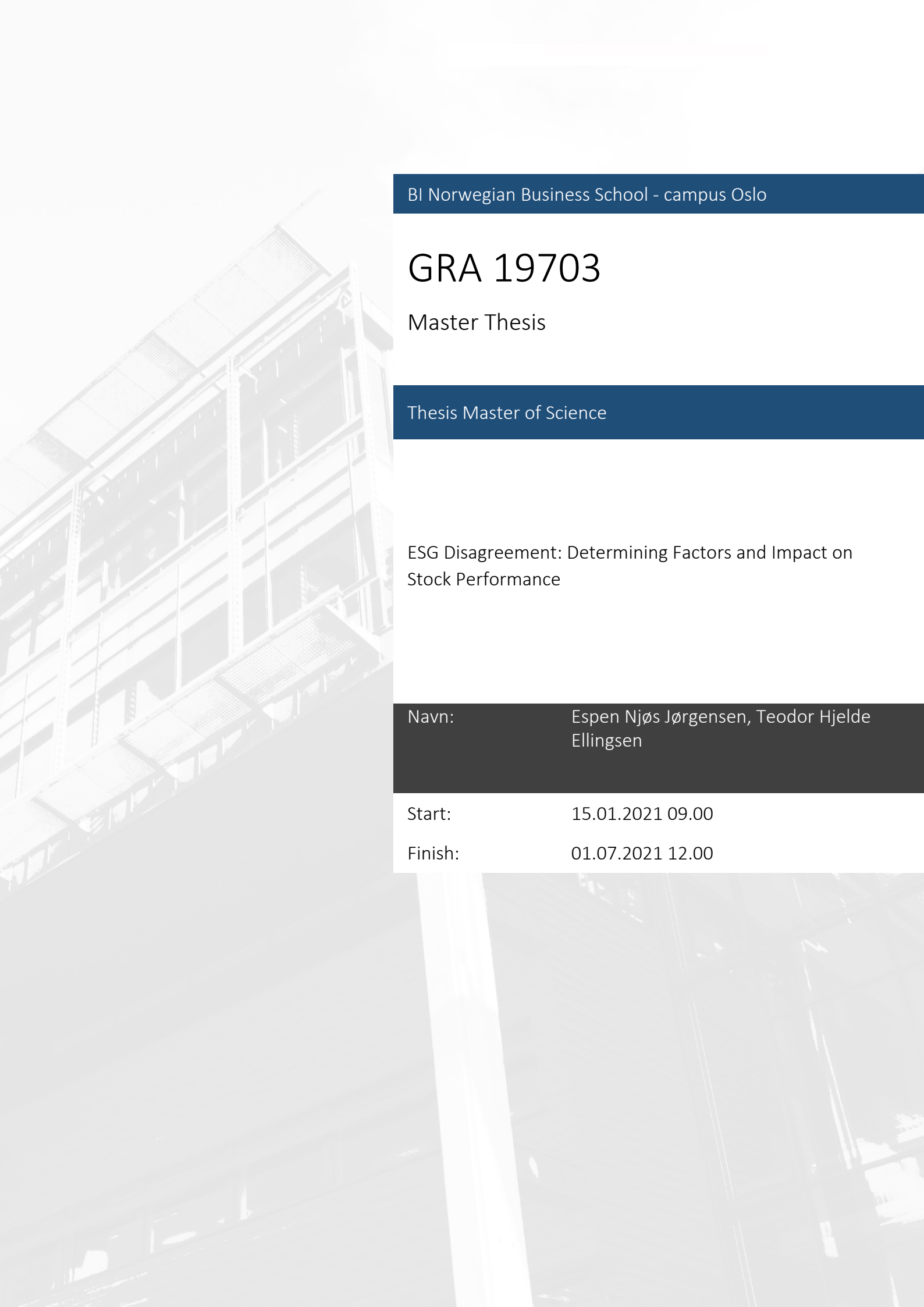

BI Norwegian Business School - campus Oslo

GRA 19703

Master Thesis

Thesis Master of Science

ESG Disagreement: Determining Factors and Impact on Stock Performance

Navn:	Espen Njøs Jørgensen, Teodor Hjelde Ellingsen

Start:	15.01.2021 09.00
Finish:	01.07.2021 12.00

ESG Disagreement: Determining Factors and Impact on Stock Performance

Master Thesis

By
Espen Njøs Jørgensen and Teodor Hjelde Ellingsen
MSc in Finance

Supervisor:
Dr. Bruno Gerard

Oslo, July 1st, 2021

Abstract

Using a sample of S&P 500 firms in the period 2010-2020. We will *first* study the rating variation of Environmental, Social, and Governance (ESG) ratings among some of the top prominent agencies. *Secondly*, we will study the characteristics of variation. *Thirdly*, we will examine the impact of the ESG rating variation on stock performance. We find that there is a large deviation between providers both cross-sectionally and over time. We discover that transparency can have an ambiguous effect on disagreement. While firm-size bias generally increases disagreement. Investors respond to the disagreement in overall and social score by undervaluing the stock, and by overvaluation in environmental disagreement. We conclude that ESG ratings continue to be a highly heterogeneous space, with clear challenges in objectivity, as well as increasing influence on the stock price.

Acknowledgments

We are thankful to our supervisor Dr. Bruno Gerard who was able to supervise us regardless of the challenges caused due to Covid-19, and for his continuous support in the process of producing this paper.

List of Contents

List of Tables and Figures ..... **iv**

Abbreviation and Definitions ..... **v**

1. Introduction and Motivation **1**

2. Background and Review of Literature **3**

2.1 ESG 3
2.2 ESG Rating Disagreement 4
2.3 Determining Factors of Disagreement 5
2.4 Heterogeneous Beliefs and Asset Pricing in Finance 7
2.5 Connecting Previous Research to This Thesis 9

3. Methodology and Hypothesis **9**

3.1 ESG Disagreement 10
3.1.1 Hypothesis 1 11
3.2 Determining Factors of Disagreement 11
3.2.1 Hypothesis 2 13
3.3 Impact of ESG Disagreement on Stock Price 14
3.3.1 Hypothesis 3 15

4. Data **16**

4.1 ESG Data 16
4.2 ESG Scoring 17
4.1 Variable and Data Description 20

5. Empirical Findings and Analysis **23**

5.1 Correlation Between Raters 23
5.2 Descriptive Statistics and Correlations on Sectors 25
5.3 Determining Factors of Disagreement 28
5.4 Effect on Stock Price 31
5.5 Limitations and Suggestions for Further Research 33

6. Summary and Conclusion **33**

Appendices **35**

List of Tables and Figures

Tables

Table 1:Rating overview 17
Table 2: Descriptive Statistics 18
Table 3: GICS Sector Sample Structure 19
Table 4: Dependent Variables 20
Table 5: Transparency Related Explanatory Variables 21
Table 6: Structure Related Explanatory Variables 22
Table 7: Value Related Explanatory Variables 22
Table 8: Correlation Results Between All Raters 24
Table 9: Descriptive Statistics on Sectors 26
Table 10: Determining Factors of Disagreement Results 29
Table 11: Impact of ESG Disagreement on Stock Price 32
Table 12: List of Variables 36
Table 13: List of ESG Databases 37
Table 14: Included Strengths and Concerns from MSCI 39
Table 15: Percentile Solution 40
Table 16: Spearman Correlation Results 41
Table 17: Pearson Correlation Results in Common Sample 42
Table 18: Pearson Correlation Results 2019/2015 44
Table 19: Pearson Correlation Results on GICS Sectors 44
Table 20: Ranked and Quartile Pearson Correlation Results on GICS Sectors 45
Table 21: Determinants of Disagreement with Range as Dependent Variable 47

Figures

Figure 1: Distribution of Sector Correlation in Quartiles 26
Figure 2: Sector Correlation Results 27

Abbreviation and Definitions

CRSP – Center for Research in Security Prices

CSR – Corporate Social Responsibility

ESG – Environmental, Social, Governance.

GICS – The Global Industry Classification Standard

ISIN – International Securities Identification Number

S&P – Standard & Poor

SRI – Socially Responsible Investing

WRDS - Wharton Research Data Services

List of Symbols

ρ – Parameter for correlation

D – Common parameter for disagreement, involving standard deviation and range

P – Parameter for stock price

$Range$ – Parameter for range on ESG disagreement

X – Vector for transparency explanatory variables

Y – Vector for structure explanatory variables

Z – Vector for value explanatory variables

i – Denotation for output by company

j – Denotation for output by time

s – Parameter for standard deviation on ESG disagreement

u – Parameter for the error term

α – Parameter for intercept

1. Introduction and Motivation

Increasingly over the last decades, companies have come under extensive pressure from regulators, investors, and other sources of influence to be responsible and sustainable. The market for sustainability to be measurable, have therefore been exponential. To tranquilize the demand, rating agencies created ESG scores. Rather than assessing a company by looking at its financial measures, ESG-scores are non-financial measures to measure the resilience to long-term, industry-material ESG risks (MSCI, 2020). Now increasingly more utilized in finance by researchers and practitioners. There as many as 98% of asset owner and investment manager signatories reported that they include ESG factors into their listed equity investments, while 91% reported ESG incorporation in fixed income and private markets (UNPRI, 2020). It turns out that ESG ratings increasingly shape the investment decisions of institutional investors representing more than $30 trillion in assets under management (GSIA, 2018). This is undoubtedly not due just to be good, evidence has shown that sustainability can also do good. High-scoring ESG firms met by adverse shocks to the market experienced greater financial solidity (Lins et al., 2017). As a result, institutional investors have become more aware of risks associated with ESG factors and expect corporations to manage issues involving ESG (Dyck et al., 2019). Subsequently, companies will have to defend their scores to investors, and even react to being rated poorly (Chatterji & Toffel 2009). Consequently, ESG agencies have over time become influential institutions that now researchers, investors, and companies rely on to make valid assessments of a firm's ESG performance. Now more than ever, with the capital ever-increasing allocation directed towards sustainable development, and only flowing through a few and large rating agencies (Avetisyan & Hockerts, 2017). ESG agencies have impromptu become the gatekeepers of green capital.

But the critical challenge for sustainable finance is in the unregulated and unsupervised nature of ESG ratings (ESMA, 2021). With no uniform requirements for how company's report ESG information and how ESG agencies measure it (Ho, 2020). The lack of binding definition creates implications as the ESG scores are

mainly based on the numbers the company itself discloses. Thus, the agencies themselves need to make independent assumptions, combined with different interpretations on scopes, measurement, and weighting factors (Rigobon et al. 2020). Increases the likelihood of disagreement. Compared to credit raters, sustainable raters display a surprising lack of agreement and can diverge from a correlation of 0.30 to 0.54 (Chatterji et al. 2016; Gibson et al. 2019). Evidently, this creates implications for researchers, practitioners, and companies that use ESG products. The variation in scores and measures can change the conclusion of the consumer differently based on which agency they have access to. This is the essence of what motivates this thesis. When acquiring a measurable instrument (e.g., ESG score, credit ratings, green bond certifications), it takes it as a given that the scoring is traceable, accurate, and free from biases. In other terms, objective. When the score then has a considerable deviation between agencies on the same firm, and that even some firms in polluting industries can obtain high environmental scores from one rating, and not on the other (Rigobon et al 2020). Accordingly, this questions the integrity of the ESG assessment, and consequently, the rating is just a subjective viewpoint of a third party. We will in this thesis measure the variation between some of the prominent agencies in the market to highlight the variation, and consequently the objectivity of ESG ratings. Simultaneously, we want to investigate whether the variation between raters can be determined by any specific biases. Additionally, as the relative influence of ESG agencies has increased, few have tested if disagreement in sustainability could also potentially become influential towards the stock price. Experimental studies have shown that sustainability can positively predict future performance (Hartzmark & Sussman, 2019). At the same time, standard asset pricing models assume that there is complete agreement among investors about the profitability distributions of future payoffs on assets (Fama & French, 2008). But when there are disagreements between raters, will that have an impact on the stock price? We build this on research about heterogeneous beliefs in the stock market. In Atmaz & Basak2018) innovation they find that stock price increases in belief dispersion while decreasing when the view is too optimistic. We, therefore, in line with their findings develop two competing hypotheses as to how ESG rating disagreement impacts stock returns - as in (Gibson et al., 2019).

2. Background and Review of Literature

We divide this section into four parts. We will begin by describing the origin and definitions of ESG, then review the world of ESG rating disagreement, and then various determinants of disagreement. Lastly connecting the ESG disagreement on heterogeneous beliefs in asset pricing.

2.1 ESG

The concepts we will discuss that originate from sustainable investments are Socially Responsible Investments (SRI), CSR, and ESG. By many authors, the terms are used synonymously with ESG, and by some to introduce theoretical models or concepts on company behavior. However, in this thesis, the concepts are primarily used to define a firm's aggregated score on its performance on sustainability measures. In the literature we cover, the faltering use of definitions is mainly because of when the research was initially published. However, along with the evolution towards today, the concepts have evolved and included more dimensions. Therefore, even when considering different investing approaches defined under branches of SRI or CSR. They will be interconnected, and ESG (investing) will be a fair term.

The origin of ESG was officially in 2005 (IFC, 2005) to regroup the three E, S, and G pillars into one concept. Based on information collected from public annual reports, NGO websites, sustainability reports, company websites, stock exchange filings, and reliable news sources, ESG scores are produced. The intermediaries that transform that information to ESG data are the rating agencies. The agencies use a distinct scoring methodology to identify risks and opportunities that are most material to that specific sector, industry, sub-industry, or firm. To assess how well a company manages its ESG risks and opportunities and subsequently rate them in form of an aggregate score (Escrig-Olmedo et al., 2019). This indicates how sustainable the company is concerning its relative performance to others.

2.2 ESG Rating Disagreement

With over 30 significant ESG agencies worldwide (KPMG, 2020), each having its distinct methodology. The essence of the challenges facing both investors and companies using ESG data is the absence of standards (Eccles et al., 2011). Opposed to financial data, ESG is relatively fresh. Thus, there is no agreement on which standards of reporting and approaches for agencies to utilize. Credit ratings are much more similar, however, expressed on an ordinal scale not precisely equivalent, but for the sake of illustration, Berg et al. (2020) find that credit rating agencies correlate 0.99, while Sindreu & Kent (2018) investigated the overlap between two distinct credit and ESG raters, which yielded a 0.82 R^2 for the credit agencies, and 0.13 R^2 for the ESG agencies. The absence of common theorization results in a complex process that leads to the disagreement between the raters themselves (Berg et al., 2020; Hawn et al., 2018). That it discredits the validity of ESG ratings itself, which researchers have debated critically (Chatterji et al., 2009; Bouten et al., 2017). Chatterji et al., (2016) studied whether the different raters gave valid assessments of the firm's social activities and performance. Using six well-established SRI raters, they found that the raters often diverged in their ratings of the same firm. They found that the raters between themselves had two factors that made them diverge. Firstly, they did not have a common theorization, on what raters assess and why it matters. Secondly, there was a low "commensurability". Meaning that what indicators are valid proxies for performance. On a high level of the categories, there was a broad agreement on the components of definitions. However, even adjusting for different definitions, the disagreement persisted. Implying that the ratings were lacking comparability in what they defined as valid proxies. Berg et al., (2020) paper deconstructed the disagreements somewhat differently. Among six ESG raters into three sources of divergence: different scope of categories, different measurement of categories, and different weights of categories. Their evidence showed that scope and measurement are the main drivers of disagreement. While weighting being less important. The same factors that Chatterji et al., (2016) found. Suggesting that the root of divergence comes from not on how they define it, but what attributes and proxies the raters base their ratings on. Consistent with Christensen et al., (2021) that

found that the raters disagreed more about the ESG outcome metrics than input metrics (policies) (Chatterji et al., 2015).

2.3 Determining Factors of Disagreement

As reviewed, ESG-raters differ in how they define and measure their assessment. We want to find if there are any determining factors of disagreement that can explain some of the variations between raters on firm-level characteristics. There has been a scarcity of research on determinants of disagreements.

We will start at the stem of what rater's base most of their assessment on—the firm's sustainability reports. The form and frequency that firms issue sustainability reports (including CSR and ESG data) differ significantly (Ioannou et al., 2015). In the 1990s, only 20 publicly listed companies issued reports that included ESG data. In 2014, the number had increased to nearly 6.000 (Cheng et al., 2014). According to G&A Institute, the constituents of the S&P 500 index that published sustainability reports were 75% in 2014, and in 2019 it grew towards 90%. As raters use their primary source in the assessment of a firm's ESG rating (Huber & Comstock, 2017). So, the importance of accuracy is essential for the ranking not to be subject to a consequential error. To weed out errors, independent assurance of reports has become a major practice in the world. In 2020 71% of the G250 companies acquired a third-party assurance of its sustainability information, a 25% increase over the last ten years (KPMG, 2017). However, there is little research on whether accounting firms improve the quality and accuracy of the reports. The research there is, find that quality of the reports is of reasonably acceptable quality (Zorio et al., 2013), if done by auditors, and not consultants (Pflugrath et al., 2011; Hodge et al., 2009),. There is uncertainty in the quality of the findings since a firm's reporting is not subject to a standardized reporting system, which leads to the firm to disclose as much or little of their ESG engagement as they want (Verrechia, 1983;Dye, 1985). [1] However, assurance providers help detect and prevent errors in sustainability reports more than

[1] This is based on voluntary disclosure theory, that under these circumstances a firm will only disclose as much as it benefits themselves. If it has many positive ESG engagements it will report extensively of them, and if not, the firm would report the bare minimum.

no assurance (Ballou et al., 2019). In 2018 the 61% of the constituents in the S&P 500 had no assurance, 36% had partial, and 3% had full assurance[2] (Welsh et al., 2018).

The lack of legally binding definition and comparability in sustainability reporting and ESG agencies makes it challenging to state how reliable and accurate they are. Giudice & Rigamonti (2020) investigated whether the ESG score changed after a scandal or misconduct. They found no significant ESG score adjustments after misconduct when the firm had their ESG reporting audited.[3] For companies that had not had their reports audited, they find significant worsening in ESG score. Auditing can therefore remove the information asymmetry between the company and the market. Minimizing the possibility of disclosing false information decreases the risk of the green-washing phenomenon (Marquis et al., 2016; Kim & Lyon, 2015), moreover reducing information costs. Implying that more accurate disclosures lessen the estimation risk that the different rating agencies must make.

Similarly, in the debt market, (Morgan, 2002) finds that more public disclosure reduces the dispersion of the opinion between credit raters. Intuitively, it also does that with the number of analysts that cover the stock. Because of the extra coverage, additional information will get to the public. Analysts also perceive high-scoring firms less pessimistically (Ioannou & Serafeim, 2015). That might indicate in absence of information, creating a simple rule in favor of the firm. There is also a conflict of interests in this section. As analysts are often directly or indirectly linked by parent company or subsidiary to the raters themselves. Tang et al. (2021) find that firms held by the same owners as the raters receive higher ESG ratings. The conflict of interest materializes, as these firms have more future negative ESG incidents. We will not investigate direct ownership. But as ownership matters, we will test whether institutional ownership affects disagreement. These results in what we call measures

[2] Partial assurance is when the assurance firm only cover some of the metrics and performance goals reported. 90% of the partial assurance group had acquired services to control environmental metrics, in most cases greenhouse gas emissions (IRRC Institute & Si2, 2018, page 29.).
[3] Partial assurance is sufficient to be labeled ESG audited.

for firm transparency. [4] Higher firm-level transparency is significantly related to the stock's liquidity, that research has found to affect the firm valuation and the cross-section of returns (Lang et al., 2012). Moreover, evidence has shown that the bigger the firm is, the greater score (Drempetic et al., 2019). Because larger firms have additional public pressure (Udayasankar, 2008), they thus must intensify their reporting and expenditures with firm size (Adams et al., 1998; Chauhan, 2014). Though, this evidence shows contradictory and ambiguous results when researching between disclosure and firm value (for positive see: e.g., Clarkson et al., 2013; Gamerschlag et al., 2011; Fatemi et al., 2018) (for negative see: e.g., de Villiers et al., 2011;Ho & Taylor, 2007). Most of the literature focuses on the effect of transparency and intensification of disclosures on one ESG rater.

On the other hand, Christensen et al., (2021) posit that the amount of disclosures explains the disagreement between multiple raters. Owing to that, an ESG-analyst may differ in how to interpret a specific metric. In the absence of a particular metric, a simple rule can be created. Such that raters are more likely to agree on the given rating. They find that ESG disclosure generally worsens ESG rating disagreement, or when there is much publicly available information, the raters disagree the most. Hahn & Kühnen, (2013) posits that firm size significantly affects the quality of sustainability reporting. It is reinforcing Christensen et al. (2020) findings that more disclosure increases the variation in ESG ratings.

2.4 Heterogeneous Beliefs and Asset Pricing in Finance

Two key characteristics of economic agents are their beliefs and preferences. Standard asset pricing models share the equivalent homothetic expectation: that all investors are assumed to have identical measures of future return and probability distribution from all securities (Sharpe, 1964). These assumptions are unrealistic. Forecasts are difficult to make, and investors have different preferences and biases. In the light of this paper, SRI is one of those biases that investors disagree on (Geczy et

[4] That is, Audited ESG reports, credit rating, number of analysts that cover the company, and institutional ownership. We will later in section 4 discuss what outcome we expect from each measure.

al., 2005). Heterogeneous beliefs are now well-recognized in the financial economics literature. There are few reasons to believe that agents agree on the true probability distribution of any observables and might even disagree more with less tangible types of non-fundamentals (Basak, 2000). An essential feature in financial markets is heterogeneity in investors' beliefs, which plays a vital role in forming security prices, dynamics, and the volume of trades amongst market participants (Basak, 2005).

Miller (1977) was one of the first to argue that uncertainty about future price levels implied that agents disagree about their point forecasts. He claimed that in practice, increased uncertainty and risk implied increased divergence of opinion. Miller illustrates this through a supply and demand curve, where stocks are limited to the outstanding amount. If the poorly informed minority has the funds to absorb the entire supply of the stock, it would be above the mean evaluation. As long as the minority can absorb the security, an increase in distortion will increase the market price. So, only when there is no disagreement about the security will the price be the average evaluation. A substantial amount of well-informed investors will prevent there being undervalued securities. But it may be a badly informed minority that has sufficient funds to make an overvalued security. Therefore, the disagreement would only reflect the optimistic (and badly informed). However, this is only possible when pessimistic are short sales constrained. Beber et al., (2010) underpinned Millers' reasoning that disagreement about future currency returns significantly impacts currency risk-premia. Chen & Epstein (2002) based their model on Millers' (1977) work, where ambiguity aversion is admissible and short-sales constraints. They find a negative relation between disagreement and stock mean return. This relation might only be valid under specific characteristics (short-sale constrained, small, illiquid, or worst-rated). Pavlova & Rigobon (2007) find empirical evidence for a model with stock prices and exchange rates with heterogeneous beliefs. Giambona et al., (2018) study of the equity risk premiums confirms that the dispersion of risk perception is relatively high with standard deviations of 2.93%.

In Atmaz & Basak (2018) innovation, they model a tractable and straightforward pure-exchange security market economy in continuous time. The economy is of

investors with heterogeneous beliefs and standard CRRA preferences that consume at a single date. They find that belief dispersion leads to higher stock volatility, trading volume, and a positive relation between the two. In addition, they show that the stock price increases (decreases), and its mean return decrease in belief dispersion when the view is optimistic (pessimistic). When the view on the stock is sufficiently optimistic, the effect dominates, causing a negative relation between disagreement and stock mean return. In their innovation, they construct measures of belief dispersion by two sufficient measures, the average bias and dispersion of belief. Consistent with their efforts, we will measure the disagreement as to the cross-sectional standard deviation and range of the ESG agency's disagreement. The measures of disagreement form the two competing hypotheses we base on the biases we want to test. First, a risk bias that implies that higher divergence leads to higher risk will result in higher returns as there is more uncertainty. Secondly, an optimism bias that higher divergence will result in lower returns due to the investor's belief that the companies actual ESG performance is captured by the most optimistic ESG rating, which accumulates to overvaluation.

2.5 Connecting Previous Research to This Thesis

This thesis will be based on Gibson et al., (2019) paper. We will measure the variation between prominent raters, however with a different set of agencies, time samples. We will track the determinants of disagreement based on the transparency factors we discussed, in addition to some structure and value-related factors that have determined similar biased relationships. Next, we will construct the test of disagreement on stock returns according to Atmas & Basak (2018) findings. If our findings are positive this creates implications for the consumers of ESG ratings, and even more for those who rely on only one raters' assessment.

3. Methodology and Hypothesis

We want to test the objectivity of ESG ratings, determinants, and their effect on stock performance. We deconstruct the research question into three parts. First, we will examine the disagreement between the raters themselves. Centered on basic

descriptive statistics and correlation, we will measure the variation between the providers in overall rating, and for each pillar, as well as on sector-level specifics. The disagreement will unfold itself into the two measurement parameters we use for the next two segments, standard deviation, and range. Second, we want to test what firm-level characteristics there are for disagreement. This can explain whether there are any significant biases or features which come into play in scoring. Third, does the disagreement influence stock performance.

3.1 ESG Disagreement

We will differentiate the correlation into an overall score and each subcategory/pillar E, S, and G. If there is not a number at the specific time and/or firm, the agency's rating will be canceled out. We will have a pairwise correlation, which will be our focus, as it maximizes the sample observations. We also present a common sample but should be given little significance as it is only present in three years. The results from the correlations test yield a scale of 1 to -1, respectively, from a perfect correlation to a perfect negative correlation. Suppose we yield a correlation (the p-value) of 0 (or close to). In that case, we find that there is enough statistical data to uphold the correlation between the variables (hence, significant).

We apply both Spearman rank and Pearson correlation tests to investigate whether the selected issuer's ratings are correlated with the others on our company sample. The difference in tests is that Spearman assesses the relationships between two variables using a monotonic function, while Pearson assesses them as a linear function. If both coefficients are +1 this implies for Pearson that when one variable increases, the other variable does consistently the same. However, if the one variable increases and the other one does as well but not the equivalent amount. Then Pearson coefficient would be less than +1. While the Spearman coefficient would still be +1. Spearman also prefers ordinal scales, while Pearson favors interval scales. Spearman can however work with both. We will present the Spearman correlation but will focus on Pearson correlation, consistent with Berg et al., 2020). However, Berg et al. (2020) does not use the initial ratings, but a transformed normalized rank. We will use both to see if the correlations are monotonic but not linear. If the Spearman correlations are

substantially larger than Pearson's, there is a monotonic but not a linear relationship between the variables. If the opposite, then this indicates that some extreme data points exert a strong influence on the calculations, and transformation is needed.

3.1.1 Hypothesis 1

Beginning with each pillar, we conjecture that the variation between raters will unfold the least in the E score. Because the environmental factor is mainly quantifiable and therefore the most objectively accurate score. Next, we expect governance as it is partly based on hard measures and partly soft data. Last, we expect the social score as it is mainly based on soft data, and therefore subject to individual subjective conclusions. Hence, we expect variations to be most significant in the social score. The overall score is based on an aggregate function on the pillars and would be expected to be somewhere in the center.

$$H_0 = \rho_E > \rho_G > \rho_S$$
$$H_A \neq \rho_E > \rho_G > \rho_S$$

3.2 Determining Factors of Disagreement

We use ESG data from six different agencies, all with four different scores, based on three different sources of extraction. Resulting in varied data at a time- and firm-level. Hence, the dataset at disposal is unbalanced. Therefore, it is irregular that every firm has a score at any given time. Nevertheless, having data structured in a panel is helpful for the goal of this thesis. The advantage is that we can control for unobservable variables across firms and time (Stock & Watson, 2015). In contrast to time-series data and cross-sectional data, panel data will be significantly more helpful in explaining the variation between raters, but also explain the characteristics that are more significant to draw robust interpretations from and are more significant than others on explaining the distortion between firms. To construct this dataset of firm-level variables and financial data, we use Compustat for accounting data and CRSP for financial data, with some exceptions that we will point out along the way. The characteristics we are going to test are based on transparency, structure, and value

factors. Transparency factors involve audit verification (AV), credit rating (CR), number of analysts covering the stock (NOA), and institutional ownership (IO). Structure-related factors and include tangibility (TAN), current ratio (CR), leverage (LEV), and capital expenditures (CE). Following, we will test Value related factors: market capitalization (MC), volume (LIQ), volatility (VOL), book to market value (BM), gross profitability of the firm (GP).

We will use a pooled panel regression with fixed effects. The fixed effects are at the time- and sector-level. This to control for unobserved heterogeneity in our sample that is constant over time. This unobserved heterogeneity will translate into the intercept coefficient. The characteristics will be the independent variables in our regression. The dependent variable (disagreement) will be tested in two ways: standard deviation and range. By standard deviation, we imply the amount of variation of a set of ratings. By range the distance between the top and bottom rating, implying that we neglect the middle scores. Both measures will be obtainable at any given time for any given firm. Respectively, the standard deviation and range are given by:

$$s_{i,j} = \sqrt{\frac{\sum\left(x_{i,j} - \bar{x}_{i,j}\right)^2}{n-1}} \qquad Range_{i,j} = MAX_{i,j} - MIN_{i,j}$$

To calculate the overall mean and standard deviation of the different ratings, thus different observations, we need to address some complications. For the measures to work, we need to calculate the overall number of observations, the overall mean, the standard deviation of the multiple panels – with their belonging given number of observations. Then the mean values and standard deviations for each panel. As there are ratings in the different times that are unbalanced than others, we apply clustered standard errors at both time and firm-level. If there is only one available rating from one agency at a given time the measures will be canceled for that given time. The regression equation is:

$$D_{i,j} = a_i + \beta_1 X(Transparency)_{i,t} + \beta_2 Y(Strucure)_{i,j} + \beta_3 Z(Value)_{i,j} + u_{i,j}$$

$$Where\ i\ is\ companies, i = 1 \dots, N\ and$$
$$j = 2010, \dots, 2020.\ D\ is\ \{s_{i,j}, Range_{i,j}\}.$$
$$X, Y, Z\ are\ vectors\ for\ the\ explanatory\ variables.$$

(1)

3.2.1 Hypothesis 2

We will divide this one research question into three parts, one for each vector of explanatory variables, but into two competing hypotheses.

It is intuitive to believe that transparency will remove the need to make subjective assessments. So that when more analysts are covering, the information will be available. Same with credit rating. As well institutional owners, demand more transparency and have shown that have a significant impact on ESG performance (Dyck et al., 2019). However, Kotsantonis & Serafeim (2019) found that higher disclosure yields higher disagreement. It is also intuitive to believe that raters would disagree less on tangible, capital intensive companies have higher emissions and thus is more measurable. Though, tangible assets are now scarce on the S&P 500 index (Elsten & Hill, 2020). While huge profitable firms have will likely have slack to invest in further pro-sustainable engagements. Drempetic et al., (2019) find that the MC has a significant influence on ESG scoring, i.e., that raters have firm size bias. Therefore, we will postulate two hypotheses on the determinants of disagreement.

The two competing hypotheses:

1. *The intuitive hypothesis*

Based on that more transparent, capital intensive, large companies result to lower disagreement, as raters have a more easily measurable ESG assessment.

H_0: We conjecture that higher transparency, structure, and value-related factors will have a negative relation to disagreement.

2. *The fallacy hypothesis.*

Evidence from Kotsantonis & Serafeim (2019) shows that higher disclosure yields higher disagreement. We posit the same but with our transparency factors. More measurements regarding structural factors will increase the discrepancies in measurements and scopes subsequently will result in higher disagreement (Rigobon et al. 2019). The larger the firm, the more complex to assess for analysts (Sadka & Scherbina, 2007).

H_A: We conjecture that higher transparency, structure, and value-related factors will have a positive relation to the disagreement

3.3 Impact of ESG Disagreement on Stock Price

We use the same approach as in firm-level characteristics, with pooled panel regression, clustered standard errors at the firm, and time level, including fixed effects. In this regression, however, the dependent variable will be the stock return, and the independent variable will be the disagreement. The independent variable (disagreement) will be tested in the same two ways: standard deviation and range. Any measures will be obtainable at any given time for any given firm. In addition, we include control measurements for characteristics that have a significant effect on stock prices, this does not imply that we use all the same control characteristics as above, but the value-related factors that have been controlled for in the elements of the disagreement section. We include only factors that have been proved to have a significant factor on the stock price. The results are presented in one panel for each rating, including all control measurements and fixed effects. We will first test both our competing hypotheses on disagreements' impact on the stock price.

$$P_{i,j} = a_i + \beta_1 D_{i,j} + \beta_2 Z(Value)_{i,j} + u_{i,j}$$

$$Where\ P\ is\ the\ stock\ price, Z\ is\ a\ Vector\ of\ control\ variables\ that\ have\ a\ significant\ relationship\ with\ the\ price.$$

(2)

Based on this, we want to see how the individual rating disagreement affects stock returns. Both rely on the literature on heterogeneous beliefs in asset pricing in an ESG context. The first hypothesis posits that higher divergence in ESG ratings should lead to higher stock returns due to the variability in ratings resulting from higher risk in the company (Atmaz & Basak, 2018). While the latter builds on optimism bias that higher divergence regarded firms result in lower future stock returns. Because investors believe that a company's true ESG performance is captured by the most optimistic ESG rating - which accumulates to overvaluation, and thus less future return in the future (Atmaz & Basak, 2018; Gibson et al., 2019). We include control measurements that have a significant impact on stock price (Gibson et al. 2019; Banz, 1981; Lang et al., 2012; Ang et al., 2006; Fama & French, 1995; Novy-Marx, 2013).

3.3.1 Hypothesis 3

The two competing hypotheses:

1. *The risk-based hypothesis*
2. *The optimism-bias hypothesis.*

To test the Risk-based hypothesis, we will run the following hypothesis:
H_0: Companies with higher disagreement in (ESG, E, S, G) ratings will experience a positive relationship between rating disagreement and future stock returns.

To test the competing Optimism-Based hypothesis, we will run the following hypothesis:

H_A: Companies with higher disagreement in (ESG, E, S, G) ratings will *NOT* experience a positive relationship between rating disagreement and future stock returns

Alternatively:

H_A.: Companies with higher divergence in (ESG, E, S, G) ratings will experience a negative relationship between rating disagreement and future stock returns

4. Data

4.1 ESG Data

We wanted to include as many as feasible raters of significant size in our sample. We have been granted access to more agencies that we have decided to include in our data set, for two reasons: (1) The access has not been sufficient, (2) Too few data points. We will use ratings from five agencies: Bloomberg, Refinitiv, S&P Global, Sustainalytics, and MSCI KLD. Some available through the school's database and a few that have been generous to provide access for the sake of this paper. We use the Eikon Terminal, Bloomberg Terminal, and WRDS for the extraction of ESG data. We decide on a sample of companies from the S&P 500 index, that yielded a notably more respective data sample regarding the overall sample of firms and belonging historical coverage compared to other regions and indexes. Some of the raters have less coverage, and some differences in the frequency they release updated ratings. Refinitiv and MSCI KLD are yearly, S&P Global is quarterly. While Bloomberg and Sustainalytics provide monthly data. However, ratings rarely vary much within a year, some even longer. We will base the scores on the annual scores. There are however providers we have limitations on the coverage. Sustainalytics and S&P Global starting from respectively, 2013 and 2016. While MSCI KLD ends in 2018. This implies that Sustainalytics and MSCI KLD will only have 2 years in a common sample. This will only have implications on the first objective of the thesis in measuring the variation between raters. Which we will point out ongoing. In measuring the correlations between providers this will not be solved as a common sample. The correlations will be tested pairwise, to best possible capture their variation. This means omitting them pairwise in their non-reported years and added back when they have coverage. Regarding capturing the best statistical significance between raters. This applies to Bloomberg and Refinitiv with a common sample of 10 years, accompanied by both relationships with MSCI of 8 years. S&P Global and Sustainalytics will have a short relation with MSCI that needs to be emphasized in the

section of variation between providers. In measuring the disagreement on firm characteristics this will limit the width of agencies of disagreement. This will be most present in 2010-2014 with only three agencies. We decide on including every rater we have access to. This opposes the idea of creating the most homogeneous sample possible, and in many other projects, this would not be possible. We do this to have a sample that has considerably more data points than it would not.

4.2 ESG Scoring

The different raters we utilize provide ESG scores from offers company ratings based on similar rating scales illustrated in Table 1.

Rating agency	Rating scale	Pillars	Source
Bloomberg	0 - 100	ESG, E, S, G	Bloomberg Terminal
Refinitiv	0 - 100	ESG, E, S, G	Eikon Terminal
S&P Global	0 - 100	ESG, E, S, G	Bloomberg Terminal
Sustainalytics	0 - 100	ESG, E, S, G	Bloomberg Terminal
MSCI KLD	-1 - +1	ESG, E, S, G*	WRDS

Table 1:Rating overview

Table 1 displays the rating agencies we utilize, their rating scale, the pillars we use, and the source we get them from. *MSCI has a different scoring scale, there is however not a recognized agreement in the research literature on how to weigh the pillars, some even argue it is not possible (Mitchell et al., 1997). We decided on following the procedure in most academic studies, summing all strengths and subtracting all weaknesses (Lins et al., 2017). By including all strengths and concerns except categories related to alcohol, military, firearms, gambling, nuclear, and tobacco as in Gibson et al. (2019). Scoring will be based on Environmental, Social (community, human rights, employee relations, diversity), and Corporate Governance. Environmental have combined feasible strengths and concerns of 23, social (including all related categories) have 33, corporate governance has 11. To get an equal aggregate weighing function we apply the following weighing factors: 34%,

24%, and 42% respectively. Similarly, as done by Ioannou & Serafeim (2015), and in Waddock & Graves (1997), we utilize a rescaling solution of equal weight to the relevant issue. This can be viewed in appendices A.3. This modification will however slightly influence MSCI KLD cross-sectional statistical context with the other rating agencies. Since the other providers have the same scoring system, we decide not to apply a conversion to percentiles which have been done in other studies to avoid fiddling with the raw data and increasing the risk of errors.

4.3 Data Sample

The original sample selected holds 505 constituents, which will vary for every vendor regarding what they cover at each specific time. We find that each vendor differs in which members they hold overall, and on only particular dates, we have a common sample.

	Bloomberg	**Refinitiv**	**S&P Global**	**Sustainalytics**	**MSCI**
Coverage	2020 - 2010	2020 - 2010	2020 - 2016	2020 - 2013	2018 - 2010
Firms	487	467	472	438	473
N	5360	5133	2358	3503	4258
Max	78.008	95.211	100	100	81.952
Min	0	2.305	0	0	21.577
Mean	33.415	53.349	46.765	51.593	44.970
Median	32.645	54.664	44.000	52.941	44.186
Std. Dev.	15.589	19.142	27.646	25.787	7.031

Table 2: Descriptive Statistics

Table 2 shows that the overall average of firms they have on total score in row firms. Sustainalytics has the overall least average coverage of 438 firms. In contrast, Bloomberg has the highest with an overall average of 487. Nevertheless, all vendors have a coverage rate of over 86%. The table also shows the apparent differences in the dataset, where S&P Global and Sustainalytics use the entire length of the scale in

their assessment of the company sample. To avoid mismatching when combining the distinct ESG scores from each agency, we apply ticker, ISIN, and company name as a common identifier. When there was a mismatch, we manually corrected the problem.

We will also test whether some sectors are more prone to disagreement than others. The structure we apply is the Global Industry Classification System (GICS) sectors, which separates the sample into 11 distinct sectors.

GICS Sector Name	Rank	Observations	% Of Total Observations
Information Technology	1	825	14.703%
Industrials	2	814	14.507%
Financials	3	726	12.939%
Health Care	4	715	12.743%
Consumer Discretionary	5	693	12.351%
Consumer Staples	6	363	6.469%
Real Estate	7	319	5.685%
Utilities	8	308	5.489%
Materials	9	308	5.489%
Communication Services	10	286	5.097%
Energy	11	253	4.509%

Table 2: GICS Sector Sample Structure

Table 3 shows the structure of the data sample in terms of GICS sectors. The sector weights make up a significant separation between the top 5 and the rest, where Information Technology (IT), Industrials, Financials, Consumer Discretionary, and Health Care accounts for 67% of the observations. It is important to consider when deciding the significance of the variation between sectors. Many eliminate Financials because many of the environmental and social policies are not likely or applicable to them, the policies apply to the companies in their loan or investment portfolio and are more likely to be significant for their performance (Eccles and Serafeim 2013). This will not apply in this thesis as we use sector fixed effects.

4.1 Variable and Data Description

GICS is an industry taxonomy developed by MSCI and Standard & Poor (S&P) in 1999. The GICS is only competing with Industry Classification Benchmark (ICB). The only difference in sector-level between the two is that GICS labels consumer businesses as consumer discretionary and consumer staples, while ICB labels them as a provider of goods, and providers of services. We apply GICS simply for unity with index providers.

Variable	Description	Database
Standard Deviation on ESG, E, S, G (SDEV)	The standard deviation on every available rating	Calculated*
Range on ESG, E, S, G	Range between the top and lowest rating	Calculated*
Price (PRICE)	Stock price on the first available day in the specific year minus the first available share price of the next annum.	CRSP

Table 3: Dependent Variables

The dependent variables we will use in this thesis are displayed in Table 4. On the standard deviation and range, there must be a minimum of two available ratings from distinct agencies, otherwise, it will be removed. Raters generally post their in-depth ratings annually at the start of the year (Huber & Comstock, 2017).To test the disagreement on the future stock price we take the first available stock price of the annual year minus the next beginning price of the next annum.

Variable	Description	Database
Audit Verification (AV)	AV is a CSR external audit verification of the environmental measures published in the firm's sustainability	Eikon Terminal
Credit Rating (CR)	S&P domestic long term issuer credit rating	WRDS
Number of Analysts (NOA)	Number of analysts covering the specific stock	Compustat/CRST
Institutional Ownership (IO)	Institutional ownership in % of the shares of a firm	Bloomberg Terminal

Table 4: Transparency Related Explanatory Variables

The financial indicators can explain the disagreements between raters through economic context. The first deals with transparency factors, displayed in Table 5. AV and IO are the only two variables that are extracted outside of Compustat and CRSP. AV from Eikon terminal, and IO from the Bloomberg terminal. AV is a CSR external audit verification of the environmental measures published in the firm's sustainability reports. The scores will be binary for AV and CR, either 1 (TRUE) or 0 (FALSE). IO is the percent of holdings owned by financial institutions. This number can in some instances be above 100 either as a result of slow updates or short selling. We divide it by 100, so in our case, it can be above 1.

Structure-related factors and involve:

Variable	Description	Database
Tangibility (TAN)	Tangible assets (PP&E) divided on the total assets	Compustat
Current Ratio (CUR)	Current assets divided on current liabilities	Compustat

Leverage (LEV)	Total long-term debt plus current debt divided on the total assets	Compustat
Capital Expenditure (CE)	Capital Expenditure divided on PP&E	Compustat

Table 5: Structure Related Explanatory Variables

In table 6 we find the structure-related explanatory factors, these are ratio measurements of the structure of the firm and will not be transformed. TAN and CUR are based on Gibson et al. 2019, while LEV and CE are from Christensen et al., 2021).

The Value-related factors measurements include:

Variable	**Description**	**Database**
Market Capitalization (MC)	Average share price multiplied with the average amount of shares outstanding in the same fiscal year	CRSP
Liquidity (LIQ)	The volume of shares traded in the fiscal year divided into average shares outstanding	CRSP
The Book to Market Ratio (BM)	Total book value of assets divided on the market capitalization	Compustat/CRSP
Gross Profitability (GP)	Revenues minus direct costs divided on the total assets	Compustat

Table 6: Value Related Explanatory Variables

In table 7 we find the structure-related explanatory factors. All variables except for BM are transformed into log. The control measurements include the value-related factors including volatility (VOLA) which is measured by the standard deviation of the share price in the last 260 days, sourced from CRSP.

5. Empirical Findings and Analysis

5.1 Correlation Between Raters

In Table 8, we set forth the correlations of ratings between ESG agencies.
The table is divided into two parts. The first part above the double lines is designated to the pairwise correlations. Meaning that we maximize the possible observations. The first four columns are for the Pearson correlation, the next four for Spearman correlations, both in the respective order, Bloomberg (BB), Refinitv (Ref.), S&P Global (S&PG), and Sustainalytics. Each panel has its score, total, and for each pillar. With the cross-sectional correlations with Refinitv, S&P Global, Sustainalytics, and MSCI at last for every panel. The last row in every panel is for the average mean correlation of each respective cross-correlation. The last row in each panel has the average correlation. The second part, under the double line, is for the common sample and is structured the same way except that the correlation panels are not structured in chronological order. This implies a Pearson correlation of when the firms have the same companies at the same time period. This regards the period 2016-2018 and should therefore give the least significance.

	Pearson correlations - Pairwise			
	BB	Ref.	S&PG	Sus.
ESG	(1)	(2)	(3)	(4)
Refinitiv	0.635			
S&P Global	0.607	0.535		
Sustainalytics	0.599	0.583	0.661	
MSCI	0.538	0.451	0.446	0.488

Average		**0.554**		
Environmental				
Refinitiv	0.583			
S&P Global	0.585	0.554		
Sustainalytics	0.497	0.498	0.570	
MSCI	0.478	0.446	0.479	0.407
Average		**0.510**		
Social				
Refinitiv	0.504			
S&P Global	0.555	0.489		
Sustainalytics	0.436	0.348	0.460	
MSCI	0.280	0.276	0.264	0.223
Average		**0.384**		
Governance				
Refinitiv	0.356			
S&P Global	0.488	0.226		
Sustainalytics	0.423	0.261	0.549	
MSCI	0.310	0.091	-0.022	-0.057
Average		**0.262**		

Table 7: Correlation Results Between All Raters

Starting in the pairwise, in other words, the full sample we see that all relationships are generally higher in the spearman correlations, which implies that the relationships are monotonic but not linear, as they are not substantially higher, transformation is not needed. Onwards in the thesis, we will only use Pearson's correlation to have consistency and comparability with the related research. Starting with total rating (ESG) we find the samples' strongest relationship between Sustainalytics and S&P Global with a correlation of 0.661. In this panel, two relationships are significant at 1% level, Bloomberg with Refinitiv and Sustainalytics. The overall rating also has the highest average correlation of 0.554 in our pairwise Pearson correlations. This is a chronologically trend downwards the panel that average correlation weakens by each rating. In the environmental panel, the average correlation slightly weakens to 0.510.

We conjectured that the environmental relationship should be higher due to its quantitative nature. ESG is likely stronger due to aggregation discrepancies in the overall assessment and weighting of the specific pillars. Also, in this panel, Bloomberg and Refinitv have a significant relationship on a 1% level. In the subsequent panel, we see that as the adjustment from a quantitative pillar towards a more qualitative pillar the average relationship weakens drastically to 0.384. In the governance panel, this effect weakens beyond the social to an average of 0.262. In this column, this swing is mainly due to the relationships regarding Refinitiv and MSCI. We see throughout the results that MSCI generally has the worst cross-correlation in every panel except on two instances, both connected to Bloomberg in respectively overall and governance panel. These results are consistent with Rigobon et al. (2019), Gibson et al. (2020), and Chatterji et al. (2016). As a sidenote which should not be given much significance because of its short time period is the common sample. Regardless of the short span, the correlations are generally, apart from a few exceptions quite similar as in the full sample. Suggesting some of the relationships does not change over time. We tested this on two phases outside of the common sample. Once in 2015 and once in 2019. This involves that S&P Global and MSCI will be removed, correspondingly. This was not the case in both models.

5.2 Descriptive Statistics and Correlations on Sectors

In essence, an ESG rating is a sector relative score assigned to individual firms. Now confirmed there is disagreement between the agencies, we want to investigate that the ESG disagreement correlates with the different observable sector characteristics.

Mean	**ESG**	**E**	**S**	**G**
Information Technology	43.918	41.370	42.528	53.725
Industrials	41.797	38.570	41.074	52.763
Financials	42.459	36.107	39.952	55.237
Health Care	45.000	41.996	42.797	56.524
Consumer Discretionary	42.168	38.936	39.973	53.351
Consumer Staples	51.765	49.979	49.719	60.012

Real Estate	44.348	40.584	43.357	57.213
Utilities	48.255	45.887	46.857	59.222
Materials	46.938	45.779	43.898	56.712
Communication Services	36.246	32.130	33.338	48.437
Energy	44.660	41.206	41.976	55.601
Average	**43.942**	**40.630**	**42.061**	**55.002**

Table 8: Descriptive Statistics on Sectors

In Table 9, we find the overall and each pillar's average rating in all GICS Sectors. There are relationships across ratings, that Consumer Staples and Utilities score higher on average, and that communication services scores significantly the lowest on all pillars relative to the other sectors. Excluding overall rating, as it is an aggregation of the other pillars. We find that the average rating across industries is the lowest for environmental, and slightly higher for social, and significantly higher for governance.

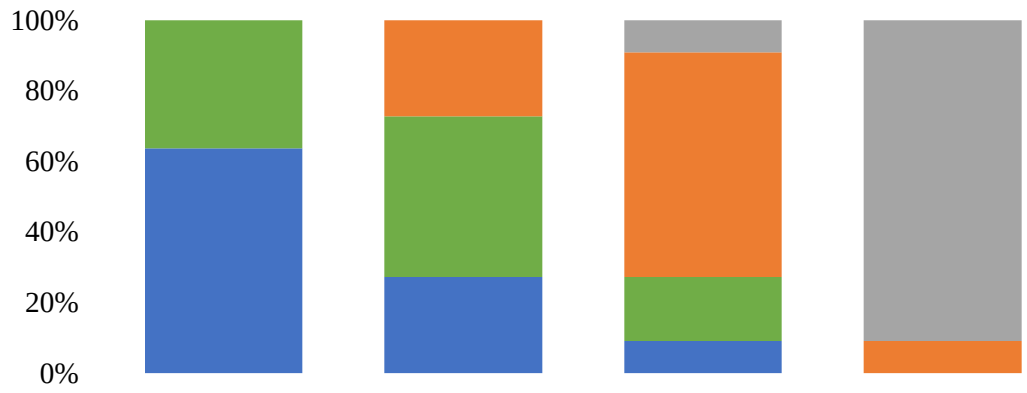

Figure 1: Distribution of Sector Correlation in Quartiles

Before assessing the sector variation across ratings, it is important to review that the distribution of each specific rating between all industries is very skewed as in the results above. As illustrated in Figure 1, dividing all scores into quartiles including all sectors makes 44 scores and 11 scores in each quartile. The first from left is the first quartile with the highest correlations. Blue is ESG, green is E, orange is S, and grey is G. In the first quartile ESG accounts for 7/11 scores, E for 4/11. The second quartile ESG and S has 3/11, while E has 5/11. Third quartile ESG and G has 1/11 each, E has

2/11, while S has 7/11. The last and most disagreeing quartile S has 1/11, while G has 10/11. The distribution of correlations across industries agencies skew more to agree on industries when scoring ESG and E than on S and G, especially visible in G as 91% of the G correlations are placed in the fourth quartile.

In Figure 2, we have the results of specific GICS sector correlations. On the vertical axis are the GICS Sector Names, which are sorted based on their ranked number of observations. The average correlations are on the horizontal axis of the charts. The stipulated line is the sector normalized average correlations across all sectors. Starting with blue is the total score, the concluding is in the order of E, S, and G. A higher correlation here implies that there is more agreement, opposed to little where agencies disagree more. The vertical lines in the same colors are the average correlations across all industries.

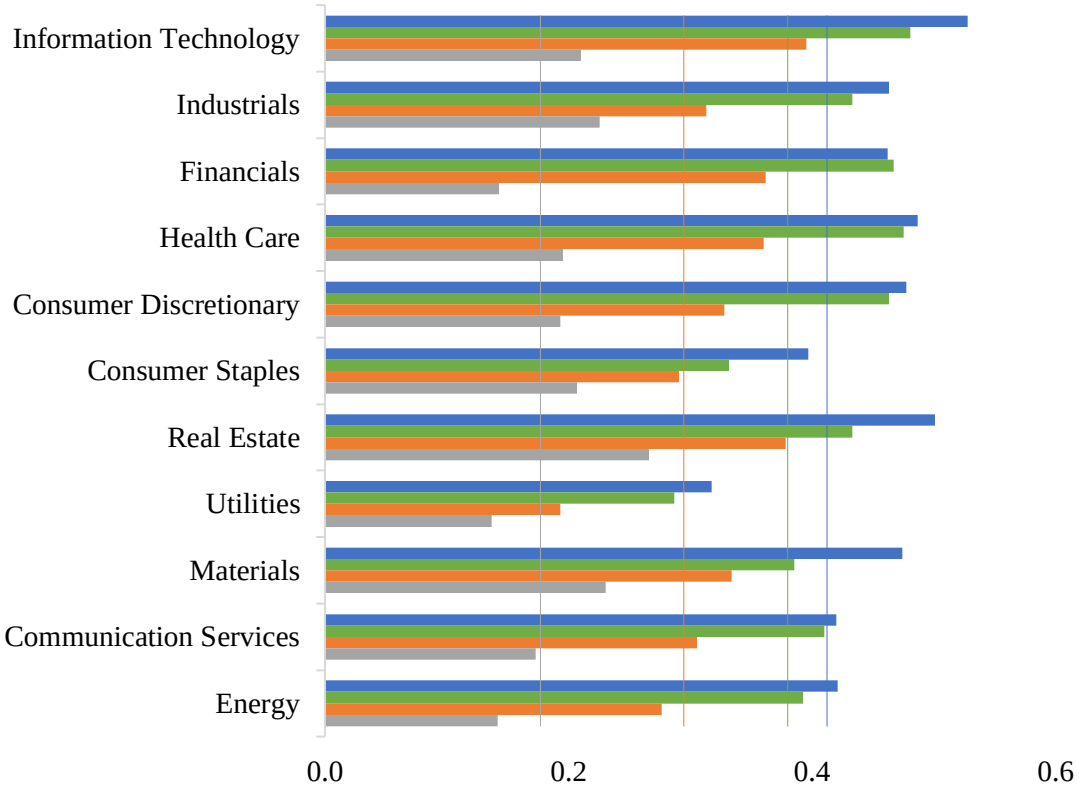

Figure 2: Sector Correlation Results

Illustrated in Figure 2 we find IT, Health Care, Financials, and Consumer Discretionary are sectors in the first quartile where agencies lean to agree on both

ESG and E ratings. All four sectors represent the four largest sectors based on market value in the S&P 500 anno 2020 and have the highest number of observations excluding industrials. The sector with the highest rate of agreement is IT, in each specific rating except governance. Real Estate has the highest on governance while being the second most agreeing on total, and social. However, the agreement is generally very low in social and governance aspects. Financials is third in environmental and social aspects; however, it is important to note that many of the environmental and social policies are not likely or applicable to them, the policies apply to the companies in their loan or investment portfolio (Eccles & Serafeim, 2013). On the other side of the scale, with the most within heterogeneity on ESG and E is Utilities and Consumer Staples. Both sectors share that they have the highest average scores across all ratings. Nonetheless, only Utilities have the lowest sector correlation in all ratings.

5.3 Determining Factors of Disagreement

We display the determining factors of disagreement here with the standard deviation as the dependent variable, and all the explanatory variables in row 1, in the enclosed panels (one for each segment), marked transparency, structure, and value. Under each of the explanatory variables in cursive is the double clustered standard errors (Std. Er.) at the time and firm-level. In the three last rows are sub-sets observations, fixed effects, and the adjusted R-squared. In columns 2-5 we report the pillars in regular order, ESG, E, S, and G.

Pillar	ESG	E	S	G
Dependent Variable:		STDEV		
(1)	(2)	(3)	(4)	(5)
Panel: Transparency				
AV	0.272	**-0.814**	**0.980**	-0.388
Std. Er.	*0.290*	*0.377*	*0.368*	*0.430*
CR	-0.257	**0.852**	-0.028	**0.826**
Std. Er.	*0.200*	*0.259*	*0.254*	*0.296*

NOA	**-0.114**	**-0.100**	-0.015	**-0.152**
Std. Er.	*0.024*	*0.031*	*0.031*	*0.036*
IO	-0.007	**-0.022**	-0.016	**0.028**
Std. Er.	*0.008*	*0.010*	*0.010*	*0.012*
Panel: Structure				
TAN	1.494	-1.067	**2.863**	3.002
Std. Er.	*1.134*	*1.473*	*1.446*	*1.679*
CUR	-0.106	-0.178	-0.170	**-0.523**
Std. Er.	*0.134*	*0.174*	*0.171*	*0.198*
LEV	0.010	0.004	0.011	0.000
Std. Er.	*0.006*	*0.007*	*0.007*	*0.008*
CE	-3.248	-7.227	-6.456	4.559
Std. Er.	*1.701*	*2.208*	*2.157*	*2.519*
Panel: Value				
MC	**3.809**	**-3.136**	**3.903**	**8.158**
Std. Er.	*0.574*	*0.738*	*0.739*	*0.850*
LIQ	-0.719	**1.331**	**1.519**	**-1.993**
Std. Er.	*0.524*	*0.656*	*0.679*	*0.775*
BM	**0.053**	**-0.070**	0.058	**0.098**
Std. Er.	*0.025*	*0.032*	*0.032*	*0.037*
GP	**-2.329**	-1.536	-2.959	-2.968
Std. Er.	*1.124*	*1.447*	*1.432*	*1.664*
Observations	5221	5218	5174	5221
Fixed Effects	Yes	Yes	Yes	Yes
Adjusted R-Squared	0.430	0.354	0.420	0.364

Table 9: Determining Factors of Disagreement Results.[5]

In table 10 we get ambiguous results overall with regards to our hypothesis. We see that structurally few variables explain the disagreement between raters, compared to

[5] We have also tested determining factors of disagreement on range, and got similar results, however after robustness checks, we found the significant variables to be explain less of the variability compared to standard deviation.

transparency and value. Starting in the column: total rating, there are four significant variables. On a 1% level significance, we have the number of analysts coverage to have a significant negative impact on the disagreement, suggesting that the disagreement lowers as the number of analysts increases. Market capitalization is also significant at a 1% level while having a positive impact on the disagreement. With book-to-market ratio have positive, gross profitability have a negative impact on disagreement at the 5% level. Indicating that disagreement increases with increasing book and market value, only offset if profitability increases. This is consistent with Sadka & Scherbina (2007) on analyst disagreement in financials (ESG are non-financial). They find that high disagreement firms have higher loadings on the market, size, and book-to-market factors than others. Suggesting that ESG-analysts also might disagree on the same factors as financial analysts do. Gibson et al. (2019) likewise find that market capitalization and gross profitability have similar results on financial characteristics on overall ESG disagreement. In the environmental pillar (column 3) all transparency factors are significant and negative except credit rating, this upholds the intuitive hypothesis that transparency reduces subjectivity, excluding the credit rating. Credit raters factor ESG into their final assessment of a firm's credit quality (S&P Global, 2021). When present, this has an increasing impact on disagreement. Similarly, with audit verification, credit ratings here are dummy variables. Audit verification is an especially valuable factor here in the environmental pillar, because in most cases the assurance only regards the environmental reporting, confirming in our case that assurance decreases E disagreement. Consistent with Ballou et al., 2019) that assurance providers are more beneficial than not in this case. On structure factors, contrary to what was anticipated, there are no significant structure-related factors in the environmental pillar. Market capitalization is a significant variable in every pillar, but only in the environmental is it negatively correlated with the disagreement. Book to market ratio is also significant and follows the sign of market capitalization. In the social pillar we have audit verification, tangibility, market capitalization, and liquidity as significant variables, all positively correlated with the disagreement. This confirms our alternative hypothesis and implies we will accept the fallacy hypothesis, that all segments have a positive relationship on disagreement. It is only in this pillar that tangibility becomes

significant. A reason might be that tangible firms are viewed as riskier in social factors is the higher risk of injuries towards employees. Companies involved in producing goods have also lagging in diversity. As ESG ratings are sector relative scores, they might have discrepancies in the measurements of sector relative severance.[6] In the governance pillar the results are ambiguous again, and interchange between a positive and negative relationship with the disagreement. The adjusted r squared environmental and governance pillars are significantly lower than their foils. This indicates that the explanatory variables are correlated with the dependent, but nevertheless do explain less of the variability in the disagreement.

5.4 Effect on Stock Price

In this section, we will test our risk-based and optimism-based hypotheses by analyzing the relationship between ESG rating disagreement and stock returns. We use a pooled panel regression model with annual stock returns as our dependent variable and ESG disagreement as our main explanatory variable. We will use two different measures of ESG disagreement. Firstly, the standard deviation of ratings at a given point in time for each firm. Secondly, the range between the highest and the lowest rating at a given point in time. Standard deviation is denoted as *Std. Dev.* and range is denoted as *Range*.

In addition to our main disagreement explanatory variables, we include sector fixed effects. To explain the cross-section of stock returns, we add control variables for standard characteristics such as market capitalization, book-to-market ratio, gross profitability, liquidity, and volume.

Sample Period	2010 - 2020
Firms Included	495
Method	Panel Least Squares

[6] According to Rigobon et al. (2019) the divergence in ESG ratings can be explained 50% by the measurement divergence.

Dependent Variable:	Price			
	(1)	(2)	(3)	(4)
Explanatory Variables:	Std. Dev.	Range	Std. Dev.	Range
	ESG		*Environmental*	
Coefficient	**1.271**	**0.528**	**-1.362**	0.134
Standard Error	***0.302***	***0.129***	***0.359***	*0.222*
Observations	5186	5205	5205	5205
Control Variables	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes
Adjusted R-Squared	0.733	0.731	0.731	0.730
	Social		*Governance*	
Coefficient	**0.666**	**0.286**	0.204	0.130
Standard Error	***0.223***	***0.096***	*0.181*	*0.079*
Observations	5145	5205	5186	5205
Control Variables	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes
Adjusted R-Squared	0.733	0.731	0.732	0.730

Table 10: Impact of ESG Disagreement on Stock Price

The results from the regression are displayed in table 11. We report the coefficient estimates for the main explanatory variable, as well as the standard error. In columns (1) and (3) we have standard deviation at the main explanatory variable, and in columns (2) and (4) we have range as the main explanatory variable. Further, the table consists of four panels, the first panel reports the results from the overall ESG rating, the remaining three panels report the results for the E, S, and G pillars.

We start by testing the risk-based hypothesis, which conjectures that companies with higher divergence in (ESG, E, S, G) ratings will experience a positive relationship between rating disagreement and future stock return. The ESG and Social panels show that disagreement in ESG and social ratings is significantly positively related to future stock returns. This evidence supports our risk-based hypothesis as in Atmaz & Basak (2018) findings.

Further, we test our optimism-based hypothesis which emphasizes that companies with higher divergence in (ESG, E, S, G) ratings will experience a negative relationship between rating disagreement and future stock returns. From the table above we see that the Environmental panel shows that disagreement in the environmental rating is significantly negatively related to future stock returns when the disagreement is measured by standard deviation. This evidence is consistent with the optimism-based hypothesis. It is also supported by Miller's (1977) research which found that investors believe that a company's true ESG score is captured by the most optimistic ESG rating resulting in an overvaluation of the stock today and lower returns in the future.

5.5 Limitations and Suggestions for Further Research

We get results that audit verification is significantly negatively correlated with disagreement in the environmental pillar. Only 6.6% having assurance over other aspects that environmental factors in their sustainability reporting in the S&P 500 (Welsh et al., 2018). Suggesting an increase in the products of assurance providers can be part of the solution in voiding the disagreement, as long as standardization is lacking. It would be interesting in examining further the effects of auditing/assurance have on the rest of the sustainability reporting, and not just the environmental. This can have an important effect for some that only rely on one providers rating.

6. Summary and Conclusion

ESG is a highly heterogeneous space. In this thesis, we determine that ESG ratings are no different. Based on a sample of companies from the S&P 500, between five prominent ESG agencies, we find significant low agreement on the overall score and for each pillar: E, S, and G. Moreover, there is also a particularly persistent distribution in which scores it is heterogeneity. The average correlation is 0.554 in overall rating, declining towards the last score and pillar, governance with an average correlation of 0.262. This relationship also applies in sector distribution. Having 64% and 36% of the total and environmental ratings placed in most "agreeing" quartile,

governance compile of 91% of the least “agreeing” quartile. It is a recurring theme that agencies disagree more when it comes to social and governance issues. It is surprising that the overall and aggregate ratings manage to evenly over the combined discrepancies and provide a score that is higher than the highest agreeing correlation of what it is accumulated from. We find sector bias in that there is significantly more agreement in the information technology and health care sector, and the most heterogeneity in the utility sector, followed by consumer staples. In essence, an ESG rating is a sector relative score assigned to an individual firm in that sector. The opposite sectors also have distinct differences in tangibility, and especially market weight in the S&P 500. We then build a set of arguments that there may be some biases, that determine disagreement, and that these come as a result of the absence of standardization and common methodologies. Across all scores, we find one variable that is significantly correlated with disagreement. In all except the environmental disagreement, it determines it in a positive increasing correlation. If the market capitalization increases in the environmental pillar, it is negative correlated with the disagreement, likewise with the book to market ratio, although not significant on the social pillar. This suggests similarly with Drempetic et al. (2019), that agencies have biases towards larger firms. Nevertheless, we get ambiguous results when investigating determinant factors of disagreement. The disagreement further creates faltering conclusions in implementing ESG scores for researchers, companies and investors based solely on which agencies product one acquires. But as, more and more capital is allocated towards sustainable investments, and the sustainable measures are ambiguous, which one do investors choose? We find that the disagreement measured in standard deviation and range is a significant factor on future stock prices in the overall rating and social rating. We posit that this can be explained by the fact that higher belief dispersion leads to uncertainty and will lead to risk-averse investors requiring higher future returns (Atmaz & Basak 2018). The other effect is however significant in the environmental score that investors believe that a company’s true ESG score is captured by the most optimistic ESG rating resulting in an overvaluation of the stock today and lower returns in the future (Miller 1977).

Appendices

A.1 – List of Variables

Variable Name	Description	Database	Construction
AV	AV is a CSR external audit verification of the environmental measures published in the firm's sustainability	Eikon Terminal	Binary: True (1), False (0)
CR	S&P domestic long term issuer credit rating	WRDS	Binary: True (1), False (0)
NOA	Number of analysts covering the specific stock	CRSP	
IO	Institutional ownership in % of the shares of a firm	Bloomberg Terminal	IO/100
TAN	The tangibility ratio. Tangible assets (PP&E) divided on total assets	Compustat	PP&E / Total assets
CUR	Current assets divided on current liabilities	Compustat	Current assets / Current liabilities
LEV	Total long-term debt plus current debt divided on the total assets	Compustat	(Long-term debt + Current debt)/ Total assets
CE	Capital Expenditure divided on PP&E	Compustat	Capex. / PP&E

MC	Average share price multiplied with the average amount of shares outstanding in the same fiscal year	CRSP	Share price x Shares outstanding
LIQ	The volume of shares traded in a fiscal year divided on average shares outstanding	CRSP	The volume of shares traded / Shares outstanding
BM	Total book value of assets divided on the market capitalization	Compustat/CRSP	Total assets divided on MC
GP	Revenues minus direct costs divided on the total assets	Compustat	(Revenues - COGS) / Total assets
VOLA	The standard deviation of the share price in the last 260 days	CRSP	
SDEV	The standard deviation on every available rating for each ESG, E, S, G	BB and Eikon Terminal, WRDS	
Range	Range between the top and the lowest rating for each ESG, E, S, G	BB and Eikon Terminal, WRDS	MAX-MIN
PRICE	Stock price on the first available day in the fiscal year minus the first available share price of the next annum.	CRSP	Share price (i) - Share price (i+1)

Table 11: List of Variables

Table 12 is a list of variables we use in the determinants of disagreement (excluding VOLA), from MC to VOLA is our control measurements when testing the impact of the disagreement on the stock price. We also display a short description, database, and construction. Note that the value determinants MC and LIQ are logged.

A.2 - List of ESG Databases

Name	Description	Database
Bloomberg	Overall (ESG), Environmental (E), Social (S), Governance (G) Score	Bloomberg Terminal
Refinitiv	Overall (ESG), Environmental (E), Social (S), Governance (G) Score	Eikon Terminal
S&P Global	Overall (ESG), Environmental (E), Social (S), Governance (G) Score	Bloomberg Terminal
Sustainalytics	Overall (ESG), Environmental (E), Social (S), Governance (G) Score	Bloomberg Terminal
MSCI	Overall (ESG), Environmental (E), Social (S), Governance (G) Score	WRDS

Table 12: List of ESG Databases

A.3 – MSCI Scoring Solution

We have three pillars, and no overall score in the MSCI KLD database. We exclude scores related to alcohol, military, firearms, gambling, nuclear, and tobacco. Likewise, as Lins et al. (2017) solution, though, we add governance.

Environmental	Social
Beneficial Products and Services	*Community*
Pollution Prevention	Charitable Giving
Recycling	Innovative Giving
Clean Energy	Other Strengths
Environment Other Strength	Negative Economic Impact
Regulatory Problems	*Human Rights*
Substantial Emissions	Human Rights Other Strength
Environment Other Concerns	Burma Concern

	Human Rights Other
Climate Change	Concerns
	Indigenous Peoples Relations
Management Systems Strength	Strength
	Freedom of Expression &
Negative Impact of Products and Services	Censorship
Land Use & Biodiversity	Human Rights Violations
Non-Carbon Releases	*Employee Relations*
Natural Capital - Water Stress	Union Relations
Natural Capital - Biodiversity & Land Use	Cash Profit Sharing
Natural Capital - Raw Material Sourcing	Employee Involvement
	Employee Strengths - Other
Supply Chain Management	Strengths
Climate Change - Financing Environmental	
Impact	Union Relations
Environmental Opportunities - Opportunities in	
Green Building	Health and Safety Concerns
Environmental Opportunities - Opportunities in	Emp. Relations Other
Renewable Energy	Concerns
Pollution & Waste - Electonic Waste	Health and Safety Strength
	Supply Chain Policies,
Climate Change - Product Carbon Footprint	Programs & Initiatives
Climate Change - Climate Change Vulnerability	Supply Chain Controversies
	Human Capital Development
Governance	Child Labor
Corp. Gov Other Concerns	Labor Management
Transparency Strength	*Diversity*
Transparency Concern	Promotion
Public Policy Strength	Board of Directors
Public Policy Concern	Work-Life Benefits
	Women and Minority
Governance Structures Controversies	Contracting

Corruption & Political Instability	Diversity Other Strength
Financial System Instability	Controversies
Controversial Investments	Non-Representation
Business Ethics	Gay and Lesbian Policies
	Board Diversity
	Board of Directors - Minorities

Table 13: Included Strengths and Concerns from MSCI

In table 14 we find the strengths and concerns we include. We include a total of 23 strengths and concerns in environmental, in social we include sub-factors as Community, Human Rights, Employee Relations, and Diversity. In Social these results into 33 strengths and concerns. In Governance, there are a total of 10 strengths and concerns. To achieve a total score, it is evident that we need to adjust the scores to achieve equal weighting.

Scale	**E**	**S**	**G**
15		97.727	
14		93.181	
13		88.634	
12		84.088	
11		79.541	
10		74.995	
9		70.449	
8		65.902	
7		61.356	
6	95.455	56.809	
5	86.363	52.263	
4	77.271	47.716	
3	68.179	43.170	

2	59.087	38.623	92.857
1	49.995	34.077	78.570
0	40.903	29.530	64.284
-1	31.811	24.984	49.997
-2	22.719	20.438	35.710
-3	13.627	15.891	21.424
-4	4.535	11.345	7.137
-5		6.798	
-6		2.252	

Table 14: Percentile Solution

In table 15 we see the assigned percentiles. The scale is significantly larger in the social aspect as there are more measures. This also changes over time, as there are periods with more assigned strengths and concerns in, for example, community than in other years. To achieve equal weighting, we need to change the percentiles for each year, we take the available strengths minus the concerns for a given firm and divide it into the available strengths and concerns. This creates a firm that can be the leader in some years with six in score in environmental, but also later be a leader with 2. We afterward assign weight that accounts equally for the total score. This is how it is similarly done in Lins et al. (2017), while their index goes from -1 - +1 ours go from 0 – 100, similarly to the other scores.

A.4 – Spearman Correlation Results

	Spearman correlations - Pairwise			
	BB	Ref.	S&PG	Sus.
ESG	(1)	(2)	(3)	(4)
Refinitiv	0.642			
S&P Global	0.609	0.542		
Sustainalytics	0.604	0.519	0.665	
MSCI	0.567	0.567	0.446	0.510
Average		**0.567**		

Environmental				
Refinitiv	0.583			
S&P Global	0.587	0.560		
Sustainalytics	0.503	0.498	0.577	
MSCI	0.504	0.504	0.490	0.428
Average			**0.523**	
Social				
Refinitiv	0.518			
S&P Global	0.564	0.500		
Sustainalytics	0.437	0.354	0.475	
MSCI	0.286	0.286	0.273	0.226
Average			**0.392**	
Governance				
Refinitiv	0.346			
S&P Global	0.459	0.211		
Sustainalytics	0.471	0.249	0.546	
MSCI	0.160	0.160	-0.026	-0.044
Average			**0.253**	

Table 15: Spearman Correlation Results

Table 16 includes the Spearman correlations done pairwise; this is to make the best use of the full sample. Spearman is best on ordinal scales, while Pearson is best at intervals. As the ratings don't make use of the full scale it is fair to say ESG ratings have a bit of both, hence why we test both Pearson and Spearman.

A.5 – Pearson Correlation Results – Common Sample

	Common sample - Pearson correlation - Complete			
	BB	Ref.	S&PG	Sus.
ESG	(1)	(2)	(3)	(4)
Refinitiv	0.610			
S&P Global	0.650	0.554		
Sustainalytics	0.590	0.530	0.690	
MSCI	0.523	0.429	0.445	0.500
Average			**0.552**	
Environmental				
Refinitiv	0.609			
S&P Global	0.599	0.551		
Sustainalytics	0.589	0.464	0.570	
MSCI	0.554	0.455	0.469	0.440
Average			**0.530**	
Social				
Refinitiv	0.442			
S&P Global	0.535	0.497		
Sustainalytics	0.444	0.345	0.499	
MSCI	0.203	0.834	0.266	0.689
Average			**0.475**	
Governance				
Refinitiv	0.283			
S&P Global	0.433	0.266		
Sustainalytics	0.368	0.299	0.575	
MSCI	0.021	0.079	-0.022	-0.047
Average			**0.225**	

Table 16: Pearson Correlation Results in Common Sample

Table 17 displays the Pearson Correlation results in the common sample. The common sample is only available in 2016-2018, so should be given little significance,

however, it is strikingly similar to the full sample, indicating that they do not change much over time.

A.6 – Pearson Correlation Results – 2019/2015

	Pearson correlations - 2019				Pearson correlations - 2015		
	BB	Ref.	S&PG		BB	Ref.	Sus.
ESG				*ESG*			
Refinitiv	0.556			Refinitiv	0.641		
S&P Global	0.621	0.527		Sus.	0.616	0.553	
Sustainalytics	0.565	0.528	0.659	MSCI	0.494	0.456	0.497
Average		**0.576**		**Average**		**0.543**	
Environmental				*Environmental*			
Refinitiv	0.571			Refinitiv	0.616		
S&P Global	0.592	0.55		Sus.	0.537	0.522	
Sus.	0.489	0.456	0.547	MSCI	0.546	0.527	0.438
Average		**0.534**		**Average**		**0.531**	
Social				*Social*			
Refinitiv	0.422			Refinitiv	0.466		
S&P Global	0.572	0.501		Sus.	0.457	0.355	
Sus.	0.461	0.372	0.448	MSCI	0.167	0.178	0.152
Average		**0.463**		**Average**		**0.296**	
Governance				*Governance*			
Refinitiv	0.26			Refinitiv	0.315		
S&P Global	0.419	0.203		Sus.	0.425	0.255	
Sus.	0.402	0.185	0.557	MSCI	-0.16	-0.03	-0.129
Average		**0.338**		**Average**		**0.113**	

Table 17: Pearson Correlation Results from 2019/2015

Table 18 displays the Pearson correlation results in both 2019 and 2015. We tested if the relationships between the raters change much over time. We tested twice out of the common sample, thus excluding S&P Global and Sustainalytics once. We concluded this was false, and that the relationships differ over time.

A.7 – Pearson Correlation Results on GOCS Sectors

Sector	**ESG**	**E**	**S**	**G**
Information Technology	0.528	0.481	0.395	0.210
Industrials	0.464	0.433	0.313	0.226
Financials	0.462	0.468	0.362	0.143
Health Care	0.487	0.476	0.360	0.196
Consumer Discretionary	0.478	0.463	0.328	0.194
Consumer Staples	0.397	0.332	0.291	0.207
Real Estate	0.501	0.433	0.379	0.266
Utilities	0.318	0.287	0.194	0.137
Materials	0.474	0.386	0.334	0.231
Communication Services	0.420	0.410	0.305	0.173
Energy	0.421	0.393	0.277	0.142

Table 18: Pearson Correlation Results on GICS Sectors

Table 19 displays the Pearson Correlation Results on the GICS sectors, for each score ESG, E, S, and G.

A.8 – Ranked and Quartile Pearson Correlation Results on GICS Sectors

Score	Sector	Correlation	Rank		Correlation	Sector	Score
ESG	I.T.	0.528	1	23	0.360	H.C.	S
ESG	R.E.	0.501	2	24	0.334	Materials	S
ESG	H.C.	0.487	3	25	0.332	C. Sta.	E
E	I.T.	0.481	4	26	0.328	C.D.	S
ESG	C.D.	0.478	5	27	0.318	Utilities	ESG
E	H.C.	0.476	6	28	0.313	Industrials	S
ESG	Materials	0.474	7	29	0.305	C.Ser.	S
E	Financials	0.468	8	30	0.291	C. Sta.	S
ESG	Industrials	0.464	9	31	0.287	Utilities	E
E	C.D.	0.463	10	32	0.277	Energy	S
ESG	Financials	0.462	11	33	0.266	R.E.	G
E	Industrials	0.433	12	34	0.231	Materials	G
E	R.E.	0.433	13	35	0.226	Industrials	G
ESG	Energy	0.421	14	36	0.210	I.T.	G
ESG	C.Ser.	0.420	15	37	0.207	C. Sta.	G
E	C.Ser.	0.410	16	38	0.196	H.C.	G
ESG	C. Sta.	0.397	17	39	0.194	C.D.	G
S	I.T.	0.395	18	40	0.194	Utilities	S
E	Energy	0.393	19	41	0.173	C.Ser.	G
E	Materials	0.386	20	42	0.143	Financials	G
S	R.E.	0.379	21	43	0.142	Energy	G
S	Financials	0.362	22	44	0.137	Utilities	G

Table 19: Ranked and Quartile Pearson Correlation Results on GICS Sectors

In table 20 we have all eleven sectors, and abbreviations for: Information Technology (I.T.), Real Estate (R.E.), Health Care (H.C.), Consumer Discretionary (C.D.), Consumer Staples (C.Sta.), Communication Services (C.Ser.) In the middle line bordered column, we see the ordered rank, where 1 is the most agreeing, while 44 is

the least agreeing. The 44 scores are then divided into quartiles of 11, illustrated by the diagonal line in the middle.

B.1 – Determinants of Disagreement - Range

Pillar	ESG	E	S	G
Dependent Variable:		Range		
(1)	(2)	(3)	(4)	(5)
Transparency				
AV	**1.762**	0.155	**4.243**	0.624
Std. Er.	*0.683*	*0.805*	*0.850*	*0.974*
CR	**-1.208**	**1.261**	-0.402	1.150
Std. Er.	*0.468*	*0.552*	*0.582*	*0.667*
NOA	-0.070	**-0.168**	0.130	-0.146
Std. Er.	*0.056*	*0.066*	*0.069*	*0.079*
IO	**0.127**	0.112	0.143	**0.180**
Std. Er.	*0.017*	*0.020*	*0.021*	*0.024*
Structure				
TAN	-4.028	-5.994	-1.370	-2.146
Std. Er.	*2.648*	*3.120*	*3.294*	*3.773*
CUR	0.041	-0.342	0.032	**-1.000**
Std. Er.	*0.313*	*0.369*	*0.390*	*0.446*
LEV	0.034	0.029	**0.033**	0.009
Std. Er.	*0.013*	*0.015*	*0.016*	*0.019*
CE	-6.149	**-15.384**	-13.708	**11.715**
Std. Er.	*4.010*	*4.725*	*4.989*	*5.716*
Value				
MC	**13.939**	**8.002**	**16.607**	**22.624**
Std. Er.	*1.322*	*1.557*	*1.644*	*1.884*
LIQ	-1.889	-0.750	**2.900**	**-3.911**
Std. Er.	*1.162*	*1.369*	*1.445*	*1.656*
BM	**0.172**	0.035	**0.226**	**0.252**
Std. Er.	*0.059*	*0.069*	*0.073*	*0.084*

GP	**-7.685**	**-6.722**	-10.208	-8.722
Std. Er.	*2.621*	*3.089*	*3.261*	*3.736*
Observations	5268	5268	5268	5268
Fixed Effects	Yes	Yes	Yes	Yes
Adjusted R-Squared	0.421	0.361	0.426	0.367

Table 20: Determinants of Disagreement with Range as Dependent Variable

In table 21 we have the results from determinants of disagreement with the range as the dependent variable, we did get similar results with SDEV however slightly lower r squared. After the robustness test, we found that the value-related factors had the most significance explaining most of the variation in the model. While the transparency and structure had a minor effect, while still being significant.

Bibliography

Adams, C. A., Hill, W.-Y., & Roberts, C. B. (1998). CORPORATE SOCIAL REPORTING PRACTICES IN WESTERN EUROPE: LEGITIMATING CORPORATE BEHAVIOUR? *The British Accounting Review*, *30*(1), 1–21. https://doi.org/10.1006/bare.1997.0060

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The Cross-Section of Volatility and Expected Returns. *The Journal of Finance*, *61*(1), 259–299. https://doi.org/10.1111/j.1540-6261.2006.00836.x

Atmaz, A., & Basak, S. (2018). Belief Dispersion in the Stock Market. *The Journal of Finance*, *73*(3), 1225–1279. https://doi.org/10.1111/jofi.12618

Ballou et al. (2019). *The Rise of Corporate Sustainability Reporting:* 25.

Banz. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, *9*(1), 3–18. https://doi.org/10.1016/0304-405X(81)90018-0

Basak, S. (2000). A model of dynamic equilibrium asset pricing with heterogeneous beliefs and extraneous risk. *Journal of Economic Dynamics and Control*, *24*(1), 63–95. https://doi.org/10.1016/S0165-1889(98)00064-5

Basak, S. (2005). Asset pricing with heterogeneous beliefs. *Journal of Banking & Finance*, *29*(11), 2849–2881. https://doi.org/10.1016/j.jbankfin.2005.02.003

Beber, A., Breedon, F., & Buraschi, A. (2010). Differences in beliefs and currency risk premiums. *Journal of Financial Economics*, *98*(3), 415–438. https://doi.org/10.1016/j.jfineco.2010.07.001

Berg, F., Kölbel, J. F., & Rigobon, R. (2020). *Aggregate Confusion: The Divergence of ESG Ratings* (SSRN Scholarly Paper ID 3438533). Social Science Research Network. https://doi.org/10.2139/ssrn.3438533

Bouten, L., Cho, C. H., Michelon, G., & Roberts, R. W. (2017). *CSR Performance Proxies in Large-Sample Studies: "Umbrella Advocates", Construct Clarity and the "Validity*

Police" (SSRN Scholarly Paper ID 3107182). Social Science Research Network. https://doi.org/10.2139/ssrn.3107182

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. https://doi.org/10.1002/smj.2407

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. https://doi.org/10.1111/j.1530-9134.2009.00210.x

Chauhan. (2014). A Relational Study of Firm's Characteristics and CSR Expenditure. *Procedia Economics and Finance*, *11*, 23–32. https://doi.org/10.1016/S2212-5671(14)00172-5

Chen, Z., & Epstein, L. (2002). Ambiguity, Risk, and Asset Returns in Continuous Time. *Econometrica*, *70*(4), 1403–1443. https://doi.org/10.1111/1468-0262.00337

Cheng, B., Ioannou, I., & Serafeim, G. (2014). Corporate social responsibility and access to finance. *Strategic Management Journal*, *35*(1), 1–23. https://doi.org/10.1002/smj.2131

Christensen, D. M., Serafeim, G., & Sikochi, A. (2021). Why is Corporate Virtue in the Eye of The Beholder? The Case of ESG Ratings. *The Accounting Review*, TAR-2019-0506. https://doi.org/10.2308/TAR-2019-0506

Clarkson, P. M., Fang, X., Li, Y., & Richardson, G. (2013). The relevance of environmental disclosures: Are such disclosures incrementally informative? *Journal of Accounting and Public Policy*, *32*(5), 410–431. https://doi.org/10.1016/j.jaccpubpol.2013.06.008

de Villiers, C., Naiker, V., & van Staden, C. J. (2011). The Effect of Board Characteristics on Firm Environmental Performance. *Journal of Management*, *37*(6), 1636–1663. https://doi.org/10.1177/0149206311411506

Drempetic et al. (2019). *The Influence of Firm Size on the ESG Score: - ProQuest.* https://www.proquest.com/docview/2215721695?fromopenview=true&pq-origsite=gscholar

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. https://doi.org/10.1016/j.jfineco.2018.08.013

Dye, R. A. (1985). Disclosure of Nonproprietary Information. *Journal of Accounting Research*, *23*(1), 123–145. https://doi.org/10.2307/2490910

Eccles, R. G., Ioannou, I., & Serafeim, G. (2011). The Impact of a Corporate Culture of Sustainability on Corporate Behavior and Performance. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1964011

Eccles, R. G., & Serafeim, G. (2013). A Tale of Two Stories: Sustainability and the Quarterly Earnings Call. *Journal of Applied Corporate Finance*, *25*(3), 8–19. https://doi.org/10.1111/jacf.12023

Elsten, C. M., & Hill, N. (2020). *Intangible Asset Market Value Study?* 3.

Escrig-Olmedo, E., Fernández-Izquierdo, M. Á., Ferrero-Ferrero, I., Rivera-Lirio, J. M., & Muñoz-Torres, M. J. (2019). Rating the Raters: Evaluating how ESG Rating Agencies Integrate Sustainability Principles. *Sustainability*, *11*(3), 915. https://doi.org/10.3390/su11030915

Fama, E. F., & French, K. R. (1995). Size and Book-to-Market Factors in Earnings and Returns. *The Journal of Finance*, *50*(1), 131–155. https://doi.org/10.1111/j.1540-6261.1995.tb05169.x

Fama, E. F., & French, K. R. (2008). Dissecting Anomalies. *The Journal of Finance*, *63*(4), 1653–1678. https://doi.org/10.1111/j.1540-6261.2008.01371.x

Fatemi, A., Glaum, M., & Kaiser, S. (2018). ESG performance and firm value: The moderating role of disclosure. *Global Finance Journal*, *38*, 45–64. https://doi.org/10.1016/j.gfj.2017.03.001

Gamerschlag, R., Möller, K., & Verbeeten, F. (2011). Determinants of voluntary CSR disclosure: Empirical evidence from Germany. *Review of Managerial Science*, *5*(2–3), 233–262. https://doi.org/10.1007/s11846-010-0052-3

Geczy, C., Stambaugh, R. F., & Levin, D. (2005). *Investing in Socially Responsible Mutual Funds* (SSRN Scholarly Paper ID 416380). Social Science Research Network. https://doi.org/10.2139/ssrn.416380

Giambona, E., Graham, J. R., Harvey, C. R., & Bodnar, G. M. (2018). The Theory and Practice of Corporate Risk Management: Evidence from the Field. *Financial Management*, *47*(4), 783–832. https://doi.org/10.1111/fima.12232

Gibson, R., Krueger, P., & Schmidt, P. S. (2019). *ESG Rating Disagreement and Stock Returns* (SSRN Scholarly Paper ID 3433728). Social Science Research Network. https://doi.org/10.2139/ssrn.3433728

Hahn, R., & Kühnen, M. (2013). Determinants of sustainability reporting: A review of results, trends, theory, and opportunities in an expanding field of research. *Journal of Cleaner Production*, *59*, 5–21. https://doi.org/10.1016/j.jclepro.2013.07.005

Hartzmark, S. M., & Sussman, A. B. (2019). Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows. *The Journal of Finance*, *74*(6), 2789–2837. https://doi.org/10.1111/jofi.12841

Hawn, O., Chatterji, A. K., & Mitchell, W. (2018). Do investors actually value sustainability? New evidence from investor reactions to the Dow Jones Sustainability Index (DJSI). *Strategic Management Journal*, *39*(4), 949–976. https://doi.org/10.1002/smj.2752

Ho, L.-C. J., & Taylor, M. E. (2007). An Empirical Analysis of Triple Bottom-Line Reporting and its Determinants: Evidence from the United States and Japan. *Journal of International Financial Management & Accounting*, *18*(2), 123–150. https://doi.org/10.1111/j.1467-646X.2007.01010.x

Hodge, K., Subramaniam, N., & Stewart, J. (2009). Assurance of Sustainability Reports: Impact on Report Users' Confidence and Perceptions of Information Credibility.

Australian Accounting Review, *19*(3), 178–194. https://doi.org/10.1111/j.1835-2561.2009.00056.x

Huber, B. M., & Comstock, M. (2017, July 27). ESG Reports and Ratings: What They Are, Why They Matter. *The Harvard Law School Forum on Corporate Governance*. https://corpgov.law.harvard.edu/2017/07/27/esg-reports-and-ratings-what-they-are-why-they-matter/

IFC. (2005). *Who Cares Wins: One Year On*. https://www.ifc.org/wps/wcm/connect/Topics_Ext_Content/IFC_External_Corporate_Site/Sustainability-At-IFC/Publications/Publications_Report_WhoCaresWins-OneYearOn__WCI__1319579364327

Ioannou, I., Li, S. X., & Serafeim, G. (2015). The Effect of Target Difficulty on Target Completion: The Case of Reducing Carbon Emissions. *The Accounting Review*, *91*(5), 1467–1492. https://doi.org/10.2308/accr-51307

Ioannou, I., & Serafeim, G. (2015). The impact of corporate social responsibility on investment recommendations: Analysts' perceptions and shifting institutional logics. *Strategic Management Journal*, *36*(7), 1053–1081. https://doi.org/10.1002/smj.2268

Kim, E.-H., & Lyon, T. P. (2015). Greenwash vs. Brownwash: Exaggeration and Undue Modesty in Corporate Sustainability Disclosure. *Organization Science*, *26*(3), 705–723. https://doi.org/10.1287/orsc.2014.0949

Kotsantonis, S., & Serafeim, G. (2019). Four Things No One Will Tell You About ESG Data. *Journal of Applied Corporate Finance*, *31*(2), 50–58. https://doi.org/10.1111/jacf.12346

KPMG. (2017). *The KPMG Survey of Corporate Responsibility Reporting 2017*. 58.

Lang, M., Lins, K. V., & Maffett, M. (2012). Transparency, Liquidity, and Valuation: International Evidence on When Transparency Matters Most. *Journal of Accounting Research*, *50*(3), 729–774. https://doi.org/10.1111/j.1475-679X.2012.00442.x

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. https://doi.org/10.1111/jofi.12505

Marquis, C., Toffel, M. W., & Zhou, Y. (2016). Scrutiny, Norms, and Selective Disclosure: A Global Study of Greenwashing. *Organization Science*, *27*(2), 483–504. https://doi.org/10.1287/orsc.2015.1039

Miller, E. M. (1977). Risk, Uncertainty, and Divergence of Opinion. *The Journal of Finance*, *32*(4), 1151–1168. https://doi.org/10.2307/2326520

Mitchell, R. K., Agle, B. R., & Wood, D. J. (1997). Toward a Theory of Stakeholder Identification and Salience: Defining the Principle of Who and What Really Counts. *The Academy of Management Review*, *22*(4), 853–886. https://doi.org/10.2307/259247

Morgan, D. P. (2002). Rating Banks: Risk and Uncertainty in an Opaque Industry. *American Economic Review*, *92*(4), 874–888. https://doi.org/10.1257/00028280260344506

MSCI. (2020). *MSCI ESG Ratings Methodology*. 17.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, *108*(1), 1–28. https://doi.org/10.1016/j.jfineco.2013.01.003

Pavlova, A., & Rigobon, R. (2007). Asset Prices and Exchange Rates. *The Review of Financial Studies*, *20*(4), 1139–1180. https://doi.org/10.1093/revfin/hhm008

Pflugrath, G., Roebuck, P., & Simnett, R. (2011). Impact of Assurance and Assurer's Professional Affiliation on Financial Analysts' Assessment of Credibility of Corporate Social Responsibility Information. *AUDITING: A Journal of Practice & Theory*, *30*(3), 239–254. https://doi.org/10.2308/ajpt-10047

Sadka, R., & Scherbina, A. (2007). Analyst Disagreement, Mispricing, and Liquidity*. *The Journal of Finance*, *62*(5), 2367–2403. https://doi.org/10.1111/j.1540-6261.2007.01278.x

Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk*. *The Journal of Finance*, *19*(3), 425–442. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x

S&P Global. (2021). *ESG in Credit Ratings*. https://www.spglobal.com/ratings/en/products-benefits/products/esg-in-credit-ratings

Stock, J. H., & Watson, M. W. (2015). Introduction to Econometrics (4th Edition). *2015*, 30.

Udayasankar, K. (2008). Corporate Social Responsibility and Firm Size. *Journal of Business Ethics*, *83*(2), 167–175. https://doi.org/10.1007/s10551-007-9609-8

Verrechia. (1983). Discretionary disclosure. *Journal of Accounting and Economics*, *5*, 179–194. https://doi.org/10.1016/0165-4101(83)90011-3

Waddock, S. A., & Graves, S. B. (1997). The Corporate Social Performance–Financial Performance Link. *Strategic Management Journal*, *18*(4), 303–319. https://doi.org/10.1002/(SICI)1097-0266(199704)18:4<303::AID-SMJ869>3.0.CO;2-G

Welsh, H., Lukomnik, J., & Kwon, S. (2018, December 3). State of Integrated and Sustainability Reporting 2018. *The Harvard Law School Forum on Corporate Governance*. https://corpgov.law.harvard.edu/2018/12/03/state-of-integrated-and-sustainability-reporting-2018/

Zorio, A., García-Benau, M. A., & Sierra, L. (2013). Sustainability Development and the Quality of Assurance Reports: Empirical Evidence. *Business Strategy and the Environment*, *22*(7), 484–500. https://doi.org/10.1002/bse.1764