Stockholm School of Economics
Department of Accounting and Financial Management
Bachelor Thesis
Spring 2023

Agree to Disagree? Disentangling ESG Rating Disagreement and its Impact on Stock Returns

Olof Gustavsson
25097@student.hhs.se

Theodor Samaras
25137@student.hhs.se

Abstract

This study investigates the phenomenon of ESG rating disagreement and its financial implications. Using a sample of the largest firms listed on the Nordic markets, we examined whether disagreement in ESG ratings from five widely used rating agencies between 2018-2022 positively affected the underlying stock returns. Specifically, a multivariate OLS regression was conducted for stock returns on ESG rating disagreement while controlling for standard firm characteristics that are known return predictors. While we confirm that ESG rating disagreement exists in the Nordics, we do not find a positive relation between rating disagreement and stock returns, as opposed to previous research within the field. Theoretical explanations concerning risk diversification and the varying financial materiality of ESG ratings strengthen our results, though empirical limitations are also acknowledged. The findings of this study are therefore indicative, but not statistically conclusive, that ESG rating disagreement does not have a stock return implication in the Nordics.

Tutor: Irina Gazizova
Keywords: ESG ratings, disagreement, stock returns, risk, uncertainty
Acknowledgements: We would like to thank our tutor, Irina Gazizova, Assistant Professor at the Department of Accounting at the Stockholm School of Economics, for her valuable support throughout the process.

Table of Contents

1. Introduction

In line with growing investor demand for sustainable practices, the scrutinization of environmental, social, and governance (ESG) related activities within companies is accelerating. In March 2022, the United Nations Principles for Responsible Investment (UN PRI) – aimed at incorporating ESG issues into investment practice – reached 4,902 signatories and an estimated total of US$121.3 trillion in AUM (Principles for Responsible Investment, 2022). Indeed, investors are increasingly considering ESG information in their investment decisions and the primary rationale appears to be a belief that it is material to investment performance (Amel-Zadeh & Serafeim, 2018). At the same time, several factors have been found to impede the integration of ESG data into the investment process. The lack of reporting standardisation and comparability across firms are considered the main obstacles for ESG integration globally (Amel-Zadeh & Serafeim, 2018).

Reflected by this movement is the rise of ESG rating agencies that aim to supply investors with third-party assessments of firms' ESG-related activities that are objective and data-driven (European Securities and Markets Authority [ESMA], 2021). These ESG ratings and scores are progressively becoming adopted by investors since they to varying degrees can support certain sustainable investing styles and help investors screen for ESG performance (Berg et al., 2022). For instance, when asked about the future of ESG investment practices, investors ranked positive screening as the most important strategy moving forward (Amel-Zadeh & Serafeim, 2018). Although various rating agencies largely cover the same ESG related activities and issues, their methodologies often differ. The lack of a regulatory framework for ESG ratings fosters variations in measurements, while the absence of standardised ESG reporting requirements across firms in the environmental, social, and governance pillars compels rating agencies to exercise judgement in producing the ratings (ESMA, 2021). In this light, rating agencies often disagree in their ratings of the same companies and hence, the phenomenon of ESG rating disagreement has emerged. In this study, we aim to further explore this rating divergence and its subsequent consequences for those who increasingly rely on them such as firms, analysts, and investors.

As pointed out by Berg et al. (2022), ESG rating disagreement has several important consequences. Most importantly, it reduces firms' incentives to improve their ESG performance as the rating agencies send them mixed signals. Their ability to forecast which ESG-improving actions will be valued by the market is consequently hampered. Therefore, the current widespread ESG rating disagreement paradoxically contradicts the rating providers' original

aim – namely to inform investors about the ESG risks and performances of companies. On the road to standardised reporting, it is thus of interest to understand the relationship between ESG rating disagreement and financial performance. Our research aims to shed light on this topic by contributing with findings from the Nordic setting.

While previous literature suggests that ESG rating disagreement positively relates to stock returns for firms on the S&P 500 stock index (Gibson Brandon et al. 2021), there has to the best of our knowledge been limited research conducted in other geographical settings. Our first contribution to this stream of literature therefore concerns transposing the findings of Gibson Brandon et al. (2021) to the Nordics. We show that ESG disagreement exists in the Nordic markets, but we do not find statistical support for a positive relationship between such rating disagreement and stock returns. In contrast to previous research within the field that studies this relationship in one country (Avramov et al., 2022; Gibson Brandon et al., 2021), we also investigate the relationship on a country-by-country basis but mostly find similar results as on the regional level. Secondly, by investigating a more recent sample period than what has been studied in the existing literature, we contribute with new findings regarding the financial implications of ESG rating disagreement. Thirdly, our study provides an initial effort at establishing connections between recent literature within the field. By interpreting the relationship between ESG rating disagreement and stock returns as driven by either risk or uncertainty, in line with Gibson Brandon et al. (2021), we additionally theorise about what effect each underlying ESG disagreement driver, as classified by Berg et al. (2022), may have on this relationship. Lastly, this study provides a summarising overview of the current Nordic ESG landscape and the development of rating divergence within it.

1.1 Purpose of study and research question

The aim of this study is to investigate the effect of diverging ESG ratings on stock returns. Previous studies have shown that increased ESG disclosure leads to higher ESG rating disagreement and that such disagreement positively relates to stock returns (Christensen et al., 2022; Gibson Brandon et al., 2021). There may therefore exist a "vicious cycle" in which companies are incentivised by stock price increases to expand their ESG disclosure, indirectly increasing rating disagreement. The increased ESG rating disagreement thereafter causes underinvestment into the actual improvement of ESG factors due to the mixed signals from the rating agencies which in the end may lead to an economic disincentive for companies to maintain sustainable practices (Berg et al., 2022). Ultimately, if such incentives exist, there is a fundamental need for re-aligning the interests. With the above reasoning in consideration, this study aims to answer the following research question:

Does ESG rating disagreement significantly affect stock returns in the Nordics?

1.2 Delimitation

The scope of this study is to investigate public firms in the Nordic setting, which limits our sample to firms listed on the stock exchanges in Sweden, Denmark, Norway, and Finland. To compute a reliable proxy for disagreement between ESG ratings, we needed to gather ratings from several providers. As such, the sample period has been restricted to five years between 2018-2022 due to the lack of ESG ratings from more than three providers per investigated firm-year for 2017 and earlier. Furthermore, due to lack of data accessibility, we consider the total ESG score from each vendor rather than the scores for individual ESG pillars.

1.3 Disposition

The remainder of this study is structured as follows. Section 2 begins by presenting previous literature and research within the field of ESG ratings, disagreement, and stock returns, to subsequently present the theoretical framework. These theories in turn lay the foundation for our hypothesis development in section 3. Section 4 describes and motivates the sample as well as the chosen empirical method. Results and analysis of the conducted statistical tests are presented in section 5, which is followed by a discussion in section 6. Lastly, section 7 presents our conclusions and research contributions, as well as suggestions for further research. References and appendices are outlined at the end of the study.

2. Literature overview and theoretical framework

The emergence of ESG ratings and its importance for investment decisions is a relatively new phenomenon. However, in recent years, the literature within the field has been expanding (Christensen et al., 2022). The first part of this section aims to highlight past contributions while drawing parallels to our study as well as navigating the current research gaps we aim to complement. The second part of this section presents our theoretical framework, explaining the relationship between ESG rating disagreement and stock returns to develop and rationalise the hypothesis, which is presented in the next section.

2.1 Literature review and previous research

2.1.1 ESG ratings

As a consequence of an increased investor demand for understanding firm's initiatives and actions in relation to ESG and sustainable practices, several "ESG rating agencies" have emerged in the last decades (Berg et al., 2022). These agencies share the common goal of providing assessments of how companies manage environmental, social, and governance risks and opportunities (Christensen et al., 2022). The ratings are ultimately intended to support the implementation of the United Nations Principle for Responsible Investment (UN PRI) by assisting investors in for example integrating ESG factors into their investment decisions, screening portfolios, and conducting due diligence (Christensen et al., 2022).

Currently, there is an ongoing debate within academia regarding the financial impact of these ESG ratings. Khan et al. (2016) found that only some of the factors that ESG rating agencies consider are financially material. However, more recent studies have shown that changes in ESG ratings may generate significant abnormal returns for the underlying stock. Specifically, Shanaev & Ghimire (2022) studied the impact of ESG rating changes on the monthly returns of US stocks and found that rating upgrades led to positive but inconsistently significant abnormal returns, while downgrades led to significantly negative abnormal returns.

Although various ratings largely cover the same ESG related activities and issues, their methodologies often differ. Generally, as outlined in a report by ESMA (2021), ESG ratings can be broadly classified into two primary categories by using the rating agencies' own definitions: (1) ESG risk ratings, that measures companies' exposure to ESG risks and subsequently how they are managed, and (2) ESG impact ratings, that measures how well

companies are committed to ESG factors. While these are the two main reported categories, a multitude of alternative products also exist.

Conceptually, ESG ratings can be compared to credit ratings, since they are both assessments of certain company-characteristics sold by third-party providers. Nonetheless, as pointed out by Berg et al. (2022), several differences exist. Firstly, the clear definition of creditworthiness as the probability of default stands in stark contrast to the definition of ESG performance that is ambiguous and highly interpretative. Secondly, since ESG reporting standards are still in their infancy and differ across jurisdictions, ESG rating agencies naturally aggregate information differently compared to credit rating agencies, which can rely on mature financial reporting standards. Thirdly, in contrast to credit ratings, ESG rating agencies are compensated by those who use their ratings rather than the companies they are rating.

2.1.2 ESG disclosure and ESG rating disagreement

Given the uncertainty surrounding the disclosure and reporting of ESG activities, and the current absence of standardised practices, there is a debate among scholars concerning how large weight one can assign ESG ratings. At an early stage, Chatterji et al. (2016) studied the convergent validity of six well-established ESG rating providers and documented a lack of rating agreement among these agencies. The authors explained these rating disagreements as an absence of common theorisation and commensurability, i.e., a lack of overlap in what providers choose to measure and in how they measure intersecting ESG activities. These findings suggest that investors should be cautious when using ESG ratings since choosing an invalid rating can lead to massive misallocations of capital (Chatterji et al., 2016).

Building on these findings, Berg et al. (2022) studied to what extent an absence of common theorisation and commensurability drive rating divergence. The authors identified three sources of divergence which they showed had varying explanatory value for ESG rating disagreement. In line with the findings of Chatterji et al. (2016), they firstly argued that providers use different sets of attributes and categories in their ratings which they referred to as "scope divergence". However, rating agencies may also measure the same attributes using different indicators – which they referred to as the "measurement divergence". Lastly, the authors suggested that rating agencies may have different views on the relative importance of each attribute used in the overall ESG score, which they referred to as a "weight divergence". According to their findings, the main driver of ESG rating disagreement is the measurement divergence which they argued to be problematic if one views ESG ratings as objective observations. Furthermore, Berg et al. (2022) also contributed to the literature by documenting

a so-called “rater effect”, which is a bias among analysts within rating agencies towards letting performance in one category influence perceived performance in other categories for the same company. This is extensively documented in psychology, where it is referred to as the “halo effect”, and is described by the authors as common within the field of ESG ratings since many of the factors that are evaluated require some degree of judgement. In simple terms, positive judgement for one ESG factor is likely to exert positive judgement on other ESG factors and on the holistic view of the firm. The authors provided two explanations for why such a rater effect exists. Firstly, rating agencies are often organised such that analysts specialise in firms rather than indicators. Secondly, certain rating agencies require firms to respond to specific questions in their questionnaires to obtain a positive factor score, irrespective of their actual factor performance.

The growing demand for ESG disclosure and reporting among companies further raises the question as to whether increased ESG disclosure necessarily leads to lower levels of disagreement. In this regard, Christensen et al. (2022), showed that firms with higher levels of ESG disclosure instead exhibit higher rating disagreement. This is counterintuitive to what has been found with financial disclosures and for credit risks where higher disclosure leads to more accurate analyses and less disagreement. In other words, this suggests that when firms disclose more information concerning their ESG performance, the rating agencies are more likely to disagree since they are provided more information to interpret differently. On the contrary to financial disclosures, the authors suggested that there is no shared understanding on what metrics to use in ESG disclosures and how to interpret them, which largely resembles the explained drivers laid out by Berg et al. (2022).

2.1.3 ESG rating disagreement in relation to stock returns

While most of the thus far mentioned literature has contributed to the field by documenting factors that explain why ESG rating disagreement exists, our research will focus on its consequences. Gibson Brandon et al. (2021) made a notable contribution in this regard by studying the effect of ESG rating disagreement on stock returns. Using seven ESG rating providers for stocks on the S&P 500 between the years 2010-2017 and measuring ESG rating disagreement as the standard deviation between ratings on the same firm-year, the authors showed that stock returns are positively related to ESG rating disagreement. These findings were rationalised by two related, yet distinct, possible explanations. The first explanation was in line with a standard risk-return argument in that higher levels of disagreement could be perceived as a source of risk for which investors demand a risk-premium. The second

explanation was that higher levels of disagreement may serve as a proxy for a specific source of Knightian uncertainty which is priced in the cross-section of stock returns. In relation to this, the authors raised some important implications. Firstly, they argued that financial analysts should adjust their estimates of a firm's equity cost of capital upward to incorporate the effect of ESG rating disagreements. Secondly, in conjunction with the first implication, CFOs should incorporate this rating disagreement effect in the cost of capital since it raises the threshold for making investment decisions.

Another recent contribution to this field of research was made by Avramov et al. (2022) who examined how ESG rating disagreement, calculated as the standard deviation between ESG ratings from six major rating agencies, affects investments in market equilibrium. The authors proposed that rating disagreement led to higher perceived market risk, higher market premium, and lower investor demand. Furthermore, by sorting a sample of US common stocks from 2002-2019 into quintile portfolios based on ESG ratings, they found that ESG rating disagreement could affect investors' demand for stocks and in turn the risk-return trade-off. Thus, though their approaches and methods differ from Gibson Brandon et al. (2021), Avramov et al. (2022) also find support for that the standard deviation of ESG ratings affects stock returns in the cross-section.

2.2 Theoretical Framework

As suggested by Christensen et al. (2022), high levels of ESG performance and disclosure significantly relates to higher levels of ESG disagreement. Such rating disagreement has subsequently been found to correspond to higher stock returns, as found by Gibson Brandon et al. (2021). Considering this, and that our analysis is conducted on the Nordic market, our study requires theorisation with regards to two main areas: (1) why there would exist ESG rating disagreement on the Nordic markets, and (2) economic interpretations of heterogeneous beliefs and its explanatory value regarding the impact of ESG rating disagreement on stock returns. However, acknowledging the current academic debate and ambiguity related to ESG ratings, we end this subsection by also laying forward theories suggesting why the relationship between ESG rating disagreement and stock returns may not be completely evident.

2.2.1 ESG rating disagreement in the Nordics

Previous research has shown that ESG rating disagreement exists in both the US and in a global context, suggesting that the phenomenon of rating divergence is widespread (Christensen et al.,

2022; Gibson Brandon et al., 2021). However, to determine whether rating divergence exists in a particular region, like in the Nordics, it is necessary to understand its current ESG landscape. For two consecutive years, 2021 and 2022, the Nordic countries have topped the global podium in an ESG index conducted by Global Risk Profile, insinuating that the region is considered a leader in ESG performance and disclosure (Global Risk Profile, 2021 & 2023). In a similar manner, the region has achieved the highest rankings in the Global Sustainable Competitiveness Index (GSCI) since its inception in 2012, strengthening its position as one of the most progressive regions in the world in terms of ESG performance (SolAbility, 2022).

In this regard, two of the findings described in **2.1.2** are of particular interest. Firstly, Christensen et al. (2022) show that high levels of ESG disclosure exacerbates ESG rating disagreement rather than resolving it. With a comparatively higher level of ESG disclosure among firms in the Nordics, is it thus expected that the region also exhibits higher ESG rating disagreement. Secondly, revisiting the "rater effect" as documented by Berg et al. (2022), high general ESG performance was found to increase the likelihood of analysts letting positive judgement for one factor exert positive judgement for another factor. With a generally higher ESG performance among Nordic firms, it may therefore be the case that considerable biases are involved in the rating process, once again increasing the probability of disagreement. In sum, the high levels of ESG performance and disclosure in the Nordics may thus theoretically cause rating disagreement. As such, it may even be the case that disagreement should be especially pronounced in the Nordics as compared to other geographical areas.

2.2.2 Heterogeneous beliefs and stock returns

As described in the introduction and in **2.1.1**, the ESG performance of firms are increasingly incorporated into investment decisions and ESG ratings have been found to significantly affect stock prices (Shanaev & Ghimire, 2022). ESG rating disagreement in turn, defined as the volatility of ESG ratings from different rating providers for the same firm-year observation, can be interpreted as heterogeneous beliefs about the true ESG-performance of a firm (Gibson Brandon et al., 2021). This belief dispersion regarding a firm's ESG rating should therefore reasonably also affect the price of the underlying stock. How and why this effect may exist can to the best of our understanding be explained as either (1) a risk-return relationship, or (2) as a source of Knightian uncertainty priced into the stock (Gibson Brandon et al., 2021; Knight, 1921). Thus, to describe and discuss the relationship between ESG rating disagreement and stock returns, it is necessary to explore the notions of risk and (Knightian) uncertainty, their distinctions, and how they may be attributed to ESG rating divergences. In this light, Anderson

et al. (2009, p.234) states a clarifying definition of both notions that "(...) an event is risky if its outcome is unknown, but the distribution of its outcomes is known, and an event is uncertain if its outcome is unknown, and the distribution of its outcomes is also unknown."

Firstly, one may view the volatility of ESG ratings as a firm-specific risk. Anderson et al. (2005) found that EPS forecast dispersion can be added as a risk factor to explain S&P 500 excess returns. This is supported by Atmaz & Basak (2018), who developed a framework which suggests that belief dispersion regarding future stock market outcomes resembles an additional risk for investors, causing them to demand a higher return when dispersion increases. These theories are built around beliefs regarding fundamental information of firms and as such, they do not directly relate to factors like ESG rating disagreement for which stock return implications are more ambiguous. Nevertheless, the underlying reasoning of these theories are considered relevant since they could partly explain why ESG rating disagreement would be perceived as a separate source of risk for which investors demand a premium. For example, if a firm is rated by several rating agencies and there exists scope- and/or measurement divergences, one may interpret this as an additional risk factor to explain stock returns. This is possible since one can find the differences in scope and measurements between the rating agencies, suggesting that the probability could theoretically be known. Specifically, assume two rating agencies have different scopes and/or measured indicators where one agency includes a specific type of ESG controversy, such as a company not meeting their CO_2 emissions target. Then, if such a controversy would take place, one may predict that a rating dispersion is probable between the two agencies. However, there would still exist ambiguity regarding the actual outcome. This example thus suggests how two drivers of ESG rating disagreement – scope and measurement divergences – can be interpreted as a source of risk.

Secondly, if ESG rating disagreement is thought of to capture uncertainty regarding the ESG performance of a firm, the volatility of ESG ratings is to be associated with ambiguity regarding the underlying probability distribution. Anderson et al. (2009) investigated this relationship and found, inter alia, that disagreement regarding market return forecasts viewed as uncertainty are highly correlated with market excess return whereas they are not when viewed as risk. Additionally, they found that the price of uncertainty is significantly positive in the cross-section. Thus, if ESG rating disagreement is indeed a proxy for ESG uncertainty, it is reasonable that investors would demand an uncertainty premium for holding the underlying stock. As an example, assume there is a divergence between two rating agencies regarding how to weigh a recent firm controversy and that the nature of this divergence cannot be known, meaning that one could not reasonably foresee or know the biases of individual analysts in the

two agencies. Thus, one could not predict whether there will be a rating dispersion between the two providers based on how much weight they will assign the different attributes and indicators related to the controversy. Such a weight divergence could therefore be interpreted as an uncertainty. Furthermore, in relation to a potential rater effect (Berg et al., 2022), a rating analyst with a positive judgement for an analysed firm may interpret positive ESG news for that firm as especially positive. However, there would still exist ambiguity regarding whether this effect would occur and as to what degree the rater effect would increase ratings. As such, another proven explanation for ESG rating disagreement – the rater effect – may encapsulate an uncertainty that could explain why ESG rating disagreement may affect stock returns.

2.2.3 Complexity regarding ESG rating disagreement and stock returns

While most of the thus far presented literature and theories suggest that ESG rating disagreement is expected to have a stock return implication, there are also theories problematising this relationship. For instance, as mentioned in section **2.1.1**, Khan et al. (2016) argue that all factors incorporated in the ESG ratings are not financially material, and as such, all parts of the ratings are not necessarily relevant to the fundamental value of firms. Furthermore, the authors found that stocks with high ratings in terms of sustainability issues defined as material by the Sustainability Accounting Standards Board (SASB) significantly outperformed firms with lower respective ratings, while firms with high ratings in terms of immaterial sustainability issues did not outperform firms with respectively lower ratings. These findings thus increase the complexity in the relation between ESG ratings and stock performance. If some factors that make up the total ESG score are not financially material, then a part of the ESG score should not reasonably affect the stock price. Since each rating agency analyses a different number of ESG factors for their overall scores, and not all agencies publicly disclose a full list of what factors are analysed, it may very well be the case that financially immaterial factors are considered. Thus, this may weaken the effect of ESG ratings on stock returns and therefore logically also the effect of ESG rating disagreement on stock returns. Put differently, if ESG ratings alone are not relevant for the stock price, the dispersion of ESG ratings therefore also becomes irrelevant in the context.

Grewal et al. (2021) complement this theorisation by stating that not all ESG issues are financially material for companies in each industry. The impact of ESG information disclosure on stock prices may even vary within industries since the exposure to ESG issues and risks naturally differ between companies. This suggests that firms scoring highly on sustainability issues defined as material by SASB may still exhibit different levels of stock price increases.

3. Hypothesis development

The combination of previous literature and theoretical frameworks contribute to identifying the focus of this study. In short, there are several reasons for why ESG disagreement exists. Rating providers use different sets of attributes and categories in their ratings, measure the same attributes differently, and assign different weights to each attribute (Berg et al., 2022). Furthermore, due to the human judgement that is required for evaluating firms' ESG performance, there is a tendency to let positive judgement for one factor cause positive judgement for another factor (Berg et al., 2022). Additionally, firms with higher levels of ESG disclosure have been found to exhibit the highest levels of rating disagreement (Christensen et al., 2022). Considering that Nordic firms are among the most progressive in the world in terms of both ESG disclosure and performance, as outlined in section **2.2.1**, ESG rating disagreement may therefore be especially pronounced in the region. Assuming ESG rating disagreement exists for the largest firms in the Nordics, we investigate whether such disagreement has implications for stock returns. Although there exists complexity in this relationship, as the financial materiality of ESG ratings can be questioned and the materiality may differ between firms in the same industry (Grewal et al., 2021; Khan et al., 2016), we still expect the disagreement among rating providers to infer either a risk or uncertainty in relation to the underlying stock. Indeed, ESG rating disagreement could either be seen as a separate source of risk that investors demand a risk premium for, or as an uncertainty that is priced in the cross-section of stock returns. Particularly, differences in scopes and measurements between rating agencies may be interpreted as sources of risk whereas divergences in weighting and the rater effect may be interpreted as uncertainty. Both frameworks suggest that rating disagreement encapsulates a positive return implication. This leads us to formulate our hypothesis:

H_1: Disagreement regarding the ESG ratings for the largest firms on the Nordic markets leads to significantly higher returns for the underlying stocks.

4. Method

4.1 Data collection and sample construction

To test our research question of whether ESG rating disagreement significantly affects stock returns in the Nordic setting, we formed a representative sample of Nordic firms. Using the screening tool in the Refinitiv Eikon database on the 13th of February 2023, we created a list of the 150 largest publicly listed Nordic firms in terms of market capitalization. To screen based on market capitalization was considered reasonable for three reasons: (1) these firms have a significant impact on the market, (2) it provides a better proxy for a "Nordic S&P 500", which makes our results more comparable to those of Gibson Brandon et al. (2021) and (3) these firms are more likely to have readily available ESG ratings than smaller firms. Four firms were excluded from the initial sample for one of two reasons. Three of them were excluded since none of the chosen rating providers had rated those firms. The fourth firm was left out since it was the only Icelandic company in the sample. Since we wanted to complement our main test of disagreement in the Nordics with tests on a country-by-country basis, it was deemed irrelevant to include only one Icelandic firm. This process left us with 146 Nordic firms listed in Sweden, Denmark, Norway, and Finland. See **Table 1** for a decomposition of the sample by country and **Table 2** for a decomposition by industry. The dependent variable of stock returns was collected by downloading calendar year annualised stock prices using FactSet on the 22nd of February and calculating the percentage change between firm-years. Stock data for the control variable "momentum" was also collected using FactSet while accounting data for the other control variables in the regressions were collected from Standard & Poor's Capital IQ on the 31st of March. Data collection for the explanatory variable of ESG rating disagreement was more strenuous as it required finding data from several rating agencies. In **Table A1** in the appendix, we define the variables used in this study.

Table 1: Sample distribution by country

Country	Firms	Observations	Percentage
Sweden	69	264	49.07%
Denmark	32	106	19.70%
Norway	23	88	16.36%
Finland	22	80	14.87%
Total	146	538	100%

Note that "Observations" refers to the firm-year observations of ESG rating disagreement, restricted to only be computed if three or more ratings were available for that firm-year.

Table 2: Sample distribution by industry

Industry classification	Firms	Observations	Percentage
Energy	4	15	2.79%
Materials	17	68	12.64%
Industrials	42	139	25.84%
Consumer Discretionary	8	34	6.32%
Consumer Staples	10	43	7.99%
Health Care	16	60	11.15%
Financials	23	84	15.61%
Information Technology	9	32	5.95%
Telecommunication Services	5	24	4.46%
Utilities	2	10	1.86%
Real Estate	10	29	5.39%
Total	146	538	100%

The Global Industry Classification Standards, GICS, are used to categorise the firms into 11 industries. Note that "Observations" refers to the firm-year observations of ESG rating disagreement, restricted to only be computed if three or more ratings were available for that firm-year.

Out of the six ESG ratings agencies that were mapped out prior to the data collection as the primary ones used globally, we managed to get access to five: (1) Refinitiv Eikon, (2) S&P Global, (3) FTSE Russell, (4) the public version of MSCI, and (5) Bloomberg. In the latest version of "Rate the Raters" – a project of the SustainAbility Institute that surveys companies and investors views on ESG raters – S&P Global and MSCI were among the top four in terms of usefulness according to the corporate respondents (Brackley et al., 2023). The other two providers in the top four, Sustainalytics and CDP, were not included since we did not have access to Sustainalytics and since the CDP score solely considers the environmental pillar. The remaining three providers included in our sample, Refinitiv Eikon, FTSE Russell, and Bloomberg, were also considered in "Rate the Raters" and valued by both investors and companies. As such, the sample is considered representative of commonly used rating providers. **Table 3** describes the five rating providers used in the sample. To ensure that the chosen ratings were comparable, we re-scaled the scores so that they were all based on a scale of 0-1. For instance, since the ratings from FTSE Russell range from 0 to 5, we divided all ratings from FTSE by 5. Regarding MSCI, who present their scores in a categorical way (CCC-AAA), we did not have access to the exact underlying numerical rating through the public version. Rather, the score for each firm was given as a letter grade with a corresponding numerical range. As such, each letter grade was assigned the average value of the ranges that each letter represented within the overall range 0-10 and divided by 10, as can be seen in note (2) in **Table 3**. Another important aspect when we chose which rating providers to include,

related to the comparability of their methodologies. To ensure comparability between the ESG rating agencies, it was crucial to examine whether they to some extent aim to measure the same aspects. Our approach has been to include, to the best of our capacity, providers that consider both ESG opportunities, risks, and controversies that can improve and/or impair the rating, and that therefore constitute the most comprehensive ratings. However, to incorporate the most widely used rating agencies, we also included Bloomberg, which is solely disclosure-oriented. In **Table A2** in the Appendix, we summarise each rating provider's methodology for further clarification.

Table 3: ESG Rating Agencies

Data provider	Sample period covered	Rating scale	Firm-year observations
Refinitiv Eikon	2018-2021	0-100	558
S&P Global	2018-2022	0-100	560
FTSE Russell	2018-2022	0-5	379
MSCI (public version)	2018-2022	CCC-AAA (0-10)	350
Bloomberg	2018-2022	0-100	603

(1) The maximum number of firm-year observation could in theory be 730 (146 firms over five years)
(2) The MSCI ESG rating scale is presented in a categorical way (Leader (AAA, AA), Average (A, BBB, BB), Laggard (B, CCC)). Each rating letter represents a range within the overall range of 0-10. Since the exact numerical rating is not retrievable through the online version of the MSCI ESG Rating, the average value of each range has been assigned to each firm. For clarity, a firm with the rating 'A', i.e., within the range 5.714-7.143, is thus given the rating ((5.714 + 7.143)/10)/2 = 0.643. The exact ranges are: CCC: 0.0-1.429, B: 1.429-2.857, BB: 2.857-4.286, BBB: 4.286-5.714, A: 5.714-7.143, AA: 7.143-8.571, AAA: 8.571-10.0 (MSCI, 2020).
(3) Some table categories are adopted from the ones used in Gibson Brandon et al. (2021).

In accordance with the reasoning by Gibson Brandon et al. (2021), we aimed to include as many data providers as possible while maintaining a useful sample period. While this naturally restricted the sample, a five-year sample from 2018-2022 was still possible to study. Years prior to 2018 for which we had data were excluded for two reasons: (1) the data was much less comprehensive, i.e., not as many companies were covered, and (2) the data was only from two out of five providers.

4.2 Research method

To investigate the research question, we used a quantitative approach to examine the potential stock return implications of ESG rating disagreement. Specifically, an ordinary least square (OLS) multivariate regression with fixed effects and clustered standard errors was used. The two key variables of interest in this empirical model are the dependent variable of annual stock returns and the independent variable of ESG rating disagreement. In line with the

methodologies used in previous research within the field, we defined ESG rating disagreement as the standard deviation of the available ratings from all five providers at yearly intervals for each firm-year observation (Avramov et al., 2022; Christensen et al., 2022; Gibson Brandon et al., 2021). Therefore, an example of such a data point may be "the standard deviation of ESG ratings provided by Refinitv Eikon, S&P Global, FTSE Russell, MSCI, and Bloomberg for Volvo AB in 2018". Important to note is that we required at least three ratings for each firm-year observation to calculate the disagreement variable. This was done to include enough providers to minimise biases from differences and similarities between the methodologies of each rating provider.

4.3 Empirical model

Before testing our hypothesis (H_1), we needed to investigate whether there even existed disagreement regarding the ESG ratings of the largest Nordic firms between the rating providers included in the sample. In line with previous literature, this was done by computing Pearson correlations between all five rating providers included in the sample and investigating the average pairwise correlations.[1] This investigation was not stated as a formal hypothesis since finding significant correlations was not the main interest but rather, it was to compare the magnitudes of the average correlations. Therefore, although its simplistic nature, this method facilitated the test of our hypothesis, which is our main test that answers the research question. To explore the relationship between ESG rating disagreement and stock returns and to test our hypothesis, we estimated the following model:

$$STOCK\_RETURN_{i,t} = \beta_0 + \beta_1 DISAGREEMENT_{i,t} + \beta_2 MC_{i,t} + \beta_3 BTM_{i,t} + \beta_4 OP_{i,t} + \beta_5 INV_{i,t} + \beta_6 MOM_{i,t} + \beta_7 INDUSTRY + \beta_8 YEAR + \beta_9 COUNTRY + \varepsilon_{i,t}$$

Where:

STOCK_RETURN: The annual stock returns for firm *i* in year *t*

DISAGREEMENT: The standard deviation of ESG ratings firm *i* received for its ESG performance in year *t*

MC: Market capitalization for firm *i* in year *t*

BTM: Book-to-market ratio for firm *i* in year *t*

[1] With previous literature, we refer to Chatterji et al. (2016), Gibson Brandon et al. (2021), Avramov et al. (2022), and Berg et al. (2022).

OP: Operating profitability for firm *i* in year *t*
INV: Investment for firm *i* in year *t*
MOM: Momentum for firm *i* in year *t*
INDUSTRY: Industry fixed effects
YEAR: Year fixed effects
COUNTRY: Country fixed effects
i: Firm unit in the cross-section
t: Calendar year as of December 31st

See **Table A1** for definitions of how these variables were constructed.

The coefficient of interest is β_1. If the coefficient estimate is positive and statistically significant at a 5% level, we reject the null hypothesis and confirm that ESG rating disagreement has a stock return implication for the largest listed firms in the Nordics.

$$H_0: \beta_1 \leq 0 \qquad H_1: \beta_1 > 0$$

Control Variables

To limit the influence of confounding variables and to enhance the internal strength and validity of our model, several control variables were collected and included in the model regression. Since the regression model explores whether ESG rating disagreement has a stock return implication, it is necessary to control for other standard stock characteristics that have been found to have high explanatory value in the cross-section of stock returns. We commenced by investigating the control variables used in previous literature within ESG rating disagreement and stock returns (Gibson Brandon et al., 2021) and comparing it to the predominant factors used in regressions to explain the cross-section of stock returns. In the end, this led to combining the factors from the Fama & French (2015) five-factor asset pricing model with the momentum effect, as documented by Jegadeesh & Titman (1993). The reasoning for including each variable is described below, while a technical definition and source for each variable can be found in **Table A1** in the Appendix.

Firstly, to control for the "size effect", i.e., that smaller stocks have, on average, had higher historical risk adjusted returns than larger stocks, the market capitalization of each firm-year observation is included as a control variable (Banz, 1981). This effect has not been proven to be linear but rather, the main effect occurs for very small firms while the difference in return

between larger firms is quite small. Thus, since the sample in this study is made up of the largest firms based on market capitalization, this may not be considered a case where firm size is the most important control variable. Nonetheless, in line with Gibson Brandon et al. (2021), we included it. Secondly, to control for the positive relationship between the ratio of a firm's book value of common equity and market value to the returns of the underlying stock, also known as the "value effect", we included the book-to-market ratio for each firm-year observation as a control variable (Fama & French, 1992). Thirdly, we included operating profitability, calculated as EBIT less interest expense divided by the book value of equity, in line with the definition by Fama & French (2015). This, since it has been proven to positively impact the cross-sectional returns of stocks, especially together with the other variables in the Fama and French five-factor model. Fourthly, the Fama and French five-factor model also includes a factor to proxy the investments made by a firm between two fiscal years, defined as the change in total assets. Investments was therefore also included as a control variable with similar reasoning as for the operating profitability (Fama & French, 2015). Lastly, the strategy of buying stocks that have performed well historically and selling those that have performed poorly historically has been shown to generate significantly positive returns when held for three to twelve months (Jegadeesh & Titman, 1993). To account for this well-documented phenomenon of autocorrelation in stock returns, we included a "momentum factor" by collecting historical stock data and calculating the relevant past returns.

After collecting the data and computing all control variables, efforts were made to bring the variables closer to the normal distribution. Market capitalization, book-to-market, operating profitability, investment, and momentum, which are all continuous variables, were winsorized at the 1st and 99th percentile to mitigate for any extreme outliers and to improve robustness. With similar reasoning, we also applied a natural logarithmic scale to market capitalization since it was retrieved in a levels format.

Fixed effects and clustering of standard errors

To account for differences across specific groups of firm-year observations, fixed effects were included in three dimensions. Firstly, to control for industry fixed effects, industry classifications from the global industry classification standards, GICS, were used. The 146 firms included in the sample were classified into 11 industry groups to minimise the possibility that the results were influenced by differences between industries. This is important since different industries may naturally have different base-levels of ESG ratings and because their stock prices are correlated due to industry-specific factors and events. Secondly, since ESG

disclosure and average ESG ratings have been documented to increase over time (Amel-Zadeh & Serafeim, 2018; Christensen et al., 2022), and since our sample covers a five-year period, year-fixed effects were applied to eliminate differences between years. Lastly, in contrast to Gibson Brandon et al. (2021), our study includes data from several countries, and thus, country-fixed effects were applied to control for potential country-specific variations.

To reduce the likelihood of correlations across observations, for instance between the explanatory variables, the standard errors have been clustered at the firm-level to ensure robustness and that the *t*-statistics are not overstated. Additional clustering on a time dimension was not considered reasonable since the sample period is denoted in years rather than months which would then imply too few clusters.

5. Results and analysis

This section presents the results of our research and is divided into four subsections. Descriptive statistics of the rating providers and all main variables are presented in **5.1**. Pearson correlations, aimed at determining the extent of disagreement, are presented in section **5.2**. Results and comments for H_1 are presented in **5.3**. To complement the findings from this subsection, we present an additional analysis in section **5.4**, investigating the hypothesis on a country-by-country basis.

5.1 Descriptive statistics

Panel A in **Table 4** provides summary statistics for the five different ESG data providers included in the sample. Out of the 730 possible firm-year observations in our sample of 146 firms over five years, we were able to collect 490 observations on average from each rating agency where MSCI provided the lowest amount at 350 observations and Bloomberg the most at 603. We note that the average ratings vary quite substantially between the rating providers, with MSCI having the highest average rating of 0.742 and S&P Global having the lowest of 0.380. This difference is visualised in **Figure 1**, which plots the average rating for each provider and year in our five-year sample period. Although the ratings are relatively stable over time, it is evident that there has been a general increase in average ratings during the sample period, disregarding the development of the FTSE ratings. This is in line with the findings of Christensen et al. (2022), who showed that the average rating for a given firm has been increasing over time. Furthermore, we note that the standard deviation of each rating provider is high, lying between 0.121 and 0.207, and that the range between the minimum and maximum ratings is large – indicating significant variation in ESG performance across the firms and that the sample consists of both high- and low performers. Three out of five rating agencies provided a maximum rating above the average maximum rating where out of the two that did not, S&P Global was very close at 0.890 while Bloomberg stands out with a low maximum value of 0.786. With a relatively narrow average interquartile range, between 0.470 and 0.702 – partly increased due to the lack of continuous ratings from MSCI – it is concluded that most of the firms retrieve ratings centred around the median.[2] In other words, the ratings data across all providers therefore seems to be approximately normally distributed.

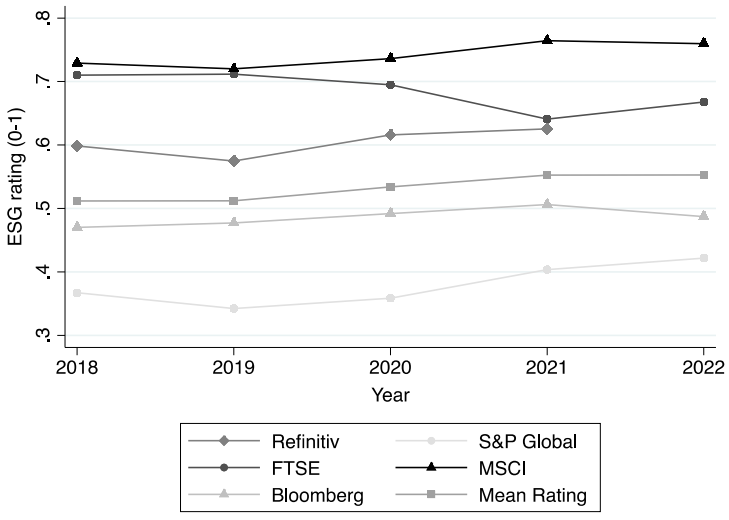
[2] Since the exact numerical ratings are not retrievable through the public version of the MSCI ESG Rating, the average value of each underlying range has been assigned to each firm. As such, the ratings can only take on these specific discrete average values. See **Table 3** for a further explanation.

Table 4: Summary statistics for the rating providers and variables included in the regressions

Panel A								
Rating agency / variable	**N**	**Mean**	**SD**	**P25**	**Median**	**P75**	**Min**	**Max**
Refinitiv Eikon	558	0.604	0.169	0.488	0.626	0.728	0.041	0.932
S&P Global	560	0.380	0.207	0.220	0.330	0.510	0.050	0.890
FTSE Russell	379	0.681	0.143	0.600	0.680	0.780	0.220	0.960
MSCI	350	0.742	0.164	0.643	0.786	0.929	0.071	0.929
Bloomberg	603	0.487	0.121	0.401	0.486	0.564	0.142	0.786
Average	490	0.579	0.161	0.470	0.582	0.702	0.105	0.899
Panel B								
STOCK_RETURN	716	0.142	0.405	-0.123	0.092	0.349	-0.587	1.753
DISAGREEMENT	538	0.171	0.058	0.126	0.172	0.216	0.020	0.350
MC	704	10.983	1.067	10.253	10.919	11.714	8.262	13.475
BTM	704	0.428	0.385	0.133	0.306	0.612	0.008	1.880
OP	669	0.197	0.178	0.083	0.180	0.274	-0.285	0.907
INV	727	0.167	0.339	0.020	0.084	0.171	-0.209	2.173
MOM	708	0.211	0.378	-0.038	0.152	0.373	-0.466	1.753

The table above provides summary statistics in two panels: Panel A displays the summary statistics for all rating providers included in the sample, while Panel B presents the summary statistics for the main variables included in the regressions after winsorization and logarithmation. Both panels show summary statistics covering the whole sample period of 2018-2022. *STOCK_RETURN* is the variable for annual stock returns. *DISAGREEMENT* is the standard deviation of the available ratings from all five providers at yearly intervals for each firm-year observation and requires at least three firm-year rating observations. This definition is possible since we have standardised the ratings from 0-1. *MC* is the natural logarithm of a firm's market capitalization. *BTM* is the book-to-market ratio. *OP* is the operating profitability. *INV* is the change in total assets between the fiscal years *t-1* and *t-2*, divided by assets in *t-2*. *MOM* denotes momentum and is the cumulative return between the years *t-2* and *t-1*. See **Table A1** in the Appendix for exact definitions of each variable.

Figure 1: The average ratings of ESG rating providers over time



Presented in the figure above are the average ESG ratings of each rating provider, converted to the scale of 0-1, across the sample period (2018-2022). The average rating across all providers is also included as a reference.

In Panel B of **Table 4**, we provide summary statistics for all the main variables included in the regressions after winsorization and logarithmation. Prior to winsorizing the continuous variables, the occurrence of outliers was high. For instance, the market capitalization of the largest firms were substantially higher than that of the median firm. As such, we validate the need for winsorizing and applying the natural logarithmic scale to market capitalization, as explained in section **4.3**. The occurrence of outliers was also found in most of the other variables, except for the disagreement variable, which solidifies the decision to winsorize these variables at the 1st and 99th percentiles before conducting the regressions and further analysis. We note that the *DISAGREEMENT* has 538 firm-year observations, indicating that the sample includes 538 firm-year observations for which three or more ratings were available, in line with our definition. The average ESG rating disagreement is 0.171 which insinuates a high variation between ratings for the same firm-year observation. Such a volatility in ratings is expected in line with both our theorisation about disagreement in the Nordics which was laid out in section **2.2.1**, and previous research (Avramov et al., 2022; Christensen et al., 2022; Gibson Brandon et al., 2021).

5.2 Pearson correlations and ESG rating disagreement

Table 5 shows the Pearson correlation coefficients between all variables included in the regression models. We find that the correlation between *STOCK_RETURN* and *DISAGREEMENT* is positive at 0.020 however insignificant and very low in magnitude, which is the opposite of what was expected according to our hypothesis development. **Figure A1** in the Appendix visualises this relationship which is further investigated in section **5.3**. What can be observed is that the *DISAGREEMENT* is significantly correlated with other variables, like *BTM* and *OP*. The correlation between *DISAGREEMENT* and *OP* is positive with a magnitude of 0.123 indicating that higher profitability is associated with higher ESG rating disagreement and vice versa. The correlation between *DISAGREEMENT* and *BTM* is on the other hand negative at -0.141, indicating that firms with high book-to-market ratios, i.e., firms more likely to be considered undervalued, have lower ESG rating disagreement, while firms with low book-to-market ratios, i.e., firms more likely to be overvalued, have higher ESG rating disagreement. The largest positive correlation, at 0.329, is found between *MOM* and *INV* while the correlation is low between *MOM* and *STOCK RETURN*, indicating that lagged changes in investments correlates with the lagged stock returns for the same period, rather than predicting future stock returns.

Table 5: Pearson correlations for the variables included in the regressions

Variable	*1)*	*2)*	*3)*	*4)*	*5)*	*6)*
1) STOCK_RETURN						
2) DISAGREEMENT	0.020					
3) MC	-0.284***	-0.060				
4) BTM	-0.017	-0.141***	-0.142***			
5) OP	0.013	0.123***	0.237***	-0.423***		
6) INV	0.009	0.058	-0.074**	-0.082**	-0.061	
7) MOM	0.067*	0.067	0.027	-0.251***	0.073*	0.329***

The table above shows the Pearson correlations coefficients for the main variables included in our regressions. *STOCK_RETURN* is the annual stock returns. *DISAGREEMENT* is the standard deviation of the available ratings from all five providers at yearly intervals for each firm-year observation and requires at least three firm-year rating observations. This definition is possible since we have standardised the ratings from 0-1. *MC* is the natural logarithm of a firm's market capitalization. *BTM* is the book-to-market ratio. *OP* is the operating profitability. *INV* is the change in total assets between the fiscal years *t-1* and *t-2*, divided by *t-2* assets. *MOM* denotes momentum and is the cumulative return between the years *t-2* and *t-1*. See **Table A1** in the Appendix for exact definitions of each variable. Significance levels are indicated by ***p<0.01, **p<0.05, *p<0.1

To evaluate whether ESG rating disagreement exists in the Nordics, Pearson correlations are also plotted between each rating provider as can be seen in **Table 6.** All correlations are highly significant and positive implying that all rating providers have a positive linear relationship with each other. This was expected since they are all providing the same type of rating on the same firm-year observation. The highest correlation, at 0.645, was found between Refinitiv Eikon and Bloomberg which is surprising considering that their methodologies are among those who differ the most.[3] Furthermore, Bloomberg has, relative to the correlations of the other providers, generally the highest correlations with all rating providers. A possible explanation for this is that Blomberg is the only rating provider in our sample that uses a disclosure-oriented measure and thus, we cannot expect disagreement stemming from divergences in measurement nor weight since they are essentially not using the same type of metrics. In contrast to this pattern, the correlation between Bloomberg and MSCI of 0.499 was below the average correlation of 0.515. However, MSCI did in general show the lowest correlations with all providers which was expected since their ratings data were much less granular and with a lower number of observations than the other providers, as described in **4.1**. To draw any conclusions about whether there exists ESG rating disagreement in the Nordics from the magnitudes of these correlations, it is necessary to benchmark against levels retrieved in previous studies, but also against another type of sold third-party assessment – credit ratings.

[3] See **Table A2** in the Appendix for detailed information about the methods of each ESG rating provider.

Table 6: Pearson correlations between ESG rating providers

Variable	*Refinitiv Eikon*	*S&P Global*	*FTSE Russell*	*MSCI*
Refinitiv Eikon				
S&P Global	0.442***			
FTSE Russell	0.448***	0.594***		
MSCI	0.476***	0.400***	0.449***	
Bloomberg	0.645***	0.589***	0.605***	0.499***
Average correlation		0.515		

The table above shows the Pearson correlations between each ESG rating provider, i.e., the linear relationship between the ratings of two providers at a time. Significance levels are indicated by ***p<0.01, **p<0.05, *p<0.1

To the best of our knowledge, important recent contributions to this field of research can be found in four major papers. Within these, the presented average correlations for the total combined ESG scores between rating providers were 0.300 in Chatterji et al. (2016), 0.447 in Gibson Brandon et al. (2021), 0.480 in Avramov et al. (2022), and 0.540 in Berg et al. (2022).[4] We therefore conclude that the average correlation of 0.515 found in this study, as presented in **Table 6**, is comparable to those found in the previous literature. As such, we present evidence that there does indeed exist ESG rating disagreement on the Nordic markets. Additionally, with these comparative numbers in mind, it seems as though the correlations are increasing over time and hence that ESG rating disagreement is decreasing over time. However, the samples used in the mentioned previous research do not overlap with ours in terms of geography, time, and to some extent the used ESG rating providers. For example, our sample differs to that of Gibson Brandon et al. (2021) when it comes to geography and time while four out of the five rating providers are corresponding.[5] Thus, to maintain prudency, we cannot conclude that ESG rating disagreement has decreased since the publishing of previous studies. Furthermore, with reference to the reasoning in **2.2.1**, we can from the comparative figures not conclude an especially pronounced ESG rating disagreement in the Nordics since our study presents higher Pearson correlations than most of the previous studies conducted in other geographical regions.

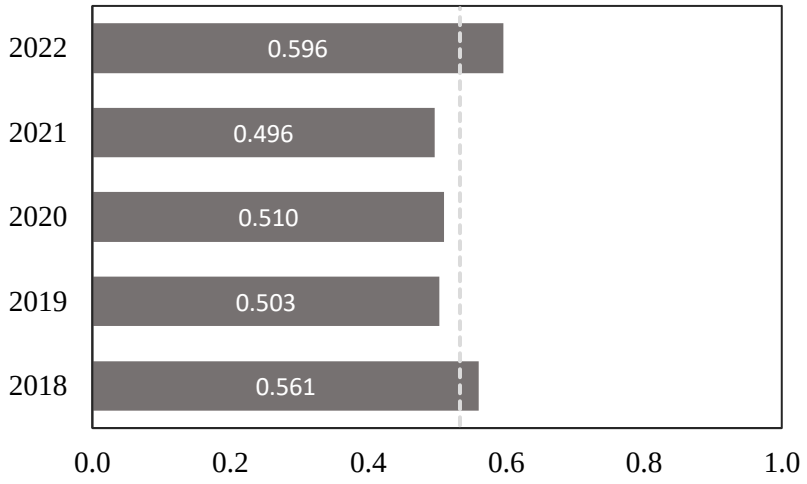
From another point of view, ESG rating agencies may conceptually be compared to credit rating agencies, such as Moody's and Standard & Poor's, as explained in **2.1.1**. Both

[4] Chatterji et al. (2016) used six leading ESG rating agencies to analyse ratings covering firms in the U.S., Canada, and Europe focusing on the period 2004-2006. Gibson Brandon et al. (2021) studied the ESG ratings from companies on the S&P 500 during 2010-2017. Avramov et al. (2022) collected ESG ratings from six major agencies on U.S. common stocks between the years 2002-2019. Berg et al. (2022) studied global aggregate ESG ratings from 2014 provided by six different rating providers.

[5] The authors included the following providers: Refinitiv Eikon (Asset4), Sustainalytics, Inrate, MSCI, Bloomberg, FTSE, and KLD.

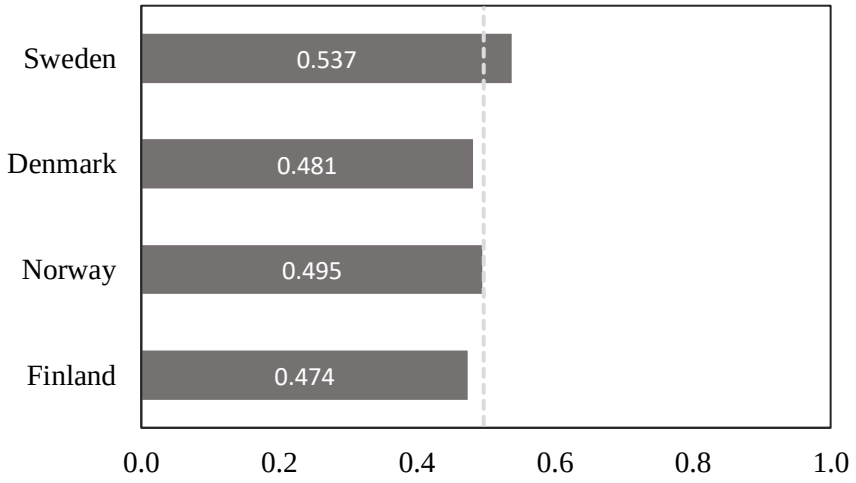
ESG rating agencies and the credit rating agencies provide a rating that allows investors to screen for the performance of firms and though they differ, this comparison serves as a meaningful benchmark in determining whether there exists a divergence in ratings. According to Berg et al. (2022), credit ratings have been found to be almost perfectly correlated at 0.99 and thus, in comparison, we conclude that our sample showed disagreement between ESG rating providers since the presented average correlation of 0.515 is substantially smaller than 0.99.

Figure 3: Average correlations between ESG rating providers over time



Presented in the figure above are the average pairwise Pearson correlations between the normalised ratings (0-1) of the ESG rating providers for each year included in the sample. The vertical line represents the average correlation across all years.

Figure 4: Average correlations between ESG rating providers by country



Presented in the figure above are the average pairwise Pearson correlations between the normalised ratings (0-1) of the ESG rating providers for each country included in the sample. The vertical line represents the average correlation across all countries.

As can be seen in **Figure 3**, there is not a clear trend in terms of average correlations between rating providers in our sample period. We note that the average correlation has increased when comparing 2018 to 2022, from 0.561 to 0.596, but also that the average correlation was stable at approximately 0.5 for the three consecutive years 2019, 2020, and 2021. Since we have a comparably lower amount of datapoints for 2022, as seen in **Table A3** in the Appendix, we are careful to draw the conclusion that the average correlations between ESG ratings of Nordic firms have increased between 2018 and 2022.

In **Figure 4**, we have plotted average correlations between the ESG rating providers in each of the four Nordic countries included in our sample to investigate whether the disagreement differs between countries. As can be seen in the figure, the ESG rating providers seem to diverge to a similar degree when rating Nordic firms, having average correlations close to 0.5. It is noted that Sweden has the highest correlation between providers at 0.537 while Finland has the lowest correlation at 0.474. The differences between countries will be further explored in the additional analysis presented in **5.4**, by investigating whether the ESG rating disagreement on the largest firms in each country positively affects the respective stock returns.

5.3 Hypothesis 1

To test our hypothesis of whether disagreement regarding the ESG ratings of the largest firms on the Nordic markets leads to significantly higher returns for the underlying stocks, we regressed the annual stock returns – *STOCK_RETURN* – on the standard deviation of ESG ratings – *DISAGREEMENT* – and the control variables: *MC, BTM, OP, INV, MOM*. The results from the baseline model, i.e., the OLS multivariate regression computed for all the years included in the sample (2018-2022), are presented in column (1) in **Table 7**. The remaining columns in the table, columns (2-6), present the results from the same regression but on a yearly basis for each year in the sample period. The baseline model (1) indicates a positive albeit small coefficient estimate of 0.044 for the *DISAGREEMENT* variable that is insignificant, suggesting that we do not find evidence that ESG rating disagreement has a stock return implication in the Nordics. Thus, we cannot reject the null hypothesis at the 5% significance level. Why the results suggest the opposite of what was expected can be due to several reasons which are further considered in the discussion section, **6.1-6.4**.

Table 7: Main regression for the whole sample period and each respective year

Variable	2018-2022 (1)	2018 (2)	2019 (3)	2020 (4)	2021 (5)	2022 (6)
DISAGREEMENT	0.044	0.880	-0.156	-0.599	-0.015	-0.013
	(0.13)	(1.65)	(-0.24)	(-1.10)	(-0.02)	(-0.03)
MC	-0.099***	-0.051*	-0.127***	-0.146***	-0.048	0.088**
	(-4.55)	(-1.91)	(-3.30)	(-3.14)	(-0.98)	(2.28)
BTM	-0.020	-0.195**	-0.147	-0.029	0.108	0.079
	(-0.39)	(-2.31)	(-1.51)	(-0.25)	(0.60)	(0.69)
OP	0.132	-0.284	-0.122	0.351	0.378	-0.120
	(1.44)	(-1.65)	(-0.44)	(1.10)	(1.98)	(-0.56)
INV	-0.045	0.106	-0.035	-0.044	-0.002	-0.107***
	(-0.91)	(0.82)	(-0.17)	(-0.44)	(-0.01)	(-2.93)
MOM	0.095	0.045	-0.035	0.513***	0.041	-0.145
	(1.49)	(0.57)	(-0.19)	(3.05)	(0.35)	(-1.64)
Year FE	Yes	No	No	No	No	No
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.307	0.422	0.128	0.379	-0.004	0.216
Observations	479	87	102	109	112	69

Presented in the table above are the results from an OLS multivariate regression of stock returns on ESG rating disagreement and control variables across the years 2018-2022 but also multivariate regressions for each year in isolation. *DISAGREEMENT* is the standard deviation of the available ratings from all five providers at yearly intervals for each firm-year observation and requires at least three firm-year rating observations. This definition is possible since we have standardised the ratings from 0-1. *MC* is the natural logarithm of a firm's market capitalization. *BTM* is the book-to-market ratio. *OP* is the operating profitability. *INV* is the change in total assets between the fiscal years *t-1* and *t-2*, divided by *t-2* assets. *MOM* denotes momentum and is the cumulative return between the years *t-2* and *t-1*. See **Table A1** in the Appendix for exact definitions of each variable. All continuous variables have been winsorized at the 1st and 99th percentile. Fixed effects are included at the year, industry, and country dimensions to varying degrees between the regressions. T-statistics are shown in parenthesis and are based on robust clustered standard errors are at the firm level. Significance levels are indicated by ***p<0.01, **p<0.05, *p<0.1.

When studying the coefficient estimates for the yearly based regressions in **Table 7**, it is noted that the insignificance also holds for each respective year in the sample period. Noteworthy is that the coefficient estimate for the 2018 regression (2) is relatively large in magnitude at 0.880 and close to being statistically significant at the 10% level (p-value of 0.104).[6] Across all six regressions, the control variables largely have expected signs that are consistent with previous studies, as outlined in section **4.3**, however the lack of significance for most of them is evident. A possible explanation for this insignificance could be that the return predictability of widely used variables have been found to decrease post-publication (McLean & Pontiff, 2016). Since

[6] The coefficient estimate of 0.880 and *t*-statistic of 1.65 can be compared to the equivalent coefficient and *t*-statistic found in the paper by Gibson Brandon et al. (2021), that was found to be 0.698 and 1.995. Note however, that this coefficient was statistically significant at the 5% level and that the regressions are not completely comparable otherwise.

all control variables included are widely known return predictors used in peer-reviewed journals within finance, accounting, and economics, this may very well be the case for our regressions. Another explanation for the insignificance of these otherwise common stock return predicting firm characteristics could be that the sample period largely coincides with the COVID-19 pandemic. Recent research shows that return predictability underwent a structural break that was consistent timing-wise with the outbreak of the COVID-19 virus (Hong et al., 2021). Hence, the insignificance could also be attributed to the highly abnormal sample period that is investigated.

The adjusted R^2, which serves as a signal of goodness-of-fit for the empirical model, is found to be 30.7% for the baseline model (1), which largely equals levels found in previous studies.[7] With regards to the yearly regressions, there is a wide difference in the empirical fit of the specified model, with some years returning a relatively high adjusted R^2 measure such as 42.2% in 2018 (2), and one year returning a negative measure of -0.4% in 2021 (5), indicating a very poorly fitted model for that year. As can be seen in **Figure A2** in the Appendix, the goodness-of-fit of each regression seems to vary substantially between the different years in the sample.

5.4 Additional analysis

To complement the findings from the regressions presented in **Table 7**, an additional analysis has been conducted by regressing stock returns on ESG rating disagreement and the mentioned control variables for each individual country included in the sample, separately. This was done for two reasons: (1) since it can contribute to an understanding of how the disagreement may vary between the countries within the studied region and, (2) since there has, to the best of our knowledge, been limited previous literature that has conducted a comparative investigation of ESG rating disagreement in relation to stock returns on both a regional and national basis. **Table 8** shows the results from these country-by-country OLS multivariate regressions that cluster the standard errors at the firm level. Fixed effects are included at the year and industry level but not at the country level since we now examine firm-year observations within one country at a time.

[7] Gibson Brandon et al. (2021) received an adjusted R^2 of 34.7% with approximately the same control variables.

Table 8: Country-by-country regression for the whole sample period (2018-2022)

Variable	Sweden (7)	Denmark (8)	Norway (9)	Finland (10)
DISAGREEMENT	0.367	1.683*	0.161	0.226
	(0.72)	(1.82)	(0.29)	(0.24)
MC	-0.127***	0.029	-0.040	-0.032
	(-3.96)	(0.43)	(-1.17)	(-0.70)
BTM	0.009	0.131	-0.048	0.217
	(0.08)	(0.19)	(-0.70)	(1.28)
OP	0.465	-0.121	0.419	0.578
	(1.52)	(-0.52)	(1.48)	(0.70)
INV	-0.092	-0.114	-0.068	-0.025
	(-1.35)	(-0.82)	(-0.61)	(-0.69)
MOM	0.135	-0.070	-0.006	-0.077
	(1.64)	(-0.51)	(-0.05)	(-0.61)
Year FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Country FE	No	No	No	No
Adj. R^2	0.394	0.200	0.537	0.082
Observations	236	96	75	72

Presented in the table above are the results from an OLS multivariate regression of stock returns on ESG rating disagreement and control variables across the years 2018-2022. However, in contrast to **Table 7**, the disagreement is now investigated at a country-basis. *DISAGREEMENT* is the standard deviation of the available ratings from all five providers at yearly intervals for each firm-year observation and requires at least three firm-year rating observations. This definition is possible since we have standardised the ratings from 0-1. *MC* is the natural logarithm of a firm's market capitalization. *BTM* is the book-to-market ratio. *OP* is the operating profitability. *INV* is the change in total assets between the fiscal years *t-1* and *t-2*, divided by *t-2* assets. *MOM* denotes momentum and is the cumulative return between the years *t-2* and *t-1*. See **Table A1** in the Appendix for exact definitions of each variable. All continuous variables have been winsorized at the 1st and 99th percentile. Fixed effects are included at the year, industry, and country dimensions to varying degrees between the regressions. T-statistics are shown in parenthesis and are based on robust clustered standard errors at the firm level. Significance levels are indicated by ***p<0.01, **p<0.05, *p<0.1.

Interestingly, the regression on the Danish firms (8) indicates a high coefficient estimate of 1.683 that is statistically significant at the 10% significance level. This suggests that ESG rating disagreement may have a positive return implication in Denmark. However, the adjusted R^2 of the model is relatively low at 20.0%. Therefore, we acknowledge that the estimate should be interpreted with a degree of prudency. The regression on the Swedish firms (7) yields a statistically insignificant coefficient estimate for the *DISAGREEMENT* variable, which is also the case for the Norwegian (9) and Finnish (10) firms. Noteworthy is that the adjusted R^2 for the Norwegian regression is quite high at 53.7%, relative to the other regressions and to the ones received in previous literature (Gibson Brandon et al., 2021). This suggests that 53.7% of the sample variation of stock returns in Norway can be explained by the independent variables. Comparably, the adjusted R^2 for the Finnish regression is very low, at only 8.2% but with a similar number of observations, which suggests that there are other independent variables that

explain stock returns on the Finnish market better. As was the case for the regression results in **Table 7**, there is again a lack of significance for all the control variables in all four regressions which can partly be explained by the same stream of literature regarding fall offs in return predictability (Hong et al., 2021; McLean & Pontiff, 2016).

6. Discussion

In this section, we discuss and provide possible explanations for our results from both an empirical and theoretical perspective. Firstly, in section **6.1**, the regression results are scrutinised from an empirical standpoint by considering choices and assumptions that may have impacted our results, as well as testing for violations of critical model assumptions. Secondly, a potential rationalisation of the results related to measurement aggregation and risk diversification is presented in section **6.2**. Thirdly, previous presented theories are revisited in section **6.3** to provide an alternative possible explanation of the retrieved results. Lastly, the section ends with a discussion about the validity and reliability of the results, presented in section **6.4**.

6.1 Robustness and reliability of assumptions

The ordinary least squares (OLS) estimate for regression coefficients produces the best linear unbiased estimate (BLUE) if certain Gauss-Markov assumptions are satisfied (Wooldridge, 2012).[8] Therefore, to avoid violations of these assumptions and to eliminate as much bias as possible in the models, we investigated the robustness of our results with statistical tests and adjusted the empirical method when necessary.

Firstly, an OLS regression assumes that the error term has constant variance, such that homoskedasticity holds. When the model instead exhibits unequal variance for the error term and there is heteroskedasticity, the standard errors and test statistics are no longer valid (Wooldridge, 2012). In other words, if heteroskedasticity is present in the data, the standard errors cannot be trusted to construct confidence intervals and *t* statistics. Thus, we performed a Breusch-Pagan Cook-Weisberg test with the null hypothesis that there is constant variance among the residuals, against the alternative hypothesis that the residuals are not distributed with equal variance. With a χ^2- value of 62.74 and p=0.000, we rejected the null hypothesis and concluded that heteroskedasticity was present in the data. To correct for this, we included robust standard errors in all regression models.[9]

Secondly, when the error terms in panel data are not independent over time, it causes autocorrelation, also known as serial correlation. This is problematic since it can cause biases

[8] Specifically, the assumptions are: (1) the underlying population model is linear in parameters, (2) a random sample is retrieved from the population, (3) independent variables are not perfectly correlated, (4) no important variable is omitted, and (5) the error term has constant variance (Wooldridge, 2012).
[9] See **Figure A2** in the Appendix for scatter plots of the residuals.

within the standard errors which impacts the efficiency of the results (Wooldridge, 2012). Therefore, a Wooldridge test for autocorrelation in panel data was conducted with the null hypothesis of no first-order autocorrelation. With a F=43.348 and p=0.000, the null hypothesis was rejected.[10] Efforts to minimise the adverse effect of this included clustering the standard errors at the firm level and including a lagged version of the dependent variable as a control variable – the momentum factor, conceptualised by Jegadeesh and Titman (1993).

Multicollinearity is a problem that can occur when the independent variables in a regression model are highly, but not perfectly, correlated which may cause the R^2 to be "close" to 1 (Woolridge, 2012). This may undermine the statistical significance of the independent variables, distorting the coefficients and inflating the confidence intervals. While multicollinearity is not a violation of the OLS assumptions, it is still necessary to control for since the aim of a regression model is to isolate the relationship between each independent variable and the dependent variable. With multicollinearity present, the regression model does not fulfil that aim. In **Table 5**, the Pearson coefficient estimates between the variables are presented, and it was noted that the highest correlation was between *INV* and *MOM* at 0.329. To further investigate whether multicollinearity was present in our model, we conducted a variance inflation factor (VIF) test. The VIF measure for the initially included control variable "market return" was close to 10 and thus potentially highly collinear with the other independent variables. After excluding it from the model, the rest of the independent variables exhibited VIF scores that were lower than 5, well below the traditional cut-off value of 10 (Wooldridge, 2012). Therefore, we concluded that we did not have a problem with multicollinearity and as such, *INV* and *MOM* could both be kept as control variables. The mentioned VIF scores are presented in **Figure A3** in the Appendix.

Since the explanatory power, as indicated by the adjusted R^2, is relatively low in all regression models, there seems to be other variables explaining stock returns that are not included in the model. Therefore, our model may be subject to an omitted variable bias, and we may have a problem with an endogeneity. However, as discussed in **5.3**, the levels of explanatory power in our regression results largely resemble the levels retrieved in previous research within the field (Gibson Brandon et al., 2021) and as such, we do not consider this to be a problem when testing the hypothesis.

[10] Note that this test was performed prior to including the momentum factor and clustering the standard errors.

6.2 Measurement aggregation and risk diversification

As described in the research method section, **4.2**, ESG rating disagreement was defined as the standard deviation of ESG ratings for a specific firm-year observation from at least three separable rating agencies. The rationale behind this definition was to restrict the sample for the average computed disagreement observation to include enough providers to lower the risks of biases stemming from the agencies' differences and similarities in methodologies. For reference, the average number of providers for each disagreement observation was 3.356 without restriction, and 4.013 with a restriction to only include firm-year observations with at least three providers. With an average of approximately four out of five providers per observed disagreement, the linear relationship between a firm's cross-sectional stock performance and its ESG rating disagreement should thus be based on a, on average, quite high proportion of the available ESG rating providers at 80%.

Revisiting the literature on the drivers of ESG rating divergence by Berg et al. (2022) and considering each rating provider's methodology, as described in **Table A2** in the Appendix, a possible methodology-related explanation of our results may be found. We consider that in general, ESG scores are built based on scopes, or categories, that are broken down in measurements consisting of different attributes and indicators. These measurements are then given different weights before being aggregated into a total ESG score, which are the ones used in this study. Each firm's aggregated ESG rating is thus based on hundreds of measurements, as exemplified by Refinitiv covering 630 firm-year observations while FTSE measures over 300 firm-year indicators (see **Table A2** in the Appendix). Furthermore, considering that these hundreds of measurements, in our setting, are multiplied with 4, the average number of agencies per observation, which holds for all 479 disagreement observations in the regression, the impact of these measurement- and respective weight divergences are, in accordance with the Law of Large Numbers, expected to cancel out.

Indeed, Berg et al. (2022) find that correlations between the categories of different ESG rating agencies tend to increase with measurement granularity, which further strengthens and develops this argument. In their paper, they propose the following reasoning: "One reason might be that category scores behave like noisy measures of an underlying latent quality so that the measurement disagreement on individual categories cancels out during aggregation." (Berg et al., 2022, p.1330). This supports the logic of why the divergence stemming from differences in measurements and weights should converge when aggregated, as discussed above, but also

shows that the disagreement driver of scope, or "category scores", may converge as a result of aggregation.

Reconsidering the concept of ESG disagreement as a risk, described in section **2.2.2**, the aggregation of divergences in measurements of ESG ratings and its implication on stock returns may be explained. Viewing the relation between ESG rating disagreement and stock returns as several relationships between the individual drivers of the disagreement and the underlying stock would imply that each driver imposes a risk to investors. When these drivers subsequently are aggregated into the total ESG score, it may thus be the case that the disagreement cancels out and the associated risks are diversified. This assumes that these risks are idiosyncratic which seems a reasonable assumption considering that these risks are specific to each measured metric. As such, when the relationship between ESG rating disagreement and stock returns is broken down into specific drivers of disagreement, it can be argued that the related risks will be diversified and thus not impose an impact on the stock. Furthermore, the cancelling out of the disagreement and the diversification of the risks does not necessarily have to be at the same level. For example, it may be the case that the disagreement cancels out partly while the associated idiosyncratic risks are diversified away almost completely. This would help explain our results since we find disagreement that has no significant impact on the stock returns.

While this partly explains why we could not find a significant positive relationship between ESG rating disagreement and stock returns, it does not consider the effects of treating disagreement as uncertainty. According to the reasoning in section **2.2.2**, part of the observed disagreement may stem from weight divergences and the rater effect, which we theorised to be classified as uncertainty. Indeed, Berg et al. (2022) suggest that the rater effect cannot be thought of as noise that will disappear when the ratings are aggregated. As such, if weight divergences and the rater effect impacted our data, they should theoretically have caused a positive impact on the stock returns. Nonetheless, considering that we did not find a significant relationship between ESG rating disagreement and stock returns, it may either be the case that the disagreement in our sample is mostly attributed to risks, which were then diversified, or that the empirical model used was not capable of encapsulating uncertainty. The latter would be in line with the findings of Anderson et al. (2009) who advocate using a two-factor model with a proxy for uncertainty, to fully capture both risks and uncertainties.

6.3 Financial materiality of ESG ratings and increased analyst awareness

While the risk diversification of each rating disagreement driver at the aggregated level may serve as an explanation of the presented results, other possible theoretical justifications of the results exist. One of the theorised explanations for why disagreement should be priced into stocks was because it could be interpreted as a proxy for firm-specific risk for which investors demand a premium. Specifically, measurement- and scope differences between rating providers were brought forward as examples of underlying risk sources. A possible explanation for why we cannot identify such a risk premium could be that all underlying factors that comprise ESG ratings are not always financially material to start with, as explained in section **2.2.3**. As such, disagreement will not be perceived as a risk that should affect the stock price. In a similar manner, weight divergences between rating agencies and a human judgement so-called rater effect were brought forward as examples of uncertainty regarding the ESG performance of a firm that investors demand an uncertainty premium for. Again, if ESG ratings are not financially material on a standalone basis, disagreement stemming from weight differences between rating providers or the rater effect should not infer an uncertainty premium.

However, while the thus far presented theories may provide possible explanations to the retrieved results in this study, they do not help to explain why previous similar research has retrieved results that do in fact suggest that ESG rating disagreement has a return implication. As noted in **2.1.3**, Gibson Brandon et al. (2021) find statistical evidence that greater disagreement in ESG ratings is positively associated with higher stock returns for firms on the S&P 500. An apparent difference between our study and theirs, is that we study another geographic region suggesting that regional differences in reporting requirements may serve as one explanation. In a report published by the American Council for Capital Formation, Doyle (2018) documented a notable disparity in average ratings between European and North American firms among the 4,150 companies investigated by Sustainalytics, indicating tendencies of a bias. This bias is explained to not reflect the quality of ESG practices but rather the quality of reporting. While European firms with more than 500 employees are mandated to publish a non-financial statement, North American firms are not (European Commission, 2023; Doyle, 2018). In line with the findings of Christensen et al. (2022), higher levels of disclosure for European firms and especially Nordic firms, as discussed in section **2.2.1**, could therefore explain part of the observed rating disagreement in this study. However, since this bias of rewarding Nordic firms with higher ratings may manifest itself differently between agencies, the arising disagreement should not be financially material and not affect the stock price.

Therefore, while Gibson Brandon et al. (2021) find rating disagreement to positively impact stock returns for US firms, for which this bias of inflated ratings does not exist, part of the disagreement found in this study on Nordic firms may be financially immaterial and thus not expected to have a stock return implication.

A complementing possible explanation for the contrasting results could be attributed to the significant advancements in ESG awareness among analysts in the Nordics in recent years. For instance, the number of EFFAS Certified ESG Analysts® increased by 47% in Sweden between 2022 and 2023 to 129 certified analysts as of 27th of March 2023 (European Federation of Financial Analysts Societies [EFFAS], 2023). This suggests that analysts in the region should be increasingly aware of the phenomenon of ESG rating disagreement and how to interpret conflicting ESG data in their valuations. An understanding of the underlying elements of why disagreement exists, for example reasons that per se do not reflect differences in ESG performance – like scope differences – can therefore result in a more effective incorporation of ESG ratings into stock prices. This potentially increased understanding could therefore help explain why ESG rating disagreement does not have a significant return implication in the Nordics. Additionally, since the sample period covered in this study is more recent than the one used in Gibson Brandon et al. (2021), this awareness may alternatively not even be restricted to a particular geographical region but rather hold generally in different parts of the world.[11]

6.4 Validity and reliability of results

The results presented in this study stand in contrast to the outlined hypothesis development and to previous research within the field and thus, there is an evident need to investigate whether the findings are valid and reliable. Ultimately, this determines the extent to which our findings could potentially be generalised and used by the financial and corporate sector, or within related fields of research.

In terms of the validity of this study, we question the relevance of the data in testing our hypothesis and the performance of our empirical model. To minimise potential sample biases and modelling errors, we gathered a sample that was random in terms of ESG ratings and considered potential issues stemming from differences and similarities in ESG rating methodologies. We also included predominant factors used to explain the cross-section of returns as control variables, while controlling for fixed effects related to countries, years, and

[11] Gibson Brandon et al. (2021) use the sample period 2010-2017 whereas this study investigates a more recent period, 2018-2022.

industries. Finally, with respect to the ordinary least squares model used, we included clustered robust standard errors, a control variable to mitigate the autocorrelation in stock returns and excluded a control variable that exhibited high multicollinearity. Nonetheless, biases may still have existed since the dataset can be considered small in relation to previous studies within the domain. For example, in comparison to Gibson Brandon et al. (2021), our main regression included 479 total observations while theirs had a total of 35,139 observations.[12]

Considering the reliability of our study, one should, given the data, assumptions, and method described above be able to replicate it and find similar results. Though there are possibilities of potential measurement errors related to the data management, regressions, and interpretations, we argue that the main source of risk regarding the reliability of this study is derived from the quality of the data. Apart from the fact that the dataset may be considered small, the five-year sample period covered in this study included at least two important macroeconomic shocks that may have impacted our results, namely the COVID-19 pandemic during 2020-2021 and the Russo-Ukrainian war during 2022-2023. This has two important implications on the reliability of our results. Firstly, the crises may have had proportionally different effects on the variables in the model suggesting that we cannot fully ensure that it accurately captures the relationship between the variation in ESG ratings and stock returns. This may partly be due to structural breaks in return predictability as discussed in section **5.3**. Secondly, the generalisability of our results is hampered since these results may not be applicable to sample periods excluding these crises. However, while the sample period may have disproportionately affected the variables in our regressions, the reliability of the ESG rating data is considered high since the Pearson correlations between the rating providers was comparable to previous literature, as concluded in section **5.2**.

[12] This difference is explained by three major reasons: (1) they were able to find more granular ESG data and were able to compute the disagreement on a monthly basis while we could retrieve yearly ESG ratings, (2) they investigated firms on the S&P 500 meaning that they studied a greater number of firms larger in size and thus with a higher likelihood of being rated, and (3) they conducted their study on a sample of seven years, while data access restricted us to five years.

7. Conclusion

In recent years, the emergence of ESG ratings has become a heavily discussed topic, especially considering that the lack of a common definition for what ESG constitutes causes rating agencies to disagree in their assessments of the same firm. This study aims to investigate whether the evident and widespread rating disagreement has any financial consequences for those who increasingly rely on them such as firms, analysts, and investors. In particular, it is investigated whether ESG rating disagreement affects stock returns for the largest firms on the Nordic markets. To address the research question, a multivariate OLS regression was conducted for stock returns on ESG rating disagreement while controlling for standard firm characteristics that are known return predictors. By comparing Pearson correlations between the ESG rating providers to levels retrieved in previous literature and correlations between credit ratings, we conclude that ESG rating disagreement exists in the Nordic markets. However, we do not find statistical support for a positive relation between such rating disagreement and stock returns at the 5% significance level. Complementing the findings from our main analysis, we also investigate the relationship on a country-by-country basis and find mostly similar results. While we do find a positive and significant relationship between ESG rating disagreement and stock returns for the Danish market at the 10% level, we conclude that the statistical inference should be interpreted with caution since the model may suffer from issues related to endogeneity.

Possible explanations of these findings can be both empirical and theoretical. A relatively restricted sample both in terms of the number of investigated firms and years covered – to encompass as many rating providers as possible – may infer problems with the power of our tests. Theoretically, when the drivers of rating disagreement are aggregated, the divergences may on average cancel out, diversifying the associated risks and impeding the risk-return impact on the underlying stock. Alternatively, the bias of rewarding Nordic firms with higher ratings due to higher levels of disclosure may give rise to disagreement that is not financially material, therefore not affecting the stock price. Additionally, analysts may be better positioned and equipped to decipher the disagreement that has financial materiality from that which does not, and effectively incorporate it in their valuations. Overall, though definite statistical conclusions cannot be drawn from the results, our findings indicate that ESG rating disagreement does not have a return implication on the Nordic markets which stands in contrast to what Gibson Brandon et al. (2021) found on the US market.

The implications of this study may be of interest for financial analysts, companies, and investors. Since we do not find a significant relationship between stock returns and ESG rating

disagreement, there should reasonably not exist scenarios in which companies are incentivised to disclose more ESG information with the sole purpose of increasing the stock price and indirectly generating ESG rating disagreement. In other words, CEOs and management should not have the incentive to mislead analysts and rating agencies by disclosing ESG information, that is not necessarily material, and create "ESG rating risks" to utilise a risk-return relationship. As such, firms should rationally only be concerned with disclosing ESG activities that regard actual improvements to increase their ESG rating rather than creating ambiguity and engaging in greenwashing. With this said, since we find rating disagreement in the Nordics, firms may still be confused as to what ESG activities are valued by the market and what truly resemble improvements. These mixed signals thus seem to continue to disincentivise actions to improve ESG related activities.

7.1 Research contributions

Through benchmarking our research question, method, sample, and findings against the existing stream of literature, we conclude that our study provides four relevant research contributions to the literature of ESG rating disagreement. Firstly, as a response to the study limitations in the paper by Gibson Brandon et al. (2021), we set out to investigate whether their results can be transposed to stock markets outside the United States for which we studied the Nordic stock markets. Furthermore, we also analysed and compared the stock return impact of ESG rating disagreement on a country-by-country basis to find potential differences between markets. Secondly, our research is conducted on a sample period between 2018 and 2022 which is more recent and does not overlap with the period of 2010-2017, studied in the paper by Gibson Brandon et al. (2021). Thirdly, this study contributes with a first attempt at linking together current prevalent research within the field. Viewing the relationship between ESG rating disagreement and stock returns as either risk or uncertainty, in line with Gibson Brandon et al. (2021), we additionally utilise the classifications of Berg et al. (2022) to break these relationships down into underlying disagreement drivers. Lastly, as ESG ratings gain increasing traction as a relatively new phenomenon, this study contributes with a summarising overview of the current Nordic ESG rating landscape by presenting data from multiple vendors and mapping out the rating divergence observed in recent years.

7.2 Suggestions for further research

During the process of conducting this study, several interesting ideas for further research have unfolded. Firstly, it may be of interest to study a larger cross-section of returns, possibly globally, to see whether ESG rating disagreement exists for all types of firms regardless of size, and whether such disagreement has any return implications. An obvious challenge with investigating smaller firms is the lack of available ratings, insinuating that such further research relies on vast data access. In this regard, we also acknowledge that the restricted data access is one of the limitations of this study. Secondly, another avenue for further research could be to investigate how much of ESG rating disagreement that can be attributed to either risk or uncertainty, that in turn is priced into stock returns. While we theorise about the effect of these underlying drivers in this study, we have not been able to quantify it. As such, researching the financial implications of each underlying driver of ESG rating disagreement would be interesting. Thirdly, since the ESG landscape is developing rapidly and growing in practitioners and followers, it is of large interest to continue study this relationship in the coming years to see whether ESG rating disagreement persists and whether it has financial implications for the firms exhibiting it.

References

Amel-Zadeh, A., & Serafeim, G. (2018). Why and How Investors Use ESG Information: Evidence from a Global Survey. *Financial Analysts Journal, 74*(3), 87-103.

Anderson, E. W., Ghysels, E., & Juergens, J. L. (2005). Do Heterogeneous Beliefs Matter for Asset Pricing? *The Review of Financial Studies, 18*(3), 875-924. 10.1093/rfs/hhi026

Anderson, E. W., Ghysels, E., & Juergens, J. L. (2009). The impact of risk and uncertainty on expected returns. *Journal of Financial Economics, 94*(2), 233-263. 10.1016/j.jfineco.2008.11.001

Atmaz, A., & Basak, S. (2018). Belief Dispersion in the Stock Market. *The Journal of Finance (New York), 73*(3), 1225-1279. 10.1111/jofi.12618

Avramov, D., Cheng, S., Lioui, A., & Tarelli, A. (2022). Sustainable investing with ESG rating uncertainty. *Journal of Financial Economics, 145*(2), 642-664. 10.1016/j.jfineco.2021.09.009

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

Berg, F., Kölbel, J.,F., & Rigobon, R. (2022). Aggregate Confusion: The Divergence of ESG Ratings. *Review of Finance, 26*(6), 1315-1344. 10.1093/rof/rfac033

Bloomberg. (2023). *ESG Data*. Retrieved February 20, 2023, from https://tinyurl.com/yyu99rrp

Brackley, A., Brock, E., Nelson, J. (2023). *Rate the Raters 2023: ESG Ratings at a Crossroads*. SustainAbility. https://tinyurl.com/yc3fcsxn

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 (John Wiley & Sons, Inc.), 37*(8), 1597-1614. 10.1002/smj.2407

Christensen, D. M., Serafeim, G., & Sikochi, A. (2022). Why is Corporate Virtue in the Eye of The Beholder? The Case of ESG Ratings. *Accounting Review, 97*(1), 147-175. 10.2308/TAR-2019-0506

Doyle, T. M. (2018). *Ratings That Don't Rate*. American Council for Capital Formation. https://tinyurl.com/5ycpyymr

European Commission. (2023). *Corporate sustainability reporting*. Retrieved April 25, 2023, from https://tinyurl.com/23hw74vs

European Federation of Financial Analysts Societies. (2023, March 27). *EFFAS Certified ESG Analyst® (CESGA) Holders List & Testimonials*. Retrieved April 7, 2023, from https://tinyurl.com/mpy5ybcy

European Securities and Markets Authority. (2021). *ESMA Report on Trends, Risks and Vulnerabilities* (No 1). https://tinyurl.com/mrx7h8br

Fama, E. F., & French, K. R. (1992). The Cross-Section of Expected Stock Returns. *The Journal of Finance, 47*(2), 427-465. 10.2307/2329112

Fama, E. F., & French, K. R. (2004). The Capital Asset Pricing Model: Theory and Evidence. *Journal of Economic Perspectives, 18*(3), 25-46. 10.1257/0895330042162430

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1-22. 10.1016/j.jfineco.2014.10.010

FTSE Russell (2020). *Integrating ESG into investments and stewardship*. https://tinyurl.com/4wm4cme5

FTSE Russell. (2023). *ESG Scores*. Retrieved February 20, 2023, from https://tinyurl.com/7r38mhsx

Gibson Brandon, R., Krueger, P., & Schmidt, P. S. (2021). ESG Rating Disagreement and Stock Returns. *Financial Analysts Journal, 77*(4), 104-127. 10.1080/0015198X.2021.1963186

Global Risk Profile. (2021, December 10). *Nordic countries at the top of the 2021 ESG Index* [Press release]. Retrieved from https://tinyurl.com/2p92szpy

Global Risk Profile (2023, March 8). *ESG Index 2022: a global decline of sustainability performance* [Press release]. Retrieved from https://tinyurl.com/mw6p2jnp

Grewal, J., Hauptmann, C., & Serafeim, G. (2021). Material Sustainability Information and Stock Price Informativeness. *Journal of Business Ethics, 171*(3), 513-544. 10.1007/s10551-020-04451-2

Hong, H., Bian Zhicun, & Chien-Chiang, L. (2021). COVID-19 and instability of stock market performance: evidence from the U.S. *Financial Innovation, 7*(1)10.1186/s40854-021-00229-1

Jegadeesh, N., & Titman, S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. Journal of Finance (Wiley-Blackwell), 48(1), 65-91. 10.1111/j.1540-6261.1993.tb04702.x

Khan, M., Serafeim, G., & Yoon, A. (2016). Corporate Sustainability: First Evidence on Materiality. *Accounting Review, 91*(6), 1697-1724. 10.2308/accr-51383

Knight, F. H. (1921). *Risk, uncertainty and profit*. Houghton Mifflin.

McLean, R. D., & Pontiff, J. (2016). Does Academic Research Destroy Stock Return Predictability? *Journal of Finance (John Wiley & Sons, Inc.), 71*(1), 5-32. 10.1111/jofi.12365

MSCI. (2020). *MSCI ESG Ratings Methodology*. https://tinyurl.com/3tkmnctv

Principles for Responsible Investment. (2022). *Annual Report 2022.* https://tinyurl.com/mu658rxw

Refinitiv. (2022). *Environmental, social and governance scores from Refinitv*. https://tinyurl.com/w4eet767

Shanaev, S., & Ghimire, B. (2022). When ESG meets AAA: The effect of ESG rating changes on stock returns. *Finance Research Letters, 46*, 102302. 10.1016/j.frl.2021.102302

SolAbility. (2022). *The Global Sustainable Competitiveness Index 2022* (11th edition). https://tinyurl.com/bdf9zun8

S&P Global. (2022a). *Environment, Social, and Governance Evaluation: Analytical Approach*. https://tinyurl.com/y5w9pdty

S&P Global. (2022b). *ESG Evaluation*. [Brochure]. https://tinyurl.com/53xa32jn

S&P Global. (2023). *ESG Scores*. Retrieved February 22, 2023, from https://tinyurl.com/4hzb5ht3

Wooldridge, J. M. (2012). *Introductory econometrics: a modern approach*. Cengage.

Appendix

Table A1: Variable description

Variable	Description	Definition	Source
Dependent variable			
STOCK_RETURN	Stock return	Calendar year annualised stock returns computed as the percentage change between stock prices	FactSet
Independent variable			
DISAGREEMENT	ESG rating disagreement	The standard deviation of the available ratings from all rating providers at yearly intervals for each firm-year observation, requiring at least three firm-year rating observations to be computed	Refinitv, S&P Global, FTSE Russell, MSCI, Bloomberg
Control variables			
MC	Market capitalization	Natural logarithm of stock price multiplied by shares outstanding at the end of December of year *t-1*	S&P Capital IQ
BTM	Book-to-market ratio	Book value of equity for the fiscal year *t-1* divided by market capitalization at the end of December of year *t-1*	S&P Capital IQ
OP	Operating profitability	EBIT for the fiscal year *t-1* minus interest expense for the fiscal year *t-1* divided by book value of equity	S&P Capital IQ
INV	Investment	The change in total assets between the fiscal years *t-1* and *t-2*, divided by *t-2* assets	S&P Capital IQ
MOM	Momentum	Cumulative return between the years *t-2* and *t-1* computed based on stock prices from the end of December each year	FactSet

The above table lists, describes and defines all variables used in our empirical model while also indicating their respective sources. The variables are classified in three groups: the dependent variable, the independent variables, and the control variables.

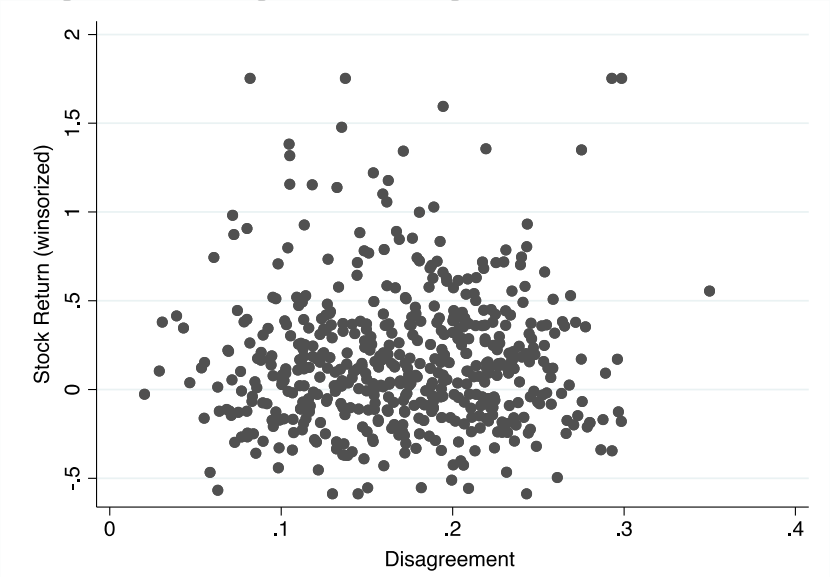
Table A2: ESG rating methodologies

ESG rating provider	Summarised methodology
Refinitv ESG Scores (Eikon)	Refinitiv provides ESG ratings for over 12,500 companies globally, covering over 85% of the global market capitalization. They produce scores using a 0-100 rating scale and sources their ESG data from annual reports, company websites, NGO websites, stock exchange filings, CSR reports, and news sources. Over 630 ESG measures are covered, and as the importance of ESG factors differs across industries, they are given different weights depending on the industry. The Refinitv ESG Combined Score is an overall company score that incorporates the reported information for each ESG pillar but also an ESG Controversies overlay. This controversy score is calculated based on 23 ESG controversy topics and aims to adjust the overall score if negative events like a scandal occur (Refinitv, 2022).
S&P Global ESG Scores	S&P Global provides ESG ratings for more than 11,500 companies representing 99% of global market capitalization (S&P Global 2023). Their rating is based on a 0-100 rating scale and their ESG data is sourced from their own web-based questionnaire – the S&P Global Corporate Sustainability Assessment (CSA) comprised of 600-1,000 total data points – but also from verified company disclosures and media/stakeholder analysis (S&P Global 2022b, 2023). The final ESG evaluation ranking is a combination of an entity's "ESG Profile" with its long-term "Preparedness". The ESG Profile assesses the exposure of the entity's operations to observable ESG risks and opportunities and includes the forward-looking effects of material ESG events. The Preparedness component assesses a company's capacity to anticipate and adapt to a variety of long-term plausible disruptions and emerging trends (e.g., environmental scenarios or technological/regulatory changes) (S&P Global 2022a).
FTSE Russell ESG	FTSE Russell provides ESG ratings for approximately 7,200 companies in 47 developed and emerging markets that are constituents of the FTSE All-World Index, FTSE All-Share Index and Russell 1000® Index (FTSE Russell 2023). FTSE Russell uses a 0-5 rating scale to 1 decimal point and builds their scores based on over 300 individual indicator assessments (FTSE Russell 2020). Their ESG company research relies on publicly disclosed information only and as such they do not send questionnaires to issuers. The most material ESG issues are given the most weight when determining a company's score, suggesting that the scores are calculated using an exposure-weighted average. The overall rating is built up by scores and exposures of both the individual pillars and identified themes. As such, their ratings encapsulate a company's exposure to, and management of, ESG issues in multiple dimensions.
MSCI ESG Ratings & Climate Search Tool	MSCI's public ESG Ratings & Climate corporate search tool provides ESG ratings for over 2,900 companies that are constituents of the MSCI All Country World Index (ACWI) (MSCI, 2020). This public version only displays a subset of the companies in their coverage universe. The ratings range from Laggard (CCC, B), Average (BB, BBB, A) to Leader (AA, AAA), and are intended to be interpreted relative to a company's industry peers. Their ratings are based on collected and standardised public data from: 1) government-, regulatory- and NGO-datasets, 2) company disclosures and documents, and 3) 3,400 media sources. MSCI's ratings aim

	to measure a company's resilience to long-term financially relevant ESG risks by considering material ESG risks and opportunities facing the company and its industry. The final score is a combination of "Key ESG Issues" and weights.
Bloomberg ESG Disclosure Score	Bloomberg provides ESG ratings for 15,000 companies in over 100 countries, covering roughly 88% of the global equity market capitalization (Bloomberg, 2023). Their ESG Disclosure Score can be summarised as disclosure-oriented since they use a 0-100 rating scale where a 0 is given to firms that do not disclose any of the ESG data included in the score and a 100 to those that disclose every data point. The scores are based on the ESG data companies report publicly and does not measure companies' performance on any data point. Environmental (E), social (S), and governance (G) pillars are equally weighted within the overall "ESG Disclosure Score", while topic weights are allocated across fields related to the issue, with quantitative fields weighted more heavily than binary fields.

This table summarises the methodologies used by each ESG rating provider included in the study.

Figure A1: Scatter plot of the independent and dependent variable



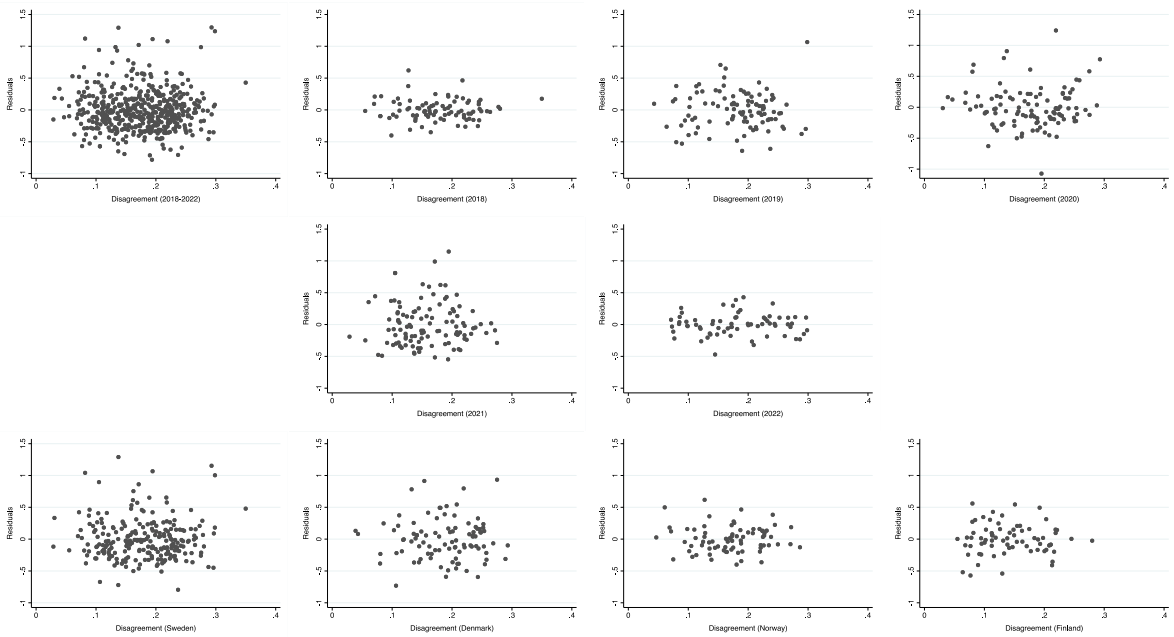
The figure above visualises the relationship between the winsorized stock returns (at the 1st and 99th percentiles) and the disagreement variable. Stock returns are computed on an annual basis and ESG rating disagreement is defined as the standard deviation of the available ratings (standardised between 0-1) from all five providers at yearly intervals for each firm-year observation and requires at least three firm-year rating observations.

Table A3: Number of observations per ESG rating agency over time

	N					
Variable	**2018**	**2019**	**2020**	**2021**	**2022**	**Total**
Refinitiv Eikon	126	141	145	146	0	558
S&P Global	96	111	117	118	118	560
FTSE Russell	58	65	75	91	90	379
MSCI	58	76	72	67	77	350
Bloomberg	139	145	145	146	28	603
Total	477	538	554	568	313	2450
DISAGREEMENT	109	119	122	123	74	538

The table above shows the number of observations per ESG rating agency over time as well as the total number of observations for each agency and year. Note that this is all the firm-year rating observations we could retrieve from each agency. The table also includes the number of observations of *DISAGREEMENT* for each year, i.e., the standard deviation between at least three rating providers.

Figure A2: Scatter plots of residuals



The figure above presents scatter plots of the residuals from each OLS multivariate regression of stock returns on ESG rating disagreement and control variables, presented in **Table 7** and **Table 8** respectively. ESG rating disagreement is defined as the standard deviation of the available ratings (standardised between 0-1) from all five providers at yearly intervals for each firm-year observation and requires at least three firm-year rating observations. The top left graph plots the residuals for the baseline regression, 2018-2022, followed by each yearly regression in isolation. The bottom row subsequently presents the residuals for the regressions of each country.

Figure A3: VIF multicollinearity tests

Panel A

Variable	VIF	1/VIF
Disagreement	1.19	0.838338
adj_MR	8.54	0.117054
adj_lnMC	1.49	0.669604
adj_BTM	2.31	0.433471
adj_OP	1.52	0.656256
adj_INV	1.24	0.808109
adj_MOM	1.40	0.712907
Country		
2	1.79	0.557819
3	1.39	0.718879
4	1.37	0.729751
Year		
2019	7.14	0.139961
2020	3.93	0.254452
2021	9.39	0.106541
2022	1.73	0.578493
Industry_11		
2	5.83	0.171550
3	8.75	0.114243
4	3.41	0.293147
5	3.90	0.256333
6	5.67	0.176306
7	4.23	0.236598
8	3.46	0.289211
9	2.57	0.388651
10	1.96	0.509314
11	3.75	0.266852
Mean VIF	3.67	

Panel B

Variable	VIF	1/VIF
Disagreement	1.19	0.838494
adj_lnMC	1.49	0.669958
adj_BTM	2.31	0.433747
adj_OP	1.52	0.656280
adj_INV	1.23	0.814936
adj_MOM	1.39	0.718922
Country		
2	1.57	0.638379
3	1.39	0.719231
4	1.34	0.745476
Year		
2019	1.80	0.556427
2020	1.85	0.541994
2021	1.84	0.544597
2022	1.73	0.579545
Industry_11		
2	5.83	0.171606
3	8.74	0.114386
4	3.41	0.293489
5	3.90	0.256389
6	5.67	0.176406
7	4.22	0.236745
8	3.45	0.289957
9	2.57	0.388728
10	1.96	0.509473
11	3.75	0.266855
Mean VIF	2.79	

(1) The figure above presents the VIF tests of the baseline model in **Table 7** – an OLS multivariate regression of the annual stock returns on *DISAGREEMENT* and control variables – before and after including market return, *MR*, as shown in Panel A and Panel B respectively. All control variables have been winsorized and are thus denoted with the prefix "adj" while MC also has been applied to a natural logarithmic scale and thus have the additional prefix "ln".
(2) Market return was considered as a control variable since it was the first documented predictor of measuring risk and the relation between expected returns and risks for stocks (Fama & French, 2004). The variable was constructed by downloading the annual returns of each respective market proxy; OMX Stockholm 30 for Sweden, OMX Copenhagen 20 for Denmark, Norway OSE OBX 25 for Norway, and OMX Helsinki 25 for Finland using FactSet. In the end, it was excluded from the model since it was highly multicollinear with the other variable in the regression (VIF score of 8.54).