ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS
MSc Economics & Business
Specialization Financial Economics

ESG Rating Disagreement and its Impact on Corporate Carbon Emissions

Author: C.W. Spek
Student number: 544038
Thesis supervisor: Dr. D.L. Urban
Second reader: [tba]
Finish date: October 2023

Preface and acknowledgements

Dear reader,

Thank for your interest in my Master Thesis, which hopefully ends my time as a student at the Erasmus School of Economics. I would like to express my gratitude towards my thesis supervisor, Dr. Daniel Urban, for his input in choosing the topic of my thesis, and for his helpful supervision throughout the thesis process. In addition, I would like to thank the co-reader for his/her effort in assessing this thesis.

I would like to thank my parents for putting their trust in during my four years of studying at the Erasmus School of Economics, and for being proud of me all this time. I would also like to thank my family, friends, colleagues, roommates and all the other acquaintances over the last four years for making my time at the Erasmus School of Economics even more enjoyable.

Enjoy the read!

Kind regards,

Christian Spek

Abstract

Environmental, social and governance (ESG) ratings play an important role in assessing the corporate social responsibility of a company. This paper investigates whether disagreement amongst different rating providers about the ESG rating of a company has an effect on the subsequent environmental performance of this firm. Additionally, it investigates which role institutional ownership plays in influencing this performance. Using a worldwide sample of 9,200 firm-year observation from four different ESG rating providers, this study reveals a significant level of disagreement in ESG ratings across these providers. Additionally, this rating disagreement does not have a straightforward effect on subsequent environmental performance, suggesting that environmental performance in firms is driven by many factors beyond just ESG ratings.

Keywords: ESG Ratings, rating disagreement, institutional ownership, carbon emissions, Big 3

JEL Classification: G24, G32, M14, Q56

Table of contents

List of tables

List of figures

List of appendix tables

1. Introduction

According to Milton Friedman, there is only one social responsibility of a firm: to increase its profits. For a very long time, environmental, social and governance (ESG) issues were considered to be irrelevant, as long as a company maximized its profits. However, over the last two decades, socially responsible investing (SRI) and corporate social responsibility (CSR) have experienced a noticeable increase in attention by managers, investors, academics, and other stakeholders. Nowadays, an investment must not only be profitable, but also sustainable. A well-known example of this sustainable investment practice is that of the largest Dutch pension fund ABP, that announced in February 2023 that it would divest completely out of its position in fossil fuel companies such as Shell, due to its continuing negative impact on the environment.[1]

Amidst this growing environmental and social awareness, managers, investors, and academics have turned to Environmental, Social and Governance (ESG) ratings as an important measure of corporate social responsibility. These ratings have proven to be a useful tool for assessing a firm's performance in these areas. However, recently multiple academic studies have questioned the validity of these ESG ratings, showing that there is often a high level of disagreement between different rating providers (e.g. Chatterji et al., 2016 and Berg et al., 2022). Additionally, the *Wall Street Journal* noted that 'investors need to dig deep' to understand the ratings from different data providers, and why they differ.[2] From this, it is clear that ESG rating disagreement is a highly relevant topic, both for managers, investors, and other stakeholders.

In prior research, a number of papers has paid attention to the question of *why* different ESG rating providers disagree with one another. However, little research has been done into the *consequences* of ESG rating disagreement. The studies that have delved into the consequences, have mainly focused on consequences for investors on the financial markets (e.g. Gibson Brandon et al., 2021 and Christensen et al., 2022). As for the consequences of ESG rating disagreement on firm-level outcomes, such as the level of carbon emissions, studies have only

[1] https://www.abp.nl/over-abp/actueel/nieuws/2023/januari/Merendeel--fossiele-beleggingen--verkocht
[2] Berg, F. (2022, November 2). Why Do ESG Ratings Vary So Widely—and How Can Investors Make Sense of Them? *WSJ*. https://www.wsj.com/articles/esg-ratings-investing-data-raters-11667229384

made implications. For example, Berg et al. (2022) suggest that due to ESG rating disagreement, managers of firms are uncertain whether investments into carbon emission reduction would be appreciated by their share- and stakeholders and are therefore disincentivized into doing so. However, until now, no definitive answer has been given to this question. Therefore, this study is centered around the following question:

What is the effect of ESG rating disagreement on the environmental performance of firms, and what role does institutional ownership play?

I contribute to existing literature on ESG rating disagreement in three ways: first, I extend the evidence on ESG rating disagreement by investigating the correlations between the four ESG rating providers available to students at the Erasmus University Rotterdam. In addition, to my knowledge, there is currently no paper that investigates the effect of ESG rating disagreement on actual ESG performance in the form of carbon emission reduction. Hence, I believe that my research contributes to the studies on ESG rating disagreement by evaluating how disagreement between ESG raters actually influences the ESG performance of firms. Prior studies have focused on the effect of institutional ownership on firms' environmental performance (e.g. Dyck et al., 2019 and Azar et al., 2019), but no prior study has incorporated the role of ESG rating disagreement in their research. Therefore, I contribute to existing literature by evaluating whether institutional ownership incentivizes firms to reduce their carbon emissions, and specifically how institutional ownership potentially alleviates the confusion caused by ESG rating disagreement.

When I compare the ESG ratings of the different data providers to one another, I find relatively low correlations between the ESG ratings of the different providers, indicating a substantial level of ESG rating disagreement consistent with prior studies on the topic. Additionally, I find that the level of rating disagreement varies considerably per rating pillar, year, and industry. Next, I find that there is no straightforward effect of ESG rating disagreement on the environmental performance of a firm in the form of a subsequent change in carbon emissions. I find a negative standalone effect of ESG rating disagreement on the subsequent change in carbon emissions, but this effect is not robust to various fixed effects specifications. When I account for the level of institutional ownership, however, I find that ESG rating disagreement has a positive association with subsequent carbon emissions, indicating that firms with higher levels of ESG rating disagreement subsequently reduce their carbon emissions to a lesser extent,

or even increase them. This effect is robust to including various fixed effects specifications but is not robust to a more restrictive sample.

When I investigate the effect of institutional ownership, I find that institutional ownership has a positive association with subsequent carbon emission change when including year and firm fixed effect, indicating that higher levels of institutional ownership are associated with a subsequent lower reduction of carbon emissions, or even an increase in carbon emissions. When I distinguish between Big 3 ownership and non-Big 3 ownership, I find that Big 3 ownership has a negative effect, while non-Big 3 ownership has a positive effect on the change in carbon emissions. These effects, however, are not all robust to different fixed effects specifications. Lastly, I find that the change in carbon emissions of firms with a higher level of institutional ownership is less affected by ESG rating disagreement. However, again, this effect is not entirely robust to different fixed effects specifications. These results indicate that a firm's choice to reduce carbon emissions is difficult to generalize and is driven by many different factors outside of ESG rating disagreement and institutional ownership.

The remainder of this paper is structured as follows: in Chapter 2, I briefly introduce ESG ratings, I lay out a literature review on ESG rating disagreement, its consequences, and the effect of institutional ownership, and I develop my hypotheses. In Chapter 3, I describe the data collection and sample construction, and I define the variables I use in my main analyses. Next, I analyze the level of ESG rating disagreement in my sample in Chapter 4. In Chapter 5, I analyze the effect of ESG rating disagreement on a firm's environmental performance, and I evaluate the effect of institutional ownership in this relation. I discuss my findings and limitations and make recommendations for future research in Chapter 6.

2. Literature review and hypothesis development

2.1. The use of ESG ratings

Over the previous decades, the interest in socially responsible investing (SRI) has experienced a noticeable increase. As of March 2022, a total of 4,395 investors representing over $120 trillion in assets under management have become signatories to the United Nations Principles for Responsible Investment (UN PRI) (PRI, 2022). Due to this increase in interest for responsible investment, the demand for information on ESG has experience a similar explosive increase in recent years (Larcker et al., 2022). Over the past two decades, a number of ESG data providers has emerged, most of which provide aggregate ratings with regards to the overall ESG performance of a company. These ESG ratings are described as 'evaluations of a company based on a comparative assessment of their quality, standard or performance on environmental, social or governance issues' (SustainAbility, 2018).

ESG ratings are commonly used by various parties (Larcker et al., 2022): asset owners use ratings to evaluate the environmental and societal impact of the firms they invest in, as well as to screen companies for investment. Similarly, institutional investors use information about firms' ESG performance to create investment products catering to the needs of their clients, as well as to examine the risk that ESG-related issues might pose to their portfolios. Furthermore, firms and their managers use ESG ratings to evaluate their own environmental and societal performance and provide their own ESG information through (voluntary) disclosures such as sustainability reports. Lastly, ESG ratings are also commonly used in research in the fields of economics, management, and finance.

2.2. ESG rating validity and disagreement

As the use of ESG ratings in investment practices, managerial decision-making and academic research has become a common practice, it is important that the ratings being used are valid. However, due to the complexity in measuring firms' non-financial or ESG performance, and the subjectivity that goes paired with it, the validity and convergence of different ESG ratings have been heavily debated in recent literature. Chatterji et al. (2009) investigate the validity of the environmental ratings from Kinder, Lydenberg, Domini Research & Analytics (KLD, now MSCI). Their results indicate that the KLD environmental ratings, especially the environmental

'concerns', were a reasonable measure for past environmental performance. In addition, the environmental ratings were able to predict *some* future environmental outcomes, albeit with a far lower explanatory power compared to the measure for past performance. Contrary to this, the environmental 'strengths' were not able to predict future environmental outcomes. These results imply that there was rather low validity in the ratings from KLD, and that KLD was not aggregating historical data in an optimal way.

In addition to analyzing the validity of a *single* ESG rating provider, recent research has focused its attention on comparing the ESG ratings of *different* rating providers and investigating the level of convergence between these ratings. From this research, it has become clear that there is an evident lack of convergence between different ESG ratings, and that different rating providers often disagree with one another. Chatterji et al. (2016) investigate the convergent validity of different ESG ratings. Noticeably, they find that the ESG ratings of six well-established providers have fairly low correlations with each other, indicating that there is low convergent validity, i.e. disagreement between different ESG raters. Chatterji et al. (2016) also delve into what drives this disagreement. They name two possible sources: a lack of common *theorization* – what rating providers believe to be socially responsible behavior – or a lack of *commensurability* – the extent to which different rating providers measure a common construct in a similar fashion. In additional analysis, Chatterji et al. (2016) find that even when adjusting for differences in theorization between different rating providers, the correlations between different ratings remain low. These results imply that different rating providers do not only have a different definition of socially responsible behavior, but they also measure similar constructs in a different way.

In recent years, multiple studies have found evidence consistent with the results of Chatterji et al. (2016) and have performed additional analyses in order to extend their results. Dorftleiner et al. (2015) compare the descriptive statistics and distributions of ASSET4 (now Refinitiv ESG), Bloomberg and KLD (now MSCI) and evaluate the correlations between both the overall ratings, and the E, S and G sub-scores. They provide evidence that the different overall ratings and sub-scores follow different distributions, and that overall ratings have rather low correlations with each other. Particular sub-scores do show somewhat higher levels of correlation within and amongst different rating providers, however the authors conclude that the different ratings are still incomparable to one another. In further research, Billio et al. (2021) compare the ratings from Sustainalytics, RobecoSAM, Refinitiv and MSCI, and find similar

disagreement among the different rating providers when considering the correlation between ratings. In addition, they harmonize and convert the different ratings into a common scale. Using this scale, the authors find that the observed disagreement is on average larger than one class, and that the percentage of agreement among the ESG ratings after applying this common scale is on average 24%. According to the authors, these measures imply strong disagreement among the different rating providers. Additionally, Gibson Brandon et al. (2021) find a low average pairwise correlation among seven different ratings providers. In addition, their results show that the average correlation is lowest for the governance dimension of ESG ratings, and highest for the environmental dimension. Furthermore, they find that correlations between ratings from some providers are remarkably high, going against common beliefs about ESG rating disagreement.

Next to providing evidence in order to confirm the extensive disagreement between different ESG rating providers, recent studies have delved into the forces driving the divergence of different ESG ratings. One of the main contributions to this area of research has come from the innovative study of Berg et al. (2022). To advance from the results of Chatterji et al. (2016), the authors provide a quantitative decomposition of ESG rating divergence based on every underlying indicator of six ESG rating providers. Berg al. (2022) provide three distinct sources of rating divergence: first of all, *scope divergence*, where different ratings are based on different sets of attributes. Next to this, *measurement divergence* refers to the situation where different rating providers measure the same attributes using different indicators. Lastly, *weight divergence* is a result of different rating providers have a different opinion on the relative importance of certain attributes. In relation to Chatterji et al. (2016), scope and weight divergence seem to fall under a lack of common theorization, while measurement divergence relates to low commensurability. When looking at the relative contributions of these sources of divergence, Berg et al. (2022) find that measurement divergence is the main contributor to rating disagreement, with more than 50% of rating divergence stemming from this source. In addition, scope divergence also has some importance, while weight divergence has only a small contribution. Furthermore, the authors discover a so-called *rater effect*, indicating that a firm that has received a high score in one category is more likely to receive high scores in other categories from a certain rating provider.

Where Berg et al. (2022) decompose rating disagreement based on the underlying data from ESG rating providers, other research add to existing literature by relating the level of rating

disagreement to firm-level characteristics. In their aforementioned study, Gibson Brandon et al. (2021) examine whether the level of rating disagreement varies in the relation to observable firm-level financial and accounting characteristics. Doing so, they find that more profitable firms, experience lower rating disagreement, while firms without a credit rating experience higher rating disagreement. Furthermore, firms with more tangible assets are subject to lower disagreement in environmental ratings, while firms with higher levels of institutional ownership experience higher disagreement in these ratings. Lastly, the authors find that firms with higher market-to-book ratios exhibit higher levels of disagreement in their social ratings. In another study, Christensen et al. (2022) investigate the role of firm-level ESG disclosure as a determinant of ESG rating disagreement. They find a strong positive relation between ESG disclosure and disagreement, suggesting that greater ESG disclosure by firms leads to greater rating disagreement. The authors provide additional evidence supportive of this notion by analyzing changes in rating disagreement after the implementation of mandatory disclosure requirements: after this implementation, affected firms experience greater rating disagreement. As a possible explanation for their findings, the authors state that a higher level of disclosure gives rise to more subjectivity, and therefore more disagreement, in determining ESG ratings.

2.3. Consequences of ESG rating disagreement

From the aforementioned research, it has become clear that ESG rating providers tend to strongly disagree with one another, and that this disagreement is related to various firm-level characteristics. However, for this research, I am not necessarily interested in *why* this rating divergence exists, but rather what are the *consequences* of this divergence.

A relatively small number of studies has focused on the real-life implications of ESG rating disagreement. The studies that have done so, have mainly looked at the consequences of rating disagreement for investors and the financial market. As an extension to their aforementioned research, Gibson Brandon et al. (2021) find that firms with greater rating disagreement experience significantly higher stock returns, with an interquartile range increase in rating disagreement being associated with an increase of 92 basis points in annual stock returns. Furthermore, they find that this relation is mainly driven by disagreement about the *environmental pillar* of ESG ratings. The authors conclude that these results are consistent with the view that disagreement about the ESG performance of a firm is perceived as an additional source of risk or uncertainty, for which risk-adverse investors require a premium. Additionally,

Christensen et al. (2022) examine stock market consequences of rating disagreement by performing short-window event studies centered around the date on which a rating provider publishes a new ESG rating. They find several results: first of all, similar to Gibson Brandon et al. (2021), the authors find that rating disagreement is positively related to market-adjusted returns. In addition, the authors provide evidence for a positive relation between rating disagreement and return volatility. Furthermore, the authors find that rating disagreement affects firms' financing choices: firms experiencing higher levels of disagreement rely more on internal financing and are less likely to raise external financing. Serafeim and Yoon (2022) take on a different perspective and investigate the predictive ability of ESG ratings concerning future ESG news and its impact on stock markets and pay specific attention to the role of rating disagreement. They find that ESG ratings demonstrate a strong predictive ability for future ESG news; however, this ability significantly weakens when there is substantial rating disagreement among rating providers. In line with the rating divergence decomposition posed by Berg et al. (2022), the authors find that the predictive power of ESG ratings diminishes for firms with significant measurement divergence. Furthermore, the results provide evidence that rating disagreement is associated with a lack of significant stock market reaction to ESG news. This suggests that in the case of a discrepancy in ESG ratings, the stock market's response to ESG-related news might not be as strong, potentially mitigating its impact on firms' stock prices.

This stream of literature shows that ESG rating disagreement has real-life consequences for both investors, managers, and other stakeholders. As suggested by Gibson Brandon et al. (2021), a key reason for these consequences is *uncertainty*. Because different rating providers have different opinions on the ESG performance of the same firm, less-informed investors are uncertain about the implications of this disagreement.

2.4. The effect of ESG rating disagreement on ESG performance

Apart from ESG rating disagreement having consequences for financial markets and its investors, disagreement about the rating of a firm might influence the subsequent environmental performance of the firm itself. As to my knowledge, little to no research has been performed into analyzing the effect of ESG rating disagreement on the following environmental performance of a firm. Due to this, it would be wise to first discuss whether, and if so how, ESG

ratings affect the subsequent environmental performance of a firm, before looking into the role that rating disagreement plays in this picture.

2.4.1. Firm response to ESG ratings

Firms may have several motives for responding to their ESG ratings. Firstly, as socially responsible investing has experienced a large gain in popularity in recent year, firms with relatively poor ESG ratings may face the risk of being unattractive candidates for investments. In addition, having a poor ESG rating might be detrimental to the overall reputation of a firm (Hamilton, 1995). Moreover, changes in ESG ratings have been found to have a relation with stocks returns: Shanaev and Ghimire (2022) find that while upgrades in ESG ratings are associated with modest positive stock returns, rating downgrades on the other hand lead to significant negative risk-adjusted returns.

Given the stakes at hand, it is surprising that only a few studies have directly researched how firms respond to their ESG ratings. Chatterji and Toffel (2010) find that firms with low initial ESG ratings showed more significant improvements in subsequent environmental performance, particularly in reducing carbon emissions, compared to firms that either did not receive an ESG rating, or compared to firms that received a more favorable rating. Adding to this, they find that this result is mainly driven by firms in industries under significant environmental scrutiny, and by firms that were subject to less costly opportunities in order to improve their environmental performance. Slager and Chapple (2016) examine firms facing exclusion from the FTSE4Good Index after the introduction of new criteria and show statistical evidence that these firms were more likely to improve their environmental performance in the following year. Following a distinct approach from the aforementioned literature, Clementino and Perkins (2020) perform a qualitative analysis into how firms react to ESG ratings. Their results paint a more differentiated picture: while the majority of the firms in their sample showed a conformist reaction to ESG ratings, their actual responses took on various forms. Some firms improved their level of ESG disclosure, were other firms increased their awareness on ESG-related issues or implemented new policies. Conversely, other firms resisted their ratings, either actively or passively. Furthermore, the majority of firms stated that ESG ratings did not meaningfully affect their environmental performance; only the firms showing active conformity to ESG ratings claimed to have improved their ESG performance. These results, albeit from a very limitative sample, show that different firms respond differently to ESG ratings.

2.4.2. The role of ESG rating disagreement

As I mentioned earlier, no study has directly researched the effect of ESG rating disagreement on the ESG performance of a firm. However, multiple studies mention that rating disagreement might have consequences on the environmental performance of firms. Berg et al. (2022) state that rating disagreement disincentivizes firms to improve their ESG performance, because these firms receive 'mixed signals' from rating provides about which actions there are supposed to undertake in order to improve their ESG performance, and about which actions would be appreciated by investors on the stock market. These mixed signals might lead firms to underinvest in improving their ESG performance. Next to this, Chatterji et al. (2016) pose that due to rating disagreement and the resulting invalidity, mangers 'lack clear guidance' on their environmental performance. Clementino and Perkins (2020) additionally mention that a number of their respondents consider that their investors look at multiple ratings before making an investment, and that this influences their own need for improving their ESG performance. In the case of different ratings, it is logical to think that these firms will feel less need for improving their ESG performance. Next to the implications of these studies, the research on the impact of rating disagreement on investors and the financial market shows that uncertainty plays a role in explaining these consequences. Next to investors, managers might experience this uncertainty about whether there is a need to improve the ESG performance of their firm. Based on these arguments, I hypothesize the following:

Hypothesis 1*: Firms with larger ESG rating disagreement improve their subsequent ESG performance to a lesser extent.*

2.5. The impact of institutional investment on ESG performance

Institutional investors have long been a driving force in shaping the financial landscape. However, in recent years, institutional investors have shifted their attention more and more towards ESG goals. According to PricewaterhouseCoopers's Asset and Wealth Management Revolution 2022 report, ESG-related assets under management are expected to grow to $33.9 trillion by the year 2026 (PwC, 2022). Next to possible social reasons, institutional investors have financial motives be active on the ESG front. Krueger et al. (2020) find that institutional investors believe that climate risks have financial implications for their portfolio firms, and that

these risks have already begun to materialize. Furthermore, these institutional investors believe that climate risks can be addressed through engagement with their portfolio firms. In addition, Hoepner et al. (2022) show that engagement on ESG issues can benefit shareholders by reducing downside risk, that this engagement is most effective when addressing environmental topics, and that firms with large downside risk reductions show a decrease in environmental incidents after engagement. This goes to show that institutional investors could play a vital role in improving the ESG performance of their portfolio firms.

A number of studies has directly investigated the effect of institutional ownership on firms' ESG performance. Dyck et al. (2019) assess whether institutional investors as a whole impact firms' ESG performance around the world. Using a sample of non-US firms, they find a positive association between the level of institutional ownership and firm-level ESG ratings. Furthermore, the authors find that this positive influence is even greater when investors are signatories to the UN PRI, and when they invest in firms with below-median initial ESG ratings. When looking into the mechanisms that investors use to positively influence firms' ESG performance, Dyck et al. (2019) find that investors focus their attention on firms they already own, and do so through predominantly private engagements. Lastly, the authors show that institutional investors are driven by financial as well as social motivations in order to improve the ESG performance of the firms they own. More specifically, the results show that the impact of institutional investors is only evident in countries with sufficiently strong social norms towards ESG issues.

Within the universe of institutional investors, Azar et al. (2021) particularly investigate the role of the "Big Three" (i.e. BlackRock, Vanguard, and State Street) with regards to the reduction of carbon emissions. When they examine the likelihood of the Big Three engaging with firms, they find that firms with higher carbon emissions have a higher probability of being the target of Big Three engagements. In addition, the authors find a significant negative relation between the level of Big Three ownership and carbon emissions, and they observe that this relation is more pronounced if firms have a higher probability of Big Three engagements. Contrary to Dyck et al. (2019) finding a negative association between institutional ownership and ESG ratings, the authors find that institutional ownership in general is not associated with a decrease in firms' carbon emissions. Furthermore, Azar et al. (2021) provide further evidence for the negative association between Big Three ownership and carbon emissions using a plausibly

exogenous change in Big Three ownership through the reconstitution of the Russell 1000 and Russell 2000 indexes.

The findings of Dyck et al. (2019) and Azar et al. (2021) show that institutional investors, and particularly the Big Three, have a positive impact on the ESG performance of their portfolio firms. The results show that these influential investors have shown an active approach in ESG-related issues, engaging with their portfolio firms in private discussions in order to drive positive changes in ESG performance. Furthermore, these investors are substantial organizations, with employees dedicated towards ESG purposes.[3] Moreover, as highlighted by Matos (2020), institutional investors need to perform their own detailed research into the environmental practices of their firms, and do not rely solely on external ESG ratings. These arguments imply that firms with a higher degree of institutional ownership exhibit lower sensitivity to ESG rating disagreement in improving their ESG performance.

In light of these arguments, I hypothesize the following:

Hypothesis 2*: The ESG performance of firms is positively affected by the level of institutional ownership.*

Hypothesis 3*: The ESG performance of firms is less affected by ESG rating disagreement in the presence of higher levels of institutional ownership.*

[3] For example, see information on investment stewardship at BlackRock: https://www.blackrock.com/corporate/insights/investment-stewardship

3. Data, sample construction and measurement

In order to test my hypotheses on how firm-level ESG rating disagreement affects subsequent environmental performance, I construct a sample with available ESG ratings over the longest possible period. In order to maximize the number of available firm-year observations, I use a sample of worldwide firms over the period from 2010 through 2019.

3.1. Data

I collect data from the four ESG data providers that are available from the Erasmus University data library: (1) Refinitiv ESG (previously Asset4),[4] (2) Sustainalytics,[5] (3) Morgan Stanley Capital International's (MSCI) Intangible Value Assessment (IVA),[6] and (4) MSCI KLD.[7] According to the 2023 'Rate the Raters' report, these data providers are widely used by both investors and corporations (SustainAbility, 2023). In addition, both Sustainalytics and MSCI are ranked as one of the highest-quality and most useful ESG ratings providers. Conversely, Refinitiv ESG has a remarkably lower ranking in both these categories.

Table 1 displays some important characteristics of these four data providers. Column 1 shows the different pillar ratings that the ratings providers supply. All four ratings providers supply an aggregated total ESG rating, as well as ratings for both the environmental, social and (corporate) governance pillars.

Column 2 shows the ratings scales used by each provider. Refinitiv ESG and Sustainalytics both use a continuous scale from 0 to 100 for their ratings, with 100 being the maximum score a firm can achieve. Similarly, MSCI IVA uses a scale from 0 to 10. Originally, MSCI KLD does

[4] Asset4 was acquired by Thomson Reuters in 2009. After the acquisition, the name of the database changed to Thomson Reuters ESG Scores, but the data was still widely known under the name Asset4. As of 2018, the ESG ratings data of Thomson Reuters are part of Refinitiv, and the data is now also known as Refinitiv ESG.

[5] After acquiring a 40% stake in Sustainalytics in 2017, Morningstar acquired the remaining 60% of Sustainalytics equity in 2020.

[6] MSCI IVA was initially created by Innovest Stategic Value Advisors, was acquired by RiskMetrics in 2009 and was ultimately taken over by MSCI when it acquired RiskMetrics in 2010.

[7] The data from MSCI KLD find its origin in Kinder, Lydenberg, Domini (KLD) & Co., Inc., which was acquired by RiskMetrics in 2009, and by MSCI in 2010. See Eccles, Lee and Stroehle (2019) for additional details on the history on KLD, and its relation to MSCI IVA.

not provide such a continuous scale for their ratings themselves. Instead, MSCI KLD employs a rating system based on dichotomous *strength* and *concern* variables within each of their seven rating categories.[8] In order to convert these variables into a continuous scale, I follow the methodology of numerous academic studies (e.g. Lins et al., 2017 and Gibson Brandon et al., 2021) and sum up both the number of strengths and concerns separately and scale these measures by the maximum number of strengths and concerns available each year. Subsequently, I create a net ESG rating by subtracting the scaled number of concerns from the scaled number of strengths. This method results in a continuous scale from –1 to 1. It is worth noting that in addition to the aforementioned seven rating categories, MSCI KLD also provides *Controversial Business Involvement* indicators for categories such as alcohol, firearms, tobacco, and nuclear power. As these indicators are not part of the 'traditional' ESG pillars, I decide not to include these indicators in the sample.

Because there are three different scales being used amongst the four different rating providers, the different *raw* ESG ratings are incomparable to each other. As a solution to this, some studies (e.g. Christensen et al., 2022) use simple arithmetic in order to re-scale the different ratings to a comparable scale. However, because the ratings from different providers are not only scaled differently, but also follow a different distribution along this scale, I believe this simple method is not sufficient. Instead, in order to reach comparability amongst the different rating providers, I apply the methodology of Gibson Brandon et al. (2021): for each rating provider and at each point in time, I convert the different firm ratings into ranked percentile scores and standardize them to have a mean of 0 and a standard deviation of 1.

Column 3 of Table 1 shows that the data of Refinitiv ESG, Sustainalytics and MSCI KLD available throughout the entire sample period. The data from MSCI IVA is available from 2013 onwards. In some cases, multiple ratings are issued to a firm within a given year. In this case, I keep only the last rating issued by the respective rating provider within this year. The four data providers cover a large sample of firms. Column 4 of Table 1 shows that both Refinitiv ESG, Sustainalytics and MSCI KLD cover around the same number of firms, ranging from around 8,500 to around 9,750. MSCI IVA, however, covers a substantially larger number of firms, with around 47,000 firms in their data sample.

[8] The different categories are Environment, Community, Human Rights, Employee Relations, Diversity, Product and Corporate Governance. Environment and Corporate Governance form their own pillar, while the remaining categories fall under the Social pillar.

Column 5 highlights the various ratings styles used by the different rating providers, which might partially explain disagreement between these providers. For example, Sustainalytics and MSCI IVA both give out *best in class* ESG ratings, while MSCI KLD provides an absolute ESG rating. Additionally, Refinitiv ESG is more focused on disclosure in their ESG ratings.

Table 1

Characteristics of rating providers

Rating provider	Pillars	Rating scale	Period covered	# of firms covered	Rating style
Refinitiv ESG	Total, E, S, G	0 – 100	2010 – 2019	9,727	Disclosure oriented
Sustainalytics	Total, E, S, G	0 – 100	2010 – 2019	9,498	Best in class
MSCI IVA	Total, E, S, G	0 – 10	2013 - 2019	47,108	Best in class
MSCI KLD	Total, E, S, G	-1 – +1	2010 – 2019	8,677	Absolute ESG ratings

In order to measure the environmental performance of firms, I gather carbon emissions data from Refinitiv ESG. Refinitiv ESG covers a worldwide cross-section of firms which totals to more than 80% of global market capitalization. Refinitiv gathers carbon emission data from information reported by firms. If a firm does not directly report its actual carbon emission data, Refinitiv estimates a firm's annual carbon emission using several models in a step-by-step procedure.[9] I obtain data on institutional ownership from the Thomson/Refinitiv database through Wharton Research Data Services (WRDS). Thomson/Refinitiv gathers the data on institutional holdings primarily from mandatory filings of the 13F form from institutional investors to the Securities and Exchange Commission. I obtain accounting and financial market data from Standard and Poor's Compustat – Capital IQ. This database provides balance sheet, income statement and stock price information for a global set of firms. As the data from Compustat is denoted in the local currency of the firm, I also obtain daily exchange rate data from Institutional Brokers' Estimate System (I/B/E/S).

3.2. Sample construction

Table 2 outlines the sample construction procedure. As shown in Table 2, I depart from the four ESG rating datasets. As I attempt to study ESG rating *disagreement*, I require that a firm-year observation is covered by at least two rating providers in order to be included in my sample. In

[9] Refinitiv ESG explains its estimation models in more detail in its online fact sheet: https://www.refinitiv.com/content/dam/marketing/en_us/documents/fact-sheets/esg-carbon-data-estimate-models-fact-sheet.pdf

addition, I require that a firm has non-missing carbon emission, institutional ownership, accounting, and financial data. This results in a final sample consisting of 9,200 firm-year observations and a number of 2,377 unique firms. Appendix A provides a description of the matching procedure.

Table 2

Sample construction procedure

Steps in the sample construction procedure:	# of firm-years	# of unique firms
Firms ESG ratings from 2+ rating providers in each firm-year	46,002	8,544
Less observations missing carbon emission data	20,267	3,984
Less observations missing accounting and financial market data	17,747	3,402
Less observations missing institutional ownership data	9,200	2,377

Note. Table 2 presents the summary of my sample construction. The sample covers the period from 2010 to 2019 for firm-years with at least two ESG ratings from Refinitiv ESG, Sustainalytics, MSCI IVA and MSCI KLD.

3.3. Measurement and descriptive statistics

In order to measure firm-level ESG rating disagreement, I define *ESG Rating Disagreement* as the standard deviation of the different ratings a given firm receives in a given year. In addition, I define *Average ESG Rating* as the average of the ratings a firm receives from the different rating providers. To measure the change in a firm's environmental performance, I define *%Δ Carbon Emissions* as the percentual change in the firm's annual greenhouse gas emissions measured in equivalents of metric tons of CO_2. As this variable is subject to heavy outliers, especially in the upper tail, I winsorize this variable at the 5 and 90 percentiles. The variable measuring the level of institutional ownership in a firm, *Institutional Holdings,* is defined for each firm-year as the fraction of a firm's equity that is held by institutional investors at that time. Similarly, *Big3 Holdings* is defined as the fraction of a firm's equity that is held at parent level by the Big 3: BlackRock, Vanguard, and State Street Global Advisors.

My tests include a vector of firm-level control variables, defined as follows: *Size* is the logarithm of a firm's total assets. I include this variable to control for the volume of a firm's operations, as well as to control for possible public pressure into the improvement of

environmental performance. *Log (Book-to-market)* is the logarithm of the book-to-market ratio of a firm, computed as the book value of equity divided by the market value of equity. I include this variable to control for a firm's growth opportunities. Furthermore, I control for a firm's past performance with two variables: *ROA* is defined as a firm's net income over its total assets, and *Tobin's Q* is defined as the sum of a firm's market capitalization, long-term debt, and debt in current liabilities, scaled by its total assets. Moreover, *Leverage* is computed as the sum of a firm's long-term debt and debt in current liabilities scaled by total assets, and *PPE* is defined as the ratio of a firm's property, plant, and equipment over its total assets. I include these two variables in order to control for a firm's credit constraints: firms with a more leveraged capital structure are subject to regular cash outflows in the form of interest payments and debt repayment, which could hinder environmentally beneficial investments. Conversely, firms with higher ratios of PPE have more assets that could be used as collateral in borrowings, which could allow for more sustainable investments. In order to mitigate the effect of outliers, I winsorize all continuous control variables at the 5 and 95 percentiles.

Table 3 provides descriptive statistics for the main variables used in my tests. As shown in Panel A, the mean of *%Δ Carbon Emissions* is equal to -1.63, with a median of -1.52. These values indicate that firms are, on average, reducing their annual carbon emissions. However, the standard deviation, as well as the values of the 25^{th} and 75^{th} percentile, show that the change in carbon emissions widely varies over firms. Panel A also shows that the level of ESG rating disagreement is relatively large, with a mean standard deviation of 0.59 across rating providers. Chapter 4, as well as Appendix B describe this variable in more detail. Furthermore, the level of institutional ownership is 42% on average, with a standard deviation of 44% and a 75^{th} percentile of 83%. These values are in line with earlier research on institutional ownership (e.g. Bena et al., 2017). Additionally, ownership by the Big 3 has a mean of 9%, a standard deviation of 10.39% and a 75^{th} percentile of 18.28%. This suggest that the Big 3 have ample opportunity to influence firm policies around the world. Panel A also shows that the firms in my sample are rather large in size, profitable, and vary in terms of leverage and tangibility.

Panel B of Table 3 shows the country composition of my sample. My sample spans a total of 61 countries, with about 55% of my sample consisting of firms from the United States, Japan, the United Kingdom, and Canada. Panel C of Table 3 shows a wide variety of industry affiliation across firms, with most firms being active in manufacturing, financials, and business equipment.

Table 3

Panel A: Descriptive statistics

Variable	Mean	Std. dev.	P25	Median	P75
%Δ Carbon Emissions	-1.63	12.48	-8.50	-1.52	5.60
ESG Rating Disagreement	0.59	0.35	0.31	0.54	0.81
Average ESG Rating	-0.04	0.82	-0.71	-0.05	0.58
Institutional Holdings	42.19	44.09	0.10	39.29	83.04
Big3 Holdings	9.01	10.39	0	2.70	18.28
Controls:					
Size	9.53	1.46	8.43	9.41	10.55
Log (Book-to-market)	-0.68	0.86	-1.26	-0.60	-0.04
ROA	0.05	0.05	0.01	0.04	0.08
Tobin's Q	1.41	1.06	0.69	1.06	1.79
Leverage	0.25	0.15	0.14	0.25	0.36
PPE	0.30	0.24	0.09	0.25	0.47

Panel B: Country composition

Country	Freq.	Pct.	Country	Freq.	Pct.
United States	2761	30.01	Russia	58	0.63
Japan	1101	11.97	Austria	55	0.60
United Kingdom	627	6.82	Jersey	54	0.59
Canada	574	6.24	Bermuda	51	0.55
France	392	4.26	New Zealand	38	0.41
Australia	291	3.16	Belgium	36	0.39
Germany	282	3.07	Indonesia	36	0.39
Taiwan	253	2.75	Luxembourg	36	0.39
South Korea	223	2.42	Portugal	35	0.38
South Africa	202	2.20	Ireland	34	0.37
Brazil	187	2.03	Israel	31	0.34
Chile	178	1.93	Greece	30	0.33
Spain	153	1.66	Philippines	26	0.28
The Netherlands	131	1.42	Colombia	15	0.16
Hong Kong	125	1.36	Hungary	15	0.16
Sweden	122	1.33	Argentina	12	0.13
Italy	110	1.20	Panama	8	0.09
Cayman Islands	95	1.03	Poland	8	0.09
China	92	1.00	Curaçao	7	0.08
India	87	0.95	Guernsey	7	0.08
Denmark	79	0.86	Qatar	7	0.08
Malaysia	74	0.80	United Arab Emirates	5	0.05
Finland	65	0.71	Czech Republic	4	0.04
Mexico	65	0.71	Malta	3	0.03
Norway	64	0.70	Papua New Guinea	3	0.03
Singapore	62	0.67	Egypt	2	0.02
Chile	60	0.65	Mauritius	2	0.02
Thailand	60	0.65	Peru	2	0.02
Turkey	59	0.64	Other	6	0.06

Table 3, continued

Panel C: Industry composition

Industry	Freq.	Pct.
Other	1499	16.29
Manufacturing	1245	13.53
Financials	1067	11.60
Business Equipment	1021	11.10
Utilities	679	7.38
Retail/Wholesale	659	7.16
Consumer Nondurables	634	6.89
Energy	577	6.27
Healthcare	535	5.82
Chemicals	526	5.72
Telecommunications	432	4.70
Consumer Durables	326	3.54

Note. Table 3 reports descriptive statistics for the variables and observations used in my tests. The sample runs from 2010 until 2019 and includes 9,200 firm-year observations across 2,377 unique firms. Panel A presents descriptive statistics for the main variables used in my tests. Panel B describes the country composition of my sample. Panel C describes the industry composition of my sample, based on Fama and French's 12 industry categories.

4. Analysis of ESG rating disagreement

One of the underlying assumptions for my hypotheses is that different ESG rating providers substantially disagree with one another. Therefore, before I move forward with empirically testing my hypotheses, I first establish whether there is ESG rating disagreement present in my sample, and whether the level of this disagreement is consistent with previous studies on the topic. In addition, I analyze whether the level of ESG rating disagreement has changed over time, and whether it varies across industries.

Table 4 provides pairwise Pearson correlations between the ESG ratings of the four rating providers in my sample, as well as the average of these correlations. Panel A of Table 4 presents the correlation coefficients for the total ESG ratings. It follows from these correlations that there is substantial disagreement amongst ESG rating providers. The average correlation is equal to 0.526. Contrary to this, the correlation between credit ratings from different providers exceeds 0.99, according to Berg et al. (2022). This shows that while credit rating providers typically have nearly identical opinions on the credit worthiness of a firm, providers of ESG ratings have widely different views on the level of a firm's corporate social responsibility. Additionally, Panel A of Table 4 shows that the level of rating disagreement varies across pairs of providers. Refinitiv ESG and Sustainalytics have the highest correlation between one another, with a coefficient of 0.680. Conversely, Refinitiv ESG and MSCI IVA have the lowest level of agreement, with a correlation of 0.349.

In Panels B through D of Table 4, the pairwise correlations are disaggregated into the environmental, social and governance pillars of ESG ratings, respectively. Here, it can be seen that the level of rating disagreement widely varies across the E, S and G pillars. The highest level of agreement is found within the environmental pillar, with an average correlation of 0.512. On the other hand, the lowest level of agreement is found within the governance pillar, with an average correlation of 0.332. Within the social pillar, the average correlation is equal to 0.412. A possible explanation for this variation between pillars is the tangibility and objectiveness of corporate social responsibility in the respective pillars. In the environmental pillar, it might be easier to judge a firm's level of corporate social responsibility by looking at objective measures such as carbon emissions, usage of water or the amount of waste. In the social and governance pillars, however, the definition of corporate social responsibility might

Table 4

Pairwise correlations of ESG ratings

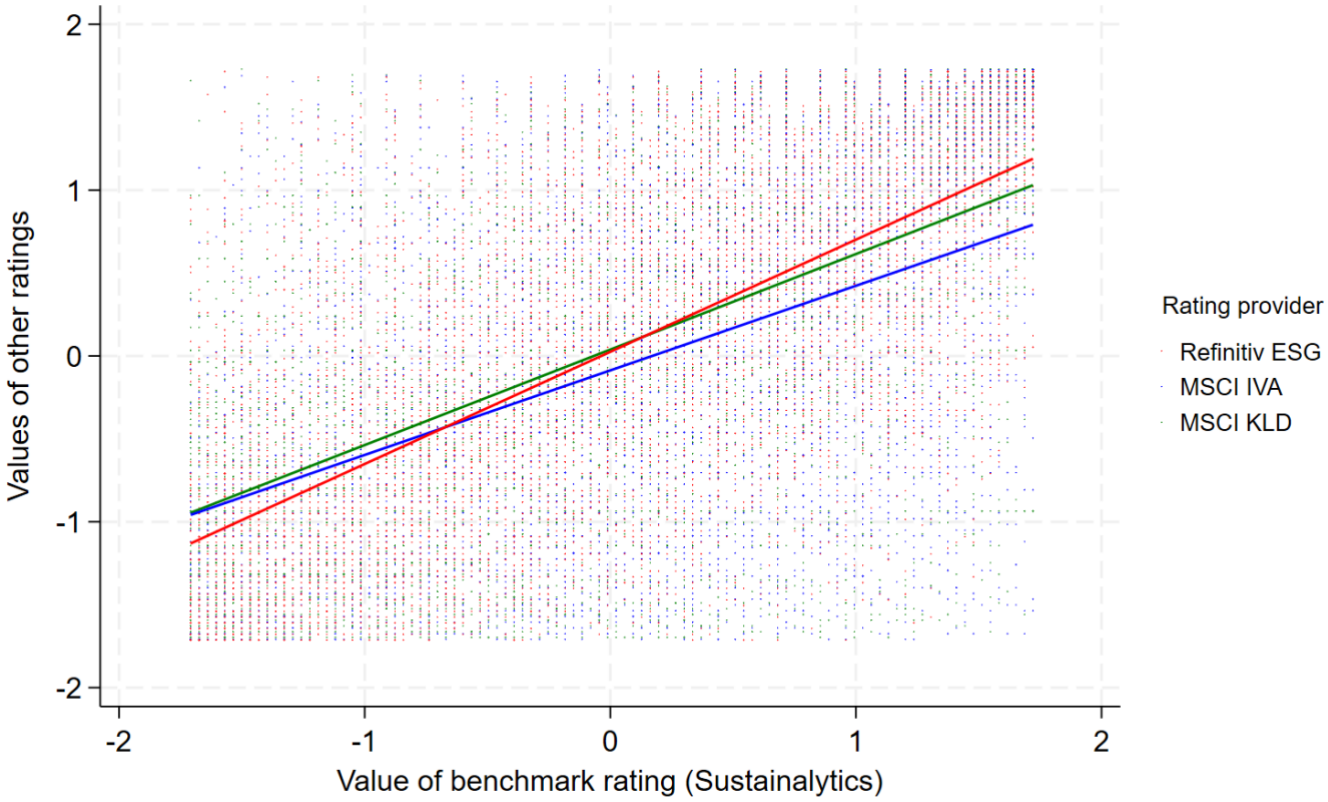
Panel A: Total pillar	Obs.	Pearson Correlations			
		Refinitiv ESG	Sustainalytics	MSCI IVA	
Refinitiv ESG	8,712				
Sustainalytics	8,547	0.680			
MSCI IVA	7,169	0.349	0.486		
MSCI KLD	7,189	0.551	0.581	0.506	
Average correlation					0.526
Panel B: Environmental pillar					
Refinitiv ESG	8,712				
Sustainalytics	8,547	0.652			
MSCI IVA	7,169	0.341	0.457		
MSCI KLD	7,189	0.560	0.638	0.433	
Average correlation					0.512
Panel C: Social pillar					
Refinitiv ESG	8,712				
Sustainalytics	8,547	0.602			
MSCI IVA	7,169	0.229	0.305		
MSCI KLD	7,189	0.489	0.483	0.386	
Average correlation					0.416
Panel D: Governance pillar					
Refinitiv ESG	8,712				
Sustainalytics	8,547	0.378			
MSCI IVA	7,169	0.188	0.329		
MSCI KLD	7,189	0.389	0.383	0.326	
Average correlation					0.332

Note. Table 4 presents pairwise Pearson correlation coefficients for each pair of data providers in my sample, as well as the average of the correlation coefficients. Panel A provides correlation coefficients for the total ESG ratings. Panels B through D provide correlation coefficients for the respective environmental, social and governance pillars.

be more subjective and subject to differing views of rating providers. This might explain why the level of rating disagreement is higher for these pillars, compared to the environmental pillar. When considering differences in sampling, the pairwise correlations I find are consistent with the results of prior studies on ESG rating disagreement from Chatterji et al. (2016), Gibson Brandon et al. (2021) and Berg et al. (2022).

Figure 1

Scatterplot of ESG ratings from different rating providers



Note. Figure 1 presents a scatterplot of the standardized values of the ESG ratings from the different rating providers in my sample. I use Sustainalytics as a benchmark rating. Refinitiv ESG is marked in red, MSCI IVA is marked in blue, and MSCI KLD is marked in yellow.

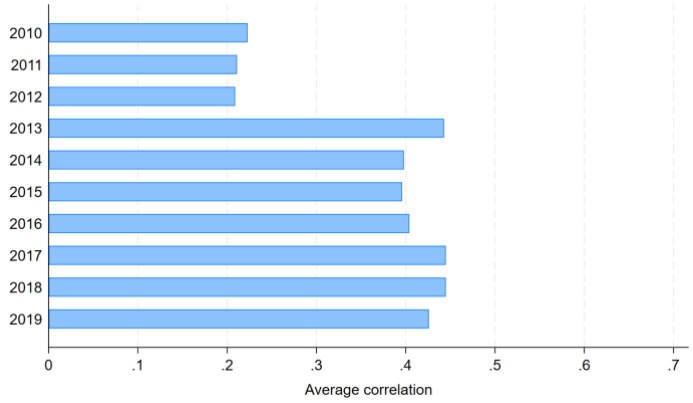
In Figure 1, I further illustrate the extent of rating disagreement by plotting the ratings from the different rating providers against each other. I take the ratings from Sustainalytics as a benchmark, as it has the highest correlation with the other rating providers. Similar to the aforementioned correlation coefficients, Figure 1 shows a positive correlation between the different rating providers. However, it also shows serious disagreement amongst rating providers. For example, when taking a value of +1 for a firm's benchmark rating (i.e. above

average), in some cases other rating providers give this firm a rating that is substantially below average. This means that in some cases, it is not possible to distinguish an above-average firm from a below-average firm, simply by looking at the different ESG ratings. As ESG ratings have been established with the purpose to evaluate a firm's ESG performance, the lack of agreement between different rating providers leaves managers, investors, and researcher with a

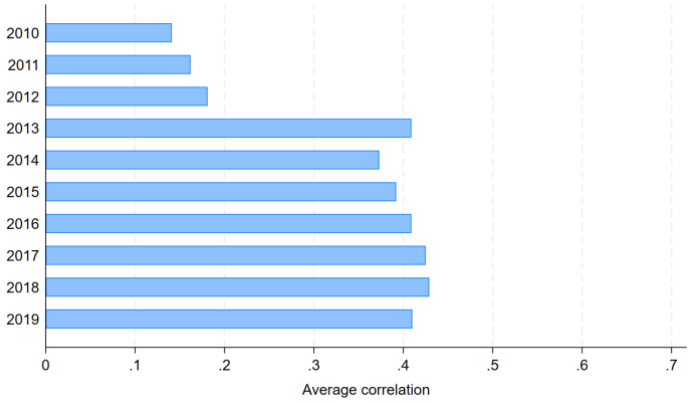
Figure 2

Average pairwise correlation per year

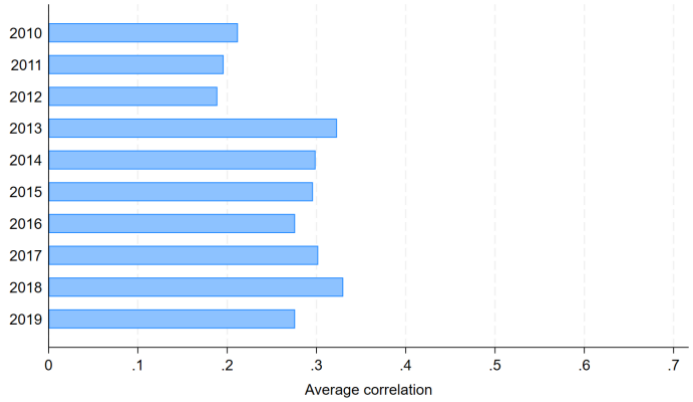
Panel A: Total pillar



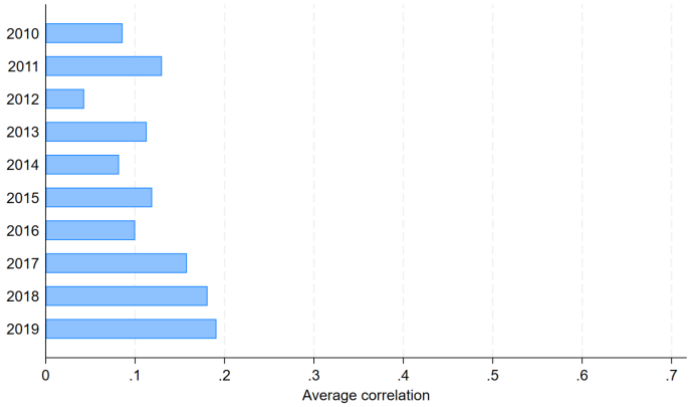
Panel B: Environmental pillar



Panel C: Social pillar



Panel D: Governance pillar



Note. Figure 2 shows bar charts of the average pairwise Pearson correlation between the ESG ratings from the different rating providers in my sample, grouped by year. Panel A provides the average correlation for the total pillar. Panels B through D provide the average correlation for the environmental, social and governance pillars, respectively.

feeling of uncertainty as to the level of corporate social responsibility within a firm.

An important question with regards to ESG rating disagreement is whether it has increased or decreased over time. While the prior evidence suggests that there is substantial ESG rating disagreement, it is possible that this disagreement has been alleviated over time, and that different rating providers have a more similar view on what constitutes corporate social responsibility. To get a view on this question, in Figure 2 I plot the average pairwise correlation per year in my sample. Here, a few observations stand out. First, Figure 2 shows that for all rating pillars, the level of rating disagreement is higher for the earlier years in my sample. A practical explanation for this might be due to the fact that the coverage of MSCI IVA in my sample is from 2013 onwards, and that the inclusion of this rating provider increases the average pairwise correlation. However, it might also suggest that over my sample period, a more common definition of corporate social responsibility has developed.

As counterevidence to the suggestion of the development of a more common standard in ESG ratings, Figure 2 shows that there is not a straightforward trend in the decrease of rating disagreement. Where in one year the average correlation between rating providers increases, it decreases in the next. This observation contraindicates an overall trend of decreasing ESG rating disagreement. For instance, for both the total, environmental and social pillar, the average correlation in 2019 is lower than it was in 2013. This also might support the findings of Christensen et al. (2022), whose results indicate an increase in ESG rating disagreement over time, instead of a decrease.

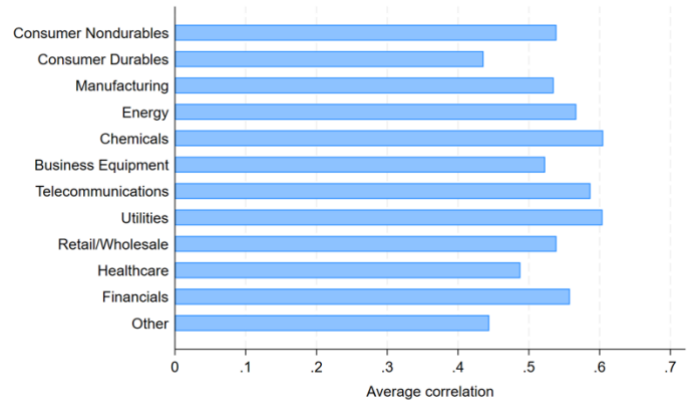
A second interesting question is how ESG rating disagreement varies across industries. In Figure 3, I group the average correlation between the different rating providers by their 12 Fama French-industry classifications.[10] Figure 3 illustrates that the level of rating disagreement does vary across different industries. Panel A of Figure 3 shows that for the total rating, disagreement in my sample is highest (i.e. average correlations are low) in the consumer durables and healthcare sectors. For both these industries, this disagreement seems to be driven by low average correlations in the social pillar (Panel C) and the governance pillar (Panel D). Contrary to this, the level of rating disagreement seems to be the lowest (i.e. average correlations are high) in the chemicals, utilities, and telecommunications sectors. This low of level of disagreement seems to be driven by relatively high average correlations in the environmental

[10] The 12 Fama French industry classifications can be found on Kenneth R. French's website: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_12_ind_port.html.

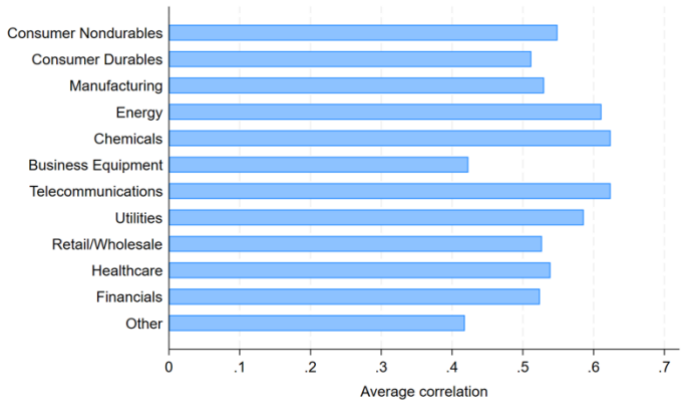
Figure 3

Average pairwise correlation per Fama French 12-industry category

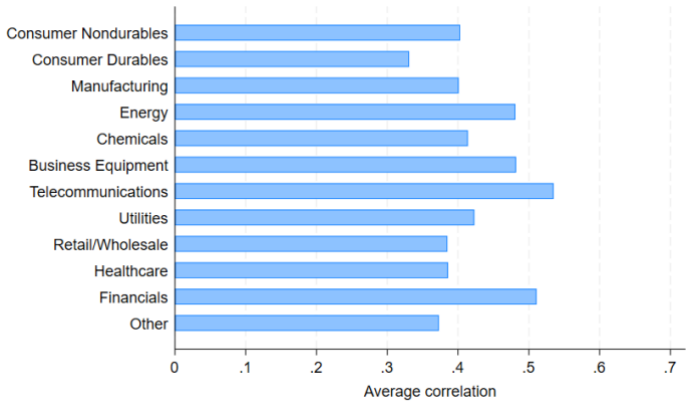
Panel A: Total pillar



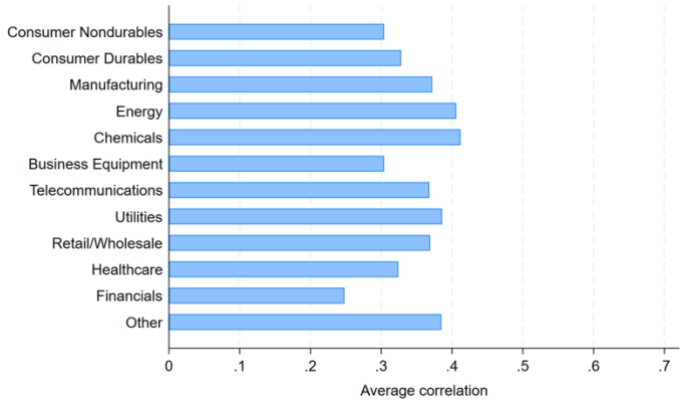
Panel B: Environmental pillar



Panel C: Social pillar



Panel D: Governance pillar



Note. Figure 2 shows bar charts of the average pairwise Pearson correlation between the ESG ratings from the different rating providers in my sample, grouped by a firm's Fama French 12-industry portfolio classification. Panel A provides the average correlation for the total pillar. Panels B through D provide the average correlation for the environmental, social and governance pillars, respectively.

pillar (Panel B) and governance pillar (Panel D) for the chemicals and utilities industries, and in the environmental pillar (Panel B) and social pillar (Panel C) for the telecommunications industry. With regards to industry-wide variation in rating disagreement, my results differ from the findings of Gibson Brandon et al. (2021). A possible explanation for this is that their sample is limited to firms from the S&P 500, while my sample contains a worldwide cross-section of firms. In addition, they include seven different rating providers in their sample, while my

sample is limited to the four rating providers that are publicly available to students from the Erasmus University.

To conclude, my findings show that in my sample, the correlation between the ratings of the different rating providers is relatively low, indicating the presence of considerable disagreement among these providers to the point where one cannot distinguish a leading firm from a laggard. The level of disagreement is naturally dependent on sample selection choices, but my findings are consistent with the findings of other studies employing a different sample selection. Additionally, the level of ESG rating disagreement in my sample varies considerably over time and across industries. These observations imply that managers, investors, and researchers need to consider the rating-year, as well as the industry a firm is active in when analyzing the different ESG ratings are firm receives, and when comparing these ratings to the ratings of other firms.

5. Empirical analysis

Now that I have established the presence of considerable ESG rating disagreement in my sample, I test the effect of this disagreement on the actual environmental performance of a firm.

5.1. The effect of ESG rating disagreement on environmental performance

I first test the standalone effect of ESG rating disagreement on the subsequent ESG performance of a firm. Here, I hypothesize the following:

Hypothesis 1*: Firms with larger ESG rating disagreement improve their subsequent ESG performance to a lesser extent.*

To investigate the relation between ESG rating disagreement and subsequent ESG performance, I estimate the following model:

$$\%\Delta\, Carbon\ emission_{i,t+1} = \alpha + \beta_1 * ESG\ Rating\ Disagreement_{i,t} + \beta_2 * Average\ ESG\ Rating_{i,t} + \Phi * Controls_{i,t} + \tau_t + \delta_i + \varepsilon_{i,t} \quad (1)$$

where *%Δ Carbon emissions, ESG Rating Disagreement, Average ESG Rating* and *Controls* are as previously defined (see Appendix C for variable definitions). Subindexes *i* and *t* refer to firm *i* and year *t*, respectively. τ_t and δ_i denote year and firm fixed effects, respectively. I cluster standard errors at the firm level. Under Hypothesis 1, I predict β_1 to be positive and statistically significant.

Table 5 presents the results of this model. For specification (1), I find that the coefficient on *ESG Rating Disagreement* is negative and statistically significant at the 10%-level. This negative association is robust to including industry and country fixed effects in specification (2). In economic terms, moving from the 25th percentile to the 75th percentile on the statistical distribution of *ESG Rating Disagreement* is associated with a subsequent carbon emission reduction of around 0.7 percentage points. However, this association is not robust to including year and firm fixed effects in specifications (3) and (4), as the coefficient on *ESG Rating Disagreement* is statistically insignificant for these specifications.

For specifications (1) – (3), the coefficients on *Average ESG Rating* are negative and statistically significant, indicating that firms with a higher average ESG rating subsequently reduce their carbon emissions to a larger extent. This negative association is robust to including both industry, country, and year fixed effects. In economic terms, a one standard deviation increase in the average ESG rating a firm receives is associated with a 0.95 to 2.17 standard deviations increase in carbon emission reduction. However, specification (4) shows that the negative association between the average ESG rating and subsequent carbon emissions is not robust to including firm fixed effects.

5.2. The effect of institutional ownership

In order to analyze the effect of the presence of institutional ownership on the subsequent change in carbon emissions for a firm, and to analyze the interaction between institutional ownership and ESG rating disagreement, I include the include the level of institutional ownership in my tests. Here, I hypothesize the following:

Hypothesis 2*: The ESG performance of firms is positively affected by the level of institutional ownership.*

Hypothesis 3*: The ESG performance of firms is less affected by ESG rating disagreement in the presence of higher levels of institutional ownership.*

In order to test these hypotheses, I expand my model under Hypothesis 1 to the following equation:

$$\begin{aligned}\%\Delta\, Carbon\ emission_{i,t+1} &= \alpha + \beta_1 * ESG\ Rating\ Disagreement_{i,t} + \beta_2 * Average\ ESG\ Rating_{i,t} \\ &+ \beta_3 * Institutional\ Ownership_{i,t} + \beta_4 * ESG\ Rating\ Disagreement_{i,t} \\ &* Institutional\ Ownership_{i,t} + \Phi * Controls_{i,t} + \tau_t + \delta_i + \varepsilon_{i,t}\end{aligned} \quad (2)$$

In this model, I include two specifications for *Institutional Ownership*. I first include the total level of institutional ownership, *Institutional Holdings*. Second, I distinguish between Big 3 ownership and non-Big3 ownership with *Big3 Holdings* and *Non-Big3 Holdings*, which is the difference between total institutional and Big 3 ownership.

Table 5

ESG rating disagreement and carbon emission change

	Dependent variable: *%Δ Carbon Emissions*			
	(1)	(2)	(3)	(4)
ESG Rating Disagreement	-0.69*	-0.73**	0.12	0.87
	(-1.77)	(-2.33)	(0.46)	(1.41)
Average ESG Rating	-2.17***	-2.06***	-0.95***	0.49
	(-12.71)	(-7.70)	(-3.08)	(0.96)
Size	0.16	0.30	-0.02	-1.73*
	(1.37)	(1.28)	(-0.06)	(-1.87)
Log (Book-to-market)	0.47	-0.75*	-0.33	-0.13
	(1.48)	(-1.92)	(-0.80)	(-0.19)
ROA	10.70**	10.37*	8.59	7.33
	(2.52)	(1.82)	(1.48)	(1.26)
Tobin's Q	1.08***	0.48	0.73	0.52
	(3.71)	(1.03)	(1.50)	(0.92)
Leverage	-1.55	-1.05	0.28	4.40
	(-1.51)	(-0.56)	(0.16)	(1.44)
PPE	3.06***	1.95*	2.29**	-3.35
	(4.73)	(1.93)	(2.18)	(-0.87)
Industry FE	NO	YES	YES	NO
Country FE	NO	YES	YES	NO
Year FE	NO	NO	YES	YES
Firm FE	NO	NO	NO	YES
N	9,200	9,094	9,094	9,200
R^2	0.03	0.05	0.08	0.05
Adjusted R^2	0.03	0.04	0.07	0.05

Note. Table 5 presents the results of the regression analyses between the level of ESG rating disagreement and subsequent changes in carbon emissions. The sample spans from 2010 to 2019 and includes 9,200 firm-year observations. All variables are defined in Appendix C. Column (1) includes no fixed effects. Column (2) includes both industry and country fixed effects. Column (3) includes both industry, country, and year fixed effects. Column (4) includes both year and firm fixed effects. Standard errors are clustered at the firm level.

*, ** and *** denote significance at the 10%, 5% and 1% levels, respectively. Intercepts are omitted. T-statistics are in parentheses.

All other specifications are the same as in Model 1. Under Hypothesis 2, I predict β_3 to be negative and statistically significant. I test Hypothesis 3 by including an interaction effect between the level of ESG rating disagreement, and the level of institutional ownership. Under Hypothesis 3, I predict that β_4 is negative and statistically significant.

Table 6 presents the results of this model. Interestingly, when I include the level of institutional ownership in Model 2, the coefficient on *ESG Rating Disagreement* is positive for all specifications, and statistically significant at the 5%-significance level for the specifications with industry, country, and year fixed effects in Columns (3) and (4), and for the specifications with year and firm fixed effects in Columns (5) and (6). In terms of economic magnitude, going from the 25th percentile to the 75th percentile on the *ESG Rating Disagreement* distribution is associated with 1.1 – 1.8 percentage point increase in subsequent carbon emissions.

Columns (1), (3) and (5) of Table 6 present the effect of the level of institutional ownership on the subsequent change in carbon emissions. Only when including year and firm fixed effects, the coefficient of *Institutional Holdings* is significant at the 10% level. However, this coefficient of *Institutional Holdings* has a positive sign, meaning that a higher level of institutional ownership is associated with an increase in subsequent carbon emissions. In Columns (2), (4) and (6) of Table 6, I split up the total level of institutional ownership into ownership by the Big 3 and other institutional ownership. For specifications (2) and (4), the coefficient on *Big3 Holdings* is negative and statistically significant at the 10% and 5% level, respectively, and robust for industry, country, and year fixed effects. In economic terms, a 1% increase in the level of Big 3 ownership is associated with a 0.09 percentage point decrease in subsequent carbon emissions. To me, this effect is quite plausible. However, when I include year and firm fixed effects in specification (6), the coefficient is no longer statistically significant.

The coefficients on *Non-Big3 Holdings* paint an opposite picture: for specifications (2) and (4), there is a positive association between the level of non-Big 3 ownership and subsequent carbon emissions, which is statistically significant at the 10% and 5% level, respectively. In terms of magnitude, a 1% increase in non-Big 3 ownership is associated with a 0.03 to 0.04 percentage point increase in subsequent carbon emissions. Again, the coefficient is no longer statistically significant when including year and firm fixed effects in specification (6).

Table 6

ESG rating disagreement, the change in carbon emissions and the effect of institutional ownership

	Dependent variable: *%Δ Carbon Emissions*					
	(1)	(2)	(3)	(4)	(5)	(6)
ESG Rating Disagreement	0.45	0.48	1.13**	1.15**	1.83**	1.79**
	(0.77)	(0.84)	(2.28)	(2.31)	(2.36)	(2.30)
Average ESG Rating	-2.09***	-2.03***	-0.91***	-0.89***	0.70	0.71
	(-7.49)	(-7.49)	(-2.89)	(-2.84)	(1.35)	(1.36)
Institutional Holdings	0.01		0.01		0.02*	
	(0.96)		(1.28)		(1.91)	
*ESG Rating Disagreement * Institutional Holdings*	-0.03**		-0.03**		-0.03*	
	(-2.60)		(-2.12)		(-1.86)	
Big3 Holdings		-0.09*		-0.09**		-0.03
		(-1.90)		(-2.15)		(-0.41)
*ESG Rating Disagreement * Big3 Holdings*		0.01		0.08		-0.03
		(0.11)		(1.58)		(-0.34)
Non-Big3 Holdings		0.03*		0.04**		0.04
		(1.82)		(2.33)		(1.58)
*ESG Rating Disagreement * Non-Big3 Holdings*		-0.04*		-0.05***		-0.02
		(-1.97)		(-2.76)		(-0.69)
Size	0.23	0.19	-0.06	-0.07	-1.76*	-1.76*
	(1.03)	(0.89)	(-0.24)	(-0.29)	(-1.90)	(-1.91)
Log (Book-to-market)	-0.73*	-0.71*	-0.33	-0.31	-0.13	-0.12
	(-1.86)	(-1.80)	(-0.78)	(-0.76)	(-0.19)	(-0.18)
ROA	10.27*	10.50*	8.55	8.54	7.61	7.72
	(1.79)	(1.85)	(1.45)	(1.46)	(1.30)	(1.32)
Tobin's Q	0.47	0.47	0.72	0.72	0.52	0.52
	(1.03)	(1.05)	(1.48)	(1.51)	(0.92)	(0.92)
Leverage	-0.96	-0.95	0.30	0.31	4.59	4.61
	(-0.51)	(-0.51)	(0.18)	(0.18)	(1.51)	(1.52)
PPE	1.99*	2.05*	2.32**	2.36**	-3.41	-3.41
	(1.92)	(1.94)	(2.16)	(2.18)	(-0.89)	(-0.89)
Industry FE	YES	YES	YES	YES	NO	NO
Country FE	YES	YES	YES	YES	NO	NO
Year FE	NO	NO	YES	YES	YES	YES
Firm FE	NO	NO	NO	NO	YES	YES
N	9,094	9,094	9,094	9,094	9,200	9,200
R^2	0.05	0.05	0.08	0.08	0.05	0.05
Adjusted R^2	0.04	0.04	0.07	0.07	0.05	0.05

Note. Table 6 presents the results of the regression analyses between ESG rating disagreement, institutional ownership, and the subsequent change in carbon emissions. All variables are defined in Appendix C. Columns (1) and (2) include industry and country fixed effects, Columns (3) and (4) include industry, country and year fixed effects, and columns (5) and (6) include year and firm fixed effects. Standard errors are clustered at the firm level.

*, ** and *** denote significance at the 10%, 5% and 1% levels, respectively. Intercepts are omitted. T-statistics are in parentheses.

Table 6 also presents results for the interaction between the level of ESG rating disagreement and the level of institutional ownership. For specifications (1) and (3), the coefficient on *ESG Rating Disagreement * Institutional Holdings* is negative and statistically significant at the 5% level. This negative association is robust to including industry, country, and year fixed effects. When including firm fixed effects in specification (5), the negative association is significant at the 10% level. In Columns (2), (4) and (6) of Table 6, I again split up institutional ownership into Big 3 ownership and non-Big 3 ownership. Here, the coefficient on *ESG Rating Disagreement * Big3 Holdings* is statistically insignificant at conventional levels. Conversely, the coefficient on *ESG Rating Disagreement * Non-Big3 Holdings* is negative and statistically significant at the 10% and 1% levels when including industry, country, and year fixed effects in specifications (2) and (4), respectively. However, when including firm and year fixed effects in the regression, the coefficient is no longer significant.

5.3. Robustness

In my tests, I already control for time-invariant cross-sectional characteristics by including various combinations of industry, country, year, and firm fixed effects. However, as shown in Table 2, the level of ESG rating disagreement heavily changes after the inclusion of MSCI IVA in my sample. Therefore, as a robustness check, I repeat my previous analyses for firms that have received ratings from all four rating providers in my sample. This restriction reduces my sample to 4,441 firm-year observations, from the year 2013 through 2019.

Table 7 presents the results of Model 1 for my restricted sample. Contrary to the results in Table 5, the coefficient of *ESG Rating Disagreement* is negative for all model specifications, and statistically significant at the 1% level for specifications (1) and (2), and at the 10% level for specification (3). However still, the coefficient is statistically insignificant when I include firm and year fixed effects. In addition, the coefficient on *Average ESG Rating* is significant and negative, and robust for including industry, country, and year fixed effects, in light with the results in Table 5. Again, this coefficient is no longer significant when including firm and year fixed effects.

The results of the repeated analysis of Model 2 are presented in Table 8. Contrary to the results in Table 6, the coefficient on *ESG Rating Disagreement* is insignificant for all specifications. However, similarly, the coefficient of *Institutional Holdings* is positive and significant at the

10% level when including year and firm fixed effects. Moreover, the coefficients on *Big3 Holdings* are again negative and significant at the 10% level when including both industry, country fixed effects, and when including industry, country, and year fixed effects. In addition, the coefficient on *Non-Big3 Holdings* is positive and statistically significant at the 10% level for the same fixed effects specifications. When looking at the interaction effects, the coefficient of *ESG Rating Disagreement * Institutional Holdings* is again negative and significant, and robust for including both industry, country, and year fixed effects.

Table 7

Robustness test Model 1 with restricted sample

	Dependent variable: *%Δ Carbon Emissions*			
	(1)	(2)	(3)	(4)
ESG Rating Disagreement	-1.72***	-2.16***	-1.05*	-0.87
	(-2.91)	(-4.43)	(-1.69)	(-0.86)
Average ESG Rating	-2.55***	-2.57***	-1.22**	0.22
	(-9.18)	(-7.27)	(-2.53)	(0.24)
Size	0.08	0.51	0.30	-1.71
	(0.45)	(1.56)	(0.89)	(-1.34)
Log (Book-to-market)	0.60	-0.47	0.03	0.66
	(1.33)	(-0.87)	(0.05)	(0.70)
ROA	12.07*	11.41	7.50	7.87
	(1.82)	(1.17)	(0.76)	(0.99)
Tobin's Q	1.22***	0.74	1.11	1.17
	(2.89)	(1.08)	(1.56)	(1.46)
Leverage	-3.50**	-2.26	-0.48	5.16
	(-2.31)	(-0.75)	(-0.18)	(1.14)
PPE	2.54***	1.59	2.00	0.45
	(2.84)	(1.10)	(1.49)	(0.08)
Industry FE	NO	YES	YES	NO
Country FE	NO	YES	YES	NO
Year FE	NO	NO	YES	YES
Firm FE	NO	NO	NO	YES
N	4,411	4,365	4,365	4,411
R^2	0.04	0.06	0.10	0.08
Adjusted R^2	0.04	0.04	0.08	0.08

Note. Table 7 presents the results of the regression analyses between the level of ESG rating disagreement and subsequent changes in carbon emissions for firms that have received a rating from all four rating providers in my sample. The restricted sample spans from 2013 to 2019 and includes 4,411 firm-year observations. All variables are defined in Appendix C. Column (1) includes no fixed effects. Column (2) includes both industry and country fixed effects. Column (3) includes both industry, country, and year fixed effects. Column (4) includes both year and firm fixed effects. Standard errors are clustered at the firm level.

*, ** and *** denote significance at the 10%, 5% and 1% levels, respectively. Intercepts are omitted. T-statistics are in parentheses.

Table 8

Robustness test Model 2 with restricted sample

	Dependent variable: *%Δ Carbon Emissions*					
	(1)	(2)	(3)	(4)	(5)	(6)
ESG Rating Disagreement	-0.74	-0.67	0.20	0.20	0.55	0.54
	(-0.91)	(-0.82)	(0.21)	(0.22)	(0.39)	(0.38)
Average ESG Rating	-2.50***	-2.37***	-1.16**	-1.11**	0.40	0.41
	(-7.14)	(-7.03)	(-2.42)	(-2.32)	(0.45)	(0.45)
Institutional Holdings	0.02		0.02		0.11*	
	(1.11)		(1.00)		(1.71)	
*ESG Rating Disagreement * Institutional Holdings*	-0.04**		-0.03*		-0.03	
	(-2.21)		(-1.95)		(-1.38)	
Big3 Holdings		-0.15*		-0.14*		0.09
		(-1.83)		(-1.87)		(0.45)
*ESG Rating Disagreement * Big3 Holdings*		0.00		0.02		-0.04
		(0.02)		(0.18)		(-0.22)
Non-Big3 Holdings		0.04*		0.04*		0.12
		(1.68)		(1.69)		(1.58)
*ESG Rating Disagreement * Non-Big3 Holdings*		-0.05		-0.05		-0.03
		(-1.43)		(-1.38)		(-0.52)
Size	0.48	0.44	0.28	0.25	-1.77	-1.77
	(1.47)	(1.38)	(0.81)	(0.73)	(-1.40)	(-1.39)
Log (Book-to-market)	-0.46	-0.42	0.03	0.06	0.79	0.79
	(-0.85)	(-0.79)	(0.06)	(0.11)	(0.84)	(0.84)
ROA	11.54	12.10	7.59	8.15	8.38	8.46
	(1.18)	(1.31)	(0.77)	(0.86)	(1.05)	(1.06)
Tobin's Q	0.73	0.71	1.10	1.08	1.27	1.27
	(1.07)	(1.10)	(1.55)	(1.60)	(1.58)	(1.58)
Leverage	-2.25	-2.05	-0.48	-0.36	5.69	5.78
	(-0.75)	(-0.71)	(-0.18)	(-0.13)	(1.25)	(1.26)
PPE	1.64	1.71	2.05	2.11	0.51	0.49
	(1.12)	(1.12)	(1.49)	(1.49)	(0.09)	(0.09)
Industry FE	YES	YES	YES	YES	NO	NO
Country FE	YES	YES	YES	YES	NO	NO
Year FE	NO	NO	YES	YES	YES	YES
Firm FE	NO	NO	NO	NO	YES	YES
N	4,365	4,365	4,365	4,365	4,411	4,411
R^2	0.06	0.06	0.10	0.10	0.08	0.08
Adjusted R^2	0.05	0.05	0.08	0.08	0.08	0.08

Note. Table 8 presents the results of the regression analyses between ESG rating disagreement, institutional ownership, and the subsequent change in carbon emissions. All variables are defined in Appendix C. Columns (1) and (2) include industry and country fixed effects, Columns (3) and (4) include industry, country and year fixed effects, and columns (5) and (6) include year and firm fixed effects. Standard errors are clustered at the firm level.

*, ** and *** denote significance at the 10%, 5% and 1% levels, respectively. Intercepts are omitted. T-statistics are in parentheses.

6. Conclusion

6.1. Discussion

This study is centered around evaluating the real-life consequences of ESG rating disagreement on firm-level activities; more specifically, to analyze the relation between ESG rating disagreement and the subsequent change in carbon emissions of a firm. As far as I am aware, this study is the first to directly investigate this relation, meaning that I did not precisely know what to expect from my results.

Before I get to my variables of interest, a different interesting result is that the average ESG rating a firm receives is significant and negatively associated with the change in carbon emissions. This is inconsistent with the results of Chatterji and Toffel (2010), who instead find that lower rated firms are more likely to reduce their carbon emissions. A possible explanation for my findings is simply that the average ESG rating a firm receives does its job, namely evaluating which firms have a good ESG performance. In this case, firms with a higher average ESG rating are actually striving for improving their ESG performance, and therefore continue to reduce their carbon emissions. Alternatively, firms might continue to reduce their carbon emissions in order to maintain their ESG rating from a strategic or investor point of view, instead of actually caring about their carbon emissions.

Under Hypothesis 1, I predicted that firms with a higher extent of ESG rating disagreement would improve their subsequent environmental performance to a lesser extent, meaning that they would exhibit smaller reductions or even increases in their subsequent carbon emissions. Here, my results paint a convoluted picture. In Model 1, the standalone effect of ESG rating disagreement on carbon emission is negative and significant without including fixed effects, and when including industry and country fixed effects. This indicates that ESG rating disagreement on its own is associated with a further reduction of carbon emissions. However, this effect is no longer significant when I include both industry, country, and year fixed effects, and when including year and firm fixed effects. This means that for Model 1, the effect of rating disagreement on carbon emissions is likely driven by these year fixed effects, and by time-invariant firm characteristics. In Model 2, however, the effect of ESG rating disagreement on carbon emission change is drastically different: when I account for the level of institutional ownership in a firm, the effect of rating disagreement on subsequent carbon emission change is significant and positive, and robust to including both industry, country, and year fixed effects,

as well to including year and firm fixed effects. This effect indicates that when controlling for institutional ownership, firms with higher levels of rating disagreement exhibit either lower subsequent reductions, or even subsequent increases in their carbon emissions. This result is in line with my prediction under Hypothesis 1.

Nonetheless, my results under Hypothesis 1 in Model 2 need to be looked at in light of the results under my other two hypotheses. Under Hypothesis 2, I predicted that a higher level of institutional ownership is associated with a subsequent reduction in carbon emissions. Contrary to this prediction, my main results in Model 2 show that when including year and firm fixed effects, a higher level of institutional ownership is associated with an *increase* in carbon emissions, instead of a reduction. This result is opposed to the notion and current trend that institutional ownership leads to an improvement of a firm's environmental performance, which is also what Dyck et al. (2019) find in their study. When I divide institutional ownership into ownership by the Big 3 and other institutional ownership, I find that ownership by the Big 3 has a negative association, albeit with weak statistical significance, with the subsequent change in carbon emissions. This association is robust to both industry, country, and year fixed effects, but becomes insignificant when I include year and firm fixed effects. Therefore, this result is not entirely consistent with the results of Azar et al. (2021), who find that a higher level of Big 3 ownership leads to a decrease in carbon emissions. However, as a sidenote, my study uses a different dependent variable, namely the change in carbon emissions, and not the total level of carbon emissions. Lastly, non-Big 3 ownership is associated with an increase in carbon emissions in Model 2. All in all, these results are not entirely in line with my prediction under Hypothesis 2.

Lastly, under Hypothesis 3, I predicted that the change in carbon emissions of a firm with higher levels of institutional ownership is less affected by ESG rating disagreement. This prediction is consistent with my results: the interaction effect between institutional ownership and ESG rating disagreement is negative and significant, and robust to including both industry, country, and year, as well as to including year and firm fixed effects. This effect indicates that the higher the level of institutional ownership within a firm, the lesser the carbon emission change of this firm is affected by ESG rating disagreement. This is in line with my prediction under Hypothesis 3. However, this interaction effects needs to be interpreted with caution, as the separate effects are not always statistically significant.

Seen together, my results for Model 2 paint an interesting picture. A potential explanation for these results might be that in my model, firms are in two ways incentivized to reduce their carbon emissions: through institutional ownership, and through the ESG ratings they receive. If the level of institutional ownership becomes higher, firms become less reliant on ESG ratings in order to evaluate their ESG performance but are instead incentivized by the institutional ownership to improve their performance through reducing their carbon emissions. Conversely, when there is a low level or absence of institutional ownership, firms observe their ESG ratings in order to decide whether they need to act. If then there is a high level of ESG rating disagreement, firms are uncertain whether they are expected to undertake action and are disincentivized to reduce their carbon emissions. Please note that this is only a potential explanation based on my results. In reality, there are many reasons why firms would want to reduce their carbon emissions, and ESG ratings and institutional ownership are only two of them. This notion is in line with the study of Clementino and Perkins (2020), who find that firms do not uniformly react to the ESG ratings they receive, but instead have their own individual views on when they need to improve their ESG performance. Also for all my models, the R^2 spans only from 0.03 to 0.10, depending on the specification, meaning that a maximum of 10 percent of the variation in carbon emission change is explained by the variables in my model.

Moreover, my robustness tests show that my results are very dependent on the selected sample, and on the fixed effects specifications. For example, when I restrict my sample to firms that have received ratings from all four rating providers in my sample, many of the aforementioned effects disappear or lose significance. In addition, the multiple fixed effects specifications show that some of the mentioned effects are likely to be driven by time-invariant industry, country, or firm characteristics. Often when including firm fixed effects in the models, the effects of ESG rating disagreement or institutional ownership loses its significance, again indicating that the choice to reduce carbon emissions is firm-dependent, and likely driven by many variables.

All things considered, my results indicate that there is considerable ESG rating disagreement amongst different rating providers. As seen in prior studies (e.g. Gibson Brandon et al., 2021, and Christensen et al., 2022), this disagreement has real-life implications for both managers of a firm, investors on the financial markets, and for academics that perform research into ESG ratings. With regards to the question whether ESG rating disagreement has consequences for the change in carbon emissions of a firm, the answer is that this is unclear. Some firms might

pay much attention to ESG ratings, and others might not. The reality is though, that ESG rating disagreement remains a remarkable phenomenon, and that time will tell what its consequences are.

6.2. Limitations and recommendations for future research

Naturally, my research does have some limitations. A first limitation is that there is no prior research on this topic. My theoretical framework on ESG rating disagreement is limited to a few papers that investigate the consequences of ESG rating disagreement on the financial market, which is different from what I attempt to study. Due to this, I did not have a reasonable expectation of the results beforehand. A second limitation is the limited availability of different ESG ratings. For my research, I would have liked to have had access to more ESG rating databases than the four I use in this study, in order to improve the generalizability and internal validity of my results. In addition, some of the ESG rating databases I use in this study contain missing variables, such as missing ratings or missing identifying variables, due to which I most likely lost many observations in the matching procedure. Especially the MSCI KLD database was somewhat problematic, with a number of missing identifying variables and complicated variable names. Lastly, a limitation to the implications of my study is that my effects describe associations between variables, and not direct causal effects. As I mention in the discussion section, there are likely many variables that affect the carbon emission change of a firm, both observed and unobserved. The omission of these variables from my model could have possibly tainted my results.

I have several recommendations for future research into this topic. First of all, as a firm's choice to reduce carbon emissions is most likely driven by many variables, I recommend performing qualitive research into what drives firms to reduce their carbon emissions. As seen in my results, many variables explain only a fraction of variation in carbon emission reduction, and qualitative research would paint a detailed picture of an individual firm's reasons behind improving their environmental performance. A second recommendation is, naturally, to make use of a possible drastic exogenous change in ESG rating disagreement. A potential example of this could when there is a development of standardized ESG rating criteria, as these would most likely decrease ESG rating disagreement. Lastly, future research could attempt to study how the different pillars and sub-pillars affect firm-level outcomes, instead of looking at the total ESG rating.

Appendix A. Matching procedure

A challenge in constructing one dataset from many different databases is to properly match the different sets. I matched on the basis of three firm identifiers: (1) CUSIP, (2) ISIN, and (3) company name. For the ESG rating providers, MSCI KLD only has the CUSIP code as identifier, while Refinitiv ESG and Sustainalytics only have the ISIN code as identifier. MSCI IVA has both CUSIP and ISIN as identifier. To prevent observations not being matched to one another due to incompatible identifiers (i.e. an observation from Refinitiv ESG could only be matched to an observation from MSCI KLD if this observation was also included in the MSCI IVA database), I also matched based on a firm's company name. In order to do so, I first removed common suffixes to the company names, such as *Inc.*, *Co.*, *Group*, etc. Next, I used a fuzzy matching procedure to match the observations from the different databases and kept the observations with a sufficiently high level of similarity between company names. Next, I matched the combined ESG rating dataset with the Refinitiv ESG carbon emission, Compustat and Thomson/Refinitiv data based on the ISIN code. I also attempted a match to these databases based on company names, but that left me with fewer remaining observations.

Appendix B. Additional descriptive statistics of the level of ESG rating disagreement

For the purposes of further analyzing the extent of disagreement amongst the ESG ratings of the rating providers available to the Erasmus University Rotterdam, I have tabulated some additional descriptive statistics of the *ESG Rating Disagreement* variable, based on the matched dataset of the four ESG rating providers in my sample. These descriptive statistics have the benefit of not being constricted to the availability of other data sources in my empirical research, such as carbon emission or accounting data.

First of all, Table B.1 shows descriptive statistics for *ESG Rating Disagreement*, grouped by the number of rating providers covering a firm at that time. This table shows that the more rating providers are covering a firm, the higher the mean of *ESG Rating Disagreement* is. This might be explained by when there are more rating providers covering a firm, the more opportunity there is for rating disagreement, as there are separate ideas on what constitutes corporate social responsibility, and how separate elements of ESG are measured and weighed.

Table B.1

ESG Rating Disagreement and the number of rating providers

	Variable: *ESG Rating Disagreement*					
Number of ESG rating providers	Obs.	Mean	Std.dev.	P25	P50	P75
2	13,325	0.554	0.453	0.190	0.438	0.812
3	20,838	0.638	0.363	0.349	0.586	0.883
4	10,061	0.660	0.340	0.387	0.641	0.905
Total	44,224	0.617	0.390	0.306	0.561	0.876

Note. Table B.1 shows descriptive statistics for the variable *ESG Rating Disagreement*, grouped by the number of ESG rating providers that a firm is covered by at that time.

Another interesting question is how the level of ESG rating disagreement relates to the time between when another rating providers start covering the same firm. Hence, Table B.2 shows descriptive statistics for *ESG Rating Disagreement*, grouped by the number of years between the first and the second ESG rating a firm receives in the available data. For this, one might suggest that the longer a first rating provider has been covering a firm, the lower the level of rating disagreement might be, due to potentially less uncertainty on the ESG performance of this firm, or due to a potential 'herding effect' between different ESG rating providers.

Table B.2

ESG Rating Disagreement and the time in between ratings

	Variable: *ESG Rating Disagreement*					
Years between first and second ESG Rating	Obs.	Mean	Std.dev.	P25	Median	P75
0	23,811	0.649	0.392	0.334	0.606	0.915
1	16,897	0.567	0.364	0.279	0.504	0.802
2	1,134	0.657	0.446	0.291	0.575	0.952
3	822	0.683	0.472	0.295	0.598	0.967
4	1,248	0.594	0.476	0.197	0.492	0.883
5	134	0.790	0.507	0.321	0.812	1.197
6	101	0.623	0.420	0.319	0.559	0.860
7	41	0.580	0.351	0.294	0.515	0.863
8	25	0.547	0.323	0.297	0.548	0.746
9	10	0.518	0.360	0.203	0.536	0.805
10	1	0.414	0	0.414	0.414	0.414

Note. Table B.1 shows descriptive statistics for the variable *ESG Rating Disagreement*, as defined in Chapter 3 and Appendix B. The variable is grouped by the number of years between the first and second ESG rating a firm receives.

Table B.2 shows that when multiple rating providers start covering a firm at the same time in the sample (i.e. the number of years between the first and second rating is 0), the mean of *ESG Rating Disagreement* is equal to 0.649. However, when there is some time between the first and second ESG rating a firm receives, the mean of *ESG Rating Disagreement* seems to get somewhat smaller. However, this is not a straightforward pattern, as the mean in the case of 5 years in between the first and second ESG rating in the sample, for example, is larger than the mean when the first two ratings are given simultaneously.

A second interesting is whether the identity of the ESG rating provider(s) a firm is first covered by makes a difference on the level of ESG rating disagreement. For example, some rating providers might have a reputation of being trustworthy and reliable, and this might influence the ESG rating of other rating providers. Table B.2 again shows descriptive statistics of the variable *ESG Rating Disagreement,* grouped by the combinations of the initial ESG rating providers covering a firm. From Table B.2, a first interesting observation is that on aggregate, the mean of *ESG Rating Disagreement* is higher when a firm is covered by more than one rating

provider initially, and highest when this firm is covered by three rating providers from the start. This again points to some extent of a herding effect, where rating providers disagree with each other to a lesser extent when one provider is already covering a firm.

Table B.3

ESG Rating Disagreement and the first rating provider(s)

	Variable: *ESG Rating Disagreement*					
Panel A: One rating provider	Obs.	Mean	Std.dev.	P25	Median	P75
MSCI IVA	1,012	0.722	0.505	0.287	0.652	1.065
Refinitiv ESG	3,468	0.646	0.423	0.297	0.583	0.937
Sustainalytics	15,933	0.557	0.363	0.273	0.492	0.783
Total	20,413	0.580	0.384	0.276	0.492	0.822
Panel B: Two rating providers						
Refinitiv ESG & MSCI IVA	409	0.729	0.459	0.372	0.675	0.985
Refinitiv ESG & MSCI KLD	1,567	0.640	0.417	0.303	0.591	0.905
Refinitiv ESG & Sustainalytics	1,106	0.580	0.384	0.267	0.518	0.835
Sustainalytics & MSCI IVA	426	0.675	0.473	0.315	0.596	0.978
Sustainalytics & MSCI KLD	1,205	0.579	0.392	0.258	0.522	0.854
MSCI IVA & MSCI KLD	7,473	0.632	0.412	0.293	0.580	0.917
Total	12,186	0.628	0.413	0.291	0.574	0.907
Panel C: Three rating providers						
Refinitiv ESG, MSCI IVA & MSCI KLD	1,326	0.723	0.377	0.429	0.706	0.986
Refinitiv ESG, Sustainalytics & MSCI IVA	255	0.731	0.353	0.435	0.712	1.010
Refinitiv ESG, Sustainalytics & MSCI KLD	9,060	0.665	0.361	0.383	0.627	0.910
Sustainalytics, MSCI IVA & MSCI KLD	623	0.642	0.448	0.277	0.579	0.971
Total	11,264	0.672	0.368	0.383	0.636	0.922
Panel D: Four rating providers						
Refinitiv ESG, Sustainalytics, MSCI IVA & MSCI KLD	361	0.660	0.330	0.388	0.657	0.902

Note. Table B.2 shows descriptive statistics for the variable *ESG Rating Disagreement*, grouped by the providers that initially cover a firm. Panel A shows descriptive statistics in the case of one initial rating providers, while Panels B through D show descriptive statistics in the case of 2 through 4 initial rating providers, respectively.

Panel A shows descriptive statistics in the case of one initial rating provider. First of all, it is important to note that due to data availability issues, the ratings of MSCI KLD could only be incorporated in my sample while being matched to one of the other rating providers. This is the reason why MSCI KLD could not be a sole initial rating provider in my sample. When looking at the other rating providers, Panel A shows that when Sustainalytics is the first ESG rating provider covering a firm, the mean of *ESG Rating Disagreement* is lowest. This might suggest that Sustainalytics is considered as a high-quality and reliable rating provider, and that other provides might 'anchor' their own ratings based on the Sustainalytics rating. In addition, when looking at Panels B and C, we see that when a firm is initially covered by either two or three providers, the mean of *ESG Rating Disagreement* is lowest when Sustainalytics in included as a rating provider. Conversely, the mean of *ESG Rating Disagreement* in Panels B and C is always highest when Refinitiv ESG is involved as an initial rating provider.

Naturally, these descriptive statistics do not provide any reliable statistical proof of a possible herding effect in ESG rating, or a causal effect between the identity of the initial ESG rating provider(s) and the level of ESG rating disagreement, but it does show some interesting patterns in the incentives of rating providers to give firms a certain ESG rating.

Appendix C. Variable definitions

Variable name	Definition	Source
%Δ Carbon Emissions	The percentual change in annual carbon emissions of a firm, measured in equivalents of metric tons of CO_2.	Refinitiv ESG
ESG Rating Disagreement	The standard deviation of the standardized values of the different ESG ratings a firm receives at time *t*.	Refinitiv ESG/ Sustainalytics / MSCI IVA / MSCI KLD
Average ESG Rating	The mean of the standardized values of the different ESG ratings a firm receives at time *t*.	Refinitiv ESG / Sustainalytics / MSCI IVA / MSCI KLD
Institutional Holdings	The percentage of the outstanding shares of a firm that is being held by institutional investors at time *t*.	Thomson/Refinitiv
Big3 Holdings	The percentage of the outstanding shares of a firm that is being held by the Big 3 – BlackRock, Vanguard, and State Street – at time *t*.	Thomson/Refinitiv
Non-Big3 Holdings	The percentage of the outstanding shares of a firm that is being held by institutional investors other than the Big 3 at time *t*.	Thomson/Refinitiv
Size	The natural logarithm of the total assets of a firm (in US$) at time *t*.	Compustat
Log (Book-to-market)	The natural logarithm of the book-to-market ratio of a firm at time *t*. The book-to-market ratio is calculated at the ratio of the book value of equity over the market value of equity of a firm.	Compustat
ROA	The net income over the total assets of a firm at time *t*.	Compustat
Tobin's Q	The sum of a firm's market value of equity, long-term debt, and debt in current liabilities over the total assets of a firm at time *t*.	Compustat
Leverage	The sum of a firm's long-term debt and debt in current liabilities over the total assets at time *t*.	Compustat
PPE	A firm's plant, property, and equipment over the total assets at time *t*.	Compustat

References

Azar, J., Duro, M., Kadach, I., & Ormazabal, G. (2021). The big three and corporate carbon emissions around the world. *Journal of Financial Economics*, *142*(2), 674-696.

Bena, J., Ferreira, M. A., Matos, P., & Pires, P. (2017). Are foreign investors locusts? The long-term effects of foreign institutional ownership. *Journal of Financial Economics*, *126*(1), 122-146.

Berg, F., Koelbel, J. F., & Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, *26*(6), 1315-1344.

Billio, M., Costola, M., Hristova, I., Latino, C., & Pelizzon, L. (2021). Inside the ESG ratings:(Dis) agreement and performance. *Corporate Social Responsibility and Environmental Management*, *28*(5), 1426-1445.

Chatterji, A. K., Durand, R., Levine, D. I., & Touboul, S. (2016). Do ratings of firms converge? Implications for managers, investors and strategy researchers. *Strategic Management Journal*, *37*(8), 1597-1614.

Chatterji, A. K., Levine, D. I., & Toffel, M. W. (2009). How well do social ratings actually measure corporate social responsibility?. *Journal of Economics & Management Strategy*, *18*(1), 125-169.

Chatterji, A. K., & Toffel, M. W. (2010). How firms respond to being rated. *Strategic Management Journal*, *31*(9), 917-945.

Christensen, D. M., Serafeim, G., & Sikochi, A. (2022). Why is corporate virtue in the eye of the beholder? The case of ESG ratings. *The Accounting Review*, *97*(1), 147-175.

Clementino, E., & Perkins, R. (2021). How do companies respond to environmental, social and governance (ESG) ratings? Evidence from Italy. *Journal of Business Ethics*, *171*, 379-397.

Dorfleitner, G., Halbritter, G., & Nguyen, M. (2015). Measuring the level and risk of corporate responsibility–An empirical comparison of different ESG rating approaches. *Journal of Asset Management*, *16*, 450-466.

Dyck, A., Lins, K. V., Roth, L., & Wagner, H. F. (2019). Do institutional investors drive corporate social responsibility? International evidence. *Journal of financial economics*, *131*(3), 693-714.

Gibson Brandon, R., Krueger, P., & Schmidt, P. S. (2021). ESG rating disagreement and stock returns. *Financial Analysts Journal*, *77*(4), 104-127.

Hamilton, J. T. (1995). Pollution as news: Media and stock market reactions to the toxics release inventory data. *Journal of environmental economics and management*, *28*(1), 98-113.

Hoepner, A. G., Oikonomou, I., Sautner, Z., Starks, L. T., & Zhou, X. (2018). ESG shareholder engagement and downside risk.

Krueger, P., Sautner, Z., & Starks, L. T. (2020). The importance of climate risks for institutional investors. *The Review of Financial Studies*, *33*(3), 1067-1111.

Larcker, D. F., Pomorski, L., Tayan, B., & Watts, E. M. (2022). ESG ratings: A Compass without direction. *Rock Center for Corporate Governance at Stanford University Working Paper Forthcoming*.

Lins, K. V., Servaes, H., & Tamayo, A. (2017). Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis. *the Journal of Finance*, *72*(4), 1785-1824

Matos, P. (2020). ESG and responsible institutional investing around the world: A critical review.

PricewaterhouseCoopers. (n.d.). *ESG-focused institutional investment seen soaring 84% to US$33.9 trillion in 2026, making up 21.5% of assets under management: PwC report*.

PwC. https://www.pwc.com/gx/en/news-room/press-releases/2022/awm-revolution-2022-report.html

Principles for Responsible Investment (PRI). (2022). *Annual Report 2021-2022*. https://dwtyzx6upklss.cloudfront.net/Uploads/b/f/m/pri_annual_report_2022_689047.pdf

Serafeim, G., & Yoon, A. (2022). Stock price reactions to ESG news: The role of ESG ratings and disagreement. *Review of accounting studies*, 1-31.

Shanaev, S., & Ghimire, B. (2022). When ESG meets AAA: The effect of ESG rating changes on stock returns. *Finance Research Letters*, *46*, 102302.

Slager, R., & Chapple, W. (2016). Carrot and stick? The role of financial market intermediaries in corporate social performance. *Business & Society*, *55*(3), 398-426.

SustainAbility. (2018). *Rate the Raters 2018: Ratings Revisited*. https://www.sustainability.com/globalassets/sustainability.com/thinking/pdfs/sa-ratetheraters_ratings-revisited_march18.pdf

SustainAbility. (2023). *Rate the Raters 2023: ESG Ratings at a Crossroads*. https://www.sustainability.com/globalassets/sustainability.com/thinking/pdfs/2023/rate-the-raters-report-april-2023.pdf