CFR Working Paper No. 23-02

"In partnership for the goals"? The (dis)agreement of SDG ratings

T. Bauckloh • J. Dobrick • A. Höck • S. Utz • M. Wagner

centre for financial research
cologne

“In partnership for the goals”? The (dis)agreement of SDG ratings

Abstract

This paper analyzes the (dis)agreement of Sustainable Development Goals (SDGs) ratings across different rating providers and implications for portfolio management. It documents a considerable level of disagreement that is particularly high for large companies and for companies from the Healthcare and the Basic Materials sector. In general, the sector in which the companies are mainly active explains a large part of the variation in disagreement measures of the SDG ratings. Moreover, we document different return characteristics and risk factor exposures of portfolios sorted according to SDG ratings of different rating providers. Overall, our analyses show that the selection of a specific SDG rating for portfolio allocation can have a crucial impact on financial and non-financial outcomes of portfolios, which bears significant implications for sustainability transitions and their financing.

1. Introduction

A recent innovation of financial service providers is to measure the contribution of companies to the 17 Sustainable Development Goals (SDGs). More precisely, SDG ratings specify whether the business activities of a company contribute positively or negatively to the SDGs. However, while the SDGs translate the otherwise rather vague term "sustainability" into a precisely defined framework, it is unclear whether SDG ratings of different rating providers coincide. The importance of evaluating whether SDG ratings from different rating providers coincide originates from their application in investment decisions. Investors use corporate sustainability assessments such as SDG ratings in portfolio management (e.g., Dimson et al. 2020). Disagreement across SDG rating providers would make portfolio allocation heavily dependent on the provider chosen. Moreover, appropriate SDG ratings might help to allocate money in a way that helps to close the existing annual investment gap of about $4 trillion to achieve the SDGs by 2030 (United Nations 2022a). The more SDG ratings coincide, the more targeted and reliable asset managers can allocate funds in line with the SDGs. A general shortage of money is certainly not an argument for the shortage of investments in business activities conforming with the SDGs since up to $7 trillion in additional funding could be derived from different sources according to the United Nations (2022b). Moreover, $4 trillion represent about 6% of the global assets under management. However, contradictory SDG ratings can, on the one hand, discourage asset managers from using these ratings and, on the other hand, misdirect funds.

Therefore, this study analyzes whether SDG ratings of different rating providers together provide a clear overview which companies contribute to the SDGs and which do not.

Anecdotal evidence suggests that SDG ratings of different rating providers vary substantially. For example, the pharmaceutical firm Bayer AG receives an SDG rating of −3.29 by MSCI and an SDG Rating of 2.5 by ISS[1]. Both rating providers create company SDG ratings between −10 and 10, and the higher the SDG contribution is, the higher the rating. Moreover, both rating providers base their calculation, among other things, on the amount of a company's revenues that contribute to one or more SDGs. Thus, it is surprising that Bayer AG is in the bottom SDG rating quartile of the MSCI company universe, but in the top quartile of the ISS company universe. To generalize this anecdotal evidence, Krippendorff's alpha[2] for our entire sample of company-level SDG ratings of 1,246 companies across four different rating providers (MSCI, Inrate, Vigeo Eiris, and ISS) is 0.39. This number implies a low agreement among SDG ratings.

Several observations stress the importance of our study. Berg et al. (2022), Chatterji et al. (2016), Christensen et al. (2022), Dimson et al. (2020) and Dorfleitner et al. (2015) have considered how assessments of the companies' performance concerning environmental, social, and governance (ESG) aspects differ across rating providers. They show that ESG ratings for the same companies deviate substantially. This disagreement in ESG ratings has, among other things, severe asset pricing implications (e.g., Avramov et al. 2022, Gibson Brandon et al. 2021, Serafeim and Yoon 2022). While the disagreement of ESG ratings can be explained by a missing common understanding of sustainability (Berg et al. 2022), SDG ratings could be expected to be aligned more strongly, since they are all based on the same predefined and detailed framework (i.e., the 17 SDGs). Therefore, a testable hypothesis is whether SDG ratings of different providers show higher agreement than ESG ratings do?[3]

To get a clearer picture of the (dis)agreement between SDG ratings of different rating providers, we look at the extent of (dis)agreement, possible drivers and consequences for asset management. In a first step, we sort the companies in our sample by the SDG ratings of the four different rating providers. Descriptive statistics show that the three highest-ranked regions and sectors in the top-SDG-quartile portfolios differ substantially, although we study the same sample of companies for each provider. A test of how many companies are grouped into the same quartile by two rating providers shows that the average agreement lies between 30% to 50%. To identify the drivers of differences in the SDG ratings, we calculate two disagreement measures of the SDG ratings across rating providers and estimate cross-sectional regressions. The results document that the primary industry sector of a

[1] Both ratings are as of 2020. The methodologies of each rater are outlined in Section 2.
[2] Krippendorff's alpha is a measure for the agreement of the same assessment provided by different raters.
[3] The correlation of ESG ratings ranges between 0.38 and 0.71 (see Berg et al. 2022).

company explain part of the variation in the disagreement measures. While the SDG ratings of the four rating providers show small disagreement in sectors such as Technology, Financials, Consumer Discretionary, and Telecommunications, the disagreement is particularly high in the Healthcare and Basic Materials sectors. For instance, the difference between the largest and smallest SDG rating for one company from the Healthcare sector is 0.715 (or 46.7% of the average sample difference) larger than the disagreement range of the companies from the Telecommunications sector. Finally, we also find disagreement in single SDGs. Nevertheless, only the disagreement in a few SDGs (2, 3, 6, 7, 8, 11, and 13) significantly influences the disagreement in the aggregated SDG rating.

In a second step, we build value-weighted quartile portfolios. Based on this sorting, we calculate the returns of each rating provider's top-minus-bottom SDG rating quartile portfolio. These zero-investment portfolios elicit that the top-SDG quartile portfolios of each rating provider have less systematic risk than the bottom-quartile portfolios. However, while some zero-investment portfolios generate a significantly positive abnormal return (based on a Fama/French 5-factor model), others show a significantly negative abnormal return. Moreover, these zero-investment portfolios also differ in terms of their factor loadings to the size, performance and investment factors.

This paper contributes to the literature on the reliability of sustainability ratings and its implications for portfolio management by analyzing the disagreement of currently available measures of companies' contributions to the 17 SDGs. It documents that SDG ratings of different providers disagree substantially. This is critical as regulators speculate and aim for the investment sector to make a significant contribution to achieving the SDGs by 2030. SDG ratings are a convenient possibility for investors to redirect their funds towards SDGs and there is initial evidence that, unlike ESG Ratings, SDG ratings capture investors' revealed sustainability preferences (van Zanten and Huij 2022). Agreement in SDG ratings might help to direct investment capital most efficiently towards solving the current challenges such as gender inequality (e.g., Brandts et al. 2021), poverty alleviation and combating corruption (e.g., Han et al. 2022). However, our analyses reveal that SDG ratings, in their current shape, largely cannot fulfill this hope.

Specifically, implications for portfolio management become visible when sorting companies based on SDG ratings of different providers as this sorting results in different portfolios with distinct risk factor exposures and portfolio returns. Therefore, investors need to be cautious when using SDG ratings for financial purposes, while researchers should know that research outcomes are heavily dependent on the chosen SDG rating provider. Moreover, companies that make use of SDG ratings to communicate progress in corporate sustainability should be aware that capital market participants might not react to their efforts due to considerably differing assessments of the company's SDG contributions across SDG rating providers.

The remainder of this paper is structured as follows. Section 2 presents the sample details and explains the main variables. Section 3 details the disagreement across SDG ratings of different providers. In the next section, we present our SDG disagreement measures and the results of regression models that explain the variation of the disagreement measures. Further, Section 5 contains performance analyses for portfolios sorted with respect to SDG ratings of different rating providers. Finally, Section 6 concludes.

2. Sample and data

2.1. Sample description

We obtain SDG ratings from four rating providers (MSCI, Inrate, Vigeo Eiris, and ISS) for the year 2020. Table 1 shows sample statistics of the rating universe of each provider as well as the intersection sample ("All"), i.e., the sample of companies for which we have SDG ratings from all four rating providers. The original rating universes comprise 8,271, 1,986, 4,280, and 6,128 companies with non-missing data for MSCI, Inrate, Vigeo Eiris, and ISS, respectively. The intersection of these individual rating universes includes 1,246 companies. The intersection sample is our main sample in the analysis. Panel A of Table 1 presents the distribution of the sample companies regarding sectors. The rating universes of each rating provider and the intersection sample show a similar sector distribution. Consumer Discretionary, Financials, and Industrials are the top 3 sectors with the highest proportion of companies within each sample. Therefore, the intersection sample "All" appears to reflect the sector distribution of the original universes of the rating providers well. Panel B contains the distribution of the sample companies regarding regions. The largest three regions in terms of coverage are Asia, Europe and North America. While Asian companies amount to more than 40% of the companies in the rating universes of MSCI and Inrate, about 45% of the companies in the rating universe of ISS are located in North America. In the intersection sample "All", 34% of the companies are from North America, 29% from Asia and 26% from Europe. Companies from other regions (Africa, Latin America & the Caribbean and Oceania) amount to about 10% of the companies in each sample. Therefore, also with respect to the distribution of the regions, the intersection sample "All" reflects the universes of each rating provider in an appropriate manner.

Panel C of Table 1 shows the distribution of company size in each sample displayed in terms of quartile breakpoints of market capitalization in million USD ($mn), retrieved for all companies from Refinitiv Eikon as of 31st Dec 2020. The median company in the intersection sample has a market capitalization of $11.6 bn. This number is substantially larger than the median size of the companies in the universes of the four rating providers. Thus, the intersection sample tends to consist of larger companies. This is reasonable since, similar to ESG ratings, most of the rating providers focus on assessing companies included in large stock market indices. If a difference in the rating universes exists, this difference could stems from smaller companies due to regional biases of the rating provider, or biases in aims and

scopes. Although our samples differ in terms of company size, we include the largest companies and a large proportion of market capitalization in the "All" sample. For instance, although the MSCI sample includes almost seven times the number of companies, our intersectional "All" sample covers about 43% of the market capitalization of the MSCI sample and 81% of the Inrate sample.

Table 1: Sample statistics

	MSCI	Inrate	Vigeo Eiris	ISS	All
Number of companies	8,270	1,968	4,280	6,128	1,246
Panel A: Sectoral distribution					
Basic Materials	7.24	7.06	7.66	6.58	5.94
Consumer Discretionary	15.49	13.52	16.71	14.59	13.64
Consumer Staples	6.60	9.25	7.38	6.12	8.75
Energy	4.24	4.98	4.91	5.39	5.62
Financials	13.83	18.04	15.02	15.94	19.02
Healthcare	10.42	7.83	7.13	11.21	7.30
Industrials	17.95	15.29	18.46	17.27	16.05
Real Estate	8.02	4.78	6.24	7.20	4.41
Technology	9.50	9.65	8.18	8.40	9.07
Telecommunications	2.90	4.42	3.62	3.07	4.49
Utilities	3.80	5.18	4.70	4.24	5.70
Panel B: Regional distribution					
Africa	1.39	1.17	2.01	0.93	1.44
Asia	40.04	44.36	28.69	21.13	29.05
Europe	19.54	22.00	31.82	23.40	26.08
North America	30.90	23.63	24.74	45.77	34.03
Latin America & the Caribbean	4.09	6.86	3.79	3.04	6.42
Oceania	4.04	1.98	8.95	5.73	2.97
Panel C: Size quartile breakpoints (in $mn)					
25th percentile	966.53	4035.04	1562.92	1024.20	5748.59
50th percentile	2561.25	8236.27	4575.02	3305.43	11603.54
75th percentile	7259.71	19415.82	12852.93	9256.76	28000.37

This table shows sample statistics of the universes of the four different SDG rating providers (Columns "MSCI" "Inrate", "Vigeo Eiris" and "ISS"). Column 5 ("All") shows the statistics of the intersection of all four universes, i.e., the sample of companies that get an SDG rating from all four rating providers. The first row contains the absolute number of companies in each sample. Panel A (B) depicts the distribution of the companies with respect to sectors (regions). All values pertaining to sector and region are in percent. Panel C contains the sample's quartile breakpoints with respect to market capitalization in $mn.

2.2. SDG ratings

The four rating providers provide SDG assessments at different levels of aggregation for each company. In the main analysis of this paper, we study one aggregated SDG rating per company and rating provider. While ISS and Vigeo Eiris provide an aggregated SDG rating, we calculate such a rating for MSCI and Inrate from their granular rating data. In the following, we explain the four rating approaches and how the aggregated SDG rating is generated.

MSCI provides SDG assessments for each of the 17 SDGs on a scale from −10 to +10. The rating with respect to one SDG is calculated by the average of the SDG Product Alignment Rating and the SDG

Operational Alignment Rating. A rating of −10 is assigned to a company that is strongly misaligned with an SDG. This can be the case when a company either generates over 50% of its revenue from activities with adverse impact related to an SDG or if it is involved in major controversies related hereto. The SDG Product Alignment Rating measures the net contribution of a company's products and services. The SDG Operational Alignment Rating assesses the impact of a company's operations. To obtain one SDG rating across all seventeen SDGs, we calculate the arithmetic mean across all seventeen SDG ratings for each company and end up with an overall SDG assessment for each company.

Inrate maps a company's revenue to 300 standardized product and service segments. These product and service segments may contribute from "very negative" (−2) to "very positive" (+2) to each SDG, leading to an SDG net alignment with respect to each of the 17 SDGs in $mn. We sum up the 17 SDG contributions for every company and end up with a net SDG contribution for every company which is positive (negative) if the sum of the contributions to each SDG is greater (smaller) than zero.

Vigeo Eiris provides one overall SDG rating for each company in their universe. A company's "Behavior Rating" is calculated which encompasses several criteria that are relevant to a company's SDG assessment framework. These criteria are weighted and the company's performance on each of these is measured. Thereafter, a decision tree determines the company's overall SDG contribution on a scale from "Highly Adverse" (−2) to "Highly Positive" (+2) by taking into account the company's geographical peers, its involvement in controversial activities and its involvement in sustainable goods and services.

ISS provides SDG assessments for 15 different objectives on a scale from −10 to +10, where −10 (+10) indicates that 100% of a company's net sales are related to products/services as well as operations that contribute negatively (positively) to one respective category. Out of these 15 single assessments, one overall rating is formed by taking the most extreme value(s) in either direction, i.e., positive or negative if there are only positive or negative ratings. If positive as well as negative ratings exist, the overall rating is computed by taking the arithmetic mean of the highest and lowest values.

Due to differences in scaling of the SDG ratings and to allow for a better comparison, we apply z-scoring and calculate a standardized SDG rating $SDG_{i,j}^{z}$ for each company i from each rating provider j

$$SDG_{i,j}^{z} = \frac{SDG_{i,j} - \mu_j}{\sigma_j} \tag{1}$$

where $SDG_{i,j}$ is the aggregated SDG rating of company i provided by rating provider j, μ_j is the cross-sectional mean of the aggregated SDG ratings of rating provider j and σ_j is the standard deviation of the aggregated SDG rating distribution of rating provider j. As a result, the distributions of the

standardized SDG rating $SDG_{i,j}^{z}$ have a mean of 0 and a standard deviation of 1 for each rating provider and are therefore directly comparable.

Table 2 contains descriptive statistics of the standardized SDG rating $SDG_{i,j}^{z}$ and sample statistics for subsamples that we grouped with respect to SDG rating quartiles for every rating provider. Rows starting with "1" show the statistics of the subsample of companies with the 25% lowest $SDG_{i,j}^{z}$ values, rows starting with "4" show the statistics of the subsample of companies with the 25% highest $SDG_{i,j}^{z}$ values. Correspondingly, quartiles 2 and 3 display the sections in between. The mean standardized SDG ratings increase with ascending quartiles. Moreover, Vigeo Eiris has the specific case of a zero standard deviation of the ratings in quartiles 2 to 4. This is due to the discreteness of Vigeo Eiris SDG ratings with only five different levels and therefore quartile groups that only contain one single level of SDG ratings and thus no variability.

Table 2 additionally contains descriptive sample statistics such as the top 3 countries, regions, and sectors. For instance, the column "Top 3 Countries" contains the list of countries (of the companies' headquarters) that occur with the highest frequency in the respective quartile. The order of the country abbreviations indicates the order in the top 3. Thus, the US is the country with the highest number of companies in Quartile 1 of the MSCI SDG ratings. The second-most companies are from Canada, and the third-most companies are from Korea. Except for Quartile 4 of the MSCI Panel, US companies take the most prominent position across all quartiles and across all rating providers, followed by Japanese companies. This can be explained by the dominant share of North American (and here particularly US) companies in the overall sample. Clustered by regions, North American companies are mostly dominant for the lower quartiles but are overtaken in some cases by Asia and Europe for higher quartiles. Concerning sectors, there seems to be more disagreement between the four rating providers: Energy is the most represented sector in the first quartiles of MSCI and Inrate whereas it is Industrials and Consumer Discretionary for ISS and Vigeo Eiris. Companies in the Healthcare sector seem to be rewarded by Vigeo Eiris and ISS methodologies with the sector taking on the top position in Quartile 4 of both rating providers even though it only accounts for 7.3% of the entire sample. Financials are the top performers for MSCI and Inrate, however, they are also the top sector for lower quartiles, which can be explained by their large share in the overall sample.

Table 2: Descriptive statistics and sample statistics per SDG rating quartile

	Mean	SD	Min	Max	Top 3 Countries	Top 3 Regions	Top 3 Sectors
					MSCI		
1	−1.22	1.19	−8.54	−0.21	US, CA, KR	North America, Asia, Europe	Ener, ConD, Util
2	0.01	0.11	−0.18	0.18	US, JP/CA	North America, Asia, Europe	Fin, Ind, ConD
3	0.40	0.11	0.21	0.57	US, JP, KR	Asia, North America, Europe	Fin, Ind, ConD
4	0.90	0.27	0.60	2.16	JP, US, FR	Asia, Europe, North America	Fin, Ind, ConS
					Inrate		
1	−1.24	0.45	−2.57	−0.65	US, JP/CA	North America, Asia, Europe	Ener, ConS, ConD
2	−0.01	0.33	−0.65	0.31	US, JP, CA	North America, Asia, Europe	Ind, ConD, Tech
3	0.41	0.07	0.31	0.53	US, JP, TW	Asia, North America, Europe	Fin, Tech, ConD
4	1.18	0.67	0.53	4.47	US, JP, CH	Europe, North America, Asia	Fin, Hc, Tel
					Vigeo Eiris		
1	−1.34	0.46	−1.92	−0.97	US, JP, CA	North America, Asia, Europe	Ind, Fin, ConD
2	−0.02	0.00	−0.02	−0.02	US, JP, CA	North America, Asia, Europe	Ind, ConD, Fin
3	0.93	0.00	0.93	0.93	US, JP, TW	Asia, Europe, North America	Fin, Tech, Ind
4	1.88	0.00	1.88	1.88	US, FR, GB	North America, Europe, Asia	Hc, Ind, ConD/Util
					ISS		
1	−1.22	0.65	−2.58	−0.33	US, CA, JP	North America, Asia, Europe	Ind/Cons, ConD,
2	−0.04	0.08	−0.30	0.01	US, JP, KR	Asia, North America, Europe	Fin, Ind, ConD
3	0.12	0.08	0.03	0.29	US, JP, CH	Europe, Asia, North America	Fin, Ind, Tech
4	1.18	0.79	0.32	2.59	US, JP, CA	North America, Europe, Asia	Hc, Tel, Tech

This table shows descriptive statistics for quartile portfolios built with SDG ratings from 2020 of the "All" sample. Mean, SD, Min, and Max denote arithmetic mean, standard deviation, minimum and maximum values of the standardized SDG ratings, respectively. Top 3 Countries, Top 3 Regions and Top 3 Sectors denote the top 3 countries, the top 3 regions and the top 3 sectors per quartile. The abbreviations CA, CH, FR, GB, JP, KR, TW and US denote Canada, Switzerland, France, Great Britain, Japan, Korea, Taiwan and the US, respectively. The abbreviations Ener, ConD, Util, Fin, Ind, ConS, Tech, Hc, and Tel denote Energy, Consumer Discretionary, Utilities, Financials, Industrials Consumer Staples, Technology, Healthcare, and Telecommunications, respectively.

3. (Dis)agreement of SDG ratings

To answer the question regarding the (dis)agreement of different rating providers on a company's SDG performance, in a first step, we present results for a general indicator of the agreement on a measurement construct provided by different actors, Krippendorff's alpha, in Table 3. Krippendorff (1998) suggests that a value higher than 0.8 indicates agreement between different rating providers on the construct being measured, with a minimum value of 0.667 recommended to at least be able to make rough statements. We find that all Krippendorff's alpha values are well below the recommended values, with a value of 0.39 for all four rating providers (Panel A). Combinations of 3 (Panel B) as well as 2 (Panel C) providers mostly yield values between 0.3 and 0.4, with the value for Inrate and ISS standing out at 0.63. Thus, in general, the agreement of different rating providers on a company's SDG performance is low.

Table 3: Krippendorff's alpha (interval)

Panel A: All raters	
	0.39
Panel B: Three raters (the named rater is the rater omitted in the calculation)	
- MSCI	0.41
- Inrate	0.34
- Vigeo Eiris	0.45
- ISS	0.35
Panel C: Pairwise	
MSCI and Inrate	0.40
MSCI and Vigeo Eiris	0.38
MSCI and ISS	0.32
Inrate and Vigeo Eiris	0.28
Inrate and ISS	0.63
Vigeo Eiris and ISS	0.33

This table shows Krippendorff's alphas for standardized SDG ratings. Panel A shows Krippendorff's alpha for the sample of companies with SDG ratings available for all four raters, Panel B shows Krippendorff's alpha for all possible combinations of three out of the four rating providers, where for instance "– MSCI" indicates the case when MSCI is not considered. Panel C shows Krippendorff's alpha values for all possible pairwise combinations between the rating providers.

In a next step, we conduct a more in-depth analysis and look at SDG quartile company matches between different rating providers. Results are provided in Table 4. In this analysis, we divide all companies included in our "All" sample into quartiles for each rating provider. We then calculate the share of companies in the quartiles of one rating provider that are also included in the respective quartiles of the other three rating providers. For example, in Panel A, only 31.14% (16.96%; 34.26%) of the companies included in the top quartile of MSCI are included in the top quartile of Inrate (Vigeo Eiris; ISS). 37 of 48 percentual matches are between 20% and 60%, with some outliers upwards as well as downwards. The highest average agreement is between Inrate and ISS, with an agreement between 60% and 70% in quartiles 1 and 4. Furthermore, the high value of 82.5 between Vigeo Eiris and ISS

Table 4: Percentual matches

Panel A: Per quartile				
	1	2	3	4
		MSCI		
N	323	313	321	289
Inrate	52.94	46.65	14.95	31.14
Vigeo Eiris	42.11	41.85	36.45	16.96
ISS	45.82	30.99	21.50	34.26
		Inrate		
N	341	442	151	312
MSCI	50.15	33.03	31.79	28.85
Vigeo Eiris	36.36	40.27	49.67	16.99
ISS	65.69	36.88	33.11	60.90
		Vigeo Eiris		
N	334	499	333	80
MSCI	40.72	26.25	35.14	61.25
Inrate	37.13	35.67	22.52	66.25
ISS	38.32	25.45	36.34	82.50
		ISS		
N	317	353	265	311
MSCI	46.69	27.48	26.04	31.83
Inrate	70.66	46.18	18.87	61.09
Vigeo Eiris	40.38	35.98	45.66	21.22
Panel B: Entire match (N = 1,246)				
MSCI & Inrate	36.52			
MSCI & Vigeo Eiris	34.75			
Inrate & Vigeo Eiris	34.51			
MSCI & ISS	33.15			
Inrate & ISS	50.32			
Vigeo Eiris & ISS	35.47			

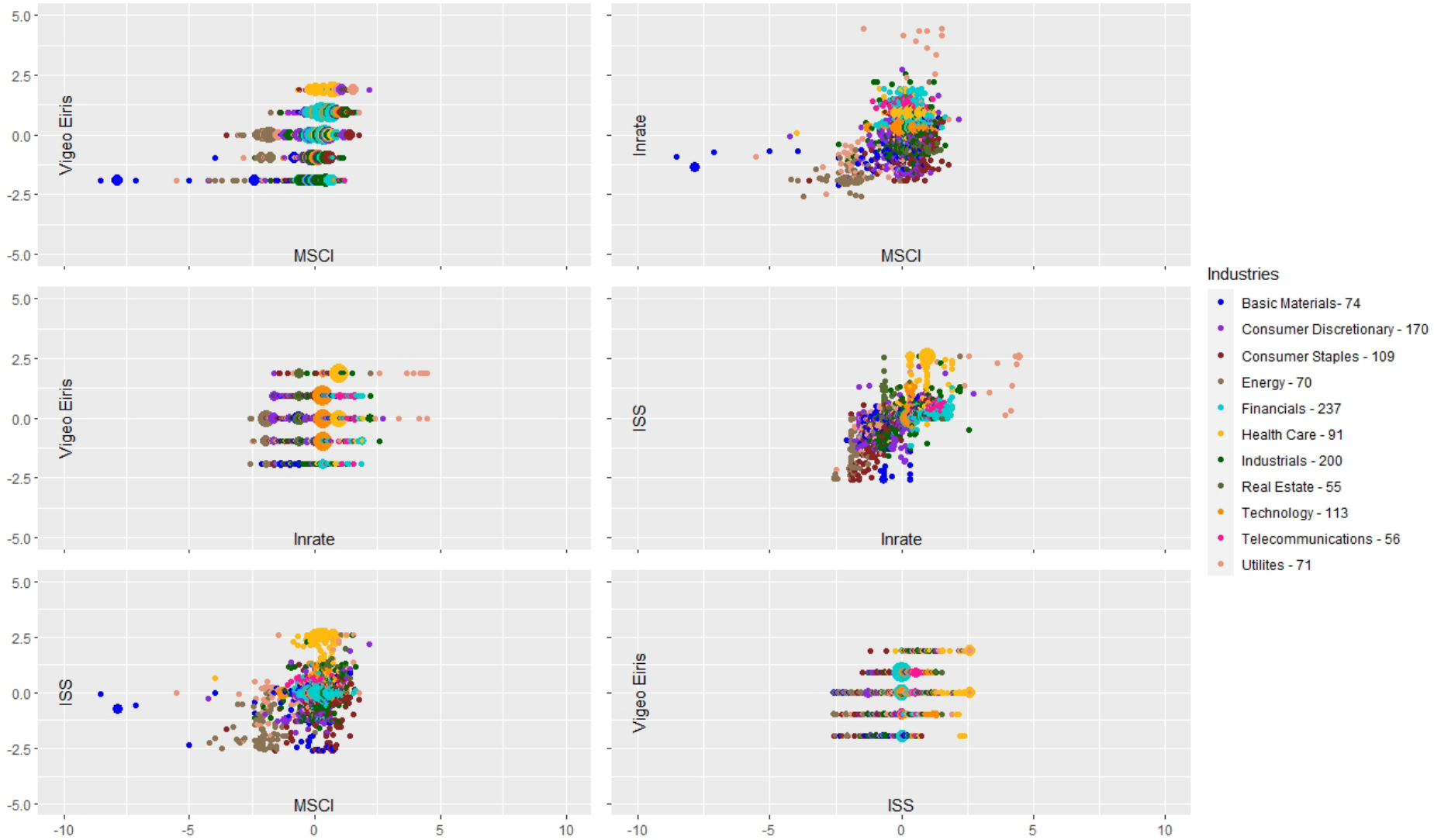
This table shows percentual matches between the rating providers for the "All" sample. Panel A shows the percentual share of companies that were assigned to the same quartile by pairs of rating providers. Panel B shows the percentual share of companies that were assigned to the same quartile across all quartiles and for all possible pairs of rating providers.

stands out. However, the fact that Vigeo Emiris only has 80 companies in quartile 4 puts the value somewhat in a different perspective: considering that about 60 of these 80 companies are in ISS's fourth quartile, which corresponds to roughly 20% of all companies in the 4th quartile of ISS, the high value of 82.5 is in fact somewhat misleading at first glance. In Panel B, we illustrate the percentual match between all possible rating provider pairs across all quartiles. That is, 36.52% of all companies were assigned to the same quartile by MSCI and Inrate. For all possible rating providers pairs, we achieve a percentual match of around 35%, except for the pair Inrate & ISS. This stresses our findings from Panel A that Inrate and ISS SDG ratings are the most similar. However, in general, Table 4 again indicates a low agreement of SDG ratings of different rating providers.

Figure 1 provides a visual presentation of the heterogeneity of SDG ratings across rating providers. It illustrates the standardized SDG ratings of all possible pairwise combinations of rating providers for

the “All” sample of 1,246 companies. The colors in this figure refer to the ICB sector classification of a company. Diameters of single points indicate the number of companies with the same value with larger points representing more companies. The legend of the figure shows the 11 ICB sectors and the corresponding number of companies. The scales of the plots are chosen to be uniform to enhance comparability. The figure indicates that SDG ratings not only differ on the company but also on the sector-level. For instance, while MSCI assigns SDG ratings between around −2 and +1 to companies from the Healthcare sector (yellow points), Vigeo Eiris assigns the highest possible SDG rating to all Healthcare companies. ISS also rates Healthcare companies very high, although the variability is greater than for Vigeo Eiris. Other sectors with SDG rating disagreement among rating providers are in particular Basic Materials, Utilities, and Energy. MSCI seems to penalize some companies in these sectors with a highly negative SDG alignment rating, whereas the other rating providers do assign some negative SDG ratings, but to a much smaller and less systematic extent. The outliers in MSCI ratings are likely to be due to their methodology of assigning ratings of −10 if a company generates over 50% of its revenue from products and services that have a negative impact on an SDG or if a company is involved in very severe controversies. Additionally, looking at the various graphs, it is also evident that there are many sectoral clusters within one rating provider’s SDG assessment, with some sectors being *per se* rated higher than other sectors. For instance, for Inrate, companies from the Financial, Telecommunications, and Healthcare sector are the top SDG contributors, whereas companies from the Energy and Utilities sectors contribute rather negatively to the SDGs.

Figure 1: Scatterplots of the standardized SDG ratings of the „All“ sample



Each scatterplot of this figure illustrates the standardized SDG ratings of the indicated two rating providers for the “All” sample. The colors indicate the ICB sector of each company and the size of the plotted dots indicates the number of companies with this tuple of SDG ratings.

4. Explaining the disagreement of SDG ratings

In a next step, similar to the approach of Dorfleitner et al. (2022) in the context of corporate social responsibility scandals, we identify determinants of the disagreement between the rating providers based on a broad set of regional, sectoral, firm-level and thematic variables. First, we explain our measures of disagreement in the overall SDG assessment by a standard set of financial and accounting indicators, sectors and regions. Thereafter, we look at how the disagreement in individual SDGs drives the overall disagreement.

We construct two disagreement measures, namely *sd* and *max–min*. For the first measure *sd*, we take all four standardized SDG ratings and calculate the standard deviation for each company in our matched "All" sample of 1,246 companies. The second measure *max–min* represents the difference of maximum and minimum standardized SDG rating for each company in our sample. We use both of these SDG disagreement measures as dependent variables in a cross-sectional regression model with a set of explanatory variables. As explanatory variables, we use logarithmized market value (MV), market to book (M/B), return on assets (ROA), leverage[4], and price to earnings (P/E) from Refinitiv EIKON as well as the ICB sector and the region of the company's headquarter. We winsorize all variables at the 5th and 95th percentile to account for outliers. All monetary values are in US$.

Table 5 presents descriptive statistics for our dependent and independent variables. The dependent variables *sd* and *max–min* have an arithmetic mean of 0.69 and 1.53, respectively, indicating that the mean standard deviation of the standardized SDG ratings of the four rating providers is on average 0.69. Equivalently, the average difference between maximum and minimum values per company and across all four rating providers is 1.53. The skewness and kurtosis of the two variables indicate that both are slightly right-skewed, i.e., there are more lower values than higher values in both disagreement measures, and strongly leptokurtic with a high number of extreme values.

Table 5: Descriptive statistics for the regression variables

	Mean	SD	Min	Max	Skew	Kurt
sd	0.69	0.37	0.05	3.86	1.85	9.30
max–min	1.53	0.84	0.10	8.47	1.76	8.26
log(MV)	9.44	1.13	5.64	14.00	0.50	0.35
M/B	3.31	3.31	0.48	13.31	1.82	2.58
ROA	0.06	0.05	−0.01	0.18	0.96	0.20
Leverage	1.09	1.09	0.00	4.05	1.41	1.18
P/E	0.25	0.18	0.07	0.76	1.56	1.77

This table presents descriptive statistics of the variables used in the regression analysis in Table 6. The data is from the "All" sample. log(MV), M/B, ROA, Leverage, P/E are winsorized at the 5% and 95% levels. Mean, SD, Min, Max, Skew, and Kurt represent the arithmetic mean, standard deviation, minimum and maximum values, skewness, and kurtosis, respectively.

[4] (Long Term Debt + Short Term Debt & Current Portion of Long Term Debt)/Common Equity*100

We estimate four sets of regression models. In the first set of models (1), we regress our SDG disagreement measures on our financial and accounting dependent variables. In the second set of models (2), we add ICB sector affiliation as categorized in Table 1 (Panel A) to the model. In the third set of models (3), we extend the previous model with headquarter region as an explanatory variable. In the fourth set of models (4), we include the disagreement measures of single SDG clusters (depicted in Table 7) as further independent variables. Table 6 shows the results of the regression analyses. For each model, we present regression coefficients and adjusted R^2 values (in %) as an overall measure for the general fit of the model. Standard errors clustered with respect to sector and region are in parentheses.

We have four main takeaways from the regression results. (1) The SDG ratings of large companies differ more than those of smaller companies. (2) The SDG ratings of companies from the sectors Healthcare, Basic Materials, Energy, Consumer Staples, Real Estate and Utilities show larger disagreement. (3) The sector a company belongs to substantially explains the disagreement of SDG ratings. (4) Finally, the aggregate disagreement is additionally driven by the disagreement in some specific SDG clusters.

In a first step, we focus on the results of the first three regression models. The coefficients for log(MV) are highly significant in all regression specifications from models (1) to (3), indicating that the disagreement between rating providers is higher the larger the company (in terms of market cap). A one standard deviation increase in the variable log(MV) corresponds to an increase of 0.056 in *sd* and 0.130 in *max–min* in model (1). Adding sector as a further explanatory variable to the regression model (2) does not change the significance levels and the coefficients of our financial variables much, however, both M/B coefficients become insignificant. The intercept term becomes significantly different from zero on a 10% level when sd is the dependent variable. We can see that sector variable adds further explanatory power to the model. Our F-tests for nested models indicate that the increment in adjusted R^2 is significant at the 1% level from model (1) to (2). Out of ten sectors, seven differ significantly from our reference sector Telecommunications. Whereas the disagreement relative to the reference sector is significantly lower in the Technology sector, it is significantly higher in the following sectors: Basic Materials, Consumer Staples, Energy, Healthcare, Real Estate and Utilities.

In model (3), we add the company's headquarter region as another explanatory variable to our model. Again, the results for our financial variables do not change much. As already visualized in Figure 1, the sector explains a fair amount of the variance in our dependent variables with adjusted R^2 values of around 18%. After adding the region to the model, our F-tests for nested models only indicate a slight increase in model quality from model (2) to (3).

Table 6: Cross-sectional regression on the SDG disagreement

	(1)		(2)		(3)		(4)	
	Dependent Variable		*Dependent Variable*		*Dependent Variable*		*Dependent Variable*	
	sd	*max–min*	*sd*	*max–min*	*sd*	*max–min*	*sd*	*max–min*
Constant	0.128	0.234	0.157*	0.306	0.273**	0.588**	0.068	0.102
	(0.115)	(0.255)	(0.094)	(0.202)	(0.132)	(0.289)	(0.105)	(0.226)
log(MV)	0.056***	0.130***	0.048***	0.113***	0.049***	0.115***	0.018*	0.018**
	(0.012)	(0.026)	(0.010)	(0.022)	(0.011)	(0.025)	(0.009)	(0.020)
M/B	-0.014**	-0.034***	-0.006	-0.017	-0.007	-0.018	-0.005	-0.014
	(0.006)	(0.012)	(0.005)	(0.011)	(0.005)	(0.011)	(0.004)	(0.009)
ROA	0.605	1.364	0.031	0.152	-0.019	0.058	0.419*	1.044*
	(0.404)	(0.893)	(0.335)	(0.734)	(0.329)	(0.724)	(0.251)	(0.548)
Leverage	0.002	0.012	0.006	0.017	0.008	0.022	0.006	0.019
	(0.011)	(0.025)	(0.010)	(0.023)	(0.010)	(0.022)	(0.01)	(0.023)
P/E	0.001	0.002	0.000	0.016	-0.006	0.008	0.015	0.053
	(0.001)	(0.002)	(0.092)	(0.198)	(0.091)	(0.196)	(0.069)	(0.147)
Tech			-0.172***	-0.398***	-0.17***	-0.392***	-0.140***	-0.303***
			(0.055)	(0.122)	(0.051)	(0.114)	(0.047)	(0.107)
BasM			0.327***	0.710***	0.306***	0.666***	0.158***	0.369***
			(0.089)	(0.188)	(0.075)	(0.159)	(0.061)	(0.129)
ConD			0.003	-0.011	0.003	-0.008	-0.023	-0.037
			(0.053)	(0.110)	(0.057)	(0.118)	(0.051)	(0.107)
ConS			0.236***	0.465***	0.229***	0.451***	0.066	0.115
			(0.043)	(0.092)	(0.046)	(0.100)	(0.052)	(0.115)
Ener			0.261***	0.528***	0.243***	0.494***	0.101*	0.212
			(0.045)	(0.099)	(0.044)	(0.099)	(0.061)	(0.132)
Fin			-0.048	-0.105	-0.062	-0.134	-0.008	0.008
			(0.044)	(0.094)	(0.043)	(0.093)	(0.044)	(0.098)
Hc			0.333***	0.722***	0.329***	0.715***	0.268***	0.626***
			(0.042)	(0.088)	(0.043)	(0.091)	(0.052)	(0.114)
Ind			0.017	0.034	0.012	0.027	0.004	0.039
			(0.041)	(0.088)	(0.041)	(0.089)	(0.043)	(0.099)
RealE			0.099*	0.248**	0.089*	0.229**	0.084*	0.084**
			(0.053)	(0.118)	(0.051)	(0.114)	(0.045)	(0.103)
Util			0.240***	0.532***	0.227***	0.507***	0.095	0.243
			(0.072)	(0.156)	(0.072)	(0.157)	(0.091)	(0.209)

	(1)		(2)		(3)		(4)	
	Dependent Variable		*Dependent Variable*		*Dependent Variable*		*Dependent Variable*	
	sd	*max–min*	*sd*	*max–min*	*sd*	*max–min*	*sd*	*max–min*
Asia					-0.127	-0.319*	-0.079*	-0.079**
					(0.080)	(0.174)	(0.045)	(0.095)
Europe					-0.125	-0.317*	-0.076*	-0.076**
					(0.078)	(0.170)	(0.044)	(0.094)
Lat. Am.					-0.076	-0.210	-0.120**	-0.302***
					(0.085)	(0.187)	(0.052)	(0.113)
No. Am.					-0.100	-0.272	-0.044	-0.145
					(0.079)	(0.171)	(0.045)	(0.096)
Oceania					0.040	0.028	0.055	0.065
					(0.113)	(0.250)	(0.084)	(0.188)
Cluster 1							0.041	0.038
							(0.025)	(0.026)
Cluster 2							0.051***	0.051***
							(0.019)	(0.019)
Cluster 3							0.063***	0.054***
							(0.018)	(0.019)
Cluster 4							0.010	0.013
							(0.014)	(0.014)
Cluster 5							-0.016	-0.008
							(0.013)	(0.013)
Cluster 6							0.151***	0.168***
							(0.036)	(0.036)
Cluster 7							0.047***	0.050***
							(0.013)	(0.014)

	(1)		(2)		(3)		(4)	
	Dependent Variable		*Dependent Variable*		*Dependent Variable*		*Dependent Variable*	
	sd	*max–min*	*sd*	*max–min*	*sd*	*max–min*	*sd*	*max–min*
Cluster 8							0.087***	0.088***
							(0.029)	(0.030)
Cluster 9							-0.003	0.001
							(0.017)	(0.018)
Cluster 10							-0.006	-0.012
							(0.034)	(0.034)
Cluster 11							0.055**	0.055**
							(0.022)	(0.022)
Cluster 12							0.009	0.011
							(0.024)	(0.026)
Cluster 13							0.096**	0.096**
							(0.040)	(0.039)
Cluster 14							0.020	0.024
							(0.018)	(0.019)
Cluster 15							0.017	0.012
							(0.021)	(0.023)
Adj. R^2	2.70	2.92	18.05***	17.36***	18.43*	17.70*	44.38***	43.70***

This table presents the results for the cross-sectional regressions on the SDG disagreement between the four rating agencies of the "All" sample. The dependent variables are standard deviation (*sd*) and the range of maximum and minimum (*max–min*) of the standardized SDG ratings per company, respectively. Models (1), (2), (3) and (4) display the results for regressions conducted with financial and accounting explanatory variables (1), plus region (2), plus industry (3), plus the 15 SDG clusters, respectively. Clustered standard errors with respect to region and industry are presented in parentheses. Log(MV), M/B, ROA and P/E denote logarithmized market value, market to book, return on assets and price-earnings ratio, respectively. Lat. Am. and No. Am. denote Latin America & the Caribbean and North America, respectively. Tech, BasM, ConD, ConS, Ener, Fin, Hc, Ind, RealE, and Util denote Technology, Basic Materials, Consumer Discretionary, Consumer Staples, Energy, Financials, Health Care, Industrials, Real Estate, and Utilities, respectively. ***, **, * denote significance at the 0.01, 0.05, 0.1 level, respectively. Adj. R^2 is multiplied by 100. M/B, ROA, Leverage, and P/E are winsorized at the 5% and 95% levels.

In the next step, we look at how the disagreement in single SDGs drives the overall SDG disagreement between the raters. To do so, we conduct further regression analyses to explain the variables *sd* and *max–min* with the same measures of disagreement in a single SDG. Since not all providers offer ratings at each individual SDG level, we use the cluster assignment approach of ISS and form 15 SDG clusters related to social or environmental aspects for each rater to not reduce the number of available observations for our analysis. For Vigeo Eiris, we take the scores of the 8 themes of the SDG Assessment framework and construct ratings for each single SDG by taking the arithmetic mean of all themes concerned with one SDG. The clusters are depicted in Table 7, with the first and second columns displaying cluster name and number, respectively. The third and fourth columns show the cluster designations by ISS and the matched SDGs from the remaining raters, respectively. SDGs 8, 9, and 17 are not assigned to any cluster. For each case in which more than one SDG is assigned to one cluster, we use the arithmetic mean of the effected SDGs to compute the cluster rating for each company. Then, we apply z-scoring to each cluster as outlined in section 2. We obtain 15 standardized SDG cluster ratings for each of our 1,246 companies in our "All" sample and calculate *sd* and *max–min* for each cluster in line with our aggregate disagreement measure.

Table 7: SDG clusters

No.	SDG cluster name	ISS variable	Inrate, MSCI, Vigeo Eiris SDG variables
1	No poverty	"Alleviating poverty"	1 (No poverty)
2	Zero hunger	"Combating hunger and malnutrition"	2 (Zero hunger)
3	Ensuring health	"Ensuring health"	3 (Good health and well-being); 6 (Clean Water & Sanitation)
4	Quality education	"Delivering education"	4 (Quality education)
5	Gender equality	"Attaining gender equality"	5 (Gender equality)
6	Providing basic services	"Providing basic services"	1, 3, 4, 6, 7, 10 (Reduced inequalities), 11
7	Peace, justice and strong institutions	Safeguarding peace"	16 (Peace, justice and strong institutions)
8	Achieving sustainable agriculture and forestry	"Achieving sustainable agriculture and forestry"	15 (Life on land), 2 (Zero Hunger)
9	Clean water and sanitation	"Conserving water"	6 (Clean water and sanitation)
10	Affordable and clean energy	"Contributing to sustainable energy use"	7 (Affordable and clean energy)
11	Sustainable cities and communities	"Promoting sustainable buildings"	11 (Sustainable cities and communities)
12	Responsible consumption and production	"Optimizing material use"	12 (Responsible consumption and production)
13	Climate action	"Mitigating climate change"	13 (Climate action)
14	Life below water	"Preserving marine ecosystems"	14 (Life below water)
15	Life on land	"Preserving terrestrial ecosystems"	15 (Life on land)

	Not assigned	8, 9, 17 (Decent work and economic growth; Industry, innovation and infrastructure; Partnership for the goals)

This table presents the mapping between the SDG clusters we use in the following analyses and the single SDGs. We follow the ISS SDG clusters. ISS provides us with SDG ratings for these clusters. For MSCI, Inrate, and Vigeo Eiris, we calculate new SDG cluster ratings based on the SDG ratings provided by these rating agencies and the mapping of the single SDG values to the SDG clusters. In clusters, in which more than one SDG dimension is assigned to the cluster, i.e., clusters 3 (Ensuring health), 6 ("Providing basic services") and 8 ("Providing basic services"), the SDG cluster ratings are calculated with the arithmetic mean of the single SDG ratings.

We use these cluster disagreement measures as independent variables along with the previously used financial variables, sectors and regions to investigate which cluster disagreements drive the aggregate disagreement controlling for common financial indicators, sector and region.

We present descriptive statistics in Table 8. The average disagreement in the clusters ranges between 0.6 and 0.8 in terms of *sd* and between 1.3 and 1.8 in terms of *max–min*[5] with the lowest disagreements in the clusters 4, 5, 7, 11 and 12. The disagreement in clusters 3, 6, 15 and 13 is somewhat higher. The standard deviation in the disagreement measures indicates that there is variation in the average disagreement across all clusters. For clusters 4, 5, 7 and 12, the maximum disagreement is quite high compared with other clusters, which, along with the high kurtosis values, indicates that the high average disagreement in these clusters might partly be driven by outliers. The skewness is positive for all cases, suggesting that there are more extreme large disagreements than small ones with respect to average disagreement.

Table 8: Descriptive statistics of the disagreement measures on SDG basis

	sd						*max–min*					
	Mean	SD	Min	Max	Skew	Kurt	Mean	SD	Min	Max	Skew	Kurt
1	0.70	0.67	0.17	8.07	4.04	26.34	1.57	1.45	0.35	16.88	3.83	23.59
2	0.67	0.74	0.08	5.43	2.72	8.44	1.47	1.61	0.18	12.13	2.73	8.69
3	0.74	0.53	0.11	4.37	1.55	3.31	1.66	1.17	0.24	9.61	1.54	3.27
4	0.65	0.73	0.15	12.35	6.91	86.04	1.44	1.56	0.33	26.05	6.47	76.15
5	0.63	0.56	0.08	11.02	10.41	170.56	1.37	1.19	0.18	22.98	9.56	150.31
6	0.76	0.51	0.06	4.52	2.47	10.67	1.70	1.13	0.13	9.71	2.41	10.04
7	0.64	0.74	0.09	10.01	5.23	39.30	1.39	1.57	0.20	20.93	5.09	37.77
8	0.74	0.63	0.11	6.96	4.89	34.37	1.67	1.39	0.24	15.63	4.67	31.80
9	0.74	0.58	0.11	6.16	4.16	26.01	1.68	1.28	0.22	12.71	3.86	22.04
10	0.71	0.58	0.07	3.26	1.24	1.16	1.53	1.25	0.14	7.03	1.24	1.20
11	0.65	0.72	0.12	8.46	4.14	27.22	1.41	1.54	0.27	17.10	4.04	25.56
12	0.63	0.71	0.05	14.53	8.50	129.43	1.35	1.51	0.12	29.28	7.95	111.05
13	0.80	0.51	0.06	3.04	1.05	0.68	1.79	1.12	0.14	6.83	1.06	0.85
14	0.72	0.64	0.10	7.43	5.12	37.54	1.59	1.36	0.23	15.20	4.73	32.28
15	0.75	0.60	0.09	4.82	3.59	17.24	1.67	1.30	0.20	10.73	3.45	16.06

[5] The correlations for each cluster pair are > 0.99 in all cases. Correlations are available upon request.

This table presents descriptive statistics of the variables used in the SDG cluster regression analysis for the companies in the "All" sample. The numbers in the first column refer to the 15 ISS SDG clusters depicted in Table 7. Descriptives statistics are presented for *sd* and *max–min*. Mean, SD, Min, Max, Skew, and Kurt represent arithmetic mean, standard deviation, minimum and maximum values, skewness and kurtosis, respectively.

Model (4) in Table 6 shows the results of the regression analysis for our model both for *sd* and *max−min.* Adding the cluster disagreements to the model increases the adjusted R^2 to 44.38 and 43.70 for *sd* and *max−min*, respectively. This increase is highly significant compared with model (3) as indicated by our F-test. As in model (1), the intercept term again becomes indistinguishable from zero. The cluster disagreements that influence our dependent variables are clusters 2, 3, 6, 7, 8, 11 and 13. Hence, the larger the disagreement between the raters in aspects related to "Combating hunger and malnutrition", "Ensuring health", "Providing basic services", "Safeguarding peace", "Achieving sustainable agriculture and forestry", "Promoting sustainable buildings" and "Mitigating climate change", the higher their disagreement in the overall SDG assessment of a company.

5. Implications of disagreement in SDG ratings for portfolio management

Our analysis so far shows the existence of a material disagreement between rating providers concerning the SDG alignment of a company. To gain further insight in the implications of the identified disagreement, we study the implications of using SDG ratings from disagreeing rating providers for portfolio management. Therefore, we conduct regressions using the 5-factor model from Fama and French (2015). Daily developed market factor returns are obtained from Kenneth French's data library[6]. We estimate regressions for each value-weighted quartile portfolio of each rating provider. To this end, we proceed as in the previous sections and use standardized SDG ratings to separate our matched sample of 1,246 companies into quartiles for each rating provider. This process results in four sets of four portfolios with average standardized SDG ratings increasing from portfolio 1 to portfolio 4.[7] In addition, we construct zero-investment (difference) portfolios by taking the difference of top and bottom portfolios, i.e., of the portfolios with the highest and lowest standardized SDG ratings. We compute daily value-weighted portfolio returns for the year 2021 and run Fama/French 5-factor models. Table 9 presents the coefficients of the regression models. Significance tests are based on Newey-West robust standard errors. The intercept coefficients are multiplied by 100. Almost all intercept terms are positive and highly significant. Thus, all portfolios yield positive abnormal risk-adjusted returns. This is since the "All" sample represents a subset of the market that performed better than the market in 2021 in general. Across all rating providers exposure to market risk declines from

[6] We thank Kenneth R. French for providing the data on https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[7] The average standardized SDG ratings per portfolio can be taken from Table 2.

portfolio 1 to 4, indicating that market risk is lower for high SDG companies. With respect to the remaining risk factors, the results are substantially different. Quartile portfolio 1 is exposed to small caps for Inrate and ISS, whereas it is exposed to large caps for MSCI and Vigeo Eiris. Quartile portfolio 4 has a large cap exposure for all rating providers except MSCI, which tilts towards small caps. For the HML factor, all lower quartile portfolios have a positive coefficient, suggesting that low SDG portfolios are mostly value stocks. For the upper quartiles, only Vigeo Eiris and ISS portfolios load significantly on HML with a negative coefficient signaling an exposure toward growth stocks. For the middle quartiles, the exposures towards HML are more unequivocal. Whereas the loadings tend to be positive and significant for MSCI and ISS, they tend to be indistinguishable from zero for Inrate and Vigeo Eiris and are even negative for some cases in absolute magnitude. Concerning the RMW factor, significant loadings are rather rare with the only significant and positive loadings on the factor for the upper quartiles with MSCI, Inrate and ISS SDG ratings. This shows that companies with high SDG ratings tend to be those with robust operating profitability. The loadings of the CMA factor are similar to those of the RMW factor, with fewer significant loadings than for the other factors. Most significant loadings are negative indicating exposure to companies that invest more conservatively. The loadings tend to become positive for upper quartiles, however, they mostly remain insignificant. Bottom quartiles never load significantly on the CMA factor.

For the regressions conducted with the difference portfolios in Panel B of Table 9 most portfolios have a positive and significant abnormal return. Therefore, using SDG ratings of MSCI, Vigeo Eiris and ISS, one could have achieved positive abnormal risk-adjusted returns with a strategy that was long in high SDG stocks and short in low SDG stocks for the investigated time series. Market risk of the long-short strategy portfolios are all significantly different from zero. This finding is in line with the results in Panel A, showing that there is a difference in market risk across quartiles. The difference portfolios based on SDG ratings from Inrate and ISS show an exposure towards large firms with values of −0.36 and −0.41, respectively, while the MSCI SDG rating portfolio has a positive and significant loading of 0.37 indicating an exposure to small caps. Consequently, depending on the rating provider one chooses, portfolios' systematic risks can be either towards large caps or towards small caps with fairly high exposures in both directions. E.g., the factor premium for SMB is −1.13% for the year 2021 and thus the difference in returns between the MSCI and Vigeo Eiris difference portfolios for the entire year is 0.88% only due to different size exposures. These different exposures could have even more tremendous return effects considering the quite low factor premium in comparison with the preceding years (3.4% and −6.15% for 2020 and 2019, respectively). For HML the risk exposures are quite similar direction-wise, i.e., negative across all rating providers, and less similar with respect to magnitude. Vigeo Eiris and ISS difference portfolios are highly exposed to growth stocks, whereas MSCI and Inrate portfolios are much less exposed to growth stocks. The loadings on RMW and CMA are each only

significant once: the MSCI and Vigeo Eiris portfolios are exposed to companies with robust operating profitability and a conservative investment style, respectively. Consequently, if one applies an SDG rating integration approach to a portfolio, depending on the rating provider, one is exposed to quite different risks and therefore also returns. Given the variety of rating providers, capital might be directed in opposing directions with the potential to even cancel out some effects completely.

Table 9: Regression results for value−weighted portfolios (daily)

Panel A: Quartile portfolios

	N	α	β_{MKT}	β_{SMB}	β_{HML}	β_{RMW}	β_{CMA}
				MSCI			
1	323	0.012***	0.844***	−0.270***	0.136***	0.026	−0.113
2	313	−0.001***	0.790***	−0.113**	0.087	0.018	−0.065
3	321	0.013***	0.782***	−0.078*	0.119***	0.026	−0.133*
4	289	0.024***	0.762***	0.100**	−0.053	0.193***	−0.080
				Inrate			
1	341	0.018***	0.839***	0.156***	0.359***	−0.023	−0.009
2	442	0.006***	0.876***	−0.158***	−0.035	0.070	−0.356***
3	151	0.024***	0.879***	−0.303***	−0.055	0.112	−0.084
4	312	0.004***	0.655***	−0.207***	0.114*	0.063	0.123
				Vigeo Eiris			
1	334	−0.009***	0.832***	−0.135***	0.285***	0.011	−0.059
2	499	0.016***	0.818***	−0.131***	0.070*	0.022	−0.276***
3	333	0.023***	0.832***	−0.096**	0.044	0.176***	−0.101
4	80	0.023***	0.631***	−0.240***	−0.284***	−0.050	0.281*
				ISS			
1	317	0.001***	0.818***	0.135***	0.369***	0.072	0.096
2	353	−0.004***	0.887***	−0.122***	0.279***	−0.032	−0.287***
3	265	0.004***	0.891***	−0.158***	0.109*	0.219***	−0.452***
4	311	0.032***	0.694***	−0.275***	−0.206***	0.017	0.110

Panel B: Difference portfolios

	N	α	β_{MKT}	β_{SMB}	β_{HML}	β_{RMW}	β_{CMA}
				MSCI			
4−1	612	0.011***	−0.082**	0.370***	−0.189***	0.167**	0.033
				Inrate			
4−1	653	−0.014***	−0.184***	−0.363***	−0.245***	0.086	0.132
				Vigeo Eiris			
4−1	414	0.032***	−0.201***	−0.105	−0.570***	−0.061	0.341**
				ISS			
4−1	628	0.031***	−0.123***	−0.410***	−0.575***	−0.055	0.014

This table shows the results for regressions with value−weighted quartile portfolios built with SDG rating data from the year 2020 for the companies in the "All" sample (1,246 companies). The regressions are conducted with daily data for the year 2021. The portfolios are rebalanced at the end of June. All returns are continuously compounded. Panel A shows regression results for quartiles 1 to 4. Panel B shows the regression results for the difference portfolios, where the high SDG rating portfolio is long and the low SDG rating portfolio is short (4−1). N, α, β_{MKT}, β_{SMB}, β_{HML}, β_{RMW}, β_{CMA} denote the number of equities, intercept of the regression and factor exposures, respectively. The intercept term is multiplied by 100. ***, **, * denote significance at the 0.01, 0.05 and 0.1 level, respectively. For quartiles 1 to 4, the null hypothesis for β_{MKT} is that β_{MKT} =1, all other coefficients are tested against 0.

6. Conclusion

Regulators expect the financial sector to make a significant contribution to achieving the SDGs by 2030. In this context, SDG ratings play a crucial role, because they enable investors to redirect their funds towards the SDGs. For corporate sustainability assessments such as ESG ratings, several studies (e.g., Berg et al. 2022, Chatterji et al. 2016, Christensen et al. 2022, Dimson et al. 2020, Dorfleitner et al. 2015) identified a large disagreement across different rating providers. Disagreement in SDG ratings makes it difficult for investors to clearly align their portfolio with the SGDs. While ESG rating heterogeneity can be explained by a missing common understanding of sustainability, SDG ratings could be expected to be aligned with a predefined and clear framework (i.e., the 17 SDGs). Therefore, this study analyzes whether SDG ratings overcome the disagreement found in ESG ratings and provide a (more) unambiguous SDG assessment of companies.

We find that the levels of (dis)agreement in SDG ratings is comparable to those of ESG ratings, with Krippendorff's alphas being well below the critical value of 0.67. Cross-sectional regression analysis reveals that the size of a company, its sector affiliation and the disagreement in certain SGDs are the main drivers of the disagreement found in aggregated SDG ratings. Furthermore, this identified disagreement leads to substantial differences in the financial outcome of SDG rating provider specific value-weighted quartile portfolios, which is quite important for investors. These portfolios differ in terms of their risk exposure to the size, the performance and the investment factors. Moreover, while some top-minus-bottom SDG rating quartile portfolios generate a significantly positive abnormal return, others have a significantly negative abnormal return.

Our study is the first, to the best of our knowledge, to comprehensively analyze the degree of (dis)agreement of SDG ratings across different rating providers, as well as the importance of this (dis)agreement for portfolio management. We provide substantial evidence that SDG rating providers do not have a unified understanding of the contribution of companies to the SDGs. This is problematic, since it calls into question the suitability of SDG ratings to align financial flows with the SDGs. The strategy of regulators to fill funding gaps in the achievement of the SDGs through private funds is thus compromised, since SDG ratings are unlikely to provide a reliable basis for identifying SDG contributions.

One possible solution to this SDG rating heterogeneity challenge could be a mapping of the EU taxonomy to SDGs. However, to date, this would only be feasible for the environmental aspects of SDGs, but not for their social aspects. However, this is difficult since several SDGs do not focus solely on environmental aspects which ultimately requires joint consideration of environmental and social aspects. Therefore, considerable improvements in the consistency of private rating providers are necessary, possibly based on extended or tightened legislation or standards of such agencies by governments or regulators. Given that this process is likely a both time-consuming and lengthy, a

preliminary remedy for the investment sector could be to consider as many providers as possible in any evaluation since this would help to average out any strong differences to a certain degree. In line with this, initial and less contentious regulation could mandate such a procedure in order to avoid any type of „cherry picking“ for any form of benign or desirable company ratings that could bias assessments. Such a regulatory “quick fix” in the interim would certainly help to achieve better targeting of financial flows to support SDG achievement. This would buy time for more comprehensive regulations whilst ensuring to help that private funds support the SDGs in the status quo.

Finally, we provide the following limitations to our paper and ideas for future research. To the best of our knowledge, our dataset provides the most extensive set of companies for which matching records of all involved rating agencies are available. However, given that rating agencies continuously extend the universe of rated companies, future research should aim to extend our research with a larger data set. Furthermore, future research could clarify if the more consistent findings identified in extant literature for the case of country-level ESG rating aggregations (e.g., Filippou and Taylor 2021) equally apply to SDG ratings. Given our extensive firm-level analysis (which would seem the appropriate initial focus), a more aggregated perspective was unfortunately beyond the scope of this study. It would nevertheless be a useful complementary analysis to more fully assess the current status of SDG ratings and to what degree they are comparable to the case of ESG ratings.

References

Avramov, Doron; Cheng, Si; Lioui, Abraham; Tarelli, Andrea (2022): Sustainable investing with ESG rating uncertainty. In *Journal of Financial Economics* 145 (2), pp. 642–664. DOI: 10.1016/j.jfineco.2021.09.009.

Berg, Florian; Kölbel, Julian F.; Rigobon, Roberto (2022): Aggregate Confusion: The Divergence of ESG Ratings. In *Review of Finance* 26 (6), pp. 1315–1344. DOI: 10.1093/rof/rfac033.

Brandts, Jordi; El Baroudi, Sabrine; Huber, Stefanie J.; Rott, Christina (2021): Gender differences in private and public goal setting. In *Journal of Economic Behavior & Organization* 192, pp. 222–247. DOI: 10.1016/j.jebo.2021.09.012.

Chatterji, Aaron K.; Durand, Rodolphe; Levine, David I.; Touboul, Samuel (2016): Do ratings of firms converge? Implications for managers, investors and strategy researchers. In *Strategic Management Journal* 37 (8), pp. 1597–1614. DOI: 10.1002/smj.2407.

Christensen, Dane M.; Serafeim, George; Sikochi, Anywhere (2022): Why is Corporate Virtue in the Eye of The Beholder? The Case of ESG Ratings. In *The Accounting Review* 97 (1), pp. 147–175. DOI: 10.2308/TAR-2019-0506.

Dimson, Elroy; Marsh, Paul; Staunton, Mike (2020): Divergent ESG Ratings. In *The Journal of Portfolio Management* 47 (1), pp. 75–87. DOI: 10.3905/jpm.2020.1.175.

Dorfleitner, Gregor; Halbritter, Gerhard; Nguyen, Mai (2015): Measuring the level and risk of corporate responsibility – An empirical comparison of different ESG rating approaches. In *Journal of Asset Management* 16 (7), pp. 450–466. DOI: 10.1057/jam.2015.31.

Dorfleitner, Gregor; Kreuzer, Christian; Sparrer, Christian (2022): To sin in secret is no sin at all: On the linkage of policy, society, culture, and firm characteristics with corporate scandals. In *Journal of Economic Behavior & Organization* 202, pp. 762–784. DOI: 10.1016/j.jebo.2022.08.027.

Fama, Eugene F.; French, Kenneth R. (2015): A five-factor asset pricing model. In *Journal of Financial Economics* 116 (1), pp. 1–22. DOI: 10.1016/j.jfineco.2014.10.010.

Filippou, Ilias; Taylor, Mark P. (2021): Pricing ethics in the foreign exchange market: Environmental, Social and Governance ratings and currency premia. In *Journal of Economic Behavior & Organization* 191, pp. 66–77. DOI: 10.1016/j.jebo.2021.08.037.

Gibson Brandon, Rajna; Krueger, Philipp; Schmidt, Peter Steffen (2021): ESG Rating Disagreement and Stock Returns. In *Financial Analysts Journal* 77 (4), pp. 104–127. DOI: 10.1080/0015198X.2021.1963186.

Han, Linsong; Li, Xun; Xu, Gang (2022): Anti-corruption and poverty alleviation: Evidence from China. In *Journal of Economic Behavior & Organization* 203, pp. 150–172. DOI: 10.1016/j.jebo.2022.09.001.

Krippendorff, Klaus (1998): Content analysis. An introduction to its methodology. 14. [print.]. Newbury Park: Sage (The Sage CommText series, 5).

Serafeim, George; Yoon, Aaron (2022): Stock price reactions to ESG news: the role of ESG ratings and disagreement. In *Review of Accounting Studies*, pp. 1–31. DOI: 10.1007/s11142-022-09675-3.

United Nations (2022a): Promotion and protection of human rights: human rights situations and reports of special rapporteurs and representatives. Available online at https://digitallibrary.un.org/record/3988295?ln=en, checked on 3/20/2023.

United Nations (2022b): Trillions needed to close finance gap on Sustainable Development Goals, says UN expert. United Nations. Available online at https://www.ohchr.org/en/press-releases/2022/10/trillions-needed-close-finance-gap-sustainable-development-goals-says-un, checked on 2/24/2023.

van Zanten, Jan Anton; Huij, Joop (2022): Corporate Sustainability Performance: Introducing an SDG Score and Testing Its Validity Relative to ESG Ratings. In *SSRN Journal*. DOI: 10.2139/ssrn.4186680.

CFR working paper series

centre for financial research cologne

CFR working papers are available for download from www.cfr-cologne.de.

2023

No.	Author(s)	Title
23-02	T. Bauckloh, J. Dobrick, A. Höck, S. Utz, M. Wagner	“In partnership for the goals”? The (dis)agreement of SDG ratings
23-01	F. Simon, S. Weibels, T. Zimmermann	Deep Parametric Portfolio Policies

2022

No.	Author(s)	Title
22-12	V. Agarwal, A. Cochardt, V. Orlov	Birth Order and Fund Manager’s Trading Behavior: Role of Sibling Rivalry
22-11	G. Cici, S. Gibson, N. Qin, A. Zhang	The Performance of Corporate Bond Mutual Funds and the Allocation of Underpriced New Issues
22-10	E. Theissen, C. Westheide	One for the Money, Two for the Show? The Number of Designated Market Makers and Liquidity
22-09	R. Campbell, P. Limbach, J. Reusche	Once Bitten, Twice Shy: Failed Deals and Subsequent M&A Cautiousness
22-08	M. Gehde-Trapp, L. Klingler	The Effect of Sentiment on Institutional Investors: A Gender Analysis
22-07	T. Bauckloh, V. Beyer, C. Klein	Does it Pay to Invest in Dirty Industries? – New Insights on the Shunned-Stock Hypothesis
22-06	J. Balthrop and G. Cici	Conflicting Incentives in the Management of 529 Plans
22-05	I. T. Ivanov, T. Zimmermann, N. W. Heinrich	Limits of Disclosure Regulation in the Municipal Bond Market
22-04	M. Ammann, A. Cochardt, S. Straumann, F. Weigert	Back to the Roots: Ancestral Origin and Mutual Fund Manager Portfolio Choice
22-03	A. Betzer, J. Gider, P. Limbach	Do Financial Advisors Matter for M&A Pre-Announcement Returns?
22-02	S. Lesmeister, P. Limbach, P.R. Rau, F. Sonnenburg	Indexing and the Performance-Flow Relation of Actively Managed Mutual Funds
22-01	T. Bauckloh, C. Klein, T. Pioch, F. Schiemann	Under Pressure: The Link between Mandatory Climate Reporting and Firms’ Carbon Performance

2021

No.	Author(s)	Title
21-11	V. Agarwal, H. Ren, K. Shen, H. Zhao	Redemption in Kind and Mutual Fund Liquidity Management
21-10	N.C. Brown, W. B. Elliott, R. Wermers, R. M. White	News or noise: Mobile internet technology and stock market activity
21-09	V. Agarwal, B. Barber, S. Cheng, A. Hameed, A. Yasuda	Private Company Valuations by Mutual Funds
21-08	T.G. Bali, H. Beckmeyer, M. Moerke, F. Weigert	Option Return Predictability with Machine Learning and Big Data
21-07	F. Chabi-Yo, M. Huggenberger, F. Weigert	Multivariate Crash Risk
21-06	V. Agarwal, H. Aslan, L. Huang, H. Ren	Political Uncertainty and Household Stock Market Participation
21-05	G. Cici, P. Zhang	On the Valuation Skills of Corporate Bond Mutual Funds
21-04	V. Agarwal, P. Hanouna, R. Moussawi, C.W. Stahel	Do ETFs Increase the Commonality in Liquidity of Underlying Stocks?
21-03	M. Saglam, T. Tuzun, R. Wermers	Do ETFs Increase Liquidity?
21-02	C. Andres, D. Bazhutov, D. Cumming, P. Limbach	Does Speculative News Hurt Productivity? Evidence from Takeover Rumors
21-01	T.G. Bali, F. Weigert	Hedge Funds and the Positive Idiosyncratic Volatility Effect

This document only covers the most recent CFR working papers. A full list can be found at www.cfr-cologne.de.

centre for financial research
cfr/university of cologne
Albertus-Magnus-Platz
D-50923 cologne
Fon +49(0)221-470-6995
Fax +49(0)221-470-3992
kempf@cfr-cologne.de
www.cfr-cologne.de