about here]

***

From the statistics Table 6 reports, we can see that financial firms are in general more CSP efficient than the manufacturing and service industries. The mean of the efficiency score of the financial sector is higher than the two other sectors. Figure 3 gives a good indication that the CSP efficiency frontiers of the manufacturing and service industries are defined by relatively few leading firms, whereas the majority of other firms in the sample are lagging behind.

Our results also illustrate that previous aggregation approaches cannot always sort out CSP-efficient firms. In Table 8, for example, we list the 34 CSP-efficient firms (i.e., DEA score equal to one) and their ranks with the weighted aggregation scores. In Table 9, we include the list of the weighted aggregation scores of the 34 firms from the bottom of DEA rankings. In Table 8, although firms with the highest 10 aggregation scores are also CSP efficient, many CSP-efficient firms are ranked below 20 percent. Thus having these "outliers" in the results means we will lose some key information in the data when adopting these aggregation approaches. Because the aggregation approach keeps the trade-off between strengths and concerns, firms that excel in particular CSP dimensions will not necessarily be considered efficient benchmarks. However, we can compare the evaluated firm to the efficient target identified by DEA, and we will be able to see the relative position of the evaluated firm for each CSP dimension.

***

[Insert Table 8 and Table 9 about here]

***

The primary reason for the difference in the rankings across methodologies is that in the aggregation approaches, firms' CSP scores depend on the firm performance within specific CSP categories, whereas DEA considers individual CSP concerns and strengths and allows for compensation across concerns or strengths in different categories. The DEA model will seek the optimal trade-offs between different concerns and strengths for the evaluated firm (i.e., weights attached to concerns and strengths). With the aggregation methods, firms will tend to receive low aggregation scores if they underperform in specific *categories* (i.e., high concern and low strength in the *same* category). For example, if a firm has a high strength and a high concern within a specific category, the final score will still be average because strengths and weaknesses cancel each other out.

In order to understand differences in ranking, we describe in detail the ranking of three firms: A, B, and C, presented in Table 8. These three firms are efficient with the DEA approach. In the aggregation method, however, only Firm A obtains a high ranking (1). Firm B and C obtain lower rankings (from 200 to 700 for both). Looking at the individual scores for each of these firms can help us understand this difference (see Table 10).

***

[Insert Table 10 about here]

***

Firm A receives low concerns and high strengths across different KLD categories. So Firm A is efficient in the DEA model, whose scores are computed based on concerns and high strengths scores in Table 10. The last column in the table shows the maximum score in the sample for individual variables; note the minimum score is zero. Firms B and C obtain lower scores in the aggregation methodology. Firm B has a comparatively low score in the Diversity category, and Firm C has a low score in the Product category (see Table 10). We can trace their low performance in these categories back to their original KLD scores in these two categories in Table 10. Firms B and C, however, both attain the efficiency status in the DEA model. As noted, DEA allows for making trade-offs between concerns or strengths. For example, Firm A is low in its Diversity strength. Then the DEA model will tend to reduce the weight for the Diversity strength and assign a higher weight to the Environmental strength, in which firm B excels. Firms B and C are both considered efficient because their relative leading status in certain concern or strength items is enough to cover those items in which they lag behind.

In contrast to Table 8, Table 9 lists the weighted aggregation scores of the firms in the bottom 34 firms of the DEA rankings. The results in Table 9 show that the DEA rankings are more consistent with those of the previous approaches than they are for efficient firms. This observation is not surprising since most of these companies score high on the concerns and low on their strengths.

We take the example of Firm D, which is ranked below 1000 in the aggregation and DEA approaches. Table 10 shows Firm D's KLD scores. Firm D has the lowest scores for the Community and Environment categories in the entire sample, and its scores for other categories are also relatively low. Hence Firm D has low rankings for the aggregation scores. Regarding the KLD disaggregated scores at the "item" level, Firm D scores low on 10 out of 12 items, which is why firm D also receives a low DEA ranking.

4.4 Statistical comparisons of CSP-KLD scores

In this section, we run three different tests to compare distribution of the CSP scores the DEA and weighted aggregation approaches generate. We first use the Wilcoxon test for medians then the Levene test for equality of variance and finally the Kendall's tau rank correlation, which measures the degree of correspondence between the two rankings. We find that DEA differs from the aggregation approaches in terms of its median and variance at the 1-percent significance level. Likewise, the aggregation methods also differ along these two parameters.

We find a positive and significant correlation between DEA and the aggregation methods. However, the correlation is stronger among the aggregation methods.

4.4.1 Test for median and variance

We use the nonparametric Wilcoxon matched-pairs signed-ranks test to examine the difference in the medians of the scores using fixed-weight specifications and DEA. In total, we perform the test for six different pairs of scores. Results of the pair-wise comparisons indicate all KLD-CSP distributions are significantly different at the 1-percent significance level. We found the DEA scores have statistically dissimilar distributions than the three types of aggregation scores (i.e., all p-values are close to 0). The finding implies the KLD-CSP distributions are sensitive to different weight configurations. Future studies should pay attention to this sensitivity.

Next we used the Levene test to assess the equality of variances. The Levene test is robust under non-normality. For all pairs of score distributions, the test rejects that any pair of score distribution has equal variance at the 1-percent significance level. This result indicates that the distributions of weighted aggregations scores tend to have different variances. We find that when combined with the Wilcoxon test results, the weighted aggregation approaches could produce sensitive distributions even with small changes in weights.

4.4.2 Correlation test

The Kendall's tau coefficient is a non-parametric statistic commonly used to measure the degree of correspondence between two rankings. Table 11 shows the Kendal's tau correlation coefficients.

***

[Insert Table 11 about here]

***

All Kendall's tau coefficients in the table are significant at the 1-percent significant level. The coefficients reveal that the ranking of the three aggregation scores are more highly correlated with each other (from 0.77 to 0.91) than with the ranking of DEA scores (around 0.49). Among the three aggregation approaches, rankings based on Ruf et al. (1998) and Waddock and Graves (1997b) are more strongly correlated because the aggregation weights they used are similar (see Table 5). The correlation between the DEA ranking and other rankings is not as high mainly because of the methodological difference the previous section explained in detail.

5 DISCUSSION AND CONCLUSION

The aggregation methodology of multiple CSP metrics most of the literature has adopted poses major methodological challenges to researchers: the aggregation score lacks comparability and interpretability, and it ignores the ordinal nature of CSP data. We tested the validity of the traditional CSP aggregation approach using the 2007 KLD data and found this approach exhibits statistically different distributions when different weights are applied. Specifically, we analyzed the mean and variance of the weighted aggregation scores and found that minor changes in the aggregation weights could lead to significant changes in these two important parameters of score distributions.

In this paper, we provide an effective methodology to circumvent these issues. We take an efficiency perspective and utilize the DEA model for ordinal data to create a single CSP efficiency index from the KLD data. In our model, the strengths and concerns of CSP are separate components of a firm's composite CSP index. Our application is distinct from the eco-efficiency studies that deal with environmental impacts rather than CSP data. Our DEA approach allows us to incorporate "soft" data represented on an ordinal scale. The DEA approach has several advantages over the linear aggregation methods: it does not require a priori weight rankings or specifications for CSP criteria; the DEA score has a direct interpretation (what is your CSP compared with the best in class?); and it can be applied to a diverse set of measures of CSP. For example, the DEA method can include soft and hard measures of performance. Researchers can also expand it to compare eco-efficiency to productive efficiency (Chen et al. 2010).

For the empirical CSP literature, the DEA efficiency score also provides an ideal proxy for CSP in econometric models as a dependent or independent variable (e.g., Delmas and Tokat 2005, Delmas et al. 2007, Simar and Wilson 2007). In many CSP studies, researchers are interested in understanding the relationship between CSP and some exogenous variables, or the relationship between the CSP components. So we should remember that the KLD data, and hence the resultant DEA scores, are meaningful in the ordinal sense only, and this measurement scale should be carried over in the subsequent analysis. Thus, for example, ordinal regression techniques may be more appropriate when the KLD scores are regarded as the dependent variable.

The output from our DEA model, the efficiency scores, can also help devise strategies at the operational level to improve the firm's CSP. For example, a CEO could refer to the efficient benchmark and make the resource and operation planning decisions to improve the firm's CSP. As many companies are paying greater attention to the CSP of their supply chains, such companies can use our model to benchmark their supply chain partners' CSP. Information such as efficiency scores and benchmark targets can prove useful in supplier base management and in monitoring the CSP of the company's supply chain.

Finally, we provide some limitations of the current study and suggestions for extension. The standard DEA model requires the evaluated firm to minimize inputs or maximize outputs proportionally to reach the efficient frontier. Recently, researchers have developed DEA models that allow for non-proportional changes in inputs or outputs (e.g., Cooper et al. 2006, chap.4). Further research could develop similar models amenable to ordinal data.

In this paper, we did not impose weight restriction constraints on the u_r and v_i in the DEA model (2). However, the evaluator may have personal preferences regarding CSP dimensions, or face an exogenously given rule on how to weight different CSP dimensions. When priorities among CSP dimensions are precisely articulated, DEA is less instrumental for CSP evaluation (e.g., environmental strength *must be* of equal importance to social strength). However, when the prioritizing relationship is fuzzy or more flexible, we may apply weight restrictions to DEA to reflect the evaluator's specific preferences for different CSP criteria. For example, the evaluator can determine that the weight for the environmental strength in the DEA model must be no less than the weight for the social strength in the DEA model. When the data lack variability or the sample is not sufficiently large, imposing weight restrictions can in general increase the discrimination of DEA results (i.e., a wider range of efficiency scores). See chapter 6 of Cooper et al. (2006) for a general discussion on weight restrictions in DEA, and Cook and Zhu (2006) for the exact implementation formulation.

Because of our methodological focus, we only use the cross-sectional 2007 KLD data in this paper. However, we can expect a firm's current CSP score to influence its future CSP scores. Moreover, the intensity and property of this dynamic effect can differ for strengths and concerns. Thus future research can conduct longitudinal analysis and investigate the dynamic interrelationships between concerns, strengths, and CSP over time, which can reveal further insights into the evolution of CSP (Chen and van Dalen 2010). Finally, another interesting direction is to combine CSP information with financial performance measures to form a more comprehensive corporate performance evaluation using DEA.

References

Agle, B., R. Mitchell, J. Sonnenfeld. 1999. Who matters to CEOs? An investigation of stakeholder attributes and salience, corporate performance, and CEO values. Academy of Management Journal 42(5): 507-525.

Albinger, H., S. Freeman. 2000. Corporate social performance and attractiveness as an employer to different job seeking populations. Journal of Business Ethics 28(3): 243-253.

Aupperle, K. E. 1991. The use of forced choice survey procedures in assessing corporate social orientation. J. E. Post eds. Research in Corporate Social Performance and Policy 12, pp. 269-280.

Backhaus, K. B. and B. A. Stone, K. Heiner. 2002. Exploring the relationship between corporate social performance and employer attractiveness. Business & Society 41(3): 292-318.

Bartkus, B., M. Glassman. 2008. Do firms practice what they preach? The relationship between mission statements and stakeholder management. Journal of Business Ethics 83(2): 207-216.

Bendheim, C. L., S. Waddock, S. Graves. 1998. Determining best practice in corporate-stakeholder relations using Data Envelopment Analysis. Business & Society 37(3): 305-338.

Berman, S., A. Wicks, S. Kotha, T. Jones. 1999. Does stakeholder orientation matter? The relationship between stakeholder management models and firm financial performance. Academy of Management Journal 42(5): 488-506.

Bird, R., A. D. Hall, F. Momente, F. Reggiani. 2007. What corporate social responsibility activities are valued by the market? Journal of Business Ethics 76(2): 189-206.

Bouquet, C., Y. Deutsch. 2008. The impact of corporate social performance on a firm's multi-nationality. Journal of Business Ethics 80(4): 755-769.

Bowman, E. H., M. Haire. 1975. A strategic posture toward corporate social responsibility. California Management Review 18(2): 49-58.

Briscoe, F., S. Safford. 2008. The Nixon-in-China effect: Activism, imitation, and the institutionalization of contentious practices. Administrative Science Quarterly 53(3): 460-491.

Bowen, F.E., P.D. Cousins, R.C. Lamming, A. C. Faruk. 2001. The role of supply management capabilities in green supply. Production and operations management 10 (2): 174-189.

Brown, B., S. Perry. 1994. Removing the financial performance halo from Fortune's "most admired" companies. Academy of Management Journal 37(5): 1347-1359.

Carroll, A.B. 1999. Corporate social responsibility: evolution of a definitional construct. Business & Society 38(3): 268-295.

Carter, C.R. 2008. Purchasing and social responsibility: a replication and extension. Journal of Supply Chain Management 40(4): 4-16.

Charnes, A., W. Cooper, E. Rhodes. 1978. Measuring the efficiency of decision making units. European Journal of Operational Research 2: 429-444.

Chatterji, A. K., D. I. Levine. 2006. Breaking down the wall of codes: Evaluating non-financial performance measurement. California Management Review 48(2): 29-51.

Chatterji, A. K., D. I. Levine, M. W. Toffel. 2009. How well do social ratings actually measure corporate social responsibility? Journal of Economics & Management Strategy 18(1): 125–169.

Chen, C.-M, M.A. Delmas, M. Montes-Sancho. 2010. Eco- vs. productive efficiency: A new approach to effective and comparative performance analysis. Working Paper. UCLA Center for Corporate Environmental Performance.

Chen, C.-M, J. van Dalen. 2010. Measuring dynamic efficiency: Theories and an integrated methodology. European Journal of Operational Research 203(3): 749-760.

Chen, J., D. Patten, R. Roberts. 2008. Corporate charitable contributions: A corporate social performance or legitimacy strategy? Journal of Business Ethics 82(1): 131-144.

Cho, C., D. Patten, R. Roberts. 2006. Corporate political strategy: An examination of the relation between political expenditures, environmental performance, and environmental disclosure. Journal of Business Ethics 67(2): 139-154.

Clarkson, M. B. E. 1995. A stakeholder framework for analyzing and evaluating corporate social performance. Academy of Management Review 20(1): 92-117.

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

Cooper, W., L. Seiford, K. Tone. 2006. Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software. Springer Verlag.

Dawkins, C. 2002. Corporate welfare, corporate citizenship, and the question of accountability. Business & Society 41 (3): 269-291.

Deckop, J., K. Merriman, S. Gupta. 2006. The effects of CEO pay structure on corporate social performance. Journal of Management 32(3): 329-342.

Delmas, M.A., Y. Tokat. 2005. Deregulation, Efficiency and Governance Structures: The U.S. Electric Utility Sector. Strategic Management Journal 26: 441-460.

Delmas, M.A, M. V. Y. Russo, M. J. Montes-Sancho. 2007. Deregulation and environmental differentiation in the electric utility industry. Strategic Management Journal 28: 189-209.

Delmas, M.A, V. Doctori-Blass. 2010. Corporate environmental performance: The trade-offs of sustainability ratings. Business and the Environment 19: 245-260.

Delquié, P. 1997. Bi-matching: A new preference assessment method to reduce compatibility effects. Management Science 43: 640–658.

Drumwright, M.E. 1994. Socially responsible organizational buying: environmental concern as a noneconomic buying criterion. Journal of Marketing 58(3): 1-19.

Dyckhoff, H., K. Allen. 2001. Measuring ecological efficiency with data envelopment analysis (DEA). European Journal of Operational Research 132(2): 312-325.

Färe, R., S. Grosskopf, C. Pasurka Jr. 2006. Social responsibility: US power plants 1985-1998. Journal of Productivity Analysis 26 (3): 259-267.

Freeman, A.M. 2003. The measurement of environmental and resource values: theory and methods. RFF press.

Gond, J.-P., A. Crane. 2010. Corporate social performance disoriented: Saving the lost paradigm? Business & Society, forthcoming.

Goto, M., M. Tsutsui. 1998. Comparison of productive and cost efficiencies among Japanese and US electric utilities. Omega 26: 177-194.

Graves, S., S. Waddock. 1994. Institutional owners and corporate social performance. Academy of Management Journal 37(4): 1034-1046.

Griffin, J. 2000. Corporate social performance: Research directions for the 21st century. Business & Society 39(4): 479-491.

Griffin, J., J. Mahon. 1997. The corporate social performance and corporate financial performance debate: Twenty-five years of incomparable research. Business & Society 36(1): 5-31.

Hillman, A., G. Keim. 2001. Shareholder value, stakeholder management, and social issues: what's the bottom line? Strategic Management Journal 22 (2): 125-139.

Hirsch, P.M., D.Z. Levin. 1999. Umbrella advocates versus validity police: a life-cycle model. Organization Science 10(2): 199-212.

Hull, C., S. Rothenberg. 2008. Firm performance: the interactions of corporate social performance with innovation and industry differentiation. Strategic Management Journal 29 (7): 781-789.

Igalens, J., J.-P. Gond. 2005. Measuring Corporate Social Performance in France: A Critical and Empirical Analysis of ARESE Data. Journal of Business Ethics 56(2): 131-148.

Johnson, R., D. Greening. 1999. The effects of corporate governance and institutional ownership types on corporate social performance. Academy of Management Journal 42(5): 564-576.

Kane, G., U, Velury, B. Ruf. 2005. Employee relations and the likelihood of occurrence of corporate financial distress. Journal of Business Finance & Accounting 32 (5-6): 1083-1105.

Kempf, A., P. Osthoff. A. Platz. 2007. The effect of socially responsible investing on portfolio performance. European Financial Management 13(5): 908-922.

Kennelly, J., E. Lewis. 2002. Degree of internationalization and corporate environmental performance: Is there a link. International Journal of Management 19(3): 478-489.

Kuosmanen, T., M. Kortelainen. 2007. Valuing environmental factors in cost–benefit analysis using data envelopment analysis, Ecological Economics 62: 56- 65.

Landier, A., V. Nair, J. Wulf. 2009. Trade-offs in staying close: Corporate decision making and geographic dispersion. Review of Financial Studies 22 (3): 1119-1148.

Loureiro, M.L, J. Lotade. 2005. Do fair trade and eco-labels in coffee wake up the consumer conscience? Ecological Economics 53(1): 129-138.

Luce, R., A. Barber, A., Hillman. 2001. Good deeds and misdeeds: A mediated model of the effect of corporate social performance on organizational attractiveness. Business & Society 40 (4): 397-415.

Majumdar, S.K., A.A Marcus. 2001. Rules versus discretion: The productivity consequences of flexible regulation. Academy of Management Journal 44 (1): 170-179.

Marquis, C., M. Glynn, G. Davis. 2007. Community isomorphism and corporate social action. The Academy of Management Review 32 (3): 925-945.

Mattingly, J., S. Berman. 2006. Measurement of corporate social action: discovering taxonomy in the Kinder Lydenburg Domini ratings data. Business & Society 45(1): 20-46.

McWilliams, A., D. Siegel. 2000. Corporate social responsibility and financial performance: Correlation or misspecification? Strategic Management Journal 21: 603-609.

McWilliams, A, D. Siegel. 2001. Corporate social responsibility: A theory of the firm perspective. Academy of Management Journal 26: 117-127.

McGuire, J., S. Dow, K. Argheyd. 2003. CEO incentives and corporate social performance. Journal of Business Ethics 45 (4): 341-359.

Mitchell, R., B. Agle, D. Wood. 1997. Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. Academy of Management Review 22(4): 853-886.

Neiling, E., E. Webb. 2009. Corporate social responsibility and financial performance: the “virtuous circle” revisited. Review of Quantitative Finance and Accounting 32: 197–209.

Neubaum, D., S. Zahra. 2006. Institutional ownership and corporate social performance: The moderating effects of investment horizon, activism, and coordination. Journal of Management 32 (1): 108-131.

Podinovski, V.V., E. Thanassoulis. 2007. Improving discrimination in data envelopment analysis: Some practical suggestions. Journal of Productivity Analysis 28(1): 117-126.

Rehbein, K., S. Waddock, S. B. Graves. 2004. Understanding shareholder activism: which corporations are targeted? Business & Society 43(3): 239-267.

Rowley, T., S. Berman. 2000. A brand new brand of corporate social performance. Business & Society 39(4): 397-418.

Ruf, B., K. Muralidhar, K. Paul. 1998. The development of a systematic, aggregate measure of corporate social performance. Journal of Management 24(1): 119-133.

Ruf, B., K. Muralidhar, R. Brown, J. Janney, K. Paul. 2001. An empirical investigation of the relationship between change in corporate social performance and financial performance: A stakeholder theory perspective. Journal of Business Ethics 32(2): 143-156.

Seuring, S., M. Müller. 2008. From a literature review to a conceptual framework for sustainable supply chain management. Journal of Cleaner Production 16(15): 1699-1710.

Shropshire, C., A. Hillman. 2007. A longitudinal study of significant change in stakeholder management. Business & Society 46(1): 63-87.

Simar, L., P.W. Wilson. 2007. Estimation and inference in two-stage, semi-parametric models of production processes. Journal of Econometrics 136(1): 31-64.

Simerly, R., 2003. An empirical examination of the relationship between management and corporate social performance. International Journal of Management (20)3: 353-359.

Srivastava, S.K. 2007. Green supply-chain management: A state-of-the-art literature review. International Journal of Management Reviews 9(1): 53-80.

Strike, V.M. and J. Gao, P. Bansal. 2006. Being good while being bad: social responsibility and the international diversification of US firms. Journal of International Business Studies 37: 850-862.

Thomas, A., R. Simerly. 1995. Internal determinants of corporate social performance: the role of top managers. D. P. Moore eds. Academy of Management Journal Best Paper Proceedings, Madison, WI. pp. 411-415.

Turban, D., D. Greening. 1996. Corporate social performance and organizational attractiveness to prospective employees. Academy of Management Journal 40(3): 658-672.

Van der Laan, G., H. Van Ees, A. Van Witteloostuijn. 2008. Corporate social and financial performance: An extended stakeholder theory, and empirical test with accounting measures. Journal of Business Ethics 79(3): 299-310.

Waddock, S., 2003. Myths and realities of social investing. Organization and Environment 16: 369-380.

Waddock, S., S. Graves. 1997a. Quality of management and quality of stakeholder relations: are they synonymous? Business & Society 36(3): 250-279.

Waddock, S., S. Graves. 1997b. The corporate social performance-financial performance link. Strategic Management Journal 18(4): 303-319.

Waldman, D., D. Siegel, M. Javidan. 2006. Components of CEO transformational leadership and corporate social responsibility. Journal of Management Studies 4(8): 1703-1725.

Webb, E. 2004. An examination of socially responsible firms' board structure. Journal of Management and Governance 8 (3): 255-277.

Wokutch, R.E., E.W. McKinney. 1991. Behavioral and perceptual measures of corporate social performance. In J. E. Post (Ed.), Research in corporate social performance and policy. Vol. 12. JAI Press, Greenwich, CT, 309-330.

Wolfe, R. 1991. The use of content analysis to assess corporate social responsibility. J. E. Post eds. Research in Corporate Social Performance and Policy 12: 281-308.

Zahra, S. A., B.M. Oviatt, K. Minyard. 1993. Effects of corporate ownership and board structure on corporate social responsibility and financial performance. Academy of Management Best Paper Proceedings, pp. 336-340.

Zhou, P., B. W. Ang, K. L. Poh. 2009. A survey of data envelopment analysis in energy and environmental studies. European Journal of Operational Research 189(1): 1-18.

Zhu, J. 2003. Imprecise data envelopment analysis (IDEA): A review and improvement with and application. European Journal of Operational Research 144: 513-529.

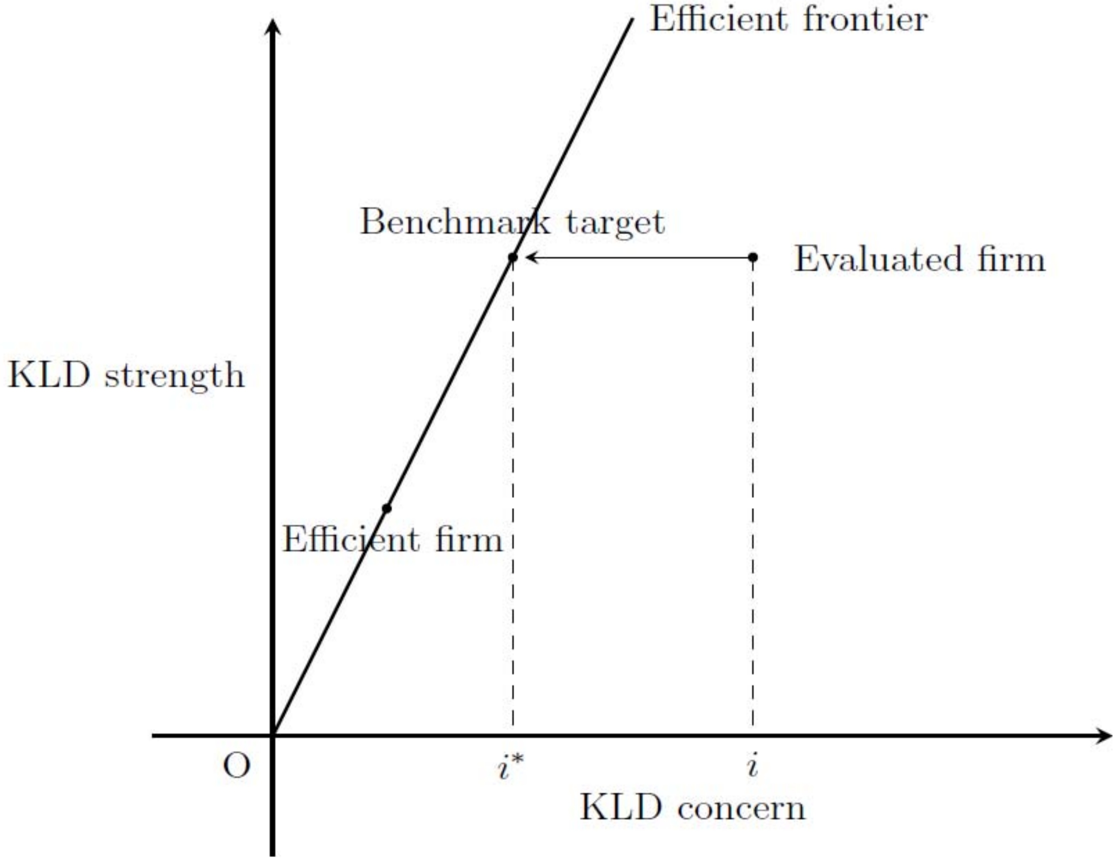


Figure 1 Graphical illustration of the DEA approach

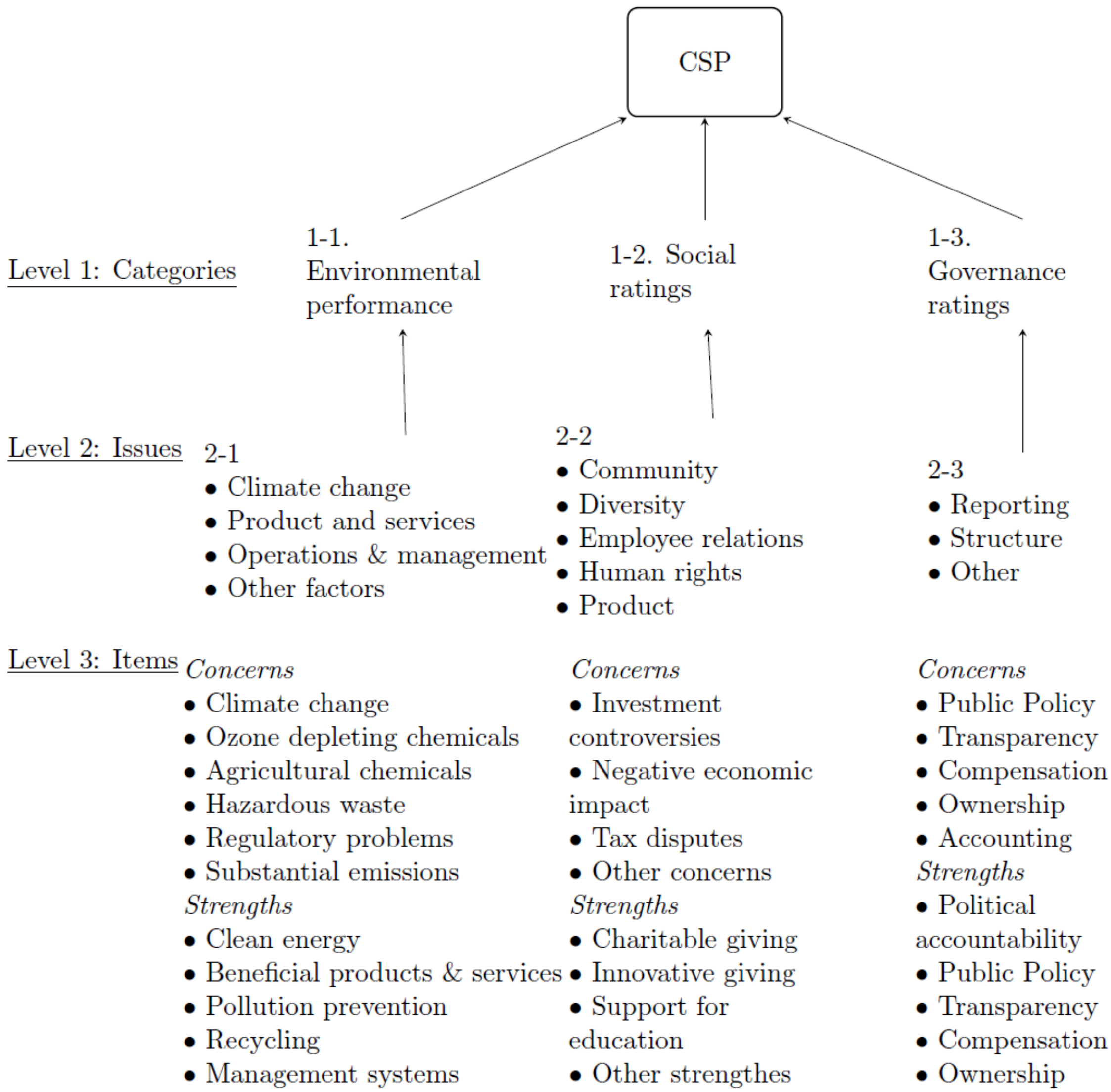


Figure 2 Illustration of the KLD structure ver. 2007

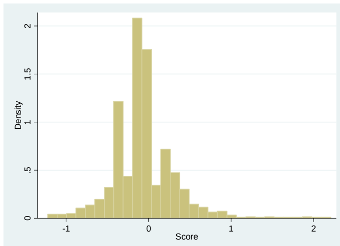
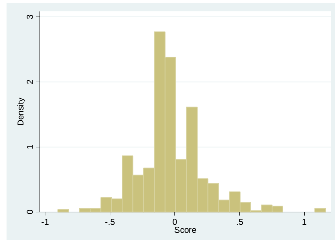
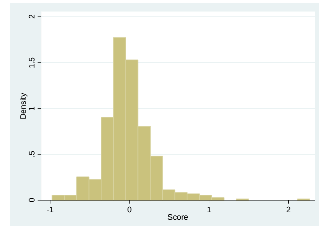
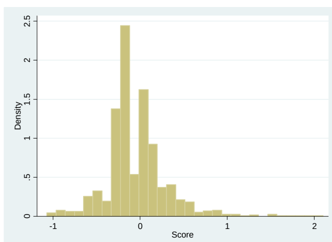
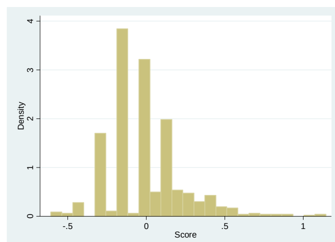
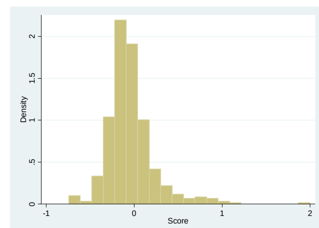
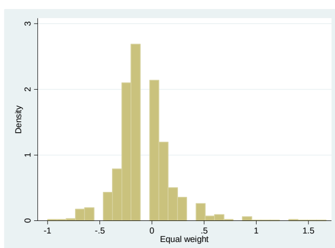
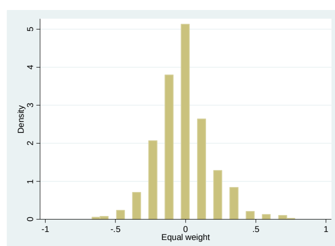
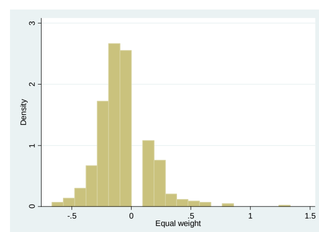
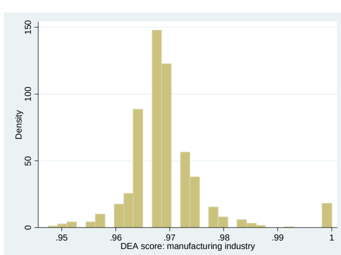
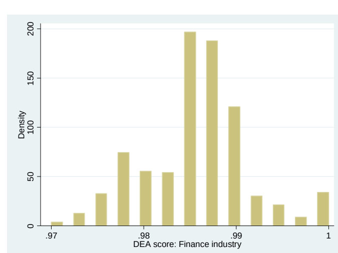
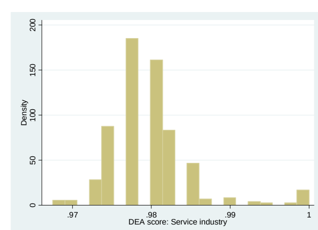
<table>
<tr><th></th><th>Manufacturing industry</th><th>Finance industry</th><th>Service industry</th></tr>
<tr><td>Ruf et al. (1998)</td>
<td></td>
<td></td>
<td></td></tr>
<tr><td>Waddock and Grave (1997)</td>
<td></td>
<td></td>
<td></td></tr>
<tr><td>Equal weights</td>
<td></td>
<td></td>
<td></td></tr>
<tr><td>DEA scores</td>
<td></td>
<td></td>
<td></td></tr>
</table>

Figure 3 2007 KLD scores using weighted aggregation

Table 1 Publications counts by journals

Journal titles	Authors	No. of papers
Journal of Business Ethics	Albinger (2000); Ruf et at. (2001); McGuire et al. (2003); Igalens and Gond (2005); Cho et al. (2006); Bartkus et al. (2007); Bouquet and Deutsch (2007); Bird et al. (2007); Chen et al. (2008); Van der Laan et al. (2008).	10
Business & Society	Griffin and Mahon (1997); Waddock and Graves (1997a); Luce et al. (2001); Backhaus et al. (2002) Dawkins (2002); Mattingly and Berman (2006); Rehbein et al. (2004); Shropshire and Hillman (2007).	8
Academy of Management Journal	Agle et al. (1999); Graves and Waddock (1994); Brown and Perry (1994); Thomas and Simerly (1995); Turban and Greening (1996); Berman et al. (1999); Johnson and Greening (1999);	7
Strategic Management Journal	Waddock and Graves (1997b); Hillman and Keim (2001); Hull and Rothenberg (2008).	3
International Journal of Management	Kennelly and Lewis (2002); Simerly (2003)	2
Journal of Management	Ruf et al. (1998); Deckop et al. (2006); Neubaum and Zahra (2006)	3
Academy of Management Review	Marquis et al. (2007)	1
Administrative Science Quarterly	Briscoe and Safford (2008)	1
Journal of International Business Studies	Strike et al. (2006)	1
Journal of Management Studies	Waldman et al. (2006)	1
Review of Financial Studies	Landier et al. (2009)	1
Others	Webb (2004); Kane et al. (2005) Kempf et al. (2007); Chatterji et al. (2009); Neiling and Webb (2009)	5
Total		43

Table 2 Example KLD strength, concern rankings, and the DEA benchmark

		Efficiency score=0.9475	
		Input-oriented benchmark target	
Categories	KLD Score	(KLD score x Efficiency score)	Rounded target
Concern			
Community	4	3.8	3
Diversity	2	1.9	1
Employee	3	2.8	2
Environment	6	5.7	5
Humanity	2	1.9	1
Product	3	2.8	2
Strength			
Community	2	2	2
Diversity	4	4	4
Employee	3	3	3
Environment	2	2	2
Humanity	1	1	1
Product	1	1	1

Table 3 KLD statistics

		No. of variables		Concerns		Strength	
Category	Issues	Concern	Strength	Mean	Std.	Mean	Std.
Environmental performance	Climate change	1	1	0.048	0.213	0.034	0.182
	Product and services	2	1	0.007	0.09	0.019	0.136
	Operations and management	3	3	0.138	0.479	0.07	0.313
	Others	1	1	0.007	0.09	0.019	0.136
Social ratings	Community	4	7	0.111	0.335	0.117	0.447
	Diversity	3	8	0.431	0.513	0.606	1.046
	Employee relations	5	6	0.524	0.71	0.275	0.6
	Human rights	4	3	0.046	0.232	0.005	0.069
	Product	4	4	0.232	0.575	0.044	0.216
Governance ratings	Reporting	2	2	0.001	0.026	0.035	0.199
	Structure	3	2	0.356	0.483	0.155	0.362
	Other	1	1	0.036	0.186	0.002	0.045

Table 4 SIC Industry classification of the KLD 2007 sample

Industry	No. of firms
Manufacturing	1072
Finance, Insurance, and Real Estate	661
Services	457
Retail Trade	192
Mining	132
Transportation, Communications, Electric, Gas, and Sanitary Services	288
Wholesale Trade	79
Construction	37
Public Administration	12
Agriculture, Forestry, and Fishing	6
Total	2936

Table 5 The aggregation weights

Category	Equal weights	Ruf et al. (1998)	Waddock and Graves (1997b)
W1-Community	0.111	0.125	0.148
W2-Diversity	0.111	0.152	0.136
W3-Employee relation	0.111	0.183	0.168
W4-Environment	0.111	0.141	0.142
W5-Human	0.111	0.152	0.136
W6-Product	0.111	0.228	0.154
W7-Nuclear power	0.111	0.089	0.074
W8-Mulitary contract	0.111	0.086	0.050
W9-South Africa	0.111	0.076	0.046

Table 6 Descriptive statistics of CSP scores

Industry		Equal weights	Ruf et al. (1998)	Waddock and Graves (1997b)	Rank DEA
Manufacturing (n=1072)	Mean	-0.0661	-0.6181	-0.0454	0.9695
	Std.	0.2780	0.3810	0.3499	0.0076
	Min	-0.9999	-1.2226	-1.0720	0.9475
	Max	1.6667	2.2069	2.1000	1.0000
Finance (n=661)	Mean	-0.0108	-0.0241	-0.0052	0.9857
	Std.	0.1944	0.2609	0.2429	0.0059
	Min	-0.6666	-0.9000	-0.6080	0.9700
	Max	0.7777	1.1638	1.1440	1.0000
Services (n=457)	Mean	-0.0557	-0.0480	-0.0335	0.9797
	Std.	0.2219	0.3237	0.2835	0.0053
	Min	-0.6667	-0.9757	-0.7440	0.9675
	Max	1.3333	2.2686	1.9980	1.0000

Table 7 Percentages of firms in our sample that have the best ranking in different KLD items

KLD concern and strength variables	% of firms in the manufacturing industry that have the best ranking	% of firms in the finance industry that have the best ranking	% of firms in the service industry that have the best ranking
KLD Concern			
Community	92.07%	83.18%	95.84%
Diversity	56.44%	62.33%	57.55%
Employee	52.71%	77.76%	60.39%
Environment	81.34%	99.39%	99.56%
Humanity	95.24%	98.94%	96.28%
Product	82.65%	83.36%	85.12%
KLD Strength			
Community	92.16%	84.72%	96.72%
Diversity	66.04%	66.41%	60.61%
Employee	72.67%	81.85%	86.21%
Environment	83.58%	99.39%	97.81%
Humanity	99.44%	99.70%	99.78%
Product	92.91%	97.88%	97.16%
Average	80.60%	86.24%	86.09%

Table 8 CSP efficient firms in the manufacturing industry and their ranks (n=1072)

Firm names	Ranking according to Ruf (1998)	Ranking according to Waddock and Graves (1997b)	Ranking according to equal weights	Ranking according to DEA scores
3M Company	26	35	26	1
Advanced Micro Devices, Inc.	11	11	9	1
Agilent Technologies, Inc.	2	3	2	1
Alcoa, Inc.	524	718	254	1
Applied Materials, Inc.	10	9	9	1
Avon Products, Inc.	12	12	15	1
Bristol-Myers Squibb Company	62	25	55	1
Coca-Cola Company	533	266	707	1
Dell Inc.	30	18	55	1
Eastman Kodak Company	41	51	59	1
Ford Motor Company	55	31	32	1
General Mills Incorporated	3	4	6	1
General Motors Corporation	162	63	59	1
Graco Inc.	52	83	59	1
Green Mountain Coffee Roasters, Inc.	8	8	4	1
Harley-Davidson, Inc.	86	93	130	1
Herman Miller, Inc.	9	10	9	1
Hewlett-Packard Company	4	2	5	1
Intel Corporation	1	1	1	1
Johnson & Johnson	17	16	9	1
Kraft Foods, Inc.	21	27	17	1
Lilly (Eli) and Company	65	34	22	1
Mattel, Inc.	87	58	33	1
Molex Incorporated	80	142	141	1
Motorola, Inc.	5	5	8	1
NIKE, Inc.	35	19	9	1
PepsiCo, Inc.	108	81	26	1
Procter & Gamble Company	18	24	17	1
Steelcase, Inc.	14	17	15	1
Texas Instruments Incorporated	7	7	7	1
Timberland Company (The)	15	15	14	1
Valero Energy Corporation	258	280	707	1
Waters Corporation	85	137	141	1
Xerox Corporation	6	6	3	1

Table 9 Some CSP inefficient firms in the manufacturing industry and their ranks (n=1072)

Firm names	Ranking according to Ruf (1998)	Ranking according to Waddock and Graves (1997b)	Ranking according to equal weights	Ranking according to DEA scores
Exxon Mobil Corp.	1069	1068	1071	1072
Tyson Foods, Inc.	1071	1072	1070	1072
Cintas Corporation	1072	1070	1049	1070
Covidien Ltd.	1062	1061	1071	1070
Exide Technologies	1060	1066	1066	1070
Koppers Holdings, Inc.	1065	1060	1049	1070
Smithfield Foods, Inc.	1064	1067	1049	1070
Brunswick Corporation	1066	1059	1066	1065
Bunge Limited	1067	1069	1049	1065
Celanese Corporation	1070	1071	1069	1065
FMC Corporation	1053	1046	1030	1065
Goodyear Tire & Rubber	1056	1052	1049	1065
Grace (W.R.) & Co.	1048	1047	1049	1065
McDermott Intl, Inc.	1063	1065	1049	1065
Pilgrim's Pride Corp.	1068	1063	1049	1065
Archer-Daniels-Midland	1054	1054	1066	1057
Chemtura Corporation	1052	1054	1049	1057
Crown Holdings, Inc.	1061	1051	1049	1057
Hercules Incorporated	1021	1021	1030	1057
Ingersoll-Rand Company	1025	1033	990	1057
NL Industries, Inc.	1024	1020	990	1057
Seaboard Corporation	1058	1062	1049	1057
ConocoPhillips	1057	1064	1049	1057
Abitibi Bowater, Inc.	1049	1053	1030	1048
AK Steel Holding Corp.	1050	1058	1049	1048
Carolina Group	1059	1034	1030	1048
Caterpillar Inc.	1038	1027	1030	1048
Cytec Industries, Inc.	1046	1024	1030	1048
Honeywell Intl, Inc.	954	968	990	1048
Huntsman Corporation	1026	1019	1030	1048
L-3 Com. Inc.	1039	1024	990	1048
Masco Corporation	1040	1034	1030	1048
Mueller Water Products	1040	1034	990	1048
Murphy Oil Corporation	1045	1056	1049	1048

Table 10 Original KLD concern and strength scores

CSP items	Firm A	Firm B	Firm C	Highest score in sample(lowest in the parentheses)	Firm D
COM-con-#	1	2	1	3	3
COM-str-#	3	2	2	4	1
(Total strength-total concern)	*2*	*0*	*1*	*4 (-2)*	*-2*
DIV-con-#	0	0	0	2	1
DIV-str-#	5	1	4	6	3
(Total strength-total concern)	*5*	*1*	*4*	*6 (-2)*	*2*
EMP-con-#	0	2	1	4	2
EMP-str-#	5	2	0	5	2
(Total strength-total concern)	*5*	*0*	*-1*	*5 (-3)*	*0*
ENV-con-#	1	4	1	5	5
ENV-str-#	3	3	2	4	1
(Total strength-total concern)	*2*	*-1*	*1*	*4 (-4)*	*-4*
HUM-con-#	0	1	2	3	1
HUM-str-#	0	0	1	1	0
(Total strength-total concern)	*0*	*-1*	*-1*	*1 (-3)*	*-1*
PRO-con-#	1	0	3	4	2
PRO-str-#	1	0	0	2	0
(Total strength-total concern)	*0*	*0*	*-3*	*2 (-4)*	*-2*

Table 11 Kendall' tau rank correlations for the pooled sample (n=2190)

	Ruf et al. (1998)	Waddock and Graves (1997b)	Equal weights	DEA score
Ruf et al. (1998)	1			
Waddock and Graves (1997b)	0.9097*	1		
Equal weights	0.7817*	0.7744*	1	
DEA score	0.4977*	0.4925*	0.4940*	1

* significant at the 1% significance level