Inside the ESG ratings: (Dis)agreement and performance

Monica Billio1 | Michele Costola1 | Iva Hristova1 | Carmelo Latino2 | Loriana Pelizzon3

1Ca’ Foscari University of Venice, Venice, Italy
2Leibniz Institute for Financial Research SAFE, Frankfurt am Main, Germany
3Leibniz Institute for Financial Research SAFE, Goethe University Frankfurt, Ca’ Foscari University of Venice and CEPR, Venice, Italy

Correspondence
Monica Billio, Ca’ Foscari University of Venice, Venice, Italy.
Email: billio@unive.it

Funding information
Leibniz Institute for Financial Research SAFE, Frankfurt: European Investment Bank (EIB) for the project “ESG-Credit.eu - ESG Factors and Climate Change for Credit Analysis and Rating”

Abstract
We analyze the ESG rating criteria used by prominent agencies and show that there is a lack of a commonality in the definition of ESG (i) characteristics, (ii) attributes and (iii) standards in defining E, S and G components. We provide evidence that heterogeneity in rating criteria can lead agencies to have opposite opinions on the same evaluated companies and that agreement across those providers is substantially low. Those alternative definitions of ESG also affect sustainable investments leading to the identification of different investment universes and consequently to the creation of different benchmarks. This implies that in the asset management industry it is extremely difficult to measure the ability of a fund manager if financial performances are strongly conditioned by the chosen ESG benchmark. Finally, we find that the disagreement in the scores provided by the rating agencies disperses the effect of preferences of ESG investors on asset prices, to the point that even when there is agreement, it has no impact on financial performances.

KEYWORDS
Corporate Social Responsibility, ESG rating agencies, sustainable investments

JEL CLASSIFICATION
M14; G24; G11

1 | INTRODUCTION

According to Milton Friedman, the main responsibility of a company is the maximization of the shareholders’ returns. For decades, environmental, social and governance (ESG) responsibilities were not considered relevant by most of the companies that have been focusing on profit maximization. Not only were ESG responsibilities believed to merely have no incidence on financial performance, but they were also perceived as a potential burden to the latter, being related to cost increases.

Nevertheless, in the last twenty years, environmental, social and governance issues revealed their influence not only on the profitability, but also on the financial viability of several firms. As a natural consequence, the process of asset allocation started evolving. Furthermore, a raising environmental, social and governance consciousness has been observed worldwide. This trend appeared as a result of the increased occurrence of extreme weather events,1 damaging infrastructures and perturbing global markets, but also because of the 2008 financial crisis, which affected both private and public sectors. Indeed, the effect of the subprime crisis has been threefold: firstly, it highlighted the relevance of investors’ decisions and therefore their inherent role; secondly, it raised the public consciousness in terms of social responsibility; thirdly, it stresses the importance of good governance practices.

Thus, even though socially responsible investment (SRI) has existed since 1920s, only recently, it has only recently experienced a
considerable surge of interest and has become a general preoccupa-
tion rather than a niche investment practice. Consequently, we have
observed the development of a new segment of the rating agencies’
market and the big three credit rating agencies (Moody’s, S&P, Fitch)
have started to include ESG evaluation in their ratings.

Nevertheless, in this new, insufficiently regulated and rapidly
expanding sector, investors, corporate managers and policy makers
need a deeper understanding of the ESG inherent particularities and a
wider knowledge of the potential impacts of ESG on the real and
financial sectors. In response, the academic research is developing at a
tremendous pace.

If a classification of this quickly expanding research was pro-
posed, four major scientific orientations could be identified. First,
a major part of the literature is aiming to investigate economic and
financial performance of ESG stocks and portfolio. In particular,
the literature concentrates on the ESG portfolio profitability, the
impact of ESG on firms’ financial performance, and the effect of
ESG scores on credit ratings. A second strand of the literature
studies and evaluates the particularities of the ESG rating score and
the adopted methodologies highlighting their disagreement.

To our knowledge, there are no papers that relate these two
aspects and investigate the implications that ESG rating disagreement
might have on ESG portfolios performance.

The present study aims to cover this void. We contribute to the
existing literature by (i) investigating the disagreement among ESG
rating agencies in terms of scores and its effects in the identification
of the constituents of ESG indexes in terms of constituents overlap
(ESG agreement portfolio) and (ii) studying the performance of the
ESG agreement with respect to non-ESG portfolios.

Firstly, we analyze the rating criteria used by nine prominent
agencies and show that there is a lack of a common characteristics,
attributes and standards in defining the E, S and G components.
Secondly, we show that the heterogeneity in the ESG industry can
lead the agencies to have opposite opinions on the evaluated
companies and that agreement, among rating agencies, is relatively
low. Moreover, we show that alternative definitions of ESG also
affect the benchmarks identification used by the sustainable invest-
ments’ industry. In relation to this, we analyse four representative
ESG indexes and show that the overlap of their common constituents is
approximately 15% (the ESG agreement portfolio). Thirdly, we
show that the low overlap of the ESG indexes (due to the disagreement in the
scores provided by the rating agencies) dis-
 perses the effect of preferences of ESG investors on asset prices, to
the point that, even for the ESG agreement portfolio, there is no
impact on the financial performances.

The paper is organized as follows. Section 2 introduces the con-
cept of ESG investing, and presents the ESG investing landscape ev-
olution. Section 3 describes the industry of ESG ratings and proposes
an assessment of their reliability including a structural comparison of
four ESG indexes. Section 4 investigates the implication of disagree-
ment, among the rating providers, on the performance of ESG and
non-ESG portfolios. Finally, the last section concludes.

2 | THE CURRENT ESG STATE OF THE ART

The term ESG was officially coined in 2004 with the publication of the
report “Who Cares Wins” by the UN Global Compact Initiative
(UN, 2004). It set the ambitious goal to regroup three of the main
ethical finance pillars: environmental, social and governance. All of them
encompass different issues and present a specific assessment target.

The environmental pillar focuses on issues such as climate change,
deforestation, air and water pollution, land exploitation and biodiver-
sity loss. Therefore, it evaluates the efforts of a company in terms of
energy efficiency, greenhouse gas emissions, waste, water and
resource management. A large part of the literature has thus naturally
tried to define the relationship between environmental and financial
performance. Derwall et al. (2004) showed that more eco-friendly
firms were benefiting from higher stock returns than their less eco-
friendly counterparts. These findings remain significant even after
several methodological controls. Manrique and Martí-Ballester (2017)
also provide similar conclusions on the basis of a sample including
2982 large firms from both developed and developing countries.

The social pillar includes aspects related, for instance, to gender
policies, protection of human rights, labor standards, workplace and
product safety, public health and income distribution, which are all
affecting employees’ satisfaction. According to Edmans (2011) there
exists a clear positive relationship between employee’s satisfaction
and long-run stock return. Namely, American companies considered
as proposing the best working condition have earned a 4-factor alpha of
3.5% per year (2.1% above the industry benchmark) in the period

At last, the governance pillar is related to aspects such as the inde-
pendence of the board of administration, shareholders’ rights, man-
gers’ remuneration, control procedures and anti-competitive practices,
as well as the respect of the law. Several studies tend to stress the
significant positive impact of these practices, like Gompers et al. (2003),
Tarmuji et al. (2016) and Velte (2017). While Tarmuji et al. (2016)
consider companies from Malaysia and Singapore, Velte (2017) focuses on
German corporations, and Gompers et al. (2003) study US firms. These
three analyses highlight the positive impact of stronger governance
practices on companies’ profitability.

The relevance of ESG on companies’ financial performance and
profitability represents one of the major strands of the current litera-
ture. A large majority of studies, concerning both industrialized coun-
tries and emerging economies, stresses the positive impact of ESG
efforts and disclosure on firms’ financial performance.

For instance, Zhao et al. (2018) focus on Chinese listed power
generation firms and find that good ESG performance can improve
KLD STATS data from 2000 to 2016, provide evidence of the positive
impact of ESG on US companies’ profitability measured through ROA,
especially for the banking sector. Ortas et al. (2015), using the ASSET4
Database including MSCI data, obtain similar results for the cases of
Spain, France and Japan, stressing the significant positive impact
of ESG performance on financial performance for companies adopting
the United Nations Global Compact (UNGC). The findings of Aureli et al. (2020) confirm the relevance of ESG disclosure on firms’ market value for 55 Dow Jones Sustainability World Index listed companies. Similarly, Giese et al. (2019) find, on the basis of MSCI ESG data, that ESG information affects positively not only companies’ valuation but also their performance. The authors even identify the precise channels for such effects, namely through: reduced capital costs, higher valuations, higher profitability, and lower exposure to tail risk. Lo and Kwan (2017) analyze the case of Hong Kong companies using the Dow Jones Sustainability Asia Pacific Index (DJSIAP) and FTSE4Good Global 100 Index (FTSE4Good) and find weak but positive evidence of the market reaction on ESG information. Furthermore, ESG initiatives present a stronger observed effect comparatively to SRI.

However, there are other studies leading to different conclusions, for instance, on the absence of a statistical impact (Landi & Sciarelli, 2019) or the presence of very weak effects of only some of the ESG pillars (Miralles-Quirós et al., 2019). While the first paper focuses on Italian companies and exploits the FTSE MIB, the second one considers a larger panel of mainly OECD members and runs its analysis on the basis of the ESG data provided by Thomson Reuters Eikon. Nevertheless, given the different methodologies accounting for ESG efforts across the different rating agencies and data providers, but also given the inherent country-specific particularities, and the ESG materiality issue, it is difficult to draw a robust conclusion on the topic. In this light, what seems evident, is the necessity of a further standardization of ESG accounting procedures, allowing investors, policy makers and scientists to evaluate at best the effects of ESG performances.

Another strand of the literature is focusing on the performance of ESG portfolios. Kempf and Osthoff (2007), Statman and Glushkov (2009), Nofsinger and Varma (2014), Henke (2016) confirm that investing in ESG-firms-based portfolios can clearly provide a performance gain. Namely, Kempf and Osthoff (2007), using data from KLD Research & Analytics, find that the strategy of buying stocks with high socially responsible ratings and sell stocks with low socially responsible ratings leads to high abnormal returns which remain significant despite transaction costs. Statman and Glushkov (2009), base their analysis on the same database and find that socially responsible investors benefit from a return advantage relative to conventional ones. Nevertheless, the authors also find that the systematic exclusion of sin stocks might be penalizing. Nofsinger and Varma (2014) consider and identify also an outperformance of socially responsible mutual funds especially during periods of market crises. According to the authors, using the Morningstar database and the Domini Social Index, this effect is especially pronounced for ESG funds using positive screening techniques and the obtained performances depend only on the socially responsible fund attributes. The work provided by Henke (2016), evaluating US and Eurozone funds (with data provided by US SIF and Euro SIF), stresses also the outperformance of ESG portfolios during crisis periods. The obtained results maintain their significance even after a large set of robustness checks. However, Yen et al. (2019) perform a similar analysis but for Asian stock markets, on the basis of ASSET4 ESG ratings, and find that “socially responsible investment” (SRI) portfolios perform better only in Japan, while in emerging Asian stock markets they are not especially rewarded. This last observation is also confirmed by Auer and Schuhmacher (2016) using the Sustainalytics ESG indicators. Namely, the authors obtain a similar performance for the Asia-Pacific region and in the United States, while in Europe, the investment performance can even be negatively affected, for certain industries and specific ESG criteria. Nevertheless, the meta-analysis performed by Friede et al. (2015), combining the findings of over 2200 individual studies, highlights that about 90% of them show a non-negative relationship between ESG and corporate financial performance (CFP) with a clear positive relationship on the large majority. Interestingly, Halbritter and Dorfleitner (2015) showed that the size and directionality of the overperformance of ESG portfolios strongly depend on the rating provider which highlights the considerable discrepancies among ESG ratings and the necessity for a stronger harmonization.

The third section of the literature evaluates the effects of ESG factors on credit ratings. Attig et al. (2013) find evidence, on the basis of MSCI ESG STATS, that firms with good social performance benefit from relatively high ratings provided by credit rating agencies. Similarly, Devalle et al. (2017) and Weber et al. (2010), confirm that firms with important environmental and sustainability performances are benefiting from higher credit rating scores. While the first study uses ESG data from Thomson Reuters Datastream, the second one is based on a questionnaire covering 58 sustainability criteria and addressed to credit officers, from 40 German banks, providing credits to small and medium-sized firms. Kiesel and Lücke (2019) demonstrate a small but clear influence of ESG performance on rating decisions, especially of the corporate governance pillar. The authors apply the LDA model identifying ESG topics in 3719 Moody’s credit rating reports. According to the work of Jang et al. (2020), focusing on the case of South Korea and using ESG data from Korea Corporate Governance Service (KCGS), ESG ratings are complementary to credit ratings as they encompass essential non-financial information and can lower the cost of debt financing, especially for small firms. Bhattacharya and Sharma (2019) attempt a similar type of analysis for the Indian market, using ESG data from Bloomberg, and observe that ESG efforts present a positive effect on credit ratings only for small and middle-level companies. As for the previous literature strands, it is difficult to draw a robust conclusion on the ESG role on credit ratings.

Thus, the presented above elements and the obtained contradictory results within the current literature, stress the necessity of reliable and harmonized ESG data.

Nonetheless, given the current urgent need for precise actions coping with ESG issues, sustainable finance’s role seems evident. In result, several ESG investment strategies have emerged. The following subsection discusses precisely these aspects.

### 2.1 ESG investment strategies

Sustainable investing is an investment approach that integrates ESG characteristics with classical investment techniques in the portfolio.
construction and management process. According to the Global Sustainable Investment Alliance (GSIA), several strategies have been emerged as global standards in the industry. The main sustainable investment strategies can be classified as follows:

1. **Negative/exclusionary screening.** Exclusion of specific unacceptable or controversial sectors or companies whose activities may harm the environment or society. According to the Global Sustainable Investment Review (GSIA, 2018), this has been the most popular strategy for open-end funds (gathering $19.8 trillion in asset under management). This achievement might be due to the ease of implementation of such a procedure, based on the identification and exclusion of the so-called “non-ESG” stocks.

2. **Positive/best-in-class screening.** Selection of the best ESG performing companies within a specific business sector, all other conditions being equal. It corresponds to the exclusion of companies not meeting certain performance thresholds.

3. **Sustainability themed investing.** Targeted investments, including only activities related to the chosen theme (clean energy, pollution reduction, low carbon emissions, water resources management, sustainable agricultural activities etc.).

4. **Impact/community investing.** Private investments dedicated to specific projects solving social and environmental issues such as renewable energy use, social housing investments etc.

5. **ESG integration.** Systematic and explicit inclusion of ESG factors into financial analysis. Given the qualitative and subjective character of this type of evaluation, the role of ESG rating agencies is crucial.

6. **Corporate engagement and stock activism.** Exercise of the shareholders’ rights aiming to influence corporate behavior through direct dialogue with corporate management and proposal submissions.

7. **Norm-based screening.** Investing only in stocks respecting minimum thresholds of ethical business practices based on international principles.

As the description provided above, these investment strategies are complex and required information availability and a deep analysis of the firms subject to these investments. Despite this complexity, in 2018, the global sustainable investing assets accounts for 30 trillions of dollars invested through these strategies and continue to increase (GSIA, 2018). This indicates a significant demand for information about the compliance of a company to the ESG properties and thus, the need of ESG ratings.

### 3 | RATING THE “UNRATABLE” AND IMPLICATIONS ON ESG INDEXES

In response to the rising demand for reliable ESG data and ESG ratings, the sustainability rating market has grown noticeably and is now in a phase of consolidation (Escrig-Olmedo et al., 2019). The observed interest is mainly a result of the fact that ESG ratings can be assimilated to corporate risk indicators, allowing for the reduction of information asymmetries (see Utz, 2017). Unlike credit ratings, ESG measurement is somehow nebulous given the lack of a common definition, reporting standards and shared characteristics among each ESG component and across rating providers. Currently, rating agencies are proposing several metrics similar to the credit ratings market, but, unlike those, ESG ratings are derived from alternative and competing definitions. Hence, a common standard for ESG is missing and makes the sustainability of a company very difficult to assess and, in some cases “unratable.”

Windolph (2011) identifies six causes that hinder a transparent and objective rating, such as: lack of standardization, lack of credibility of information, bias, trade-offs, lack of transparency, and lack of independence. Moreover, during the last years, ESG rating agencies have not only integrated new criteria into their assessment models in order to respond to new global challenges (Escrig-Olmedo et al., 2019) but also the assessment criteria are changing rapidly, making the evaluating process even more complex.

The conceptual framework is therefore very fragile. Nevertheless, this paper offers an alternative view on three main challenges (and the implications that may arise): (i) agreement and disagreement in the ESG ratings, (ii) agreement and disagreement in the ESG indexes and (iii) implications of (ii) on financial performance.

#### 3.1 | How agencies judge corporates: ESG ratings

We investigate how the ESG criteria are evaluated, and if the data offered by the agencies is reliable. In doing so, the present research firstly provides an overview of the most relevant issues faced by the agencies. Secondly, it illustrates the used methodologies, as well as the business assessment process, in order to draw conclusions on the achieved consensus among all players.

Rating agencies have developed their own assessment methodology to evaluate ESG engagement. Given the different methodologies, Table 1 offers an overview of the factors that major ESG rating providers are considering within their assessments. All the information has been extracted and pooled from companies’ websites. The goal of this first analysis is to check whether the agencies converge in terms of methodology and if the ratings they provide are consistent.

As Table 1 shows, the first difference is related to data sources. The identification of sensitive data points is, in fact, a crucial prerequisite for a good rating. The principal sources for all agencies are publicly available information, such as companies’ reports and websites. However, the sources of information change from one rating agency to another. For example, ISS-Oekom and Bloomberg leverage on direct contact with the company, while Thomson Reuters also considers stock market registrations. RobecoSAM, on the other hand, uses a completely different approach. The ESG rating assessment for this provider begins with the invitation of the world’s largest listed companies to participate in a questionnaire which, along with the provided information, will also assess all the missing information.
<table>
<thead>
<tr>
<th>MSCI</th>
<th>Vigeo-Eiris</th>
<th>Refinitiv</th>
<th>Sustainalytics</th>
<th>ISS Oekom</th>
<th>RobecoSAM</th>
<th>ECPI</th>
<th>Bloomberg</th>
<th>FTSE Russell</th>
</tr>
</thead>
<tbody>
<tr>
<td>Headquarter</td>
<td>New York, United States</td>
<td>Paris, France</td>
<td>Toronto, Canada</td>
<td>Amsterdam, Netherlands</td>
<td>Munich, Germany</td>
<td>Zurich, Switzerland</td>
<td>Milan, Italy</td>
<td>New York, United States</td>
</tr>
<tr>
<td>Sources</td>
<td>Company disclosure, 1600+ Media sources, 100+ specialized dataset</td>
<td>Company disclosure, Recommendation, Conventions</td>
<td>Company websites, Company reports, NGO websites, Media and news, Stock exchange filings</td>
<td>Public disclosure, Media and news, NGO reports</td>
<td>Publicly available information, Interview with stakeholders, information on company policies and practices, company direct contact</td>
<td>Survey approach</td>
<td>Company reports, Media and news, Regulatory data, Bloomberg and Thomson Reuters, University networks</td>
<td>Company reports, Publicly available information, Company direct contact</td>
</tr>
<tr>
<td>N. Criteria</td>
<td>37</td>
<td>38</td>
<td>178</td>
<td>155</td>
<td>100</td>
<td>74</td>
<td>80/86</td>
<td>120</td>
</tr>
<tr>
<td>Main risk factors</td>
<td>Environmental</td>
<td>Human resources, Human rights, Environment, Business behavior, Community, Involvement, Corporate governance</td>
<td>Environmental</td>
<td>Resource use, Emission, Innovation</td>
<td>Social</td>
<td>Workforce, Human rights, Community, Product responsibility</td>
<td>Governance</td>
<td>Management, Shareholders, CSR strategy</td>
</tr>
</tbody>
</table>

TABLE 1 Key differences between ESG rating agencies using publicly available information on agencies websites.
However, although sometimes sources of information tend to converge, the way they are processed may also be different.

Besides, the number of assessed indicators changes among the different raters. MSCI and FTSE Russell represent the extreme cases since they assess respectively 37 and 300 ESG criteria. Other agencies, instead assess different metrics in relation to the industry the company belongs to (see Sustainalytics and RobecoSAM). The main risk factors summarize any single score from the indicators used in the previous step. At this stage, we noticed considerable structural divergences. For instance, ECPI considers two sustainable dimensions since it incorporates the social dimension in the governance sphere. RobecoSAM substitutes the governance dimension with an economic one (which includes also corporate governance).

Finally, the difficulty in achieving a general definition of ESG materiality,7 eventually triggering the weighting mechanism of the assessed criteria, may generate further divergence in the overall rating. As Table 1 shows, all the examined ESG rating agencies have developed a proprietary definition of materiality. Consequently, the weighting procedures vary considerably. Furthermore, we observe that almost all rating agencies adjust their final rating by integrating issues, specific to the considered industry, but very few of them publish the assigned weights transparently. Thus, the discrepancy across ratings may simply reside in the consideration of different components weighted differently.

In order to further illustrate the existing disparity, we provide an example including four ESG providers: Sustainalytics, RobecoSAM, Refinitiv and MSCI.8 Table 2 reports the ESG rating for four companies (Verizon Communications Inc., Nissan Motor Co., Ltd., Oracle Corp. JPN and Goodman Group). The rating scale for Sustainalytics, RobecoSAM and Refinitiv ranges from 0 to 100 while MSCI provides a classical taxonomy ranging from CCC to AAA scale.

As shown in Table 2, the ratings differ considerably across the providers. For instance, in the case of Nissan Motor Company, Sustainalytics and MSCI assign a low rating while RobecoSAM and Refinitiv include it among the best-in-class. These discrepancies can be explained by the fact that both Sustainalytics and MSCI focus on the managerial perspective investigating the firm’s exposure to ESG risk and its mitigation strategy (e.g., unmanaged ESG risks and potentially higher production costs for its carbon-intensive products). On the other hand, despite giving a low ESG controversial score, Refinitiv judges Nissan positively, especially in the environmental and governance spheres. Finally, RobecoSAM overweights the environmental dimension for the automotive industry. In particular, operational eco-efficiency and climate strategy account for 20% of the overall rating. As a result, the ranking given by RobecoSAM awards Nissan’s cleantech innovation capacity compared to the other analyzed competitors (32 companies).

Similar considerations can be formulated for the remaining companies reported in the table and for which we observe diverging ESG ratings inherently related to the concerned metrics (i.e., Verizon Communications is very high in Sustainalytics and very low in RobecoSAM while it is in the middle range in MSCI and Refinitiv). This simple
Example clearly highlights that the currently applied metrics can lead to contradictory evaluations.

In order to better understand the ratings’ divergence across providers, we consider a large set of 1049 companies listed in the MSCI World Index with available ESG ratings from the four providers mentioned above.\(^9\) Firstly, we compute the rank correlation among the ESG ratings for the four considered providers.\(^{10}\) In relation to this, we harmonize and convert the rating classes among the four considered providers by applying a common scale ranging from 1 to 7. The conversion table is reported in the Appendix A. The conversion scale described above (Bottom panel in Table 3).

<table>
<thead>
<tr>
<th>Company</th>
<th>Sustainalytics</th>
<th>RobecoSAM</th>
<th>Refinitiv</th>
<th>MSCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nissan Motor Co., Ltd</td>
<td>6</td>
<td>77</td>
<td>72</td>
<td>CCC</td>
</tr>
<tr>
<td>Verizon Communications Inc.</td>
<td>91</td>
<td>20</td>
<td>67</td>
<td>BB</td>
</tr>
<tr>
<td>Oracle Corp. Jpn</td>
<td>78</td>
<td>8</td>
<td>63</td>
<td>BB</td>
</tr>
<tr>
<td>Goodman Group</td>
<td>86</td>
<td>21</td>
<td>58</td>
<td>AA</td>
</tr>
</tbody>
</table>

**TABLE 2** Example of divergence in ESG ratings for Verizon Communications Inc., Nissan Motor Co., Ltd, Oracle Corp. Jpn and Goodman Group

The correlation between ESG ratings is on average 0.58 and ranges from 0.43 (between Refinitiv and MSCI) to 0.69 (between Refinitiv and RobecoSAM). Similarly to Berg et al. (2019), our results confirm a disagreement among the different rating providers, for the same considered companies.\(^11\) Given the obvious discrepancies in terms of methodology and ESG ratings assigned to the same firms, it is important to identify the channels that might affect the ESG indexes construction. As noted also by Engle et al. (2019), heterogeneity in the ESG scores can lead to the formation of different ESG portfolios due to disagreement in the ratings. In fact, if the implemented metrics vary among the rating providers, there will also be disagreement in the selection of the ESG indexes constituents. We investigate this issue in detail in the next subsection.

### 3.2 ESG indexes agreement

The quantity of actively managed portfolios combining environmental, social and governance factors has grown considerably. According to iShares (2019), more than 1000 indexes, including ESG preferences expressed by the investors, are now available on the market.

Given the lack of a globally accepted standard methodology, we investigate the discrepancies among the ESG indexes, provided by the four ESG rating agencies considered above (i.e., Sustainalytics, RobecoSAM, Refinitiv and MSCI) and focusing on developed markets investment set.

Table 4 specifies the names of the four ESG indexes, their associated ESG Rating providers and the number of their constituents. Despite the different dimensions in terms of constituents, the considered indexes remain quite similar in terms of sectorial and geographical composition. As Figure 1 shows, at the sectorial level, all the four indexes have a similar composition. For each class the average difference ranges from 1% to 3%. The largest difference is 5% in consumer discretionary (STOXX and MSCI) and Information Technology (Refinitiv and MSCI). At the geographical level, the average difference is 15% (North America), 16% (Europe) and 6% (Asia Pacific). As can be seen from Figure 2, the difference between North America and Europe can be explained mainly by STOXX and MSCI. Based on the average exposition, the first index is over-(under)exposed to Europe (North America) while the second is less (more) exposed to North America (Asia Pacific).

The sectorial and geographical analysis of the ESG indexes composition highlight the presence of some geographical differences,
Despite the significant overlap in terms of geographical coverage. In line with the previous section, we aim to measure the agreement between the rating providers also from this perspective. To investigate the degree of similarity on the selection of the constituents, that is, the agreement, we compute the overlap coefficient by using the Szymkiewicz–Simpson coefficient. The latter corresponds to the size of the intersection between two indexes divided by the size of the smallest one.12

Table 5 shows that the agreement rate among the evaluated indexes is low and ranges from 35% to 59%. In this respect, the lowest agreement rate is observed between the Refinitiv and the Dow Jones indexes while the highest is observed between STOXX and MSCI indexes. The overall overlap coefficient is 15% (48 constituents), which confirms that the divergence is greater when considering all the indexes together.  

Therefore, having an index where there is agreement between the four ESG rating providers would imply to consider all the firms that have been included in all the four ESG indexes. This is exactly the portfolio coming from the overall overlap (the ESG agreement portfolio). Conversely, an index composed of the firms that have been considered only by one rating agency in its index would imply disagreement (the ESG disagreement portfolio). To check if the two portfolios are coherent, we compute the MAE and the percentage of observed agreement on the ratings of the constituents across the four rating agencies. The MAE of the ESG agreement (disagreement) portfolio is on average 0.73 (1.44) and ranges from 0.46 (1.18) to 0.90 (1.79). The percentage of observed agreement of the ESG agreement (disagreement) portfolio is on average 46% (22%) and ranges from 33% (17%) to 68% (27%). Results confirm that the consensus among the rating providers is stronger in the ESG agreement portfolio.

However, the observed low agreement rate in ESG indexes might be related to the differences in terms of geographical exposition rather than to ESG ratings. For this reason, we perform an additional analysis, by computing the overlap measures across the different geographical areas. Thus, we aim to evaluate how agreement and disagreement rates might evolve after controlling the geographical allocation of the ESG indexes constituents. It is worth noting that we are considering ESG global indexes that might be allocated with different intensities within the same geographical areas. This might imply that the overlap coefficient is low just because the two indexes focus on different geographical areas.13

---

**TABLE 4** List of the considered four rating providers and their corresponding ESG indices including the number of their constituents (last column)

<table>
<thead>
<tr>
<th>Label</th>
<th>Rating Provider</th>
<th>ESG Index</th>
<th>Number of Constituents</th>
</tr>
</thead>
<tbody>
<tr>
<td>STOXX</td>
<td>Sustainalytics</td>
<td>STOXX Global ESG Leaders</td>
<td>431</td>
</tr>
<tr>
<td>Dow Jones</td>
<td>RobecoSAM</td>
<td>Dow Jones Sustainability World</td>
<td>317</td>
</tr>
<tr>
<td>Refinitiv</td>
<td>Refinitiv</td>
<td>Refinitiv Global ESG</td>
<td>404</td>
</tr>
<tr>
<td>MSCI</td>
<td>MSCI</td>
<td>MSCI World ESG Leaders</td>
<td>777</td>
</tr>
</tbody>
</table>
Clearly, we expect an increase in the agreement rate for those indexes with a similar exposition in a given area with respect to the total composition. To further investigate this issue, we normalize the geographical allocation and look for the overlap. The analysis shows that for the ESG indexes that have a significant difference in the geographical allocation (e.g., STOXX and MSCI) the overlap coefficient increases by 10% for North America, 5% for Europe and 1% for Asia. However, for the indexes with very similar exposition, for example, Refinitiv and MSCI in North America (difference of 4%), Dow Jones and Refinitiv in Europe (3%), and STOXX and Refinitiv (0%) and Dow Jones and Refinitiv (5%) in Asia Pacific, the agreement decreases for Refinitiv and MSCI in North America by −3% and for STOXX and Refinitive in Asia Pacific by −7% while, it increases for Dow Jones and Refinitive in Europe, +19%, and Dow Jones and Refinitiv in Asia Pacific, +5%.

Therefore, the geographical allocation might influence the measured overlap effect, but on average such impact remains largely marginal. Thus, we conclude that the disagreement across ESG indexes persists even after controlling the overlap coefficient for the geographical area.

The above described results highlight plausible doubts on the achievement of a consensus among ESG ratings, since it is evident that ESG rating agencies provide heterogeneous information to the market not only in terms of ESG ratings but also in terms of ESG indexes and their inherent constituents. The lack of a globally accepted standard methodology and a minimum level of technical requirements create two major issues. From one perspective, investors face considerable difficulties in selecting ESG targets for investment, and conversely, companies encounter significant difficulties in identifying the characteristics they should comply with, in order to be included into ESG indexes. This assessment further motivates us to analyse how the agreement and disagreement concerning the inclusion of ESG indexes constituents can affect the financial performance of ESG portfolios with respect to a nonESG counterpart.

### Table 5

<table>
<thead>
<tr>
<th></th>
<th>STOXX</th>
<th>Dow Jones</th>
<th>Refinitiv</th>
<th>MSCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>STOXX</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Dow Jones</td>
<td>50%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Refinitiv</td>
<td>43%</td>
<td>35%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MSCI</td>
<td>59%</td>
<td>49%</td>
<td>50%</td>
<td>-</td>
</tr>
</tbody>
</table>

Note: The indices include companies from the developed markets that exhibit high ESG performance according to the related provider. The overall overlap coefficient is 15%.

In this section we aim to assess whether the agreement of the inclusion of stocks in an ESG index by the four considered ESG rating agencies (ESG agreement portfolio) affects the financial performance,
that is, if it is generating an extra-performance after controlling for the remuneration of financial risk factors. The performance measure that captures this feature best is the Jensen-alpha. We aim to also check whether the performance of the ESG agreement portfolio are better than a nonESG portfolio.

### 4.1 Building the portfolios

The ESG and the nonESG portfolios are built as follows:

1. **The ESG agreement.** We consider the common constituents of the four indexes described in Table 4. Coherently with the previous analysis, if a firm is included as the constituent of an index, it is considered as an ESG leader according to the corresponding rating provider. Consequently, there is agreement when a stock is included in all the four indexes. This portfolio is originated from the overall overlap discussed in Section 3.2, and includes 48 constituents.

2. **The nonESG portfolio.** “nonESG” stocks have a broader definition. Indeed, they are not only related to a certain type of industry, but they also depend on the firm’s characteristics and attributes (E, S and G components). We build the nonESG portfolio following a negative screening approach which considers all the stocks that are excluded from the ESG investment universe. In order to identify these nonESG firms we consider all firms that do not comply with the principles of the UNGC which is where the concept of ESG originated. We define the nonESG portfolio as follows. Firstly, we consider the constituents of the MSCI World Index which represent a good proxy of the investment universe of the developed markets. Secondly, we consider the constituents of the MSCI World ESG screened index which excludes firms from the MSCI World index that are not in compliance with the UNGC principles and are involved in controversial or unacceptable activities which may hurt the environment or society as a whole. Finally, the nonESG portfolio is obtained as the complement of the MSCI World ESG screened and includes 119 constituents. None of these 119 constituents have been included in any of the other three ESG indexes.

### 4.2 Results

Since the term ESG was first coined in 2005, it is reasonable to divide the analysis into two periods in order to detect the presence of a structural change in the financial performance of ESG and nonESG portfolios. The first period ranges from December 1999 to December 2004 while the second one starts in January 2005 and ends in January 2020. The total returns for the constituents have been downloaded from Elkon/Datastream at monthly frequency. The portfolios are equally weighted.

Table 6 includes both the performance measures for the ESG agreement and the nonESG portfolios. The second portfolio shows higher Sharpe and Sortino ratios in the first period (1.22 and 1.53) with respect to the ESG portfolio (0.25 and 0.38) while in the second period the two ratios are more similar. The same conclusions hold true for the Omega ratio and the Max Drawdown and Var at 5%.

This simple analysis indicates that the nonESG portfolio is better preferable than the ESG agreement portfolio from the financial point of view. Nevertheless, it also suggests that there might have been a change in investors' preferences or tastes after 2005 with the introduction of the ESG firm’s characteristics. Indeed, the performances of the two portfolios are closer in the second period. However, the presence of a Jensen-alpha that captures the ESG component in the market could only be investigated after controlling for the remuneration of the usual financial risk factors.

In regard to this, we estimate the Jensen-alpha of both portfolios representing ESG agreement stocks and the corresponding counterpart (nonESG portfolio) using the Carhart four-factor model (Carhart, 1997). The model is an extension of the Fama–French three-factor model (Fama & French, 1992) and allows one to disentangle the portfolio performance measuring the impact of (1) the market risk, (2) the outperformance of small versus big companies, (3) the outperformance of high book/market versus small book/market companies and (4) the momentum factor. After controlling for these factors, the difference on the alphas between ESG and its counterpart should reveal if an ESG impact does exist on stock performance.

The estimated results are reported in Table 7 for both considered periods. The first column in the table shows the estimation for the ESG agreement portfolio while the second shows the estimation for the nonESG portfolio for the period 2005–2004. In order to measure whether the observed difference between the two portfolios is statistically significant, we estimate a long/short portfolio which is built between the ESG agreement portfolio (long position) and the nonESG portfolio (short position). After controlling for the four-risk factors, a positive (negative) significant alpha would imply that the ESG portfolio outperforms (is outperformed by) the nonESG portfolio, while a non-significant alpha, would imply that there is no statistical difference in the two portfolios. The result is reported in the third column.

### Table 6 Performance measures for the ESG agreement and the nonESG portfolios for the considered periods

<table>
<thead>
<tr>
<th></th>
<th>Sharpe ratio</th>
<th>Sortino ratio</th>
<th>Omega ratio</th>
<th>MaxDD</th>
<th>VaR 5%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ESG</td>
<td>nonESG</td>
<td>ESG</td>
<td>nonESG</td>
<td>ESG</td>
</tr>
<tr>
<td>2000–2004</td>
<td>0.254</td>
<td>1.224</td>
<td>0.377</td>
<td>1.535</td>
<td>1.209</td>
</tr>
<tr>
<td>2005–2019</td>
<td>0.870</td>
<td>1.044</td>
<td>1.153</td>
<td>1.287</td>
<td>1.953</td>
</tr>
<tr>
<td>2000–2019</td>
<td>0.691</td>
<td>1.092</td>
<td>0.945</td>
<td>1.361</td>
<td>1.694</td>
</tr>
</tbody>
</table>

Note: The first three columns provide the annualized Sharpe Ratio, Sortino Ratio and Omega Ratio respectively. The forth column provides the Max Drawdown (MaxDD) and the last provides the Value at Risk at 5%.

---

1. Billio et al.

2. Jensen-alpha.

The following three columns describe the same type of analysis but for the period 2005–2019. In the first period, the alphas are not significant neither for the ESG agreement nor for the nonESG portfolios. However, in the second period, both the ESG and nonESG portfolios provide a similar positive and significant alpha of 0.59% and 0.53%, respectively. Conversely, the alphas on the long-short portfolio are not significant for both periods and hence, it appears that, there are no significant portfolio performance differences after controlling for the four risk factors. According to our findings, the observed results are a consequence of the previously detected scores’ disagreement among the rating agencies, to the extent that even when there is agreement, the ESG effect is diluted and it has no impact on performances.

Nevertheless, the ESG agreement and the nonESG portfolios show some differences in terms of their exposition to the four factors. In the first period, there is a negative and significant difference with the high minus low (HML) factor and the winners minus losers in the previous 12 months (WML) factor. The first factor indicates that the nonESG portfolio is exposed to growth stocks while the second highlights that the ESG portfolio is negatively related to the momentum. The latter is confirmed in the second period where there is also a positive significant difference in terms of market exposition for the ESG portfolio. This is also consistent with the findings in Monasterolo and De Angelis (2020) where the green asset class shows a market exposition that is larger than the brown one. Therefore, this requires some robustness checks to validate the results and assess their consistency. This procedure is especially important in order to avoid that our results being biased due to portfolio characteristics not related to ESG features.

### 4.3 Robustness Analysis

To validate our results, we perform four robustness analyses on the ESG portfolio.

1. **The sectorial analysis.** We perform an analysis by distinguishing among the different sectors that might be more or less exposed to the impact of ESG risks. In the same manner, we repeat the long-short portfolio exercise and investigate for any differences in the alphas between the sectorial ESG agreement and the nonESG portfolios. Results are included in Appendix D and show that, except for the communication services sector in 2005–2019, the alphas are not statistically significant.

2. **The ESG matched portfolio.** Another potential issue that might affect our results, concerns the Carhart model that might not be able to fully capture the portfolio characteristics due to a mismatch on the sectorial and geographical composition (see Figure E1 in the Appendix). To address this issue, we have performed an additional analysis using a matched ESG agreement portfolio with the nonESG portfolio through the propensity score approach (Heckman et al., 1998). Results are included in Appendix E and confirm that the alpha in the long short portfolio is not statistically significant. Moreover, the two portfolios are very similar in terms of exposition to the four factors.

3. **The rebalanced portfolio.** An additional constraint lies in the non-synchronized data availability, that is, the indexes and their constituents' data start at different points of time. For the analysis, we have selected the constituents of the indexes in 2019 as reference in order to include the pre-ESG era for all the four considered indexes. As a robustness check, we have considered a rebalanced ESG agreement portfolio which tracks the periodic revision in all the four considered indexes. The commonly available period for the four indexes starts from January 2012. Results are included in the Appendix F. The analysis shows that the portfolio rebalance does not affect our results.

4. **ESG indexes commonly available period.** As highlighted above, ESG indexes started in different periods. To assess if the different starting dates affect our results, we run the Jensen-alpha...
performance test for the 2012–2019 period. This complementary test confirms our previous results: the Jensen-alpha for the long-short ESG-nonESG portfolio is not statistically different from zero. Detailed results are reported in the Appendix F.

5. The ESG disagreement portfolio. We also consider the disagreement among the rating providers and build an alternative portfolio: the ESG disagreement portfolio (i.e., the one where the stocks considered only have one ESG rating agency that included it in its ESG index). Also in this case, if investors’ tastes change towards more sustainable investments, we should find a difference in terms of alphas with respect to the ESG agreement portfolio. Results are included in Appendix G and confirm no statistical difference in terms of alphas between the ESG agreement and the ESG disagreement portfolios.

5 | CONCLUSION

Within the last two decades a growing awareness on climate change and environmental problems have influenced the public opinion and it has urged governments and firms to integrate these aspects in their regulatory decisions and firm strategies. At the same time, governance and social issues also became more pressing after several negative corporate disclosures and the global financial crisis. According to the Forum for Sustainable and Responsible Investment (USSIF), the sustainable and responsible investments have grown by more than 38% in the period 2016–18. Without any doubt, the investment sector is experiencing a structural change that will significantly affect all the financial sectors and the other industries. The main reason is the higher weight assigned to the ESG criteria by asset managers during the selection process and the use of ESG indexes as benchmark for performance analysis. Overall, the involved stakeholders are facing a considerable need for precise information on companies’ ESG merit. This new type of information is provided by the recent specialized market of ESG ratings.

So far, literature has concentrated on the disagreement of rating agencies or on the financial performance of ESG investments. In this paper, we link this two aspects and investigate the implications that disagreement might have on ESG portfolios performance.

We show that there is a lack of common metrics among rating agencies in the definition of ESG (characteristics, attributes and standards) and that heterogeneity in judgment can lead agencies to assign even opposite ratings to a given company.

Moreover, this heterogeneity is problematic even to the investment industry itself, since the identification of sustainable investments portfolios and therefore the choice of the relevant benchmarks (ESG indexes) depends on the ratings originated from these metrics. ESG indexes represent the target of ESG passive investment fund strategies or of Exchanged Traded Funds as well as an important input for the asset allocation and stock selection strategies of ESG active portfolios (i.e., active managed funds). Consequently, the performance measurement of these investment instruments could be seriously affected just on the basis of the chosen benchmark.

We provide evidence on this aspect and confirm that the observed ESG rating heterogeneity leads to the identification of alternative benchmarks, since the common component in terms of constituents is represented by a relatively small sample of firms. The quota is 15% for the ESG index with the smallest number of constituents and 6% of the largest one. In this respect, we further develop our analysis and investigate the financial implication of ESG disagreement among rating agencies. Namely, we evaluate the financial performance of a portfolio composed by the common stocks to all the considered rating providers and their respective ESG indexes.

Our findings show that observed ESG disagreement, among rating agencies, disperses the effect of preferences of ESG investors on asset prices (see Merton, 1987), to the point that even when there is agreement, the latter is so weak that has no impact on the financial performances of ESG portfolios.

The theoretical motivation of this result is that the invested amount is so limited, that does not involve a significant financial impact. In our view, financial performances would be different if all major ESG rating agencies agree on the definition of a set of common metrics. This would lead to more homogeneous stock selections and consequently, to the identification of an unique benchmark or a set of ESG indexes with a significant overlap. Thus, both active and passive ESG investment funds would have the opportunity to concentrate their investments on the same stocks and thus, generate a significant impact on asset prices (as predicted by Merton, 1987). Instead, we find that, even if sustainable and responsible investments have grown significantly in the last years, there is no difference in financial performances with respect to the non-ESG counterpart. Moreover, we show that the same conclusion can be drawn from the stock portfolio analysis, for both cases where ESG indexes agree and where they do not.

Accordingly to our findings, there is an evident need for a more precise definition of ESG investment benchmarks. Namely the creation of a road map for the adoption of common practices and rules in sustainable indicators’ reporting (comparable to the IFRS for the accounting standards) should be considered by policy makers and regulators.

The present study provides a first step of investigation. In order to have a complete and detailed overview, it would be interesting to disentangle the role of the three E/S/G components with regard to the agreement and disagreement among ESG rating agencies. A further step of research could focus on the potential effects of a convergence process over time.

ACKNOWLEDGMENTS

The authors thank two anonymous reviewers and the editors, Robert Engle, Marina Brogi, Nicola Cucari and Valentina Lagasio, for their insightful comments and suggestions. They also acknowledge the financial support of European Investment Bank (EIB) for the project “ESG-Credit.eu - ESG Factors and Climate Change for Credit Analysis and Rating”. Carmelo Latino and Loriana Pelizzon gratefully acknowledge research and financial support from the Leibniz Institute for Financial Research SAFE, Frankfurt. The usual disclaimer applies.
ENDNOTES

1 Such as: droughts, floods, hurricanes, wildfires not only in tropical regions but also in temperate climate zones.

2 Derwall et al. (2004); Kempf and Osthoff (2007); Statman and Glushkov (2009); Nofsinger and Varma (2014); Henke (2016); Yen et al. (2019); Auer and Schuhmacher (2016); Friede et al. (2015).

3 Zhao et al. (2018); Brogi and Lagasio (2019); Ortas et al. (2015); Aureli et al. (2020); Giese et al. (2019); Lo and Kwan (2017); Landi and Scarel (2019); Miralles-Quirós et al. (2019); Bruder et al. (2019).

4 Attig et al. (2013); Devalle et al. (2017); Weber et al. (2010); Kiesel and Lücke (2019); Jang et al. (2020); Bhattacharya and Sharma (2019).

5 Escrig-Olmedo et al. (2010); Escrig-Olmedo et al. (2019); Berg et al. (2019).

6 Both matching information from Morningstar, Bloomberg, Novethics (France), Ecoreporter (Germany) and SRI Fund Advice (US).

7 Broadly speaking, the concept of ESG materiality can be defined as the principle used to identify the most relevant ESG factors affecting the business model of a given company.

8 We consider these four ESG rating providers with available data in Eikon/Datastream and Bloomberg.

9 Data has been downloaded using Eikon/Datastream and Bloomberg.

10 The MAE provides a measure of errors between paired observations describing the same phenomenon, and is equal to \( \frac{1}{n} \sum_{i=1}^{n} |R_i - R^*_i| \), where \( R_i \) and \( R^*_i \) are the ratings on firm \( i \) by rating provider \( A \) and \( B \), respectively, \( n \) is the total number of the considered firms.

11 Additionally, we have also performed a correlation analysis of the ESG ratings provided by the four rating agencies for the 1049 firms included in the MSCI word index. Our analysis shows that correlations of the ESG ratings are on average 0.57 and range from 0.44 to 0.67, thus confirming the results of Berg et al. (2019).

12 The overlap coefficient known as the Szymkiewicz–Simpson coefficient is computed such as \( \frac{\sum_{i \in I} \min(I_i)}{|I|} \), where \( I \) and \( I_i \) are the ESG indexes, and \( |I| \) indicates the cardinality of the set.

13 The same issue could arise for sectors. However, Figure 1 shows that all the four indexes have a similar composition at the sectoral level.

14 Detailed results are included in the Appendix B.

15 Further details are available at: https://www.msci.com/esg-screened-indexes.

16 The descriptive statistics for the ESG agreement portfolio and the nonESG portfolio are included in Appendix C.

17 The Sharpe ratio is defined as \( \frac{\mu - r_f}{\sigma_d} \) and the Sortino ratio as \( \frac{\mu - r_f}{\sigma_d} \), where \( \sigma_d \) is the standard deviation of the downside. The Omega ratio is defined as \( \frac{\int_{r_f}^{\infty} (1-F(x)) dx}{\int_{r_f}^{\infty} F(x) dx} \). We set \( \theta = 0 \) to investigate the probability of having a positive return against negative ones.

18 The four factors have been downloaded from Ken French’s website. The descriptive statistics are available in Appendix C.

19 The Dow Jones Sustainability World index and the STOXX Global ESG Leaders index were launched in 1999 and 2001, respectively. The Refinitiv Global ESG index and the STOXX Global ESG Leaders were launched in 2004 and 2007, respectively. Through Eikon/Datastream and Bloomberg, we had access to the historical composition of MSCI World ESG leaders only after 2012.


REFERENCES


21 The propensity score analysis is conducted using PSMATCH2 proposed by Leuven and Sianesi (2003) in STATA 14.1.


APPENDIX A.: Harmonization of the rating classes for the four providers

We harmonize and convert the rating classes among the four considered rating providers by applying a common scale ranging from 1 to 7. The conversion table is reported in Table A1.

**TABLE A1** The conversion of the rating classes according to the four ESG rating agencies

<table>
<thead>
<tr>
<th>Score</th>
<th>MSCI</th>
<th>Sustainalytics</th>
<th>RobecoSAM</th>
<th>Refinitiv</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CCC</td>
<td>0–14</td>
<td>0–14</td>
<td>0–14</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>15–28</td>
<td>15–28</td>
<td>15–28</td>
</tr>
<tr>
<td>3</td>
<td>BBB</td>
<td>29–42</td>
<td>29–42</td>
<td>29–42</td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td>58–71</td>
<td>58–71</td>
<td>58–71</td>
</tr>
<tr>
<td>5</td>
<td>AA</td>
<td>72–85</td>
<td>72–85</td>
<td>72–85</td>
</tr>
<tr>
<td>6</td>
<td>AAA</td>
<td>86–100</td>
<td>86–100</td>
<td>86–100</td>
</tr>
</tbody>
</table>

**TABLE B1** Overlap coefficients among four ESG indices for North America

<table>
<thead>
<tr>
<th>STOXX</th>
<th>DOW JONES</th>
<th>REFINITIV</th>
<th>MSCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>STOXX</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOW JONES</td>
<td>27%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>REFINITIV</td>
<td>50%</td>
<td>66%</td>
<td></td>
</tr>
<tr>
<td>MSCI</td>
<td>71%</td>
<td>53%</td>
<td>47%</td>
</tr>
</tbody>
</table>

**APPENDIX B.: Overlap indices according the geographical area**

In this section, we provide the overlap coefficients for the three macro areas: North America, Europe and Asia Pacific. The overlap coefficients are shown in Tables B1–B3, and are obtained as the size of the intersection between two indices divided by the smaller size of the two. Results confirm disagreement also at a geographical level.

**APPENDIX C.: Descriptive statistics**

Table C1 reports the descriptive statistics for the ESG agreement and the nonESG portfolios and Table C2 for the four factors (Carhart, 1997), respectively. Figure C1 provides the cumulative returns of the two portfolios over the two periods. The nonESG portfolio shows a higher cumulative return during the first part of the period while in the second part the spread becomes narrow. Finally, we include in Tables C3 and C4 the financial indicators for the constituents of both the portfolios.
TABLE C1  Descriptive statistics (%) at monthly frequency for the ESG and the nonESG agreement portfolios in the considered periods

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>SD</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ESG</td>
<td>nonESG</td>
<td>ESG</td>
<td>nonESG</td>
</tr>
<tr>
<td>2000–2004</td>
<td>0.36%</td>
<td>1.17%</td>
<td>4.90%</td>
<td>3.31%</td>
</tr>
<tr>
<td>2005–2019</td>
<td>1.02%</td>
<td>0.96%</td>
<td>4.07%</td>
<td>3.17%</td>
</tr>
<tr>
<td>2000–2019</td>
<td>0.86%</td>
<td>1.01%</td>
<td>4.29%</td>
<td>3.20%</td>
</tr>
</tbody>
</table>

Note: The first column provides the average return (mean), the second provides the standard deviation (SD), the third provides the minimum return (min) and the last provides the maximum return (max).

TABLE C2  Descriptive statistics (%) at monthly frequency for the four factors (Carhart, 1997) in the considered periods

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>SD</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MKT-Rf</td>
<td>SMB</td>
<td>HML</td>
<td>WML</td>
</tr>
<tr>
<td>2000–2004</td>
<td>–0.15%</td>
<td>4.60%</td>
<td>–10.35%</td>
<td>8.71%</td>
</tr>
<tr>
<td>2005–2019</td>
<td>0.59%</td>
<td>4.30%</td>
<td>–19.51%</td>
<td>11.41%</td>
</tr>
</tbody>
</table>

Note: The factors are: Excess market return (MKT-Rf), the book-to-market factor (HML) the size factor (SMB), and the winners minus losers (WML). The first column provides the average returns (mean), the second provides the standard deviation (SD), the third provides the minimum return (min) and the last provides the maximum return (max).

FIGURE C1  Cumulative returns for the ESG agreement (red line) and nonESG (blue line) portfolios during the considered periods [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE C3  Financial indicators of the constituents of the ESG agreement portfolio

<table>
<thead>
<tr>
<th></th>
<th>Financial leverage</th>
<th>Tot. revenues</th>
<th>Market value</th>
<th>EPS</th>
<th>EBITDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>12.51</td>
<td>6,284,085</td>
<td>8183</td>
<td>0.24</td>
<td>1,329,500</td>
</tr>
<tr>
<td>25%</td>
<td>25.02</td>
<td>11,588,000</td>
<td>16,858</td>
<td>1.31</td>
<td>2,176,250</td>
</tr>
<tr>
<td>50%</td>
<td>40.33</td>
<td>23,640,000</td>
<td>48,181</td>
<td>2.44</td>
<td>4,869,500</td>
</tr>
<tr>
<td>75%</td>
<td>62.43</td>
<td>74,142,000</td>
<td>115,118</td>
<td>9.58</td>
<td>11,623,250</td>
</tr>
<tr>
<td>90%</td>
<td>79.68</td>
<td>169,568,845</td>
<td>510,698</td>
<td>32.15</td>
<td>55,527,900</td>
</tr>
<tr>
<td>Avg</td>
<td>44.14</td>
<td>248,780,227</td>
<td>440,440</td>
<td>24.78</td>
<td>59,240,652</td>
</tr>
</tbody>
</table>

Note: The table shows the data for the 10th, 25th, 50th, 75th and 90th percentiles. The last row shows the average values.

Financial leverage, total revenues, market value, earning per shares (EPS) and earning before interest taxes depreciation & amortization (EBITDA). We use the debt-to-capital ratio to measure the financial leverage. Market value is displayed in millions of dollars.
APPENDIX D.: Performance analysis on the sectorial portfolios

In this section, we report the Jensen’s Alpha of the long-short ESG agreement portfolios and the nonESG portfolio for each sector. Results are included in Table D1 and confirm that there is no statistical difference among the two portfolios, except for Communication Services in the period 2005–2019.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Financials</td>
<td>0.0024</td>
<td>0.0038</td>
</tr>
<tr>
<td>Industrials</td>
<td>0.0038</td>
<td>0.0004</td>
</tr>
<tr>
<td>Consumer Discretionary</td>
<td>-0.0057</td>
<td>0.0032</td>
</tr>
<tr>
<td>Communication Services</td>
<td>0.0157</td>
<td>0.0061*</td>
</tr>
<tr>
<td>Information Technology</td>
<td>-</td>
<td>0.0052</td>
</tr>
<tr>
<td>Consumer staples</td>
<td>0.012</td>
<td>-0.0017</td>
</tr>
<tr>
<td>Utilities</td>
<td>0.0007</td>
<td>-0.0022</td>
</tr>
<tr>
<td>Health Care</td>
<td>0.0142</td>
<td>0.0017</td>
</tr>
<tr>
<td>Materials</td>
<td>0.0049</td>
<td>0.0017</td>
</tr>
</tbody>
</table>

Note: Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively. Robust standard errors (HAC) have been implemented.

APPENDIX E.: Performance analysis with a matched ESG portfolios

In this section, we report the performance analysis of the matched ESG agreement portfolio and the nonESG portfolio. In respect to this, we implement the propensity score approach using the nearest neighbor algorithm (see e.g., Heckman et al., 1998).21 The selection of the ESG “twin” for every constituent of the nonESG portfolio is based on the Geographical location, Sector and market capitalization. To perform the propensity score match, given the different dimension of the two portfolios (48 for the ESG and 119 for nonESG), we extend the agreement to the stocks that are included at least by two rating agencies in their index. Figures E1 and E2 show the geographical and sectorial composition before and after the match, respectively.

Table E1 shows the estimates of the Carhart four-factor model. Results confirm that the alphas from the long (MESG)–short (nonESG) portfolio is not statistically significant and that the two portfolios are very similar in terms of exposition to the four factors.
TABLE E1 Estimates of the Carhart four-factor model for the matched ESG agreement (MESG) with the nonESG portfolio, and the nonESG portfolio for the two periods, using monthly returns

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MESG</td>
<td>nonESG</td>
<td>MESG-nonESG</td>
<td>MESG</td>
<td>nonESG</td>
<td>MESG-nonESG</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.0041</td>
<td>0.0037</td>
<td>0.0010</td>
<td>0.0054***</td>
<td>0.0053***</td>
<td>0.0005</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Mkt-Rf</td>
<td>0.4972***</td>
<td>0.6669***</td>
<td>−0.0721</td>
<td>0.7152***</td>
<td>0.6838***</td>
<td>0.0749***</td>
</tr>
<tr>
<td></td>
<td>(0.130)</td>
<td>(0.074)</td>
<td>(0.084)</td>
<td>(0.041)</td>
<td>(0.027)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>SMB</td>
<td>0.2523*</td>
<td>0.0353</td>
<td>0.2131</td>
<td>−0.1086</td>
<td>−0.0714</td>
<td>−0.0304</td>
</tr>
<tr>
<td></td>
<td>(0.151)</td>
<td>(0.077)</td>
<td>(0.140)</td>
<td>(0.081)</td>
<td>(0.070)</td>
<td>(0.060)</td>
</tr>
<tr>
<td>HML</td>
<td>0.4331***</td>
<td>0.5479***</td>
<td>−0.0308</td>
<td>0.0688</td>
<td>0.1326**</td>
<td>−0.0491</td>
</tr>
<tr>
<td></td>
<td>(0.133)</td>
<td>(0.088)</td>
<td>(0.066)</td>
<td>(0.078)</td>
<td>(0.067)</td>
<td>(0.052)</td>
</tr>
<tr>
<td>WML</td>
<td>−0.1405*</td>
<td>(0.082)</td>
<td>−0.0811</td>
<td>0.0716</td>
<td>0.0721**</td>
<td>0.0082</td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.056)</td>
<td>(0.056)</td>
<td>(0.045)</td>
<td>(0.034)</td>
<td>(0.031)</td>
</tr>
<tr>
<td>F-Statistic</td>
<td>9.252</td>
<td>59.04</td>
<td>0.7485</td>
<td>88.99</td>
<td>203.5</td>
<td>2.092</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.406</td>
<td>0.752</td>
<td>0.006</td>
<td>0.779</td>
<td>0.831</td>
<td>0.046</td>
</tr>
<tr>
<td>Obs.</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>180</td>
<td>180</td>
<td>180</td>
</tr>
</tbody>
</table>

Note: The third and sixth columns (MESG-nonESG) include the estimate on the long (MESG) – short (nonESG) portfolio. Robust standard errors (HAC) are reported within brackets. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.
APPENDIX F.: Performance analysis with rebalanced portfolios

In this section, we construct the rebalanced ESG agreement and disagreement portfolios which annually track the revision of the four considered indices from January 2012 to December 2019. As shown in Table F1, the alphas are not statistically significant as in the case of the fixed portfolios (no revision) included in Table F2.

<table>
<thead>
<tr>
<th>Table F1</th>
<th>Estimates of the Carhart four-factor model for the rebalanced ESG agreement (ESG AGR), the rebalanced ESG disagreement (ESG DIS) and the nonESG portfolios in the period 2012–2019</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ESG AGR</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.0045*** (0.001)</td>
</tr>
<tr>
<td>Mkt-Rf</td>
<td>0.7816*** (0.045)</td>
</tr>
<tr>
<td>SMB</td>
<td>-0.3756*** (0.098)</td>
</tr>
<tr>
<td>HML</td>
<td>-0.0361 (0.080)</td>
</tr>
<tr>
<td>WML</td>
<td>-0.0263 (0.062)</td>
</tr>
<tr>
<td>F-Statistic</td>
<td>104.4</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.773</td>
</tr>
<tr>
<td>Obs.</td>
<td>96</td>
</tr>
</tbody>
</table>

Note: The last two columns (ESG AGR-nonESG and ESG AGR-ESG DIS) include the estimate on the long – short portfolio. Robust standard errors (HAC) are reported within brackets. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.

APPENDIX G.: Performance analysis with the ESG disagreement portfolio

In this section, we build the ESG disagreement portfolio (i.e., the one where the stocks considered only have one rating agency that included it in its index) and compare it with the performance of the ESG agreement portfolio. Also in this case, if investors’ tastes change towards more sustainable investments, we might expect a difference in terms of alphas. Results in Table G1 show that there is no statistical difference among the alphas of the two portfolios.

<table>
<thead>
<tr>
<th>Table G1</th>
<th>Estimates of the Carhart four-factor model for the (fixed) ESG agreement (ESG AGR), the fixed ESG disagreement (ESG DIS) and the nonESG portfolios in the period 2012–2019</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ESG AGR</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.0047*** (0.001)</td>
</tr>
<tr>
<td>Mkt-Rf</td>
<td>0.8756*** (0.039)</td>
</tr>
<tr>
<td>SMB</td>
<td>-0.2613** (0.104)</td>
</tr>
<tr>
<td>HML</td>
<td>-0.0990 (0.090)</td>
</tr>
<tr>
<td>WML</td>
<td>-0.0487 (0.067)</td>
</tr>
<tr>
<td>F-Statistic</td>
<td>135.9</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.792</td>
</tr>
<tr>
<td>Obs.</td>
<td>96</td>
</tr>
</tbody>
</table>

Note: The last two columns (ESG AGR-nonESG and ESG AGR-ESG DIS) include the estimate on the long – short portfolio. Robust standard errors (HAC) are reported within brackets. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.
### Table G1: Estimates of the Carhart four-factor model for the long-short portfolio between the ESG agreement and the ESG disagreement portfolios

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha</td>
<td>–0.0052 (0.003)</td>
<td>–0.002 (0.002)</td>
</tr>
<tr>
<td>Mkt-Rf</td>
<td>–0.0095 (0.100)</td>
<td>–0.0423 (0.029)</td>
</tr>
<tr>
<td>SMB</td>
<td>–0.1575 (0.415)</td>
<td>–0.1949 (0.148)</td>
</tr>
<tr>
<td>HML</td>
<td>–0.3258*** (0.117)</td>
<td>–0.0655 (0.069)</td>
</tr>
<tr>
<td>WML</td>
<td>–0.0675 (0.061)</td>
<td>0.0191 (0.049)</td>
</tr>
<tr>
<td>F-Statistic</td>
<td>3.578</td>
<td>2.691</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.188</td>
<td>0.017</td>
</tr>
<tr>
<td>Obs.</td>
<td>60</td>
<td>180</td>
</tr>
</tbody>
</table>

Note: The first column refers to the long-short portfolio in the period 2000–2004 while the second refers to the portfolio in the period 2005–2019. Robust standard errors (HAC) are reported within brackets. Statistical significance is denoted by ‘***’, ‘**’, and ‘*’ at the 1%, 5%, and 10% level, respectively.