Original Article

Measuring the level and risk of corporate responsibility
– An empirical comparison of different ESG rating approaches

Gregor Dorfleitner
is a Full Professor of Finance at the University of Regensburg, Germany. He holds a degree in Mathematics and Business Administration from the University of Augsburg, where he also completed his PhD thesis in 1998. He has published various articles in international journals such as *European Journal of Operational Research, Applied Financial Economics, Quantitative Finance, Insurance: Mathematics & Economics, International Journal of Theoretical & Applied Finance, Journal of Asset Management* and the *Journal of Banking & Finance*.

Gerhard Halbritter
currently works as a Management Consultant at zeb.rolfes.schierenbeck.associates GmbH with focus on the financial services industry. In 2015 he completed his PhD at the University of Regensburg. He also holds master's degrees in Business Administration and Economics. His research interests include asset pricing and socially responsible investments.

Mai Nguyen
is a Research Assistant and PhD student at the Department of Finance at the University of Regensburg, Germany. She has received her master's degree in Business Administration from the University of Regensburg. Her research interests include socially responsible investments.

Correspondence: Gregor Dorfleitner, Department of Finance, University of Regensburg, Universitätsstraße 31, Regensburg D-93053, Germany

ABSTRACT   This article empirically compares different rating approaches of corporate social performance (CSP) using environmental, social and corporate governance (ESG) scores of three important sustainability rating providers. We both comment on the ESG level and the risk of changes in a company's ESG level. These ratings are highly relevant to managers and investors comprising ESG issues into their decision-making process. Furthermore, a number of empirical studies, focused on corporate social responsibility topics, are based on certain ESG scoring approaches. The data set includes ESG data for more than 8500 companies worldwide. The article suggests an evident lack in the convergence of ESG measurement concepts. The different ratings neither coincide in distribution nor in risk. Therefore, all CSP stakeholders are encouraged to critically evaluate the validity of the particular ESG scoring model.

INTRODUCTION

In the last two decades the interest in socially responsible investing (SRI), also known as ethical investing or sustainable investing, has increased enormously around the world. The concept of SRI comprises the inclusion of non-financial issues such as environmental, social and corporate governance (ESG) aspects in the investment selection process. ESG ratings provided by specialized rating institutions play an important role in the decision-making process of managers and investors who care about social responsibility as well as for numerous studies. In contrast to the first few articles analyzing the validity and convergence of different ESG metrics, we refer to all three ESG dimensions and empirically compare three ESG rating approaches with respect to the question whether the final ESG appraisal coincides. In addition, we are the first to explore the question of ESG risk in the sense of changes in a company's ESG level over time. This is of relevance to some investors, fund managers and subject to newer methods on portfolio selection based on financial and non-financial information.

Socially responsible (SR) investments are typically identified within a screening process. One can generally distinguish between a negative and a positive approach. The negative method excludes companies involved in controversial business sectors such as alcohol, gambling, tobacco or weapons. As opposed to this, a positive screening process selects companies based on corporate social performance (CSP) measures referring to the assessment of corporate social responsibility (CSR) decisions (see, for example, McWilliams and Siegel, 2001; Renneboog et al, 2008; Luo and Bhattacharya, 2009). Since a positive screening process requires extensive research, specialized rating institutions offer so-called ESG ratings, which evaluate a company's efforts to implement environmental, social and governance issues. Some of the most important ESG rating providers are ASSET4 by Thomson Reuters,

Ethical Investment Research Service (EIRIS), Kinder Lydenberg Domini & Co. (KLD) by MSCI and Sustainability Asset Management Group (SAM). Other providers, such as Bloomberg Sustainability, report ESG disclosure scores.

Managers and investors use these ESG ratings within their decision-making process. Furthermore, a wide range of empirical papers is based on ESG ratings. Chatterji and Toffel (2010) discover that firms and their managers react on being rated. Companies with poor KLD ratings improve the environmental performance to a higher extent than high-rated competitors. Every multinational corporation spends a significant amount of money on CSR activities. The annual statements report upon these endeavors. Managers are facing growing pressure from various stakeholder groups to implement social responsibility (Crilly et al, 2012). Socially responsible investors and providers of financial products, such as mutual funds and indices, also rely on ESG ratings within the portfolio selection process. In this context, the ratings can either be used in an absolute or in a best-in-class version, which benchmarks each company against its peer group. Since 2014, the US market for assets under professional management following SRI strategies has exceeded US$6.57 trillion (US SIF, 2014).

Owing to this impressive development, ESG ratings are utilized for a large number of empirical studies, which contribute important implications for managers and investors. Several papers examine the impact of social screening on the financial performance. While early research commonly utilizes SRI mutual funds or indices (for example, Bauer et al, 2005; Bello, 2005; Kreander et al, 2005; Statman, 2006; Schröder, 2007), today a growing number of ESG ratings allow a more accurate analysis of the relationship between companies' social and financial performance.

Most studies rely on the CSP measurements offered by KLD on the account of a long data history and a large US company sample. While Kempf and

Osthoff (2007) and Statman and Glushkov (2009) provide evidence of the fact that stocks with high ESG scores achieve excess returns, Galema *et al* (2008) and Mănescu (2011) do not suggest a statistically significant difference in risk–adjusted returns between screened and conventional portfolios. Fernando *et al* (2009) review in which way firms' environmental efforts affect various aspects such as a company's market value or the cost of equity capital. They conclude that the stock market rewards neither positive nor negative environmental performers. On the contrary, the results of Chava (2014), Ghoul *et al* (2011) and Goss and Roberts (2011) suggest that the engagement of companies in environmental issues and, respectively, CSR significantly lower the cost of equity or debt. Cai *et al* (2012) question whether firms in controversial industries can be socially responsible and therefore determine the relationship between firm value and CSR engagement for these entities. They achieve the result that the CSR activity of companies in sinful business sectors does have a positive impact on market value. In terms of financial risk, Oikonomou *et al* (2012) find a negative but insignificant relation between CSP and systematic risk, while corporate social irresponsibility is positively correlated with measures of systematic risk for S&P 500 entities. Conversely, Bouslah *et al* (2013) come to the conclusion that the CSP of US firms in the observation period of 1991–2007 mainly impacts idiosyncratic risk.

Many studies use other ESG ratings such as ASSET4, Bloomberg or SAM. Eccles *et al* (2014) examine the effect of CSP on the long-term financial performance of 180 US companies and conclude that 'high sustainability' firms significantly outperform their counterparts in the longer term. The effect of ESG aspects on financial performance and risk of companies in the United Kingdom is investigated by Humphrey *et al* (2011), who infer that firms with high scores do not outperform entities with poor scores with regard to their

systematic risk, book-to-market ratios or performance. Bird *et al* (2012) provide evidence that CSR activities lead to a higher market valuation in the European market while the impact in the United States, Japan and Australia is lower. Cheng *et al* (2014) suggest that companies with superior CSP face significantly lower capital constraints. Wimmer (2013) examines the long-term persistence of ESG scores in screened mutual funds and concludes that the ESG rating levels change after approximately 3 years. Ioannou and Serafeim (2012) show that political institutions, legal and labor market institutions are the main determinants for a firm's CSP level. Other approaches deal with the extension of the conventional portfolio optimizing theory by adding a sustainability dimension. Utz *et al* (2014), Dorfleitner and Utz (2012) and Jessen (2012) are examples of approaches that use ESG scores that measure social or sustainability returns, and integrate these into the theory of portfolio choice.

As managers, investors and researchers rely on ESG ratings and there is no regulation on how the providers have to assess the CSP exactly, it is crucial that the ratings are valid. From a practical point of view it is also desirable for them to coincide largely with respect to the final ESG appraisal. Coincidence does not necessarily require (roughly) identical values. Even if the ratings vary in location and scale, the implications could be equivalent. Chatterji *et al* (2009) investigate the validity of environmental ratings by KLD. The findings indicate that especially the environmental concerns constitute a fairly good measure for the past environmental performance. As opposed to this, the ratings are not able to predict future environmental developments. Semenova and Hassel (2014) compare the environmental performance measurement of ASSET4, Global Engagement Services and KLD. The results provide evidence that the ratings feature some common dimensions but do not converge on aggregate. The findings are mainly because of a different scoring

methodology of the providers. Since these implications are highly relevant for all stakeholders interested in CSP, the analysis should comprise all ESG dimensions. Chatterji et al (2015) take up this point and examine correlations between overall ESG rating data of ASSET4 and KLD as well as different SRI indices. Furthermore, they study whether the ratings are consistent with the composition of social indices. The authors suggest a low convergence among various ESG ratings.

This article extends the existing evidence relating to the convergence of CSP measurements with a qualitative and quantitative analysis of three ESG rating data sets, namely ASSET4, Bloomberg and KLD. Compared with Chatterji et al (2015), we focus both on the overall ESG rating and the individual environmental, social and governance performance scores. We compare descriptive statistics and distributions of the three rating approaches and evaluate correlations among the providers as well as the particular pillars. Despite the fact that all data providers have developed measures for the CSP with regard to ESG issues, the outcomes differ noticeably. ASSET4 and KLD do not coincide in terms of their ESG appraisal. Although Bloomberg provides an assessment of the ESG disclosure level, their ratings generally show a higher resemblance to those of ASSET4. This indicates that ASSET4's valuation methodology depends on firms' disclosure on the one hand, but also that companies with good CSP tend to disclose more non-financial information on the other hand. Beyond that, large corporations are more likely to achieve higher ESG ratings because of enhanced reporting activities. The implications are crucial for managers interested in CSP, socially responsible investors and researchers applying a single ESG concept.

When creating portfolios, SR investors may not only consider the current ESG rating of a company but also the risk that the future CSP of the company could be lower than the current one. This idea is modeled by the approaches of Dorfleitner et al (2012) and Dorfleitner and Utz (2012), who establish the notion of a risky social or sustainability return in portfolio theory. ESG risk is therefore defined as the risk of changes in the average ESG level of an investors' portfolio. This article further contributes to the literature as it is the first to explore whether the riskiness of various providers' ESG scores coincides (largely). Concerning the risk of different ESG scores, the results indicate a similar diversity as found for the ESG level.

The remainder of this article proceeds as follows. We first describe the ESG data and qualitatively compare the three rating approaches. In the second step, we analyze whether the ratings coincide quantitatively. Afterwards, we confer the risks involved with the CSP measurements of the three providers. Finally, the conclusion follows with suggestions for future research.

DATA

In this study, we use annual ESG rating data of publicly traded companies provided by several sources: the ASSET4 database of Thomson Reuters' Datastream, the KLD ratings provided by MSCI ESG STATS (Statistical Tool for Analyzing Trends in Social and Environmental Performance) and the ESG data set of Bloomberg Sustainability. For simplicity, we use the abbreviations ASSET4, KLD and Bloomberg for the three different ESG data providers. Since we wish to compare and analyze the ratings, we first clarify the coverage universe behind the different rating sources. As noted by Chatterji et al (2015), rating agencies may not only define CSR differently but also measure CSP in various ways. Therefore, we provide a short overview of the valuation approaches and the underlying ESG data in the following.

ASSET4 commenced evaluation of the environmental, social and governance performance of approximately 1000 firms in 2002 and provides ESG scores of

over 4300 companies worldwide at the beginning of 2014. The coverage universe comprises listed companies ranging from the S&P 500, Russell 1000, MSCI Europe, FTSE 250, ASX 300 to the MSCI World Index and the 250 MSCI Emerging Markets companies. The KLD rating universe covers the 500 largest US companies and all stocks in the Domini 400 Social Index (a total of roughly 650 stocks) since 1991. All firms listed in the Russell 1000 were added at the beginning of 2001. Since 2003 the KLD coverage includes the largest 3000 US companies by market capitalization. According to Bloomberg, the rating agency has researched 20 000 of the most actively traded public companies. ESG data is provided for approximately 4100 firms in 52 countries. In our study, we use the full available data of KLD and ASSET4. With respect to the Bloomberg data, we were only able to acquire the ESG scores of companies that are also rated by ASSET4. In summary, the total sample ranges from 2002 to 2012 and comprises 8561 corporations.

Generally, the entire assessment of a firm's socially responsible engagement is captured by a total ESG score that is a combination of sub-criteria with regard to environmental, social and corporate governance aspects. In addition, ASSET4 offers an economic dimension that reflects a company's capability to generate sustainable growth and to create shareholder value in the long run by using best management practice. Rating agencies score a company's non-financial performance by means of positive and negative indicators according to their reports and additional public information. In contrast to financial ratings neither a general consensus nor a set of rules exists on how to measure CSP. Therefore, it is necessary to have a closer look at the scoring methods and the underlying data of non-financial information used by the rating agencies.

The KLD rating model includes binary indicators of seven ESG related groups: environmental, governance and social issues such as community, human rights, employee

relations, diversity and customers. According to MSCI ESG Research the ESG information is classified, on the one hand, into strength indicators that comprise beneficial environmental, social and corporate governance dimensions and, on the other hand, concerns that measure the severity of controversies and negative non-financial impacts. The existence of a strength or a concern is indicated by 1. Otherwise, an absence is denoted by 0. The database also includes business involvement data for controversial business issues such as alcohol, gambling and tobacco, which can be interpreted as being an exclusionary screen. The total number of indicators provided by KLD increased from 71 in 2002 to 80 in the time period from 2007 to 2009. Significant rating methodology changes that comprise, among others, industry-specific ESG rating templates were introduced in 2010. As a result, the total number of ESG indicators dropped to 62 because of a substantial consolidation of the ESG rating indicators, but increased to 70 in 2012. Unlike with the other two rating agencies, neither a total score of the individual ESG sub-criteria nor an overall ESG score is available. Hence, the rating results of KLD are basically presented by binary ESG indicators.

In comparison, the ASSET4 data set is based on over 850 binary data points related to aspects of sustainability reporting in the 2014 research cycle. These data points are aggregated in over 250 ESG 'Key Performance Indicators' (KPIs). The KPIs are then combined into 18 'category scores' that are sub-components of the so-called 'pillars' representing 4 areas of a company's social performance: economic, environmental, social and corporate governance. Each pillar can attain any value between 0 and 100. Hence, a high (low) score indicates a strong (poor) performance in this category in comparison with all rated companies. As the notion 'pillars' describes the subordinated scores with regard to the ESG sub-criteria, these terms are used interchangeably

in this article. In contrast to KLD, scores for the individual ESG pillars as well as an overall score, which is also called 'Integrated Rating' (IR) by ASSET4, are available. The total ESG score (IR) is composed of the four equally weighted pillars. According to Thomson Reuters' Datastream, all scores are normalized by using z-scoring and benchmarked against the complete universe of 4000 companies.

Comparable to ASSET4, the scoring model by Bloomberg incorporates more than 100 data points related to ESG. The so-called 'Total ESG Disclosure Score', which is a combination of the environmental, social and governance disclosure scores, is tailored to different business sectors to evaluate each company on data points that are relevant to its industry. The disclosure scores range from 0.1 for companies that expose a minimum number of data points to 100 for those that disclose every sustainability data point. Although the ESG disclosure scores by Bloomberg only indicate the extent to which a company is transparent in terms of its non-financial reporting, Eccles et al (2014) show that firms with higher ESG scores also expose a significantly higher percentage of non-financial information. Hence, the Bloomberg data provides by this means, similarly to ASSET4 and KLD, an insight into a company's level of sustainability. This being the case, the disclosure scores are used as ESG score equivalents in this study.

Since KLD does not supply total scores concerning the ESG factors in the same way as ASSET4 or Bloomberg, it is necessary to aggregate the indicators to ensure comparability. A popular method used in the academic literature on KLD data is to subtract the sum of concerns from the sum of strengths (see, for example, Waddock and Graves, 1997; Johnson and Greening, 1999; Ruf et al, 2001; Chatterji et al, 2009). However, as the number of indicators varies across years and dimensions, the comparability is limited. Hence, we follow the alternative approach introduced by Kempf and Osthoff (2007) that allows for relative scores. For this purpose,

we transform the concerns into strengths by taking the opposite binary value. In order to obtain a total score for each individual ESG sub-criterion, we total the binary indicators of every category and normalize the resulting amount to a range from 0 to 100. Therefore, the results are in the same range as the ASSET4 and Bloomberg scores. There are different ways to achieve standardization: by scaling the scores either from the lowest possible to the highest possible or from the lowest in the data set to the highest in the data set. Since the number of KLD's indicators changes over the observation period, we choose the second option. The weighted average of all ESG-related indicators is symbolized by ESG. The term 'Overall Score', symbolized by ESG$\star$, denotes the calculated total score, which also comprises KLD's rating of involvement in controversial industries (for example, tobacco, firearms).

In this study, the score variables regarding the ESG sub-criteria for all data sets are labeled as ENV (environmental), SOC (social), GOV (corporate governance) and ECN (economic) whereas the total ESG score that includes all three factors, or four pillars in the case of ASSET4, is indicated by ESG.

Although various ESG rating institutions attempt to measure CSP, not only the valuation approaches vary but also the underlying data used differs in quantity and consequently in information depth. Summarizing the above, Bloomberg, KLD and ASSET4 define CSP by the three ESG dimensions. Moreover, ASSET4 defines additionally an economic dimension that refers to a company's framework in order to improve employee and customer satisfaction as well as various economic issues such as insider trading, financial transparency, published profit warnings, brand value or the number of patents owned by the company. These aspects, however, represent a minor share of the total underlying data. Furthermore, most of these economic issues are also covered by the other raters through their three ESG sub-scores, albeit

in a less detailed way. As previously described, the most obvious difference between the valuation models is the number of data points. While ASSET4 scores are calculated on the basis of more than 850 binary data points in the research cycle 2014, KLD presents its ratings through a maximum of 62–80 indicators during the observation period. With respect to the Bloomberg approach, the ESG scores are largely based on the Global Reporting Initiative standards that comprises, analogically to ASSET4, binary data points.

In the environmental dimension, all three data providers basically rate companies in terms of the same issues such as emissions, water, waste, resource reduction, biodiversity and the environmental impact of a firm's products and services. However, there are notable differences in the interpretation of environmental issues. For example, animal testing is solely covered by ASSET4, whereas the question of whether a company's practice complies with environmental regulations is only considered by Bloomberg and KLD.

Consistently for all providers, the social dimension accounts for approximately half of the total data points and, respectively, binary indicators. In this regard, employment quality, occupational health and safety, diversity and opportunity, human rights and product responsibility are primarily evaluated. But when considered in detail, one can see enormous differences in what and to which extent social performance is measured. While KLD evaluates health and safety through only two indicators with regard whether the firm has strong health and safety programs or whether it is involved in controversies, ASSET4 and Bloomberg assess it in a more detailed way. For example, through rating policies or programs on serious diseases and specific pandemics, or the number of injury and fatalities reported by employees and contractors while working for the company. Another example is that KLD does not consider aspects of training and development opportunities for the employees. As ASSET4 has the most extensive data quantity, various

issues such as the indirect economic impact on local communities, the compliance with local regulations regarding political contributions, the level of general business ethics or tax fraud are additionally covered compared with the other raters.

With respect to the governance score, KLD provides the smallest amount of information. This includes reporting quality, public policy, governance structure and business ethics controversies indicators. Beyond that, the other databases also account for aspects relating to shareholder and stakeholder engagement or board functions and structure. The latter, however, is considered by KLD in the social dimension.

Summarizing, although the three institutions rate CSP considering approximately the same aspects, there remain some distinct differences in the method and in the conception of the CSR appraisal. These are especially the level of detail at which the ESG criteria are rated in certain points and the consideration of specific sustainability aspects. However, if a company's entire performance regarding its social responsibility is superior in one rating approach, it should nonetheless obtain comparably good ratings in all other ones, irrelevant of how CSP is measured.

ESG LEVEL: DISTRIBUTION AND CORRELATION

To investigate whether there is any difference between the three ESG rating approaches we compare the distributions and the descriptive statistics in the first step.

Figure 1 presents the histograms of the total ESG scores and the sub-criteria ENV, SOC, GOV and ECN, respectively. It shows the distributions of the three different full samples that include all scores of the available rated companies during the entire observation period. Unsurprisingly, since the valuation approaches are distinct, as described in the previous section, the distribution of the three data sets show no similarity.

Whereas the ESG scores of ASSET4 show a bimodal distribution with a tendency toward lower and higher values, the distribution of the Bloomberg ESG scores exhibits a right skew starting from a low value. In comparison, the ESG scores of KLD are mostly concentrated in the range between 60 and 80. The corresponding descriptive statistics are presented in Table 1. Using a random effects panel model we confirm the results of the descriptive analysis finding significant differences in the average ESG level of the three rating concepts.

As valuation methods of the rating institutions may change during the period under consideration, we also compare the average ESG rating results of the companies from different sources. Table 2 shows the descriptive statistics of the mean ESG scores of all rated firms averaged over the
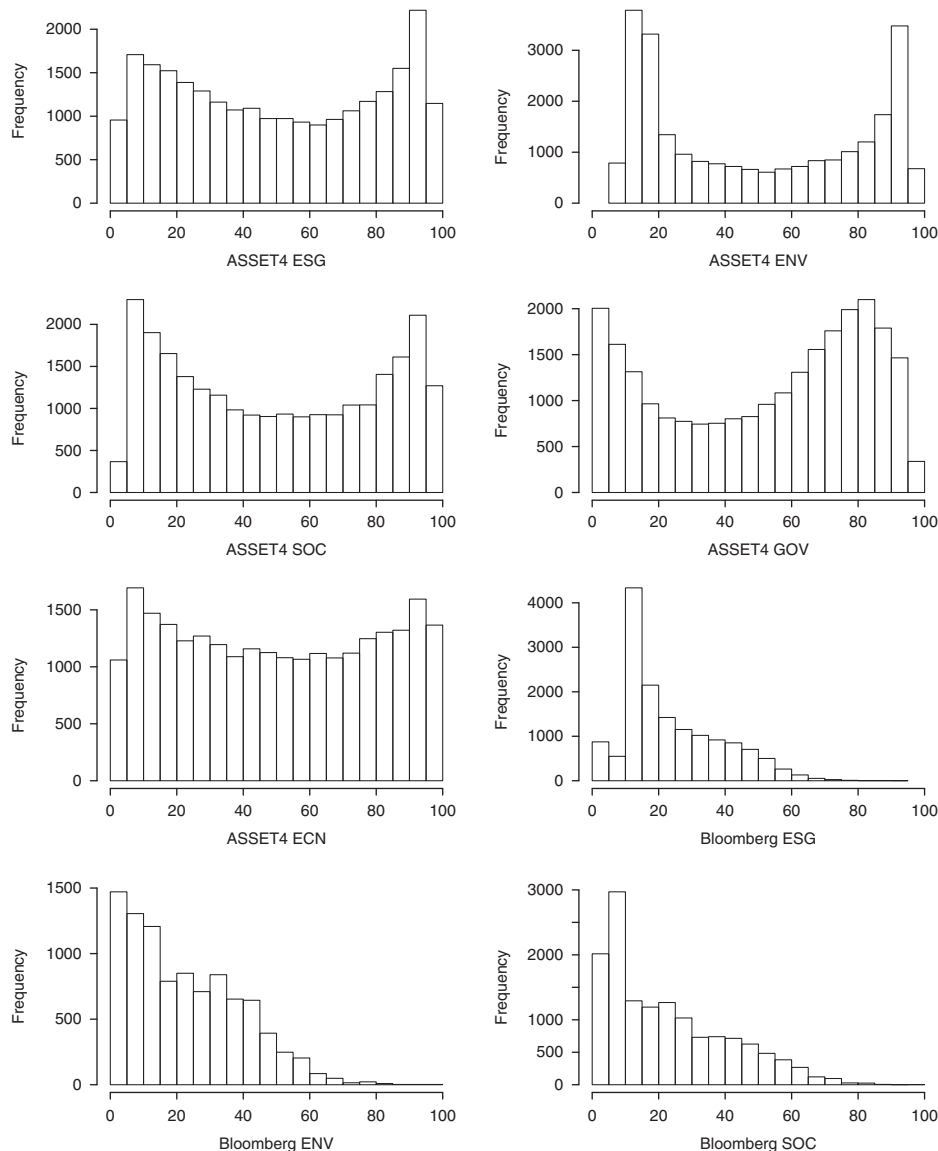

Figure 1: Score histograms.
Notes: This figure presents the histograms for each score type of the three rating agencies ASSET4, Bloomberg and KLD. The histograms include the full panel data set across all companies. The horizontal axis denotes the rating score between 0 and 100, while the vertical axis shows the absolute frequency.
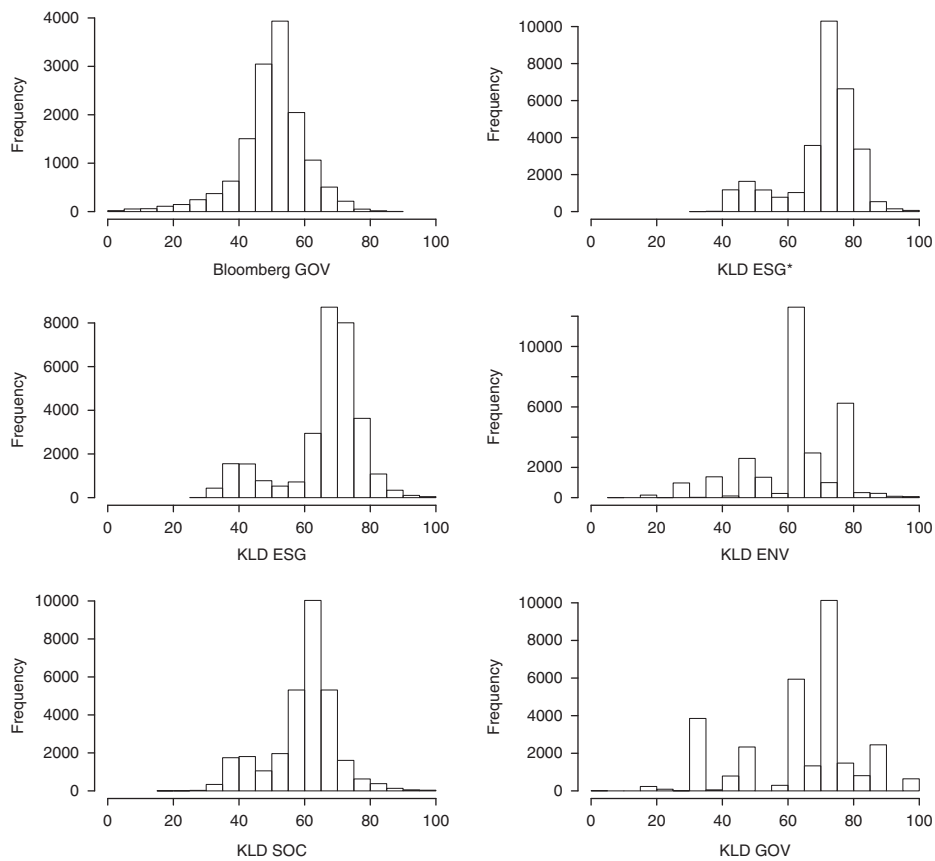
Figure 1: (*Continued*)

Table 1: Descriptive statistics: Full samples

	n	Mean	SD	Skew	Kurt	Min	0.25	Med	0.75	Max
ASSET4										
ESG	24 952	50.16	30.55	0.04	1.59	2.38	21.68	48.55	79.82	98.57
ENV	24 952	49.51	31.88	0.16	1.38	8.86	17.21	44.75	83.65	97.18
SOC	24 952	49.90	30.85	0.07	1.54	3.37	20.06	48.25	80.60	98.95
GOV	24 952	51.94	30.23	−0.31	1.67	1.31	22.01	59.23	78.65	98.62
ECN	24 952	49.89	29.93	0.03	1.67	1.07	22.55	49.19	77.37	99.13
Bloomberg										
ESG	14 968	23.83	14.70	0.85	3.10	0.75	12.92	19.01	33.47	92.56
ENV	9499	22.93	16.56	0.68	2.80	0.78	9.09	19.79	34.88	97.67
SOC	13 990	23.15	18.08	0.87	2.88	3.13	8.77	17.54	35.00	96.49
GOV	14 024	49.95	10.05	−0.86	5.82	3.57	46.43	51.79	55.36	85.71
KLD										
ESG	30 433	66.32	12.06	−1.12	3.77	25.58	64.58	68.75	74.36	100.00
ESG*	30 433	70.36	10.55	−1.08	3.71	32.65	68.52	72.22	77.55	100.00
ENV	30 433	63.27	13.54	−0.96	3.91	9.09	63.64	63.64	72.73	100.00
SOC	30 433	59.54	10.48	−0.60	3.80	17.86	56.25	61.29	65.52	100.00
GOV	30 433	64.59	17.28	−0.61	2.76	0.00	57.14	71.43	75.00	100.00

Notes: This table presents the descriptive statistics of the full samples. It shows the total number of observations as well as the mean, the standard deviation, the skewness, the kurtosis, the minimum, the quartiles and the maximum of all available data by rating agency.

Table 2: Descriptive statistics: Cross-sectional data

	n	Mean	SD	Skew	Kurt	Min	0.25	Med	0.75	Max
ASSET4										
ESG	4356	44.25	27.14	0.24	1.83	3.11	20.57	41.07	67.56	97.44
ENV	4356	44.07	28.14	0.39	1.70	10.10	17.03	37.82	69.80	96.59
SOC	4356	44.98	27.87	0.27	1.73	4.14	19.37	40.90	70.23	97.15
GOV	4356	47.97	27.99	−0.25	1.64	1.53	20.46	54.08	72.74	95.28
ECN	4356	45.67	24.77	0.11	1.99	1.46	25.05	45.02	65.41	98.01
Bloomberg										
ESG	3411	21.33	13.36	0.85	3.14	0.83	12.54	16.87	29.92	70.91
ENV	2277	20.14	14.56	0.72	2.76	0.78	7.44	16.96	30.36	75.45
SOC	3192	20.77	16.09	0.99	3.11	3.13	7.81	15.44	30.70	80.48
GOV	3113	48.41	9.88	−1.26	6.10	3.57	44.90	50.00	53.57	76.49
KLD										
ESG	5323	66.42	8.38	−0.70	4.22	34.23	62.34	66.74	72.09	98.60
ESG*	5323	70.47	7.36	−0.66	4.12	42.63	66.81	70.67	75.51	98.77
ENV	5323	64.02	9.46	−0.53	4.00	18.18	59.07	63.64	70.57	100.00
SOC	5323	59.14	7.34	−0.16	4.58	32.14	55.24	59.54	63.47	94.71
GOV	5323	65.11	10.86	−0.42	4.04	16.67	59.92	65.22	71.61	100.00

Notes: This table presents the descriptive statistics of the mean ESG scores of all available rated companies averaged over the observation period. It shows the total number of observations, the mean, the standard deviation, the skewness, the kurtosis, the minimum, the quartiles and the maximum of the averaged scores over the available times series of each rated company.

observation period. In view of the mean values, there is no noticeable disparity between the analysis on the basis of the full samples and the averaged company scores. The mean values are slightly lower on average, as the descriptive statistics are calculated on the basis of averaged figures. More remarkable are the standard deviations, the minimum and the maximum, respectively. Obviously the ASSET4 data indicates the biggest variability as the standard deviations are the highest while the KLD standard deviations are the lowest. Hence, it is evident that the variability of the ASSET4 scores is the largest of the three data sets as the scoring model has the largest data basis. Thus, ASSET4 provides a much higher database depth and therefore a more differentiated evaluation. It comes as no surprise that the KLD scores have the lowest deviation, since the values are calculated solely on the basis of 62–80 binary indicators during the observation period.

To examine whether the average score over all rated companies varies widely from year to year, we also compute the standard deviations of the yearly average scores based on the full sample of one data provider.

As ASSET4 benchmarks rating results against all other companies of the coverage universe by using z-scoring, the yearly average scores provided by ASSET4 show the least fluctuation from year to year. The standard deviations varies between 1.26 and 2.29 compared with the values between 2.71 and 6.26 for the Bloomberg data and, respectively, compared with the values between 8.03 and 12.51 in the case of KLD's ratings. Generally, a high standard deviation can reflect both changes in the valuation approach and fluctuations in the CSP of the rated universe from year to year. As described in the section 'Data', the high standard deviations of the KLD data are partially because of the varying total number of indicators and the rating methodology changes in 2010.

Apart from the distinct rating approaches and the obvious differences in level and scale of the ESG scores, we investigate whether good or poor CSP is still rated similarly according to the three data sources. For this purpose, we generate three sub-samples each containing only companies for which scores are available from the two data sets to be compared. In the first step, the average

Table 3: Percentage of companies by market capitalization in quartile groups

	ESG				ENV			
	1	2	3	4	1	2	3	4
ASSET4								
Large	29	17	23	32	18	20	25	36
Mid	18	30	29	24	28	27	26	19
Small	28	32	24	15	32	31	23	14
Bloomberg								
Large	31	12	24	33	15	23	28	34
Mid	15	38	25	22	28	26	23	22
Small	29	28	27	16	39	27	21	13
KLD								
Large	17	19	28	35	20	15	33	33
Mid	15	28	34	24	18	23	35	25
Small	30	26	21	22	29	23	25	23

	SOC				GOV			
	1	2	3	4	1	2	3	4
ASSET4								
Large	21	19	24	36	46	24	13	18
Mid	23	30	26	21	11	24	33	32
Small	34	28	25	13	09	28	35	28
Bloomberg								
Large	24	19	26	32	28	17	18	37
Mid	25	31	22	22	20	25	34	22
Small	25	28	28	18	28	34	23	15
KLD								
Large	17	21	26	36	28	23	23	27
Mid	14	24	27	35	31	28	29	11
Small	31	27	24	19	23	25	25	28

Notes: This table presents the rankings by market capitalization. The companies are ranked in quartile groups based on their average scores. Group 1 contains the 25 per cent of all companies with the lowest average score, whereas Group 4 includes the 25 per cent of all companies with the highest average score. The rows show the portion of companies depending on their market capitalization size small (less than $2 billion), medium (between $2 billion or $10 billion) or large in these four groups for each score type and rating agency, respectively.

company scores are ranked within the ASSET4, Bloomberg and KLD ratings and referred to quartile groups. Table 3 reports the results depending on the market capitalization size large (more than $10 billion), medium (between $2 billion and $10 billion) and small (less than $2 billion). It shows that large companies tend to receive better scores except in the case of the corporate governance performance according to the ASSET4 rating. These findings are consistent with those of Humphrey *et al* (2011) that investigate the effect of ESG issues on corporate financial performance and risk.

Next, we compare the quartile groups of the three rating agencies and calculate the percentage of companies that are assigned to the same quartile group according to the different data sources. Table 4 displays the results. ASSET4 and Bloomberg ratings coincide most highly in all sub-scores whereby the closest match occurs in the ESG total score. This means that approximately half of the companies are referred to the same quartile groups according to the ESG score of ASSET4 and Bloomberg. The lowest relation can mostly be detected between Bloomberg and KLD data.

Table 4: Percentage of matching quartiles

	ESG				ENV		
	ASSET4	Bloomberg	KLD		ASSET4	Bloomberg	KLD
ASSET4	100.00	—	—	ASSET4	100.00	—	—
Bloomberg	50.65	100.00	—	Bloomberg	49.68	100.00	—
KLD	37.90	37.70	100.00	KLD	36.65	31.55	100.00
	SOC				GOV		
	ASSET4	Bloomberg	KLD		ASSET4	Bloomberg	KLD
ASSET4	100.00	—	—	ASSET4	100.00	—	—
Bloomberg	47.76	100.00	—	Bloomberg	43.46	100.00	—
KLD	37.54	31.52	100.00	KLD	31.14	32.79	100.00

Notes: This table reports the percentage of companies that are assigned to the same quartile group on the basis of their scores according to the three rating agencies. Each matrix shows the matching percentage of the different score types for every combination.

In order to get a more profound understanding of the convergence of different ESG ratings we also determine correlations between the scores. As the data set has a panel structure, we do not compute them directly. Since the cross-sectional dimension of the data sets is considerably larger than the time-series dimension, we first calculate the cross-sectional correlations among all rating types for every year. For each score combination we use a sub-sample containing identical companies. Years with less than 300 overlapping companies are excluded from the analysis. In addition, we conduct pairwise correlation tests and compute the mean of the annual correlations.

Table 5 presents the average correlations among all different sub-scores of the three rating providers. Although the table shows correlations between all score combinations, we focus our analysis on the combinations with regard to the same provider or the same sub-criterion.

It appears that the total ESG scores of ASSET4, Bloomberg and KLD are highly correlated with their own sub-scores. These correlations range between 0.41 and 0.93 and are all significant on a 1 per cent level. The corporate governance pillar is generally less strongly correlated with the overall score and the other pillars within one provider. ASSET4 additionally accounts for economic issues that show a high dependency level on the the total and social score.

When comparing the relationship between the ESG scores of different sources, ASSET4 and Bloomberg show the highest resemblance. This is particularly true for the total score ESG. The sub-criteria of the two raters also exhibit significant positive correlations between 0.47 and 0.60. The KLD rating data indeed shows little relation to the other two data sources with values between 0.05 and 0.39. In particular the connection between the environmental score of ASSET4 and KLD is extremely slight. This is consistent with the evidence of Semenova and Hassel (2014), who explore environmental ratings and Chatterji *et al* (2015), who analyze the convergence of ESG scores in the time period between 2004 and 2006. Overall, the results show that there is some resemblance between the scores of ASSET4 and Bloomberg. KLD also shows a certain linkage, even though it is less strong.

In summary, the three rating approaches lead to distinct results. Generally, the levels are different. The correlation analysis also shows that the difference cannot simply be described by an affine linear transformation. This leads to the conclusion that ESG scores of different providers do not coincide largely and are therefore incomparable on the whole.

Table 5: ESG level correlations

	ASSET4					Bloomberg			
	ESG	*ENV*	*SOC*	*GOV*	*ECN*	*ESG*	*ENV*	*SOC*	*GOV*
ASSET4									
ESG	1.00								
ENV	0.78***	1.00							
SOC	0.89***	0.74***	1.00						
GOV	0.54***	0.13**	0.29***	1.00					
ECN	0.78***	0.45***	0.61***	0.32***	1.00				
Bloomberg									
ESG	0.62***	0.58***	0.62***	0.31***	0.41***	1.00			
ENV	0.51***	0.56***	0.55***	0.13**	0.35***	0.93***	1.00		
SOC	0.54***	0.54***	0.60***	0.15**	0.34***	0.88***	0.69***	1.00	
GOV	0.38***	0.25***	0.31***	0.47***	0.25***	0.53***	0.34***	0.34***	1.00
KLD									
ESG	0.27**	0.23**	0.29**	0.16**	0.21**	0.39***	0.33***	0.32***	0.29***
ESG*	0.25**	0.21**	0.28**	0.15**	0.20**	0.37***	0.33***	0.31***	0.28***
ENV	0.05*	0.04*	0.07*	0.03	0.05	0.21*	0.17*	0.15*	0.14*
SOC	0.31***	0.27***	0.34***	0.19**	0.23***	0.37***	0.31***	0.31***	0.28***
GOV	0.07*	0.10	0.06*	0.05*	0.07	0.23*	0.21*	0.23*	0.19*

	KLD				
	ESG	*ESG**	*ENV*	*SOC*	*GOV*
KLD					
ESG	1.00				
ESG*	0.99***	1.00			
ENV	0.54***	0.54***	1.00		
SOC	0.89***	0.88***	0.25***	1.00	
GOV	0.41***	0.41***	0.18***	0.09*	1.00

Notes: This table presents the average annual cross-sectional correlations among all different score types of the three rating providers. For each combination we only consider years in which we have overlapping data for at least 300 companies. *** indicates that the correlation is significantly different from 0 on a 1 per cent level for all years. ** and * state that this is the case for at least 2/3 and 1/3 of the years, respectively. Otherwise 1/3 or less of the annual tests are significant on a 1 per cent level.

ESG RISK

In this section we investigate whether the ESG risk of different rating providers coincide. This is of special interest for socially responsible investors who seek to achieve a portfolio with a certain ESG level. Managers of SRI funds such as the Pax MSCI International ESG Index Fund are subject to determined minimum ESG requirements. As a consequence, large variations in a company's ESG assessment can significantly affect the fund composition. Yahoo Inc. may serve as an example for the downside risk of a firm's ESG appraisal. The total ASSET4 ESG score of Yahoo reaches a peak in 2004 with a value of 64.66. Afterwards, the score declines to 19.28 in 2006. Two years later the value

increases to 80.60, while in 2011 the overall ESG score is 43.18. This development is first and foremost driven by variations in the social pillar. As this example shows, a comprehensive comparison of ESG rating approaches needs to consider both the ESG level and risk. Owing to short time series and possible heteroskedasticity problems, we abstain from estimating the volatility of these ESG scores. Furthermore, negative developments are more important for investors in terms of risk. In order to take into account the magnitude of negative changes in ESG ratings, we use the concept of the lower partial moment of order one, which was introduced by Bawa (1975) and Jean (1975).

We only use data that includes observations of at least 3 consecutive years.

Analogously to Fishburn (1977) we use relative changes in order to account for the different score levels of the three data providers. By setting the target value to 0, we aim at measuring the magnitude of loss. Therefore, we also refer to this risk measure as the 'expected loss', which can be formalized as $E(\max\{0, 0-rsc\})$, where rsc symbolizes the relative score change. In order to compare the expected loss between the different data sets, we do not only estimate the expected loss within the score information of one provider, but also use sub-samples of those companies that are also covered by another rating agency in the same years. Table 6 presents the results.

At first glance, it appears that the results based on the three data sets are not comparable to a greater extent. While the average ESG risks according to ASSET4's full sample vary around 7 per cent, the outcomes of the other two providers exhibit a larger variation. Furthermore, it is not possible to consistently determine which score is subject to the highest or lowest risk. According to ASSET4 and Bloomberg the social sub-score is disposed to highest risk, while the same applies to the corporate governance dimension in the case of KLD. Moreover, it is striking that the results calculated on the basis of the ASSET4 sub-samples including only companies that are also covered by Bloomberg in the same time period are generally lower. A similar connection with the other data sets does not exist. Hence, the average expected losses based on the full sample and the sub-samples do mostly not coincide. In summary, the degree of ESG risk appears to be highly dependent on the data source.

Since the ESG scores of ASSET4 and KLD are calculated on the basis of the best rating of each year, the expected loss may depend on outliers. Therefore, we also compute pairwise correlations of the expected losses based on the cross-section of companies, for each of which we can estimate the expected loss in each of the four dimensions from the available time series. Table 7 reports the results.

Table 6: Average expected losses

	ASSET4		Bloomberg		KLD	
	n	*Average (%)*	*n*	*Average (%)*	*n*	*Average (%)*
ESG						
ASSET4	24 952	*7.52*	12 736	5.15	6962	4.47
Bloomberg	12 736	5.60	14 968	*4.59*	4518	5.02
KLD	6962	7.77	4518	5.27	30 433	*5.10*
ENV						
ASSET4	24 952	*6.58*	8414	7.63	6962	5.29
Bloomberg	8414	3.99	9499	*8.01*	2128	5.29
KLD	6962	7.00	2128	8.98	30 433	*5.57*
SOC						
ASSET4	24 952	*7.64*	12 183	8.53	6962	4.96
Bloomberg	12 183	5.49	13 990	*8.25*	4267	5.15
KLD	6962	8.82	4267	10.66	30 433	*5.27*
GOV						
ASSET4	24 952	*7.51*	11 830	3.29	6962	6.54
Bloomberg	11 830	5.80	14 024	*3.42*	4514	7.20
KLD	6962	5.22	4514	3.02	30 433	*7.33*

Notes: The columns ASSET4, Bloomberg and KLD report the number of observations that vary in size depending on the sub-samples used and the corresponding average expected losses. The average expected losses based on the ESG information of one data provider are marked in italics. In the ASSET4 column, the first row exposes the mean expected loss calculated on the basis of the full sample of ASSET4, whereas the second row shows the average expected loss based on the sub-sample including only companies that are also covered by Bloomberg in the same time period.

Table 7: Expected loss correlations

	ASSET4				Bloomberg				KLD			
	ESG	ENV	SOC	GOV	ESG	ENV	SOC	GOV	ESG	ENV	SOC	GOV
ASSET4												
ESG	1.00	—	—	—	—	—	—	—	—	—	—	—
ENV	0.45***	1.00	—	—	—	—	—	—	—	—	—	—
SOC	0.54***	0.38***	1.00	—	—	—	—	—	—	—	—	—
GOV	0.39***	0.18***	0.23***	1.00	—	—	—	—	—	—	—	—
Bloomberg												
ESG	0.07*	−0.01	0.01	0.04	1.00	—	—	—	—	—	—	—
ENV	0.14***	0.20***	0.16***	0.02	0.70***	1.00	—	—	—	—	—	—
SOC	0.11**	0.03	0.14***	0.07	0.68***	0.46***	1.00	—	—	—	—	—
GOV	0.10**	0.02	−0.02	0.10**	0.52***	0.22***	0.28***	1.00	—	—	—	—
KLD												
ESG	−0.04	−0.14***	−0.11***	−0.03	−0.07	−0.02	0.06	−0.05	1.00	—	—	—
ENV	−0.07**	−0.08***	−0.14***	−0.01	−0.13	−0.09	0.03	−0.11	0.79***	1.00	—	—
SOC	−0.03	−0.09***	−0.08***	−0.05	−0.08	−0.01	−0.05	−0.06	0.89***	0.66***	1.00	—
GOV	0.05	0.03	−0.02	0.04	0.02	0.00	0.10	−0.09	0.73***	0.57***	0.63***	1.00

Notes: This table presents the pairwise correlations of the expected losses. For each combination the correlations of the expected losses are computed by sub-samples that contain observations from different rating agencies in the same year for identical companies. ***, ** and * indicate a significance level of 1, 5 and 10 per cent, respectively.

Comparably to the outcomes of the level correlations in section 'ESG level', the correlations between the expected loss of the total ESG score and the individual pillars within one data provider are significantly positive throughout. Consistently for all data sets, the relationship between the corporate governance risk and the total ESG risk is the lowest. However, for many combinations the different rating concepts do not indicate significant expected loss correlations. This is especially true for the relationship between KLD's and Bloomberg's ESG scores. At this point, the highest significant correlation, amounting to only 20 per cent, is between the environmental score of ASSET4 and Bloomberg. Apart from this, the remaining highly significant values at the level of 1 per cent are negative or relatively low. Nevertheless, these results do not indicate a remarkable relationship between the ESG risks of the different ESG score providers.

To sum up, the expected loss is highly dependent on the data source. Hence, hardly any relationship between the average ESG risks of the three data sets exists. Thus, the ESG scores of different providers do not coincide largely with respect to the ESG risk in terms of the expected percentage change of downward variations. Comparable to the results on level correlations, the ESG risk of KLD's scores show the least or no correlation with the other two data sets.

CONCLUSION

Providing a comprehensive comparison of three widely used ESG rating approaches, our study is highly relevant for managers, investors and researchers interested in CSP. We expand upon important previous evidence of Semenova and Hassel (2014) and Chatterji *et al* (2015), focusing on an in-depth analysis of both the rating methodology and their descriptive statistics, distribution and ESG level. We also put special emphasis on the particular ESG dimensions. Furthermore, our study comprises a risk component. This is the first paper to examine ESG risk in the sense of a downward variation of the scores over time, which is of special interest to investors and subject to newer methods on portfolio selection based on financial and non-financial information.

The results suggest an evident lack in the convergence of ESG measurements. First, the qualitative evaluation of the three rating methodologies reveals obvious distinctions in the scoring approaches as well as the CSR definition. This does not only lead to differences in the complexity of the CSP assessment but also in the degree of transparency. While ASSET4 sheds light on various issues regarding social responsibility through qualitative and quantitative questions, KLD combines multiple aspects in one indicator without reporting upon the specific assessment criteria. Although the CSP concepts of the three providers are generally based on similar aspects regarding environmental, social and governance dimensions, the different composition and weighting of the indicators lead to significant distinctions in the final ESG appraisal. This is especially true for the social pillar.

Second, the descriptive statistics confirm the obvious differences in level between the three ESG score providers. Owing to enhanced reporting activities, large firms generally obtain higher scores. By performing a quartile analysis, we find that ASSET4 and Bloomberg ratings are, to a large extent, in line for approximately half of the companies. KLD only shares the same quartile groups for about one-third of the firms. The correlation analysis provides evidence of the fact that the total ESG scores of ASSET4, Bloomberg and KLD are significantly positively correlated with regard to their environmental and social scores. However, KLD shows low correlations to the scores of the other providers. In terms of the particular ESG dimensions, corporate governance aspects are least strongly connected to the other pillars.

Third and lastly, the ESG risk analysis demonstrates that the expected loss is highly dependent on the underlying data basis. Thus, hardly any correlation exists between the different data sets in terms of ESG risk. Similarly to the ESG level analysis, the ESG risk of KLD's ESG scores show the least resemblance to the other databases. Overall, the results do not indicate a remarkable coincidence between the expected losses of the three data providers.

The results provide highly relevant inferences for all empirical studies using a single ESG rating concept. Further research should investigate whether implications of previous papers are robust considering an exchange of the ESG data set. Halbritter and Dorfleitner (2015) already examine this point and investigate whether the relationship between the financial and social performance is dependent on the underlying rating concept. In addition, our findings are of special interest to managers and investors who rely on ESG ratings within their decision-making process. They are advised to accurately analyze whether the ESG rating concept is consistent with their personal goals. Eventually we have to face the diversity of CSR.

ACKNOWLEDGEMENTS

REFERENCES
Bauer, R., Koedijk, K. and Roger, O. (2005) International evidence on ethical mutual fund performance and investment style. *Journal of Banking & Finance* 29(7): 1751–1767.

Bawa, V.S. (1975) Optimal rules for ordering uncertain prospects. *Journal of Financial Economics* 2(1): 95–121.

Bello, Z.Y. (2005) Socially responsible investing and portfolio diversification. *Journal of Financial Research* 28(1): 41–57.

Bird, R., Momenté, F. and Reggiani, F. (2012) The market acceptance of corporate social responsibility: A comparison across six countries regions. *Australian Journal of Management* 37(2): 153–168.

Bouslah, K., Kryzanowski, L. and M'Zali, B. (2013) The impact of the dimensions of social performance on firm risk. *Journal of Banking & Finance* 37(4): 1258–1273.

Cai, Y., Jo, H. and Pan, C. (2012) Doing well while doing bad? CSR in controversial industry sectors. *Journal of Business Ethics* 108(4): 467–480.

Chatterji, A., Durand, R., Levine, D. and Touboul, S. (2015) Do ratings of firms converge? Implications for strategy research. *Strategic Management Journal*. advance online publication August 11, doi:10.1002/smj.2407.

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

Chatterji, A. and Toffel, M. (2010) How firms respond to being rated. *Strategic Management Journal* 31(9): 917–945.

Chava, S. (2014) Environmental externalities and cost of capital. *Management Science* 60(9): 2223–2247.

Cheng, B., Ioannou, I. and Serafeim, G. (2014) Corporate social responsibility and access to finance. *Strategic Management Journal* 35(1): 1–23.

Crilly, D., Zollo, M. and Hansen, M. (2012) Faking it or muddling through? Understanding decoupling in response to stakeholder pressures. *Academy of Management Journal* 55(6): 1429–1448.

Dorfleitner, G., Leidl, M. and Reeder, J. (2012) Theory of social returns in portfolio choice with application to microfinance. *Journal of Asset Management* 13(6): 348–400.

Dorfleitner, G. and Utz, S. (2012) Safety first portfolio choice based on financial and sustainability returns. *European Journal of Operational Research* 221(1): 155–164.

Eccles, R.G., Ioannou, I. and Serafeim, G. (2014) The impact of corporate sustainability on organizational processes and performance. *Management Science* 60(11): 2835–2857.

Fernando, C.S., Sharfman, M.P. and Uysal, V.B. (2009) Does Greenness Matter? Environmental Performance, Ownership Structure and Analyst Coverage. Norman: University of Oklahoma, Working Paper.

Fishburn, P.C. (1977) Mean-risk analysis with risk associated with below-target returns. *The American Economic Review* 67(2): 116–126.

Galema, R., Plantinga, A. and Scholtens, B. (2008) The stocks at stake: Return and risk in socially responsible investment. *Journal of Banking & Finance* 32(12): 2646–2654.

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

Goss, A. and Roberts, G.S. (2011) The impact of corporate social responsibility on the cost of bank loans. *Journal of Banking & Finance* 35(7): 1794–1810.

Halbritter, G. and Dorfleitner, G. (2015) The wages of social responsibility – Where are they? A critical review of ESG investing. *Review of Financial Economics* 26(3): 25–35.

Humphrey, J.E., Lee, D.D. and Shen, Y. (2011) The independent effects of environmental, social and governance initiatives on the performance of UK firms. *Australian Journal of Managment* 37(2): 135–151.

Ioannou, I. and Serafeim, G. (2012) What drives corporate social performance? The role of nation-level institutions. *Journal of International Business Studies* 43(9): 834–864.

Jean, W.H. (1975) Comparison of moment and stochastic dominance ranking methods. *Journal of Financial and Quantitative Analysis* 10(1): 151–161.

Jessen, P. (2012) Optimal responsible investment. *Applied Financial Economics* 22(21): 1827–1840.

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

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

Kreander, N., Gray, G., Power, D.M. and Sinclair, C.D. (2005) Evaluating the performance of ethical and non-ethical funds: A matched pair analysis. *Journal of Business Finance & Accounting* 32(7–8): 1465–1493.

Luo, X. and Bhattacharya, C.B. (2009) The debate over doing good: Corporate social performance, strategic marketing levers, and firm-idiosyncratic risk. *Journal of Marketing* 73(6): 198–213.

Mănescu, C. (2011) Stock returns in relation to environmental, social and governance performance: Mispricing or compensation for risk? *Sustainable Development* 19(2): 95–118.

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

Oikonomou, I., Brooks, C. and Pavelin, S. (2012) The impact of corporate social performance on financial risk and utility: A longitudinal analysis. *Financial Management* 41(2): 483–515.

Renneboog, L., Horst, J.T. and Zhang, C. (2008) Socially responsible investments: Institutional aspects, performance, and investor behavior. *Journal of Banking & Finance* 32(9): 1723–1742.

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

Schröder, M. (2007) Is there a difference? The performance characteristics of SRI equity indices. *Journal of Business Finance & Accounting* 34(1): 331–348.

Semenova, N. and Hassel, L.G. (2014) On the validity of environmental performance metrics. *Journal of Business Ethics*. advance online publication August 17, doi:10.1007/s10551-014-2323-4.

Statman, M. (2006) Socially responsible indexes: Composition, performance, and tracking error. *The Journal of Portfolio Management* 32(3): 100–109.

Statman, M. and Glushkov, D. (2009) The wages of social responsibility. *Financial Analysts Journal* 65(4): 33–46.

US SIF (2014) Report on US Sustainable, Responsible and Impact Investing Trends 2014. Forum for Sustainable and Responsible Investment, US SIF Foundation, http://www.ussif.org.

Utz, S., Wimmer, M., Hirschberger, M. and Steuer, R.E. (2014) Tri-criterion inverse portfolio optimization with application to socially responsible mutual funds. *European Journal of Operational Research* 234(2): 491–498.

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

Wimmer, M. (2013) ESG-persistence in socially responsible mutual funds. *Journal of Management and Sustainability* 3(1): 9–15.