

A Work Project, presented as part of the requirements for the Award of a Master Degree in
Finance from the NOVA – School of Business and Economics.

**ESG RATINGS DIVERGENCE & DEVELOPING A PROPRIETARY ESG
SCORECARD**

LUCAS BONNIKE – 27794

A Project carried out on the Master in Finance Program, under the supervision of:

Gonçalo Sommer Ribeiro

January 3rd, 2020

Abstract

This paper analyses the divergence in Environmental, Social, and Governance (ESG) ratings between two leading providers: Sustainalytics and RobecoSAM. It finds that there are significant divergences in ratings between the two providers and that the rating correlation between them is much lower when compared to the correlation observed between credit rating agency scores. This applies to both their composite ESG ratings and their unidimensional Environmental, Social, and Governance scores. Additionally, this paper attempts to build a transparent, reliable, and cost-effective proprietary ESG scorecard by using a limited amount of data points, however the resulting scores are weakly correlated to the agency ratings.

Keywords

ESG Ratings; ESG Rating Agencies; Proprietary ESG Scorecards; ESG Scoring; Sustainable Investments; Environmental, Social, and Governance Performance; Corporate Sustainability

This work used infrastructure and resources funded by *Fundação para a Ciência e a Tecnologia* (UID/ECO/00124/2013, UID/ECO/00124/2019 and *Social Sciences DataLab*, Project 22209), *POR Lisboa* (LISBOA-01-0145-FEDER-007722 and *Social Sciences DataLab*, Project 22209) and *POR Norte* (*Social Sciences DataLab*, Project 22209).

Table of Contents

Introduction	3
Literature Review	5
Data	11
Methodology	16
Empirical Results	17
Conclusion	20
References	21

Introduction

In March 2019, the European Commission agreed on new rules related to sustainable investments and sustainability risks in an effort to “put environmental, social, and governance (ESG) considerations at the heart of the financial system” (European Parliament and Council, 2018). This new regulation includes a requirement for financial market participants to consider financially material ESG factors in their investment decision-making. Any firm operating in Europe that receives a mandate from their clients or beneficiaries to make investment decisions on their behalf is impacted, including Undertakings for Collective Investment in Transferable Securities (UCITS, or “mutual funds”), Alternative Investment Fund Managers (AIFMs, or “hedge funds”), as well as private equity and venture capital funds. This is yet another incentive for asset managers to incorporate ESG data in their investment decision-making process.

Regulatory concerns aside, there has also been a strong push from asset owners to increase the focus placed on ESG-related issues within the investment decision-making process. A worldwide survey of 118 large asset owners found that 84% are “pursuing or considering pursuing ESG integration in their investment process” and that 60% of those who have integrated ESG in their investment process have only been doing so for less than 4 years (Morgan Stanley, 2018). However, according to the same survey, 23% of asset owners cite the quality of ESG and sustainability data as the top challenge for sustainable investing, only second to the need for evidence of ESG-related financial performance (24%). The importance of this issue is corroborated by another survey which found that more than half (51%) of the 89 asset owners they surveyed said that the “lack of quality ESG data” was their biggest challenge in allocating to ESG-oriented investments (CFA Society New York, 2019). These results highlight the importance of easily available and quality ESG data to financial practitioners. Consequently, it is not surprising that the industry has seen an increase in ESG data offering and the emergence of so-called “ESG rating agencies” in recent years. These providers score

companies based on their E, S, and G performance and sell their evaluations for a fee, similarly to credit rating agencies in the credit risk space.

Although much of the sustainable investment literature has historically been focused on investigating the existence of a potential ESG risk factor and developing investment strategies based on this factor, research on the validity of ESG data sources is more recent. The aim of this paper is twofold: first, investigate whether evaluations made by ESG rating agencies are in line with each other or whether there are unexpected divergences which may reduce the relevance of these ratings; and second, investigate whether developing a proprietary ESG scoring model is feasible and whether it could be used as a replacement for traditional ESG rating services. The latter point is particularly important given that access to sustainability data from one of the large providers can cost up to \$50,000 (Thompson, 2019).

First, attention is placed on reviewing existing literature related to two ESG topics: (1) a review of research analysing evidence as to whether ESG performance is a relevant risk factor and whether investment strategies can be designed around this factor, or, in other words: “why should asset managers care about the quality of ESG metrics?”; and (2) a review of past research analysing issues with and divergences between ESG rating agencies, or, in other words: “are ratings really a straightforward option to evaluate ESG performance?”. Second, a computational comparison of the ratings of two leading ESG providers (Sustainalytics and RobecoSAM) is performed and the divergences between the two are analysed at various levels. Additionally, an attempt is made at creating a transparent and reliable ESG scoring methodology based on a limited number of alternative data points, and the resulting scores are compared to the ratings provided by Sustainalytics and RobecoSAM. An overview of the datasets, methodology, and results relevant to each of these two research points will be provided. Finally, propositions for further research will be offered alongside concluding comments.

Literature review

Before placing our attention on ESG ratings and understanding whether the different ESG rating agencies agree with each other, it is important to understand why an asset manager may need quality ESG metrics such as ratings in the first place. One of the early areas of focus of the ESG literature has been to investigate whether ESG factors could impact financial performance. This is the first topic discussed in this section.

An extensive review of more than 2,000 ESG-related empirical studies revealed that a large majority of historical research suggest a positive ESG and financial performance relationship (Friede, Busch, and Bassen, 2015). However, the relationship between ESG criteria and financial performance does not necessarily need to be positive for asset managers to pay attention to ESG metrics. Instead, and ignoring altruistic concerns, so long as this relationship is not proven to be insignificant, market participants should have an interest in data such as ESG ratings. However, for the sake of objectivity, one should also note the existence of several studies that show that active ESG selection of high- or low-rated stocks does not provide unequivocal superior risk-adjusted performance (Otten, Bauer, and Koedijk, 2002; Auer and Schuhmacher, 2016). Finally, it is important to consider that a potential relationship between ESG criteria and financial performance is not necessarily static in time and, just like for any other factors, this relationship may vary over time as any mispricing may be reduced or traded away once discovered (McLean and Pontiff, 2012).

A paper published by Nagy, Kassam, and Lee in 2016 evaluated two different equity trading strategies constructed using MSCI's ESG ratings, one of the leading providers of ESG data. Both strategies were benchmarked to the MSCI World Index with a sample period running between February 2007 and March 2015.

First, they constructed an “ESG Tilt” portfolio which was overweight on stocks with higher ESG ratings. This strategy assumes that ESG ratings are linked to future stock performance. The authors use two arguments to make this assumption: first, companies that integrate ESG considerations into their operations are less likely to sustain losses due to ESG issues such as environmental fines or labor disputes, and second, these companies are better able to make use of new ESG-related opportunities like clean technologies, which may lead to a competitive advantage reflected in stock performance. The authors found that almost 90% of the portfolio’s active risk was idiosyncratic and that this strategy delivered a positive active return over the sample period of 1.06% per year, of which 40% was attributable to specific risk (and consequently, possibly to ESG-related factors). On the other hand, a large portion of the active return was also attributable to style factors, which, according to the authors of the study, may be due to the fact that the ESG Tilt strategy led to a bias in favor of lower idiosyncratic volatility and mid-cap stocks, and to a tilt away from value stocks.

Second, they constructed an “ESG Momentum” portfolio which was overweight on stocks that increased their ESG rating during the last 12 months. Note, the sample period was shorter by a year due to this 12-month lag. According to the authors, this portfolio is intrinsically more short-term focused than the ESG Tilt strategy as the market may react quickly to a change in ESG rating while the advantages of a better-rated portfolio are expected to manifest themselves over the longer run. This strategy generated an active return of 2.23% annually, with about 80% of the active risk coming from stock-specific sources. Of those 223bps of active return, 132 were attributable to stock-specific return (which may possibly be ESG-related). Again, styles had a rather large contribution of 72bps while industry selection contributed for 44bps. According to the authors, this strategy led to a positive tilt to mid-cap stocks and the momentum factor. They argue that stocks that increased their ESG rating over the last 12 months also tended to perform

well over the last 12 months, therefore the ESG Momentum strategy captures some of the positive performance associated with the momentum factor.

The authors conclude that both model portfolios outperformed the MSCI World Index over the sample period while also increasing their ESG profile. In both cases, a significant portion came from stock-specific sources which could indirectly be attributed to ESG factors. However, it is important to note that both strategies led to significant systematic return contributions, so it appears that ESG factors may be intrinsically linked to other risk factors.

Another paper, published in 2016 by Ashwin Kumar et al., hypothesises that companies that integrate ESG practices into their operations are less vulnerable to reputation, political, and regulatory risk which leads to lower volatility of cash flows and profitability. In their words, “doing the right thing means you are less exposed in the long run”. They attempt to make a quantitative demonstration that ESG positive companies bear less risk compared to non-ESG stocks, which then results in higher risk-adjusted returns. To do so, they compared the members of the Dow Jones Sustainability Index (DJSI), one of the oldest and most recognised ESG indices that contains the “best-in-class” ESG stocks, to 809 companies that were not part of the DJSI and were instead representative of average market performance around the globe. It is interesting to note that members of the DJSI are selected based on the RobecoSAM ESG ratings, which are discussed alongside Sustainalytics ratings in a latter section. The authors formed 12 industry-specific portfolios to avoid industry bias, and equally weighted all stocks within each portfolio to eliminate market capitalisation bias. First, they found that high-ESG portfolios show lower volatility in their stock performances compared to their peers in the same industry, on average by 28.67% less annually. This applied to all 12 industries studied, however some industries were more strongly impacted by this effect, notably in the energy industry with a difference of almost 50.75% between ESG and non-ESG portfolios, annually. According to the authors, this difference is a risk premium that non-ESG companies face and that should be

considered by investors when making investment decisions. Since lower risk usually means less return, one may be tempted to think that the ESG portfolios provided less return. However, the authors found that even with lower risk, the ESG investments performed better than the non-ESG portfolios, by 6.12% annually, on average. The industries of energy, healthcare, and foods & beverage showed the highest advantage regarding the positive impact of good ESG practices on their return. It is interesting to note, however, that some industries were negatively impacted by better ESG practices, notably the automobiles, insurance, and banking industry. Combining the two sides of investing, risk and return, the authors found that the lower risk brought by better ESG practices actually improved the risk-adjusted return of the investments in these companies, or, in other words, resulted in higher Sharpe ratios by an average of 7.67%. This also applied to Treynor ratios, that were higher by an average of 11.81%. The authors conclude that ESG stocks tend to be less risky and offer more efficient risk-return profiles than non-ESG stocks.

These two articles show that investors should have an interest in ESG metrics (such as ESG ratings) as they may have an impact on financial performance. However, with the multitude of ESG rating agencies available, one may wonder whether their evaluations are worthwhile and whether they all agree with each other. Consistency between the ESG rating agencies would provide more weight to these measures and potentially motivate greater attention to ESG considerations. The remainder of this section is dedicated to reviewing literature investigating issues related to divergence between the ESG rating agencies.

The first paper discussed investigates the divergence of ESG rating across five leading rating agencies (Berg, Koelbel, and Rigobon, 2019). The authors found that the correlations between the five ESG agencies in their sample were on average 0.61 and ranged from 0.42 to 0.73, which is very low compared to correlations between major credit rating agencies such as Moody's and Standard & Poor's, which correlate at 0.99. This issue leads to three major

consequences according to them. First, ESG performance is unlikely to be properly reflected in financial markets if investors receive conflicting information from the different agencies. Second, companies are unsure how to improve their ESG performance as they received mixed signals from the different sources. Finally, the divergence of ratings also poses a challenge for empirical research as using one provider versus another may lead to vastly different results and conclusions. The authors' goal is not only to understand how much the ratings diverge between the agencies, but also understand why they diverge. Note, contrary to the research reviewed here, this paper ignores the underlying reasons and instead assumes, potentially naively, that ESG ratings should be in line as, after all, one should objectively be able to tell whether a company is doing well or poorly on environmental, social, and governance issues. Berg, Koelbel, and Rigobon (2019) attempt to explain the observed dispersion in ratings by decomposing it into three elements: (1) the scope of attributes used by the agencies that together constitute the ESG rating (or "scope divergence"), (2) the way these attributes are measured (or "measurement divergence"), and (3) the way the different attributes are aggregated to form the final ESG rating (or "weight divergence"). They found that 53% of the difference between rating agencies stems from measurement divergence, 44% is due to scope divergence, and 3% is due to weight divergence. Finally, the authors also found what they call a "rater effect", which means that a significant portion of the measurement divergence is rater-specific and not category-specific, i.e. a firm that performs well in one category for one rater is more likely to perform well in all the other categories for the same rater, and vice versa.

The last paper to be reviewed in this section articulates that ESG ratings have rarely been evaluated. Additionally, if the ratings are invalid, then trillions of dollars of capital are potentially being misallocated, and several academic studies may also be invalid (Chatterji et al., 2015). They assessed the agreement of six well-established rating agencies. The authors specifically used two elements to evaluate ratings: (1) "common theorisation", the conceptual

definition of what needs to be measured and why it matters, and (2) “commensurability”, the extent to which raters get similar answers when measuring the same constructs. The former is closely related to the concept of “scope divergence”, while the latter is similar to that of “measurement divergence”, both of which were defined in the aforementioned study (Berg, Koelbel, and Rigobon, 2019). The authors found that overall convergence was low, although similarities exist between raters based in the United States and those based in Europe, which suggests that geographically proximate raters may have similar theorisations and/or higher commensurability of ESG. Even after adjusting for theorisation differences, they found that commensurability was low, which meant that all or most raters had high measurement errors when trying to measure similar attributes. The authors concluded that users of ESG ratings should be cautious in drawing conclusions about firms based on this data. The authors also concluded that the rating agencies should validate their own data to ensure the effective measurement of ESG attributes, and urged academics to acknowledge issues with ESG ratings such as measurement errors, and potentially correct them using statistical methods to ensure the validity of their research.

This literature review helps understand two key points. First, ESG metrics may be important tools for asset managers as several studies have assessed the existence of a relationship between ESG and financial performance. Second, although ESG ratings are widely used by finance practitioners and in the academic world, it appears that the different agencies may not always agree on what constitutes good ESG behaviour and how to measure it.

The remainder of this paper is first dedicated to quantifying divergences between two leading providers, Sustainalytics and RobecoSAM. Second, this paper hypothesises that information contained in ESG ratings can potentially be replicated using a transparent ESG scorecard and using only a limited number of datapoints. After all, there exists several

techniques and “rules of thumb” to evaluate credit risk for unrated companies, and this paper hopes to find a similar procedure in the ESG space.

Data

Due to the growing interest in ESG, the sustainability rating market has grown considerably in the last decade (Escrig-Olmedo et al., 2019). ESG rating and information provider agencies (also called corporate social responsibility (CSR) ratings, social ratings, sustainability ratings, or SRI ratings agencies) have emerged in response to the growing demand for social and environmental information to better allocate capital in sustainable companies. ESG rating agencies assess the sustainability of companies based on information obtained from the companies themselves, usually through questionnaires, surveys, or public disclosures. In the last 10 years, the ESG rating agency market has gone through several phases of consolidation, with mergers and acquisitions being common amongst the existing ESG rating agencies (Escrig-Olmedo et al., 2019). One such example of acquisition is the absorption of several ESG research providers (RiskMetrics, ISS, Innovest, KLD Research & Analytics, and GMI Ratings) by MSCI to form their ESG branch, “MSCI ESG Research”. Other notable examples include the acquisition of Sustainalytics by Morningstar (a well-established data provider in the area of funds) in 2010, and the purchase of Vigeo-Eiris by Moody’s in early 2019 (Berg, Koelbel, and Rigobon, 2019). Additionally, in November 2019, RobecoSAM agreed to transfer their ESG ratings business to S&P Global. Finally, many of the traditional financial information providers such as Bloomberg and Thomson Reuters have started offering ESG data to their clients. Currently, there exist six leading providers of ESG metrics: MSCI, Sustainalytics, RobecoSAM, Bloomberg, Thomson Reuters, and Vigeo-Eiris. This research focuses on two of these six, namely Sustainalytics and RobecoSAM, due to the fact that a large sample of their ratings are freely available through Bloomberg’s database. Moreover, both

ratings are provided as percentile scores on the Bloomberg Terminal, which makes the comparison easier.

RobecoSAM gathers information from over 2,500 companies around the world through an annual survey, the RobecoSAM Corporate Sustainability Assessment (CSA) since 1999 (RobecoSAM, 2019). This survey forms the research backbone for the construction of the Dow Jones Sustainability Indices, which are widely used as financial benchmarks in the ESG space. According to them, gathering information from the company directly allows for a much deeper analysis of sustainability than frameworks based on public disclosure alone. This survey explores three different dimensions, which they call Environmental, Social, and Economic. All three dimensions cover both industry-specific and financially relevant sustainability criteria. These intangible factors have been chosen due to their clear correlation to past financial performance (RobecoSAM, 2019). Companies are allocated more or less points based on their response to the survey, and these points are weighted and summed to form both unidimensional (E, S, and G) scores and a final aggregate score, the “Total Sustainability Score” ranging from 0-100 (RobecoSAM, 2019). The Environmental dimension includes themes such as environmental reporting, operational eco-efficiency, biodiversity, and climate strategy. The Social dimension includes topics such as corporate citizenship and philanthropy, human capital development, talent attraction and retention, and occupational health and safety. Finally, the Economic dimension covers corporate governance themes (it thus corresponds to the “G” of ESG) such as business ethics, risk and crisis management, and tax strategy (RobecoSAM, 2019). RobecoSAM’s unidimensional and total scores are then transformed into percentile ranks which are freely available to Bloomberg Terminal users. Therefore, a score of 99 means that the company is in the top 1% of companies in terms of E, S, G, or overall ESG performance.

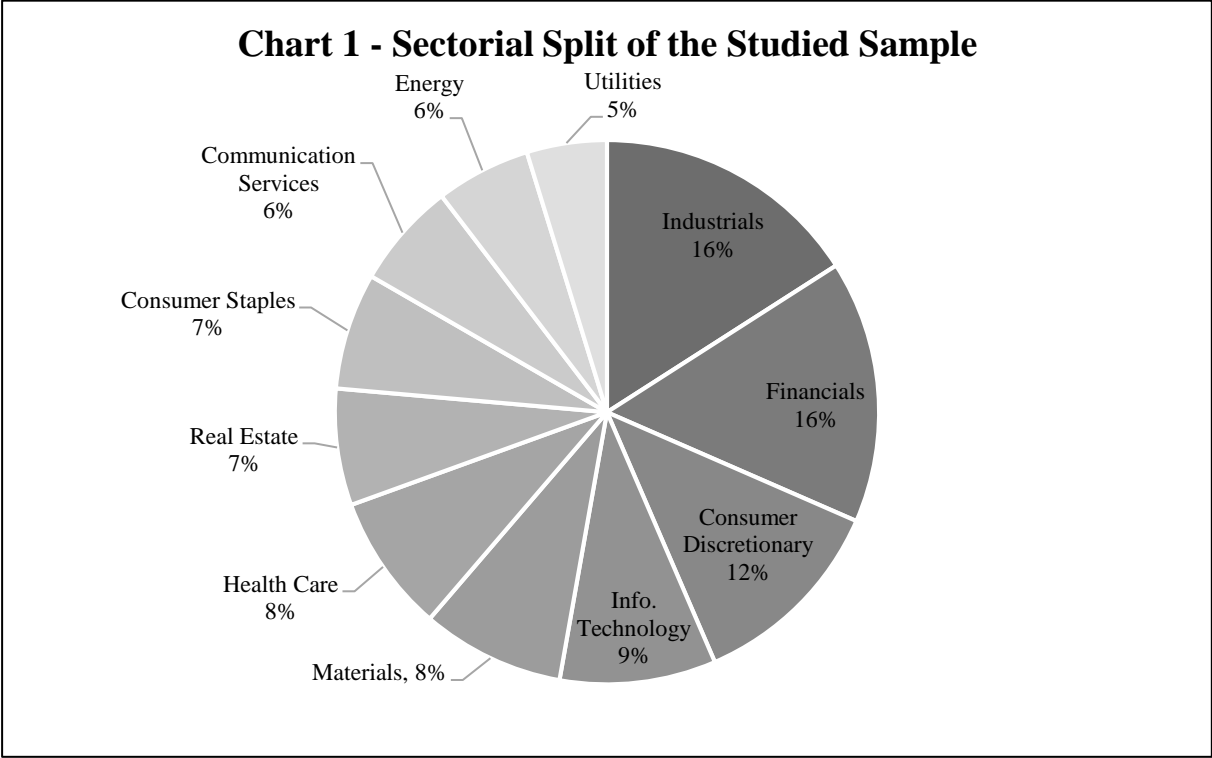
Sustainalytics has been a leading provider of ESG products and services for over 25 years (Sustainalytics, 2019). They provide an “ESG Risk Rating” for over 12,000 companies.

This Risk Rating represents units of unmanaged ESG risk, with lower scores representing less unmanaged risk and, consequently, better ESG practices. Therefore, ratings provided by Sustainalytics have two dimensions: “exposure”, how much the company is exposed to ESG issues, and “management”, how well the company manages these issues (Sustainalytics, 2019). Based on these scores, companies are grouped into one of five risk categories (negligible, low, medium, high, severe) which can be used to make cross-sectoral comparisons of companies (Sustainalytics, 2018). Similarly to RobecoSAM, Sustainalytics emphasises that the issues considered within the ESG Risk Ratings are selected based on their materiality. They also take into account issues that are considered material from an ESG perspective in the long-run, even if the financial consequences are not fully measurable today (Sustainalytics, 2019). The ratings are made up of three building blocks: corporate governance, material ESG issues (MEIs), and “idiosyncratic ESG issues”. MEIs include, for example, themes such as “Human Capital” which themselves include topics such as employee recruitment, development, diversity, and labour relations, as well as various environment-related themes (Sustainalytics, 2019). According to Sustainalytics, idiosyncratic ESG issues are those issues that are unpredictable or unexpected, and unrelated to the industry or business model, such as accounting scandals. Based on how much risk a company is exposed to and how well it manages this risk, Sustainalytics aggregates and weighs each theme to come up with an overall ESG score. They also compute E, S, and G cluster scores. A sample of the 12,000 covered companies is available to Bloomberg Terminal users as percentile scores.

In addition to RobecoSAM and Sustainalytics’ ratings that are retrieved via Bloomberg, the following research also makes use of data published by Bloomberg themselves. First, it uses Bloomberg’s “disclosure scores”. These scores are not ESG ratings; instead, they are an indication of how much ESG data the companies disclose. Scores range from 0.1 for companies that disclose a minimum amount of data to 100 for those that disclose every data point collected

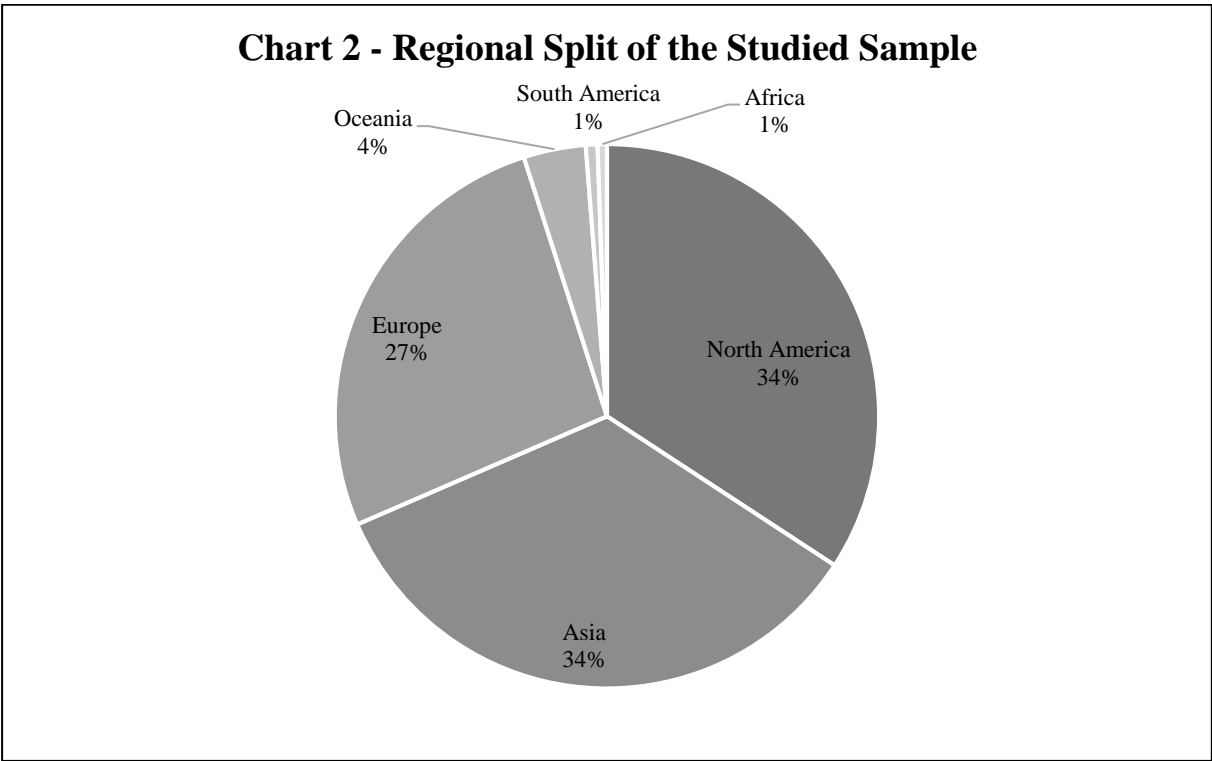
by Bloomberg. Each datapoint is weighted in terms of importance (e.g. “Greenhouse Gas Emissions” has a greater weight than other disclosures). Finally, Bloomberg makes available a wide range of ESG-related datapoints (e.g. water usage and waste production, workforce diversity figures, information on compensation and board membership, etc.). These datapoints are used in a latter part of this research in an attempt to replicate ESG ratings in a transparent and straightforward way.

The sample studied here consists of 2,165 firms for which both the RobecoSAM and Sustainalytics ratings are available on the Bloomberg Terminal, domiciled in close to 50 different countries and spanning 11 different sectors. The sectorial split, as can be seen in the below chart (Chart 1), is roughly equal although Industrials and Financials companies are slightly more represented, each accounting for about 16% of the total sample.

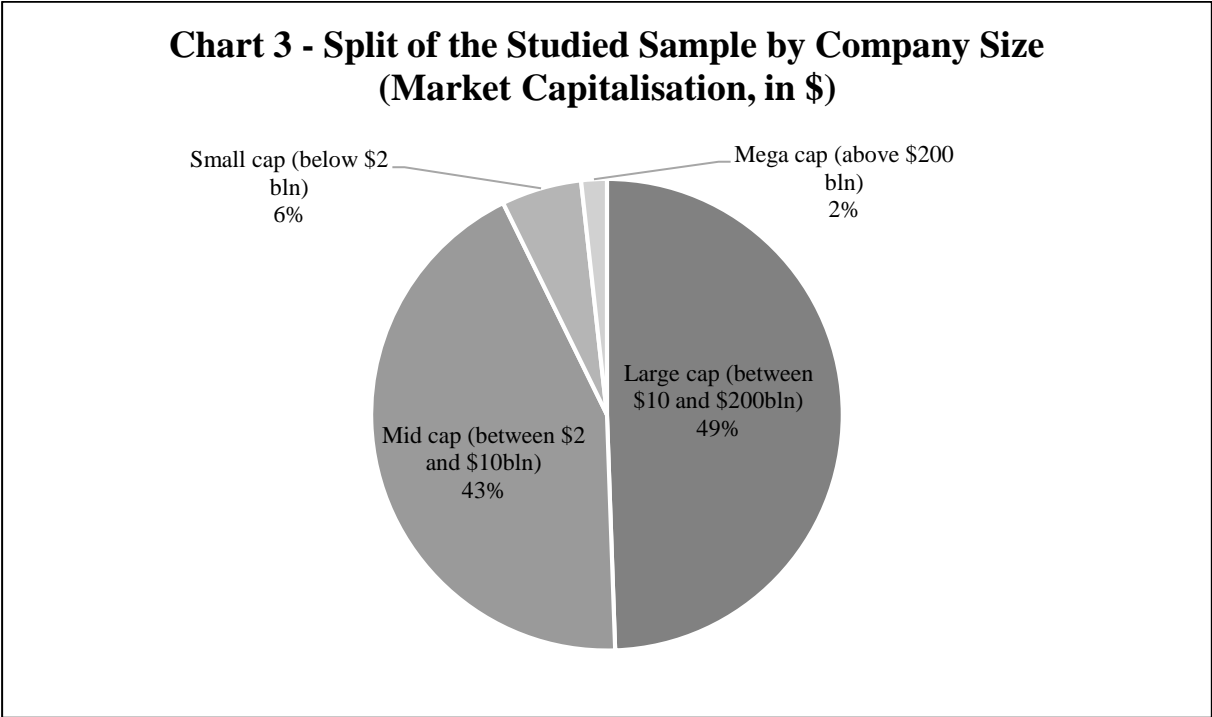


Additionally, a geographical split is shown in Chart 2. Almost 95% of the companies studied here are domiciled either in North America, Asia, or Europe. It is unclear whether this split is representative of the coverage of the two chosen rating agencies, or if larger companies (which

are more likely to be found in one of the three continents mentioned above) are prioritised in the publication of the agencies' ratings on the Bloomberg Terminal.



Finally, the chart below (Chart 3) shows a split by company size. Small cap companies seem to be underrepresented in the sample.



Methodology

This section describes the methodology used to compare the two ESG rating agencies as well as an attempt at creating a proprietary ESG scorecard.

First, statistical information on ESG ratings contained in the sample are presented. A Pearson correlation coefficient is computed between RobecoSAM and Sustainalytics scores. This is done at the ESG level, but also at the unidimensional E, S, and G levels. Additionally, low correlation may be due to the fact that the agencies have slight disagreements on all companies, or it could be due to a few very large disagreements. To get a better view on this, the correlation between the 100 firms with the lowest and highest divergence at the ESG level are analysed separately and Pearson correlation coefficients are presented for those. Finally, the Bloomberg “ESG disclosures scores” are used to understand whether the rating agencies disagree on companies that disclose less ESG information (i.e. that have a low disclosure score) or whether the disagreements are independent of the quantity of data disclosed by the firms.

Second, this paper hypothesises that the ratings can be replicated using a limited number of datapoints. Nine indicators, three belonging to each category (E, S, and G) are chosen as follows: Environmental (total greenhouse gas emissions per unit of sales, total waste generated per unit of sales, and water consumed per unit of sales), Social (binary values representing whether: (1) the firm has established ethical guidelines and/or a compliance policy for its employees, (2) the firm has disclosed efforts to improve the management of employee health and safety, and (3) whether the company has implemented any initiatives to reduce social risks in its supply chain such as child labour), and Governance (percentage of non-executive directors on board, percentage of female executives on board, and total executive compensation as a percentage of operating expenses). Unfortunately, due to the lack of available data the sample is reduced to around 450 firms that have all the required datapoints. Once all the datapoints

have been collected, an attempt is made to use a simple 1/3 weighting in each category to establish E, S, and G ratings, where each positive datapoint (e.g. % of female executives on board) counts positively towards the category score and each negative data point (e.g. greenhouse gas emissions per sales) counts negatively towards the category score. An overall ESG score is also computed using 1/3 weighting between the unidimensional scores. Finally, the set of custom scores are transformed into percentile scores to be compared to Sustainalytics and RobecoSAM’s ratings.

Empirical Results

The table below (Table 1) shows descriptive statistics of the Sustainalytics and RobecoSAM ratings sample.

Table 1 – Descriptive Statistics of ESG Ratings in the Studied Sample

	Sustainalytics	RobecoSAM
Observations	2,165	2,165
Mean	50.84	42.74
Standard Deviation	29.19	30.68
Minimum	0.00	0.00
Median	50.98	39.00
Maximum	100.00	100.00

Table 2 shows the correlations between Sustainalytics and RobecoSAM’s ratings at the ESG, E, S, and G levels. The correlations are in line with Berg, Koelbel, and Rigobon’s findings (2019), although it seems that Sustainalytics and RobecoSAM are more correlated than average. As a reminder, they previously found that the correlations between the ESG agencies in their sample were on average 0.61 and ranged from 0.42 to 0.73.

Table 2 – Correlations Between Sustainalytics and RobecoSAM's Ratings

Dimension	Correlation
ESG	72%
E	69%
S	63%
G	60%

However, the correlations are much lower than the correlation between credit rating agencies such as Moody's and Standard & Poor's, which correlate at 0.99 according to Berg, Koelbel, and Rigobon (2019).

To better understand where the disagreements come from, the sample is split further into two groups, the 100 firms with the highest agreement (computed as the absolute difference between the ESG scores of the two providers) and the 100 firms with the lowest agreement between the two providers at the ESG level.

Large firms such as Nestlé SA, SAP SE, Philip Morris International, and Bayer AG are found in the group with the highest agreements, but the sample also contains small and mid-cap stocks. The average absolute disagreement in this group is only of 0.48 percentage points, and the correlations between the two sets of ESG ratings is close to 0.99, which indicates that when rating agencies agree on an assessment, they agree very strongly.

Interestingly, the group composed of the 100 firms with the lowest agreement also contains large companies such as Intel Corp., Verizon Communications Inc., Honda Motors Co Ltd and Mondelez International Inc. Company size does not seem to be a good predictor of agreement or disagreement between the ESG ratings agency as both groups have a median market capitalisation of around \$8.5bln. This is comparable to (although slightly below) the median size of the full sample which is around \$10bln. The average absolute disagreement in the second group is of 58 percentage points, indicating that when the rating agencies disagree, they disagree

very strongly. Therefore, it seems that the relatively low correlation between Sustainalytics and RobecoSAM's ratings may be impacted by large disagreements rather than many small disagreements. This may be due to different theorisations of what constitutes good ESG behaviour, as shown by previous research (Berg, Koelbel, and Rigobon (2019); Chatterji et al., 2015).

Finally, it seems like the amount of ESG data disclosed by companies, as measured by the Bloomberg disclosure scores, is only very weakly correlated to the extent to which the rating agencies agree or disagree. This is true at the ESG level (correlation of 0.10 between the Bloomberg ESG disclosure scores and the absolute score difference) but also at the E, S, and G level (correlation of 0.07, 0.07, and 0.08, respectively). Again, this might point to divergences in the theorisation of ESG rather than an issue with the measurement of the data, as the agencies disagree even when they have access to a large amount of ESG information to make their assessment.

The remainder of this section is dedicated to presenting the results of the attempt at creating a transparent ESG scorecard using raw ESG datapoints. Again, the sample only contains 450 firms this time, as the 9 indicators chosen are not available for all 2,165 firms of the sample on Bloomberg (however, they were also partly picked based on availability, so one can assume that they are some of the most widely available ESG datapoints available). This indicates that ESG rating agencies must have access to more data than what is publicly available, which indicates added value.

Unfortunately, the 1/3 weighting methodology does not yield good results. The correlations between the resulting percentile scores and Sustainalytics and RobecoSAM's ratings are presented in Table 3. Note, however, that the correlation between Sustainalytics and

RobecoSAM’s ESG ratings in the reduced sample of 450 sizes is lower than the full sample, at only 60%.

Table 3 – Correlations Between Custom 1/3 Scores and Sustainalytics/RobecoSAM Ratings

Dimension	Sustainalytics	RobecoSAM
ESG	24%	19%
E	16%	19%
S	25%	25%
G	19%	7%

These low correlations show that the ESG ratings cannot be easily approximated using only 9 equally weighted ESG metrics. It remains unclear whether one could get better results by using a different set of datapoints or more advanced statistical methods due to the lack of publicly disclosed data available on Bloomberg. The sample was already drastically reduced from 2,165 to 450 firms with 9 metrics, so adding more metrics would only further reduce the sample size and thus the predictive power of this research.

Conclusion

This paper has shown that the need for quality ESG assessments is of growing importance to the financial industry. Unfortunately, this research has also shown that ESG rating agencies are weakly correlated, which corroborates the findings of previous research (Berg, Koelbel, and Rigobon, 2019; Chatterji et al., 2015) and casts a doubt on the validity of ESG ratings. However, an attempt at creating a custom ESG scorecard using a limited set of datapoints and a transparent weighting methodology did not succeed in replicating the ESG ratings provided by the ESG agencies. Moreover, the amount of publicly disclosed information

remains scarce and out of 2,165 firms that were rated by both Sustainalytics and RobecoSAM, only 450 had information for all 9 ESG metrics used to compute the custom ESG scores. Therefore, ESG ratings do seem to provide added value in terms of coverage of companies that do not publicly disclose much ESG information, and they must have access to information that is not easily available elsewhere.

However, this paper and prior research have shown that even the concept of what “good ESG performance” is does not seem to be aligned amongst the rating agencies as they are not able to agree on ESG assessments even when the companies disclose a wide range of ESG-related information. Therefore, further research should first be made in an attempt to standardize the concept of ESG and provide a stable assessment framework for companies, ESG rating agencies, and financial practitioners across the globe. In 2017, the United Nations-supported Principles for Responsible Investment (UNPRI), a global network of investors, published their own “Blueprint for Responsible Investment” which aims to align ESG expectations. Today, more than \$70tn of assets are managed by close to 1,200 signatories of the PRI, including market leaders such as BlackRock, Vanguard, or State Street (UNPRI, 2017). This should hopefully lead to more standardization and alignment on the part of the ESG rating agencies, which will ensure that all market participants are able to evaluate ESG opportunities correctly.

References

Ashwin Kumar, N. C., Camille Smith, Leïla Badis, Nan Wang, Paz Ambrosy, and Rodrigo Tavares. 2016. "ESG Factors And Risk-Adjusted Performance: A New Quantitative Model". *Journal Of Sustainable Finance & Investment* 6 (4): 292-300.

Auer, Benjamin R., and Frank Schuhmacher. 2016. "Do Socially (Ir)Responsible Investments Pay? New Evidence From International ESG Data". *The Quarterly Review Of Economics And Finance* 59: 51-62.

Berg, Florian, Julian Koelbel, and Roberto Rigobon. 2019. "Aggregate Confusion: The Divergence of ESG Ratings". *MIT Sloan Research Paper* 5822 (19).

CFA Society New York. 2019. "Asset Owners Making Significant Allocations To ESG With More Capital On Their Way, According To CFA Society New York Survey." <https://www.cfany.org/2nd-esg-summit-press-release/>

Chatterji, Aaron K., Rodolphe Durand, David I. Levine, and Samuel Touboul. 2015. "Do Ratings Of Firms Converge? Implications For Managers, Investors And Strategy Researchers". *Strategic Management Journal* 37 (8): 1597-1614.

Escrig-Olmedo, Elena, María Fernández-Izquierdo, Idoia Ferrero-Ferrero, Juana Rivera-Lirio, and María Muñoz-Torres. 2019. "Rating The Raters: Evaluating How ESG Rating Agencies Integrate Sustainability Principles". *Sustainability* 11 (3).

European Parliament and Council. 2018. "Proposal for a regulation of the European Parliament and of the Council on disclosures relating to sustainable investments and sustainability risks". COM (2018) 354 final.

Friede, Gunnar, Timo Busch, and Alexander Bassen. 2015. "ESG And Financial Performance: Aggregated Evidence From More Than 2000 Empirical Studies". *Journal Of Sustainable Finance & Investment* 5 (4): 210-233.

McLean, R. David, and Jeffrey E. Pontiff. 2012. "Does Academic Research Destroy Stock Return Predictability?". *Journal Of Finance*.

Morgan Stanley. 2018. "Asset Owners Embrace Sustainability ". *Sustainable Signals*.
<https://www.morganstanley.com/assets/pdfs/sustainable-signals-asset-owners-2018-survey.pdf>

Nagy, Zoltán, Altaf Kassam, and Linda-Eling Lee. 2016. "Can ESG Add Alpha? An Analysis Of ESG Tilt And Momentum Strategies". *The Journal Of Investing* 25 (2): 113-124.

Otten, Roger, Rob Bauer, and Kees C.G. Koedijk. 2002. "International Evidence On Ethical Mutual Fund Performance And Investment Style". *LIFE Working Paper* 02 (59).

RobecoSAM. 2019. "RobecoSAM's Corporate Sustainability Assessment Methodology". *Measuring Intangibles*.
https://assessments.robecosam.com/documents/measuring_intangibles_csa-methodology.pdf

Sustainalytics. 2019. "Methodology". *The ESG Risk Ratings*.

<https://www.sustainalytics.com/esg-ratings/> (see "Methodology")

Sustainalytics. 2018. "Moving up the innovation curve". *The ESG Risk Ratings*.

<https://www.sustainalytics.com/esg-ratings/> (see "White paper")

Thompson, Jennifer. 2019. "ESG Rating Agencies Fulfil The Need For Knowhow". *Financial*

Times. <https://www.ft.com/content/2cd37df8-a973-3f94-b498-09ee1a6ba53b>

UNPRI, 2017. "Genesis, Assumptions and Process". *A Blueprint for Responsible Investment*.

<https://www.unpri.org/download?ac=1919>