

DOES THE CHOICE OF CSR PERFORMANCE RATING PROVIDER MATTER?

Lies Bouten
IESEG School of Management
3, rue de la Digue
59000 Lille, France
E-mail : l.bouten@ieseg.fr

Charles H. Cho
ESSEC Business School
1 AVENUE BERNARD HIRSCH
CS 50105 CERGY
95021 CERGY PONTOISE CEDEX, FRANCE
E-mail: cho@essec.edu

Giovanna Michelin
Univeristy of Exeter Business School
Streatham Court, Rennes Drive
EX4 3BN Exeter, UK
E-mail: g.michelon@exeter.ac.uk

Robin W. Roberts
Kenneth G. Dixon School of Accounting
University of Central Florida
4000 Central Florida Blvd
Orlando, FL 32816-1400
E-mail: rroberts@ucf.edu

Any comments, feedback and suggestions welcome.

Please do not quote, cite, distribute or copy without the consent of the authors.

January 2016

DOES THE CHOICE OF CSR PERFORMANCE RATING PROVIDER MATTER?

ABSTRACT

At present, there are three major suppliers of corporate social responsibility (CSR) performance ratings: Thomson Reuters (ASSET4); MSCI (ESG Intangible Value Assessment (IVA)); and Sustainalytics (ESG Indicator). Almost all of the large-scale empirical studies dealing with CSR reporting use only one of the three providers as the source for their CSR performance measures. Given the subjective processes underlying the development of CSR performance ratings, prior research is beginning to question the construct validity of the ratings as well as their robustness across providers. The purpose of our study is to review the CSR performance constructs and proxies employed in studies published in a select set of influential journals and working paper series, and to analyze how the results of CSR studies may be influenced by the selection of CSR performance proxies. We believe that this research is critically important because the social and environmental accounting literature is building on significant results generated by different proxies for CSR performance.

DOES THE CHOICE OF CSR PERFORMANCE RATING PROVIDER MATTER?

INTRODUCTION

During the last two decades, many academic studies have been using raw data or ratings provided by social rating agencies to construct proxies of corporate social responsibility (CSR) performance. These CSR rating providers are private firms specialized in evaluating the corporate governance, environmental and social performance of publicly-traded corporations. As the demand for CSR information increased, the number of information providers also increased. Currently, there are three major suppliers of CSR ratings: (1) Thomson Reuters (ASSET4); (2) MSCI (ESG Intangible Value Assessment (IVA), formerly known as KLD); and (3) Sustainalytics (ESG Indicator).

A recent study of Chatterji et al. (forthcoming) demonstrates the lack of agreement across the ratings provided by six agencies: KLD, Asset4, Innovest, DJSI, FTSE4Good and Calvert. Such lack of convergence seems to originate from two sources: (1) the lack of a common theorization of CSR, i.e. raters' beliefs about the meaning of (corporate) social responsibility and (2) the lack of commensurability, i.e. differences in how raters measure the same construct. Hence, raters choose to measure different items, and also measure overlapping items differently. Bendell (2010) argues that the various rating agencies have different understandings of the concept of CSR. However, van den Heuvel (2012) suggests that some of the (reconstructed) overall and sub-scores of ASSET4 and Sustainalytics are sufficiently correlated, which led the author to conclude that these ratings, at least partially, measure the same construct. If differences in theorization and commensurability might be present, robustness across providers is not guaranteed. Therefore, the choice of CSR performance rating providers used to construct a proxy for a variable used in academic research, might impact the results; that is, proxies originating from different social rating agencies might lead to different academic results. This paper attempts to measure the extent

to which this is the case. This objective is highly relevant since almost all of the large-scale empirical studies dealing with CSR disclosures use at least one of the three providers as their main source for measures of CSR performance. In particular, our objective is to unpack the impact of using proxies based upon ratings provided by different social rating agencies by gaining a deeper understanding of both the theorization used by the different ratings providers and the differences in commensurability.

In order to do so, we use a multi-method approach (Cooper and Morgan, 2008). First, we conduct a series of analyses to identify (i) the impact of the differences in theorizations and commensurability between the rating providers and (ii) the impact of using different proxies originating from the same ratings provider. Second, we complement the statistical output with interview narratives. We have conducted semi-structured interviews with analysts and former analysts from these rating agencies and financial analysts who use(d) these ratings. The evidence suggests that inferences are highly dependent on the specific CSR proxy ratings as well as on the choice of the rating provider, thus posing a heightened concern over the adequacy of CSR performance constructs. Moreover, the findings illustrate that these CSR performance scores are largely based upon self-disclosed company information.

We believe that this study contributes to the academic debate on CSR performance ratings in three significant ways. First, we examine the research consequences that may arise from relying on a particular set of ratings. More specifically, we investigate the sensitivity of some prior research findings in relation to the choice of CSR performance rating providers. We argue that this research is critically important because the CSR reporting literature is building on significant results generated by different proxies originating from different CSR performance rating providers. Consequently, our research contributes to the literature by showing that the choice of CSR performance rating providers significantly matters, hence

allowing us to provide guidance to future researchers regarding potential biases that may arise from relying on one provider. The second contribution arises from providing indications regarding the extent to which the calculation of a proxy using data from a specific provider can influence research findings. The third contribution is that this study illustrates that authors should be cautious when using CSR rating based proxies for measuring CSR performance in CSR disclosure studies since the majority of their CSR performance indicators are derived from self-reported company information.

LITERATURE REVIEW

The databases of the three major suppliers of CSR performance ratings – Thomson Reuters (ASSET4), MSCI (ESG Intangible Value Assessment (IVA), formerly known as KLD) and Sustainalytics (ESG Indicator) – have been widely used across different disciplines. Table 1 provides an overview of academic studies published in the majors¹ of accounting, finance and management (see Swanson, 2010), *Accounting, Organizations and Society*, and the working paper series of the Harvard Business School that used a proxy based upon one of these three external ratings. This overview indicates that these ratings are most often used as a proxy for corporate responsibility performance (CSP) (e.g., Barnett & Solomon, 2012; Cheng et al., 2013; Dhaliwal et al., 2011; Graves and Waddock, 1994; McWilliams and Siegel, 2011; Surroca et al., 2013). Table 1 also indicates that these ratings are often used to measure stakeholder management in the management literature (e.g., Choi and Wang, 2009; Coombs and Gilley, 2005; Hillman and Keim, 2001; Kacperczyk, 2009). In addition, these ratings are sometimes used as proxies for transparency (Lewis et al., 2014); environmental risk management (Sharfman & Fernando, 2008); employee-firm relation (Bae

¹ Following Swanson (2010), the majors in accounting are: *Journal of Accounting and Economics*, *Journal of Accounting Research*, *The Accounting Review*, *Contemporary Accounting Research*. The majors in Finance are: *Journal of Finance*, *Journal of Financial Economics*, *Journal of Finance and Quantitative Analysis*. Finally, the majors in management are: *Academy of Management Journal*, *Academy of Management Review*, *Administrative Science Quarterly* and *Strategic Management Journal*.

et al., 2011; Luo et al., 2014), social irresponsibility (Muller & Krausler, 2011; Hoi et al., 2013; Wong et al., 2009) and integrated reporting (Serafeim, 2014a).

[Insert Table 1 about here]

Deckop et al. (2006, p. 334) describe KLD (current MSCI) as “the largest multidimensional CSP [corporate social performance] database available to the public”. First used as a proxy for CSP by Graves and Waddock in 1994, KLD data has been referred to as “the de facto [CSP] research standard at the moment” (Waddock, 2003, p. 369) and as the standard for quantitative measurement of corporate social actions (Mattingly and Berman, 2006). Consequently, KLD social ratings are among the most influential and the most widely accepted CSR measures used by academics (Chatterji et al., 2009). However, Table 1 indicates that while MSCI ratings (the former KLD ratings) retain the dominant position, ASSET4 and Sustainalytics ratings are also gaining in importance in the academic literature.

Originally, these ratings were developed, in large part, to reduce information asymmetry between firms and social investors by providing social investors information that reveals the extent to which firms’ behaviors are socially responsible (Chatterji et al., 2009). As a result, these external ratings have been constructed for benchmark reasons, i.e., to provide a parsimonious index that condenses the social, environmental and economic evaluation of a corporation into one number that reflects its CSR efforts.

Although it is challenging to summarize a company’s CSR performance into one or several scores, benchmarking brings many advantages. According to Graafland et al. (2004), benchmarking has not only the potential to enhance transparency, accountability and cross-company comparison, it also is a systematic approach that makes it possible to judge CSR performance in a straightforward way, as it allows everyone to judge the CSR performance of a company by evaluating only one or a few scores. Furthermore, some authors argue that benchmarking by external independent rating agencies might guarantee a more objective

view of CSR performance than would be provided by the company itself (Graves and Waddock, 1994; Sharfman, 1996). MSCI, for instance, uses a combination of surveys, financial statements, and articles in the popular press and academic journals, as well as government reports, to assess social performance (see, e.g. Kim et al., 2012). Although the construction of ratings remains subjective, analysts must value each action, including those with an outcome that is difficult to quantify (Graafland et al., 2004; Sharfman, 1996). Consequently, concerns have been raised regarding subjectivity as well as the omission of issues that are difficult to measure (Entine, 2003).

Given the subjective processes underlying the development of CSR ratings (Sharfman, 1996), prior and current research is beginning to question the construct validity of these ratings (Chatterji et al., 2009; Sharfman, 1996; Szwajkowski and Figlewicz, 1999). Chatterji et al. (2009) examined how well the KLD ratings provide transparency about past and likely future environmental performance. They found that while KLD “concern” ratings constitute fairly good summaries of past environmental performance, KLD environmental “strength” ratings do not accurately predict future environmental performance. Sharfman’s (1996) method for assessing construct validity was criterion validation. In this form of validation, one correlates the results of the measure in question with some other accepted method of measuring the CSP construct. More specifically, Sharfman (1996) investigates the correlation between the KLD scores, data from the Fortune magazine corporate reputation survey, and data derived from the holdings lists of 11 “social choice” mutual stock funds. The findings indicate that KLD scores correlate with other criterion variables. Consequently, the study of Sharfman (1996) suggests that KLD ratings correlate sufficiently with other measures of corporate social performance. Szwajkowski and Figlewicz (1999) also examined the convergent validity of KLD and the Fortune reputation survey. Since direct measurement of construct validity is nearly impossible, these researchers check for discriminant and

convergent validity. If both databases provide valid measurements, then their ratings should tend to converge regarding the common dimensions (convergent validity). Conversely, attributes that are conceptually independent of each other should be relatively unrelated, both within each database and between databases (discriminant validity). They conclude that their test results show sufficient convergent and discriminant validity. Although, these studies have found that CSR performance ratings – in particular KLD ratings – are reasonable proxies for CSR performance, researchers are starting to question the robustness of these CSR ratings across providers (Chatterji et al., in press; van den Heuvel, 2012). The fact that the providers are not fully transparent about their methods or research approaches (see Bendell, 2010) makes the comparison between CSR performance ratings a highly complicated exercise.

After selecting a specific rating provider, academic researchers face unique challenges to create a single score (Sharfman, 1996). Indeed, researchers need to decide whether they will include the overall CSR score (if provided) or whether they will construct a CSR performance index themselves. In the latter case, researchers have to decide on the number of dimensions to be included, the weighting of these different dimensions, etc. Table 2 provides an overview of how the studies that used the MSCI database, which does not provide an overall CSR performance score, constructed their own total CSR performance score. The wide variety of possibilities raises the question and potential concern about the extent to which the construction of the proxy influences the final results. Subtracting the concerns² from the strengths to arrive at a single net CSP score is a common (see, e.g., Graves and Waddock, 1994; Griffin and Mahon, 1997; Waddock and Graves, 1997; Johnson and Greening, 1999; Ruf et al., 2001) but highly debated practice since recent research shows that the KLD strengths and concerns measures are theoretically distinct and represent two independent constructs even though they may correlate with each other (Mattingly and

² KLD examines a number of positive indicators (strengths) and negative indicators (concerns) in each non-exclusionary dimension, but evaluates only negative indicators in each exclusionary dimension.

Berman, 2006; Strike et al., 2006). More specifically, these studies argue that a firm with five strengths and five concerns is fundamentally different from a firm with one strength and one concern, a distinction that may be lost in summing strengths and concerns (Chatterji et al., 2009). Walls et al. (2012) argue that combining strengths and concerns might lead to non-significant or spurious results. Therefore, recent studies (e.g., Flammer, 2013; Walls et al., 2012; Godfrey et al., 2009) use a total strengths and a total concerns score.

[Insert Table 2 about here]

CSR RATINGS: METHODOLOGIES AND METRICS

The findings of Chatterji et al. (in press) suggest that differences in CSR ratings originating from different rating providers may be explained by differences in the theorization of CSR. According to them, an analysis of common theorization requires researchers to examine: (i) which high level categories rating providers take into account, (ii) whether they screen out particular industries, and (iii) whether they normalize the ratings across industries. Below, we highlight these three elements for ASSET4, MSCI and Sustainalytics. However, given that most studies do not make *direct* usage of the ratings (see Tables 1 and 2) but rather compute a CSR index score based upon the raw data provided by these rating agencies, we also highlight the different methods and variables that are used to measure the similar constructs. The latter refers to differences in commensurability. The most important elements are summarized in Tables 3 and 4.

[Insert Table 3 about here]

[Insert Table 4 about here]

ASSET4

The Thomson Reuters' ASSET4 database provides a comprehensive set of *economic, environmental, social, and governance* (ESG) scores since 2002. It covers about 4000 companies all over the world and historical data is available for about 1000 companies. While

ASSET4 does not use sector screens, the ESG data is collected by trained research analysts from company reports as well as news sources, stock exchange filings, and non-governmental organizations. The collection of information includes over 750 individual data points, which are subsequently combined into over 250 key performance indicators, based on a default equal-weighted framework. In other words, after gathering the data points, analysts transform qualitative data into consistent units of quantitative data. The key performance indicators (KPIs) are then aggregated into a framework of 18 categories (see Table 3) grouped within 4 pillars (*Economic, Social, Environmental and Governance*) and then integrated into an overall single score. Weights play an important role in both the integration of the subdomains in the four pillars and the integration of the four pillars in the overall score. Typically, data for environmental performance includes information on the reduction of emissions, such as CO₂ and other gas emissions, biodiversity protection, waste recycled, pollution controversies; product innovations, environmental footprint reductions, environmental R&D expenditures, environmental labels; resource use reductions, and water recycled. Social performance encompasses data on employees such as turnover, accidents, injury rates, training hours, or women; community involvement such as cash and in-kind donations, lending, political contributions, or employee volunteer activities; product responsibility such as quality certifications, or social labels; and controversial products such as tobacco or alcohol. Corporate governance indicators refer to executive compensation, board experience, board diversity, anti-takeover devices and compensation controversies.

Importantly, ASSET4 differentiates between driver indicators and outcome indicators (see Table 4). While driver indicators are capturing the information about availability of policies and processes that companies have committed to, outcome indicators measure the *results* of the policies and company actions. Controversies are also considered as outcome indicators. Finally, ASSET4 provides standardized scores (z-scores) for indicators,

categories, pillars and the overall score, using all underlying data points and comparing them across all companies covered. The z-scores are essentially a measure comparing one company with a given benchmark and it expresses the value in units of standard deviation of that value from the mean value of all companies. For example, Cheng et al. (2013) use the annual z-scores for the pillars and constructed a composite CSR index by assigning equal weights to each of the three pillars.

MSCI

The MSCI CSR ratings database, formerly known as KLD, rates the *social, environmental* and *governance* performance of companies on more than 280 data points using a proprietary rating system that delivers 50 ESG indicators. The ratings are not normalized across industries. The database includes the 3000 largest US companies since 2003 while historical data up to 1991 is available for all firms in the Standard and Poor's 500, as well as for the MSCI KLD 400 Social Index. *Screens* are used for firms with military concerns, tobacco concerns, and nuclear power concerns (Chatterji et al., forthcoming). The KLD database has become the standard for quantitative measurement of companies' CSR initiatives. While extensively used in scholarly research to operationalize companies' CSR initiatives (e.g. Chatterji et al., 2009), most studies (see Table 2) created their own index based upon the raw data supplied by KLD through the MSCI ESG STATS that is available for academics. The creation of indexes occurs because MSCI does not provide an overall or total CSR performance score. The ESG indicators are grouped into seven areas: (1) environment, (2) community, (3) human rights, (4) employee relations, (5) diversity, (6) product, and (7) corporate governance (see Table 3). Strength and concern scores in each of the seven areas of CSR are based on an assessment made by KLD analysts after direct communications with company officers and visits to company facilities, as well as extensive investigation of public records, including more than 14,000 global media sources, company

websites, reports from governmental agencies and NGOs, financial reports filed with regulatory agencies, and company annual reports. The KLD database thus indicates the presence or absence of strengths and weaknesses in each of the seven areas of CSR (see Table 4). The descriptions of strengths and weaknesses seem to suggest that “strength” scoring is mainly based upon “strategic” indicators while “concerns” scoring is mainly based on “performance” indicators, as illustrated by the example descriptions below:

Strength – Waste Management (ENV-str-B)

This indicator evaluates companies that are at risk of incurring liabilities associated with pollution, contamination, and the emission of toxic and carcinogenic substances, and/or companies that produce or sell electronic products face risks associated with recycling and/or disposal of end-of-life electronic products. Companies with strong programs and track records of reducing emissions and waste, and/or that proactively address electronic waste concerns by establishing comprehensive and well-managed product recovery and recycling programs, score higher. Companies that create large volumes of toxic and carcinogenic emissions or waste, yet lack programs or policies to reduce or control these substances and have experienced recent incidents of contamination, or electronic product companies with a strictly compliance-driven approach, score lower (MSCI ESG STATS, User Guide & Ratings Definition, 2013, p. 6).

Concern – Operational Waste (ENV-con-I)

This indicator measures the severity of controversies related to the impact of a firm’s non-hazardous operational waste. Factors affecting this evaluation include, but are not limited to, a history of involvement in environmental impact-related legal cases, widespread or egregious impacts of the firm’s non-hazardous waste streams, resistance to improved practices, and criticism by NGOs and/or other third-party observers (MSCI ESG STATS, User Guide & Ratings Definition, 2013, p. 8).

Sustainalytics

Sustainalytics ratings are more recent. They cover more than 4000 companies worldwide since 2009. They evaluate companies on their *ESG performance* using targeted sector-specific ESG indicators. For each material aspect, Sustainalytics provides (i) measures of preparedness in which they assess the organization’s capability to deal with ESG aspects by looking at policies, managements systems, programs and targets; (ii) measures of disclosure in which they consider the company’s transparency on preparedness and performance via sustainability reporting and the implementation of key reporting and

verification standards; and (iii) quantitative performance measures in which they contemplate a company's performance based on quantitative social and environmental metrics, including both absolute and relative terms (see Table 4). A fourth element of assessment – qualitative performance measures – is used for the controversy assessment. Data is gathered using primary and secondary sources as well as third-party data providers. A set of core and sector specific metrics are analyzed, scored and weighted to determine a company's overall ESG performance. Each category, for instance supply chain monitoring, is granted a raw score (between 0 and 100). Both the raw score and the weighted scores of the different categories of each of the three domains (governance, environmental and social) are available to researchers. Their proprietary rating system includes between 60 and 100 indicators weighted according to the industry in which firms operate. Sustainalytics analysts use the following sources in their assessment—annual reports, financial reports, 10-K reports, CSR reports, SCR sections on websites, news releases, news sources such as Bloomberg, newspapers and NGOs, as well as company's feedback. For example, Surroca et al. (2010) measured CSR performance by constructing a weighted sum of the scores of a selection of five stakeholders groups (i.e., employees, customers, suppliers, the community and the environment) using the corresponding Sustainalytics weights averaged by sector and country. However, in Surroca et al. (2013), the global CSR score, which is the sum of all items averaged by their corresponding weight and rated on a scale from 0 to 100, is used to measure the same construct.

RESEARCH QUESTIONS AND METHODS

Based on the literature review and the analysis of methodologies and metrics of the three major CSR performance rating providers, we pose the following three research questions:

- 1) To what extent does the choice of the CSR performance rating provider influence research results and conclusions?*

- 2) *To what extent do CSR performance ratings rely on CSR disclosures?*
- 3) *To what extent does the choice of a CSR performance rating proxy originating within the same database influence research results and conclusions?*

To answer these research questions, we adopt a multi-method approach by performing statistical analyses and conducting interviews (Cooper and Morgan, 2008). This combined approach allows us to complement the statistical output with more nuanced explanations (Miles and Huberman, 1994). The details of the quantitative/archival approach are discussed first, followed by the details of the qualitative analysis.

Research Methods: Quantitative Analyses

In order to investigate RQ1, we test whether alternative CSR ratings proxies affect the significance of empirical results when used for research purposes and, in order to analyze RQ3, we examine the sensitivity of the research results to the construction of the CSR performance proxy. In RQ3 we focus particularly on the MSCI database since it is most commonly used in research (see Table 1).

We carry out these analyses using two different samples. First, we use the data collected by Cho et al. (2010) for their study of the persuasiveness of language used by U.S. companies in environmental disclosures in their 10-K annual reports. Their study included 190 disclosing companies, listed on the Standards and Poor's 500 Index. The main dependent variables are thus not "simple" disclosure scores, but scores that indicate disclosure characteristics: environmental disclosure optimism (EDO) and environmental disclosure certainty (EDC). Their findings suggest that the disclosures of worse environmental performers exhibit significantly more "optimism" and less "certainty" than the companies that have a better environmental performance. In the Cho et al. (2010) study, the corporate environmental performance score reflects the number of environmental Concerns as registered by KLD Research and Analytics, Inc. (now MSCI). Consequently, their

environmental performance Concern scores range from zero, representing no major Concerns, to a maximum Concern score of seven.

We replicate the Cho et al. (2010) study several times to examine (i) whether the choice of the CSR rating provider influences the results, i.e. whether the environmental proxies calculated using the ASSET4 database³ and the MSCI database lead to the same results (RQ 1) and (ii) if the calculation of the environmental performance proxy based upon the MSCI database influences the results (RQ3).

Second, we investigate whether there is a relationship between CSR disclosure and CSR performance using a sub-sample of Cho et al. (2015) which is covered by the three ratings providers. In this analysis, we rely on the total scores of the three CSR ratings providers: Total Concerns, Total Strengths and net score by MSCI, the average of the environmental, social and governance pillar scores for ASSET4 and the total weighted score for Sustainalytics. We then compare the sensitivity of these results to the construction of the CSR performance proxy using alternative measures from within the MSCI database.

Our dependent variable is the disclosure score collected by Cho et al. (2015) for their 2010 sample, which included 213 firms of the Fortune 500 industrials. The data were collected using the disclosure scale of CSR information items from the Ernst and Ernst (1978) survey (see Appendix 1 for further details on this measure). We identified the controls based on literature on the determinants of CSR disclosure (Cho and Patten, 2007; Clarkson et al. 2008; Cho et al. 2015). We include size (SizeRank), measured as the relative rank of the company within the sample for its year based on total revenues (Cho et al. 2015). Companies belonging to the extractive, paper, chemicals, petroleum, and metals industries are classified as members of an environmental sensitive industry (ESI) by using a one/zero indicator

³ We can assess the sensitivity of the findings of the Cho et al. 2010 study only using ASSET4, as the study relies on disclosure data for the fiscal year 2002, not yet covered by Sustainalytics

variable. We further control for profitability (ROA), measured as EBIT over total assets (Al-Tuwaijiri et al. 2004).

RQ 2 examines the extent to which CSR performance ratings rely on self-disclosed CSR information. We conduct this analysis by investigating the relation between CSR disclosure and CSR performance metrics provided by respectively MSCI, ASSET4 and Sustainalytics. Our independent variables are built on a careful understanding of the methodologies for the three CSR ratings providers reported in the previous section. We begin by separately identifying the disclosure-related and the performance-related measures for each CSR rating provider. Based upon the description of the databases, we consider Drivers and Outcomes for ASSET4, Strengths and Concerns for MSCI, and the Preparedness, Disclosure and Performance measures for Sustainalytics (see Table 4). With the exception of MSCI, which provides the total Strengths and Concerns for each company, the other rating providers do not provide an overall score for Drivers vs. Outcomes (ASSET4) or Preparedness, Disclosure and Performance (Sustainalytics⁴), therefore we build these scores on our own. For ASSET4, within each pillar (environmental, social and governance), we determined the average z-scores of all the indicators for Drivers vs. Outcomes⁵. For Sustainalytics, within each category (environmental, social and governance) we determine the average of the raw scores of all the indicators for Preparedness, Disclosure and Performance. In our analysis we then employ a single average score across the pillars (ASSET4) or categories (Sustainalytics)⁶.

Merging observations in the Cho et al. (2015) sample that are rated by MSCI, ASSET4 and Sustainalytics leads to 173 observations. We further lose eight observations due

⁴ Because Sustainalytics provides intra-year scores, we first determine an average score for the year for each raw indicator.

⁵ Because ASSET4 value or data indicators have different measurement units, we have no option but to use the z-scores.

⁶ Our main inference does not change if we consider the average of all Driver vs. Outcome indicators – ignoring the breakdown by pillar for ASSET4 or the average of all Preparedness, Disclosure and Performance indicators – ignoring the breakdown by category for Sustainalytics.

to missing performance data in the Sustainalytics database. Thus our final sample for the investigation of RQ1 includes 165 observations. We run the following regression models:

$$(1) \text{ Disclosure} = \alpha_1 + \alpha_2 \text{ MSCI strenghts} + \alpha_3 \text{ MSCI concerns} + \text{Controls} + \varepsilon_1$$

$$(2) \text{ Disclosure} = \beta_1 + \beta_2 \text{ ASSET4 drivers} + \beta_3 \text{ MSCI ASSET4 outcomes} + \text{Controls} + \varepsilon_2$$

$$(3) \text{ Disclosure} =$$

$$\gamma_1 + \gamma_2 \text{ Sustainalytics Disclosure} + \gamma_3 \text{ Sustainalytics Performance} \\ + \gamma_4 \text{ Sustainalytics Preparedness} + \text{Controls} + \varepsilon_1$$

Because of their reliance upon self-reported disclosure, we expect α_2 , β_2 , γ_2 and γ_4 to be positively associated with the disclosure score, whereas we do not expect α_3 , β_3 and γ_3 to be associated with the disclosure score.

Research Methods: Qualitative Insights

In order to provide some potential explanations for our quantitative findings, semi-structured interviews were conducted with current analysts from ASSET4 and Sustainalytics, one former analyst from MSCI, and one former financial analyst who dealt with all three rating providers⁷. These interviews were performed to gain relevant insights into the theorization of the different rating providers (see Chatterji et al., forthcoming), and help us better understand any statistical differences found in our quantitative analyses. The interview guide is available in Appendix 2. The interview guide was structured around questions related to differences in theorization and commensurability. Semi-structured interviews were used to gain information and facts, but also to obtain personal insights and subjective views (Czarniawska, 2004). Interviews were conducted on the respondents' company premises (with two exceptions) and had an average duration of 45 minutes. Interviews were carried out between December 2014 and March 2015. The current transcripts have been read and coded by one researcher. However, during the main study, three researchers will code each

⁷ We are currently conducting additional interviews.

transcript using NVivo. The codes will be derived initially from the interview protocol and prior literature review (Yin, 2003).

RESULTS

RQ1: Influence of rating provider (MSCI vs. ASSET4)

To analyze the potential influence of using a particular ratings provider when performing large-sample CSR studies, we selected the dataset of Cho et al. (2010), which uses environmental disclosure optimism and environmental disclosure certainty as its dependent variables and MSCI environmental Concern scores as a proxy for environmental performance. We use a sub-sample of Cho et al. (2010) as we consider only the companies that are rated by both MSCI and ASSET4 ($n = 129$). The correlation table (Table 5, Panel A) contains a striking result. Indeed, the net environmental performance score calculated by subtracting total Concerns from total Strengths in the MSCI database is significantly negatively correlated to the total environmental performance score as supplied by ASSET4. This finding can be explained by the fact that both the environmental Concerns and the environmental Strengths scores arising from MSCI are significantly positively related to the final environmental performance score as supplied by ASSET4. Furthermore, the table shows that the environmental Concerns and environmental Strengths scores as supplied by MSCI are not correlated. Finally, the correlation shows that MSCI environmental Concerns (Strengths) are negatively (positively) correlated to the Net environmental score. This finding is expected since it originates from the calculation of the net environmental performance score.

[Insert Table 5 about here]

Columns (1) and (2) of Panel B (in Table 5) replicate Cho et al. (2010) using their original sample. MSCI Environmental Concern scores positively influence the environmental disclosure optimism score ($p = 0.058$) while they negatively influence the environmental

disclosure certainty score ($p = 0.042$). Consequently, the authors conclude that the disclosures of worse environmental performers exhibit significantly more “optimism” and less “certainty” than those of companies that have a better environmental performance.

Columns (3) and (4) re-run the same regression model as in (1) and (2) respectively, but only for the sub-sample of observations included both in MSCI and ASSET4, to ensure comparability. Columns (5) and (6) report the results for the same regression models if we use the ASSET4 Environmental Score rather than the MSCI environmental Concerns. Although MSCI environmental Concern scores are not associated with optimism for the sub-sample of 129 companies covered by both databases, it still positively and significantly associated with the environmental disclosure certainty score. Evidence also indicates that the ASSET4 environmental score is not significantly related to either the environmental disclosure optimism score or the environmental disclosure certainty score. Thus, caution is needed when comparing the results of studies that use different environmental or CSR performance indicators based on different CSR rating providers.

Our interview findings are helpful in interpreting our quantitative results. The fact that social rating providers use different themes and subthemes when measuring CSR (or environmental performance) suggests that there might be different theorizations of CSR (or environmental performance) (see Chatterji et al., forthcoming). These differences in theorization are confirmed by our respondents. For example, respondent B1 stated: *“Yes I would say company by company and country and country, the theory of corporate social responsibility that drives measurement is very different.”*

Most rating providers are vague when communicating their definition of CSR, appearing hesitant to make their CSR theorization too explicit:

“They're very careful, very clever. They won't give their definition of corporate social responsibility. ... So they use a pandering, kind of a fuzzy definition of what they believe corporate social responsibility could be for you as an asset manager or an asset owner. Because they don't want, they want to work with everybody.” (D1)

Nevertheless, these rather implicit differences in theorization might generate differences in measurement: *“These [theorizations] ... define the share of the positive and negative. Define the focus on absolute or relative types of analysis.... And the profiles of the analysts are different, and the profiles of the founders are different.”* (B1).

Interestingly, there has been a shift in theorizations of CSR in the field. In particular, the theorization evolved into the direction of a fiduciary approach, in which the possible financial impact of CSR is central:

“This new theory of corporate social responsibility, it got diffused across the sector for all companies that also decided to sell to the mainstream. That means in Jantzi in Canada, ASSET4 later on, Inovent, and KLD around 2001 they started shifting to this or at least combining their original theory of corporate social responsibility with this new theory of corporate social responsibility which is, it's about financial performance.” (B1)

According to respondent B1, the United Nations Principles for Responsible Investment [UNPRI] has played an important role in the diffusion of this new theorization of CSR. This shift in theorization of CSR might, however, bias longitudinal studies that measure the impact of some proxy of CSR performance. For example, one respondent disclosed that a change in the measurement protocols occurred following this shift in theorization. In particular, this social rating agency wants to adjust its measurement system in a way that provides the asset managers more information concerning the possible financial implications of CSR: *“...the alpha component is very important. So that's why it's very important to have an offer and to well describe more the opportunity side of the ESG.”* (C1). The respondent relates this evolution in their methodology directly to the change in theorization, i.e. a change in *“the point of view of society regarding CSR”* (C1):

“So corporate social responsibility started as something negative, say you now start at the beginning, just to exclude the most controversial companies or the companies involved in the sin activities, so not ethical activities. Then it went into a risk driven methodology so there were a lot of indicators that could be like I don't know, policies for bribery and corruption, and now we are thinking about we are let's say it's the third stage which is opportunities. So take into account how corporate social

responsibility can be a source of improvement, of opportunities to get more market exposure, to get more market share, to reduce the risks of course but at the same time to attract more opportunities for the company.”

MSCI, and in particular KLD, which is considered one of the oldest social rating providers, originally used a theorization of CSR that can be labelled as “*social contract responsibility*” (B1). This label means that KLD wanted to offer a “*tool of accountability*” (B1). Hence, they were more focused on the negative aspects of a company’s CSR performance, i.e. the concerns. According to respondent B1, this approach also implies that that they did not care as much about the ability to compare performance across companies, but cared more about “*holding them to account compared to a set of absolutes and best practice sort of social contract.*”(B1). This approach might explain why KLD did not normalize their ratings by industry.

The new theorization of CSR has not been successfully implemented in every rating agency:

“So [social rating agency 1] did a change in its course but still the tools are very much based on the theory of corporate social responsibility that's values driven. For some such as [social rating agency 2] there's been a gradual change of theory, change of tools, change of analyst types. For companies such as [social rating agency 1] there hasn't been much transition.” (B1)

“[Social rating agency] is much more culturally anchored. They had this vision and ideology and so there are some social rating agencies which have always been influenced by some ideologies, ethically somehow. So that would remain.” (D1)

Taken together, these findings seem to indicate that although at the discourse level all of the CSR ratings agencies seem to have evolved towards this new differentiation of CSR, different rating agencies might continue to have different “silent” or “implicit” theorizations of CSR. These differences in theorization might impact the information they collect, how they score this information and how they assign weights to the different scores, i.e. how they combine the indicators into sub-scores and global scores (see *infra*). Thus, comparing data from different social rating agencies ‘*doesn't make sense*’ (B1) if social rating providers utilize

different underlying theorization of CSR. Consequently, the interview evidence suggests that we should not compare studies that use CSR proxies originating from different databases, since different CSR rating providers have different theorizations of CSR. This holds true because, while in one rating provider the accountability view might still be dominant, the financial or the opportunity view might have become dominant for another rating provider. Moreover, the findings highlight that it might be difficult to use a CSR proxy, for example the overall CSR score in a longitudinal study, because the theorizations of CSR rating providers and the methods they use to calculate overall scores varies over time. However, when looking into the databases, the scores the companies get assigned by the social rating agencies look rather stable over time. Two respondents explained this by highlighting that they ‘force’ stability into the data, as ‘*an intentional attempt to keep consistency*’ (B1):

“What does it mean when the methodology is turned around? So okay, let's look at how the ranking looks. Well it's a mess. Everything is different. We cannot show this to the client. So let's go look at our weights, look at the stuff, because we want, we can have some changes that we can justify. We cannot have everything changed. Because for ranking, one of the values is continuity. Stability. If you're not stable, you don't have values for comparison between year and year, you lose your value, meaning that even though in the back end, everything changes, in the front end you have to keep some stability and that's, as a result when you see (unintelligible) this latest, it looks so good, even though [X] has changed so much, this firm has had overall some stable movements in the ratings. That's not because of reflection of stability of the platform. It's because they have to project stability. So even though things change behind the scene, they manipulate it to make sure the outcome.” (B1)

“There will be a new company profile, it will be issued in 2015 so in a few months, and this will give more space to this opportunity side. But it's the first, don't say that it's done. We want to go, but at the same time we can't let's say evolve our methodology too fast compared to the market because we have 350 clients, and so we have to take each account that we have to have something progressive in terms of innovation. It shouldn't be a quantum leap.” (C1)

This leads us to question the extent to which CSR ratings measure CSR performance.

Since Sustainability metrics are not available for the Cho et al. (2010) sample, we investigate whether there is a relationship between CSR disclosure and CSR performance using a sub-sample of Cho et al. (2015), which has data available from the three ratings

providers. Models (1), (2) and (3) in Panel B, Table 6 report evidence of the relationship between CSR disclosure (as measured in Cho et al. 2015) and CSR performance (RQ2), relying on the total scores of ASSETS 4 and Sustainalytics, i.e. the average of the environmental, social and governance pillar scores for ASSET4 (Column 2, Panel B) and the total weighted score for Sustainalytics (Column 3, Panel B). For MSCI a total score was constructed by subtracting the total Concerns from the total Strengths. Hence, models (1), (2) and (3) use an alternative measure of CSR performance, originating from the three CSR ratings providers. Panel A, Table 6 illustrates that all overall scores are significantly positively correlated. Panel B, Table 6 shows that all overall scores are positively related to the disclosure variable (see Table 6). The nature of the dependent variable, i.e. a disclosure variable, might explain why we now find positive associations with the total scores of all three CSR rating providers.

Our interviews allow us to further unpack this result from our archival analysis. An important point that determines the commensurability is the sources rating providers use to gather their data. While one rating provider screened only the publicly available information, the other two providers interacted with the companies:

“We don’t include private information, so we don’t engage with the company directly. And if they provide us with data that is not available publicly, we will not include it in our assessment.” (A1)

Although the other two rating providers directly interviewed the companies they were analyzing or asked for their feedback, their ratings were still highly driven by company-provided disclosures: *“I can say about 60% of the data comes from the company that we cannot verify really” (B1)*. This suggests that the level of disclosure impacts the performance score that companies get assigned. The interview findings corroborate that using proxies from CSR rating providers to measure the impact of ‘CSR performance’ on ‘CSR disclosure’ might lead to spurious findings.

[Insert Table 6 about here]

RQ2: CSR performance ratings and reliance on CSR disclosures

To examine whether the disclosure-driven indicators are driving the positive associations found between the CSR performance indicators and CSR disclosure in Table 6, we adjusted the analysis by splitting up the total CSR performance score into disclosure-related and performance-related measures for each CSR rating provider (see Table 4). Table 7, Panel A illustrates the descriptive statistics (mean and standard deviation) and correlation analysis for the sample of 165 companies that are included in the Cho et al. (2015) paper and are also covered – with no missing information – in MSCI, ASSET4 and Sustainalytics. The various sub-ratings are all significantly correlated with each other, with the exception of MSCI Concerns and Sustainalytics Performance. It is also worth pointing out that the correlation coefficients between MSCI Concerns and the other ratings seem to be relatively smaller. Table 7, Panel B reports the regression results for equation (1), (2) and (3). With respect to MSCI ratings, as expected, we find that MSCI Strengths are significantly and positively associated with the disclosure score, while the coefficient of the MSCI Concerns score is not statistically different from zero. With respect to model (2), we find that both ASSET4 Drivers and Outcome are positively associated with the disclosure score, so our expectations that ASSET4 Outcome indicators rely less on self-reported disclosure were not met. Finally, with respect to Sustainalytics ratings, results of our analysis are reported in column (3). As expected, the coefficient for the Performance sub-ratings is not associated with the disclosure score, indicating that the ratings depicting performance tend to be less reliant on self-reported disclosure.

[Insert Table 7 about here]

Overall, this analysis reveals that disclosure-driven indicators are highly correlated with the disclosure data. Hence, overall scores, which are largely driven by disclosure data, might be questionable independent variables in CSR disclosure studies. With respect to the more ‘pure’ performance indicators, it is striking that the sign of the Sustainalytics and Asset4 variables are the same as the MSCI Concerns. Indeed, although the coefficient of the Sustainalytics performance variable is not significant, we would have expected opposite signs, since a higher performance indicator in the Sustainalytics performance suggests better performance, while a higher MSCI Concerns score reflects worse performance.

According to the interviews, it is important to note that even more ‘pure’ performance data still rely on the disclosures of the companies that are being rated. It follows, therefore, that a non-disclosing company is more likely to have a low performance score, which might explain the positive although non-significant sign of the coefficient. Indeed, respondent A1 confirmed that if the company is not disclosing the CSR information they are looking for, the company will obtain a low score for transparency and will not get a performance score assigned. As a result, since the extent of disclosure seems to drive the scoring by social rating agencies, companies start to disclose more information:

“Many times though, companies call us and ask us to look at their data, look at their own information to understand how they can improve their public disclosure.” (A1)

“I think in terms of communications it has had some implications in terms of how they report. I think it's structured a little in a certain way, it's made the companies in the image of the methodology—their disclosure...because they want to have a better score and as a result they want to make sure they give the right data.” (B1)

Although companies’ self-reported information is extensively used by all rating providers, companies do obtain different scores from different rating providers due to (i) differences in theorization and (ii) also due to the fact that different analysts are rating the companies. Indeed, the interviews suggest that the rating system is highly influenced by the quality and the background of the analysts:

“And then the analysts varied a lot. So there were analysts who were activists before, who worked at the United Nations before. Two analysts who were more MBA style, finance style that were later. So the more the company tried to target the mainstream in the 2000s, the more they hired this type of analyst. And their views about what's the firm's responsibility and what we are measuring was really different, depending on who the analyst was.” (B1)

“So we are all measuring the same things but who is measuring? So if you take a company where the average experience is fifteen months, you will have a different result. If you take a company where the average experience is four years, five years, you will have another result.” (C1)

Although social rating agencies write scoring guidelines for the analysis, to “avoid the point of view” (C1), analysts have still some freedom when rating:

“How do we convert this data to a number? From 0 to 10. And that's very much up to, we receive the scoring guidelines but a lot of it is analyst judgment. Should I give this 6? Or 7 or 8? Or 5? Scoring guidelines left a lot to use with regards to what is best practice, what is worst practice, how does this fall in the middle? So the first weakness comes when for any, the most reliable type of indicators, how do I convert this piece of data that the company has the same CEO and chairman for the past seven years. How do you convert it to a number? And I think, I'm hopeful that companies such as [X], they've gotten better at it.” (B1)

“At X there was a pressure for the financial CSR. [...]It did affect the way we wrote the profiles. It did affect the way we talked to the companies. We would talk always about financial materiality. Is there any residual effect of our personal beliefs? Definitely. Because they will come out on the hidden part of the work. Where is the hidden part of the work? It's in excel files, nobody sees. It's in the way we look at the detailed indicators. [...] And at X we were under a lot of pressure for framing and shifting towards financial materiality. Frame but again as I told you, I have so many examples of analysts' ideology still playing an important role.” (B1).

The described observations highlight that the ratings are, to a significant extent, the outcome of a subjective process. This observation might explain the commensurability differences across ratings providers, since subjectivity means that even two analysts in the same rating agency might arrive at different conclusions regarding a company's CSR scores:

“And the analysts had a lot of flexibility. [...]So as a result of this delegation to the analyst level, that means the theories of corporate social responsibility in the minds of the analysts, had major implications for how they measure it.” (B1)

An alternative explanation for this finding might be that analysts of different CSR rating agencies “crosscheck” or compare their scores with those of other ratings providers on the overall level to avoid that they are issuing a completely different score:

“So there is a positive to this difference. You can say our data is better, that's why it's different. And then there is a negative side because if the other provider you consider it a good provider, then, or if your ratings are too different, or even as an analyst you have little self-confidence because you just came out of an MBA program, then you will say you know what? That means my ranking is not good. I'm going to go check my data and ...And I've seen it happen. Around me. So what does this mean? Even if you find similarities, it's not because the underlying reality has been looked at in the same way. It's because they have been adjusting to each other.” (B1)

RQ3: Influence of the choice of the proxy originating from the MSCI database

The findings thus far led us to further investigate the extent to which the choice of a CSR performance rating proxy originating from the same database influences research results. Model (4), (5) and (6) in Table 6 report evidence on the sensitivity of the results on the association between CSR disclosure and CSR performance to the construction of the CSR performance proxy, using different measures from the MSCI database. We replicate column (1) using alternatively: MSCI Strengths and Concerns, MSCI Strengths, or MSCI Concerns. In line with evidence from Table 7, we find that MSCI Concerns are not associated with the disclosure score, while MSCI Strengths appears to be consistently and positively associated with CSR disclosure. Comparing Tables 5 & 7 further suggest that using the Sustainalytics Total score would lead the author to draw different conclusions than when using only the Sustainalytics Performance Score.

We further explore this research question, by relying once again on the replication of the Cho et al. (2010) study. Table 7 Panel C replicates the study by considering alternatively the CSR proxies originating from the MSCI database⁸. For convenience, once again Columns (1) and (2) report the replication of the original Cho et al. (2010) results. Columns (3) and (4) indicate that, in contrast to environmental Concerns scores, environmental Strength scores are

⁸ In this analysis we consider the original sample in Cho et al. (2010), with n. of observations equal to 190

not significantly related to either their disclosure optimism or certainty scores. Consequently, using this proxy for environmental performance would have led to the conclusion that there is no relation between environmental performance and disclosure optimism or certainty scores. If the two former proxies would have been included in the same regression, Columns (5) and (6), evidence would be aligned with the original Cho et al. (2010) paper. Finally, we measure environmental performance using the MSCI Net Environmental Score. This last proxy was calculated by subtracting the environmental Concern scores from those exhibiting environmental Strengths. We note a change in the sign of coefficients in both model (7) and (8), although the relation between MSCI environmental net score and optimism is not significant. The use of the net environmental performance score would have indicated that while this score is positively associated with the certainty score ($p = 0.016$), it is not significantly associated with the optimism score. Consequently, evidence suggests that the use of different proxies originating within the KLD database for environmental performance leads to different results and different research conclusions.

While Concern scores reflect past performance, the Strength scores mainly reflect the amount of disclosed information by companies. Thus, it might be that companies that disclose more information for legitimacy reasons have both high Concern and Strength scores. This indicates that researchers must carefully justify why they use a certain proxy of CSR performance of a specific rating provider in their study. For example, Cho et al. (2010) selected Concern scores as their proxy for CSR performance because they grounded their empirical work in Legitimacy theory and impression management.

The interview findings and Table 4 help us to understand why this finding is not surprising. Indeed, the different components analyzed by the different rating providers attempt to measure different concepts. While the MSCI environmental concerns, as a “*tool of accountability*” (B1) highlight the negative issues, the strengths focus on strategic indicators,

which involve opportunities. Table 7 also indicates that if we do not use the overall score provided by Sustainalytics and ASSET4 in our regression, but rather the individual components of their scores together, we can draw different conclusions. This finding suggests that before using an overall score it is important to reflect on how this score was developed. In particular, respondent B1 argues that *“These [theorizations] define the weights of their indicators”*. Respondent C1 confirms that the theorization of CSR might not only influence the “measures”, but also the “weightings” they are using:

“Well it's both because we have a matrix and in this matrix we have the horizontal axis which takes into account all so let's say mainstream impacts, can be market share, financial, business impact let's say. And on the vertical axis we take into account the stakeholder impacts, social, environment. [...] So in this way, in the weighting system we take into account, and the way we weight indicators, we take into account financial aspects.” (C1)

Interviewee B1 even labelled the assignment of weights as the *“next level of shakiness”* (B1), since it is especially analyst-sensitive or analyst-driven:

“So if I get a sector to analyze, I get food and beverage sector. It's my, I look at the Excel file and in the Excel file I see that the overall weight for environmental issues for food and beverages is 60%. The overall weight for social issues in 40%. Well this is significant because if I adjust this a little, the weight of all indicators, all 100 something indicators will change. And it was at our disposal. I would discuss it with the sector head, saying look, I think last time that they didn't judge is correctly. I think social should have a bigger weight in this sector because of the effect that they have on water, because access to water in poor countries and because of the issue of obesity in the food and beverage sector, I'd convince her and we'd make it 50% 50%. Was it a negotiation? Of course.” (B1).

The above paragraphs highlight why different social rating providers might assign the same companies different scores. The theorization of CSR not only impacts the dimensions of CSR they study, but also the information they gather, their way of quantifying the information, the distribution of the weights, the type of analysts they hire. This might explain why analyst B1 revealed that *“inside the company, we laughed about it [the fact that academia bought it and used it for research]”*. Moreover, the way in which the same concepts are measured might be different, even within the same company, depending upon

the analyst doing the assessment. Respondent C1 believed that the main reason for the differences in scoring between the social rating providers is mainly driven by the expertise and qualification of the analysts, and to a smaller extent by the fact that they apply different weighting systems. The incommensurability of the ratings seems to originate from the complexity of the most basic task in a rating agency, i.e. turning information into quantitative scores:

“I mean if I had to justify the gap in terms of (unintelligible) I would say that 75% of this gap is explained by the experience of the analysts and the 25% you can put the fact that the rating agencies may have access to many other sources of information, the fact that the rating agency will have more feedback for the (unintelligible) global, you will have better feedback from companies if you are local.” (C1)

DISCUSSION AND CONCLUSIONS

CSR researchers in accounting, finance, and management have utilized a variety of theories, research methods, and data sources to investigate relationships among corporate environmental performance, financial performance, social performance, and their associated disclosures. As the investment community’s interest in corporate social and environmental performance has grown, alternative sources for obtaining independent measures of CSR performance also have grown. Thus, researchers who undertake large-sample empirical studies of CSR performance must make important decisions regarding the CSR constructs and proxies to use in their statistical analyses. Currently, there are three major providers of independent CSR performance ratings (ASSET4, MSCI, Sustainalytics). The overall purpose of our study was to review the CSR constructs and proxies employed in studies published in a select set of influential journals and working paper series, and to provide an example of how the results of CSR studies may be influenced by the selection of CSR proxies. We used data from Cho et al. (2010) and Cho et al. (2015) to illustrate these concerns.

Our review of CSR papers found that the MSCI (formally KLD) data have been the source of independent CSR ratings for almost all of the studies we examined. Sustainalytics

and ASSET4 data sources, however, are being employed in some of the most recently completed studies that we reviewed. As detailed in Tables 1 and 2, previous research has used MSCI ratings to calculate a number of different CSR performance measures, the most important ones being a net strength-concern score and a focus only on strength or concern scores. The justifications for adopting a specific calculation of CSR performance also varied. We do not intend to imply that the justifications provided are weak, but that they are sometimes idiosyncratic to the particular research question being investigated. For example, studies concerned with corporations' treatment of employees would include only the employee relations and diversity dimensions of the MSCI measures. Other studies would calculate industry-adjusted CSR performance measures because comparisons of performance within specific industries were the focus of the analysis.

Our analysis highlights that using CSR performance proxies originating from different CSR rating providers might lead authors to draw different conclusions. Indeed, the main question is whether these CSR rating providers measure CSR performance in the same way. Drawing on the interview findings, our study indicates that (i) different CSR rating providers have different theorizations of CSR and might therefore 'measure' CSR performance differently and (ii) that the measurement process is highly subjective, which implies that even two analysts from the same rating agencies might score the CSR performance of a company differently. Thus, we suggest that caution is required when developing a specific CSR proxy based on prior research and also is needed when interpreting findings based on the data supplied by one CSR ratings provider.

The CSR performance calculations using MSCI data that were used by researchers when examining some type of "overall" CSR performance (e.g., social performance or environmental performance), also varied across studies. For example, these studies may have calculated an overall net score (strengths minus concerns), included total strengths and total

concerns separately in their regressions, or collapsed the ratings into dummy variables. It should be noted that some of these studies performed robustness checks or sensitivity analyses. When a study focuses on overall performance and MSCI data are used, it seems imperative that alternative specifications of the overall performance proxy are considered unless there are clear theoretical links between the construct being investigated and the specific empirical proxy utilized. Although some rating agencies like ASSET4 or Sustainalytics provide a total or overall score, researchers should be aware that their different theorizations of CSR might imply a significantly different weighting scheme. Again, robustness checks or sensitivity analyses are imperative.

As more researchers use ASSET4 or Sustainalytics data, there is a heightened concern over the appropriateness of building an empirical model that uses CSR performance constructs that were found significant in prior research when using data from a different CSR performance rating provider. For example, the MSCI data are generally used to calculate nominal scores. Each strength measure and each concern measure usually are assumed to be of equal importance and equally applicable to corporations regardless of industry. ASSET4 data, on the other hand, standardizes its overall “pillar” scores by industry. Thus, for a specific corporation, its MSCI environmental “net” score may represent a very different view of its environmental performance than would its ASSET4 environmental overall “pillar” score. We see evidence of this potential problem when performing alternative specifications of CSR performance in our example analysis using the Cho et al. (2010) data. Cho et al. (2015) focus on environmental concern scores because their analysis is grounded in the legitimacy/impression management literature. Thus, the justification for using only concern scores is derived from theory. Had Cho et al. used net scores, either from MSCI or ASSET4 data, or included both total MSCI strength and concern scores, we show that their results would have been different in some respects. In conclusion, further research is needed to

address the construct validity of the corporate CSR performance ratings provided by MSCI, ASSET4, and Sustainalytics, and to help determine whether or not these providers are measuring CSR constructs that can be used interchangeably to advance large-sample, empirical work in CSR performance and reporting.

APPENDIX 1

Disclosure scale of CSR information items from the Ernst & Ernst (1978) survey used in Cho et al. (2015)

Environment

1. Pollution control
2. Prevention or repair of damage to the environment
3. Conservation of natural resources
4. Other environmental disclosures

Energy

5. Conservation of energy
6. Energy efficiency of products
7. Other energy-related disclosures

Fair Business Practices

8. Employment of minorities
9. Advancement of minorities
10. Employment of women
11. Advancement of women
12. Employment of other special-interest groups
13. Support for minority businesses in the U.S.
14. Socially responsible business practices abroad
15. Other statements on fair business practices

Human Resources

16. Employee health and safety
17. Employee training
18. Other human resource disclosures

Community Involvement

19. Community activities
20. Health-related activities
21. Education and the arts
22. Other community activity disclosures

Products

23. Safety
24. Reducing pollution arising from use of product
25. Other product-related disclosures

APPENDIX 2

1. There are different definitions of CSR. Can you describe what you exactly try to measure, i.e. the definition of CSR you have in mind.
 - Can you tell something about the history of this rating agency. Did they always have the exact same definition of CSR as you just described? If not, why did it evolve over time? What was the impact of this on your methodology?
2. How are ratings produced? How do analysts build their assessment?
 - What are the main drivers of the scoring assessment? (financial risk vs. normative approach)
 - Do you only measure social and environmental performance or do you also try to assess possible financial impacts of social and environmental performance?
 - Do you agree with the following quote of a former Analyst of Vigeo?
What is the impact of this on the methodology of your rating agency:
“You’re not here to save the planet, as they say. You work for investors. So you really do need a vision of the economic impacts on firms and how they can impact their business.”
 - How is data collected? Interaction with firms? How often do you assign scores? To which extent do you rely on the disclosures of companies themselves?
 - How are weights assigned, if any?
 - Outcome/output indicators vs. processes/strategy?
 - How analysts build their assessment is a rather subjective process, how do you deal with that?
3. Have there been any changes over the last several years in terms of methodology, day-to-day practices and so on?
 - Who decides about a change in the methodology?
 - Can you explain the most important changes in terms of day-to-day practices and why they took place?
 - Can you see any externalities of these ratings (on investors or firm behavior?)
 - Is there a driver that has become predominant?
 - What has been the cultural/ideological influence?
4. Who is their (main) targeted audience? Who are the ratings mainly thought for? Who are your users/clients?
 - What investment strategy do they serve?
 - What do you think about academia relying on these proxies to measure CSR performance?
5. Do you think the ratings are comparable across providers? To which extent? Can you explain why?
 - Would you say that you all have the same definition of CSR? If not, could you highlight some important differences?
 - Are you aware that different providers have different methodologies and what these are? What do you know and what is your opinion about them?
 - Can the differences between the providers mainly be explained by differences

in the definitions of CSR or by the fact that it is very difficult to measure CSR performance?

- How do you try to distinguish yourself from the other providers, why should they buy your product?

REFERENCES

- Agle, B. R., R. K. Mitchell and J. A. 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.
- Bae, K.-H., J.-K. Kang and J. Wang. 2011. Employee treatment and firm leverage: a test of stakeholder theory of capital structure. *Journal of Financial Economics* 100 (1), 130-153.
- Barnett, M. L. and R. M. Solomon. 2012. Does it pay to be really good? Addressing the shape of the relationship between social and financial performance. *Strategic Management Journal*, 33 (11) 1304-1320.
- Bendell, J. 2010. ESG analysis in deep water. *Journal of Corporate Citizenship*, 40, 6-25.
- Berman, S. C., A. C. Wicks, S. Kotha and T. M. Jones. 1999. Does stakeholder orientation matter? The relationship between stakeholder management models and firm financial performance. *Academy of Management Journal*, 42 (5), 488 – 506.
- Briscoe, F. and S. Safford. 2008. The Nixon-in-China Effect: Activism, Imitation, and the Institutionalization of Contentious Practices. *Administrative Science Quarterly* 53, 460-491.
- Chatterji, A. K., Durand, R., Levine, D. and S. Touboul. Forthcoming. Do Ratings of Firms Converge? Implications for Strategy Research, *Strategic Management Journal*.
- Chatterji, A. K., Levine, A. K., & Toffel, M. W. (2009). How Well do Social Ratings Actually Measure Corporate Social Responsibility? *Journal of Economics & Management Strategy*, 18 (1), 125-169.
- Chatterji, A. K. and M. W. Toffel. 2010. How firms respond to being rated. *Strategic Management Journal* 31 (9) 917-945.
- Cheng, B., I. Ioannou and G. Serafeim (2013). Corporate social responsibility and access to finance. *Strategic Management Journal*, 35 (1), 1-23.
- Chin, M. K., D. C. Hambrick and L. K. Treviño. 2013. Political Ideologies of CEOs: The Influence of Executives' Values on Corporate Social Responsibility, *Administrative Science Quarterly*, 58 (2)197–232.
- Cho, C. H., Michelon, G., Patten, D. M. and R. W. Roberts. 2015. CSR disclosure: the more things change ... ? *Accounting, Auditing and Accountability Journal* 28 (1): 14-35.
- Cho, C. H., and D. M. Patten. 2007. The role of environmental disclosures as tools of legitimacy: A research note. *Accounting, Organizations and Society* 32(7–8): 639–647.
- Cho, C.H., R. W. Roberts, and D. M. Patten. 2010. The language of U.S. corporate environmental disclosure. *Accounting, Organizations and Society* 35 (4): 431-443.
- Choi, J. and H. Wang. 2009. Stakeholder relations and the persistence of corporate financial performance. *Strategic Management Journal* 30 (8) 895-907.
- Cohen, J.R., L. Holder-Webb, L. Nath, and D. Wood. 2012. Corporate reporting of nonfinancial leading indicators of economic performance and sustainability. *Accounting Horizons*, 26(1), 65-90.
- Coombs, J. E. and K. M. Gilley. 2005. Stakeholder management as a predictor of CEO compensation: main effects and interactions with financial performance. *Strategic Management Journal*, 26 (9), 827-840.
- Cooper, D and W. Morgan. 2008. Case study research in accounting. *Accounting Horizons*, 22 (2), 159-178.
- Deng, X., J.-K. Kang and B. S. Low. 2013. Corporate social responsibility and stakeholder value maximization: evidence from mergers. *Journal of Financial Economics* 110 (1), 87-109.

- Dhaliwal, D. S., O. Z. Li, A. Tsang, and Y. G. Yang. 2011. Voluntary nonfinancial disclosure and the cost of equity capital: The initiation of corporate social responsibility reporting. *The Accounting Review* 86 (1): 59-100.
- Flammer, C. 2013. Corporate social responsibility and shareholder reaction: The environmental awareness of investors. *Academy of Management Journal*, 56 (3), 758-781.
- Gao, F., L. L. Lisic and I. Zhang. 2014. Commitment to social good and insider trading. *Journal of Accounting and Economics* 57 (2-3) 149-175.
- Godfrey, P. C., C. B. Merrill and J. M. Hansen. 2009. The relationship between corporate social responsibility and shareholder value: an empirical test of the risk management hypothesis. *Strategic Management Journal* 30 (4) 425-445.
- Graves, S. B. and S. A. Waddock. 1994. Institutional owners and corporate social performance. *Academy of Management Journal* 37 (4), 1034-1046.
- Hillman, A. J. and Keim, G. D. 2001. Shareholder value, stakeholder management, and social issues: What's the bottom line? *Strategic Management Journal* 22 (2), 125-139.
- Hillman, A. J., C. Shropshire and A. A. Cannella Jr. 2007. Organizational predictors of women on corporate boards. *Academy of Management Journal* 50 (4) 941-952.
- Hoi, C. K., Q. Wu and H. Zhang. 2013. Is Corporate Social Responsibility (CSR) Associated with Tax Avoidance? Evidence from Irresponsible CSR Activities. *The Accounting Review* 88 (6), 2025-2059.
- Hong, H. and Kostovetsky, L. 2012. Red and blue investing: Values and finance. *Journal of Financial Economics* 103 (1) 1-19.
- Hull, C. E. and S. Rothenberg. 2008. Firm performance: The interactions of corporate social performance with innovation and industry differentiation. *Strategic Management Journal* 29 (7) 781-789.
- Ionnaou, I. and G. Serafeim. 2010. The impact of corporate social responsibility on investment recommendations. *HBS Working Paper* 11-017.
- Ionnaou, I. and G. Serafeim. 2012. The consequences of Mandatory Corporate Sustainability Reporting. *HBS Working Paper* 11-100.
- Jayachandran, S., K. Kalaiganam and M. Eilert. 2013. Product and environmental social performance: varying effect on firm performance. *Strategic Management Journal*, 34 (10) 1255-1264.
- Johnson, R. A. and D. W. Greening. 1999. The effects of corporate governance and institutional ownership types on corporate social performance. *Academy of Management Journal* 42 (5) 564-576.
- Kang, J. 2013. The relationship between corporate diversification and corporate social performance. *Strategic Management Journal*, 34 (1), 94-109.
- Kacperczyk, A. 2009. With greater power comes greater responsibility? Takeover protection and corporate attention to stakeholders. *Strategic Management Journal* 30 (3) 261-285.
- Khan, M., Serafeim, G; and A. Yoon. 2015. Corporate Sustainability: First evidence on materiality. *HBS Working Paper* 15-073.
- Kim, Y., M. S. Park and B. Wier. 2012. Is Earnings Quality Associated with Corporate Social Responsibility? *The Accounting Review*, 87 (3) 761-796.
- Koh, P.-S., C. Qian and H. Wang. 2014. Firm litigation risk and the insurance value of corporate social performance. *Strategic Management Journal*, 35 (10) 1464-1482.
- Lewis, B. W., J. L. Walls and G. W. S. Dowell. 2014. Difference in degrees: CEO characteristics and firm environmental disclosure. *Strategic Management Journal* 35 (5) 712-722.

- Luo, X., V. K. Kanuri and M. Andrews. 2014. How does CEO tenure matter? The mediating role of firm-employee and firm-customer relationships. *Strategic Management Journal*, 35(4), 492-511.
- Lys, T., J. P. Naughton and C. Wang. 2013. Signaling Through Corporate Accountability Reporting. Conference on "Research in Corporate Accountability Reporting", Harvard Business School, January 18, 2013.
- Matsumura, E. M., Prakash, R. and S. C. Vera-Muñoz. 2014. Firm-Value Effects of Carbon Emissions and Carbon Disclosures. *The Accounting Review*, 89 (2) 695-724.
- McWilliams, A. and D. Siegel. 2000. Corporate social responsibility and financial performance: correlation or misspecification? *Strategic Management Journal*, 21 (5), 603-609.
- Miles, M.B. and A.M. Huberman. 1994. Qualitative data analysis: an expanded sourcebook. Thousand Oaks: Sage Publications.
- Muller, A. and R. Kräussl. 2001. Doing good deeds in times of need: a strategic perspective on corporate disaster donations, *Strategic Management Journal*, 32 (9) 911-929.
- Parthiban, D., M. Bloom and A. J. Hillman. 2007. Investor activism, managerial responsiveness and corporate social performance. *Strategic Management Journal* 28 (1), 91-100.
- Patel, P. C. and D. Cooper. 2014. The harder they fall, the faster they rise: Approach and avoidance focus in Narcissistic CEOs. *Strategic Management Journal* 35 (10) 1528-1540.
- Serafeim, G. 2014a. Integrated Reporting and Investor Clientele. *HBS Working Paper* 14-069.
- Serafeim, G. 2014b. The role of the corporation in society: An alternative view and opportunities for further research. *HBS Working Paper* 14-110.
- Sharfman, M. 1996. The Construct Validity of the KLD Performance Data Ratings. *Journal of Business Ethics*, 15 (3), 287-297.
- Sharfman, M. P. and C. S. Fernando. 2008. Environmental risk management and the cost of capital. *Strategic Management Journal*, 29 (6), 569-592.
- Strike, V.M., J. Gao and P. Bansal. 2006. Being good while being bad: social responsibility and the international diversification of U.S. firms. *Journal of International Business Studies* 37(6), 850-862.
- Surroca, J., J. A. Tribó and S. A. Zahra. 2013. Stakeholder pressure on MNEs and the transfer of socially irresponsible practices to subsidiaries. *Academy of Management Journal*, 56 (2), 549-572.
- Surroca, J., J. A. Tribó and S. A. Waddock. 2010. Corporate responsibility and financial performance: the role of intangible resources. *Strategic Management Journal* 31 (5) 463-490.
- Swanson, E. P. 2004. Publishing in the Majors: A Comparison of Accounting, Finance, Management and Marketing. *Contemporary Accounting Research* 21 (1) 223-255.
- Turban, D. B. and D. W. Greening. 1996. Corporate social performance and organizational attractiveness to prospective employees. *Academy of Management Journal*, 40 (3) 658-672.
- van den Heuvel, R. 2012. How Robust are CSR Benchmarks? Comparing ASSET4 with Sustainalytics. Masters' Thesis in Economics, Tilburg University.
- Waddock, S. A. and S. B Graves. 1997. The corporate social performance-financial performance link. *Strategic Management Journal*, 18 (4), 303-319.
- Walls, J. L., P. Berrone and P. H. Phan. 2012. Corporate Governance and environmental performance: is there really a link?, *Strategic Management Journal*, 33 (8), 885-913.

- Wang, H. C., J. He, J. T. Mahoney. 2009. Firm-specific knowledge resources and competitive advantage: the roles of economic – and relationship –based employee governance mechanisms. *Strategic Management Journal* 30 (12) 1265-1285.
- Wong E. M., M. E. Ormiston, and P. E. Tetlock. 2011. The effects of top management team integrative complexity and decentralized decision making on corporate social performance. *Academy of Management Journal*, 54 (6) 1207-1228.

Table 1

	ASSET4	MSCI (former KLD)	Sustainalytics	As proxy for
Khan et al. (2015)		X		Sustainability performance
Gao et al. (2014)		X		To measure CSR orientation as a proxy for reputational costs
Matsumura et al. (2014)		X		Environmental performance
Lys et al. (JAE conference)	X			Measure of CSR expenditure (A2IR) produced by ASSET4
Patel & Cooper (2014)		X		CSR
Serafeim (2014a)	X			Uses on specific score to measure Integrated Reporting
Serafeim (2014b)	X			Uses overall social and environmental scores to measure social and environmental commitment and performance
Hoi et al. (2013)		X		To measure irresponsible CSR activities
Koh et al. (2014)		X		CSP
Surroca et al. (2013)			X	Corporate responsibility performance (treatment of stkh)
Flammer (2013)		X		Environmental corporate social responsibility
Luo et al. (2014)		X		Firm-employee relationship Firm-customer relationship
Cheng et al. (2013)	X			CSR performance
Jayachandran et al. (2013)		X		Product social performance Environmental social performance
Deng et al. (2013)		X		Social performance
Lewis et al. (2014)		X		Transparency strength Climate change concern as a proxy for a firm's prior environmental performance
Chin et al. (2013)		X		CSR profile
Kang (2013)		X		Corporate social performance
Kim et al. (2012)		X		CSR
Barnett & Solomon (2012)		X		CSP – the net KLD score is an assessment of the overall level of social responsibility and so proxies for stakeholder influence capacity
Walls et al. (2012)		X		Environmental performance
Hong & Kostovetsky (2012)		X		Social responsibility
Ioannou & Serafeim (2012) unpublished HBS	X			ESG performance
Muller & Kraussl (2011)		X		Social irresponsibility
Bae et al. (2011)		X		Firm's employee treatment

Wong et al. (2011)		X		Corporate social performance
Dhaliwal et al. (2011)		X		Social performance
Ioannou & Serafeim (2010) unpublished HBS		X		CSR
Chatterji & Toffel (2010)		X		Rating
Surroca et al. (2010)			X	Corporate Social Performance
Cho et al. (2010)		X		Environmental performance
Godfrey et al. (2009)		X		CSR participation / firm engagement
Choi & Wang (2009)		X		Aggregate measure of stkh relations
Kacperczyk (2009)		X		Measure corporate attention to stakeholders
Sharfman & Fernando (2008)		X		Environmental risk management
Briscoe & Safford (2008)		X		CSR score
Hull & Rothenberg (2008)		X		Corporate Social Performance
Parthiban et al. (2007)		X		Corporate Social Performance
Cho & Patten (2007)		X		Environmental performance
Coombs & Gilley (2005)		X		Stakeholder management
Hillman & Keim (2001)		X		Stakeholder management
McWilliams & Siegel (2000)		X		Corporate social performance
Berman et al. (1999)		X		To assess stkh relationships / stkh posture
Johnson & Greening (1999)		X		Two dimensions of CSR performance: people and a product quality
Agle et al. (1999)		X		Corporate social performance
Waddock & Graves (1997)		X		Corporate social performance
Turban & Greening (1996)		X		Corporate social performance
Graves & Waddock (1994)		X		Corporate social performance
Hillman et al. (2007)		x		Only women perspective
Wang et al. (2009)		x		Only firm-employee relationship

Table 2

MSCI (ESG) – KLD	Dim.	S & C	0/1/2	w	Dummy	Comments
Khan et al. (2015)	7	Net	1	/	No	<ul style="list-style-type: none"> • KLD material strengths – KLD material concerns • KLD immaterial strengths – KLD immaterial concerns
Gao et al. (2014)	5 + Add concerns in screens	Net	1	/	Both	<ul style="list-style-type: none"> • CSR conscious: if TS-TC is positive • Not hr & cg category • Net score for each of the five dimensions
Matsumura et al. (2014)	Only env.	Both	1	/	No	<ul style="list-style-type: none"> • Environmental strengths • Environmental concerns
Patel & Cooper (2014)	5	n.c.	Z-scores	n.c.	No	Only the following information is provided in the paper: Firms with higher corporate social responsibility (z-scores across Community, Diversity, Employee Relations, Environment and Product/Customer categories in the KLD database
Koh et al. (2014)	5	Net	Z-score	/	No	Each dimension was standardized to make the scores directly comparable across dimensions and over time (Mattingly and Berman, 2006). The KLD index evaluates each of the five dimensions in terms of ‘strengths’ and ‘concerns.’ The weaknesses were subtracted from the strengths to arrive at a net score for each dimension.
Flammer (2013)	Only env.	Separately	0/1	/	No	Environmental concerns: 0/1 Environmental strengths: 0/1
Deng et al. (2013)	7	Net	+1/-1?	/	No	If a firm conducts a good deed (a harm) listed as strength (concern) indicator, it gains (loses) one point.
Chin et al. (2013)	6	Net	1	/	No	<ul style="list-style-type: none"> • The net score is an overall score
Kang (2013)	7	Net	1	/	No	<ul style="list-style-type: none"> • Robustness checks • 7 categories individually as well • Separate strengths and concern scores

Kim et al. (2012)	5	Net	1	/	No	<ul style="list-style-type: none"> • CSR firm if TS – TC is positive, used in descriptive tables • Exclusion of cg category • Sensitivity check: Total strengths and total concerns separately in regression • Net score per category
Barnett & Solomon (2012)	7 + socially controversial business activities	Net	-1/+1/0 (7) -1/0 (controversial)	/	No	<ul style="list-style-type: none"> • Robustness checks: • Using different weightings • Based on only the 7 dimensions
Walls et al. (2012)	Only environment	Separately	1	/	No	
Hong & Kostovetsky (2012)	4	Net & Separately	-1/+1	/	No	Robustness check: 7 dimensions Ratings for a firm in each category are obtained by adding one point for each strength and subtracting one point for each concern.
Wong et al. (2011)	7	Net overall	-1/+1 & average	/	No	<ul style="list-style-type: none"> • Controlled for industry • Robustness checks: • Only 5 dimension • Total strengths (7) and total concerns (7) separately
Dhaliwal et al. (2011)	7	Only S	1	/	No	Industry-adjusted
	7	Only S	1	/	Yes	Compared to industry median
Ioannou & Serafeim (2010) unpublished HBS	6	Both	1	/	No	
Chatterji & Toffel (2010)	Only env.	Both	1	/	Yes	Poor = no strengths, only concerns Mixed or good= only strengths or both strengths and concerns

Cho et al. (2010)	Only env.	Only C	1	/	No	
Godfrey et al. (2009)	6	Separately	1	/	Yes	Final measures a bit different than in other studies CSR participation if one of the 42 items is positive CSR negative level: number of negative items
Choi & Wang (2009)	5	Net	Z-score	/	No	To arrive at a net score for each dimension, the total number of concerns was subtracted from the total number of strengths (Graves, Waddock, and Kelly, 2005). In addition, each dimension was standardized to make the scores across dimensions directly comparable (Mattingly and Berman, 2006). The average of standardized scores on the five dimensions with an equal weight was then used to obtain an aggregate stakeholder relations score (Hillman and Keim, 2001).
Kacperczyk (2009)	5	Only S	1	/	Yes	If at least one strength over the 5 dimensions => 1, otherwise zero Robustness check: Variable: total number of strengths Dummy for each stakeholder
Briscoe & Safford (2008)	4 pos & 4 neg items	Net	n.c.	n.c.	No	<ul style="list-style-type: none"> • Completely different topic • Four positive and four negative items selected by author
Hull & Rothenberg (2008)	8?	Only C?	1/0?	Yes via subcat	No	Very vague
Parthiban et al. (2007)	5	Net	+2/-2	n.c.	No	<ul style="list-style-type: none"> • Adding the strengths and subtracting the weaknesses for a composite measure • KLD score for each category are on a five point-scale from significant concern to significant strength.
Cho & Patten (2007)	Only env.	Only C	1	/	Yes	
Coombs & Gilley (2005)	5	Net Per dim	+2/-2	/	No	Given ratings on five-point Likert type scales No overall measure

Hillman & Keim (2001)	5	Net?	-2/+2	no	No	The KLD categories are rated on a scale ranging from -2 (major concern), -1 (concern), 0 (neutral), +1 (strength), +2 (major strength)
McWilliams & Siegel (2000)	8	All	/	/	Yes (in DSJI or not)	<ul style="list-style-type: none"> • Very rough measurement, just in DSJI or not
Berman et al. (1999)	5	Net per dim.	-2/+2	/	No	<ul style="list-style-type: none"> • All 5 net scores are separately included in the regression • No overall net score • Bizar: KLD measures firm actions toward each of the five stkh groups using a five point likert scale ranging from -2 to +2
Johnson & Greening (1999)	5 → 2	NC	NC	NC	No	<ul style="list-style-type: none"> • Two dimensions of CSR performance: people (women and minorities, community and employee relations) and a product quality (product & environment) dimension • Industry adjusted
Agle et al. (1999)	4	NC	-2/+2 contr. indus		No	<ul style="list-style-type: none"> • Five point scale • 4 dim: combined employee rel and treatment of women and minorities into one dimension • Controlled for industry effects • Averaged ratings over 2 years • Say that they calculated overall rate, but did not really use it in their analyses
Waddock & Graves (1997)	Eight dimensions	Net	-2/+2	Yes		<ul style="list-style-type: none"> • First five: five step scale • Next 3: -2/-1/0 • Weightings are used
Turban & Greening (1996)	5	Net per dimension	-2/+2	/	No	<ul style="list-style-type: none"> • All 5 dimensions included • No overall net score
Graves & Waddock (1994)	8	Net	-2/+2	Yes	No	<ul style="list-style-type: none"> • First five: five step scale • Next 3: -2/-1/0 • Weightings are used

Hoi et al. (2013)	7	Only C	1	No	No & Yes	First concerns score, but then High Neg CSR
Jayachandran et al. (2013)	2 Product & environment	Net	N.C.	No	No	Product social performance Environmental social performance
Sharfman and Fernando (2008)	Only environment	Separately	N.C.	No	No	To calculate the KLD score they averaged the strengths and concerns separately. Combination with TRI data
Muller and Krausler (2011)	7	Only C	1	/	/	Social irresponsibility: average number of concerns for each firm (per year) over the 2000-2004 period Social responsibility (not core measure): average number of strengths for each firm (per year) over the 2000-2004 period

Notes: w = weights, s = strengths, c = concerns, n.c. = not clear , TS = total strengths, TC = total concerns; Separately: strengths and concerns appear in different regressions, Both: strengths and concerns appear in the same regression

Table 3: Theorization of CSR

ASSET4	MSCI	Sustainalytics
<p>Environmental Performance: Resource Reduction Emission Reduction Product Innovation</p>	<p>Environment: <i>Strengths:</i> Environmental Opportunities Waste Management Packaging Materials & Waste Climate Change Environmental Management Systems Water Stress Biodiversity & Land Use Raw Material Sourcing Other Strength <i>Concerns:</i> Regulatory Compliance Toxic Spills & Releases Climate Change Impact of Products & Services Biodiversity & Land Use Operational Waste Supply Chain Management Waste Management Other concern</p>	<p>Environmental: Operations Supply Chain Products and Services</p>
<p>Social Performance: Employment Quality Health & Safety Training & Development Diversity Human Rights Community Product Responsibility</p>	<p>Social: Community <i>Strengths:</i> Innovative Giving Community Engagement <i>Concerns:</i> Community Impact Human Rights <i>Strengths:</i> Indigenous Peoples Relations Strength Human Rights Policies & Initiatives</p>	<p>Social: Employees Supply Chain Customers Community and Philanthropy</p>

	<p><i>Concerns:</i> Support for Controversial Regimes Freedom of Expression & Censorship Human Right Violations Other Concern</p> <p>Employee Relations</p> <p><i>Strengths:</i> Union Relations Cash Profit Sharing Employee involvement Employee Health & Safety Supply Chain Labor Standards Compensation & Benefits Employee Relations Professional Development Human Capital Management:</p> <p><i>Concerns:</i> Union Relations Employee Health & Safety Supply Chain Child Labor Labor-Management Relations</p> <p>Diversity</p> <p><i>Strengths:</i> Board of Directors-Gender Women & Minority Contracting Employment of Underrepresented Groups</p> <p><i>Concerns:</i> Workforce diversity Board of Directors –Gender Board of Directors – Minorities</p> <p>Product</p> <p><i>Strengths:</i> Quality Social opportunities</p>	
--	--	--

	<p>Access to Finance</p> <p><i>Concerns:</i></p> <p>Product Quality & Safety</p> <p>Marketing & Advertising</p> <p>Anticompetitive practices</p> <p>Customer Relations</p> <p>Other concerns</p>	
<p>Corporate Governance Performance:</p> <p>Board Structure</p> <p>Compensation Policy</p> <p>Board Functions</p> <p>Shareholder rights</p> <p>Vision and Strategy</p>	<p>Governance</p> <p><i>Strengths:</i></p> <p>Reporting Quality</p> <p>Corruption & Political Instability</p> <p>Financial System Instability</p> <p><i>Concerns:</i></p> <p>Reporting Quality</p> <p>Governance Structures</p> <p>Controversial Investments</p> <p>Business Ethics</p> <p>Other concerns</p>	<p>Governance:</p> <p>Business Ethics</p> <p>Corporate Governance</p> <p>Public Policy</p>
<p>Economic Performance:</p> <p>Client Loyalty</p> <p>Performance</p> <p>Shareholder Loyalty</p>	/	

Table 4: Main differences related to commensurability

ASSET4	MSCI	Sustainalytics
<p>Driver indicators: capture the information about availability of policies and processes that companies have committed to.</p> <p>Outcome indicators: measure the results of the policies and company actions. Controversies are also considered as outcome indicators.</p>	<p>Strengths: mainly based upon strategic indicators</p> <p>Concerns: mainly based on performance indicators</p>	<p>Preparedness measures: assesses the organizations capability to deal with these aspects by looking at policies, managements systems, programs and targets</p> <p>Disclosure measures: consider the company's transparency on preparedness and performance via sustainability reporting and the implementation of key reporting and verification standards</p> <p>Quantitative performance measures: company's performance based on quantitative social and environmental metrics, including both absolute and relative terms.</p> <p>Qualitative performance measures: used for the controversy assessment.</p>
<p>The ESG data is collected by specially trained research analysts using news sources, stock exchange filings, non-governmental organizations as well as from company reports.</p>	<p>The ratings are based on the assessment made by KLD analysts after direct communications with company officers and visits to company facilities, and extensive investigation of public records, including more than 14,000 global media sources, company websites, reports from governmental agencies and NGOs, financial reports filed with regulatory agencies, and company annual reports.</p>	<p>Sustainalytics analysts use the following sources in their assessment: annual reports, financial reports, 10-K reports, CSR reports, SCR sections on websites, news releases, news sources such as Bloomberg, newspapers and NGOs, as well as company's feedback.</p>

Table 5.
Panel A. Correlation Matrix

	1	2	3	4	5	6
1 EDO	1					
2 EDC	0.210*	1				
3 MSCI Env. Concerns	0.221*	0.05	1			
4 MSCI Env. Strengths	0.214*	0.085	0.05	1		
5 MSCI Env. Net Score	-0.121	-0.013	-0.916***	0.355***	1	
6 ASSET4 Env. Score	0.261**	0.192*	0.301***	0.178*	-0.211*	1

EDO = the 2002 environmental disclosure “optimism” score; EDC = the 2002 environmental disclosure “certainty” score; MSCI Env. Concerns = the 2002 environmental Concerns score from KLD; MSCI Env. Strengths = the 2002 environmental Strengths score from KLD; MSCI Env. Net Score = environmental Strengths – environmental Concerns; EnvASSET4 Env. Score = the 2002 total environmental performance score as supplied by ASSET4

* p < 0.10; ** p < 0.05; *** p < 0.01.

Panel B. Replicating Cho et al. 2010, across CSR rating providers

	(1) EDO	(2) EDC	(3) EDO	(4) EDC	(5) EDO	(6) EDC
MSCI Env. Concerns	0.399* [1.912]	-2.293** [-2.050]	0.159 [0.675]	-2.343* [-1.701]		
ASSET4 Env. Score					0.011 [1.174]	0.010 [0.191]
Controls	YES	YES	YES	YES	YES	YES
Observations	190	190	129	129	129	129
R-squared	0.104	0.251	0.171	0.341	0.177	0.325

Variables defined in notes to Panel A; * p < 0.10; ** p < 0.05; *** p < 0.01.

Panel. C. Replicating Cho et al. 2010, across MSCI proxies

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	EDO	EDC	EDO	EDC	EDO	EDC	EDO	EDC
MSCI Env. Concerns	0.399* [1.912]	-2.293** [-2.050]			0.365* [1.735]	-2.499** [-2.217]		
MSCI Env. Strengths			0.581 [1.395]	2.264 [1.009]	0.480 [1.148]	2.953 [1.318]		
MSCI Env. Net							-0.226 [-1.136]	2.574** [2.440]
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	190	190	190	190	190	190	190	190
R-squared	0.104	0.251	0.095	0.238	0.110	0.258	0.092	0.258

Variables defined in notes to Panel A; * p < 0.10; ** p < 0.05; *** p < 0.01.

Table 6.
Panel A. Descriptive Statistics

	N. of obs = 173	Mean	SD	1	2	3	4	5	6	7	8	9
1	Disclosure score	15.231	6.143	1								
2	MSCI Net score	2.474	5.184	0.406***	1							
3	ASSET4 Total Score	75.689	17.863	0.582***	0.498***	1						
4	Sustainalytics Total Score	56.304	7.108	0.433***	0.582***	0.557***	1					
5	MSCI Strengths	7.231	5.057	0.533***	0.829***	0.605***	0.633***	1				
6	MSCI Concerns	4.757	2.994	0.197**	-0.331***	0.160*	0.061	0.253***	1			
7	SizeRank	102.289	61.152	-0.261***	-0.058	-0.261***	-0.260***	-0.388***	-0.556***	1		
8	ROA	0.116	0.065	-0.029	0.221**	0.146	0.103	0.154*	-0.124	0.007	1	
9	ESI	0.318	0.467	0.063	0.156*	0.08	0.05	0.185*	0.043	-0.105	0.013	1

Table 6. Relationship between CSR disclosure and CSR performance, using Cho et al. 2015

	(1)	(2)	(3)	(4)	(5)	(6)
MSCI Net	0.501*** [6.030]					
ASSET4 Total Score		0.196*** [8.803]				
Sustainalytics Total Score			0.344*** [5.597]			
MSCI Concerns				0.057 [0.349]	0.150 [0.810]	
MSCI Strengths				0.648*** [7.371]		0.650*** [7.441]
SizeRank	-0.024*** [-3.506]	-0.011* [-1.723]	-0.015** [-2.160]	-0.004 [-0.509]	-0.022** [-2.400]	-0.006 [-0.800]
ROA	-11.395* [-1.748]	-10.507* [-1.791]	-6.570 [-1.015]	-10.078 [-1.610]	-1.818 [-0.257]	-10.418* [-1.689]
ESI	-0.340 [-0.377]	0.105 [0.129]	0.372 [0.412]	-0.521 [-0.601]	0.499 [0.508]	-0.530 [-0.613]
Constant	17.872*** [15.606]	2.749 [1.357]	-1.929 [-0.515]	12.043*** [6.554]	16.790*** [8.501]	12.486*** [9.432]
Observations	173	173	173	173	173	173
R-squared	0.236	0.364	0.216	0.301	0.074	0.301

t-statistics in brackets

*** p<0.01, ** p<0.05, * p<0.1

Table 7.
Panel A. Descriptive statistics

	N. of obs = 165	Mean	SD	1	2	3	4	5	6	7	8	9	10	11
1	Disclosure score	15.521	5.990	1										
2	MSCI Concerns	4.812	3.019	0.185*	1									
3	MSCI Strengths	7.430	5.017	0.512***	0.234**	1								
4	ASSET4 Drivers	62.839	8.973	0.565***	0.251**	0.656***	1							
5	ASSET4 Outcomes	55.231	5.911	0.546***	0.229**	0.708***	0.866***	1						
6	Sustainal. Disclosure	41.385	20.602	0.452***	0.343***	0.605***	0.589***	0.638***	1					
7	Sustainal. Performance	38.229	12.787	0.236**	0.114	0.418***	0.322***	0.423***	0.444***	1				
8	Sustainal. Preparedness	51.053	10.176	0.403***	0.158*	0.588***	0.594***	0.585***	0.497***	0.238**	1			
9	SizeRank	101.00	60.759	-0.248**	-0.595***	-0.410***	-0.354***	-0.412***	-0.391***	-0.213**	-0.287***	1		
10	ROA	0.119	0.065	-0.088	-0.166*	0.111	0.096	0.092	0.057	0.122	0.029	0.024	1	
11	ESI	0.333	0.473	0.033	0.031	0.165*	0.138	0.197*	0.208**	0.178*	0.054	-0.094	-0.018	1

*** p<0.01, ** p<0.05, * p<0.1

Table 7.
Panel B. Multivariate analysis with CSR performance across different providers

	(1)	(2)	(3)
MSCI Concerns	0.062 [0.369]		
MSCI Strengths	0.625*** [7.001]		
ASSET4 Drivers		0.246*** [2.888]	
ASSET4 Outcomes		0.249* [1.862]	
Sustainalytics Disclosure			0.091*** [3.529]
Sustainalytics Performance			0.027 [0.747]
Sustainalytics Preparedness			0.133*** [2.846]
SizeRank	-0.002 [-0.189]	-0.002 [-0.270]	-0.005 [-0.680]
ROA	-13.081** [-2.053]	-13.588** [-2.291]	-11.034* [-1.733]
ESI	-0.743 [-0.863]	-0.896 [-1.087]	-0.784 [-0.882]
Constant	12.554*** [6.260]	-11.619*** [-2.660]	6.018** [2.213]
Observations	165	165	165
R-squared	0.289	0.358	0.268

t-statistics in brackets

*** p<0.01, ** p<0.05, * p<0.1

Table 5 Panel B shows the results for the following regression:

$$\text{Disclosure Score} = a_0 + a_1 \text{ CSR dimensions} + a_2 \text{ Size Rank} + a_3 \text{ ROA} + a_4 \text{ ESI}$$

Where Disclosure Score is the disclosure score as measured in Cho et al. (2015); CSR dimensions are alternatively broken down into: total MSCI Concerns and total MSCI Strengths (column 1), ASSET4 Drivers and ASSET4 Outcomes (column 2), Disclosures, Performance and Preparedness by Sustainalytics (column 3); Size is measured as the relative rank of the company within the sample for its year based on total revenues (Cho et al. 2015); ROA is equal to operating profit/total assets; ESI is a dummy variable equal to 1 if the company operates in the extractive, paper, chemicals, petroleum, and metals industries.