How Robust are CSR Benchmarks?

Comparing ASSET4 with Sustainalytics

Master thesis of the department of Economics

Faculty of Economics and Business Administration

Tilburg University

Name:	Remco van den Heuvel
ANR:	577166
Supervisors:	prof. Dr. J.J. Graafland (Tilburg University)
	H. Smid (Tilburg University)
Second Reader:	Dr. J.M.M. Derwall (Tilburg University)
Programme:	Msc. Economics
Department:	Economics
Word count:	12.773
Date:	18-10-2012

Acknowledgements

For the past 7 years I have had the pleasure of studying in Tilburg at Tilburg University. Since I finished my bachelors in Business Law last august I knew that this thesis would be my final act as a student. I have enjoyed working on it despite some minor setbacks. Especially the week where I could not get a stable connection with Datastream was very frustrating. When looking at the end result lying in front of me I am proud of what I have accomplished as a student and as a person during my time in Tilburg.

I would especially like to thank prof. Johan Graafland for guiding me through the process of writing this thesis. With always providing sharp insights, useful suggestions and helping with setting realistic deadlines he has been a great thesis supervisor. After hearing all the terrible stories from my colleagues I am especially thankful for the response time in email conversations, which greatly helped in making steps forward. With respect to the data I have to thank Hugo Smid for his assistance and insights. The conversations we had while waiting on Datastream to produce yet another list of indicators made this process much less burdensome. Overall I hope that my endeavors on comparing these two ratings will help their future research so that I can say that I actually will have made a useful contribution.

Without the moral and financial support of my parents I would not have lasted this long and I would not have been able to enjoy my student life as I have done now. I probably would not have started with my second bachelors and I probably would not have gone to Moscow, Dublin and Kopenhagen on study trips and to Gerlos, Val Thorens, Risoul, the USA, Mexico, Belize and Guetemala on vacations. Thank you for making all of this possible and I hope I can do the same for my children in the far future.

Although all these trips were nice and studying at Tilburg was a great experience in the end it were the people that made it awesome. I have enjoyed living at the Burchtstraat 12 with Matthijs and Kim, who after a period of 7 years feel like brothers to me. I would also specifically like to name Tias, who I have been friends with since we met at the introduction camps of the economic faculty and will hopefully remain so for the rest of my life. It were these people and many others like all members of the Ronde Tafel, active members from my study association Asset First International and other friends that made the past 7 years the best years of my life.

Finally I would like to thank my girlfriend Elise who, despite currently studying in Sydney Australia, has been of great support to me. She has shown me what real persistence is when it comes to studying and I will be forever grateful for this. Upon graduation I will jump on an airplane and travel together down-under with you.

Table of Contents

1. Introduction

When trying to measure the quality of Corporate Social Responsibility (CSR) of a company the first problem that emerges from the literature is that there is no consensus on the definition of CSR (Dahlsrud, 2006). The ambiguity could be a natural consequence of the fact that CSR reflects the role of a business in society, which is constantly changing. Besides the change over time there is a difference in values and norms between regions, countries and continents. Consequently these differences lead to alternating interpretations of the phenomenon called CSR. Given the lack of a clear and widely accepted definition of CSR, it is hard to develop a methodology that allows comparison or benchmarking of CSR of different companies in the world.

Nonetheless, during the last decade a large number of investment related companies have set up ratings that try to capture the quality of CSR and create an index which can be used to make investment decisions, justify company policies and shape the view that consumers have of companies. All these ratings use different methodologies, variables and assumptions but if there is divergence in the results this may undermine the whole concept of these ratings, e.g. providing transparency with respect to investment decisions. The CSR quality of a company will only be correctly perceived by the public, if its social and environmental value creation is transparent (Graafland, Eiffinger, & Smid, 2004). If scores of several ratings on one company contradict one another, how can an investor judge which number is more reliable? There may be various causes of divergence of scores. It could simply be differences in the type of indicators used to operationalize CSR or differences in the weights attached to different indicators. It could also stem from more constructive aspects like the methods used for gathering the information, or even at the most fundamental level of defining what CSR actually consists of. If these CSR rating companies do not agree on what they all claim to be measuring, what is an investor to do when confronted with information from these rating companies? An important question is therefore: do ratings of different rating companies indeed diverge and if so, what causes these divergences?

This research will look at two indices, the ASSET4 index from Thomson Reuters and the Sustainalytics index, and compare their methodologies and outcomes. By looking at companies that are present in both indices a conclusion can be made as to what extent these two indices agree and disagree on the quality of CSR of those companies. In addition to simply looking and analysing the rating outcomes an attempt will be made to construct new scores by aligning the manner in which companies are rated to see if this will improve the level of convergence. This leads to the following main research question:

How robust is the CSR rating of a company when one compares the Sustainalytics index and the ASSET4 Index?

In order to analyse this main research question, we have to encounter various sub-questions:

1. What are the general characteristics of the ratings of Sustainalytics and ASSET4?

2. What are the variables used in these two indexes and how are they weighed?

3. What is the methodology behind the gathering of the information on the variables and the use of them in getting to a final rating?

4. If there are differences between the CSR rating of Sustainalytics and ASSET4, can we explain their causes?

5. Can we improve the level of convergence by aligning scoring methods?

This research will proceed as follows. Chapter two will provide a theoretical framework. Next, chapter three will discuss the methodology and data. In chapter four the empirical analyses will be presented. First the datasets will be dissected to the indicator level after which we will rebuild new scores to see if the use of equivalent indicators and weights increases the level of convergence between the overall ratings of Sustainalytics and Asset 4. In the final chapters I present the limitations, recommendations and conclusions of this study.

2. Theoretical framework

This chapter presents the theoretical framework that provides the background for the statistical analyses in chapter four. First, we discuss some research that show that CSR ratings do affect market behaviour, including that of the companies that are rated. Next, we turn to measurement theory, which sheds light on how convergence between CSR ratings is defined and how convergence is related to validity.

2.1 Do CSR ratings impact market behaviour?

The importance of this research depends on the influence that CSR ratings have on the investment decisions of players in financial markets, be it individual consumers or large institutional investors. If investors neglect CSR (ratings) in their investment decisions there is little practical use for research into the robustness of rating systems. If CSR ratings however do play a role there is a need for in-depth knowledge on how these ratings are constructed and applied. Therefore this section will cover an overview of literature on the effects of CSR ratings on investments decisions of firms and investors.

Chatterji, Levine and Toffel (2009) look at the usage of CSR rankings and motives behind social investors and stakeholders' decisions and distinguishe between four kinds of motives namely: financial, deontological, consequentialist and expressive. The first motive behind the decision to invest in a socially responsible company is that it performs financially better. This is based on prior research that had examined how CSR can benefit companies by attracting socially responsible consumers (Bagnoli & Watts, 2003), reducing the threat of regulation (Maxwell et al, 2000), improving their reputations with consumers (Lev et al, 2006) and reducing concern from activists and non-governmental organizations (Baron, 2001). The deontological motive applies to investors who look from a moral point of view at their investments and thus do not wish to profit from unethical or heinous actions (Rosen et al, 1991). The example given in the paper by Rosen and his co-authors (1991) is that the Methodist Church's stock market investments have kept themselves from investing in firms that are involved or associated with alcohol and gambling. Consequentialist motives stem from the will of investors and consumers to reward behaviour that has good consequences for society and to punish bad behaviour. The last motive, called expressive, means that stakeholders use their transactions to express their personal identity to themselves and to others (Williams, 2007). Their socially responsible investment is an extension of their identity and they worry that negative social performance taints companies and thus by extension themselves.

A striking example of how CSR ratings may influence behaviour is TIAA-CREF, the largest US retirement fund, who decided to sell 1.2 million shares of Coke in July 2006 after KLD Research & Analytics removed it from its list of socially responsible companies (Wilbert, 2006). These shares were worth $54.2 million and were part of social choice account portfolios.

The reason behind this removal was that KLD had concerns about the company's labour and environmental practices in the developing world.

Although this is a very appealing example it could have been an isolated incident and we now turn to academic research that has been done on the topic. Cellier and Chollet (2010) have looked at the relationship between CSR ratings and stock prices and financial performance. They research stock price fluctuations around CSR rating announcements to see if these rating announcements are taken into account by financial analysts and investors. They establish that CSR rating announcements have an overall significant positive impact on the stock market, which confirms that CSR rating announcements are incorporated in stock prices. Earlier, Hamilton (1995) found a significant negative impact of the release of information on the use of toxic chemical on stock prices in the US. Furthermore, Klassen and McLaughlin (1996) found significant positive abnormal returns after a firm receives environmental performance awards, and significant negative returns after environmental crises.

In a recent paper Gregory et al. (2011) also found that indicators of corporate social responsibility are valued by markets. Their research shows that the community, diversity and employee aspects of CSR performance, measured by KLD data, are consistently positively valued by markets. The market valuation is expressed as firm value. They also note that environmental factors have increased in importance over time. Their overall conclusion is that companies with superior CSR performance, on average, are rewarded by the market with higher valuations.

Besides reactions in the stock market on CSR ratings company themselves also respond to being rated. Mackenzie and Rees (2011) examined this by looking at companies that were in danger of being excluded from the FTSE4Good index. Some firms were approached and told that they were in danger of being excluded while the others, the control group, were not. They then investigated if there was a difference in future performance between these two groups as measured by the FTSE4Good environmental scores. The companies that were lobbied indeed performed better compared to the control group. Another research by Chatterji and Toffel (2009) researched if poor CSR scores encouraged better performance. Their sample consisted of 588 US firms from the S&P small or mid-capitalisation indices over the period of 1999 to 2004. They establish that firms that performed badly on the KLD rating improved their environmental performance, which was measured by the amount of toxic emissions, compared to firms with higher rankings or firms that were excluded.

The research shows that firms do react on being rated. The papers mentioned though only examined the "negative" part of the spectrum, e.g. firms that perform badly and/or are in danger of being excluded from future rating. What still needs to be looked at is how firms respond after they are rated positively. Will they try to improve further or will they slow down their efforts?

2.2 Measurement theory

Although in 2005 $2 trillion was invested in portfolios using socially responsible investment criteria (Chatterji & Levine, 2007) up until now there is no uniform methodology on how to measure "social responsibility". Opponents therefore argue that raters cannot truly conclude as to which firms are socially responsible. Because of a lack of reliable methods, CSR rating companies ultimately mislead stakeholders (Entine, 2003). In order to evaluate this claim, there is need of a thorough evaluation of the validity of the ratings. One way to do so is to compare the outcomes of various rating systems.

When two CSR ratings rate the same set of companies you would expect a high degree of resemblance between the two outcomes. But like in a regular market there can be imitators and differentiators. When you observe success of a rival rating company you might wonder why their product or service is valued better by the market and move towards their market position by copying desirable aspects, e.g. the market converges. On the contrary, there is the possibility to position yourself in a niche by differentiating away from your competitors. This also holds for CSR ratings.

A paper by Chatterji and Levine (2007) describes two families of theories of convergence. The first refers to convergence with high validity. Convergence with high validity does not only require that two ratings produce similar results, but also that both ratings are reliable, i.e. that both use high quality measurement methods and data.

The second family refers to convergence with low validity. The theory stems from neo-institutional theory and the critical assumption here is that organisations seek legitimacy in their environment, rather than strictly focusing on efficiency (Scott, 1995, Staw & Epstein, 2000). According to this theory pressure to imitate others is highest when the underlying mechanisms, in this case how CSR ratings are constructed, are unclear to decision makers. The result can be herd-like behaviour even when most decision-makers have information that would recommend a different outcome (Banerjee, 1992). Another important factor is that companies that are measured often try to influence the ratings (Meyer & Gupta, 1994). They try to highlight relatively small CSR accomplishments or try to compensate a negative CSR event by a big positive one, which is called "greenwashing" (Lyon & Maxwell, 2011). Furthermore a significant share of the data used in CSR ratings comes from the rated companies themselves (Chatterji et all., 2009). Firms need to influence almost all rating firms since large deviations between scores would raise questions. So convergence between ratings is likely to take place but the question is whether this is due to high quality of measurement and data, or if this is due to factors like imitation and greenwashing.

In contrast to convergence there are also families of differentiation. Again there is a version with high validity. A possible explanation for divergence could be that different CSR ratings try to satisfy the demands of different groups of clients. A CSR rating aimed at filtering out which companies to invest in for a religious group may yield different ratings compared to a CSR rating for groups that just want the best investment deals. The first group will most likely

not invest in arms and tobacco while the second simply looks at return to investment. Another reason why different rating outcomes may still be valid comes from the segmentation possibilities of the ratings in combination with the use of different benchmarks. For instance some ratings divide their sample up in countries, industries, sub-industries and even further, while others use a different segmentation. Several ratings, but not all, measure social performance relative to a reference group, but this reference group may differ, dependent on the segmentation used. All these factors could lead to differentiation in outcomes.

The last possibility leads to differentiation with low validity. This may be the result if raters do not agree on the definition of CSR as well as the measurement tools. The result is a large divergence in results. What remains to be determined is which, if any, of the raters are of good quality and use the right measurement tools and which are not. Since you only know that they do not resemble each other this will require adding extra data in the form of another rating of which the quality is already determined or benchmarking with respect to objective outcome variables.

2.3 External verification

Already in 1996 Scharfman (1996) was looking into the construct validity of a CSR rating. In this case the subject was the Kinder, Lydenberg & Domini (KLD) social performance ratings data. The starting point of this paper is stated as:"Since the evaluations in the KLD database are essentially only the opinions of the principals of that firm, researchers may be concerned about the validity of these ratings. Validity is the one major concern with the KLD ratings. Since the data have only recently been made available to researchers, no studies have appeared that have validated these data". For this reason Scharfman (1996) wanted to test the validity of KLD data by comparing them to other more accepted measures of CSR, or Corporate Social Performance (CSP) as they called it. They take the Fortune corporate reputation survey as their benchmark and compare it to the KLD scores they created. Because the Fortune data may not truly be representative of CSP but rather a measure of reputation, the research included another CSP measure. Data was collected from a holdings list best known as "social choice" mutual funds. A score was given based on how many times a firm was part of a fund portfolio, assuming that "if a firm was chosen to be part of the portfolio it was a better average firm in terms of CSP". They find that their KLD scores correlate with the two other CSP indicators with values ranging from 0.18 to 0.55. From these correlations they answer their question if the KLD ratings correlate sufficiently with other measures of corporate social performance with a qualified "yes". They also correlate the external scores on each other, with very low results (r=0.13), to conclude that they may have been tapping different parts of the CSP construct. The overall conclusion is that their KLD ratings seem to be capturing at least part of the same construct as do other measures of corporate social performance and hence can be considered, to that extent, valid. Scharfman (1996) stresses, however, that validity is a continuous phenomenon, which means that one can only estimate to what degree a measure is valid.

The paper mentioned before by Chatterji, Levine and Toffel (2009) researched the validity of CSR rankings by comparing them with underlying past and future performance. They investigate 588 US firms for the period of 1991-2003 for which they have KLD data. To evaluate the CSR rankings they use a set of performance metrics, including toxic emissions, the dollar value of penalties, the number of penalties, reported chemical or oil spills and permit denials or shut ins. Furthermore, they control for company size and type of industry. With respect to past performance they find a statistically significant relationship between KLD's environmental concerns and past environmental performance. However, when looking at future environmental performance the empirical results show that KLD has less explanatory power than historical performance itself. A statistically significant relationship is found when using KLD sub-scores to predict later emissions and regulatory violations. These relationships are stronger than was the case when these were predicted using historical performance, company size, and industry.

Mackenzie and Rees (2011), also mentioned before, looked at FTSE4Good corporate social responsibility scores of 1825 companies in 25 countries and compared them with ASSET4 ESG scores. Their main goal is to determine whether CSR is higher when there is an "open society" in a company, which is characterized by a lack of entrenched shareholders, good internal governance, accountability through access to the FTSE4Good ranking system and whether voice and accountability is high. Overall CSR scores and sub-scores provided by FTSE4Good are the dependent variables in their study. In the midst of their research they set out to validate their results using ASSET4 data. They correlate the two ratings to confirm that the dimensions measured are the same. After they compensate for the fact that ASSET4 measures a company's performance without risk adjustment they find, according to them, strong correlations. The ASSET4 corporate governance sub-score correlates strongly with FTSE4Good measures of corporate governance (0.53) and with another component they call controlling bribery (0.39). The ASSET4 environmental measure has correlations of 0.57 with two FTSE4Good components named environmental management and climate change. The social pillar correlated with levels of 0.23 and 0.54 with socially related FTSE4Good components. From these findings they conclude:"Thus there is clearly a strong and statistically significant measure of agreement between the scores from the two sources". A second step is taken to validate their choice for FTSE4good. They rerun their statistical test after they replaced the FTSE4Good scores with ASSET4 ESG scores. The results mirror the results produced before thus further confirming their results.

Heyes and Ceton (2009) compare KLD rating to Fortune's ranking of "America's Most Admired Companies" to if 'actual' CSR (measured by KLD) relates to 'perceived' CSR (approximated by Fortune's ranking). When they plotted their 242 matches against each other the following correlation graph emerged shown in figure 1. The companies are spread out over the graph and this is reflected by the r which is only 14%. The conclusion of this research is that CSP reputations are hardly representative of true CSP using the KLD scores.

Figure 1: Actual vs. Perceived CSP
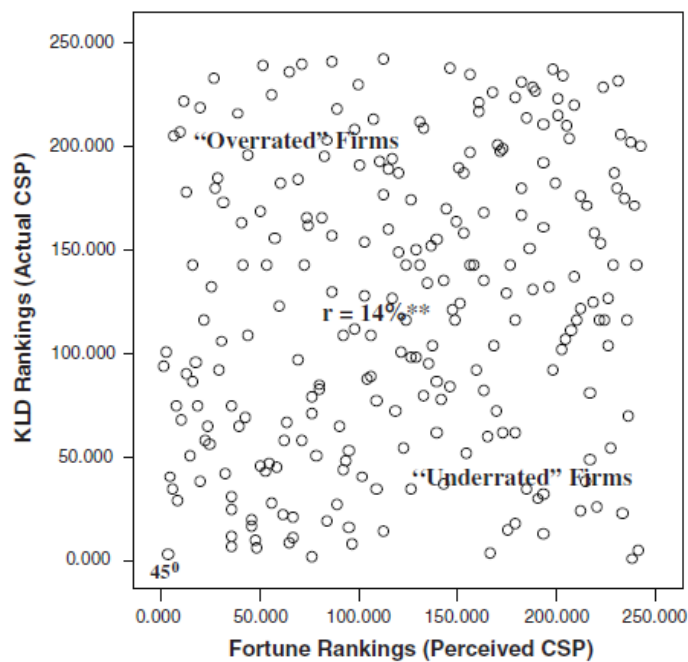
3. Methodology and Data

3.1 Methodology

Our sample consists of those firms that were assessed in the years 2008, 2009 and 2010 both by Sustainalytics and ASSET4. We merge the two databases using the company's ISIN codes. We assess the convergent validity, i.e. the level of similarity, of the two ratings using Pearson correlations. This type of validation states that if the measures of constructs are both theoretically and empirically related to each other (Scharfman, 1996). If they do, this will provide some evidence to conclude that the two different CSR metrics are related to the same construct.

To unravel the differences between the ASSET4 and the Sutainalytics ratings this study will apply a top-down analysis. The starting point is the following formula which describes how a total CSR score for a single year is created for a single company on the basis of the sub-scores:

(1) $$Score = \sum_{i=1}^{n} w_i . s_i$$

In this equation n is the number of sub-scores, w is the weight attached to each score and s is the sub-score itself. By multiplying the weights with the scores and adding them up the final CSR rating is calculated. Both ASSET4 and Sustainalytics provide final scores in their datasets and the first step is to correlate these scores and see how the variance in one can predict values of the other, e.g. the R^2.

Next, we dissect the overall scores and see if the correlation can be improved if the weights used are identical. The first step is to look at the w in formula 1 and see if the correlation can be improved if the weights used are identical. Because ASSET4 uses an equal weights system, this step involves the manipulation of the Sustainalytics weights by also equalising them. Furthermore, as can be seen in figure 1 and 2 below, ASSET4 adopts an economic sub-score while Sustainalytics does not. After manipulating the two overall CSR scores for these two differences the following formula appears:

(2) $$Overall\ Score = \frac{1}{3} * Social_{score} + \frac{1}{3} * Environmental_{score} + \frac{1}{3} * Governance_{score}$$

The goal is to see if after making these changes that make the two ratings theoretically more coherent, we can improve the correlation and prediction power between the two ratings: $H1: A4CSR_{Score} = SUSCSR_{Score}$ $H0: A4CSR_{Score} \neq SUSCSR_{Score}$. As a complementary strategy we also apply the Sustainalytics sector weights to the ASSET4 sub-scores to further confirm if applying the same weights improves the correlation or not.

The next step investigates possible differences in the sub-scores themselves for the social, governance and environment pillars. This is a relative straightforward process since both ratings provide the sub-scores for the environmental, social and governance pillars and these will be compared in the same manner as done before, e.g. bi-variate correlations and R^2. This will provide a first insight into whether possible differences in the overall score stem from a single sub-score, a combination of sub-scores or from random noise.

Moving down a level we will look at how the sub-scores themselves are created. Each sub-score is the product of underlying categories multiplied by their respective weights (for the categories see figure 1 and figure 2 below). However, due to the fact that there is hardly any overlap between the categories of the two ratings, this level will be skipped and the research directly jumps to the lowest level.

At the lowest level of aggregation each category consists of a number of indicators multiplied with their respective weights. To compare the sub-scores, the methodology is used as described above. Before we can start constructing new scores based on comparable weights and indices a comparison will have to be made of how the indices are defined at this level. For each Sustainalytics indicator a matching ASSET4 indicator will be searched for by using the indicator definition files of both ratings. After the identification of a suitable indicator they will also be correlated on each other. This will indicate the degree that theoretically equivalent indicators are empirically related to each other.

Now that we have two sets of indicators with scores constructed in a similar fashion we can next calculate new sub-scores using equal weights. In this phase we encountered, however, a serious problem of lack of data due to the fact that ASSET4 does not provide information on all matched indicators. As a result, the number of observations strongly declined rendering any comparison with Sustainalytics useless. As a remedy, we decided to make a selection of the indicators that will be included in these new sub-scores, based on the number of companies for which these indicators are available in the ASSET4 database. Based on this methodology, we calculated new sub-scores for Sustainalytics and ASSET4 based on exactly similar underlying indicators and weights and compare the results as described before. After the creation of new pillar sub-scores the construction of new overall CSR scores is straightforward by plugging them into formula 2 and redo the correlation tests.

Besides this analysis of convergence between Sustainalytics and ASSET4, we also performed an external verification of the two ratings systems. For this purpose, we related some sub indicators of Sustainalytics and ASSET4 to objective outcome variables, for instance total CO_2 emissions, which the ASSET4 database holds. We correlate these to the matching value indicators controlling for company size, region and sector with the preposition that this outcome variable would be a good benchmark for the indicators.

In this research we will not set any bottom level of correlation on beforehand but we do now have a feeling of what would be sufficient and what not due to the last part of the theoretical framework. Fortunately we use data on 3 years so this already provides a base for comparison. Combine this with the fact that we will not only correlate overall CSR scores and pillar sub-

scores, but also a large amount of individual indicators and there will be a large amount of reference data.

3.2 DATA

3.2.1 Sustainalytics

The company was formed from the Canadian founded Jantzi Research Inc., and the European founded Sustainalytics. These two companies merged in 2009 to become, as they call it themselves, a global leader in ESG research and analysis. The companies in their dataset are based on several sources in which they distinguish the core universe from the research universe. The former is the MSCI World Index complemented by a selection of relevant local indices. In the data from 2010 these include the S&P 500, the TSX Index, the Jantzi Social Index and the AEX[1]. The research universe consists of all the companies in the core universe, supplemented by companies that result from specific client requests, which are mainly institutional investors. Companies are analysed by local research partners using one consistent, sound methodology that consists of 110-160 indicators per company. The company reviews the variables needed for their research on a quarterly basis.

3.2.1.1 Construction

The issues that Sustainalytics researches can be constructed like a tree (see Figure 2). Three main pillars are distinguished: Environment, Social, and Governance. Below this level several categories are distinguished like Operations, Products and Services and Employees. Below the categories we find various indicators. Research is ultimately conducted at the indicator level. Sustainalytics use two kinds of indicator templates. A full template is used for all companies on the MSCI World Index. Junior templates may be assigned to companies outside of the MSCI, when those companies have lower market cap thresholds (below 2 billion) and less exposure and/or interest. Junior templates contain significantly fewer indicators than full templates.

In general, two classes of indicators are distinguished. Core indicators are those used by all companies in all peer groups for which the senior template is applied. A template can be adjusted to company specific characteristics by deactivating an indicator when it is not considered relevant given the specific activities of the company and/or the geographic locations of a company. When information is not available on a company it is still given a score for an

---

[1] The **MCSI World Index** is a stock market index of 1500 world stocks. This index is often used as a common benchmark for 'world' or 'global' stock funds. The index includes a collection of stocks of all the developed markets in the world, as defined by MSCI. The index includes securities from 23 countries but excludes stocks from emerging economies. The **S&P 500** index comprises the 500 large-cap common stocks traded in the United States. The stocks are publicly held companies that are traded on the NYSE Euronext or the NASDAQ OMX. The **Jantzi Social Index** is a common stock index that consists of 60 Canadian companies that pass a set of broadly based environmental, social, and governance rating criteria. The **AEX** is a Dutch index of the top 25-30 companies.

indicator. The exception to this rule is when the lack of data is not due to the lack of cooperation of a company. For example, for the indicator that looks into whether companies have responded to the questionnaire of the Carbon Disclosure Project (CDP), the indicator is deactivated for

Figure 2: Sustainalytics Framework
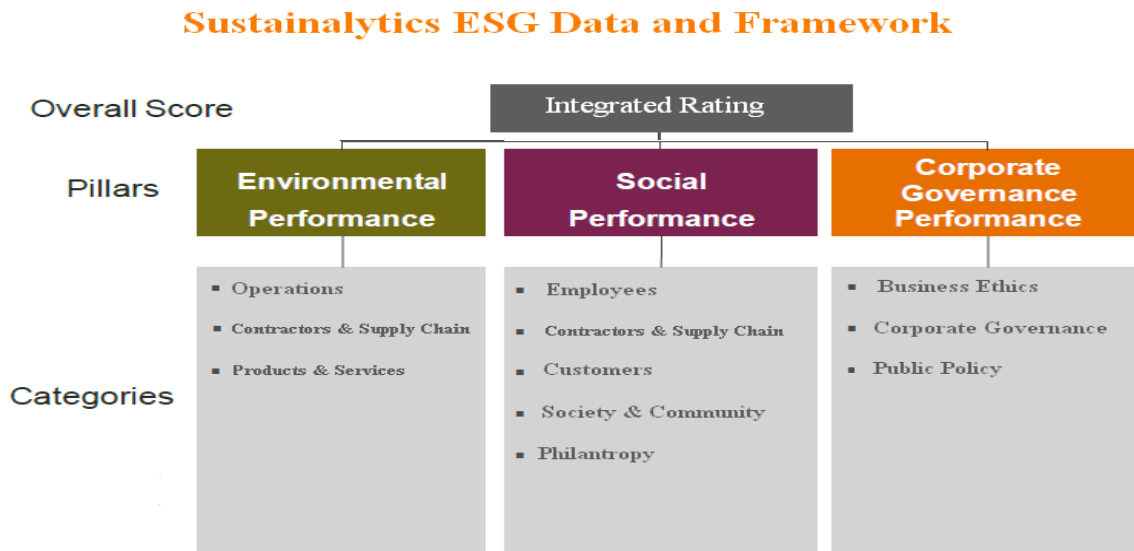
companies that were not contacted by the CDP.

Within each of the two classes of indicators (core and sector-specific), Sustainalytics roughly distinguishes indicators on policies and policy statements, programs and management systems and on outcomes. The scopes of the various kinds of CSR commitments and implementations are also taken into account when the analysts judge a company.

3.2.1.2 Controversies

The Sustainalytics ESG platform contains 10 special indicators that assess whether companies are involved in certain controversies or incidents. These indicators are found for every topic, except Philanthropy. Controversy indicators are seen as very important and therefore carry a weight of 3% in the Sustainalytics default weight matrix (appendix 1): any negative assessment therefore will reduce the score and ranking of a company substantially. These indicators are used by many clients for screening purposes. Therefore, companies with controversies will receive much attention from clients and a very consistent and monthly assessment is needed.

3.2.1.3 Aggregating scores

To calculate the total score of the company, as well as aggregate scores on the three themes and sector scores, Sustainalytics uses a default weight matrix. Clients might change this default

matrix according to their needs when they want to put emphasis on elements they consider important.

Sustainalytics uses customized weights for the indicators to calculate the total ESG rating for every company. These weights are uniquely defined for every peer group. Sustainalytics groups the evaluated companies into 42 different peer industry groups. Every peer group is assessed for a fixed amount of core indicators. In addition, sector-specific indicators are assigned to every peer group. So every peer group is evaluated by a different set of indicators. The senior companies are assessed for about 65 to 80 indicators and the junior companies for 40 to 50 indicators. The companies are assigned a raw score between 0 and 100 for the core indicators and the relevant sector-specific indicators. The overall company score is the weighted average of all raw scores of the relevant peer group indicators.

The weight of the topic is simply the summation of the weights of the relevant indicators for the peer group. Logically, the total weight of a theme is the summation of the weights of the topics. The total weight of the themes adds up to 100%. To give an impression about the weights of the themes, the averaged weights of the themes Governance, Social, and Environment are respectively 25.8%, 38.4%, and 35.8% for the total dataset of senior companies. These weights are different for every peer group, depending on the relevance of the theme for the total company score. By providing all this information about their dataset, Sustainalytics provides their customers a transparent customizable service, which still requires certain skill to apply.

3.2.2 ASSET4

ASSET4 has created a database that is said to provide transparent, objective, and auditable extra-financial information and to offer a comprehensive platform for establishing benchmarks for the assessment of corporate performance (Schafer et al., 2006). They were founded in 2003, taken over by Thomson Reuters in 2009 and their headquarters are located in Zurich Switzerland. In 2009 investors that represent more than $2.5 trillion assets used the ASSET4 data. The principal customers of ASSET4 come from the financial sector. They claim to support the transparency of the rating methodology that facilitates understanding the process by which they calculate their scores and sub-scores. The ASSET4 universe includes over 3000 public world companies and covers major indices: S&P 500, MSCI World Index, Nasdaq, FTSE350 and MSCI World Index. The company collects and analyzes data from company reports, company websites, NGO websites, newspapers, journals, and trade publications but the sources of most ESG data are CSR reports created by the companies themselves. All data must be objective and publically available, though analysts are permitted to contact company investor relations offices to learn the location of public data.

ASSET4 has collected data and scored companies on ESG principles since the fiscal year 2002. Analysts at ASSET4 are assigned a company, where the quality of their work is linked to a bonus. There are guidelines to fill in the documented answers and customer contact is handled by an analyst from a different team. They use a yearly unbalanced panel with systematic

environmental, social and governance data. For these scores they use 278 key performance indicators and over 750 individual data points. The key performance indicators are then

Figure 3: ASSET4 Framework
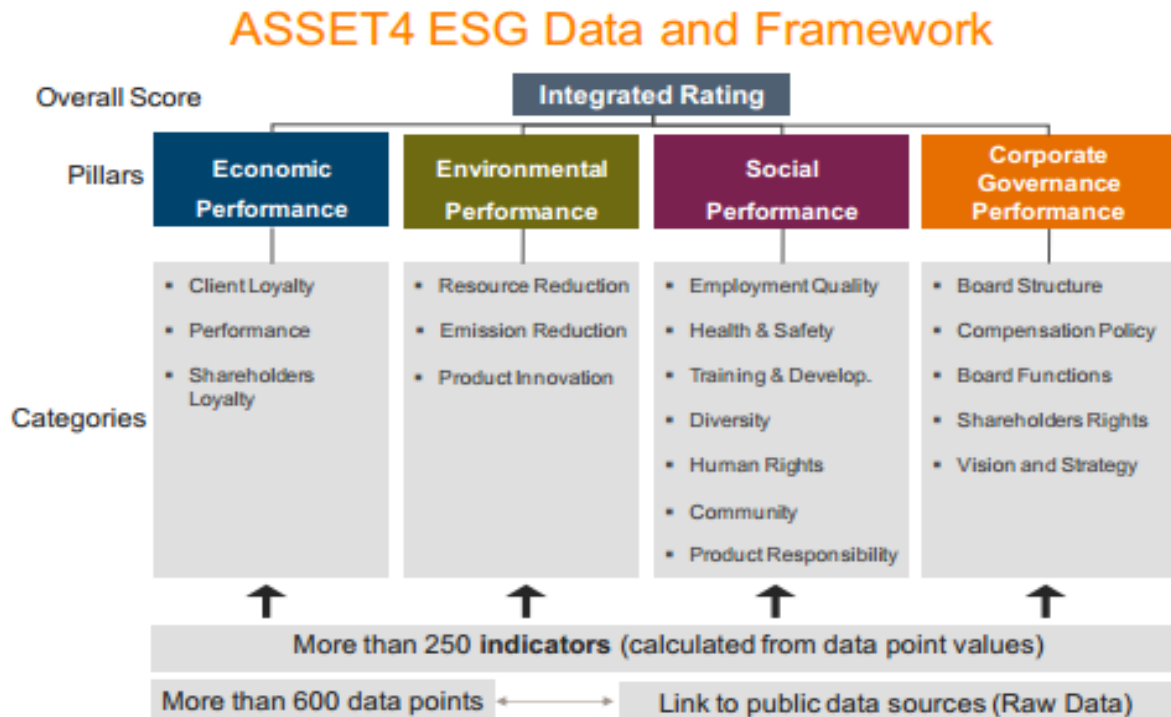
combined into eighteen category scores (Figure 3), which serve as subcomponents of four pillars. Each of the eighteen categories receives a score between 0 and 1, with high scores indicating strong performance in the category. The overall company score, which ASSET4 calls the Integrated Rating, is computed by blending the four pillar scores. A definition or explanation of the word blending was not found.

3.2.3 Assigning value to indicators

One of the steps in our research is to investigate to what extent Sustainalytics and ASSET4 have similar indicators at the disaggregate level. In order to compare these indicators, we need insight into how Sustainalytics and ASSET4 construct these indicators.

Sustainalytics uses scores for each indicator between 0 and 100. The indicator definitions show that Sustainalytics uses this range to express the quality of policies, guidelines and programs. Although this potentially raises the informational content of the score because it requires in-depth research, it also leads to an increase in the level of subjectivity in the score. The people who do the actual rating are given a guideline, which shows them how to score certain

aspects as positive or negative but the definitions are not closed and thus provides room for opinions and personal views.

This can be illustrated by one example, the Sustainalytics indicator S.4.2.2 named Community Involvement. To get a score of 100 for this indicator the following description is given in the Sustainalytics indicators structure report: "*The company has strong and detailed community involvement programs*". Besides this short definition a rater also has the following more elaborate definition:

> "*Select this answer category if the company has a program or otherwise systematically involves communities in major new projects or expansions of activities. If, for example, the company has a department that deals with these issues company-wide, we would usually recommend selecting this answer category. We should, in general, have a good understanding of what the company does in order to select this answer category. So the quality of reporting may affect whether this answer category is selected or not.*"

Even with these kinds of definitions by his/her side, there is still space for the rater for subjective judgment. Words used in the guideline, such as *usually recommend, we should in general, may affect*, illustrate subjective elements in the measurement of the indicators.

This is partially different from how ASSET4 approaches the issue. ASSET4 uses a multiple step approach to create "score" indicators.[2] First, they answer questions by yes/no or number values, like percentages and dollar amounts, and from several of these questions a value is created from which a score is formed. This is illustrated by the example in Table 1. The name of the example indicator is *Score - Community/Policy* which is the match of the Sustainalytics example given before. The following slightly complicated process does not hold for all indicators. Most are simpler and have fewer layers of questions but to get an idea of how ASSET4 constructs its scores a more complicated example will provide a better insight.

Table 1: ASSET4 yes/no questions

Yes/No questions
Does the company have a policy to strive to be a good corporate citizen?
Does the company have a policy to respect the rights of indigenous people?
Does the company have a policy to strive to increase the indirect economic impact it has on local communities?
Does the company have a policy to strive to be a fair competitor?
Does the company have a policy to avoid bribery and corruption at all its operations?
Does the company have a policy to comply with local regulations regarding political contributions?
Does the company have a general, all-purpose policy regarding business ethics?
Does the company have a policy to limit activities in undemocratic countries abusing human

---

[2] It must be noted however that the process described below is what we think the process is that ASSET4 uses as we can see from the materials at hand.

rights?
Does the company have a policy to treat suppliers and contractors as key business partners?

An ASSET4 rater has much less room for subjective elements in answering these questions by only allowing a yes or a no and not directly attaching a score to an observed element. There is still some room, which can be seen in use of the word "strive", which suggests good intentions but not necessarily good outcomes. After these questions are answered an overall conclusion is drawn by adding them up into the following indicator which also is answered with a yes or a no: "*Does the company have a community reputation policy regarding the various elements driving its global reputation and license to operate*?". The third step then is a transformation of the previously answered question into the *Value – Community/Policy* which allows for the last step towards the score indicator which comprises of creating a score which is relative to the whole ASSET4 universe. A problem with this last step is that it is not clear how the Value is translated into a score indicator.

A problem that both raters face is the fact that when a company has guidelines, policies or programs in place which even have deadlines and quantifiable objectives in it, it still does not show if a company actually actively tries to adhere to the principles specified in these documents. To be able to give a complete picture each company would have to be visited by an inspector from a rating agency which then would inspect the whole company and especially look at the outcome of all these documents. This of course is not a realistic approach since costs would be too high and time too short to keep the data up to date. Still it is important to keep in the back of your mind that all these documents might not show the complete truth or simply be window dressing to get a positive rating.

3.2.4 Descriptives

This section will give an overview of the descriptive statistics of the Sustainalytics and ASSET4 dataset. This will provide a first acquaintance with the two datasets before we merge them for the analyses. The information will be provided for the overall ESG scores and the pillar sub-scores for the years 2008, 2009 and 2010.

Several things are worth mentioning after looking at the overall ESG scores of the two raters (table2). First, Sustainalytics has a much smaller range of final scores, between 28.85 and 87.46, compared to ASSET4 which has scores between 2.36 and 97.50. This naturally leads to a much higher standard deviation for the latter. If an investor would look at a company with an overall 2009 Sustainalytics CSR score of 40 without the additional information from table 2 he might draw the wrong conclusions. He could reason that a score of 40 is almost average while it actually is a relative low score given the fact that the lowest score of that year is 33.59. This illustrates that information of the distribution in scores is needed to put the numbers in perspective. Second, the averages are still reasonable close to each other. Actually, this is not in line with what we expected, because Sustainalytics gives a score of 0 when there is no

information available, whereas ASSET4 does not include an indicator on which it has no information. We would expect the scores of Sustainalytics to be pulled downwards by this approach but this is apparently not the case. Third, there is a large increase in the number of firms between year 2009 and 2010 for the Sustainalytics dataset. This is due to the merger between Sustainalytics and Jantzi. This is only good for our analyses since it will most likely lead to a larger overlap in firms.

Table 2: Overall ESG scores

	N	Minimum	Maximum	Mean	Std. Deviation
Sus-08	1135	29.10	78.30	51.43	9.06
Sus-09	1152	33.59	86.69	55.30	9.67
Sus-10	4117	28.85	87.46	51.95	8.85
A4-08	2921	2.36	97.50	49.96	30.66
A4-09	3353	2.91	97.35	49.62	30.76
A4-10	3876	3.09	96.84	49.91	31.00

Tables 3, 4 and 5 show the same information for the sub-scores of the social, environmental and governance pillars. The same pattern emerges from the descriptive statistics of the sub-scores with the exception of the Sustainalytics environmental score of 2010 that ranges between 8.57 and 94.00.

Table 3: Social sub-scores

	N	Minimum	Maximum	Mean	Std. Deviation
Sus-08	1135	25.56	82.71	49.63	11.25
Sus-09	1152	26.67	89.43	54.06	11.78
Sus-10	4117	25.88	89.42	51.62	10.33
A4-08	2921	3.36	97.93	49.76	30.90
A4-09	3353	3.87	97.69	49.43	30.97
A4-10	3876	3.59	97.58	49.41	31.12

Table 4: Environmental sub-scores

	N	Minimum	Maximum	Mean	Std. Deviation
Sus-08	1135	21.50	91.67	49.65	11.81
Sus-09	1152	25.37	91.02	52.45	12.25
Sus-10	4117	8.57	94.00	46.22	12.53
A4-08	2921	9.59	94.10	49.54	32.07
A4-09	3353	9.43	94.26	49.32	32.05
A4-10	3876	8.90	94.75	49.31	31.98

Table 5: Governance sub-scores

	N	Minimum	Maximum	Mean	Std. Deviation
Sus-08	1135	30.80	94.40	57.18	9.69
Sus-09	1152	33.63	94.40	61.68	10.72
Sus-10	4117	26.28	94.40	60.96	10.25
A4-08	2921	1.39	97.06	51.72	30.25
A4-09	3353	1.33	97.03	51.55	30.17
A4-10	3876	1.73	96.38	51.82	30.37

4. Analysis

This chapter will contain the most important part of this thesis, the analysis. Following the steps described in the methodology the ASSET4 and Sustainalytics CSR ratings will be unravelled as far as possible to get to explanations why the two would be different or similar. The first part will cover the overall CSR scores after which the focus will be on sub-scores and the underlying indicators.

4.1 Overall CSR scores

As a starting point the overall CSR scores of ASSET4 and Sustainalytics are to be compared. Plotting them against each other gives figures 4, 5 and 6. In these pictures EW20.. refers to ASSET4 and CompanyScore.. refers to Sustainalytics.

For the year 2008 there are 1023 companies that have received a rating by both companies. When correlated with each other the Pearson coefficient is 0,664, with significance at the 0.01 level (2-tailed) (see Table 6). When regressed the R^2 is 0.441, which means that 44% of the total variation in the ASSET4 CSR scores is explained by the variation in the Sustainalytics CSR score. For 2009 there are 1028 observations and the Pearson correlation coefficient is 0.657 and the R^2 is 0.432. For 2010 there are 3032 observations. The Pearson correlation coefficient is 0.634 and the R^2 is 0.402.

Because ASSET4 uses four sub-scores and Sustainalytics three, the content of the overall scores differs. Furthermore, ASSET4 uses an equal weights system while Sustainalytics uses custom weights. To correct for these two differences new overall CSR scores were created for both ratings using a weight of a third for the social, environmental and governance sub-scores, leaving out the economic pillar. Table 6 shows that the increase in theoretical coherence improves the empirical convergence between the ASSET4 and Sustainalytics ratings. For all years, the Pearson correlation coefficient slightly increases. If we replace the ASSET4 weights by the weights used by Sustainalytics, the results even further improve[34]. Why there is a larger increase than in the equal weights case is not clear since basically the same effect takes place, e.g. the weights are equalised. It could simply be the result of the fact that the companies that are left out due to the absence of a peer-group classification were relative outliers and leaving them out would thus result in better correlations.

---

[3] In appendix 2 three examples of the Sustainalytics default weight matrix can be found. For instance companies that are in the automobiles industry have a weight of 25% for governance, 30% for social and 45 % for environmental while companies that are in textiles, apparels and luxury goods have respective weights of 25%, 45% and 30%. These same industry classifications are now applied

[4] Note that the number of observations for all three years is lower. This is due to the fact that not all companies are given a peer group reference in the data.

Figure 4
Overall 2008
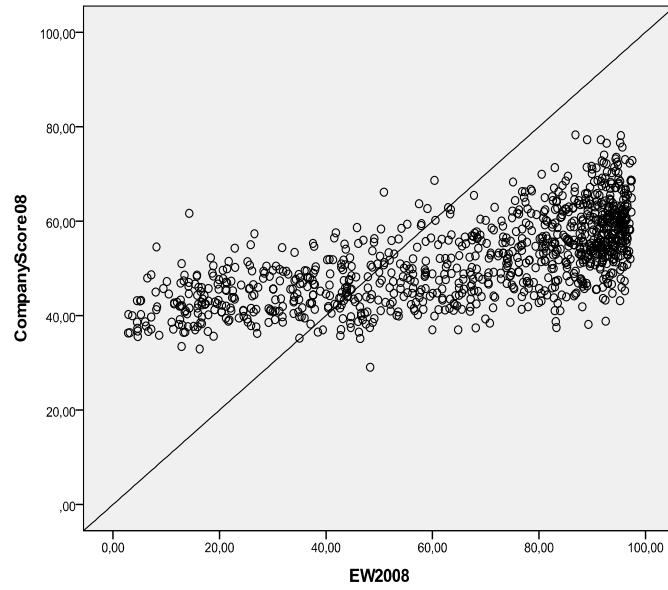
Figure 5
Overall 2009
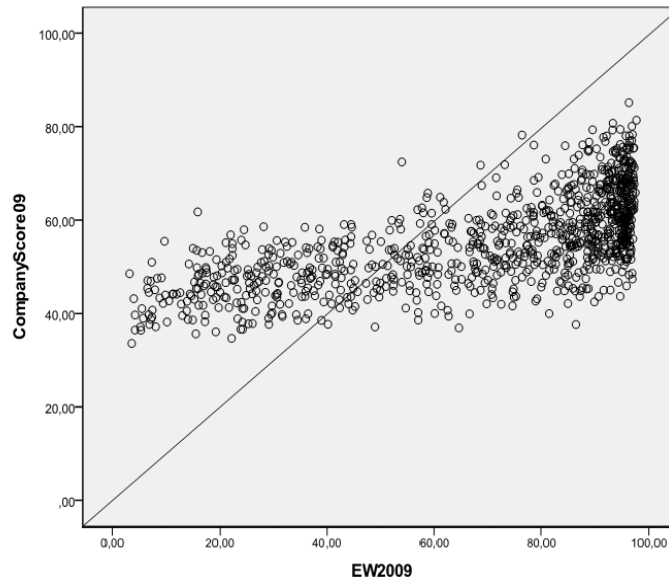
Figure 6
Overall 2010
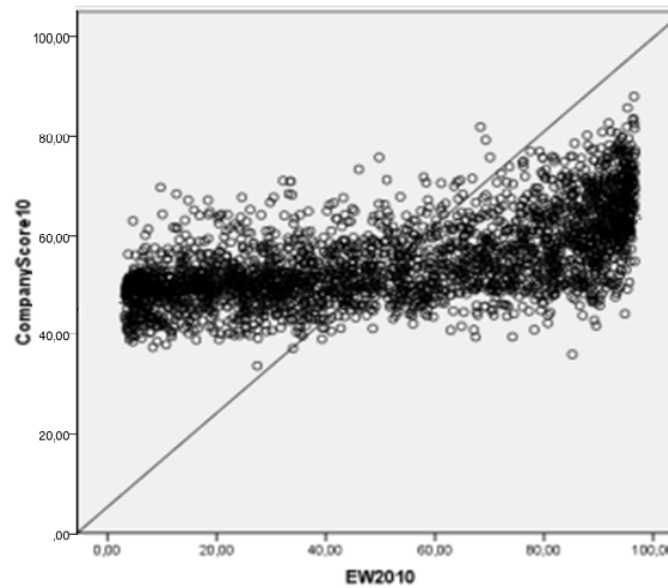
Table 6: Overall score correlations

		Unadjusted CSR score	Equal Weights 1/3	Sustainalytics Weights
2008	Pearson	0.664**	0.682**	0.703**
	R^2	0.441	0.465	0.495
2009	Pearson	0.657**	0.705**	0.712**
	R^2	0.432	0.497	0.512
2010	Pearson	0.634**	0.671**	0.748**
	R^2	0.402	0.450	0.559

**. Correlation is significant at the 0.01 level (2-tailed)

4.2 Sub-scores comparison

Both CSR ratings also provide sub-scores for the social, environmental and governance pillars. The next step of the analysis involves comparing these scores to see if there are differences between the sub-scores in the level of correlations and R^2's. Table 7 provides an overview of the results.

Table 7: Sub-score correlations

		Social	Governance	Environmental
2008	Pearson	0.617**	0.376**	0.648**
	R^2	0.380	0.141	0.420
2009	Pearson	0.627**	0.449**	0.640**
	R^2	0.393	0.201	0.409
2010	Pearson	0.474**	0.452**	0.639**
	R^2	0.225	0.204	0.408

**. Correlation is significant at the 0.01 level (2-tailed)

Table 7 shows that the Sustainalytics and ASSET4 ratings show high correlations for the social and environmental sub-scores, but for governance pillar the correlation coefficients are much lower. Furthermore, the environmental results remain stable over the three measured years while for the social and governance pillar there are fluctuations. Especially in 2010 the correlation for the social score drops significantly. Why this is the case remains unclear.

To see how large the impact is of the governance factor in distorting the similarities between the social and environmental pillars a new overall CSR score is created based solely on these two sub-scores, both with an equal weight of a half attached to them. The results can be found in table 8 under the column equal weights 1/2.

Table 8: Overall score correlations (2)

		Unadjusted CSR score	Equal Weights 1/3	Equal Weights 1/2	Sustainalytics Weights
2008	Pearson	0.664**	0.682**	0.728**	0.703**
	R^2	0.441	0.465	0.530	0.495
2009	Pearson	0.657**	0.705**	0.715**	0.712**
	R^2	0.432	0.497	0.512	0.512
2010	Pearson	0.634**	0.671**	0.638**	0.748**
	R^2	0.402	0.450	0.407	0.559

**. Correlation is significant at the 0.01 level (2-tailed)

For 2008 and 2009 there is an improvement compared to the 1/3 equal weights case in correlation but for 2010 the correlation goes down, so still no answer can be given to the question if the governance pillar is an interfering factor or not.

4.3 Categories

Figures 2 and 3 showed earlier that each pillar consists of a number of categories. Table 9 provides an overview of categories provided by ASSET4 and Sustainalytics.[5]

Table 9: Categories

Social		Governance		Environmental	
ASSET4	**Sustainalytics**	**ASSET4**	**Sustainalytics**	**ASSET4**	**Sustainalytics**
Training & Development	Employees	Board structure	Business ethics	Resource reduction	Contractors & Supply Chain
Health & safety	Contractors & Supply chain	Compensation policy	Corporate Governance	Emissions Reduction	Operations
Product Responsibility	Customers	Board Functions		Product innovation	Products & Services
Community	Society & community	Shareholder rights			
Human rights	Philantropy	Vision & strategy			
Employment quality					
Diversity					

Table 9 shows that, although there are a few good matches in categories, for the most it is not possible to make a good match between the categories. Furthermore, there is a difference in the number of categories used and for these two reasons this layer will not be analyzed.

---

[5] The economic pillar of ASSET4 is left out due to the fact that Sustainalytics does not use a similar sub-score.

The next step thus will be comparing the individual indicators of the two ratings, matching them. After this the construction of the new sub- and overall scores can be attempted.

4.4 Indicators

The most comprehensive step in this research involves looking at the individual indicators of both ASSET4 and Sustainalytics and trying to match them in an attempt to improve the theoretical coherence between the two rating basis and to test whether this improves the empirical convergence. Before we make a comparison of the content of individual indicators at the disaggregate level, we first explain in more detail the methodology that ASSET4 and Sustainalytics employ at this level. Next we will search for matches between the sets of indicators.

4.4.1 General differences

The Sustainalytics indicator list was the starting point for this part of the research. For each Sustainalytics indicator an as good as possible match was looked for in the ASSET4 database based on the definitions provided. The end result of this process can be found in appendices 2 to 12, which provides an overview of the matched indicators from both ratings. As a guide table 10 shows an example of an indicator match. In bold are the indicator codes as they can be found in the actual datasets and in italics the name of the indicator. Whereas Sustainalytics uses mostly a score between 0-100, for ASSET4 it can be a yes/no, true/false, a percentage or also a score between 0-100, which they define as follows: "*A "Score" is a number between 0 and 100 showing how the company performs compared to the entire ASSET4 universe based on the "Value" in the related indicator*". For Sustainalytics however the score can only be one of several options from which there is no deviation possible. For some variables the rationale for the match is obvious, while for others it is not clear at a first glance. Therefore, in the third column a short rationale is provided why the ASSET4 indicator is matched with the Sustainalytics indicator. In the same column you see how the ASSET4 values are transformed. In Table 10 the ASSET4 value of Yes is transformed in the Sustainalytics score of 100 and no to 0. A certain degree of arbitrariness is included in the transformation process due to the limitation of available background information on the side of ASSET4.

Table 10: Example

Sustainalytics indicator	ASSET4 indicator	Rationale for match	Pearson Correlations		
			2008	**2009**	**2010**
G.1.1 *Policy on Bribery and Corruption*	**SOCODP0017**: *Community Reputation Policy Elements/Bribery and Corruption*: Yes/No	Clear Match: Yes=100 No=0	0,535[**] N=962	0,549[**] N=54	0,547[**] N=968

In the far right column Pearson correlations of each individual pair of indicators for the years 2008, 2009 and 2010 can be found. There are large deviations in these correlations which could be the result of several elements. First, the matches of definitions may not be perfect. Although some definitions resemble each other greatly, in some cases the indicator was chosen that came closest. Although the definitions in these cases could resemble each other reasonably the underlying data could be completely different and thus lead to low correlations. Second, the transformation process itself can create low correlations. When ASSET4 uses a yes/no approach and Sustainalytics uses 5 possibilities of scores between 0 and 100 , information will be lost when trying to convert one into the other. In general the correlations are much better when ASSET4 score indicators are used because these required no transformation.

Besides the correlations also the number of overlapping observations for the three years is given. For the generic Sustainalytics indicators, there are many overlapping firms with numbers between 900 and 1100 for 2008 and 2009 and over 1500 for 2010. For industry specific indicators, there are much less overlapping firms.

Sustainalytics has for every category an indicator on controversies and incidents where none of both provides the maximum score of 100. They make a distinction between 5 levels of seriousness of the controversies or incidents. ASSET4 does have information on controversies but only provides the amount of incidents that a business is linked to. No additional information on the size is provided so there is no reasonable transformation possible and the choice is made to leave these indicators out.

4.4.2 Matched Governance indicators

Out of the 31 Sustainalytics governance indicators 21 ASSET4 matches were found (appendices 2-4). The first thing that is noticed is that by far not all of the matches of the Sustainalytics governance indicators originate from the corporate governance pillar of ASSET4. For example, for the category business ethics only one indicator from ASSET4 comes from the corporate governance pillar which shows that the two ratings do not agree on indicator placement. Most of the matches fall under the ASSET4 social pillar. It is not up to this research to decide if an indicator on bribery and corruption should fall under the governance or social pillar but it does show that part of the differences in correlations on sub-indicators could stem from the placement of indicators.

By far the best correlations of the 65 total matches can be found in the two indicators that both ASSET4 and Sustainalytics answer in the same form.Indicators G.3 and G.1.3.3 are answered in a yes/no format by both raters. The highest correlation equals 0,926 which can be interpreted as an extremely good match.

As mentioned a low number of overlapping observations can stem from the fact that the indicator is industry or sector specific. The last three indicators in this category deal with clinical trials, animal welfare and animal testing and clearly apply solely to the pharmaceutical industry. Indicators for which no match in the ASSET4 dataset could be found are on genetic engineering, tax transparency and money laundering (for a complete list see appendix 13)

In contrast to the business ethics category, ASSET4 and Sustainalytics do agree on the placement of the indicators in the corporate governance category. All indicators under the Sustainalytics corporate governance category originate from the ASSET4 corporate governance pillar. There is quite some agreement between the definitions of the matching indicators but this is not seen in the correlations despite of the fact that the number of observations is high in all except one case. This leads us to the conclusion that they try measure the same theoretical CSR element but in a different way. Indicators for which no match was found are about the disclosure of directors' biographies, in house teams dedicated to responsible investment and an oversight of ESG issues. These types of indicators illustrate the Sustainalytics approach by looking at processes and indirect indicators of CSR performance within a company instead of looking at objective numbers.

Lastly in the public policy category the first indicator has reasonable correlations even though the ASSET4 indicator is a yes/no indicator, but the second has extremely low non-significant ones.

4.4.3 Social

Although ASSET4 and Sustainalytics do agree on the fact that the indicators belong in the social pillar there are only 16 matches out of the 54 possible. Only for the first category a large number of matches are identified, but for the remaining categories matchings are scarce (appendix 5-9). This is not what was expected after the initial correlations on the overall scores and sub-scores. From these analyses it was found that the environmental and the social pillar had better correlations than the governance one so that one would expect more agreement on the underlying indicators.

In the social category employees (appendix 5) only the correlations of the first two pairs of indicators are good. What is striking is that an indicator like employee turnover rate produces relatively low correlations, around .450, while it would seem that it is a simple number that can be looked up in the books, which leaves no space for any subjective elements. The indicator S.1.4 on trade union representation is a similar objective number and this one produces significantly better results with significant correlations of .726. This research can provide no answers as why these differences occur but it does illustrate that even at the simplest level indicators that say to measure the same thing can come up with significantly different outcomes. Indicators for which no match were found include formal policy on working conditions, formal policies on the elimination of discrimination and programs to increase workforce diversity (Appendix 13).

Out of the nine indicators in the Sustainalytics category contractors & supply chain (appendix 6) only one match could be made which did provide a good correlation with a large number of overlapping observations. Sustainalytics thoroughly tries to analyze the supply chain system by looking at quality of supply chain standards, supply chain monitoring systems, health

& safety objectives with respect to the supply chain and supply chain audits where ASSET4 takes no explicit interest in this.

Just like in the contractors & supply chain category the customer category (appendix 7) only provided one match. The lack of matches stems from the fact that the Sustainalytics indicators are sector specific. They include policy statements on data privacy for software companies, outsourcing of core editorial tasks for the publishing industry and delayed flights for the aviation. It seems that ASSET4 tries to have one general formula to calculate its scores while Sustainalytics applies different formulas for different industries.

In the category society & community (appendix 8) there is again only one clear match, with a relative high correlation. All the other correlations are not very convincing due to either the transformation process or the non-perfect match of indicator definitions. Also the number of overlapping observations is not very high due to the industry specific nature of the indicators.

In the philanthropy category it is striking again that with such a simple indicator the correlation does not even rise above 0.700. In this case it is a yes/no indicator in which the question is asked if a company has a corporate foundation or not. When looking at the definitions a possible explanation is that some companies do have corporate foundations but these do not provide funding of any kind.

4.4.4 Environmental

For the total environmental pillar there were 26 indicator matches out of the 54 possible (appendix 10-12). Just as with the social pillar we expect higher levels of correlations than in the governance pillar since the correlations were significantly higher when comparing the sub-scores. Again there are large differences in the correlation levels within each category. For the operations (appendix 10) category almost half of the matches provide significant correlations with the two highest both relating to the Environmental Management System with correlation around 0.600. For Sustainalytics indicators on programs & targets of some sort the correlations drop below 0.300 which again illustrates the difference in approaches.

Just as before we only find a small amount of matches when we look at the Sustainalytics contractors & supply chain (appendix 11) category. There is only one indicator with a large number of overlapping observations, which is on environmental supply chain management, and this provides good correlations between 0.347 and 0.425. The other two indicators are apparently too specific as shown by the low number of overlapping observations and the non-significant results.

The final category on products & services (appendix 12) provided six matches with good results. The numbers of observations are not very high leading us to conclude that the indicators are sector/industry specific or just very specific in general. The matched indicators cover subjects like genetically modified organisms, investment in sustainable buildings, environmental considerations in R&D and automobile fleet CO_2 emissions(for car producers).

4.5 New sub-scores and overall scores.

In order to create new sub-scores and overall scores first a selection needed to be made which indicators would be used and which ones would be dropped. The critical factor here is if the number of overlapping observations is high enough. Because not for every indicator there is information on each firm, there will be a loss in firms when trying to apply a single formula to calculate the new scores and thus the starting number of firms must be high enough to compensate for this. First for each pillar a new score will be created and after this the new overall scores can be calculated and presented.

4.5.1 Governance

Since there are large deviations in the number of usable matches within the different categories it is not feasible to apply the Sustainalytics weights. If we would use their weight setup some indicators would get a relatively large weight if they were the only one left in a category while others would have to "share" a category. That is why we make the choice not to use the categories and instead give each indicator an equal weight in calculating the new pillar sub-score. What we end with is a simple formula that applies to all companies, for which we know that each indicator has the same weight attached to it. After dropping the obvious mismatches on the list due to the low amount of matching observations the following variables presented in their Sustainalytics form with their respective number of matching observations are included in the calculations for the new governance sub-score:

Table 11: Governance indicators used

Sus. Indicator name	# observations 2008	# observations 2009	# observations 2010
G.1.1 Policy on bribery and corruption	962	954	968
G.1.3 Signatory to UN global Compact	962	964	968
G.2.1 CSR reporting Quality	962	1038	2904
G.2.2 External verification of CSR reporting	1023	1001	1922
G.2.3 Disclosure of Directors'Remunerations	962	947	1743
G.2.6 Executive Compenstation Tied to ESG Performance	962	947	1743
G.2.8 Separation of board Chair and CEO roles	962	980	2904
G.2.12 Compensation Committee independence	846	835	1380
G.3.1 Policy on Political involvement and contributions	962	947	1743

For both Sustainalytics and ASSET4 all indicators will be given a weight of 1/9 after which the following correlations for the governance sub-scores were created:

Table 12: New governance sub-score correlations

Governance	2008	2009	2010
Pearson correlations	.660[**]	.680[**]	.728[**]
R^2	.436	.462	.530
N	164	125	151

What is striking is the low number of observations we are left with after putting all these indicators together due to the lack of information in the ASSET4 dataset. When we look back at the correlation of the original governance scores in table 3 a significant improvement can be seen. Where in the original analyses the R^2 did not get higher than .204 we now see levels that are more than twice as high. This means that after the series of transformations and reductions the strength of the relations and the predictive powers of the variables, e.g. the two governance sub-scores, have increased significantly.

It should be noted, however, that selection effects can bias the results. If we calculate the correlations of the original governance sub-scores using only the companies that are also used in table 12 we can test if the selection effects have an impact. The results are shown in Table 13.

Table 13: Selection effects

Governance	2008	2009	2010
Pearson correlations	.347[**]	.266[**]	.474[**]
R^2	.120	.172	.224
N	164	125	151

When we compare this to the correlations of the complete set of overlapping companies the results show us that selection effects barely have influence. For 2009 the deviation is noticeable different but for the other two years they come very close. The correlations were .376, .449 and .452 in the complete set comparison respectively.

4.5.2 Environmental

For the environmental pillar the following indicators were selected to calculate the new environmental pillar sub-score:

Table 14: Environmental indicators used

Sus. Indicator name	# observations 2008	# observations 2009	# observations 2010
E.1.2 Environmental Management System	862	980	2904
E.1.3.2 Programs & Targets to reduce hazardous waste generation	314	322	969
E.1.3.4 Programs & Targets to reduce water use	343	348	1112
E.1.7 Programs and Targets to reduce direct GHG emissions	962	980	2904
E.2.1.1 Programs and targets for environmental improvement of suppliers	403	393	733

Compared to the previous table on selected governance indicators the number of observations fluctuates much more between different indicators. For both Sustainalytics and ASSET4 all indicators will be given a weight of 1/5 after which the following correlations for the environmental pillar were calculated.

Table 15: New environmental sub-score correlations

Environmental	2008	2009	2010
Pearson correlations	.537[**]	.516[**]	.621[**]
R^2	.289	.266	.385
N	170	166	328

Just as in the governance pillar there is a large information loss due to the lack of overlap in data for the companies. Compared to the new governance sub-scores the correlations and the R^2's are lower where at the beginning of this research the original correlations were much better for the environmental pillar. Also compared to the original environmental sub-scores correlations reported in table 3 they are lower. This again could be due to selection effects and in order to check this Table 16 presents the correlation results for the original environmental sub-scores only for the companies that we used in table 15.

Table 16: Selection effects (2)

Environmental	2008	2009	2010
Pearson correlations	.403[**]	.468[**]	.622[**]
R^2	.162	.219	.387
N	170	166	328

The differences with table 6 are larger than was the case with the governance pillar. Especially for 2008 and 2009 we see that the correlations and R^2's of the original sub-scores for the smaller sample of companies are much lower than was the case for the complete set. Hence, we can conclude that the selection effects play an important role in this case. If we take account of these, comparison of Table 15 and Table 16 shows that the greater alignment of the content of the environmental sub-index of Sustainalytics and ASSET4 improves the empirical relationship between the environmental ratings significantly.

4.5.3 Social

The same approach is used for the social pillar. The following variables with their respective number of observations were selected to calculate the new social pillar subs-core:

Table 17: Social indicators used

Sus. Indicator name	# observations 2008	# observations 2009	# observations 2010
S.1.1 Policy on Freedom of association	962	956	2017
S.1.5 Employee turnover rate	1023	1001	1922
S.2.1 Scope of social supply chain standards	962	980	2904
S.5.2 Corporate foundation	962	947	1743

For both Sustainalytics and ASSET4 all indicators will be given a weight of 1/4 after which the following correlations for the social pillar were created:

Table 18: New social sub-score correlations

Social	2008	2009	2010
Pearson correlations	.759[**]	.753[**]	.744[**]
R^2	.576	.567	.553
N	962	947	1743

The number of overlapping observations is much higher than in the governance and environmental pillar and this is accompanied by higher correlations and R^2's. The high number of observations makes the selection effects less important and the results more trustworthy. There are, however, only four variables that were selected so it could simply again be random e.g. the companies match very well on these four indicators but when more are added the results diverge again. Still to check for selection effects as with the other two pillars we correlate the original social sub-scores only using the companies we used in table 18.

Table 19: Selection effects (3)

Environmental	2008	2009	2010
Pearson correlations	.492[**]	.520[**]	.492[**]
R^2	.242	.270	.242
N	962	947	1743

If we compare the results with the original correlations from table 6 (with correlation coefficients of .617, .627 and .474 respectively), we see that these correlations are lower than for the complete set for 2008 and 2009, Hence, again selection effects do play a role and comparison of Table 18 and 19 shows that the increase in correlation when calculating the new sub-scores is even better. For 2010 the results are similar so there is no selection effect here, which means that the smaller sample represents the set accurately.

4.5.4 Overall scores

From the previously calculated new sub-scores new overall scores were calculated. All three pillars were given an equal weight and from this the following correlations emerged:

Table 20: New overall score correlations

Overall	2008	2009	2010
Pearson correlations	.614[**]	.711[**]	.837[**]
R^2	.377	.506	.700
N	32	25	25

These results mean that for the year 2008 there were for only 32 companies data on all of the 18 indicators in both ASSET4 and Sustainalytics. Although for 2008 and 2009 the results are not very different from the original results that can be found in table 4 the correlation and the R^2 for 2010 are the highest found which suggests that the approach taken is a step in the right direction. But the extreme low number of observations cannot be ignored and future research will have to take this into account by perhaps using a larger number of formulas instead of the one size fits all approach used in this research. Just as with the sub-scores we check for selection effects. Table 21 presents the correlations of the original overall scores for the companies that we were left with of table 20. These numbers strongly confirm the large increase in correlations after the reconstruction of the scores due to the higher theoretical coherence.

Table 21: Selection effects (4)

Overall	2008	2009	2010
Pearson correlations	.123	.382	.386
R^2	.015	.146	.149
N	32	25	25

4.5.5 Preliminary conclusions

In the chapter 2 in the section on external verification we have seen what levels of correlation academic papers have used to draw conclusions upon as to when two elements correlate sufficiently. For instance, Scharfman (1999) found correlations between 0.18 and 0.55 while Mackenzie and Rees (2011) found correlations between 0.23 and 0.57. Both conclude that the correlations are sufficient. When looking at the original overall scores, the original sub-scores and all the reconstructed scores we see significant higher correlations with few exceptions. Especially for the governance sub-scores correlations substantially improved after aligning the indicators at the disaggregate level. We have seen that this cannot be the result of selection effects.

Based on the correlations found and the academic papers discussed we can only conclude that the two ratings correlate more than sufficient and that they try to measure, at least partially, a part of the same construct. This is confirmed after looking at the correlations of the reconstructed overall and sub-scores.

4.5.6 External verification

Up until now we have only been looking at the two ratings compared to each other. In order to provide an estimation which of the two is a better predictor of CSR performance we apply an approach comparable to Sharfman (1996) on the construct validity of the KLD ratings data. They use an external source, Fortune data, to test the validity of the data. In this research we pick certain objective outcome indicators from the ASSET4 data system and we correlate these against indicators from the two ratings with a similar content. In this part of the research we use partial regressions controlling for region, sector and for company size. The first two are controlled for by using dummies and the latter by using the ln of revenues. Furthermore we take the averages of all three years to take into account the time effect. The complete results of these analyses can be found in table 22.

At a first glance it would appear that ASSET4 scores better since it has significant results in 6 out of 10 outcome variables while Sustainalytics only has 3, including employee fatalities with a significance of 0.007. A critical note here is that the data for the outcome variables was extracted from the ASSET4 database. Chances are that these outcome variables are incorporated in some manner in the indicators used, either subjectively or objectively. If we for instance look at table 21 it is logical to argue that the outcome variable *total injury rate employees* is incorporated in the ASSET4 indicator *ASSET4 score health&safety&injuries.* With this said Sustainalytics does a good job in keeping up with ASSET4 at the variables where both are significant. For this reason, we will not go as far as proclaiming ASSET4 the winner. When taking the above mentioned elements into consideration Sustainalytics also does a good job in explaining some of the outcome indicators and future research with true objective outcome variables obtained from an independent source will have to decide which one is better.

Table 22: Outcome variable comparison

| | ASSET4 | | Sustainalytics | |
Outcome	Description	Correlation	Description	Correlation
Co2 equivalents emissions total[a]	Score Greenhouse Gas emissions	-.587***	Carbon Intensity	-.272***
C02 Equivalents emissions total[a]	Score Greenhouse Gas emissions	-.587***	Carbon Intensity trend	.030
Waste Total[b]	Score emission reduction / waste	-.642***	Waste Intensity	.041
Water Withdrawal[b]	Score resource reduction/water use	-.034	Water intensity	.086
Trade union representation[a]	Score employment quality	.985***	Percentage covered by CBA	.716***
Total injury rate employees[b]	Score health & safety &injuries	.972***	Trend in lost time incident rate	.080
Employee accidents[b]	Score health & safety & injuries	-.588***	Trend in lost time incident rate	.150
Employee fatalities[b]	Score health & safety & injuries	-.062	Number of fatalities	-.324**
Lost time injury rate employees[b]	Score health & safety / lost days	-.026	Trend in lost time incident rate	-.071
Employee lost working days[b]	Score health & safety / lost days	.026	Trend in lost time incident rate	-.071
[a] Controlled for: Company size (ln of revenues), region and sector				
[b] Controlled for: Company size and region				
*** significant <0.001				
** significant <0.05				

4.5.7 Transparency

Rating agencies in general take a position above industries and sectors. To deliver their product they require firms to cooperate and thus provide information, e.g. become more transparent. The irony is that the transparency of ratings agencies themselves is not all too clear. On the one hand this makes sense since they have the right to protect their unique "formulas", which are their sources of income.

However, it is relevant for this study what the processes are behind the creation of scores. For Sustainalytics this becomes clear by looking at the files provided. Although there is a level of subjectivity included a company should be able to recognize its score on different indicators. For ASSET4 this is a different story. From the information provided we can only assume what the steps are in the score creation process and what happens with underlying indicators. Especially the step from the yes/no questions to the creation of values and scores remains

unclear. We do not know if the weights attached to different indicators differ or even if some are left out or not. What this comes down to is: You do not know exactly what you are buying when you purchase access to the ASSET4 data set. The creation of the scores, their main source of income, remains a black box. Users and companies that are being rated just have to assume that this is a "fair" process and that these scores reflect the actual underlying values accurately. As said before scores are their unique selling point and with that their main source of income but this lack in transparency makes it difficult to rate the rater.

On the other hand, at the individual indicator level ASSET4 does better when it comes to transparency. By using yes/no answers it becomes immediately clear what is going on and there is no reason for doubt. By adding more subjective elements the level of transparency in the Sustainalytics ratings goes down for the user.

5. Conclusion

This research sets out to unravel the differences and similarities between the Thomson Reuters ASSET4 ESG and the Sustainalytics datasets. Both claim to measure Corporate Social Responsibility and provide assistance to investors by giving them a handle in the form of data and numbers.

The second chapter provides the theoretical framework of this research. First, we analyze how ratings influence investment decisions and the behavior of rated firms. Next, we researched papers that have done similar kinds of research in the past to see if a benchmark could be set as to what levels of correlations would be sufficient. For this purpose Scharfman (1999) and Mackenzie & Rees (2011) were very helpful. The first paper is the standard when it comes to measuring the construct validity of CSR ratings. The second paper tries to validate its choice for FTSE4Good, another CSR rating, by correlating it to the ASSET4 CSR data that this research also uses.

We applied a top-down approach when working through our analyses. This meant starting with the overall CSR scores, social, environmental and governance sub-scores as provided by the two ratings. We correlated these scores and calculated the R^2's so that we had our first results and benchmark for the rest of the research. The correlations of the overall scores varied between 0.634 and 0.664 while for the sub-scores there was a difference with on the one hand the social and environmental pillar and on the other hand the governance pillar. For the first two, correlations were found between 0.474 and 0.648, while for the governance this was between 0.376 and 0.452. There were however factors that interfered in making a fair comparison between the overall scores. ASSET4 uses a fourth pillar sub-score and equal weights, while Sustainalytics uses custom weights, which were defined by the industry a firm is in. After we corrected for these two differences, correlations between overall scores improved to levels between 0.671 and 0.682. Compared to the correlations found in the literature the scores are good to very good and as Scharfman would have put it: They correlate more than sufficiently. On the individual indicator level we still remain with many unmatched indicators which suggests that different companies put different elements under the CSR umbrella. This lack of agreement is not shown in the overall and sub-score correlations since these provide a strong indication that they try to measure at least part of the same construct.

For the construction of the pillar sub-scores both ratings use a number of categories. Because there was not enough overlap between these categories, the research moved one level lower to the level of individual indicators. On the basis of the Sustainalytics indicator list pairs of indicators were searched for. The ASSET4 indicator was then transformed into one that was comparable to the Sustainalytics partner. Due to the fact that ASSET4 uses different kinds of setups to answer questions, for instance yes/no or scores, and Sustainalytics solely uses scores, choices had to be made as to what value would be transformed in what manner. An ASSET4 score indicator was left unchanged, whereas ASSET4 indicators with a yes/no indicator were transformed into a 100/0 score. After the transformation each pair of indicators was correlated and the consequences of the transformation choices could be seen. The yes/no indicators score

much lower in general than the score indicators. This loss of information has to be taken into account when evaluating the following steps of the process.

New sub- and overall scores could now be created. Choices had to be made as to what indicators would be included in the calculations. The deciding factor was the number of observations. Because we use a single formula to calculate the new sub- and overall scores, and not one for every industry, we need information for every indicator included in the calculation or else the result would be a blank. The result is that sector and industry specific indicators were not included. The end result is that we used 18 indicators from the three pillars to calculate these new scores. We were left with 32, 25 and 25 companies for which we calculated the new scores for 2008, 2009 and 2010. The resulting significant correlations were .614, .711 and .837 for the three years. Although the number of firms is very low the highest correlation of a score was found in this last step. Although the answer is far from conclusive we can say that when the two rating companies use similar indicators, with similar scores and similar weights attached to them, they move closer towards each other and it seems they are measuring a part of the same construct, e.g. Corporate Social Responsibility.

The last part of the analysis tried to prove objectively which of the two ratings was "better" by taking outcome variables and correlating these to near-lying indicators from the two ratings and controlling for company size, region and industry. Although ASSET4 scored better it must be said that these outcome variables were taken from the ASSET4 dataset and the results are thus to be taken cautiously. Where both produced significant correlations Sustainalytics stayed relatively close to the ASSET 4 correlations and thus proving its strength against arguably unfair competition.

Concluding, our analysis shows that there is substantial convergence between Sustainalytics and ASSET4. The convergence improves if we reconstruct the ratings by aligning weights and indicators. However, although it may increase the confidence in the ratings of Sustainalytics and ASSET4 now it has been shown that both generate comparable results, we do not yet know for sure that they both give a good approximation of the corporate social responsibility of a company. Because it might be that both ratings do not completely reflect the concept of CSR. Even if the two ratings show reasonable similarity, this does not prove that there is a uniform CSR definition. Actually, we have seen that there are a lot of differences in the sets of indicators that Sustainalytics and ASSET4 are using. Still, overall the correlation is quite good. This high correlation therefore does not provide us with a definite idea how CSR should ideally be measured.

On the other hand, if we assume that there exists a concept of CSR that can be objectively measured, it seems encouraging that two methods that try to do so provide similar results. These similarities (and lack of similarities at some points) show that there is a larger (respectively lower) chance that these parts do (do not) constitute an element of CSR and that this element is objectively measurable. To increase this chance future research should encompass more than two ratings. If the same process would then be followed as this research has, a large step could be made towards finding common ground for measuring CSR.

6. Recommendations

The biggest problem of research into the robustness of ESG ratings as a measurement of CSR is still the lack of data in the ratings used and the treatment of these gaps in the data. As we have seen in this research the number of firms we were left with after the construction of the new overall CSR scores was very limited. What ESG rating bureaus could agree on is for instance how to treat companies on which there is no data available for indicators. Sustainalytics gives these companies a score of 0 while ASSET4 does not use these blank indicators in the calculation of their scores. If there is a large amount of these indicators it does manipulate the scores. There could be an industry standard on how to treat this and this would at least take this bias out of the comparison.

A related aspect, but more aimed at researchers instead of rating bureaus, that could result in better findings would be not to use a single formula to calculate new scores. In this research for the total group of companies indicators were chosen with a large number of overlapping observations. If for each separate industry you would do the matching of indicators and create new scores based on the number of observations per industry you would end up with a much larger amount of sub-scores and overall scores. This would encompass a great deal of work but it could be worth the effort.

Also we have seen that Sustainalyics changed their methods in 2010 which probably led to the deviating scores and correlations compared to 2008 and 2009. In upcoming years it has to be checked if these changes are structural or if there was something else going on this specific year.

7. References

Bagnoli, M., & Watts, S. G. (2003). Selling to Socially Responsible Consumers: Competition and The Private Provision of Public Goods. *Journal of Economics & Management Strategy , 12* (3), 419-445.

Banerjee, A. (1992). A Simple Model of Herd Behaviour. *Quarterly Journal of Economics , 107* (3), 797-817.

Baron, D. (2001). Private Politics, Corporate Social Responsibility and Integrated Strategy. *Journal of Economics and Management Strategy , 10* (1), 7-45.

Chatterji, A., & Levine, D. (2008). Imitate or Differentiate? Evaluating the Validity of Corporate Social Responsibility Ratings. *Working Paper Series* , 1-55.

Chatterji, A., & Toffel, M. W. (2010). How Firms Respond to Being Rated. *Strategic Management Journal , 31* (9), 917-945.

Chatterji, Levine, A. K., & Toffel, M. W. (2010). How Well do Social Ratings Actually Measure Corporate Social Responsibility? *Journal of Economics & Management Strategy , 18* (1), 125-169.

Chollet, P., & Cellier, A. (2011). The Impact of Corporate Social Responsibility Rating Announcements on European Stock Prices. *International Conference of the French Finance Association(AFFI).* Paris.

Dahlsrud, A. (2008). How Corporate Social Responsibility is Defined: An Analysis of 37 Definitions. *Corporate Social Responsibility and Environmental Management , 15* (1), 1-13.

Entine, J. (2003). The Myth of Social Investing: A Critique of its Practice and Consequences for Corporate Social Performance Research. *Organization & Environment , 16* (3), 352-368.

Graafland, J. J., Eijffinger, S. C., & Smid, H. (2004). Benchmarking of Corporate Social Responsibility: Methodological Problems and Robustness. *Journal of Business Ethics , 1-2* (53), 137-152.

Gregory, A., Whittaker, J., & Yan, X. (2008). Do Markets Value Corporate Social Responsibility? *Working Paper* .

Hamilton, J. T. (1995). Pollution as News: Media and Stock Market Reactions to the Toxic Release Inventory Data. *Journal of Environmental Economics and Management , 28* (1), 98-114.

Klassen, R. D., & McLaughlin, C. P. (1996). The Impact of Environmental Management on Firm Performance. *Management Science , 42* (8), 1199-1214.

Leb, B., & Petrovits, C. (2010). Is Doing Good Good for you? Yes, Charitable Contributions Enhance Revenue Growth. *Strategic Management Journal , 31* (2), 182-200.

Liston-Heyes, C., & Ceton, G. (2009). An Investigation of Real Versus Perceived CSP in S&P-500 Firms. *Journal of Business Ethics , 89* (2), 283-296.

Lyon, T. P., & Maxwell, J. W. (2011). Greenwash: Corporate Environmental Disclosure under Threat of Audit. *Journal of Economics & Management Strategy , 20* (1), 3-41.

Maxwell, J. W., Lyon, T. P., & Hackett, S. C. (2000). Self-Regulation and Social-Welfare: The Political Economy of Corporate Environmentalism. *Journal of Law and Economics , 43* (2), 583-617.

Meyer, M. W., & Gupta, V. (1994). The Performance Paradox. *Research in Organizational Behavior , 16*, 309-369.

Rees, B., & Mackenzie, C. (2011). Corporate Social Responsibility and the Open Society. *Working Paper* , 1-44.

Rosen, B. N., Sandler, D. M., & Shani, D. (1991). Social Issues and Socially Responsible Investment Behavior: A Preliminary Empirical Investigation. *Journal of Consumer Affairs , 25* (2), 221-234.

Scharfman, M. (1996). The Construct Validity of the KLD Performance Data Ratings. *Journal of Business Ethics , 15* (3), 287-297.

Scott, W. R., (1995). *Institutions and Organizations.* California: Sage Publications.

Staw, B. M., & Epstein, L. D. (2000). What Bandwagons Bring: Effects of Popular Management Techniques on Corporate Performance, Reputation, and CEO Pay. *Administrative Science Quarterly , 45* (3), 523-556.

Wilbert, C. (2006, July 19). Social Responsibility of Coke Questioned. *Atlanta Journal Constitution* , p. 4.

Williams, G. (2007). Some Determinants of the Socially Responsible Investment Decision: A Cross-Country Study. *Journal of Behavioral Finance , 8* (1), 43-57.

8. Appendices

Appendix 1:Three examples of Sustainalytics default weight matrix

Peer Groups	Food & Staples Retailing	Health Care Providers, Equipment & Services	Auto Components
Business Ethics	2	2	8
Corporate Governance	2	2	12
Public Policy	1	1	5
Governance	5	5	5
Employees	3	15	20
Contractors & Supply Chain	3	5	5
Customers	1	8	6
Society & Community	1	7	4
Philanthropy	1	5	5
Social	9	8	8
Operations	6	25	4
Contractors & Supply Chain	5	7	1
Products & Services	4	3	2
Environment	6	7	7
Total	**20**	**20**	**20**
Business Ethics	10,0%	10,0%	8,0%
Corporate Governance	10,0%	10,0%	12,0%
Public Policy	5,0%	5,0%	5,0%
Governance	25,0%	25,0%	25,0%
Employees	15,0%	15,0%	20,0%
Contractors & Supply Chain	15,0%	5,0%	5,0%
Customers	5,0%	8,0%	6,0%
Society & Community	5,0%	7,0%	4,0%
Philanthropy	5,0%	5,0%	5,0%
Social	45,0%	40,0%	40,0%
Operations	12,0%	25,0%	20,0%
Contractors & Supply Chain	10,0%	7,0%	5,0%
Products & Services	8,0%	3,0%	10,0%
Environment	30,0%	35,0%	35,0%
Total	**100%**	**100%**	**100%**

Appendix 2: Indicator comparison Pillar-Governance   Category-Business Ethics

Sustainalytics indicator	ASSET4 indicator	Rationale for match	Pearson Correlations		
			2008	**2009**	**2010**
G.1.1 *Policy on Bribery and Corruption*	**SOCODP0017**: *Community Reputation Policy Elements/Bribery and Corruption*: Yes/No	Clear Match: Yes=100 No=0	0.535** N=962	0.549** N=954	0.547** N=968
G.1.1.1 *Programs to Combat bribery and Corruption*	**SOCODP008**: *Bribery and Corruption Training*: Yes/No	Oke Match: Yes=100 No=0	0.252 N=33	0.381* N=33	0.252 N=33
G.1.2 *Whistleblower programs*	**SOCODP011**: *Whistleblower Protection*: Yes/No	Clear Match: Yes=100 No=0	0.254** N=767	0.311** N=778	0.235** N=780
G.1.3 *Signatory to UN Global Compact*	**CGVSDP020**: *Global Compact*: Yes/No	Clear Match: Yes=100 No=0	0.826** N=962	0.885** N=964	0.935** N=968
G.1.3.2 *Policy on Responsible Investment*	**SOPRDP052**: *SRI Socially Responsible Investments*: True/False	Good Match: Yes=100 No=0	0.454** N=169	0.558** N=163	0.415** N=179
G.1.3.3 *Member of UNEP Finance Initiative*	**ENERDP008**: *Emission Reduction UNEP Finance Initiative (FI) Statements*: Yes/No	Clear Match: Yes=100 No=0	0.896** N=169	0.926** N=171	0.742** N=171
G.1.3.4 *Membership in Initiatives Promoting Sustainable buildings*	**ENRRO07S**: *Score - Resource Reduction/Green Buildings*	Oke Match: 0-25= 0 25-75=50 75-100=100	0.896** N=169	0.558** N=163	0.415** N=179
G.1.3.5 *Equator Principles and related Reporting.*	**ENPIDP036**: *Equator Principles*: Yes/No	Clear Match: Yes=100 No=0	0.817** N=38	0.738** N=59	0.799** N=67
G.1.4.3 *Policy on Animal Testing*	**ENPIDP060**: *Animal Testing Reduction*: Yes/No.	Clear Match: Yes=100 No=0	0.428** N=161	0.372** N=156	0.612** N=448
G.1.4.4 *Policy on Animal Welfare*	**ENPIDP059**: *Animal Testing guidelines*: Yes/No.	Good Match: Yes=100 No=0	0.128 N=57	0.093 N=54	0.132 N=145
G.1.4.6 *Clinical Trial Protocals*	**SOPRO06S**: *Score - Product Responsibility/Clinical Trials and Research Guidelines*	Clear Match: No transformation needed.	0.520** N=31	0.687** N=35	0.460** N=88

Appendix 3: Indicator comparison Pillar-Governance   Category-Corporate Governance

Sustainalytics indicator	ASSET4 indicator	Rationale for match	Pearson Correlations		
			2008	2009	2010
G.2.1 *CSR Reporting Quality*	**CGVSO06S:** *Score - Vision and Strategy/GRI Report*	Clear Match: No transformation needed	0.417** N=962	0.642** N=1038	0.607** N=2904
G.2.2 *External Verification of CSR Reporting*	**CGVSO08S:** *Score - Vision and Strategy/CSR Reporting Auditor*	Clear Match: No transformation needed	0.707** N=1023	0.687** N=1001	0.639** N=1922
G.2.3 *Disclosure of Directors' Remuneration*	**CGCPD01S:** *Score - Compensation Policy/Policy*	Clear Match: No transformation needed	0.287** N=962	0.395** N=947	0.591** N=1743
G.2.6 *Executive Compensation Tied to ESG Performance*	**CGCPO09S:** *Score - Compensation Policy/Sustainability Compensation Incentives*	Clear Match: No transformation needed	0.254** N=962	0.310** N=947	0.350** N=1743
G.2.7 *Board Diversity*	**CGBSO03S:** *Score - Board Structure/Board Diversity*	Oke Match: No transformation needed.	0.173** N=810	0.205** N=794	0.238** N=1286
G.2.8 *Separation of Board Chair and CEO Roles*	**CGBSO09S:** *Score - Board Structure/CEO-Chairman Separation*	Clear Match: No transformation needed.	0.708** N=962	0.728** N=980	0.648** N=2904
G.2.10 *Audit Committee Independence*	**CGBFDP018:** *Audit Committee Independence:* Percentage.	Good match but we only know the percentage so we take that as the score.	0.454** N=169	0.558** N=163	0.415** N=179
G.2.12 *Compensation Committee Independence*	**CGBFDP020:** *Compensation Committee Independence:* Percentage.	Good match but we only know the percentage so we take that as the score.	0.223** N=846	0.387** N=835	0.525** N=1380

Appendix 4: Indicator comparison Pillar-Governance   Category-Public Policy

Sustainalytics indicator	ASSET4 indicator	Rationale for match	Pearson Correlations		
			2008	**2009**	**2010**
G.3.1 *Policy on Political Involvement and Contributions*	**SOCODP0018:** *Community Reputation Policy Elements/Political contribution:* Yes/No.	Clear Match: Yes=100 No=0	0.471[**] N=962	0.454[**] N=947	0.482[**] N=1743
G.3.2 *Total Value of Political Contributions*	**SOCODP035:** *Lobbying Political Contributions:* Number.	Good Match except that ASSET4 is for one year.So to compensate the numbers of the Sustainalytics are divided by 3: 0-166.666=50 0=100 >166.000=0	0.009 N=1023	0.104[**] N=1001	0.033[**] N=1922

Appendix 5: Indicator comparison Pillar-Social   Category-Employees

Sustainalytics indicator	ASSET4 indicator	Rationale for match	Pearson Correlations		
			2008	**2009**	**2010**
S.1.1 *Policy on Freedom of Association*	**SOHRDP0011:** *Human Rights Policy Elements/Freedom of Association:* Yes/No	Good Match: Yes=100 No=0	0.721[**] N=962	0.680[**] N=956	0.605[**] N=2017
S.1.4 *Percentage of Employees Covered by Collective Bargaining Agreements*	**SOEQDP031:** *Trade Union Representation:* Percentage.	Clear Match: We use the percentage as the score.	0.677[**] N=184	0.695[**] N=303	0.726[**] N=565
S.1.5 *Employee Turnover Rate*	**SOEQDP034:** *Turnover of Employees:* Percentage.	Clear Match: No data=0 >10%=0 5%-10%=50 <5%=100	0.180[**] N=1023	0.469[**] N=1001	0.448[**] N=1922
S.1.6 *Top Employer*	**SOEQDP028**: *Employment Awards*: Number.	Three or more awards are proxies for being on an employer list: 0=0 1-3=75 >3=100	0.224[**] N=228	0.311[**] N=189	0.131[**] N=369
S.1.6.1 *Employee* trainings days per employee.	**SOTDDP018:** *Average Training Hours*: Number.	Good Match. ASSET4 indicator needs to be divided by 8(working hours in a day): 0=0 <5=0 5-10=50 10-15=75	1.0[**] N=2	0.612 N=5	0.771[**] N=12

			>15=100			
S.1.6.2 *Programmes and Targets to Reduce Health and Safety Incidents*	**SOHSD01S:** *Score - Health & Safety /Policy*	Clear Match: No transformation needed.	0.193** N=365	0.175** N=374	0.234** N=1149
S.1.6.3 *Programmes to Adress HIV/Aids Among its Workforce*	**SOHSO03S:** *Score - Health & Safety/HIV-AIDS Programme.*	Clear : No transformation needed.	0.451 N=14	0.02 N=11	0.66 N=26
S.1.6.6 *Number of Fatalities*	**SOHSO01S:** *Score - Health & Safety/Injuries*	Not a clear match but the ASSET4 indicator can be a good proxy: No transformation needed	-0.342 N=5	-0.074 N=169	-0.103 N=296

Appendix 6: Indicator comparison Pillar-Social   Category-Contractors &   Supply Chain

Sustainalytics indicator	ASSET4 indicator	Rationale for match	Pearson Correlations		
			2008	**2009**	**2010**
S.2.1 *Scope of Social Supply Chain Standards*	**SOHRDP0014:** *Human Rights Policy Elements/Supply Chain Human Rights:* Yes/No	Not a clear match: Yes=75 No=0	0.657** N=962	0.665** N=980	0.649** N=2904

Appendix 7: Indicator comparison Pillar-Social   Category-Customers

Sustainalytics indicator	ASSET4 indicator	Rationale for match	Pearson Correlations		
			2008	**2009**	**2010**
S.3.1.1 *Pubic Position Statement on Responsible Marketing*	**SOPRDP0016:** *Product Responsibility Policy Elements/Responsible Marketing:* Yes/No	Clear Match: Yes=100 No=0	0.551** N=137	0.502** N=130	0.476** N=202

Appendix 8: Indicator comparison Pillar-Social   Category-Society & Community

Sustainalytics indicator	ASSET4 indicator	Rationale for match	Pearson Correlations		
			2008	**2009**	**2010**
S.4.2.1 *Policy on Human*	**SOHRD01S**: *Score - Human Rights/Policy*	Clear Match: No transformation needed.	0.578** N=224	0.575** N=228	0.520** N=731
S.4.2.2 *Community Involvement Programs*	**SOCOD01S:** *Score - Community/Policy*	Oke match. Being a good corporate citizen is a good proxy of being involved in a community: No transformation	0.249** N=292	0.255** N=298	0.205** N=973

			needed.			
S.4.2.9 *Policy on Indigenous People and Land Rights*	**SOCODP0064:** *Community Reputation Code of Conduct/Indigenous People*: Yes/No.	Good match: Yes=100 No=0		0.328** N=82	0.389** N=83	0.158** N=330
S.4.2.12 *Programmes to Address Digital Divide*	**SOPRDP025:** *Product Access Low Price*: Yes/No.	Oke Match: Yes=100 No=0		0.400** N=87	0.470** N=88	0.164** N=271

Appendix 9: Indicator comparison Pillar-Social   Category-Philantropy

Sustainalytics indicator	ASSET4 indicator	Rationale for match	Pearson Correlations		
			2008	**2009**	**2010**
S.5.2 *Corporate Foundation*	**SOCODP038:** *Company Foundation*: Yes/No.	Clear Match: Yes=100 No=0	0.658** N=962	0.691** N=947	0.684** N=1743

Appendix 10: Indicator comparison Pillar-Environmental   Category-Operations

Sustainalytics indicator	ASSET4 indicator	Rationale for match	Pearson Correlations		
			2008	**2009**	**2010**
E.1.2 *Environmental Management System*	**ENERDP0013**: *Emission Reduction Policy Elements/Environmental Management Systems*: Yes/No	Clear Match: Yes=100 No=0	0.609** N=962	0.643** N=980	0.555** N=2904
E.1.2.1 *Programs and Targets to Protect Biodiversity*	**ENERDP0162:** *Emission Reduction Objectives /Biodiversity*: Yes/No.	Good match: Yes=100 No=0	No data in ASSET4		
E.1.2.6 *Waste Intensity*	**ENERO10S:** *Score - Emission Reduction/Waste*	Good match: No transformation needed	-0.066 N=187	-0.175* N=83	-0.185** N=207
E.1.2.7 *Water Intensity*	**ENERO13S:** *Score - Emission Reduction/Discharge into Water System*	Water pollutant is an element of external cost of water-related impacts so it is taken as an estimator.	-0.153 N=67	0.052 N=83	-0.211* N=330
E.1.2.8 *Percentage of Certified Forests Under Own Management*	**ENPIDP043**: *Labeled Wood Percentage*	Clear Match: 0=0 0%-25%=25 25%-50%=50 50%-75%=75	For the 3 companies that matched Sustainalytics gives the same score so correlation is not applicable.		

		75%-100%=100			
E.1.3 *External Certification of EMS*	**ENERDP075**: *Environmental Management System Certified Percent*:	Clear Match: 0=0 0%-50%=25 50%-75%=50 75%-90%=75 >90%=100	0.571** N=133	0.649** N=187	0.572** N=308
E.1.3.2 *Programs & Targets to Reduces Hazardous Waste Generation*	**ENERO14S**: *Score - Emission Reduction/Waste Reduction*	Clear Match: No transformation needed.	0.272** N=314	0.253** N=322	0.284** N=969
E.1.3.3 *Programs & Targets to Reduces Air Emissions*	**ENERDP039**: *HAP Emissions Reduction*: Yes/No.	HAP's are not completely equal to air emissions but as an approximation: Yes=100 No=0	0.193** N=255	0.224** N=260	0.234** N=882
E.1.3.4 *Programs &Targets to Reduces Water Use*	**ENERDP066**: *Wastewater Reduction*: Yes/No	Clear Match: Yes=100 No=0	0.221** N=343	0.223** N=348	0.272** N=1112
E.1.4 *Environmental Fines and Non-monetary Sanctions*	**ENERDP090**: *Environmental Controversy Costs*: Number.	The ASSET4 number is for one year so we multiply it by 3 to create an comparable indicator: >400.000=0 200.000-400.000=25 10.000-200.000=50 0-10.000=75 0=100	0.440** N=121	0.293 N=44	0.421** N=69
E.1.7 *Programs and Targets to Reduce Direct GHG emissions*	**ENERO05S**: *Score - Emission Reduction/CO2 Reduction*	CO_2 is only one of several Green House Gasses but it is a good indicator of the intensions of a company: No transformation needed	0.431** N=962	0.439** N=980	0.441** N=2904
E.1.8 *Programs and Targets to Increase Renewable Energy*	**ENRRO06S**: *Score - Resource Reduction/Renewable Energy Use*	Oke match: No transformation needed.	0.075 N=109	0.115 N=124	-0.070 N=189
E.1.11 *% Primary Energy Use from Renewables*	**ENRRO06S**: *Value - Resource Reduction/Renewable Energy Use*	Clear Match: No transformation needed.	0.252** N=109	0.358** N=124	0.223** N=189

Appendix 11: Indicator Comparison Pillar – Environmental Category - Contractors & Supply Chain

Sustainalytics indicator	ASSET4 indicator	Rationale for match	Pearson Correlations		
			2008	2009	2010
E.2.1.1 *Programs and Targets for Environmental Improvement of Suppliers*	**ENRRO11S**: *Score - Resource Reduction/Environmental Supply Chain Management*	Clear Match: No transformation needed.	0.425[**] N=403	0.386[**] N=393	0.347[**] N=733
E.2.1.7 *Data on Percentage of Recycled/Re-used Raw Material Used*	**ENRRDP027**: *Materials Recycled and Reused Total*: Percentage.	Clear Match: No data=0 0=0 0%-50%=50 50%-100%=100	0.759 N=6	0.988[**] N=3	-0.379 N=13
E.2.1.9 *Programs and Targets to Promote Sustainable Food Products*	**ENPIO15S**: *Score - Product Innovation/Organic Products*	Oke match: 0=0 0-25=25 25-75=75 75-100=100	0.422[*] N=23	0.316 N=23	0.306 N=34

Appendix 12: Indicator comparison Pillar-Environmental Category-Products & Services

Sustainalytics indicator	ASSET4 indicator	Rationale for match	Pearson Correlations		
			2008	2009	2010
E.3.1.3 *Automobile Fleet Average CO2 Emissions*	**ENPIDP029**: *Fleet CO2 Emissions*: Number.	Clear Match: >180=0 170-179=20 No data =30 160-169=40 150-159=70 140-149=90 130-139=95 120-129=97 <120=100	0.700 N=7	0.513 N=7	0.138 N=11
E.3.1.6 *Systematic Integration of Environmental Considerations at R&D Stage (Eco-design)*	**ENPID01S**: *Score - Product Innovation/Policy*:	In the ASSET4 indicator eco-design is only a part of the whole but as an approximation it is good: No transformation needed.	… … For 2008 ASSET4 has no data.	0.344[**] N=195	0.297[**] N=485
E.3.1.7 *Programs and Targets for End-of-Life Product Management*	**ENPIO16S**: *Score - Product Innovation/Product Impact Minimization*:	Oke match: No transformation needed.	0.366[**] N=130	0.400[**] N=127	0.350[**] N=359

E.3.1.9 *Policy on Use of Genetically Modified Organisms (GMO) in Products*	**ENPIO17S**: *Score - Product Innovation/GMO Free Products*	Clear Match: No transformation needed.	0.261* N=68	0.464** N=69	0.108 N=155
E.3.1.11 *Assets Under Management in Responsible Investment*	**ENPIO09S**: *Score - Product Innovation/Environment al Asset Management*:	Sutainalytics uses percentage and ASSET4 score so no transformation needed.	0.513** N=165	0.504** N=159	0.599** N=418
E.3.1.14 *Share of Property Portfolio Invested in Sustainable Buildings*	**ENRRO07S**: *Score - Resource Reduction/Green Buildings*:	Sutainalytics uses percentage and Asset4 score so no transformation needed.	0.364* N=46	0.250 N=47	0.251** N=161

Appendix 13: List of Sustainalytics indicators with no match.

Governance

G.1.3.1	Signatory to UN Principles for Responsible Investment
G.1.4	Tax Transparency
G.1.4.1	Policy on Money Laundering
G.1.4.5	Policy on Genetic Engineering
G.1.5	Business Ethics Related Controversies or Incidents
G.2.4	Disclosure of Directors´ Biographies
G.2.5	Oversight of ESG Issues
G.2.5.1	In-house Team Dedicated to Responsible Investment/Finance
G.2.9	Board Independence
G.2.11	Non-Audit Fees Relative to Audit Fees
G.2.13	Governance Related Controversies or Incidents
G.3.3.1	Transparency on Payments to Host Governments
G.3.4	Public Policy Related Controversies or Incidents

Social

S.1.1.1	Formal Policy on Working Conditions
S.1.2	Formal Policy on the Elimination of Discrimination
S.1.3	Programs to Increase Workforce Diversity
S.1.5.1	Percentage of Temporary Workers
S.1.6.1	Employee Training
S.1.6.4	Health and Safety Certifications
S.1.6.5	Trend in Lost-Time Incident Rate
S.1.7	Employee Related Controversies or Incidents
S.2.1.1	Quality of Social Supply Chain Standards
S.2.1.2	Membership in the Electronic Industry Citizenship Coalition
S.2.1.3	Policy on Sourcing of Coltan
S.2.2	Supply Chain Monitoring System
S.2.2.1	Supply Chain Audits
S.2.2.2	Reporting on Supply Chain Monitoring and Enforcement
S.2.2.3	External Social Certification of Suppliers
S.2.2.4	Fair Trade Products
S.2.3	Contractors and Supply Chain Related Controversies or Incidents
S.3.1.2	Public Policy Statement on Advertising Ethics
S.3.1.3	Policy Statement on Data Privacy
S.3.1.4	Programs to Minimize Health Impact of Electronic and Magnetic Fields
S.3.1.5	Outsourcing of Core Editorial Tasks
S.3.1.6	Corporate Wide Editorial Guidelines

S.3.1.7	Policy on Conflicts of Interests
S.3.1.8	Percentage of Flights Delayed More than 15 Minutes
S.3.1.9	Public Position Statement on Health Consequences of Products
S.3.1.10	Periodic Occupier Satisfaction Surveys
S.3.1.11	Programs and Targets to Reduces Energy/Water use by Customers
S.3.1.12	Adherence to WHO Ethical Criteria for Medicinal Drug Promotion
S.3.2.1	External QMS Certifications
S.3.3	Customer Related Controversies or Incidents
S.4.1	Activities in Sensitive Countries
S.4.2.3	Programs and Targets to Promote Access to Financial Services for Disadvantaged People
S.4.2.4	Policies and Management Systems on Access to Medicines
S.4.2.5	Programs and Initiatives to Develop Medicines for Neglected Diseases
S.4.2.6	Equitable Pricing Programs for Medicines
S.4.2.7	Policies on Access to Health Care
S.4.2.8	Programs to Support Independent Media
S.4.2.10	Policies and Programs to Promote Access to Basic Services
S.4.2.11	Local Community Development Programs
S.4.2.13	Policy on Drug Donations
S.4.2.14	Value of Drug Donations Relative to EBIT
S.4.3	Society and Community Related Controversies or Incidents
S.5.1	Guidelines for Philanthropic Activities and Primary Areas of Support
S.5.3	Percent Cash Donations of NEBT

Environmental

E.1.1	Formal Environmental Policy
E.1.1.1	Reporting Quality Environmental Data
E.1.2.2	Guidelines and Reporting on Closure and Rehabilitation of Sites
E.1.2.3	Environmental and Social Impact Assessments
E.1.2.4	Oil Spill Reporting and Performance
E.1.3.5	Other Programs to Reduces Key Environmental Impacts
E.1.5	Participation in Carbon Disclosure Project
E.1.6	Scope of Corporate Reporting on GHG Emissions
E.1.7.1	Programs and Targets To Improve the Environmental Performance of Own Logistics and Vehicle Fleets
E.1.7.2	Programs and Targets to Phase out CFCs and HFCs in Refrigeration Equipment
E.1.9	Carbon Intensity
E.1.10	Carbon Intensity Trend
E.1.12	Operations Related Controversies or Incidents
E.2.1	Formal Policy or Program on Green Procurement
E.2.1.2	External Environmental Certification Suppliers
E.2.1.3	Programs and Targets to Stimulate Sustainable Agriculture
E.2.1.4	Programs and Targets to Stimulate Sustainable Aquaculture/Fisheries
E.2.1.5	Food Beverage and Tobacco Industry Initiatives
E.2.1.6	Programs and Targets to Reduces GHG Emissions from Outsources Logistics Services
E.2.1.8	Data on Percentage of FSC Certified Wood/Pulp as Raw Material
E.2.1.10	Food Retail Initiatives
E.2.2	Contractors and Supply Chain Related Controversies or Incidents
E.3.1.1	Sustainability Related Products and Services
E.3.1.2	Revenue from Clean Technology or Climate Friendly Products
E.3.1.4	Trend Automobile Fleet Average Fleet Efficiency
E.3.1.5	Products to Improve Sustainability of Transport Vehicles
E.3.1.8	Organic Products
E.3.1.10	Environmental and Social Standards in Credit and Loan Business
E.3.1.12	Use of Life Cycle Analysis for New Real Estate Projects
E.3.1.13	Programs and Targets to Increase Investments in Sustainable Buildings
E.3.1.15	Sustainability Related Financial Services
E.3.1.16	Products with Important Environmental/Human Concerns
E.3.1.17	Carbon Intensity of Energy Mix
E.3.2	Products and Services Related Controversies or Incidents