Disagreement among ESG rating agencies: shall we be worried?

Lopez, Claude and Contreras, Oscar and Bendix, Joseph

Milken Institute, Milken Institute, Milken Institute

20 September 2020

DISAGREEMENT AMONG ESG RATING AGENCIES: SHALL WE BE WORRIED?

Claude Lopez, PhD, Oscar Contreras, PhD, and Joseph Bendix

Milken Institute, Research Department

1. EXECUTIVE SUMMARY

Environmental, “E”, and governance, “G”, issues such as climate change and board composition have received a lot of attention in recent years, leading public firms to disclose related information to help/influence investors’ decision-making process. Some companies and investors have also focused on social issues such as treatment of employees, worker safety, or a company's contribution to community. Yet, these issues have been less prominent in investing decisions. However, recent events—from the on-going social unrest to the impact of the pandemic on health and social inequalities —are changing this landscape, elevating the "S" in ESG investment considerations.

In the absence of a structured framework to report and monitor firms’ efforts on these dimensions, the burden lies on them to communicate on their initiatives and on investors to try to monitor them. New technologies, such as big data analysis or AI, can help process a larger set of information from different sources such as communication strategies of firms or other alternative sources. However, there is the need of defining a core set of variables that would capture these efforts as being part of a long-term strategy beyond the reaction to current events. ESG rating agencies could then process this information and provide their assessment of the firms.

In this study, we show that using a common set of variables would partially resolve inconsistencies and the lack of comparability across rating providers that often confuse investors. Furthermore, we dissociate the impact of the rating agencies’ different focus on “E”, “S” or “G” from that of using different data. While the former, if properly disclosed, can be useful as it allows investors to choose what rating will be more in line with their preferences, the latter necessarily requires harmonization of the data collected.

Using information publicly available, we illustrate how difficult it is to understand or predict some of the existing ratings. Yet, we are also able to identify some commonalities: all rating considered agree on the worst performers. They also reach some consensus when measuring risks arising from governance factors, especially for Corporate Social Responsibility Strategy and Management. Corporate Social Responsibility Strategy includes variables that reflect a company's practices to communicate that it integrates the economic (financial), social and environmental dimensions into its day-to-day decision-making processes. Management includes variables that measure a company's commitment and effectiveness towards following best practice corporate governance principles.

Overall, our study has two main implications when it comes to assessing how well-equipped firms are to deal with ESG risks. First, there is a need for standardization of the data, starting with common disclosure standards, aligning the different ESG disclosure standards existing across the world. The resulting harmonization of the data to be collected would allow rationalizing the reporting burden on the firms while increasing the quality of the data collected. One option would be a centralized data collection via a portal, where each firm would report its information. The different rating agencies, regulators, and other ESG data providers or users would then pull information from this data hub. Ultimately this will increase the firms' participation while improving the rating agencies' credibility in the eyes of investors.

Creating harmonized and high-quality data is only part of the solution. The second implication of our study is the importance of transparency when it comes to the methodologies used to calculate the rating or the focus of the rating. In other words, are "E", "S", and "G" factors equally important? Or does the rating focus mostly on one dimension? Each method of aggregate of the data lead to a different rating, even when using the same data. Having different emphasis across different rating agencies can provide useful information as long as the difference reflects a clear prioritization from the rating agencies, emphasizing the ESG issues they deem the most important. If that is the case, the agencies need to be transparent about it with the rating users, investors, or firms, which in turn will decide which rating is most in line with their priorities.

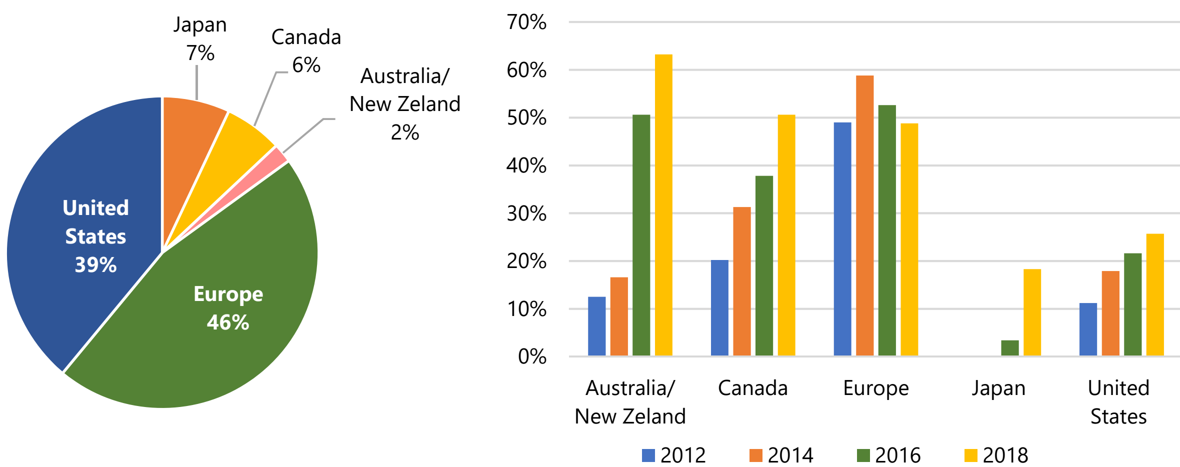
2. INTRODUCTION

Asset owners and managers are increasingly incorporating Environmental, Social, and Governance (ESG) factors into their financial analysis and decision-making processes. According to the Global Sustainable Investment Alliance—an international agency that collects information across Europe, the United States, Canada, Japan, Australia, and New Zealand—, the value of assets under management with an explicit ESG mandate reached US$30.7 trillion at the beginning of 2018, an impressive 34 percent increase relative to 2016. Investment strategies that explicitly incorporate ESG criteria now command a significant fraction of all professionally managed assets across all these regions, ranging from about 18 percent in Japan to more than 50 percent in Canada, Australia, and New Zealand (see Figure 1.b).[1]

Figure 1. Professionally Managed Assets with an ESG Mandate

a) Fraction of Global ESG Investing by Region (as of 2018)

b) ESG Investing as a Fraction of Total Assets under Management, 2012 - 2018



[1] The volume of assets under management with an ESG focus can vary a lot depending on what is included. The numbers in the GSIA report should be considered as broad estimates, as they include multiple investment strategies.

However, when carefully measured, ESG-focused funds remain a low percentage of the total assets under management at the world's largest asset managers (see Table 1). The lack of offering may be one of the explanations (see Figure 2 and Appendix 1).

The increasing focus on ESG investing has led to the rise in the number and relative influence of ESG rating agencies. By providing cost-effective information services on a company's environmental, social, and governance performance, and therefore by potentially alleviating informational asymmetries among market participants, these agencies can play a crucial role in consolidating the market for ESG-related financial instruments. Better information on ESG practices is critical in allowing investors to incorporate these criteria into their decisions in a way that correctly reflects their preferences. Moreover, an independent assessment of a company's environmental, social, and governance performance can also present companies with an opportunity to differentiate themselves, potentially influencing them to adopt better practices to avert downgrades or improve their scores.[2]

Some market participants remain skeptical of the value of the information provided by available ESG rating agencies. A recent survey conducted by Sustainalytics, a major provider of ESG research and ratings, found that while many investors regularly use ratings to inform their decisions, they are usually challenged and sometimes frustrated by them.[3] Inconsistencies and lack of comparability across rating providers, in particular, have created confusion among investors and become a barrier to greater adoption of ESG investing.[4] These discrepancies across ESG ratings also affect company managers, who not only face less pressure to improve their ESG performance but also find it harder to identify appropriate strategies to do so.

Differences across ESG scores can naturally emerge if rating providers adopt different definitions of ESG performance, and as a result, end up assessing different dimensions. Some agencies, for example, may equate ESG performance with a company's ability to comply with specific ethical standards. In contrast, others may do it with a company's ability to manage financially material risks and opportunities arising from ESG factors. To a certain extent, the availability of ratings with different definitions is natural, given the

[2] For an analysis of this "monitoring effect" in a corporate governance context, see Grimminger and Di Benedetta (2013).
[3] Wong and Petroy (2020).
[4] BNP Paribas (2019).

subjective nature of ESG criteria. But more important, it might be required to satisfy assets owners and managers with different needs and motivations. Thus, the focus should not be on agreeing on a single definition but on data-standardization, achieving greater clarity in the labeling of the ratings and more transparency regarding their objectives. These would allow market participants to differentiate products better and to determine whether a particular definition aligns well with their goals.

Inconsistencies across ESG rating agencies are not only an issue of definitions. At least two other reasons can lead rating providers to differing scores on the same company. First, rating providers may disagree on how to measure the same ESG factor. Despite efforts by multiple standard-setting organizations, there is no universally accepted approach to measuring non-financial indicators. Rating agencies employ hundreds of ESG-related variables. Some of them come from company reports and regulatory filings and, therefore, should be consistent across agencies. Yet, many others are privately obtained through interviews or questionnaires and third-party independent reports with potentially conflicting approaches. Second, even if agencies agree on how to measure different ESG-related factors, each ESG agency has developed its own methodology to decide what ESG-related indicators to consider and how to aggregate them into an overall score.

Besides documenting the extend of the disagreement among ESG scores, in this report, we provide some insights into the drivers behind the inconsistencies. We contrast the impact of the data used and of the methodologies. We agree that the lack of data standardization is an issue for both investors and assessed firms, and it should be resolved by harmonization of the data collected and streamlining of the process. However, it is less clear to us that the difference in methodology is a negative thing if it reflects each rating agencies prioritization or specialization in a particular dimension, E, S or G. In that case, it has to be transparent regarding what it is choice providing rating user with a better understanding of the underlying process that least to the assessment. Overall, we hope to inform market participants on how to contextualize and critically evaluate discrepancies in ESG scores and offer some useful information on how to potentially address them.

Our analysis focuses on rating agencies that employ the same definition of ESG performance: a company's ability to manage financially material risks and opportunities arising from ESG factors. As mentioned before, this allows us to concentrate on differences arising from two sources of disagreement: how agencies measure ESG factors and the methodologies they use to aggregate them into a single score.

We shed light on the sources behind the disagreement among ESG rating agencies using an indirect approach. Our indirect method relies on machine learning techniques to identify and estimate the relationship between the ESG ratings and a set of explanatory variables publicly available—which do not (necessarily) coincide with the ones used by the rating agencies. We then compare the identified relationships across the three different rating agencies using various methods. Finally, we assess the ability of our estimated ratings to replicate the disagreement observed among the agencies' ratings.

While all the agencies considered in our study use the same definition of ESG performance, we observe that their ratings coincide -- at a level similar to that usually observed for other types of ratings -- only across the worst performers, which represent a relatively small number of firms. Overall, we not only find substantial discrepancies among rating providers, but we also show that such inconsistencies cannot be easily explain based on information readily available to investors.

It is difficult for an investor to understand or predict a ranking as both the methodology and the data used are different and unavailable for review. Yet the predictive power analysis and the contribution analysis in our study indicate relative consistency across rating agencies on the appropriate way to measure financially material risks arising from governance factors.

More broadly, our analysis has two main implications when it comes to assessing how well-equipped firms are to deal with ESG related risk. First, the use of standardized data will lead to more comparable ratings. This would benefit both the firms being evaluated and the investors using these evaluations, as it would lead to a clearer link between the information and its impact on the assessment. A firm could then decide the appropriate strategy to improve its rating, and an investor would understand the implications of the rating in terms of ESG risk management. Several ESG disclosure standards already exist across the world, which is why an alignment around a common disclosure framework is necessary.[5,6] As our results show, there may be some dimensions, such as governance, for which some consensus may already exist.

The second implication of our analysis is that the lack of standardized data is only part of the problem: the other critical factor is that each rating agency has its own method to aggregate the information. There is a benefit in having different emphasis across different rating agencies. However, such a diversified set of information is useful to the rating users if they are able to understand what each rating is capturing. Only in that case, the users can then decide how to use the information when defining an investment strategy or in making strategic decisions to improve a rating.

The remainder of this report is organized as follows. To establish common terminology, we begin with a discussion of the definition of ESG investing. We then document the extent of disagreement over ESG scores across three major rating agencies at different levels of aggregation. Next, we use machine learning techniques to better understands how the various rating agencies assess a company's ESG performance based on a set of explanatory variables publicly available. Finally, we offer some conclusions drawn from our analysis.

[5] Novick (2020) discusses several issues related to the convergence of ESG disclosure.

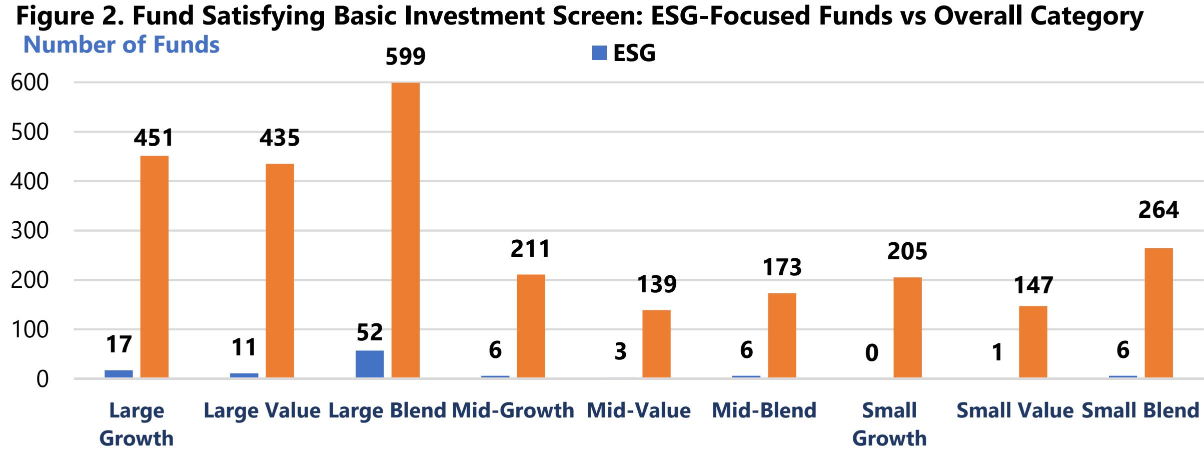
[6] See Clarkin et al. (2020) for a list of the different initiatives across the globe.

Table 1. Assets Under Management in ESG-Focused Funds

Company	AUM ($US Billions)	ESG Investment ($US Billions)	ESG AUM Percentage
BlackRock	$6,470.00	$17.58	0.27%
Vanguard	$6,200.00	$9.54	0.15%
UBS	$3,260.00	$0.29	0.01%
Fidelity	$2,900.00	$0.67	0.02%
State Street	$2,690.00	$0.17	0.01%
Allianz	$2,490.00	$0.21	0.01%
Capital Group	$2,060.00	$0.00	0.00%
JP Morgan Asset Management	$1,900.00	$0.08	0.00%
Goldman Sachs	$1,859.00	$0.13	0.01%
Bank of New York Mellon	$1,800.00	$0.36	0.02%
PIMCO	$1,780.00	$1.96	0.11%
Amundi	$1,653.00	$0.32	0.02%
Prudential Financial	$1,481.00	$0.00	0.00%
AXA Group	$879.00	$0.00	0.00%
Morgan Stanley	$552.00	$6.72	1.22%

Source: Morningstar Direct (7/5/2020).
Note: Funds classified as ESG explicitly stated in their mandates that the investments were chosen primarily for their ESG-risk mitigating characteristics. Keywords in the primary investment mandate also include impact investing, gender/ethnic diversification, and environmental sustainability.

Figure 2. Fund Satisfying Basic Investment Screen: ESG-Focused Funds vs Overall Category



Source: Morningstar Direct (7/5/2020).
Note: Out of 288 ESG-focused funds identified by Morningstar in the U.S., only 104 would pass a simple investment screen commonly employed by fund-of-fund managers: at least three years of historical returns and a fund size over US$50 million (Lauricella, 2020).

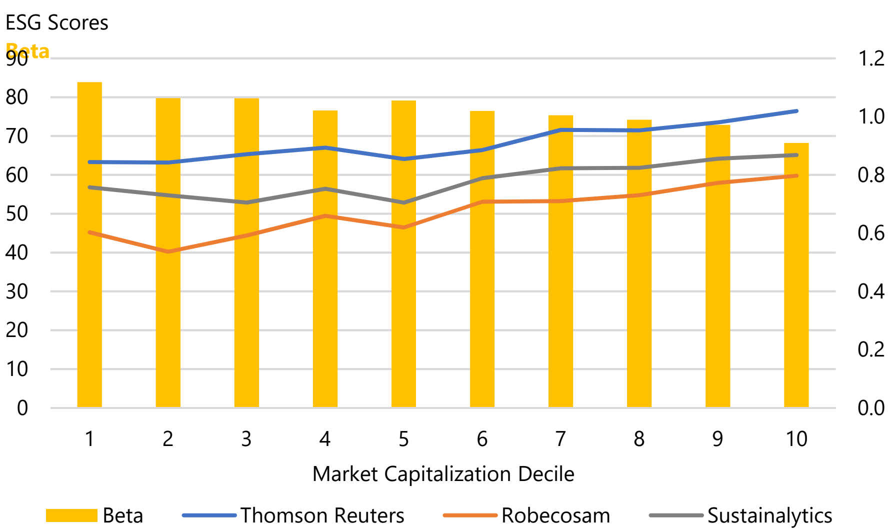
****

BOX 3 **ESG SCORES LEVELS AND FINANCIAL VARIABLES**

This box illustrates the relationship between ESG score levels and some widely used financial variables for the studied group of companies.

After sorting the firms from the largest (10th decile) to the smallest (1st decile) based on their market capitalization, Figure 3 plots the average Beta (a measure of a particular asset's volatility relative to the risk of general systemic market movement) and the average ESG scores for the three rating agencies. All three rating agencies award higher average scores to larger companies. These same firms show overall more resilience (lower Beta) to risks, including ESG ones.

Figure 3. ESG Scores and Beta by Market Capitalization Decile



Source: Authors' calculations using Bloomberg and Refinitiv Eikon data.

---

3. WHAT IS ESG INVESTING?

Although there is no universally excepted definition, ESG investing is widely understood as an investment approach that goes beyond the analysis of traditional financial indicators by incorporating Environmental, Social, and Governance (ESG) factors into the investment process (i.e., the process of selecting and managing an investment portfolio). Of course, ESG considerations are not entirely new, and, in various ways, many investors have long incorporated some of these issues into their investment frameworks. The modern reference to ESG investing, however, denotes a more explicit and systematic integration of ESG factors into the investment process, as opposed to a more informal, less structured approach.

Investors can have multiple motivations

Investors seek to integrate ESG factors into their financial decisions for various (not mutually exclusive) reasons. (See Box 2 for a list of factors commonly referred to as ESG.)

- Some investors may consider that ESG data can help paint a broader picture of a company's operating environment. Accordingly, they rely on ESG investing to identify and manage risks and opportunities that cannot be easily detected through standard financial analysis—that is, as *a source of financial value*. According to Dan Hanson, former managing director at BlackRock, "*ESG is a proxy for risk that is not priced in, and companies that better manage these risks can deliver returns with greater certainty…*"[7] Reducing exposure to polluters or companies with poor waste management policies, for example, can help mitigate regulatory risk, whereas screening for good social practices (such as workplace culture, human rights protection, or corporate community engagement) can reduce exposure to scandals that could damage a company's reputation.[8]
- Other investors rely on ESG investing to meet their values (e.g., ethical, religious, political, or cultural) or to promote specific environmental, social, or governance outcomes they deem desirable. Investors, for instance, may integrate ESG factors into their financial decisions to identify and exclude companies engaging in practices they find morally questionable, such as low labor standards or human rights violations. These investors might seek to advance their non-financial

[7] Cited in Koehler and Hespenheide (2013).
[8] For studies on the relationship between ESG performance and profitability, see Friede et al. (2015) and, more recently, Verheyden et al. (2016).

objectives without hampering financial objectives. In some cases, they might even be willing to sacrifice financial returns to achieve their non-financial goals. A recent survey conducted by UBS among asset owners across 46 countries found that "doing good for society and the environment" is among the top four drivers behind ESG investing. [9]

- And still others, such as institutional investors or financial advisors acting on behalf of a third party, may rely on ESG criteria to satisfy specific legal requirements. One of the world's largest investment funds, for example, the Norwegian Government Pension Fund Global, is mandated to avoid companies that contribute or are responsible for "*serious or systematic human rights violations,…, serious violations of the rights of individuals in situations of war or conflict, severe environmental damage,…, gross corruption, [or] other particularly serious violations of fundamental ethical norms.*"[10]

Multiple labels for similar issues

Despite its growing popularity, there are substantial terminological and conceptual inconsistencies surrounding ESG investing. Phrases such as sustainable, responsible, or socially responsible investing are sometimes conflated or used interchangeably with the term ESG investing, while other times, they are used to denote related but conceptually different ideas. Understandably, the broad array of terms used to describe various ESG approaches, together with a lack of consistency in their use, has created confusion among investors. A recent survey conducted by State Street Global Advisors found that over half of those investors already implementing some type of ESG strategy within their portfolio were struggling with a lack of clarity around ESG terminology in their organizations.[11]

To reduce confusion among investors, and because the common theme underlying all the different labels is an emphasis on ESG issues, we believe that the more neutral term ESG investing is appropriate. Accordingly, we see ESG investing as an *umbrella term*—one that, as mentioned above, refers to an investment approach that involves some type of environmental, social, or governance consideration, that

[9] See, for example, Fritsch (2019).
[10] Norway' Ministry of Finance (2019).
[11] State Street Global Advisors (2018).

can have various motivations, and that, depending on the investor's goals, resources, and circumstances, may result in different strategies.[12]

[12] For a detailed discussion on how to incorporate ESG factors into the investment process, see Grim and Berkowitz (2018).

BOX 2:

ENVIRONMENTAL, SOCIAL, AND GOVERNANCE (ESG) FACTORS

Broadly defined, environmental factors focus on a company's environmental impact, social factors examine how it manages relationships with different stakeholders—such as customers, employees, suppliers, and the communities within which it operates—, and governance factors deal with a company's leadership, internal controls, and shareholder rights.

ESG factors cover a wide range of topics and the relevant issues are likely to depend on the company being analyzed, its industry, and, ultimately, on the investor's preferences and objectives. For these reasons, it should not be surprising that a definitive list of ESG factors does not exist.

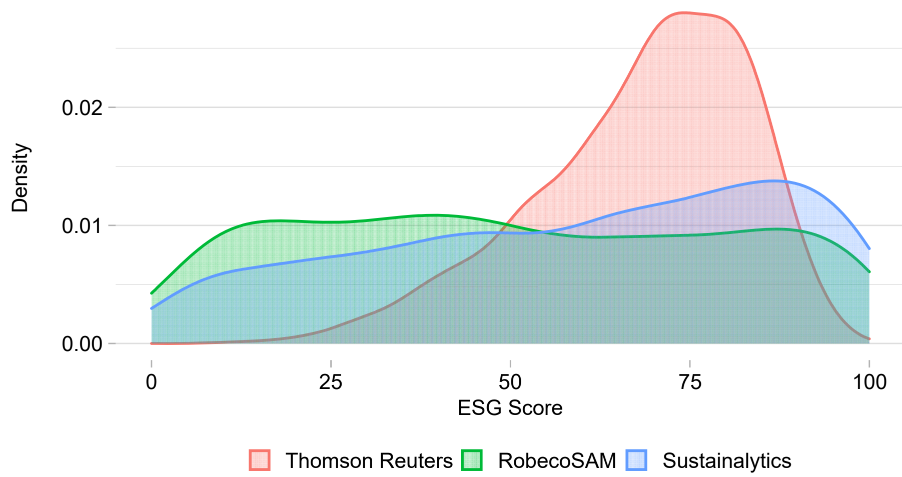
The table below displays some examples of well-known ESG factors.

Environmental	Social	Governance
- Climate change policies, plans, and disclosure practices - Air and water pollution - Deforestation - Biodiversity impact - Water stress - Waste and hazardous materials management - Usage of renewable Energy	- Community engagement - Human rights - Labor practices - Product safety - Data security and customer privacy - Diversity and inclusion - Customer relations - Ethical supply chain sourcing	- Management structure - Executive compensation - Board composition - Business integrity - Transparency - Bribery and corruption - Lobbying - Whistleblower schemes - Shareholder relations

4. DISAGREEMENT AMONG ESG RATING AGENCIES

Our analysis considers three major rating agencies that emphasize the financial materiality of ESG factors when measuring a company's ESG performance: RobecoSAM, Sustainalytics, and Thomson Reuters.[13] As mentioned in the introduction, considering only rating agencies that agree on a definition of ESG performance allows us to concentrate on differences arising from how agencies measure ESG factors and the methodologies they use to aggregate them into a single score. Our sample contains annual information on 943 firms for the year 2018, the latest for which all three ESG scores were available.[14] The data were collected from Bloomberg and Refinitiv-Eikon.

Figure 4. ESG Score Distributions



Source: Bloomberg and Refinitiv Eikon.

[13] According to Gaffuri (2017), RobecoSAM's methodology seeks to identify "…any [ESG] factor which might have a present or future impact on companies' value drivers, competitive position, and thus on long-term shareholder value creation." According to Sutainalytics (2019), its rating "measure[s] the degree to which a company's economic value is at risk driven by ESG factors." And according to Thomson Reuters (2017), its rating helps to "easily identify companies with…exposure to ESG risks."

[14] To construct our sample of firms, we started with the 2000 largest companies by market capitalization. We then excluded companies for which we were unable to procure information on all three different ESG scores, as well as companies for which a substantial fraction of the explanatory variables used in the following section was missing. For multiannual scores, we consider the last available for 2018.

Table 2. Correlations between ESG Ratings

Pair of Scores	Correlation
RobecoSAM vs Sustainalytics	0.72
RobecoSAM vs Thomson Reuters	0.65
Sustainalytics vs Thomson Reuters	0.65

Source: Authors' calculations using Bloomberg and Refinitiv Eikon data.
Note: The correlations are the Pearson product-moment correlation coefficients.

A simple glance at the distributions of ESG ratings (see Figure 4) confirms that the agencies' assessments of the firms are quite different: most of Thomson Reuters scores are concentrated around high values, between 50 to 80, while RobecoSAM and Sustainalytics spread them mostly evenly between10 and 90.

The pairwise correlations, reported in Table 2, confirm that RobecoSAM and Sustainalytics tend to agree the most in their assessment with a correlation of 0.72. Yet, this level of agreement is significantly lower than the one usually encountered among credit rating, with an average correlation of 0.986.[15]

4.1. DISAGREEMENT BY ECONOMIC SECTOR

A look at the economic sectors, with Figure 5 for correlations and Table 3 for a short description of the sectors including their unique regulatory and financial characteristics, allows us to derive more granular insights on the differences: [16]

- The overall level of agreement among rating agencies (i.e., the average pairwise correlation between ESG scores) varies substantially across sectors, ranging from 0.50 in Energy to 0.77 in Technology.

[15] For other studies reporting correlations among ESG rating agencies, see Berg et al. (2020), Gibson et al. (2019), and State Street Global Advisors (2019).
[16] We use the Thomson Reuters Business Classification to assign each company into one of ten different economic sectors.

- The highest within-sector heterogeneity in the level of agreement among rating agencies occurs in the sector with the lowest correlation, Energy.[17] The companies in this sector may be harder to evaluate, as they are highly regulated or because significant investments in infrastructure make it harder to identify the relevant ESG risks and the appropriate strategies to deal with those risks.
- Sectors with a higher level of agreement among rating agencies, such as Financials, Technology, and Cyclical Consumer Goods & Services, seem to have less emphasis on environmental factors, particularly the first two. This could indicate, for example, more consistency across rating agencies on the appropriate way to measure financially material risks arising from social and governance factors.

Overall, the three rating agencies give very different ESG scores, with a correlation below 0.5, to more than 60% of the firms. In contrast, they have a very similar assessment, with a correlation of 0.95 or more, for only 10% of the firms, the worst-performing ones. (See Appendices 3 and 4 for an analysis of disagreement by market capitalization decile and at the firm level.)

[17] The higher heterogeneity in the Energy sector should be taken carefully, for it is also one of the sectors with the lowest number of observations (48).

Substantial discrepancies in ESG scores across rating agencies is a problem for both investors and companies. Investors may have difficulties in integrating ESG factors into their portfolios in a manner that reflects their preferences. Companies could be discouraged from improving their ESG performance, as they may not be able to identify an appropriate strategy to do so, or they may find the outcome too uncertain and not worth the investment.

Figure 5. Correlations between ESG Scores by Economic Sector

Source: Authors' calculations using Bloomberg and Refinitiv Eikon data.

Table 3. Economic Sectors: Description and Unique Features

Sector	Description	Unique Financial & Regulatory Characteristics
Basic Material (68)	Companies involved in the discovery, development, and processing of raw materials, including mining and metal refining, chemicals, and packaging. Recognizable names in the sector include Ecolab, Dupont, and Dow.	Companies in this sector supply most of the materials used in construction. Thus, they are sensitive to changes in the business cycle and tend to thrive when the economy is strong, exhibiting a rather high Beta of 1.13 on average.
Consumer Cyclical (120)	Companies that produce elastic, or non-essential goods and services purchased by individuals and households such as Automobiles (Ford/GM), Specialty Retailers (Amazon), Hotels & Entertainment (Marriott International), and Media-Publishing (ViacomCBS).	Compared to the Consumer Non-Cyclical sector, the Consumer Cyclical sector has higher profit margins than the, but its demand is more sensitive to the business cycle. The sector has a rather reactive Beta to the market, at 1.17. Consumer Non-Cyclical companies trade at the lowest sector average of 2.1x sales.
Consumer Non-Cyclical (82)	Companies that produce inelastic or essential goods and services purchased by individuals and households. Industries within the sector include Food and Drug Retailers (e.g., Walmart), Food and Tobacco producers (e.g., General Mills), Beverage producers (e.g., Coca-Cola), and Personal & Household Products/Services (e.g., Proctor & Gamble).	Within the Consumer Non-Cyclical sector, businesses provide goods/services that have a relatively inelastic demand. Due to this inelasticity, Consumer Non-Cyclical companies can employ larger debt levels relative to other sectors, utilizing leverage to increase ROE. Consumer Non-Cyclicals exhibit a comparatively smaller average Beta at just .65.
Energy (48)	The energy sector includes companies involved in the exploration and development of oil or gas reserves, oil and gas drilling, and refining like recognizable names Exxon Mobil, Chevron, Occidental Petroleum and Schlumberger.	Companies in the Energy sector incur large capital expenditure costs to create and maintain core business activity infrastructure. Energy providers are extremely susceptible to output pricing and supply and demand shocks, leading to the highest average Beta across sectors (at 1.36). The industry also pays out the largest dividend yield to investors, averaging 7.06% on an annual basis.
Financials (226)	The largest represented sector in the S&P 500 by number of firms. It includes large banking institutions (e.g., JP Morgan Chase and Bank of America), payment services (e.g., American Express), as well as insurance and asset management institutions (e.g., BlackRock and MetLife).	The Financials sector treats debt fundamentally different from all other economic sectors, utilizing it as a revenue-generating asset from a lender/investor perspective. This creates the widest discrepancy between enterprise value and market capitalization at 2.09:1 ratio among the economic sectors. Financials is more volatile than the overall market, with an average Beta of 1.08. Return on Equity for the sector was 12.01%, below the sector-agnostic average of 27%. The Financials sector is also highly regulated and therefore affected by governmental decisions.

Healthcare (83)	The Healthcare sector consists of companies that provide medical services (UnitedHealth Group/Cigna), healthcare equipment and devices (Johnson & Johnson/Thermo Fisher Scientific), and Pharmaceuticals/Biotechnology (Gilead/Pfizer/Merck).	Because of the necessity of its products, the Healthcare sector has a Beta (.98) that most closely mirrors the S&P500, while generating the 2nd highest average ROE at 31%. Influenced by outliers within the highly volatile biotechnology industry, the Healthcare sector has by far the largest average EV/EBIT valuation multiple at 111x, ranging from 7x to 7,152x. The sector also exhibits the second highest average Price-to Earnings ratio at 38, partly due to the highly regulated FDA approval process (with successful drug patents allowing for monopolies on certain drug/treatment advancements that possess pricing power to recoup R&D costs).
Industrials (132)	Enterprises that produce machinery (Boeing/Caterpillar), passenger and material transportation (Delta/UPS), and Aerospace & Defense (Lockheed Martin/Raytheon) all fall under the industrials umbrella.	The most diverse sector in terms of products or services, Industrials exhibits the largest range of ROE in the S&P500, returning anywhere between -225% and +766%. Industrials also exhibits comparatively lower valuation multiples on average: 14x EV/EBITDA, 16x EV/EBIT, 2.8x EV/Sales, & 21 P/E.
Technology (96)	The Technology sector offers a wide range of products and services for both customers and other businesses. Industries within the Technology sector include Software & IT (e.g., Microsoft), Communications & Networking (e.g., Facebook), Computers, Phones, Household Electronics (e.g., Apple), and Office Equipment (e.g., Cisco).	The Technology sector is unique in a myriad of ways, and contrary to other sectors, profit takes a back seat to growth, and operating metrics are not as pertinent to the valuation discussion. Because of this growth focus, operators in this sector tend to shy away from debt financing, exhibiting a comparatively low 82% Debt-to-Equity Ratio on average for 2018. The propensity for equity financing provides for larger cash-on-hand in the balance sheet, making it the only sector in the S&P500 who's average Market Capitalization is actually greater than the Enterprise Value of the firm. Strong cash infusions through equity offerings allow tech companies to possess the largest average Current and Quick Ratios on the balance sheet, at 2.35 and 2.14, respectively. The Technology sector is characterized by high average valuation multiples, trading at 22x EBITDA, 5.5x Sales, and 52x Earnings, the highest of any sector.
Telecom (29)	The Telecommunications sector consists of companies that transmit data in words, voice, audio, or video across the globe. Recognizable names in the sector include AT&T, Verizon, T-Mobile, and CenturyLink.	While the sector remains concentrated, it is moving towards a more decentralized system with less regulation and barriers to entry. Beta is much lower than average at .62. Because firms often operate on a subscription and revenue recognition model, dividend yields are larger than in most other economic sectors at

		an average of 5.52% yield per year, second only to Energy.
Utilities (59)	The Utilities sector includes companies that provide basic amenities, such as water, sewage services, electricity, dams, and natural gas. Some of the larger names in the sector are Nextera Energy, Duke Energy, Edison International, and Sempra.	Utilities are part of the public service landscape and therefore heavily regulated. It typically offer stable and consistent dividends (4.47%), coupled with less price volatility relative to equity markets, possessing the smallest average Beta at .61. Because of the inelastic nature of the products and services provided, Utilities companies do not need the same type of balance sheet cash cushion required in other Economic Sectors, allowing them to possess the lowest average Quick and Current Ratios of any sector at .85 and .93, respectively.

Source: Authors' calculations using Bloomberg and Refinitiv Eikon data.
Notes: The number in parenthesis below the sector name indicates the number of companies in our sample.

5. WHAT IS DRIVING THE DISCREPANCIES IN ESG SCORES?

Understanding what drives these discrepancies is essential to make sense of them. Not having access to the raw data or to the detailed methodologies employed by the different ESG rating agencies, our analysis of their disagreement relies on an indirect approach that uses publicly available information. It consists of three steps:

1. **Collection of publicly available ESG and other indicators for the firms studied**. A total of 207 ESG indicators (58 related to environmental factors, 70 to social factors, and 79 to corporate governance factors), as well as 35 financial variables and information on both headquarters location and economic sector.[18]
2. **Estimation of the relation between the ESG ratings and the explanatory variables**. Standard econometric techniques cannot easily handle a large number of variables, and they usually require specifying a particular structure on the relationships among variables. As an alternative, we use a machine learning technique called random forest. Random forest models

[18] The data were collected from Refinitiv-Eikon, a major provider of financial news and information. A detailed list of all the explanatory variables is provided in available upon request.

can accommodate complex, non-linear patterns and can handle different types of variables efficiently.[19]

3. **Comparison of the estimation results across rating agencies** that look at three distinct and complementary angles: (i) the variables' ability to predict the ESG scores, (ii) their contribution to the ratings predicted by our estimation, and (iii) the importance of variables interaction when predicting the ESG scores. Exercises (i) and (ii) tell us how informative individual variables are regarding the content of the ratings. On the other hand, (iii) tell us something about how that information is aggregated into a single score (not how agencies actually do it, but how it is done in terms of the estimated relations between ratings and explanatory variables). Finally, we compare the disagreement among the predicted ESG ratings with the one observed among the agencies' ratings.

[19] In contrast to other algorithms, random forest models also generate an internal measure of the model's ability to predict previously unseen observations, thereby eliminating the need to use a separate dataset to evaluate their performance.

BOX 4

RANDOM FOREST MODELS: A PRIMER

A random forest is a machine learning algorithm. It combines the outcomes of a large number of individual decision trees to generate a single prediction, either by calculating the average (when the prediction variable is continuous) or by implementing a "majority vote" (when the prediction variable is categorical).[20] Unsurprisingly, the model is called a forest because it relies on a multiplicity of decision trees. But what exactly is a decision tree? Why do we need many of them? And in what sense is the forest random?

A decision tree is a predictive algorithm that, as its name implies, uses a tree-like structure to predict the value of a target variable using a set of explanatory variables. A decision tree starts with a single node, which then branches into possible outcomes based on the value of one of the explanatory variables. Each of those outcomes leads to additional nodes, which once again branch off into other possibilities based on another explanatory variable, giving it a tree-like shape. This process continues until a terminal node is reached, which leads to no additional sub-nodes and contains our prediction for the variable of interest. Decisions regarding what explanatory variables to use at each node, and how to use them to split the tree, are taken sequentially (from top to bottom) and are based on the gain in precision induced by the split.

Although decision trees provide a very intuitive modeling approach, they tend to perform poorly when predicting previously unseen observations (i.e., observations that were not used to estimate the model). This poor performance occurs because decision trees suffer from a problem called "high variance." Since decision tree models are incredibly flexible, they tend to overfit the data used to estimate them. As a result, decision trees tend to capture not only the actual relationship between predictors and outcome but also the noise contained in the sample (which results in poor predictive performance).

Various techniques (such as pruning, minimum node size, and maximum number of terminal nodes) can mitigate overfitting, but estimating a random forest is one of the most common approaches. The basic idea is simple: by combining a large number of "imperfect" decision trees, we can "average out" their

[20] For a detailed discussion, see Breiman (2001).

individual mistakes and dramatically improve the accuracy of our predictions. This approach, however, requires that each decision tree in the forest be different so that it provides new information. It is here where the "random" part of the model becomes relevant. Although ideally we would like to estimate each decision trees using a different sample from the population of interest, this is rarely feasible. Instead, we can achieve something similar by injecting randomness into the tree-growing process by doing the following: 1) estimating each tree using a different random sample with replacement drawn from the original dataset, and 2) every time we must decide how to split a node, limiting the search to a randomly selected subset of explanatory variables.

5.1. IS IT ABOUT THE DATA?

We use data publicly available on the firms to identify what information the ESG ratings are capturing. Although these variables do not necessarily coincide with those employed by the rating agencies, we can expect them to be related to the various ESG ratings—and therefore to be informative about their content. Furthermore, using the same variables across the ratings allow us to indirectly assess the impact of standardization of the information.

5.1.1. VARIABLE'S PREDICTIVE POWER[21]

One way to do that is by assessing the ability of the explanatory variables, individually or grouped, to predict the ESG scores provided by the rating agencies.

[21] Our analysis is based on two of the most widely used measures, Mean Decrease in Impurity and Perturbation Importance, using Li *et al*. (2019) and Breiman (2001), respectively.

Table 4. Top Ten Predictors for ESG Scores

Thomson Reuters	RobecoSAM	Sustainalytics
	Environmental Variables	
• Targets Emissions • Resource Reduction Policy • Policy Emissions • Environmental Supply Chain Management • Policy Environmental Supply Chain • Environment Management Training • Policy Energy Efficiency	• Targets Emissions • Renewable Energy Use • Resource Reduction Targets	• Targets Emissions • Renewable Energy Use • Environmental Supply Chain Management • Policy Environmental Supply Chain • Resource Reduction Targets
	Social Variables	
• Flexible Working Hours	• ILO Fundamental Human Rights • Human Rights Contractor • Policy Human Rights	• ILO Fundamental Human Rights • Human Rights Contractor
	Governance Variables	
• CSR Reporting • Independent Board Members	• CSR Reporting • Stakeholder Engagement • Global Compact Signatory • Board Gender Diversity	• CSR Reporting • Stakeholder Engagement • Global Compact Signatory

Source: Authors' calculations using Bloomberg and Refinitiv Eikon data.

First, focusing on the top ten variables with the highest predictive power for each of the ESG scores, Table 4 shows that: [22]

- *The factors have different predictive power across the ratings.* Although environmental factors seem to be important predictors for all three ESG scores, they are disproportionally so for Thomson Reuter. By contrast, RobecoSAM and Sustainalytics appear to offer a more balanced picture across environmental, social, and governance indicators.

[22] The top predictors were chosen by ranking all explanatory variables in ascending order according to each of our two measures and selecting the first ten variables to appear in both rankings.

- *Very few factors overlap across the three agencies*. Of the top ten predictors, only two are common among all rating providers: *Targets Emissions* and *CSR Reporting*.[23] However, RobecoSAM and Sustainalytics share eight common top predictors.

Second, we extend the analysis to all the variables. To do so we aggregate them in categories when assessing their predictive power for the different ratings. Figure 6 reports the outcome when considering five broad categories: environmental, social, governance, financial, and others. Figure 7 expands the analysis to 18 subcategories: three environmental, four social, three governance, six financial, and two related to other factors.

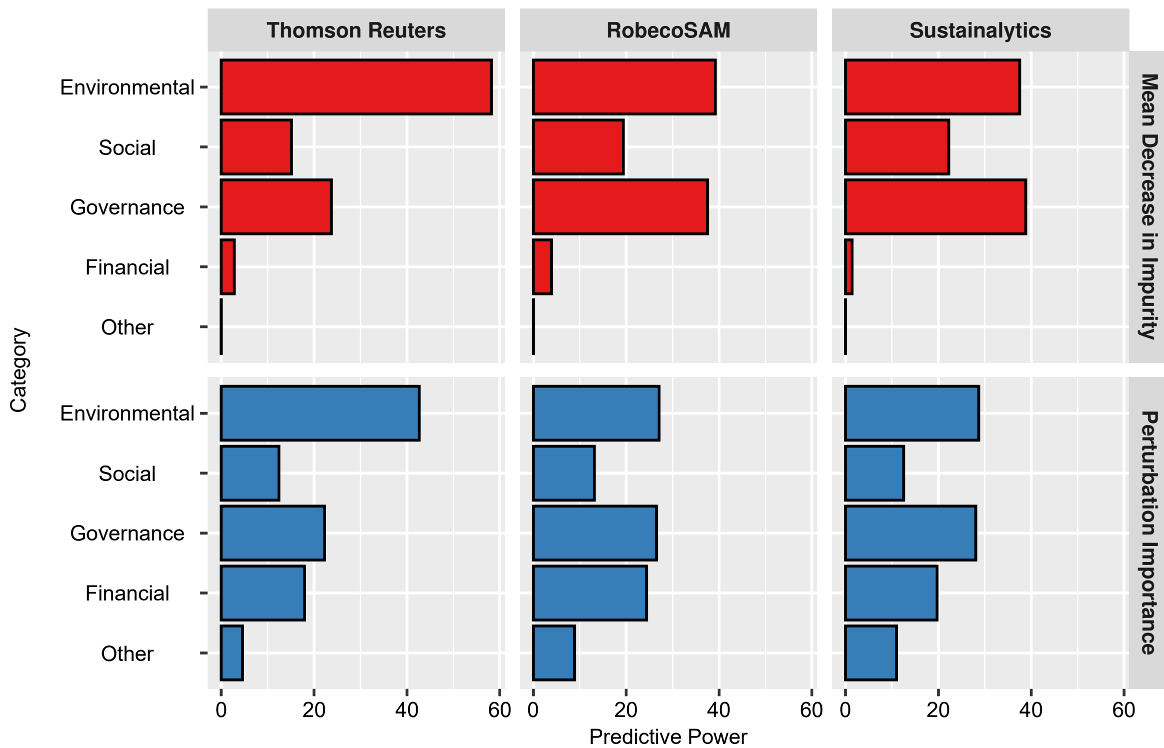
- **The overall environmental and governance factors have the highest predictive power** for all three ESG scores, followed by social and financial considerations—in no particular order—and, finally, by other factors.
- ***Emissions* and *Resource Use* have the most predictive power for environmental factors**. *Emissions* refers to variables that measure a company's commitment and effectiveness towards reducing environmental emissions in the production and operational processes. *Resource Use* refers to variables that reflect a company's performance and capacity to reduce the use of materials, Energy, or water, and to find more eco-efficient solutions by improving supply chain management. The subcategory *Innovation*, which includes variables that reflect a company's capacity to reduce its environmental impact through new environmental technologies and processes, shows little power.
- **CSR *Strategy* and *Management* capture most of the predictive power of Governance factors across the ratings**.[24] Yet, *Management* is significantly more relevant than *CSR Strategy* in predicting Sustainalytics ESG scores. The results also confirm our previous finding that the relative importance of environmental variables is significantly higher for Thomson Reuters that for the other two rating agencies.

[23] *Targets Emissions* measures whether a company has set and achieved short-term and long-term reduction targets to reduce emissions to land, air, or water from business operations. *CSR Reporting* measures a company's efforts to publish a report on Corporate Social Responsibility, Health, and Safety, or Sustainability issues.

[24] *CSR Strategy* includes variables that reflect a company's practices to communicate that it integrates the economic (financial), social and environmental dimensions into its day-to-day decision-making processes. On the other hand, *Management* includes variables that measure a company's commitment and effectiveness towards following best practice corporate governance principles.

- **Among social variables, *Human Rights* and *Workforce* have the highest predictive power across all agencies**, whereas *Product Responsibility* has the lowest.[25] However, while *Workforce* is the most critical social subcategory for Thomson Reuter, *Human Rights* is the top predictor for RobecoSAM and Sustainalytics.

Figure 6. Predictive Power by Category

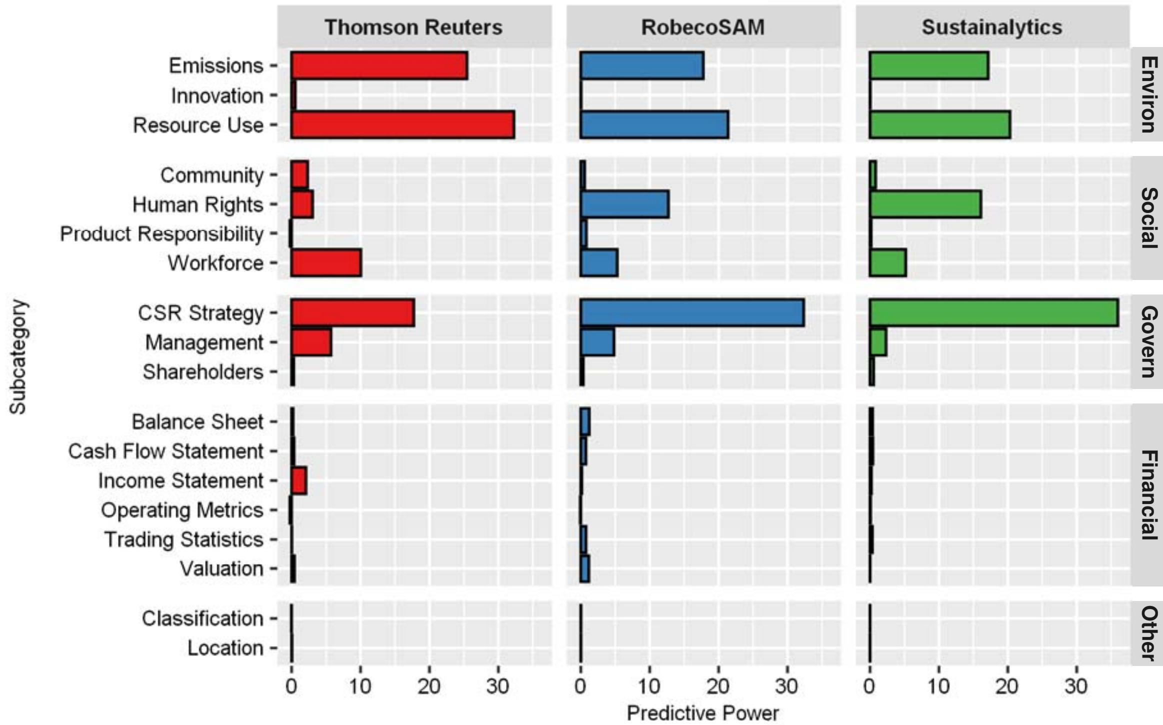


Source: Authors' calculations using Bloomberg and Refinitiv Eikon data.

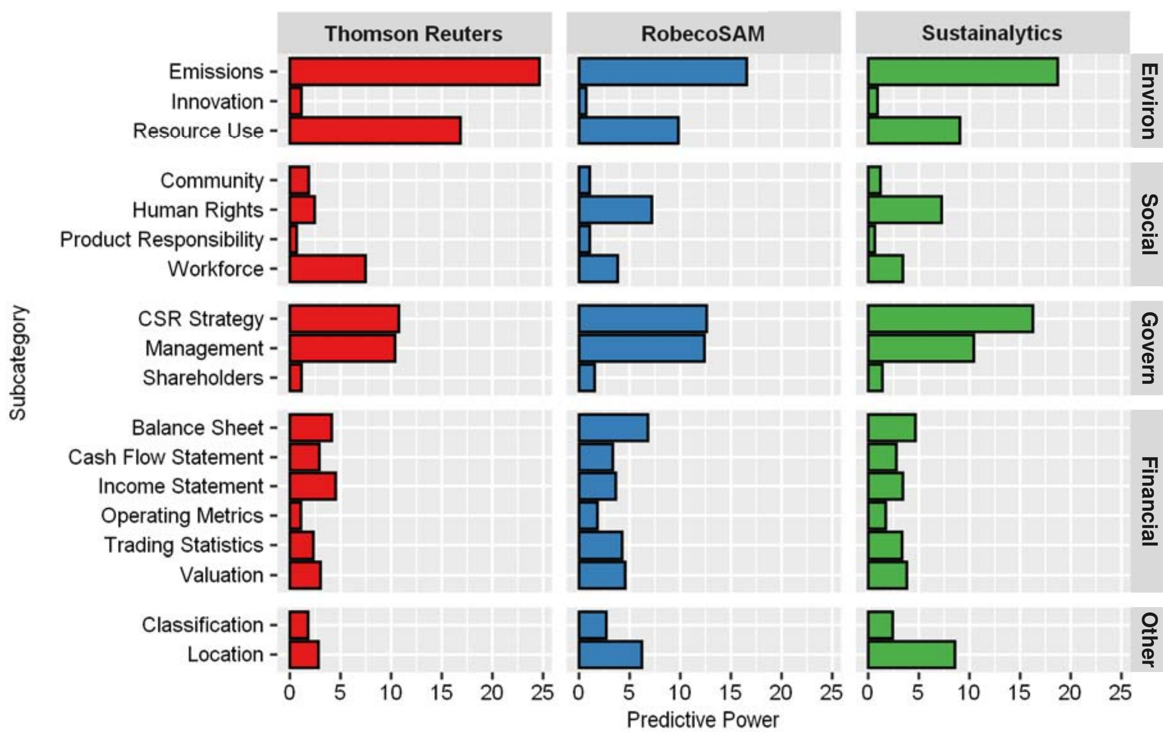
[25] *Human Rights* include variables that measure a company's effectiveness towards respecting the fundamental human rights conventions. *Workforce* refers to variables that reflect a company's effectiveness towards job satisfaction, a healthy and safe workplace, maintaining diversity and equal opportunities and development opportunities for its workforce. *Product Responsibility* includes variables that reflect a company's capacity to produce quality goods and services integrating the customer's health and safety, integrity and data privacy.

Figure 7. Predictive Power by Subcategory

a) Measure 1: Mean Decrease in Impurity



b) Measure 2: Perturbation Impurity



Source: Authors' calculations using Bloomberg and Refinitiv Eikon data.

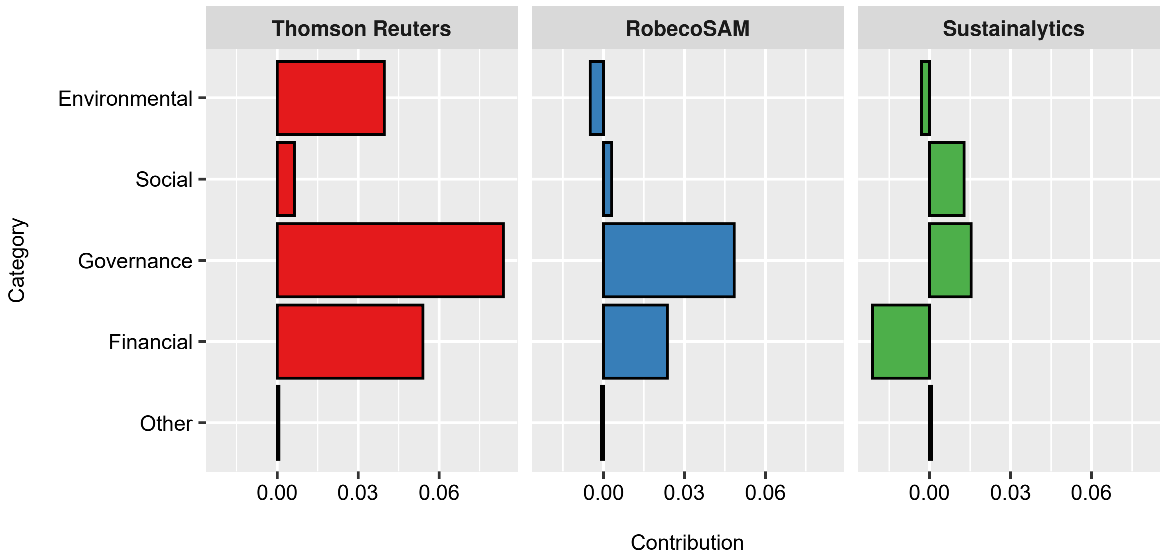
5.1.2. VARIABLE CONTRIBUTIONS

The other way to identify what information the ESG ratings are capturing is to evaluate how much each variable contributes to the predicted ESG rating. To do so, we use the predictive power of the variables to generate new ESG ratings. We then estimate the actual contribution of each group of variables to these predicted ESG ratings. Figure 8 reports the results for the categories and Figure 9 for the subcategories.[26] Ultimately, this allows us to identify how much the different factors matter when calculating the various ratings, based on the information derived from the Machine Leaning analysis:

- **Governance and financial variables are the top two contributors for all three agencies.** Governance is the category whose importance is robust across the two analyses: prediction power of a category as well as contribution to the predicted score. Yet, its magnitude varies significantly across rating providers.
- ***Management* and *CSR Strategy* drive the contribution of governance**, in line with the previous analysis. Yet, *CSR Strategy* contributes negatively to the predicted Sustainalytics score.
- ***Balance Sheet* and *Cash Flow Statement* drive the contribution of financial variables.** And both are negatively related to the predicted Sustainalytics score.
- **Environment variables are still important for the predicted Thomson Reuters score**, **especially *Emissions* and *Resources Use*.**
- ***Workforce* remains an important sub-category for social variables**, in line with the previous analysis.

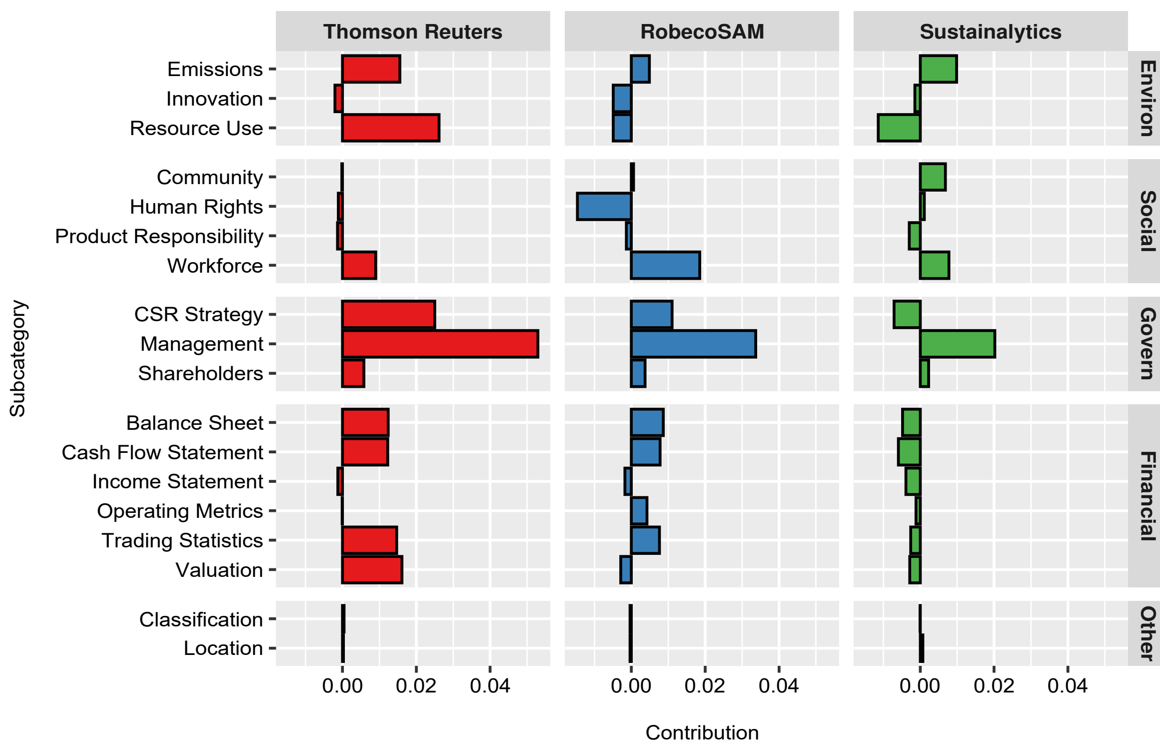
[26] See Appendix 7 for more details on how variable contributions are calculated.

Figure 8. Contribution to Predicted ESG Scores by Category



Source: Authors' calculations using Bloomberg and Refinitiv Eikon data.

Figure 9. Contribution to Predicted ESG Scores by Subcategory



Source: Authors' calculations using Bloomberg and Refinitiv Eikon data.

5.2. IS IT ABOUT THE METHODS?

Beyond the variables, the methods used to aggregate the information differ from one rating to another. We illustrate this point by looking at how the variables interact. Finally, we show how challenging it is for investors to rationalize and understand the discrepancies across agencies by comparing the rating we have generated with the one provided by the agencies.

5.2.1. VARIABLE INTERACTIONS

Looking at the interaction among variables or group of variables provides insights on how the different agencies' aggregate information impacts the ratings.

We use the estimated random forests to determine whether—and to what extent—the different explanatory variables interact with each other when predicting the ESG scores. The overall interaction (see Figures 10 and 11) is quite different across the ratings, especially at the subcategory level. For example, while the overall interaction effects of environmental variables are concentrated on the subcategory *Resource Use* for the predicted Thomson Reuters and RobecoSAM ratings, they appear to be (roughly) evenly divided between *Emission* and *Resource Use* for Sustainalytics. Similarly, although the overall interaction effects associated with governance variables seem to be concentrated on the subcategory *Management* for the predicted Sustainalytics rating, they are more evenly distributed between *Management* and *Shareholders* for RobecoSAM and (to a lesser extent) for Thomson Reuters.

Figures 12 and 13 focus on the pairwise interaction by category and subcategories.[27] These pairwise effects measure the extent to which variables belonging to one group interact with variables in another group. As expected, the results show significant differences across rating agencies. For the predicted Thomson Reuters, for example, most pairwise interaction effects are relatively weak and evenly distributed across categories and subcategories.

[27] Following Friedman and Popescu (2008), we estimate variable interaction effects by decomposing the prediction function into main and interaction effects and measuring how much of the variance in the model's predictions depends on the latter.

By contrast, pairwise interaction effects appear to be relatively larger and more concentrated for the other two predicted ratings. In the case of RobecoSAM, the most substantial pairwise interaction effects are between financial and governance variables (especially between Valuation and Management), within financial variables (driven by the interaction between *Balance Sheet* and *Operating Metrics)*, between environmental and social variables (mostly driven by the interaction between *Emissions* and *Product Responsibility)*, and between finance variables and other (*Valuation* and *Location).*

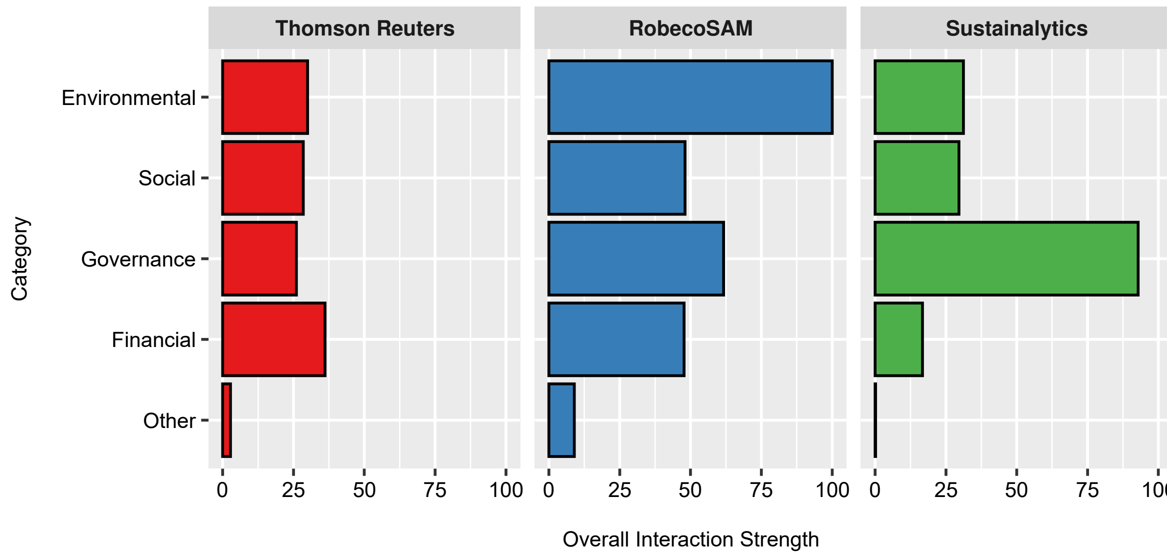
Similarly, for the predicted Sustainalytics rating, there are significant interaction effects between governance and environment (*Management* and *Resource Use),* between governance and finance (*Human Rights* and *Balance Sheet)*, and within governance (variables in the *Management* subcategory).

Perhaps surprisingly, the interaction between *Classification* (which includes a company's economic sector) and all the environmental, social, and governance subcategories appears to be very weak. This result is at odds with the use of sector-specific methodologies, a claim made by all three rating agencies in our sample.[28]

Using standardized data, our analysis shows how the information processing matters for the ratings. Yet harmonization of the methods is not necessarily the solution. Not being able to reconcile the ratings due to their different data treatment is not an issue as long as the difference reflects the rating agencies' priorities, emphasizing the ESG issues they deem the most important. If that is the case, these choices need to be shared with the rating users, investors, or firms, which in turn will decide which rating is more in line with their priorities.

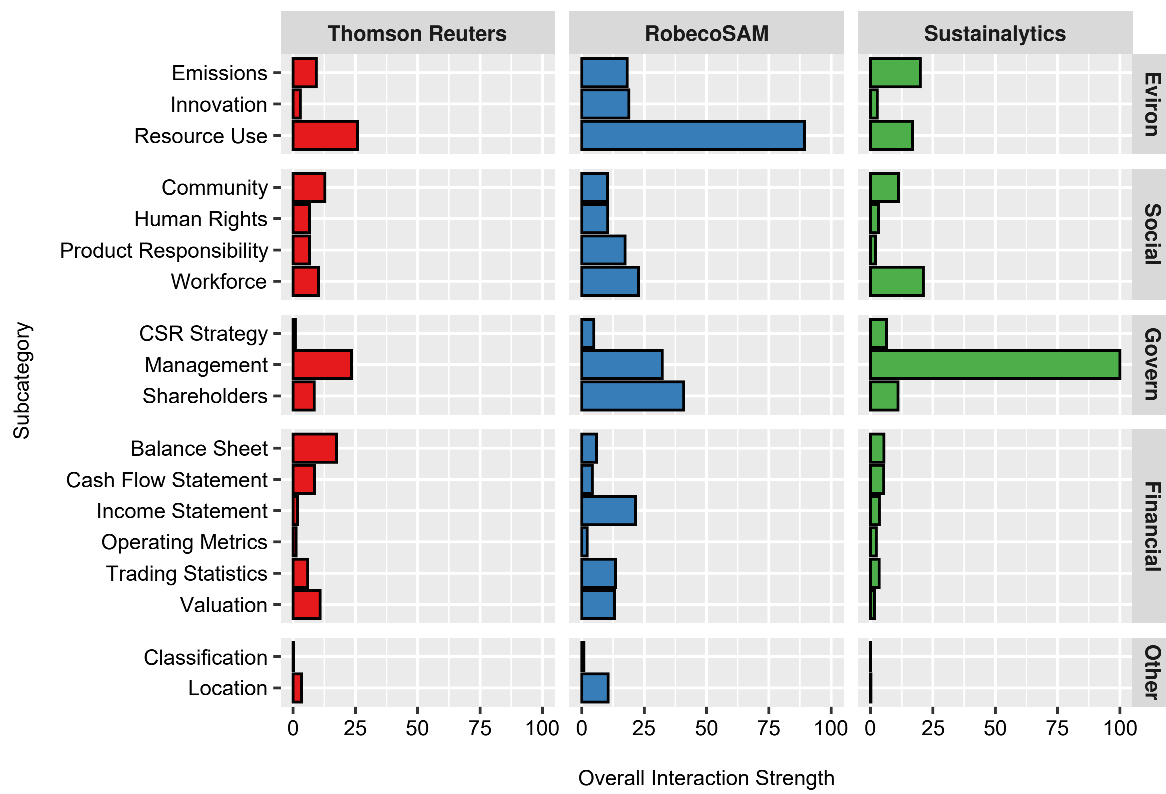
[28] See Gaffuri (2017, p.11), Sutainalytics (2019, p. 5-6), and Thomson Reuters (2018, p. 6).

Figure 10. Overall Interaction Strength by Category



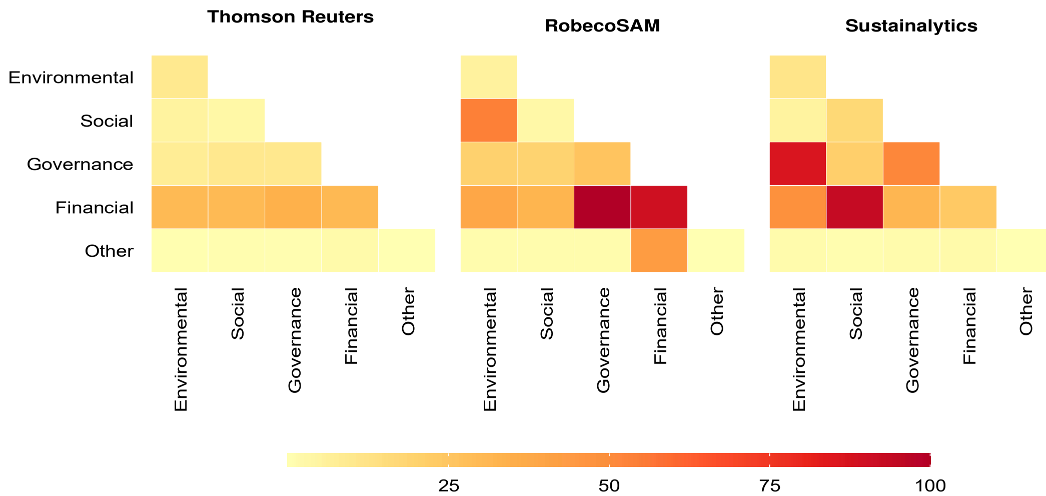
Source: Authors' calculations using Bloomberg and Refinitiv Eikon data.

Figure 11. Overall Interaction Strength by Subcategory



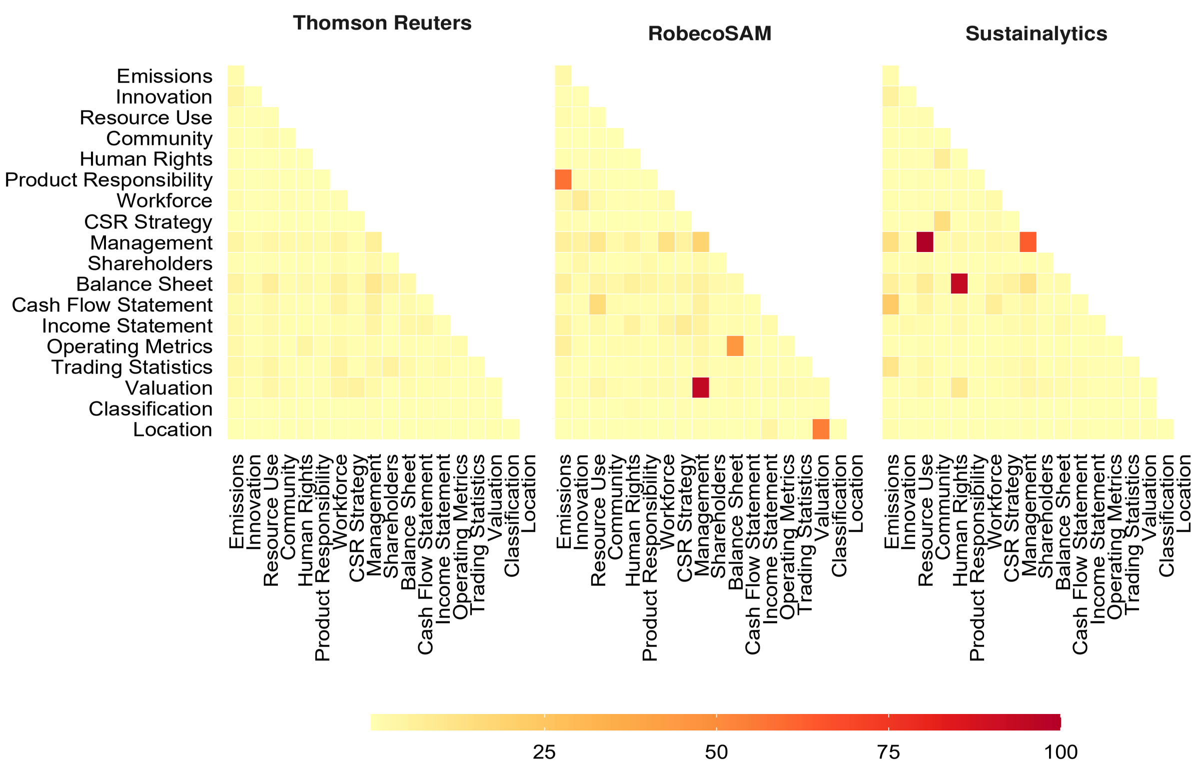
Source: Authors' calculations using Bloomberg and Refinitiv Eikon data.

Figure 12. Pairwise Interaction Strength by Category



Source: Authors' calculations using Bloomberg and Refinitiv Eikon data.

Figure 13. Pairwise Interaction Strength by Category



Source: Authors' calculations using Bloomberg and Refinitiv Eikon data.

5.3. RATINGS: OBSERVED VERSUS GENERATED

To conclude our analysis, we check the ability of the generated ratings to replicate the level of disagreement observed between the actual ESG rating of the agencies. Comparing predicted and observed levels of disagreement offers valuable information to investors: it captures the difficulty in predicting and understanding the discrepancies across ESG scores based on information readily available to market participants.

Table 5. Correlations between ESG Ratings: Observed and Predicted

	Observed ESG Scores	Predicted ESG Scores
RobecoSAM vs Sustainalytics	0.72	0.87
RobecoSAM vs Thomson Reuters	0.65	0.82
Sustainalytics vs Thomson Reuters	0.65	0.79

Source: Authors' calculations using Bloomberg and Refinitiv Eikon data.
Note: The correlations are the Pearson product-moment correlation coefficients.

Table 5 reports correlation coefficients for each possible pair of ESG scores as predicted by the estimated random forests and as observed in the data. For all three pairs, the correlations between predicted scores are greater than those observed in the actual ESG rating of the agencies. Using similar data while allowing for different methods to process it strengthens the convergence across the ratings, confirming that the use of standardized data will lead to more comparable ratings.[29]

.

[29] Novick (2020)

6. CONCLUSIONS

Inconsistencies and the lack of comparability across rating providers often confuse investors. While differences across ESG scores can naturally emerge if rating providers adopt different definitions of ESG performance, our analysis shows that differences arise even when the rating agencies declare using a similar definition. Thus, the focus when it comes to ESG ratings should not be on agreeing on a single definition, but on standardization of the data, achieving greater clarity in the labeling of the ratings and more transparency regarding their objectives.

Our analysis illustrates how difficult it is to understand or predict the ratings. It shows that most of the discrepancies among rating providers cannot be easily explained by information readily available to investors or other users of these ratings. Yet, two clear outcomes emerge:

- The three ratings strongly agree on who are the worst performers, with a correlation higher than 0.95.
- The three ratings reach some consensus when measuring risks arising from governance factors, especially for Corporate Social Responsibility Strategy and Management. The first subcategory of governance includes variables that reflect a company's practices to communicate that it integrates the economic (financial), social and environmental dimensions into its day-to-day decision-making processes. The second one includes variables that measure a company's commitment and effectiveness towards following best practice and corporate governance principles.

Overall, our study has two main implications when it comes to assessing how well-equipped firms are to deal with ESG risks. First, there is a need for standardization of the data. The use of standardized data will help to reconcile the ratings, at least partially. This emphasizes the lack of common disclosure standards, and the importance of aligning the different ESG disclosure standards existing across the world. The resulting harmonization of the data to be collected would allow rationalizing the reporting burden on the firms while increasing the quality of the data collected. One option could be to have a centralized data collection, a portal, where each firm would report its information. The different rating agencies, regulators, and other ESG data providers or users would then pull information from this data hub. Ultimately this will increase the firms' participation while improving the rating agencies' credibility in the eyes of investors.

Creating harmonized and high-quality data is only part of the solution. The second implication of our study is the importance of transparency when it comes to the methodologies used to calculate the rating or the focus of the rating. Are E, S, and G factors equally important? Or is the rating focusing mostly on one dimension? Our study highlights the importance of the different methodologies used by the rating agencies to aggregate the data and their impact on the ratings. Having different emphasis across different rating agencies can provide useful information if the difference reflects a clear prioritization from the agencies, emphasizing the ESG issues they deem the most important. If that is the case, the agencies need to be transparent about it with the rating users, investors, or firms, which in turn will decide which rating is most in line with their priorities.

7. REFERENCES

Berg, F., J. Koelbel, and R. Rigobon. 2020. "Aggregate Confusion: The Divergence of ESG Ratings." MIT Sloan Working Paper 5822-19.

BNP Paribas. 2019. *The ESG Global Survey 2019*.

Breiman, L. 2001. "Random Forests." *Machine learning*.

Clarkin, C. M., M. Sawyer, and J. L. Levin. 2020. "The Rise of Standardized ESG Disclosure Frameworks in the United States." Harvard Law School Forum on Corporate Governance. Monday, June 22.

Friede, G., T. Busch, and A. Bassen. 2015. "ESG and Financial Performance: Aggregated Evidence from More than 2000 Empirical Studies." *Journal of Sustainable Finance & Investment*, *5*(4), 210-233.

Friedman, J. H., and B. E. Popescu. 2008. "Predictive Learning via Rule Ensembles." *The Annals of Applied Statistics*.

Fritsch, D. 2019. *ESG: Do you or Don't you? Are the World's Asset Owners Thinking about Long-Term, Sustainable Finance?* Responsible Investor Research, UBS.

Gaffuri, Jvan. 2017. *RobecoSAM CSA*. Lisbon.

Gibson, R., P. Krueger, N. Riand, and P. S. Schmidt. 2019. "ESG Rating Disagreement and Stock Returns." ECGI Working Paper Series in Finance, N° 651/2020.

Global Sustainable Investment Alliance. 2019. *2018 Global Sustainable Investment Review*.

Grim, M. D., and D. B. Berkowitz. 2018. *ESG, SRI, and Impact Investing: A Primer for Decision-Making*. Vanguard Research.

Grimminger, A., and P. D. Benedetta. 2013. *Raising the Bar on Corporate Governance: A Study of Eight Stock Exchange Indices*. World Bank Group.

Koehler, D., and E. Hespenheide. 2013. "Finding the Value in Environmental, Social and Governance Performance." *Deloitte Review* Vol. 12 (January).

Lauricella, T. 2020. "4 Lessons from BlackRock's Bold Sustainable Investing Statement." Morningstar. https://www.morningstar.com.au/funds/article/4-lessons-from-blackrocks-bold-sustainable-in/198904

Li, X., Y. Wang, S. Basu, K. Kumbier, and B. Yu. 2019. "A Debiased MDI Feature Importance Measure for Random Forests." *Advances in Neural Information Processing Systems*.

Novick, B. 2020. "Towards a Common Language for Sustainable Investing." Harvard Law School Forum on Corporate Governance. Wednesday, January 22

State Street Global Advisors. 2018. *Performing for the Future. ESG's Place in Investment Portfolios. Today and Tomorrow.* 2017 ESG Survey Report.

State Street Global Advisors. 2019. *The ESG Data Challenge*. March.

Sustainalytics. 2019. *The ESG Risk Rating: Methodology Abstract*. Version 2.0.

Thomson Reuters. 2018. *Thomson Reuters ESG Scores*.

US SIF Foundation. 2018. *Report on US Sustainable, Responsible and Impact Investing Trends*.

Verheyden, T., R. G. Eccles, and A. Feiner. 2016. "ESG for All? The Impact of ESG Screening on Return, Risk, and Diversification." *Journal of Applied Corporate Finance*, 28(2), 47-55.

Wong, C., and E. Petroy. 2020. *Rate the Raters 2020: Investor Survey and Interview Results*. Sustainability.

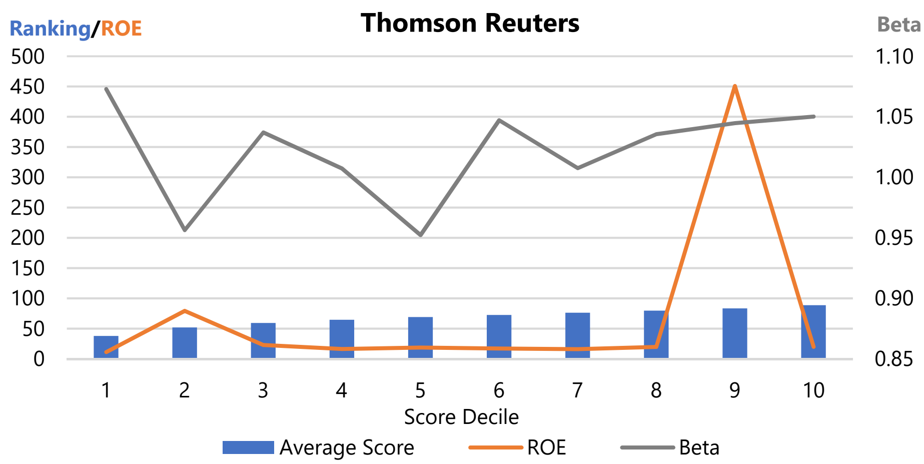
APPENDIX

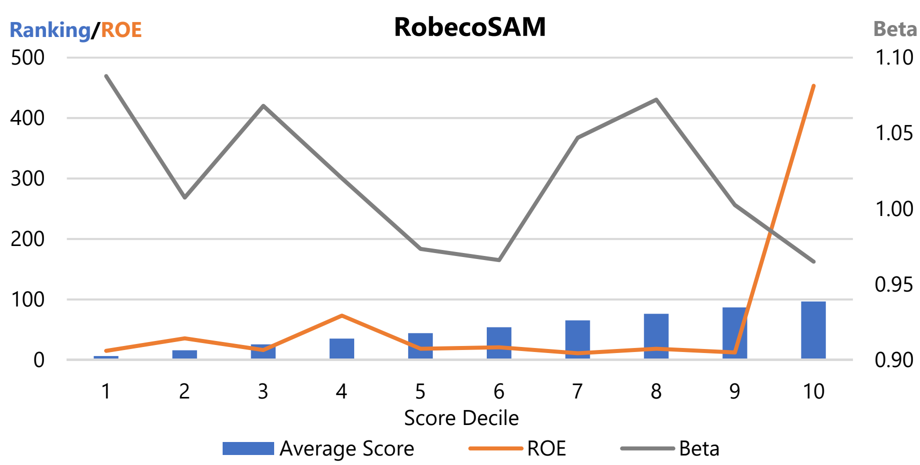
APPENDIX 1. NUMBER OF ESG-FOCUSED FUNDS IN LARGEST ASSET MANAGEMENT FIRMS

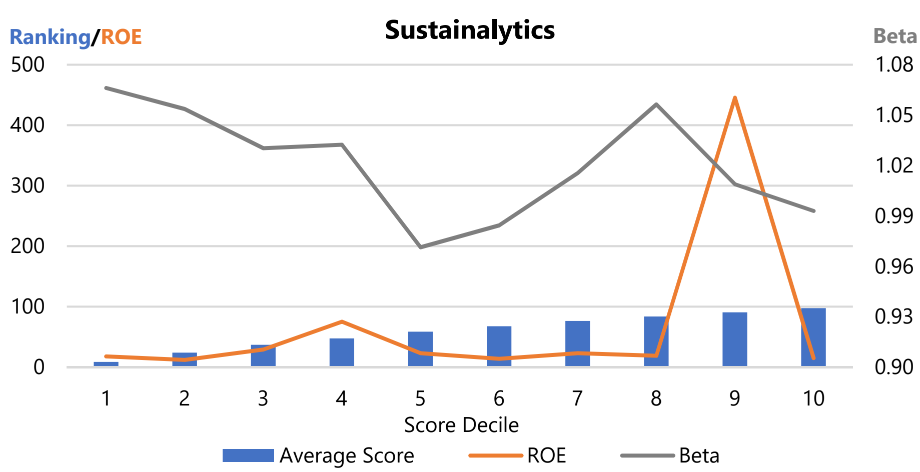
Company	# of Funds (All Share Classes)	# of ESG Funds	Percentage ESG
BlackRock	1038	18	1.73%
Vanguard	207	6	2.90%
UBS	26	4	15.38%
Fidelity	318	5	1.57%
State Street	140	2	1.43%
Allianz	51	3	5.88%
Capital Group	62	0	0.00%
JP Morgan Asset Management	197	2	1.02%
Goldman Sachs	104	2	1.92%
Bank of New York Mellon	205	8	3.90%
PIMCO	146	14	9.59%
Amundi	136	5	3.68%
Prudential Financial	322	0	0.00%
AXA Group	10	0	0.00%
Morgan Stanley	262	7	2.67%

Source: Morningstar Direct (7/5/2020).
Note: Funds classified as ESG explicitly stated in their mandates that investments were chosen primarily for their ESG-risk mitigating characteristics. Keywords in the primary investment mandate also included impact investing, gender/ethnic diversification, and environmental sustainability.

APPENDIX 2. ROE AND BETA BY ESG SCORE DECILE



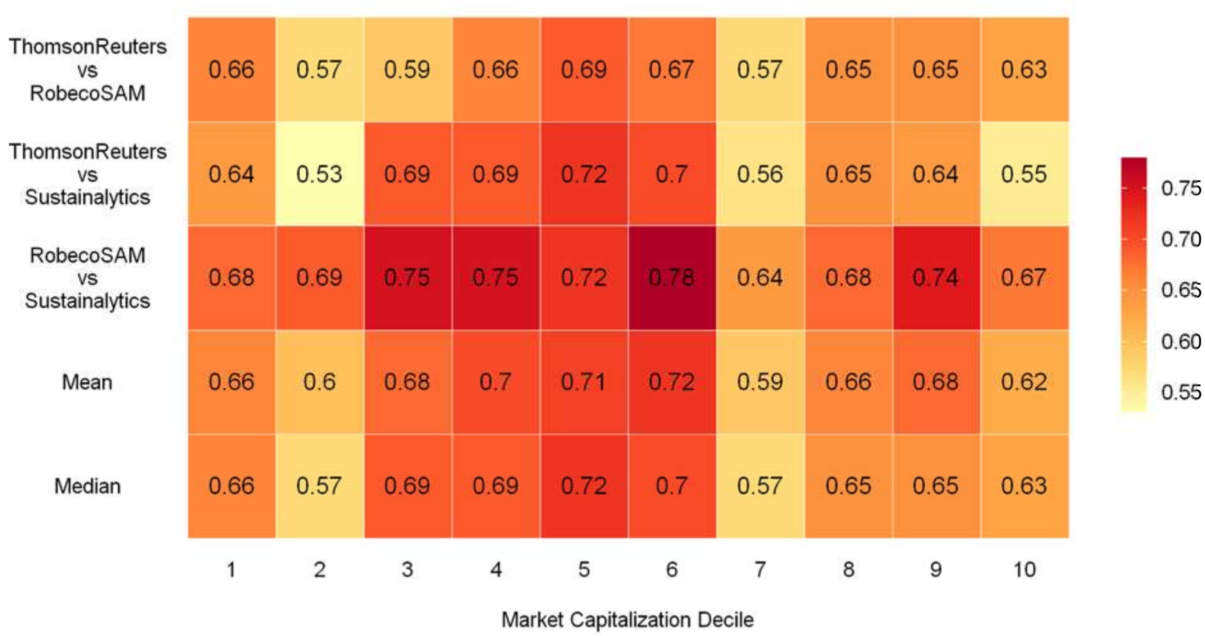




APPENDIX 3. DISAGREEMENT AMONG RATING AGENCIES AND MARKET CAPITALIZATION

The analysis in the main text indicates that the extent of disagreement among ESG rating agencies varies substantially across firms. To better understand what is driving this heterogeneity, this appendix shows correlations for each pair of ESG scores after dividing companies into deciles based on their market capitalization. Figures A.1 below shows the results of the exercise. First, consistent with our previous findings (both when we pool all firms and when we divide them by economic sector), RobecoSAM and Sustainalytics exhibit the highest pairwise correlation across all market capitalization deciles. Second, all pairwise correlations follow a relatively similar pattern as we move from companies with low market capitalization to companies with high market capitalization. Third, the relationship between the (average) level of agreement among rating agencies and the level of market capitalization is not monotonic. The level of agreement among rating agencies appears to be slightly higher for companies with intermediate levels of market capitalizations (i.e., deciles 4, 5, and 6) than for companies with low or large levels (especially those in deciles 2, 7, and 10). Taken together, the results suggest that it does not exist a clear relationship between the level of market capitalization and the degree of agreement among rating agencies in our sample.

Figure A.1. Correlations between ESG Scores by Market Capitalization Decile

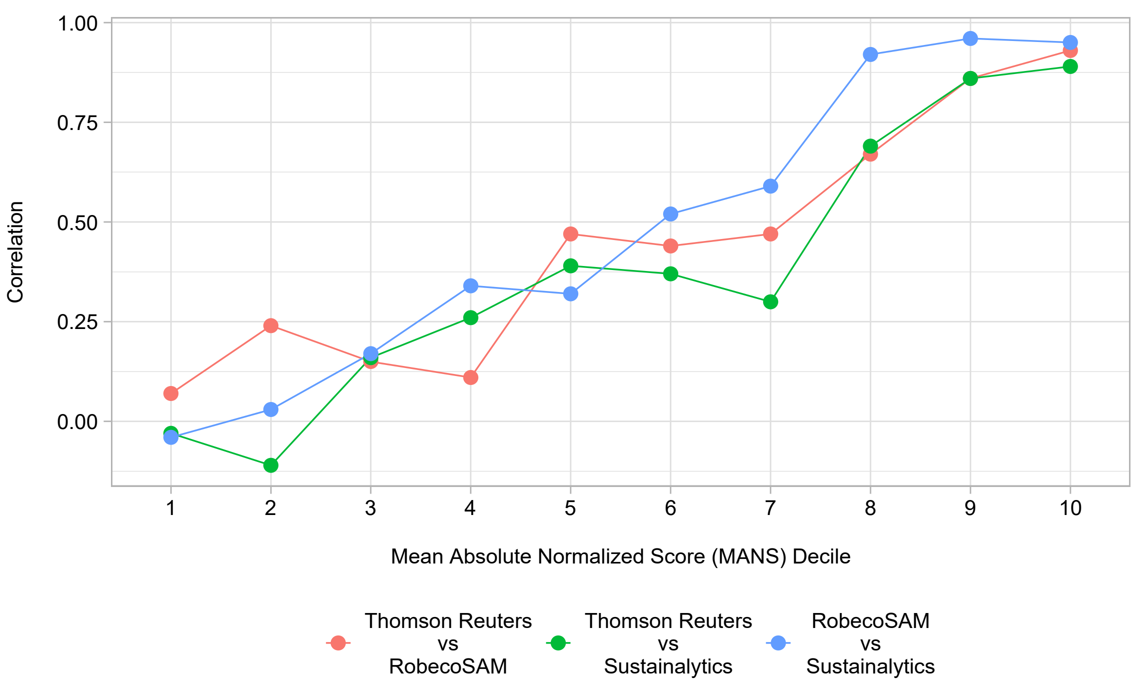


Source: Authors' calculations using Bloomberg and Refinitiv Eikon data.

APPENDIX 4. DISAGREEMENT AMONG RATING AGENCIES AT THE FIRM-LEVEL

Figure A.2 explores how disagreement varies across individual firms. It shows correlations between ESG scores after grouping companies based on an individual measure of "disagreement among rating agencies." [30] Surprisingly, the results reveal that the extent of the inconsistencies among rating providers varies substantially across firms. Indeed, if disagreement among agencies were roughly constant across all companies, the curve in Figure A.2 would be relatively flat. Instead, the average correlation between ESG

Figure A.2. Correlations between ESG Scores by Deciles based on Firm-Specific Measure of Disagreement

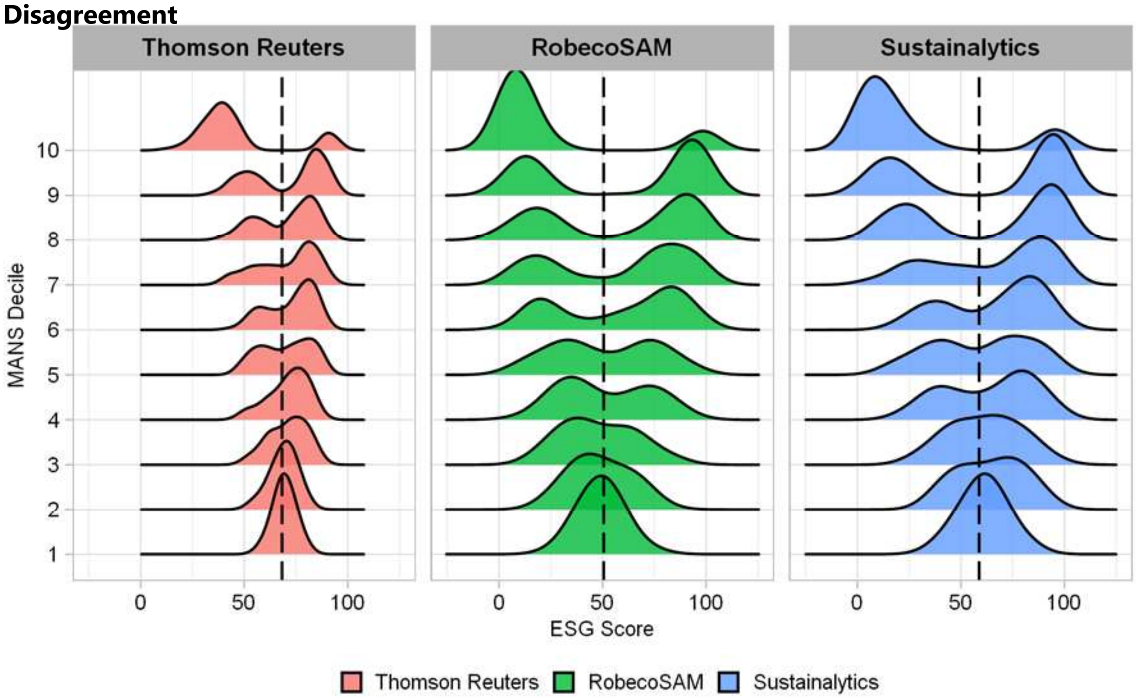


[30] To calculate our firm-level measure of disagreement, we first normalize all ESG scores by subtracting off their respective means and dividing them by their respective standard deviations. For each company in our sample, we then calculate the mean of the absolute value of the normalized scores across all three rating agencies. The resulting number is our firm-level measure of disagreement. For a similar exercise, see Berg et al. (2019).

scores increases from a value of about 0 (for companies in the first decile) to a value slightly above 0.9 (for companies in the top decile).

As discussed in section 4, economic sectors explain part of the variation in disagreement across firms. Figure A.3, which plots the distributions of ESG scores after grouping companies based on our firm-specific measure of disagreement, offers two additional insights. First, as the firm-level measure of inconsistencies increases, ESG scores move away from their respective means (i.e., the vertical dotted lines). Thus, the level of agreement among rating agencies appears to be higher for companies whose scores are away from the mean (i.e., "relatively good" and "relatively bad" firms) than it is for companies whose scores are close to the average. Second, for all three rating agencies, most companies in the top decile of our firm-specific measure of disagreement have extremely low ESG scores, indicating that the strongest agreement among rating providers occurs across the worst performers.

Figure A.3. ESG Score Distributions by Deciles based on our Firm-Specific Measure of Disagreement



Source: Authors' calculations using Bloomberg and Refinitiv Eikon data.
Note: The vertical dotted line represents the overall average score for each of the rating agencies.

APPENDIX 5. TOP INDIVIDUAL PREDICTORS FOR ESG SCORES

Variable	Category	Definition
Environment Management Training	Environmental	Does the company train its employees on environmental issues?
Environmental Supply Chain Management	Environmental	Does the company use environmental criteria (ISO 14000, energy consumption, etc.) in the selection process of its suppliers or sourcing partners?
Policy Emissions	Environmental	Does the company have a policy to improve emission reduction?
Policy Energy Efficiency	Environmental	Does the company have a policy to improve its energy efficiency?
Policy Environmental Supply Chain	Environmental	Does the company have a policy to include its supply chain in the company's efforts to lessen its overall environmental impact?
Renewable Energy Use	Environmental	Does the company make use of renewable Energy?
Resource Reduction Policy	Environmental	Does the company have a policy for reducing the use of natural resources or to lessen the environmental impact of its supply chain?
Resource Reduction Targets	Environmental	Does the company set specific objectives to be achieved on resource efficiency?
Targets Emissions	Environmental	Has the company set targets or objectives to be achieved on emission reduction?
Flexible Working Hours	Social	Does the company claim to provide flexible working hours or working hours that promote a work-life balance?
Fundamental Human Rights ILO UN	Social	Does the company claim to comply with the fundamental human rights convention of the ILO or support the UN declaration of human rights?
Human Rights Contractor	Social	Does the company report or show to use human rights criteria in the selection or monitoring process of its suppliers or sourcing partners?
Policy Human Rights	Social	Does the company have a policy to ensure the respect of human rights in general?
Board Gender Diversity	Governance	Percentage of female on the board.

CSR Reporting	Governance	Does the company publish a separate CSR/H&S/Sustainability report or publish a section in its annual report on CSR/H&S/Sustainability issues?
Global Compact Signatory	Governance	Has the company signed the UN Global Compact? The UN GC is a non-binding united nations pact to encourage businesses worldwide to adopt sustainable and socially responsible policies, and to report on their implementation.
Independent Board Members	Governance	Percentage of independent board members as reported by the company.
Stakeholder Engagement	Governance	Does the company explain how it engages with its stakeholders? How is it involving the stakeholders in its decision-making process?

APPENDIX 6. VARIABLE SUBCATEGORIES

Category	Subcategory	Subcategory Definition
Environmental	Emissions	Variables that measure a company's commitment and effectiveness towards reducing environmental emissions in its production and operational processes.
Environmental	Innovation	Variables that reflect a company's capacity to reduce the environmental costs and burdens for its customers, thereby creating new market opportunities through new environmental technologies and processes or eco-designed products.
Environmental	Resource Use	Variables that reflect a company's performance and capacity to reduce the use of materials, Energy or water, and to find more eco-efficient solutions by improving supply chain management.
Social	Community	Variables that reflect a company's commitment to being a good citizen, protecting public health and respecting business ethics.
Social	Human Rights	Variables that reflect a company's effectiveness in terms of respecting fundamental human rights conventions.
Social	Product Responsibility	Variables that reflect a company's capacity to produce quality goods and services, integrating the customer's health and safety, integrity, and data privacy.
Social	Workforce	Variables that measure company's effectiveness in terms of providing job satisfaction, a healthy and safe workplace, maintaining diversity and equal opportunities and development opportunities for its workforce.
Governance	CSR Strategy	Variables that reflect a company's practices to communicate that it integrates economic (financial), social and environmental dimensions into its day-to-day decision-making processes.
Governance	Management	Variables that measure a company's commitment and effectiveness towards following best practice corporate governance principles.
Governance	Shareholders	Variables that measure a company's effectiveness towards equal treatment of shareholders and the use of anti-takeover devices.
Financial	Balance Sheet	Variables that reflect a company's assets, liabilities, and shareholders' equity.
Financial	Cash Flow Statement	Variables that summarizes the amount of cash and cash equivalents entering and leaving a company.
Financial	Income Statement	Variables that measure a company's revenues and expenses during a particular period. It also includes variables indicating how the revenues are transformed into the net income or net profit.

Financial	Operating Metrics	Variables that illustrate a company's overall performance such as Return on Equity, Return on Assets, and EBITDA.
Financial	Trading Statistics	Variables that reflect the trading of a company's stock. It includes variables such as monthly Sharpe Ratio, volatility, institutional Ownership, 200-day Price PCT Change, and liquidity measures.
Financial	Valuation Metrics	Variables that reflect and are related to a company's valuation such as market capitalization, enterprise value, P/E Ratio, P/EG Ratio, Beta, and dividend yield.
Others	Classification	Economic sector according to the Thomson Reuters Business Classification.
Others	Location	Country of headquarters, also known as Country of Domicile.

APPENDIX 7. CALCULATING VARIABLE CONTRIBUTIONS

To understand how variable contributions are calculated in a random forest model, notice that given a set of independent variables or predictors, we can estimate how the value of the prediction changes after every split in each decision tree. Since each split is associated with a variable, and since the split either adds or subtracts to the predicted value given in the previous node, the final prediction can be decomposed as the sum of the variable contributions plus the "bias" (i.e., the model's prediction at the beginning of the decision tree). After averaging over all the individual decision trees in the random forest model, the final prediction can be decomposed as follows:

$$\text{prediction}(x) = \text{bias} + \text{contribution}(1, x) + \ldots + \text{contribution}(n, x)$$

where

- x is a set of predictors,
- *bias* is the model's prediction before using any predictor (usually the mean of the variable we want to predict in the original dataset),
- *contribution(j, x)* is the contribution of variable j to the final prediction, and
- *n* is the number of predictors.

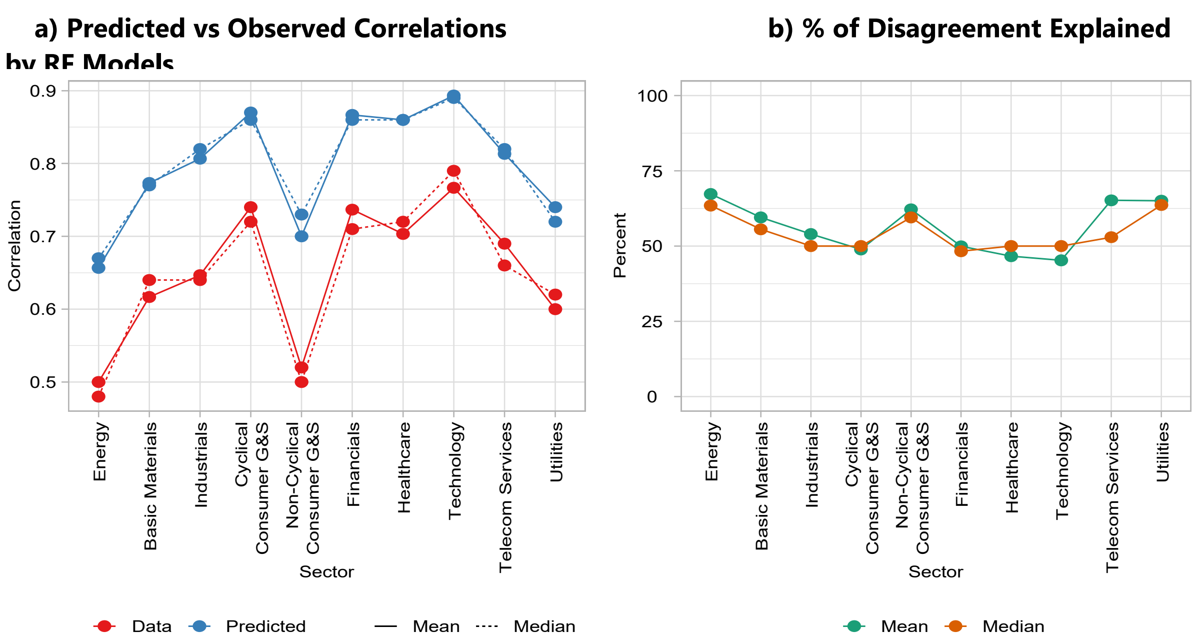
Although the previous expression is superficially similar to a linear regression, the coefficients of a linear regression are fixed, with a single constant for every variable. For the random forest model, by contrast, the contribution of each variable is a complex function. One that also depends on all the other variables, which together determine the decision path that generates the prediction and thus the contributions that are passed along the way. [31]

APPENDIX 8. OBSERVED VERSUS GENERATED RATINGS AND FIRMS' CHARACTERISTICS

[31] For a detailed discussion of the methodology, see https://blog.datadive.net/interpreting-random-forests/

In this appendix, we explore whether the ability of our model to account for the disagreement among ESG rating agencies varies with some of the firms' characteristics. To this end, we divide companies by economic sector and market capitalization decile, and then compare the mean and median correlations between the ESG scores observed in the actual data with those predicted by the random forest models. Figures A.4(a) and A.5(a) shows the results of the exercise. The results suggest that the random forests do a reasonably good job at capturing *variations* in the level of disagreement among rating agencies across sectors and market capitalization deciles, but that they tend to underpredict the level itself. Thus, the figures seem to indicate that the importance of factors not captured by the random forests in explaining the disagreement among rating agencies remains significant across all economic sectors and market capitalization deciles. This last point is confirmed by Figures A.5(b) and 4.5(b), which display the fraction of disagreement explained by the random forest models for each economics sector and market capitalization decile. The figures show that the ability of the random forests to account for the disagreement among agencies ranges from 45.2% to 67.3% across economic sectors and from 46.6% to 59.5% across market capitalization. Although the specific numbers may vary, the overall picture seems to confirm that the models can account

Figure A.4. Explanatory Power of Random Forest Models by Economic Sector



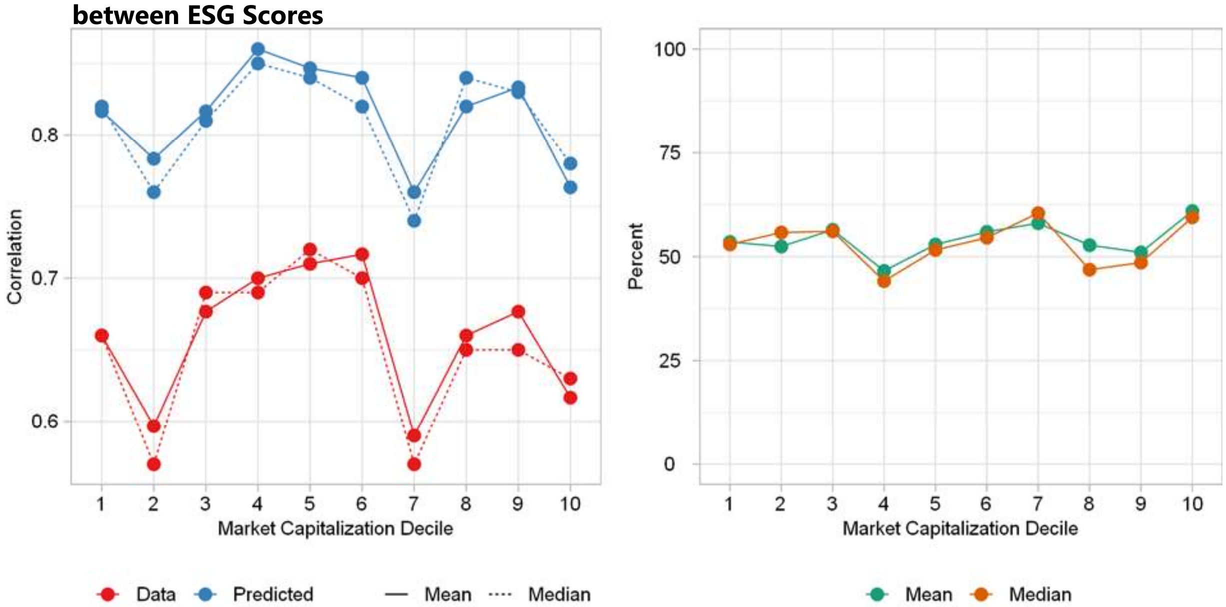
Source: Authors' calculations using Bloomberg and Refinitiv Eikon data.

for around half of the observed disagreement among rating agencies.

Figure A.5. Explanatory Power of Random Forest Models by Market Capitalization Decile

a) Predicted vs Observed Correlations **b) % of Disagreement Explained by RF Models**

between ESG Scores



Source: Authors' calculations using Bloomberg and Refinitiv Eikon data.

AUTHORS

Dr. Claude Lopez is the Head of the Research Department at Milken Institute, where she leads data-driven efforts aimed at influencing global policy issues on International Finance, Health Economics, and Regional Economics. She is an active member of the T20 task force on international financial architecture for stability and development, and a contributor to the W20 (Women 20), two advisory committees to the G20.

Lopez has over 20 years of experience in academic and policy research in the US and abroad. Before joining the Institute, Lopez headed multiple research teams at the Banque de France, the nation's central bank, and was professor of economics at the University of Cincinnati. She has an MS in econometrics from Toulouse School of Economics and a PhD in economics from the University of Houston

Dr. Oscar Contreras is an economist specializing in International Finance within the Research Department at the Milken Institute. He has experience in macroeconomics and financial markets, particularly in emerging economies.

Before joining the Institute, Contreras was a Senior Economist at the Research Department of the Central Bank of Mexico, where he was responsible for providing macroeconomic analysis on monetary policy issues and conducting original research on macroeconomic and financial topics. Prior to the Bank, he was an Assistant Professor at the University College for Financial Studies (CUNEF) in Madrid, Spain. Contreras holds a PhD in Economics from Northwestern University and a BA (Hons) in Economics from the Center for Economics Research and Teaching (CIDE) in Mexico City.

Joseph Bendix is a Research Analyst in International Finance within the Research Department at the Milken Institute. His work focuses on topics related to systemic risk, capital flows, and investment. More specifically, he is in charge of identifying and analyzing the market-level data sets in many of the research reports produced by the Institute. Bendix lends his experience to developing presentations for the Institute conferences held throughout the year. He holds a bachelor’s in economics and a Master’s of Science in Finance from the University of San Diego

ACKNOWLEDGEMENTS

The authors would like to thank Keith Savard, Jihad Dagher, Bill Lee, Perry Wong, Jakob Wilhelmus, and Jose Siaba Serrate for numerous insightful discussions related to this topic. They also would like to thank participants at the MI Research Department brown bag for useful discussions. Claude Lopez would like to thank participants at the T20 Summits Japan and Saudi Arabia.