Inside the ESG Ratings:
(Dis)agreement and performance [*]

Monica Billio[†]

Michele Costola[‡]

Iva Hristova[§]

Carmelo Latino[¶]

Loriana Pelizzon[‖]

November 5, 2020

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

[†]Ca' Foscari University of Venice, billio@unive.it.

[‡]Ca' Foscari University of Venice, michele.costola@unive.it.

[§]Ca' Foscari University of Venice, iva.hristovapegoraro@unive.it.

[¶]SAFE, latino@safe-frankfurt.de.

[‖]SAFE and Ca' Foscari University of Venice, pelizzon@safe.uni-frankfurt.de.

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 preoccupation 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

---

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

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 proposed, 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,[2] the impact of ESG on firms' financial performance,[3] and the effect of ESG scores on credit ratings.[4] A second strand of the literature studies and evaluates the particularities of the ESG rating score and the adopted methodologies highlighting their disagreement.[5]

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 investments' 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

---

[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 Sciarelli (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).

to 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 for the ESG agreement portfolio, there is no impact on the financial performances.

The paper is organized as follows. Section 2 introduces the concept of ESG investing, and presents the ESG investing landscape evolution. 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 disagreement, 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 biodiversity 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's satisfaction. According to Edmans (2011)

there exists a clear positive relationship between employees 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 1984-2009.

At last, the *governance pillar* is related to aspects such as the independence of the board of administration, shareholders' rights, managers' 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 literature. A large majority of studies, concerning both industrialized countries 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 financial performance. Brogi and Lagasio (2019), using MSCI ESG 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 Pacic 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 socially responsible investments (SRI).

However, there are other studies leading to different conclusions, for instance, on the absence of a statistical impact (Landi and 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)[6], 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

---

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

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:

- *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 "nonESG" stocks.

- *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.

- *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.).

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

- *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.

- *Corporate engagement and stock activism.* Exercise of the shareholders' rights aiming to influence corporate behaviour through direct dialogue with corporate management and proposal submissions.

- *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.

INSERT TABLE 1 HERE

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-Okeom 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. 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 Russel 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 RebecoSAM). 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, specic 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 Unt). 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.

INSERT TABLE 2 HERE

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

---

[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.

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 to measure the divergence between the considered rating agencies (Top panel in Table 3).

INSERT TABLE 3 HERE

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. Secondly, we compute the mean absolute error (MAE) in order to evaluate the magnitude of their evaluation differences in term of classes. (Mid panel in Table 3).[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 MAEs of the ESG ratings are on average 1.32 and range from 1.11 to 1.59. Given that the total number of rating classes is 7, it implies that the observed disagreement among rating providers is on average larger than one class. Thirdly, we compute the percentage of agreement among the ESG ratings, describing the proportion of firms in the sample, for which the ESG rating agencies agree

---

[9]Data has been downloaded using Eikon/Datastream, Bloomberg and MSCI ESG database. We thank an anonymous reviewer for the suggestion.

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

to provide the same ESG rating after applying the conversion common scale described above (Bottom panel in Table 3). The table shows that the agreement of the ESG ratings is on average 24% and ranges from 19% (between RobecoSAM and MSCI) to 28% (between Sustainalytics and RobecoSAM). Thus, both measures confirm the existence of a strong evaluation 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 1,000 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

---

[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).

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).

<div align="center">INSERT FIGURE 1 AND FIGURE 2 HERE</div>

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, i.e. 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

---

[12]The overlap coefficient known as the Szymkiewicz–Simpson coefficient is computed such as $\frac{|I_i \cap I_j|}{\min(|I_i|,|I_j|)}$, where I_i and I_j are the ESG indexes, and $|\cdot|$ indicates the cardinality of the set.

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]

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, e.g. 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.[14]

---

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

[14]Detailed results are included in the Appendix B.

Thus, we conclude that the disagreement across ESG indexes persists even after controlling the overlap coefficient for the geographical area.

<center>INSERT TABLE 5 HERE</center>

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.

4 The performance of ESG agreement and nonESG portfolios

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, i.e. 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:

<center>16</center>

- **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.

- **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 that do not comply with the principles of the United Nations Global Compact (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 United Nations Global Compact (UNGC) principles and are involved in controversial or unacceptable activities which may hurt the environment or society as a whole.[15] 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

---

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

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 Eikon/Datastream at monthly frequency.[16] 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%.[17]

<center>INSERT TABLE 6 HERE</center>

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 and 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.[18] After controlling for these

---

[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 $\frac{r_p - r_f}{\sigma}$ and the Sortino ratio as $\frac{r_p - r_f}{\sigma_d}$ where σ_d is the standard deviation of the downside. The Omega ratio is defined as $\frac{\int_\theta^{\inf}[1-F(x)]dx}{\int_{\inf}^\theta F(x)dx}$. 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.

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 (short position). After controlling for the four-risk factors, a positive (negative) significant alpha would imply that the ESG portfolio overperforms (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. 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.

<center>INSERT TABLE 7 HERE</center>

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

<center>19</center>

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.

- **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.

- **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 E.1 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.

- **The rebalanced portfolio.** An additional constraint lies in the non synchronised data availability, i.e the indexes and their constituents' data start at different points of time.[19] For the analysis, we have selected the constituents of the indexes in 2019

---

[19]The Dow Jones Sustainability World index and the STOXX Global ESG Leaders index were

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.

- **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.

- **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

---

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.

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.[20] 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%

---

[20]The 2018 Report and Highlights On US Sustainable, Responsible And Impact Investing Trends is available at `https://www.ussif.org/trends`.

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.

References

Attig, N., El Ghoul, S., Guedhami, O., Suh, J., 2013. Corporate Social Responsibility and Credit Ratings. Journal of Business Ethics 117, 679–694.

Auer, B. R., Schuhmacher, F., 2016. Do socially (ir)responsible investments pay? New evidence from international ESG data. The Quarterly Review of Economics and Finance 59, 51–62.

Aureli, S., Gigli, S., Medei, R., Supino, E., 2020. The value relevance of environmental, social, and governance disclosure: Evidence from Dow Jones Sustainability World Index listed companies. Corporate Social Responsibility and Environmental Management 27, 43– 52.

Berg, F., Kølbel, J., Rigobon, R., 2019. Aggregate Confusion: The Divergence of ESG Ratings. MIT Sloan Research Paper No. 5822-19 .

Bhattacharya, S., Sharma, D., 2019. Do environment, social and governance performance impact credit ratings: a study from India. International Journal of Ethics and Systems 35, 466–484.

Brogi, M., Lagasio, V., 2019. Environmental, Social, and Governance and company profitability: Are financial intermediaries different? Corporate Social Responsibility and Environmental Management 26, 576– 587.

Bruder, B., Cheikh, Y., Deixonne, F., Zheng, B., 2019. Integration of ESG in Asset Allocation. Available at SSRN 3473874 .

Carhart, M. M., 1997. On persistence in mutual fund performance. The Journal of Finance 52, 57–82.

Derwall, J., Bauer, R., Guenster, N., Koedijk, K. G., 2004. The Eco-Efficiency Premium Puzzle. Financial Analysts Journal 61, 51–63.

Devalle, A., Simona, F., Cantino, V., 2017. The Linkage between ESG Performance and Credit Ratings: A Firm-Level Perspective Analysis. International Journal of Business and Management 12.

Edmans, A., 2011. Does the Stock Market Fully Value Intangibles? Employee Satisfaction and Equity Prices. Journal of Financial Economics 101(3), 621–640.

Engle, R. F., Giglio, S., Kelly, B. T., Lee, H., Stroebel, J., 2019. Hedging climate change news. Tech. rep., National Bureau of Economic Research.

Escrig-Olmedo, E., Fernandez-Izquierdo, M., Ferrero-Ferrero, I., Rivera-Lirio, J., Muñoz-Torres, M., 2019. Rating the Raters: Evaluating how ESG Rating Agencies Integrate Sustainability Principles. Sustainability 11, 915.

Escrig-Olmedo, E., Muñoz-Torres, M. J., Fernandez-Izquierdo, M. A., 2010. Socially responsible investing: sustainability indices, ESG rating and information provider agencies. International Journal of Sustainable Economy 2, 442–461.

Fama, E. F., French, K. R., 1992. The cross-section of expected stock returns. The Journal of Finance 47, 427–465.

Friede, G., Busch, T., Bassen, A., 2015. ESG and Financial Performance: Aggregated Evidence from More than 2000 Empirical Studies. Journal of Sustainable Finance & Investment 5, 210–233.

Giese, G., Lee, L.-E., Melas, D., Nagy, Z., Nishikawa, L., 2019. Foundations of ESG Investing: How ESG Affects Equity Valuation, Risk, and Performance. The Journal of Portfolio Management 45, 69–83.

Gompers, P. A., Ishii, J. L., Metrick, A., 2003. Corporate Governance and Equity Prices. Quarterly Journal of Economics 118, 108–155.

GSIA, 2018. Global Sustainable Investment Review. Tech. rep., GSIA.

Halbritter, G., Dorfleitner, G., 2015. The wages of social responsibility — where are they? A critical review of ESG investing. Review of Financial Economics 26, 25–35.

Heckman, J. J., Ichimura, H., Todd, P., 1998. Matching as an econometric evaluation estimator. The review of Economic Studies 65, 261–294.

Henke, H.-M., 2016. The effect of social screening on bond mutual fund performance. Journal of Banking & Finance 67, 89–84.

iShares, 2019. An Evolution in ESG Indexing. Tech. rep., BlackRock.

Jang, G.-Y., Kang, H.-G., Lee, J.-Y., Bae, K., 2020. ESG Scores and the Credit Market. Sustainability 12, 3456.

Kempf, A., Osthoff, P., 2007. The Effect of Socially Responsible Investing on Portfolio Performance. European Financial Management 13, 908–922.

Kiesel, F., Lücke, F., 2019. ESG in credit ratings and the impact on financial markets. Financial Markets, Institutions & Instruments 28, 263– 290.

Landi, G., Sciarelli, M., 2019. Towards a more ethical market: the impact of ESG rating on corporate financial performance. Social Responsibility Journal 15, 11–27.

Leuven, E., Sianesi, B., 2003. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. SocioRePEc .

Lo, K. Y., Kwan, C. L., 2017. The Effect of Environmental, Social, Governance and Sustainability Initiatives on Stock Value – Examining Market Response to Initiatives Undertaken by Listed Companies. Corporate Social Responsibility and Environmental Management 24, 606– 619.

Manrique, S., Martí-Ballester, C., 2017. Analyzing the Effect of Corporate Environmental Performance on Corporate Financial Performance in Developed and Developing Countries. Sustainability 9, 1–30.

Merton, R., 1987. A Simple Model of Capital Market Equilibrium with Incomplete Information. The Journal of Finance 42, 483–510.

Miralles-Quirós, M. M., Miralles-Quirós, J. L., Redondo-Hernández, J., 2019. The impact of environmental, social, and governance performance on stock prices: Evidence from the banking industry. Corporate Social Responsibility and Environmental Management 26, 1446–1456.

Monasterolo, I., De Angelis, L., 2020. Blind to carbon risk? An analysis of stock market reaction to the Paris Agreement. Ecological Economics 170, 106571.

Nofsinger, J., Varma, A., 2014. Socially responsible funds and market crises. Journal of Banking & Finance 48, 180–193.

Ortas, E., Álvarez, I., Garayar, A., 2015. The Environmental, Social, Governance, and Financial Performance Effects on Companies that Adopt the United Nations Global Compact. Sustainability 7, 1932–1956.

Statman, M., Glushkov, D., 2009. The Wages of Social Responsibility. Financial Analysts Journal 65.

Tarmuji, I., Maelah, R., Tarmuji, N. H., 2016. The Impact of Environmental, Social and Governance Practices (ESG) on Economic Performance: Evidence from ESG Score. International Journal of Trade, Economics and Finance 7, 67–74.

UN, 2004. Who Cares Wins: Connecting financial markets to a changing world. Tech. rep., United Nations Global Compact.

Utz, S., 2017. Corporate scandals and the reliability of ESG assessments: Evidence from an international sample. Review of Managerial Science 13, 483–511.

Velte, P., 2017. Does ESG Performance Have an Impact on Financial Performance? Evidence from Germany 8, 169–178.

Weber, O., Scholz, R. W., Michalik, G., 2010. Incorporating sustainability criteria into credit risk management. Business Strategy and the Environment 19, 39–50.

Windolph, S. E., 2011. Assessing Corporate Sustainability Through Ratings: Challenges and Their Causes. Journal of Environmental sustainability 1, 5.

Yen, M.-F., Shiu, Y.-M., Wang, C.-F., 2019. Socially Responsible Investment returns and news: Evidence from Asia. Corporate Social Responsibility and Environmental Management 26, 1565– 1578.

Zhao, C., Guo, Y., Yuan, J., Wu, M., Li, D., Zhou, Y., Kang, J., 2018. ESG and Corporate Financial Performance: Empirical Evidence from China's Listed Power Generation Companies. Sustainability 10.

List of Figures

List of Tables

	MSCI	VIGEO-EIRIS	REFINITIV	SUSTAINALYTICS	ISS OEKOM	ROBECOSAM	ECPI	BLOOMBERG	FTSE RUSSELL
RATING SCORE	CCC to AAA	– to ++	D- to A+ and 0 to 100	0 to 100	D- to A+	0 to 100	F to EEE	0 to 100	0 to 5
HISTORY	1990	1983	2002	1992	1985	1995	1997	2008	2001
HEADQUARTER	New York, United States	Paris, France	Toronto, Canada	Amsterdam, Netherlands	Munich, Germany	Zurich, Switzerland	Milan, Italy	New York, United States	London, United Kingdom
SOURCES	Company disclosure, 1600+ Media sources, 100+ specialized dataset	Company disclosure, Recommendation, Conventions	Company websites, Company reports, NGO Websites, Media and news, Stock Exchange filings	Public disclosure, Media and news, NGO reports	Publicly available information, Interview with stakeholders, information on company policies and practices, company direct contact	Survey approach	Company reports, Media and news, Regulatory data, Bloomberg and Thomson Reuters, University networks	Company reports, Publicly available information, Company direct contact	Publicly available information, Company direct contact, Other sources (governments and NGOs)
N. CRITERIA	37	38	178	155	100	74	80/86	120	300
MAIN RISK FACTORS	**Environmental** Climate Change, Natural Resources, Pollution And Waste Management, Environmental Opportunities **Social** Product Liability, Human Capital, Stakeholder Needs, Social Opportunities **Governance** Corporate Behavior, Corporate Governance	Human Resources, Human Rights, Environment, Business Behavior, Community Involvement, Corporate Governance	**Environmental** Resource Use, Emission, Innovation **Social** Workforce, Human Rights, Community, Product Responsibility **Governance** Management, Shareholders, CSR Strategy	Industry-Specific indicators. Factors Change According To The Industrial Group To Which A Company Belongs	**Environment** Climate Change Strategy, Ecoefficiency, Energy Mgmt, Env. Impact of Product, Env. Mgmt, Water Risk And Impact **Social** Equal Opportunities, Freedom of Association, Health And Safety, Human Rights, Product Responsibility, Social Impact of Product, Supply Chain Mgmt, Taxes **Governance** Business Ethics, Compliance, Independence of The Board, Remuneration, Shareholder Democracy and Structure	About 21 Industry-Specific Indicators. Three Main Dimensions: **Economic** (38/100) **Environmental** (27/100) **Social** (35/100)	**Environmental** Environmental Strategy Policy Environmental Mgmt Products Production Process **Social & Governance** Employees And Human Capital Community Relations Markets Corporate Governance & Shareholder	**Environmental** Carbon Emissions, Climate Change Effect, Pollution, Waste Disposal, Renewable Energy, Resource Depletion **Social** Supply Chain, Political Contributions, Discrimination, Diversity, Community Relations, Human Rights **Governance** Cumulative Voting, Executive Compensation, Shareholders' Rights, Takeover Defense, Staggered Boards, Independent Directors	**Environmental** Biodiversity, Climate Change, Pollution & Resources, Water Security, Supply Chain **Social** Labor standards, Human Rights & Community, Health & Safety, Customer Responsibility, Supply Chain **Governance** Tax Transparency, Risk Management, Corporate Governance, Anti-Corruption
MATERIALITY AND WEIGHTING	Proprietary Definition. Analysis on material risks and opportunities for all the GICS sub-sectors	Proprietary Definition. Based on principles developed by International Bodies.	Proprietary Definition. Standard weighting for all the categories Environmental = 34%, Social 35.5%, Governance= 30.5%	Proprietary Definition of Materiality - subindustry level. Assessment of the potentially material issues in the future	Proprietary Definition. Selection and weighting of 5 key issues per sector and 800 detailed Industry-Specific Criteria	Proprietary Definition. Disclosure of criteria and weighting of the 61 industries analyzed	Proprietary Definition. Based on principles developed by International Bodies (eg. UN Global Compact Initiative and UN PRI)	Proprietary Definition. Based on principles developed by International Bodies (eg. GRI, CDP, SASM for three industries. FSB Task Force on Climate-related Financial Disclosures)	Proprietary Definition. Ratings are calculated using an Exposure-weighted average. Alignement with the UN Sustainable Development Goals (SDGs)

Table 2: Example of divergence in ESG ratings for Verizon Communications Inc., Nissan Motor Co., Ltd, Oracle Corp. JPN and Goodman Group Unt.

Company	Sustainalytics	RobecoSAM	Refinitiv	MSCI
Nissan Motor Co., Ltd	6	77	72	CCC
Verizon Communications Inc.	91	20	67	BB
Oracle Corp. Jpn	78	8	63	BB
Goodman Group Unt	86	21	58	AA

Table 3: Rank correlation (Top panel), mean absolute error (Mid panel) and the percentage of observed agreement (Bottom panel) among the considered rating providers. On average, the rank correlation, the mean absolute error and the percentage of the observed agreement are equal to 0.58, 1.32 and 24%, respectively.

	Sustainalytics	RobecoSAM	Refinitiv	MSCI
Sustainalytics	-			
RobecoSAM	0.69	-		
Refinitiv	0.64	0.69	-	
MSCI	0.53	0.45	0.43	-
	Sustainalytics	RobecoSAM	Refinitiv	MSCI
Sustainalytics	-			
RobecoSAM	1.2812	-		
Refinitiv	1.1878	1.3937	-	
MSCI	1.3260	1.5872	1.1144	-
	Sustainalytics	RobecoSAM	Refinitiv	MSCI
Sustainalytics	-			
RobecoSAM	28.22%	-		
Refinitiv	23.74%	20.59%	-	
MSCI	25.36%	19.46%	27.93%	-

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

Label	Rating Provider	ESG Index	Number of Constituents
STOXX	Sustainalytics	STOXX Global ESG Leaders	431
Dow Jones	RobecoSAM	Dow Jones Sustainability World	317
Refinitiv	Refinitiv	Refinitiv Global ESG	404
MSCI	MSCI	MSCI World ESG Leaders	777

Table 5: Overlap coefficients among four ESG indices. The indices include companies from the developed markets that exhibit high ESG performance according to the related provider. The overall overlap coefficient is 15%.

	STOXX	Dow Jones	Refinitiv	MSCI
STOXX	-			
Dow Jones	50%	-		
Refinitiv	43%	35%	-	
MSCI	59%	49%	50%	-

Table 6: Performance measures for the ESG agreement and the nonESG portfolios for the considered periods. 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%.

	Sharpe ratio		*Sortino ratio*		*Omega ratio*		*MaxDD*		*VaR 5%*	
	ESG	nonESG	ESG	nonESG	ESG	nonESG	ESG	nonESG	ESG	nonESG
2000-2004	0.254	1.224	0.377	1.535	1.209	2.441	-31.9%	-18.9%	-7.3%	-5.0%
2005-2019	0.870	1.044	1.153	1.287	1.953	2.144	-47.7%	-34.5%	-6.7%	-4.7%
2000-2019	0.691	1.092	0.945	1.361	1.694	2.217	-47.7%	-34.5%	-6.9%	-4.9%

Table 7: Estimates of the Carhart four-factor model for the ESG agreement and the nonESG portfolios for the two periods. The last column (ESG-nonESG) includes the estimate on the long (ESG) – short (nonESG) portfolio.

	2000-2004			2005-2019		
	ESG	nonESG	ESG–nonESG	ESG	nonESG	ESG–nonESG
Alpha	0.0049	0.0037	0.0012	0.0059***	0.0053***	0.0006
	(0.004)	(0.003)	(0.005)	(0.001)	(0.001)	(0.001)
Mkt-Rf	0.808***	0.6669***	0.1415	0.8552***	0.6838***	0.1713***
	(0.093)	(0.074)	(0.087)	(0.035)	(0.027)	(0.038)
SMB	0.1888	0.0353	0.1536	-0.1364*	-0.0714	-0.065
	(0.182)	(0.077)	(0.454)	(0.075)	(0.070)	(0.088)
HML	-0.0177	0.5479***	-0.5655***	0.0352	0.1326**	-0.0975
	(0.133)	(0.088)	(0.126)	(0.080)	(0.067)	(0.092)
WML	-0.241***	-0.0811	-0.1599**	-0.1187*	0.0721**	-0.1908***
	(0.062)	(0.056)	(0.079)	(0.067)	(0.034)	(0.055)
F-Statistic	45.63	59.04	18.97	202.5	203.5	11.37
$\bar{R}^2$	0.77	0.752	0.54	0.821	0.831	0.252
Obs.	60	60	60	180	180	180

Robust Standard errors (HAC) are reported within brackets.
Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.

Appendices

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 A.1.

Table A.1: The conversion of the rating classes according to the the four ESG Rating Agencies. Sustainalytics, RobecoSAM and Refinitiv provide a percentile rating while the MSCI' rating scale ranges from CCC to AAA. The final score (first column) we consider ranges from 1 (lowest rating) to 7 (highest rating).

Score	MSCI	Sustainalytics	RobecoSAM	Refinitiv
1	CCC	0-14	0-14	0-14
2	B	15-28	15-28	15-28
3	BB	29-42	29-42	29-42
4	BBB	43-57	43-57	43-57
5	A	58-71	58-71	58-71
6	AA	72-85	72-85	72-85
7	AAA	86-100	86-100	86-100

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 Figures B.1-B.3, 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.

Table B.1: Overlap coefficients among four ESG indices for North America. The indices include companies from North America that exhibit high ESG performance according to the related provider.

	STOXX	DOW JONES	REFINITIV	MSCI
STOXX				
DOW JONES	27%			
REFINITIV	50%	66%		
MSCI	71%	53%	47%	

Table B.2: Overlap coefficients among four ESG indices for Europe. The indices include companies from Europe that exhibit high ESG performance according to the related provider.

	STOXX	DOW JONES	REFINITIV	MSCI
STOXX				
DOW JONES	77%			
REFINITIV	65%	56%		
MSCI	64%	43%	46%	

Table B.3: Overlap coefficients among four ESG indices for Asia Pacific. The indices include companies from Asia Pacific that exhibit high ESG performance according to the related provider.

	STOXX	DOW JONES	REFINITIV	MSCI
STOXX				
DOW JONES	35%			
REFINITIV	36%	39%		
MSCI	60%	21%	60%	

C Descriptive Statistics

Table C.1 reports the descriptive statistics for the ESG agreement and the nonESG portfolios and Table C.2 for the four factors (Carhart, 1997), respectively. Figure C.1 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 C.3 and C.4 the financial indicators for the constituents of both the portfolios.

Table C.1: Descriptive statistics (%) at monthly frequency for the ESG and the nonESG agreement portfolios in the considered periods. The first column provides the average return (mean), the second provides the standard deviation (std), the third provides the minimum return (min) and the last provides the maximum return (max).

	mean		std		min		max	
	ESG	nonESG	ESG	nonESG	ESG	nonESG	ESG	nonESG
2000-2004	0.36%	1.17%	4.90%	3.31%	-12.70%	-8.55%	12.14%	6.65%
2005-2019	1.02%	0.96%	4.07%	3.17%	-15.53%	-12.79%	17.19%	8.56%
2000-2019	0.86%	1.01%	4.29%	3.20%	-15.53%	-12.79%	17.19%	8.56%

Table C.2: Descriptive statistics (%) at monthly frequency for the four factors (Carhart, 1997) in the considered periods. 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 (std), the third provides the minimum return (min) and the last provides the maximum return (max).

	mean	std	min	max
	2000-2004			
MKT-Rf	-0.15%	4.60%	-10.35%	8.71%
SMB	0.44%	3.07%	-10.11%	10.85%
HML	1.65%	3.68%	-10.08%	12.20%
WML	0.24%	6.10%	-14.01%	17.77%
	2005-2019			
MKT-Rf	0.59%	4.30%	-19.51%	11.41%
SMF	-0.04%	1.42%	-3.52%	3.82%
HML	-0.05%	1.71%	-4.61%	4.59%
WML	0.40%	3.33%	-24.26%	9.22%

Figure C.1: Cumulative returns for the ESG agreement (red line) and nonESG (blue line) portfolios during the considered periods.
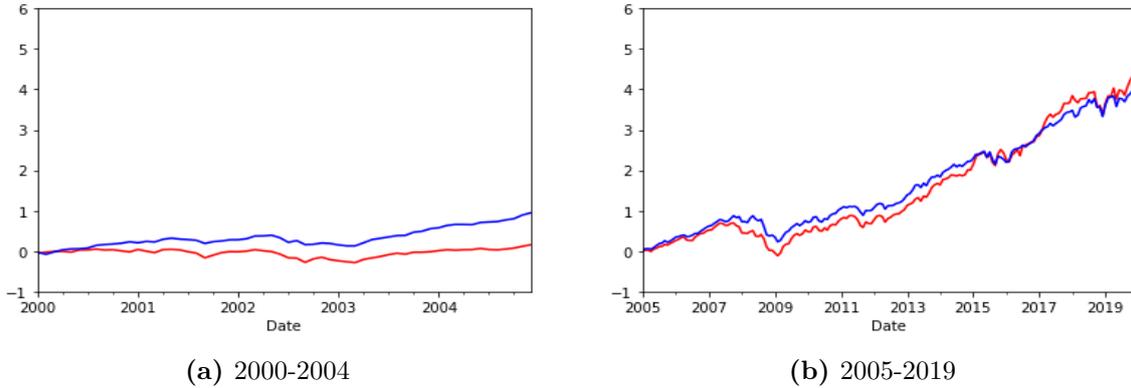

(a) 2000-2004

(b) 2005-2019

Table C.3: Financial indicators of the constituents of the the ESG Agreement Portfolio. The table shows the data for the 10th, 25th, 50th, 75th and 90th percentiles. The last row shows the average values.

	Financial Leverage	Tot. Revenues	Market Value	EPS	EBITDA
10%	12.51	6,284,085	8,183	0.24	1,329,500
25%	25.02	11,588,000	16,858	1.31	2,176,250
50%	40.33	23,640,000	48,181	2.44	4,869,500
75%	62.43	74,142,000	115,118	9.58	11,623,250
90%	79.68	169,568,845	510,698	32.15	55,527,900
Avg	44.14	248,780,227	440,440	24.78	59,240,652

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.

Table C.4: Financial indicators of the constituents of the nonESG Portfolio. The table shows the data for the 10th, 25th, 50th, 75th and 90th percentiles. The last row shows the average values.

	Financial Leverage	Tot. Revenues	Market Value	EPS	EBITDA
10%	16.33	4,926,900	7,012	0.00	902,524
25%	30.71	9,770,250	11,160	0.80	1,636,708
50%	47.27	18,353,000	25,445	3.59	4,094,000
75%	61.72	77,500,498	105,166	14.28	13,390,058
90%	72.26	627,100,000	548,283	137.86	97,292,200
Avg	47.57	358,737,239	194,613	49.58	45,726,947

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.

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 Figure D.1 and confirm that there is no statistical difference among the two portfolios, except for Communication Services in the period 2005-2019.

Table D.1: Estimates of the Jensen's alphas of the long-short ESG agreement portfolios for each sector and the nonESG portfolio for the considered periods.

	2000-2004	2005-2019
Financials	0.0024	0.0038
Industrials	0.0038	0.0004
Consumer discrectionary	-0.0057	0.0032
Communication Services	0.0157	0.0061*
Information Technology	-	0.0052
Consumer straples	0.012	-0.0017
Utilities	0.0007	-0.0022
Health Care	0.0142	0.0017
Materials	0.0049	0.0017

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

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 capitalisation. 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 E.1 and E.2 show the geographical and sectorial composition before and after the match, respectively.

Table E.1 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.

Figure E.1: The composition of the ESG agreement (red bars) and nonESG (blue bars) portfolios.
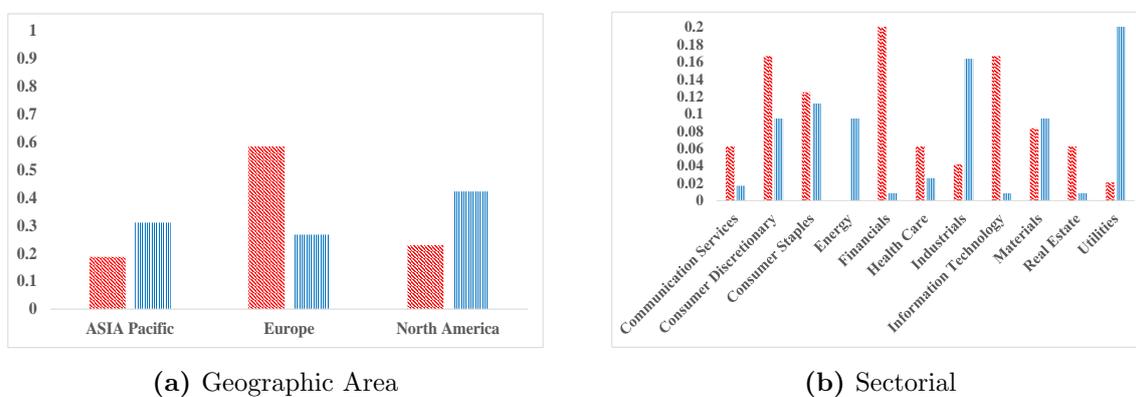
(a) Geographic Area

(b) Sectorial

Figure E.2: The composition of the ESG agreement matched (red bars) with the nonESG portfolio, and the nonESG portfolio (blue bars).

(a) Geographic Area

(b) Sectorial
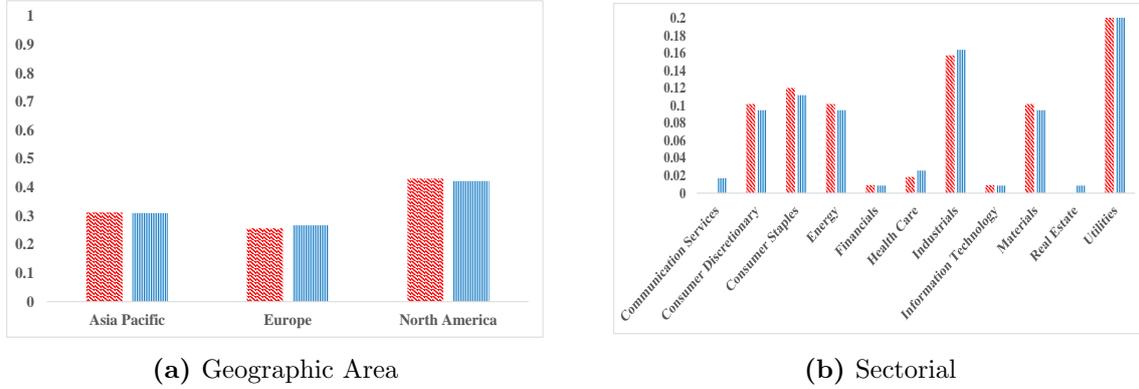
Table E.1: Estimates of the Carhart four-factor Model for the matched ESG agreement (MESG) with the nonESG portfolio, and the nonESG portfolio for the two periods, using monthly returns. The third and sixth columns (MESG-nonESG) include the estimate on the long (MESG) – short (nonESG) portfolio.

	2000-2004			2005-2019		
	MESG	nonESG	MESG-nonESG	MESG	nonESG	MESG-nonESG
Alpha	0.0041	0.0037	0.0010	0.0054***	0.0053***	0.0005
	(0.004)	(0.003)	(0.004)	(0.001)	(0.001)	(0.001)
Mkt-Rf	0.4972***	0.6669***	-0.0721	0.7152***	0.6838***	0.0749***
	(0.130)	(0.074)	(0.084)	(0.041)	(0.027)	(0.027)
SMB	0.2523*	0.0353	0.2131	-0.1086	-0.0714	-0.0304
	(0.151)	(0.077)	(0.140)	(0.081)	(0.070)	(0.060)
HML	0.4331***	0.5479***	-0.0308	0.0688	0.1326**	-0.0491
	(0.133)	(0.088)	(0.086)	(0.078)	(0.067)	(0.052)
WML	-0.1405*	-0.0811	-0.0741	0.0716	0.0721**	0.0082
	(0.082)	(0.056)	(0.056)	(0.045)	(0.034)	(0.031)
F-Statistic	9.252	59.04	0.7485	88.99	203.5	2.092
$\bar{R}^2$	0.406	0.752	-0.006	0.779	0.831	0.046
Obs.	60	60	60	180	180	180

Robust Standard errors (HAC) are reported within brackets.
Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.

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 F.1, the alphas are not statistically significant as in the case of the fixed portfolios (no revision) included in Table F.2.

Table F.1: 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. The last two columns (ESG AGR-nonESG and ESG AGR-ESG DIS) include the estimate on the long – short portfolio.

	ESG AGR	ESG DIS	nonESG	ESG AGR-nonESG	ESG AGR-ESG DIS
Alpha	0.0045***	0.0042***	0.034**	0.001	0.0003
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Mkt-Rf	0.7816***	0.9040***	0.7037***	0.0826*	-0.1177***
	(0.045)	(0.027)	(0.042)	(0.043)	(0.035)
SMB	-0.3756***	-0.0145	-0.2145*	-0.1621	-0.3621***
	(0.098)	(0.052)	(0.104)	(0.104)	(0.088)
HML	-0.0361	0.0705	0.1299	-0.1597**	-0.1003
	(0.080)	(0.058)	(0.080)	(0.080)	(0.066)
WML	-0.0263	-0.0619	0.0420	-0.0656	0.0383
	(0.062)	(0.041)	(0.063)	(0.081)	(0.047)
F-Statistic	104.4	395.5	97.9	5.166	16.1
$\bar{R}^2$	0.773	0.931	0.773	0.047	0.255
Obs.	96	96	96	96	96

Robust Standard errors (HAC) are reported within brackets.
Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.

Table F.2: 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. The last two columns (ESG AGR-nonESG and ESG AGR-ESG DIS) include the estimate on the long – short portfolio.

	ESG AGR	ESG DIS	nonESG	ESG AGR-nonESG	ESG AGR-ESG DIS
Alpha	0.0047***	0.0083***	0.0034**	0.0013	-0.0035
	(0.001)	(0.003)	(0.001)	(0.002)	(0.003)
Mkt-Rf	0.8756***	0.8530***	0.7037***	0.1719***	0.0226
	(0.039)	(0.062)	(0.042)	(0.047)	(0.059)
SMB	-0.2613**	-0.2941	-0.2145*	-0.0468	0.0328
	(0.104)	(0.278)	(0.104)	(0.128)	(0.294)
HML	-0.0990	0.0266	0.1299	-0.2289**	-0.1257
	(0.090)	(0.064)	(0.080)	(0.094)	(0.092)
WML	-0.0487	-0.2156	0.0420	-0.0907	0.1669
	(0.067)	(0.134)	(0.063)	(0.092)	(0.144)
F-Statistic	135.9	137.3	97.9	6.558	3.112
$\bar{R}^2$	0.792	0.705	0.773	0.125	0.022
Obs.	96	96	96	96	96

Robust Standard errors (HAC) are reported within brackets.
Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.

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 G.1 show that there is no statistical difference among the alphas of the two portfolios.

Table G.1: Estimates of the Carhart four-factor Model for the long-short portfolio between the ESG agreement and the ESG disagreement portfolios. The first column refers the long-short portfolio in the period 2000-2004 while the second refers to the portfolio in the period 2005-2019.

	2000-2004	2005-2019
Alpha	-0.0052	-0.002
	(0.003)	(0.002)
Mkt-Rf	-0.0095	-0.0423
	(0.100)	(0.029)
SMB	-0.1575	-0.1949
	(0.415)	(0.148)
HML	-0.3258***	-0.0655
	(0.117)	(0.069)
WML	-0.0675	0.0191
	(0.061)	(0.049)
F-Statistic	3.578	2.691
$\bar{R}^2$	0.188	0.017
Obs.	60	180

Robust Standard errors (HAC) are reported within brackets.
Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.