Lisbon School of Economics & Management
Universidade de Lisboa

MASTER

ACCOUNTING

MASTER´S FINAL WORK

DISSERTATION

FIRM-LEVEL DRIVERS OF ESG-RATING DIVERGENCE

LARISSA RUMES

SUPERVISION:
CRISTINA GAIO

01-2025

Lisbon School of Economics & Management
Universidade de Lisboa

MASTER

ACCOUNTING

MASTER´S FINAL WORK

DISSERTATION

FIRM-LEVEL DRIVERS OF ESG-RATING DIVERGENCE

LARISSA RUMES

SUPERVISION:

CRISTINA GAIO

01-2025

GLOSSARY

CapEx	–	Capital Expenditure.
CEO	–	Chief Executive Officer.
ESG	–	Environmental, Social, Governance.
ESG-score	–	Environmental, Social, Governance Score.
ESGC-score	–	Environmental, Social, Governance Controversies Score.
EU	–	European Union.
FE	–	Fixed Effects.
FTSE	–	Financial Times Stock Exchange Group; ESG-rating provider.
GHG	–	Greenhouse Gas.
ISS	–	Institutional Shareholder Services Inc.; ESG-rating provider.
KLD	–	Kinder, Lydenberg, Domini & Co; ESG-rating provider.
MSCI	–	Morgan Stanley Capital International; ESG-rating provider.
OECD	–	Organization for Economic Co-operation and Development.
OLS	–	Ordinary Least Squares.
ROA	–	Return on Assets.
S&P	–	Standard & Poor's Global; ESG-rating provider.
US	–	United States.
VIF	–	Variance Inflation Factor.

ABSTRACT

Despite their undeniable relevance for both the academic and business sphere, ESG-ratings have been proven to be diverging. This can lead to consequences on capital allocation, investment behavior, volatility in stock returns, the use of ESG-ratings and trust in ESG data while increasing the risk of greenwashing. This research examines the impact of firm size, financial profitability and average ESG-performance of firms on the extent of ESG-rating disagreement in order to provide further insight into firm-level drivers of ESG-rating divergence. The aim of this research is not to fully explain rating divergence, but rather to identify firm-level drivers as the starting point for further analysis. Using OLS regression analyses of panel data, the divergence of Bloomberg and Refinitiv scores was evaluated for the overall ESG-score as well as for every pillar dimension. Firstly, the data provides evidence on the existence of ESG-rating divergence. The findings indicate a significant reducing impact of financial profitability on the extent of disagreement. Higher rating divergence is especially associated with greater firm size within the social pillar. For the overall ESG-score even a reduction of disagreement is related to higher firm size whereas rating agency specific effects of Bloomberg and Refinitiv cannot be completely out ruled. Average ESG-performance in the current and previous period shows an increasing impact on rating divergence. The analyses underline the necessity of an increase in transparency of rating agencies' methodologies to generate a more profound understanding of their differences and conduct more targeted analyses. Moreover, the results underline the need for standardization in definition and disclosure of ESG data.

JEL CODES: C12; C23; M14.

KEYWORDS: Corporate Sustainability; Divergence; ESG; ESG-Ratings; Rating Disagreement.

\

RESUMO

Apesar da sua inegável relevância tanto para a esfera académica como para a esfera empresarial, está provado que os ratings ESG são divergentes. Isso pode gerar consequências importantes, como impactos na alocação de capital, no comportamento de investidores, na volatilidade dos retornos das ações, no uso das avaliações ESG e na confiança nos dados, além de aumentar o risco de práticas de *greenwashing*. Este estudo examina o impacto da dimensão da empresa, da rendibilidade financeira e do desempenho ESG médio das empresas no grau de discordância dos ratings ESG, a fim de proporcionar uma visão mais aprofundada dos fatores que determinam a divergência das notações ESG a nível da empresa. O objetivo não é oferecer uma explicação completa para as divergências, mas sim identificar elementos ao nível das empresas que sirvam como base para estudos futuros. Utilizando análises de regressão OLS de dados de painel, a divergência entre as classificações da Bloomberg e da Refinitiv foi avaliada para a classificação global ESG e para cada dimensão dos pilares (ambiental, social e de governança). Os resultados confirmam que há, de facto, diferenças nas notações ESG. Os resultados indicam uma redução significativa do impacto da rendibilidade financeira no grau de desacordo. Por outro lado, empresas maiores tendem a ter maior discrepância nas classificações relacionadas ao pilar social, enquanto, para a pontuação geral, o aumento do tamanho da empresa pode reduzir as diferenças, embora o impacto específico de cada agência (Bloomberg e Refinitiv) não possa ser totalmente descartado. Outro ponto observado foi que um desempenho médio em ESG, tanto no período atual como no anterior, contribui para aumentar a divergência entre as classificações. Esses resultados reforçam a importância de uma maior transparência nas metodologias usadas pelas agências de classificação, permitindo um entendimento mais claro de suas diferenças e análises mais focadas. Além disso, os resultados evidenciam a necessidade urgente de padronizar a definição e a divulgação de dados ESG.

JEL CÓDIGOS: C12; C23; M14.

PALAVRAS-CHAVE: Sustentabilidade Empresarial; Divergência; ESG; Ratings ESG; Discordância de Rating.

.

TABLE OF CONTENTS

Glossary	i
Abstract, Keywords, JEL Codes	ii
Table of Contents	iv
Index of Tables	v
Acknowledgements	vi
1. Introduction	1
2. Literature Review and Hypothesis development	3
2.1. ESG-ratings & theoretical Frameworks	3
2.2. Rating agencies	6
2.3. Divergence of ESG-ratings & Hypothesis development	7
3. Methodology and Data	13
3.1. Data	13
3.2. Methodology	15
3.3. Descriptive Statistics	16
4. Analyses and Results	22
4.1. Analyses and Regression Results	22
4.2. Robustness Tests and additional Tests	28
5. Conclusion	31
References	34
Appendices	46

INDEX OF TABLES

Table I:	Sample by industry and country	17
Table II:	Descriptive statistics Model 1	18
Table III:	Correlations ESG-Scores	19
Table IV:	Correlations Model 1	19
Table V:	Correlations Model 1 pillar-level	20
Table VI:	Descriptive statistics Model 2	21
Table VII:	Correlations Model 2	21
Table VIII:	Regression Results Model 1	24
Table IX:	Regression Results Model 1: country and industry fixed effects	25
Table X:	Regression Results Model 1: year fixed effects	26
Table XI:	Regression Results Model 2	27
Table XII:	Robustness Test Refinitiv ESG-Score	29
Table XIII:	Regression Results for Model 1.1 and 1.2 supplemented with CONTROV variable	30

Acknowledgements

I thank my family, my partner and my friends for being a constant source of support and energy during the process of writing my thesis. Thank you for always having an open ear and being there for me.

Many thanks beyond that go to my supervisor, Prof. Cristina Gaio, for attentively helping me with her experience and knowledge no matter which questions I had. Thank you for providing such valuable support at any time.

1. INTRODUCTION

The attention attributed to environmental, social, and governmental (ESG) aspects has significantly increased. ESG aspects have reached a position of undeniable relevance in the context of investors (In et al., 2019), business organizations and the public interest (Jámbor & Zanócz, 2023). In 2022, $30.3 trillion have been invested in sustainable assets worldwide. Sustainable investment assets under management on non-US markets have grown by 20% from 2020 to 2022 (GSIA, 2022). Companies are under pressure to disclose and communicate E, S, and G information (Lee & Raschke, 2023; Schaltegger & Hörisch, 2017).

With the increase in interest in ESG data and performance, it is necessary to find a reliable measurement instrument of ESG-performance of corporations. Agreeing on universally accepted measurements of non-financial information can be challenging in lack of a common definition, objective measurement tools and external benchmarks (Tang et al., 2022; Tarquinio & Posadas, 2020). As a way of obtaining insight investors heavily rely on ESG-ratings from sustainability rating agencies to obtain an external evaluation of a companies' ESG-performance (Berg et al., 2022; Drempetic et al., 2020; Eccles & Stroehle, 2018; Sultana et al., 2018). Hartzmark & Sussman (2019) demonstrate that investors value sustainability since ESG-ratings influence the fund flow of mutual fund investors. Investment decisions based on external ESG-ratings show how ratings contribute to the allocation and transfer of capital. Moreover, a variety of academic studies rely on ESG-ratings from sustainability rating agencies as indicator for ESG-performance in their analysis (see, e.g. Albuquerque et al., 2019; Drempetic et al., 2020; Lins et al., 2017; Sharma et al., 2024). ESG-ratings, therefore, influence the academic conclusions drawn from analyses conducted based on these ratings.

However, research papers have found the results of rating agencies to be diverging for ratings of the same corporation (Berg et al., 2022; Capizzi et al., 2021; Chatterji et al., 2016; Christensen et al., 2022). The ESG-scores of Tesla are one of the most prominent examples, being rated at the top by MSCI, the bottom area by FTSE and medium by Sustainalytics (Dimson et al., 2020). This implies that ESG-ratings currently are inconsistent and not fully comparable (In et al., 2019). Evaluating the true value of ESG-ratings is crucial to sustainable investing (Tang et al., 2022). Wrongful, incomplete, or

inconsistent measurement of ESG-performance in the form of ESG-ratings can lead to severe consequences. Capital may not be allocated to the most sustainable companies (Drempetic et al., 2020; Hartzmark & Sussman, 2019). ESG-rating uncertainty may present a barrier to the use of ESG-ratings in investment processes (Amel Zadeh & Serafeim, 2018) or even discourage investors from undertaking ESG investments (Avramov et al., 2022).

Low quality in ESG disclosures and ESG-rating divergence can increase the probability of future greenwashing (Biju et al., 2023; Wu et al., 2020). Rating disagreement is proven to hinder direct market reactions to positive ESG news (Serafeim & Yoon, 2023). This implies that the differences in ratings lead to questions on quality and trustworthiness of ESG-ratings (Liu, 2022) especially in the light of growing concerns of greenwashing. The concerns of investors manifested in stock reactions associated with ESG-rating divergence. Higher divergence of ratings was found to be positively correlated with increased return on stock thus influencing the equity cost of capital (Gibson Brandon et al., 2021), higher volatility in returns and a decrease in the likelihood of obtaining external financing (Christensen et al., 2022). Moreover, company ESG-scores may determine which companies surpass the threshold to be included in sustainability stock indices (Hedesström et al., 2011). Authors argue that the divergence can signal that ESG-ratings do not effectively measure actual sustainability performance and can lead to inefficient capital allocation (Drempetic et al., 2020).

Conclusions reached in academic research relying on ESG-ratings in their analysis might need to be questioned and future research requires a thorough evaluation of sustainability performance measurement. The inconsistency of ESG-ratings fosters a lack of trust in the ratings as ESG-performance measurement tool and highlights the need to draw conclusions on their basis exercising cautiousness (Chatterji et al., 2016). Evidence was found that distrust in the ratings can even lead to investors deterring from the use of ratings for their decision-making (Jonsdottir et al., 2022).

Given the significance of ESG-ratings in both, academic and business context, and the potential consequences of the divergence of ESG-ratings, it is crucial to explore the primary factors that lead to these divergences and increase the differences between ESG-ratings from different agencies. Some research has already been conducted aiming at

understanding and explaining the main reasons of disagreement among rating agencies (e.g. Berg et al., 2022; Christensen et al., 2022; Dumrose et al., 2022; Liu, 2022). However, the authors call for further research on the drivers of raters' divergence. The following research will focus on investigating potential firm-level drivers of ESG-rating divergence, including firm size, financial profitability and ESG-performance, by applying regression analysis to panel data.

The results of this research are a step on the way to generating a better understanding of drivers of differences among rating agencies' scores, allowing for a more accurate assessment of these differences and more conscious use of ESG-ratings. Higher financial profitability plays an important role in explaining lower rating disagreement. It cannot be generalized that higher firm size is conductive to higher score divergence. On average higher ESG-scores displayed an increasing impact on the extent of disagreement. The analyses, beyond this, demonstrate the relevance of further increasing transparency and introducing standardizing regulations and common theorization.

The remainder of this report is structured as follows: Section 2, Literature Review & Hypothesis Development ,provides an overview over prior related research and the stated hypothesis, section 3, Methodology and Data, gives an overview over the research design, the chosen model and methodology applied, the collection and design of data as well as descriptive statistics. Section 4 outlines the results of the analysis and the outcomes of the robustness tests performed. Finally, section 5 concludes and highlights any limitations given.

2. LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

2.1. *ESG-ratings & theoretical frameworks*

ESG metrics are a set of standards of environmental, social and governance factors to achieve long-term sustainability which are among others applied to evaluate the sustainable performance of corporations used by potential investors in the light of sustainable finance (Sultana et al., 2018). Environmental parameters in general show the company in the context of its environment and nature, social factors describe matters with a relationship of a corporation with its employees, vendors, customers, the public, and other agents it engages with. Governance factors cover the corporate governance structure

and governance practices, such as leadership, risk management, executive compensation, and ethics (Oprean-Stan et al., 2020; Sultana et al., 2018). Soppe (2004) defines sustainable corporate finance as "a financial policy that strives for triple-bottom-line performance measurement" (p. 221) aiming at long-term financial objectives. The triple-bottom-line framework has three different pillars: social, environmental and economic (Elkington, 1997). According to Elkington (1997) the bottom-line accounting on profit or loss is supplemented by accounting on human capital and natural capital to measure performance from a broader and more holistic perspective.

ESG-ratings are an assessment of a company's performance in terms of environmental, social and governance matters (Tang et al., 2022). They have in practice become commonly applied as measurement tool of ESG-performance in both, (sustainable) investment practices as well as economic, management, and finance research (Albuquerque et al., 2019; Drempetic et al., 2020; Gibson Brandon et al., 2021; Lins et al., 2017; Sharma et al., 2024), allowing to compare companies in terms of sustainable performance (Berg et al., 2022) by capturing a company's ESG-performance as accurately as possible (Kotsantonis & Serafeim, 2019). The ESG-rating providers, therefore, take an increasing role as "nonfinancial information intermediar[ies]" (Tsang et al., 2023, p.18).

Besides their practical relevance at hand, ESG-ratings are also relevant from several theoretical perspectives. When economic agents assess the sustainability performance of a company, they may face information asymmetries (Rischkowsky & Döring, 2008) as they often do not have complete access to information or distrust the information provided (Windolph, 2011). This goes back to the seminal research of Jensen & Meckling (1976) on the conflict of interests between distinct contract parties known as the principle-agency theory. Signaling these suggests that in these cases of information asymmetry the company-side party will seek to credibly distribute information to a third party (Connelly et al., 2011; Spence, 2009). Cooperating with ESG-rating agencies and proactively disclosing ESG information can have a signaling effect on stakeholders (Hahn & Kühnen, 2013). In this context, ESG-rating agencies can function as reliable third-party intermediary to provide relevant information to these stakeholders (Chatterji & Levine, 2006; Gregory, 2024; Rischkowsky & Döring, 2008; Windolph, 2011) and decrease information asymmetries (Eccles & Stroehle, 2018).

Freeman (2010) defined stakeholders as "groups without whose support the organization would cease to exist" (p. 31). A fundamental part of Freeman's stakeholder theory is the need of organizations to acknowledge the joint needs and expectations of distinct stakeholders with relations to the corporation and to maximize the value creation for them to create overall corporate value (Freeman et al., 2007). Stakeholders' demand for reliable non-financial information has increased in the past (Jámbor & Zanócz, 2023). In recent years, interactions with stakeholders have increasingly been characterized as cooperative (Dathe et al., 2022). Transferred to the concept of ESG, this can mean that companies aim at showing a good ESG-performance to demonstrate the understanding and fulfilment of interests, expectations, and perspectives of a wide group of stakeholders (Hahn & Kühnen, 2013). This can contribute to a reduction in cost of capital (El Ghoul et al., 2011), better access to financing and reduced capital constraints (Cheng et al., 2014), and improved customer satisfaction (Servaes & Tamayo, 2013). In this context, ESG-ratings may function as a tool to demonstrate the effective consideration of ESG-related stakeholder needs through an external third-party measurement.

Furthermore, Meyer & Rowan (1977) and DiMaggio & Powell (1983) state that as part of neo-institutional theory companies conform with societal rules, norms, and expectations to seek legitimacy in the eyes of their environment. The authors argue that only by reaching this status of legitimacy the organizational survival can be secured (Hasse & Krücken, 2009; Meyer & Rowan, 1977) and benefits may be optimized. Firms are motivated by the external environment to make ESG disclosures (De Villiers & Alexander, 2014). Legitimacy theory, moreover, states that firms right to exist, is subject to a legitimization by society (Dowling & Pfeffer, 1975; Suchman, 1995). Expressing an acceptable business conduct in ESG matters can aim at establishing a status of legitimacy (Schaltegger & Hörisch, 2017). In fact, the legitimacy-seeking characteristic is a dominating driver of corporate social responsibility practices. Incorporating sustainability practices and sustainability disclosure can therefore be a tool to form the observed legitimacy of a companies' environment (Campbell et al., 2003). Investors may however rely on external information provided in the form of ESG-ratings in evaluating the legitimacy as it may be perceived of higher credibility and reliability (Döpfner, 2016). The divergence of such ESG-ratings threatens the above-described theoretical relevance

of sustainability disclosures and ESG-ratings as the differences lead to distrust and incomparability (Jonsdottir et al., 2022).

2.2. Rating agencies

The demand for external and reliable information on ESG-performance of corporations highlights the relevance of ESG-rating agencies. After first emerging in the 1980s the market for ESG-ratings grew in accordance with the growing interest in sustainable investments (Berg et al., 2022). Sustainability ratings on the independent rater market are often conducted by various organizations, such as banks, specialized rating agencies, and screening departments (Windolph, 2011). Some of the main well-established ESG-rating agencies are among others Refinitiv (previously Asset4), Sustainalytics, Bloomberg, FTSE, ISS, KLD (since 2010 part of MSCI), S&P Global, and MSCI (Berg et al., 2022; Escrig-Olmedo et al., 2019). KLD is one of the historically most used ratings in the academic sphere (Berg et al., 2022).

In contrast to credit ratings, ESG-ratings are commissioned by the investors and not by the companies themselves. However, the rating agency often is directly provided with information from the company which is used in the evaluation along with other external sources (Gregory, 2024), interviews, and own research (Tsang et al., 2023). Therefore, specific data sources may differ between rating agencies. Moreover, companies may report a variety of different metrics and data points for the same ESG issue (Kotsantonis & Serafeim, 2019). In the process of generating an ESG-score the raters will typically (1) identify indicators of sustainability, (2) collect and (3) assess data and finally (4) quantify qualitative data through their individual ranking methodology (European Commission, 2020). However, the specific categories and sub-topics identified, the measurement, sector-specific topics, collection process and sources, final evaluation methodologies and the consideration of ESG risks differ among the raters. In case of missing company information, the raters may have different approaches of imputation to fill the gaps (Kotsantonis & Serafeim, 2019). Imputation models may, for instance, include regression methods, machine learning-based predictive mean matching, rule-based approaches, and estimation-based input-output models. In addition, differences in the classification of ESG-performance due to different benchmarks and peer group definitions exist. In fact,

changes in Refinitiv's methodology have led to inconsistent ESG-ratings and significant differences between original and revised Refinitiv scores over time (Berg et al., 2021).

The lack in transparency of raters on the methodologies and practices applied has been subject to criticism (see, e.g. Berg et al., 2022; Chatterji & Levine, 2006; Escrig-Olmedo et al., 2019; Tang et al., 2022) as it has a negative impact on the comparability, trustworthiness, and understandability of ESG-ratings (Jonsdottir et al., 2022). The lack of a clear definition of ESG and a standardized global ESG reporting framework as well as objective measurement tools or external benchmarks (Amel Zadeh & Serafeim, 2018; Eccles & Stroehle, 2018; Tang et al., 2022) attributes ratings a certain interpretational characteristic (Berg et al., 2022) which imposes conceptual difficulties on information providers (Eccles & Stroehle, 2018). Hedesström et al. (2011) in their analysis, for instance, point out the missing consensus on what acting environmentally sustainable means. The lack of standardization and transparency on methodologies and procedures has a negative impact on the comparability amongst ESG-ratings (Jonsdottir et al., 2022). Despite their empirical relevance ESG-ratings yet are not subject to regulations as credit ratings on the other hand are (Tang et al., 2022).

2.3. *Divergence of ESG ratings & Hypothesis development*

Despite the high relevance of ESG-ratings they are often criticized due to the divergence among different raters which has among others been identified by Chatterji et al. (2016). Chatterji et al. (2016) investigated common theorization and commensurability of ratings as preconditions of rating convergence. The authors define common theorization as agreement on the understanding of what corporate social responsibility means. Whereas, commensurability is the way of measuring the given information. Chatterji et al. (2016) conclude that rating agencies show both, low common theorization and low commensurability meaning that rating agencies apply different methodologies and definitions of ESG-performance and different methods to measure the performance. In total, the work demonstrates a low convergence of ESG-ratings amongst different rating agencies. In analysing the relationship between distinct indicators and ESG-ratings several papers have come to conclusions supporting these results (see e.g., Billio et al., 2021; Christensen et al., 2022; Gibson Brandon et al., 2021) leading to a consensus on ESG-rating divergence.

The work of Chatterji et al. (2016) has been followed by further research on the drivers of the differences in ESG-ratings. The conclusions of Berg et al. (2022) also support the hypothesis of rating divergence. In investigating reasons of the divergence, the authors decompose rating differences into three sources of divergence: scope, measurement, and weight. Scope divergences stem from different ESG-indicators being included measured in the evaluation, measurement divergences are the result of measuring the same indicator differently. Finally, weight divergences stem from assigning different relative weights to the different components of the overall ESG score, such as (sub-) topics, indicators and pillars. According to the results of Berg et al. (2022) measurement divergences, especially in the categories climate risk management, product safety, corporate governance, corruption, and environmental management systems, are the main driver of differences in ESG-ratings contributing to the divergences by 56%. Correspondingly, rating divergence was concluded to be attributed to differences in the scope of rating agencies by 38% and to differences in the weighting of topics and pillars by 6%. In their analysis of environmental pillar scores in the car manufacturing and paper and forestry industry, Hedesström et al. (2011) also identified differences attributable to the weighting of environmental topics. In their analysis of the influence of female CEOs on ESG-ratings Aabo & Giorici (2023) came to different conclusions per rating agency.

ESG-ratings are based on a variety of different metrics and may differ in scope, measurement, and weightings (Berg et al., 2022; Lee & Raschke, 2023). Attempts to standardize and regulate the metrics to report can contribute to reducing the divergence. Dumrose et al. (2022), for instance, demonstrate that the EU Taxonomy, which was introduced in the past years, to some extent has an explanatory value for environmental pillar ratings and thus might contribute to addressing measurement differences. The study shows that standardized mandatory ESG reporting requirements might be helpful in reducing confusion and contributing to a greater consensus on presentation and measurement of environmental performance of companies.

Other research papers focussed on the relevance of firm size, data availability, and the amount of information disclosed on rating divergence. Firm size appears to be of significant explanatory value for determining ESG-ratings (Tang et al., 2022). Drempetic et al. (2020), who investigated the impact of firm size on Refinitiv ratings, concluded on

a significant positive relationship of the two variables. Tang et al. (2022) reached the same conclusions in an analysis of KLD ESG-ratings. In his analysis on the relationship between firm size and ESG-ratings Gregory (2024) concluded on an overall positive relationship of company size and their ratings and, moreover, found indications for evidence of rating divergence. The authors point out that one reason for this may be that larger firms have access to higher absolute budgets and thus have the capabilities to collect and disclose more information. Larger companies are believed to provide more extensive and voluntary ESG information (Gallo & Christensen, 2011; Tamimi & Sebastianelli, 2017), also due to higher public interest and pressure (Hahn & Kühnen, 2013; Naser et al., 2006; Prado-Lorenzo & Garcia-Sanchez, 2010; Veronica Siregar & Bachtiar, 2010). This aligns with the signaling theory as well as the stakeholder theory as larger companies may be exposed to a larger and more diverse set of stakeholders and thus may feel higher pressure to voluntarily disclose ESG information (Hahn & Kühnen, 2013).

According to Christensen et al. (2022) and Liu (2022), ESG disclosure has a positive impact on rating divergence, suggesting that increased disclosures in fact result in greater disagreement among rating agencies. This is especially true for the "E" (environmental) and "S" (social) pillars. One reason for this may be that a greater amount of information leads to different ways of interpreting the underlying data set(s) and therefore to disagreements (Christensen et al., 2022; Cookson & Niessner, 2020). With more information available more ways of measuring the information are possible, which was proven to be a source of differences in ESG-ratings (see Berg et al., 2022). In contrast, Kimbrough et al. (2024), who analysed voluntary primary source management-provided disclosures, found a negative relationship to rating disagreement for US companies. A reason for the differences to the results of Christensen et al. (2022) and Liu (2022) might be that these authors considered information from various sources including not directly management-provided data. Research has found evidence on a significant positive relationship between company size and resources to provide ESG information (Drempetic et al., 2020) meaning large firms are more likely to have the capabilities to disclose more information. In addition to financial resources, large firms have a higher degree of knowledge on sustainability management tools (Hörisch et al., 2015) and may face a higher degree of public exposure and pressure to make sustainability information publicly

available as a proactive approach towards criticism and pressure (Hackston & Milne, 1996; Michelon, 2011). Moreover, large firms are often more complex and diversified (Gibson Brandon et al., 2021) which can add another level of complexity to the evaluation of ESG-performance. Given the observed relationships regarding disclosure and rating divergence on the one hand and the potential influence of firm size on data availability on the other hand, the question regarding the relationship between ESG-rating divergence and firm size may arise. This leads to H1, ESG-rating divergence being the variable of interest:

H1: *Firm size has a positive effect on ESG-rating divergence.*

In previous research, different proxies have been used to measure firm size. Some authors measured firm size solely in terms of total assets (see Christensen et al., 2022; Crespi & Migliavacca, 2020), others in terms of market capitalization, revenue, and total employees (see Drempetic et al., 2020; Gregory, 2024). To analyse the given relationship of H1 in depth two variating metrics of firm size (total assets and market capitalization) will be considered in the analyses.

In the past several studies have been conducted analysing the link between ESG-performance and financial performance (Friede et al., 2015; Rahi et al., 2024; Schaltegger & Hörisch, 2017). The results have varied depending on the research design and have led to inconclusive results. Some argue that enhanced environmental and social performance lead to favourable advantages for firm performance, for instance, by increasing the firms' reputation, innovation, competitiveness or cost savings (Ameer & Othman, 2012; Bartolacci et al., 2020; Lee & Raschke, 2023; Porter & Van Der Linde, 1995; Russo & Fouts, 1997; Waddock & Graves, 1997). On the contrary, opponents of this approach are concerned about the decreasing effect of costs of environmental measures on firm profitability, efficiency, and financial performance (Friedman, M. (1970); Hedesström et al., 2011; Walley & Whitehead, 1994). Taken together, research suggests an overall positive or neutral impact of sustainability practices and disclosures on financial company performance (Bartolacci et al., 2020; Friede et al., 2015; Rahi et al., 2024).

As part of the slack theory a virtuous cycle with bidirectional causality between sustainable and financial performance has been suggested implying that not only may sustainable performance favour increases in financial performance but also vice versa

(Ameer & Othman, 2012; Boso et al., 2017; Orlitzky et al., 2003). Higher slack resources are associated with a higher impact on corporate sustainability (Rahi et al., 2024). In support of the slack theory, Waddock & Graves (1997) found a positive cause-and-effect connection of financial performance (as cause) and sustainable performance. Firm profitability is often seen as a relevant indicator for slack resources (Xiao et al., 2018). Firms with higher financial performance may have more resources available to engage in ESG activities (McGuire et al., 1988). Greater slack resources are associated with a higher level of substantial investments in environmental and social strategic decisions (Perez-Batres et al., 2012) and can lead to more innovation (Nohria & Gulati, 1996). Greater retained profit has been proven to be related to greater social performance (Waddock & Graves, 1997). Crespi & Migliavacca (2020) and Kimbrough et al. (2024) found evidence supporting the hypothesis that profitability may have a positive effect on ESG-ratings.

As part of previous research on ESG-rating divergence financial profitability measures have often been included as control variables. Gibson Brandon et al. (2021) found a negative relationship between gross profitability and rating disagreement, which they attribute to a greater consensus on more profitable firms. Raters may be more likely to agree on more profitable companies as they may be more innovative and can re-invest their profit in ESG activities. This aligns with the argumentation of the slack resource theory. Therefore, H2 is derived as follows:

H2: *Company profitability has a negative impact on ESG-rating divergence.*

Raters' subjectivity appears to be another key factor influencing ESG-ratings. Tang et al. (2022), for instance, have shown how KLD allocated higher scores to companies owned by the same institutional investors implying that raters' ownership can be a determinant of ESG-ratings. This rater-specific bias can contribute to differences in ESG-ratings across rating agencies. Berg et al. (2022), moreover, prove that a significant rater effect exists meaning that a positive evaluation in one area and a raters' overall (positive) company view can bias the rater and thus lead to more positive evaluations in other categories and, by that, the overall company score. Especially areas requiring a high degree of judgement may be susceptible to the raters' bias. According to Berg et al. (2022) a possible explanation is that analysts at rating agencies often evaluate a whole company and not just one category. This shows that the differences in ratings partly arise from

structural rater- and firm-specific patterns which are rooted in the organization of the rating agencies rather than being mere randomly distributed differences (Berg et al., 2022). The authors thereby demonstrate that raters themselves have a significant impact on the scores. These results are coherent with the halo theory transferred to the concept of ratings. The halo theory describes the tendency of raters to (subconsciously) align detailed evaluations with the priorly obtained holistic view of the entity (Balzer & Sulsky, 1992; Murphy et al., 1993; Thorndike, 1920). It goes back to the Thorndike definition of halo effect: "marked tendency to think of the person in general as rather good or rather inferior and to colour the judgments of the [performance] by this general feeling" (Thorndike, 1920, p.25). Raters' decisions are influenced by their overall positive or negative view of the individual to be rated. Similarly, financial auditors tend to be influenced by judgements obtained on a high-level in subsequent more detailed evaluations (O'Donnell & Schultz, 2005).

In addition, there has already been some evidence on the negative relationship between ESG-performance and ESG-rating divergence (Christensen et al., 2022; Kimbrough et al., 2024; Liu, 2022) implying that raters disagree more on companies which on average have a lower ESG-performance. These findings align with the results of Bonsall & Miller (2017) and Cantor & Packer (1994) demonstrating a higher disagreement of credit ratings for companies with on average lower credit ratings. Lee & Raschke (2023) have demonstrated that a lower ESG-performance is positively associated with greenwashing. Controversial opinions on these companies might therefore be more likely. Additionally, the consideration of controversial events in the evaluation may differ among the rating agencies (see Bloomberg 2023; Refinitiv, 2022).

In summary, rater effects have proven to be incorporated in ESG-ratings which means that the scores are not completely free from biases (Berg et al., 2022), which can influence the overall rating. Moreover, credit raters and ESG raters tend to disagree more on companies with lower ESG-performance (Bonsall & Miller, 2017; Cantor & Packer, 1994; Christensen et al., 2022; Kimbrough et al., 2024). There appears to be a higher consensus on companies with on average higher ESG-performance. Considering the rater-specific effect in scores divergence could be higher for companies with on average lower ESG-ratings. This leads to H3.1:

H3.1*: Lower average ESG performance in the current period has a positive impact on ESG-rating divergence.*

Analysts may be influenced not just by the current company evaluation but also by their subconscious incorporation of past experiences. Past ratings may have led to an image and holistic view of a company being formed contributing to future halo effects (Balzer & Sulsky, 1992; Murphy et al., 1993; Thorndike, 1920). Prior ratings can influence subsequent evaluations (Christensen et al., 2022). Therefore, as variation to H3.1, H3.2 is derived as follows:

H3.2*: Lower average ESG performance in the previous period has a positive impact on ESG-rating divergence in the following period.*

3. METHODOLOGY AND DATA

3.1. Data

The analyses will be done on the divergence of Bloomberg and Refinitiv ESG-ratings. According to an analysis published in an OECD report Bloomberg and Refinitiv ESG-scores are among the three most relevant ESG-rating providers (Boffo & Patalano, 2020). Based on panel data of European listed companies for 2016-2022 the variable of interest (rating divergence $ESG\_Dis_{i,t}$) is calculated as absolute difference between the ESG-ratings of Bloomberg and Refinitiv for a given firm i in year t as the dispersion is only evaluated for two raters (Kimbrough et al., 2024). As Bloomberg applies a scale ranging from 0-10, whereas Refinitiv scores range from 0-100, the ratings are rescaled to a common scale of 0-10. Based on the obtained data from Bloomberg and Refinitiv ratings a common sample is constructed. To be included in the sample firm-year ESG-ratings are required to be available from both, Refinitiv and Bloomberg. The analysis will be done for the overall ESG-rating as well as the decomposed ratings for environmental, social, and governmental pillars as disclosure levels can, for instance, vary for environmental, social and governmental disclosures (Tamimi & Sebastianelli, 2017) and pillar weightings can influence overall rating differences (Lee et al., 2023).

To allow a thorough analysis of the divergence of Bloomberg and Refinitiv in the following an understanding of the main differences and similarities in the approaches of the two rating agencies is developed. Slight differences can, firstly, be found in the

definition of the ESG-scores of each provider. Bloomberg considers their scores as an indicator of "a company's management of financially material ESG issues" (Bloomberg, 2023, p.2), whereas Refinitiv defines ESG-scores as a measurement of "the company's ESG-performance based on verifiable reported data in the public domain" (Refinitiv, 2022, p.8). Both definitions highlight different aspects of their ratings.

Refinitiv's ESG-score is calculated by taking a weighted average of the three pillar scores, which are themselves derived from a weighted sum of ten category scores that include hundreds of ESG measures (Refinitiv, 2022). Bloomberg's ESG-scores also consist of separate environmental, social, and governance pillar ratings calculated based on separate field, issue, and theme scores, which are aggregated to an overall weighted ESG-score (Bloomberg, 2023). Therefore, the score providers have each created a scoring taxonomy which structures the data points, fields and (sub-) issues considered. The individual data fields, their measurement, and the aggregation of data points may differ from each other. Bloomberg and Refinitiv scores are both based on publicly available company data (Bloomberg, 2023; Refinitiv, 2022), so that, in principle, analysts of each side have access to the same data.

Both rating agencies include industry-specific adjustments: Refinitiv's respective weights of the environmental and social pillar are industry-specific (Refinitiv, 2022). Bloomberg's social and environmental pillar score include industry-specific fields (Bloomberg, 2020), while governance pillar scores consider country- and market-specific policies and rules (Bloomberg, 2023). The exact treatment of industry-specialities may lead to differences amongst the two rating agencies.

Both score providers measure the ESG-performance relative to the respective country/industry peer groups (Bloomberg, 2023; Refinitiv, 2022). Peer groups e.g. have an impact on individual relative field and pillar weightings of Bloomberg scores (Bloomberg 2023). Despite both involving peer groups in defining "good" performance discrepancies in the specification of peer groups can lead to differences in weightings and measurements.

Beyond the ESG-score, Refinitiv also publishes ESGC-scores, which consider controversies and negative incidents the company has been involved in. In case no controversies have become public, the ESG-score equals the ESGC-score. The ESGC-

score can therefore be a more comprehensive measurement as it also includes negative aspects (Tang et al., 2022). Compared to Refinitiv's ESGC-score, Bloomberg does not specifically screen for controversial events. According to their methodology they, however, incorporate matters which can negatively impact a company's financial performance (Bloomberg, 2023). For this, the analysis will compare Refinitiv's ESGC-score with Bloomberg's ESG-score to ensure comparability.

Finally, Refinitiv and Bloomberg scores are based on a materiality approach (Bloomberg, 2023; Refinitiv, 2022), whereas the two approaches of defining materiality may be different. In lack of transparency on the materiality evaluation as well as judgement being involved it is difficult to thoroughly compare this process step for Bloomberg and Refinitiv.

The sample was constructed by extracting data for firms from the following indices: Amsterdam Exchange Index, Bel 20 Index, Deutsche Börse DAX Index, FTSE 100 Index, CAC 40 Index, OMX Stockholm_PI, Swiss Exchange Index, OMX Copenhagen 20 Index, IBEX 35 Index (Spain), OMX Helsinki, ISEQ Overall Price Index. All firm-level data as well as the Refinitiv ESG-ratings were extracted from Eikon Datastream. Bloomberg ESG-scores were obtained from the Bloomberg Terminal. To adjust for the impact of extreme outliers, all continuous variables except for $ESG\_Dis_{i,t}$ and $S\_Dis_{i,t}$ have been winsorized at the 1%-level. Leverage is the only variable winsorized at the 2%-level. Duplicates, errors and firm-years with missing data were removed.

3.2. Methodology

Following a deductive approach, the objective of the analysis is to test H1-H3 applying quantitative methods. Before analysing the divergence of the ratings, the absence of an extensive correlation between Bloomberg and Refinitiv ratings will be verified by reviewing the pairwise correlation coefficient. The given hypothesises will be tested by applying multiple linear regression analysis to panel data (see Christensen et al., 2022; Kimbrough et al., 2024; Liu, 2022). For the analysis Stata will be utilized as software. The general regression model will be regressed in pooled OLS and is specified as follows, i indexing the firm, t indexing the year.:

$$ESG\_Dis_{i,t} = \beta_0 + \beta_1 SIZE_{i,t} + \beta_2 PROF_{i,t} + \beta_3 ESG\_PERF_{i,t} + \Sigma\beta_4 CONTROL + \epsilon_{i,t} \quad (1)$$

The proxy for rating divergence ($ESG\_Dis_{i,t}$), is measured as absolute difference between the ESG-ratings of Bloomberg and Refinitiv for a given firm i in year t and is calculated at the year-end. $SIZE_{i,t}$ is the independent variable measuring firm size for a firm i in year t. First, firm size is calculated as natural logarithm of total assets (Model 1.1: $SIZE1_{i,t}$) (see Chen et al., 2017; Drempetic et al., 2020). As a variation, firm size will be measured as the natural logarithm of market capitalization (Model 1.2: $SIZE2_{i,t}$) (see Christensen et al., 2022; Drempetic et al., 2020; Liu, 2022). The logarithm is applied for better comparison between the firms. $PROF_{i,t}$ is the proxy for firm profitability measured in terms of return on assets of firm i in year t. To test H3.1 $ESG\_PERF_{i,t}$ is the average ESG-performance measured as average ESG-rating of a firm i in year t (see Christensen et al., 2022; Liu, 2022). In order to test H3.2 the model is modified as shown in the following Model 2 by lagging the average ESG-performance of a firm i by one year. Due to the use of a lagged variable the first year of panel data cannot be used in the analysis. Variations in the size proxy analogue to Model 1 lead to the sub-models 2.1 and 2.2.

$$ESG\_Dis_{i,t} = \beta_0 + \beta_1 SIZE_{i,t} + \beta_2 PROF_{i,t} + \beta_4 ESG\_PERF_{i,(t-1)} + \Sigma\beta_5 CONTROL + \epsilon_{i,t} \quad (2)$$

The model will be controlled for industry, country, and year effects (see Kimbrough et al., 2024; Liu, 2022). Additionally, leverage (LEV), as indicator for capital structure, book-to-market ratio (BTM), as an indicator for growth opportunities (see Christensen et al., 2022; Liu, 2022), and the natural logarithm of capital expenditure ($CAPEX$), as indicator for innovation, (Kimbrough et al., 2024) are included as control variables ($CONTROL$). Standard errors are clustered at the firm-year level. To ensure the absence of extensive multicollinearity among the variables, VIF is used.

Please refer to Appendix A for an overview of the variable definitions.

Robustness tests addressing the decisions of firm size and financial profitability proxies and the choice of Refinitiv's ESGC-score over the ESG-score and pooled OLS as method of regression analyses will be performed.

3.3.Descriptive Statistics

The sample shown in Table I contains companies for which ESG-scores from both rating agencies are available. By excluding firm years with missing data, the resulting

panel becomes imbalanced, with 200 companies having complete information for each year. The constructed sample includes 1,639 firm-year observations in total. Table I presents the data, covering 279 firms across 13 countries and 11 industries.

Table I: Sample by industry and country.

Companies by industry		**Companies by country of headquarter**	
Industry	**No. of companies**	**Country**	**No. of companies**
Basic Materials	25	*Belgium*	18
Consumer Discretionary	38	*Denmark*	19
Consumer Staples	18	*Finland*	22
Energy	7	*France*	32
Financials	53	*Germany*	33
Health Care	30	*Ireland; Republic of*	9
Industrials	58	*Luxembourg*	2
Real Estate	13	*Mexico*	1
Technology	13	*Netherlands*	24
Telecommunications	11	*Spain*	27
Utilities	13	*Sweden*	23
Total	279	*Switzerland*	17
		United Kingdom	52
		Total	279

Table II highlights the summarized statistics for the Bloomberg and Refinitiv scores and the variables used in Model 1. As the table shows, Refinitiv ratings in the sample seem to have a higher range than Bloomberg ratings. Refinitiv ratings have a higher mean, and the maximum score exceeds the maximum Bloomberg score by 1.24. Among the three pillars, the G-pillar has the narrowest range, spanning from 3.19 to 8.91, compared to the E-pillar, which covers the full scale from 0 to 10. In contrast, the Refinitiv pillar ratings do not exhibit such pronounced differences in range. ESG_DIFF is defined as absolute difference between Bloomberg and Refinitiv scores. Therefore, there should be no negative values which is confirmed by the summarized statistics. Differences between the ratings range from none to 6.86 with a mean of 2.463. Interestingly, the mean differences between Bloomberg and Refinitiv ratings vary across the three pillar scores. The social pillar shows the largest discrepancy, with a mean difference of 4.408, while the governance pillar has a much smaller mean difference of 1.546. This suggests that the variations between the rating agencies are more pronounced in specific areas of evaluation. Consequently, in the analyses, not only the overall ESG-score but also a break down at the pillar level for a more detailed comparison will be assessed. $SIZE1$ depicting the natural logarithm of total assets shows a similar mean to $SIZE2$, the natural logarithm

of market capitalization. Errors in all logarithmic variables were avoided by adding 1 before applying the natural logarithm. The profitability indicator ROA shows both negative and positive figures, as net losses result in an overall negative ratio. Errors caused by negative signs in both the numerator and denominator, which could falsely suggest positive profitability, are prevented by the variable's design. While the return may be negative, total assets in the sample are positive in all cases. The same applies for the debt-to-equity (LEV) and book-to-market (BTM) ratio. Shareholders' equity may be negative; however, debt or market capitalization are shown with positive signs in all instances. ESG_PERF is a function of the Bloomberg and Refinitiv scores average and, therefore, reflects the summary statistics of the two ratings.

Table II: Descriptive statistics Model 1.

Descriptive statistics of regression variables								
Variable	**N**	**Mean**	**Min**	**Max**	**SD**	**Perc. 25**	**Med.**	**Perc. 75**
Bloomberg ESG Score	1,639	4.142	0.90	8.27	1.186	3.33	4.13	4.97
Bloomberg E-Pillar	1,639	3.854	0.00	10.00	2.149	2.23	4.02	5.39
Bloomberg S-Pillar	1,639	3.143	0.00	9.29	1.627	1.95	2.87	4.06
Bloomberg G-Pillar	1,639	6.391	3.19	8.91	1.124	5.58	6.37	7.24
Refinitiv ESGC Score	1,639	6.472	0.39	9.51	1.561	5.45	6.66	7.67
Refinitiv E-Pillar	1,639	6.995	0.00	9.90	2.135	6.02	7.51	8.56
Refinitiv S-Pillar	1,639	7.475	0.24	9.82	1.758	6.60	7.85	8.79
Refinitiv G-Pillar	1,639	6.444	0.52	9.86	2.067	5.06	6.85	8.07
ESG_DIFF	1,639	2.463	0.00	6.86	1.334	1.45	2.48	3.41
E_DIFF	1,639	3.359	0.02	8.69	2.058	1.81	3.19	4.66
S_DIFF	1,639	4.408	0.00	8.59	1.946	3.07	4.59	5.80
G_DIFF	1,639	1.546	0.03	4.45	1.063	0.68	1.38	2.25
SIZE1	1,639	23.776	20.18	28.01	1.663	22.54	23.52	24.82
SIZE2	1,639	23.405	20.84	26.03	1.104	22.63	23.37	24.12
PROF	1,639	5.353	-11.25	30.28	6.122	1.62	4.50	7.79
ESG_PERF	1,639	5.312	1.92	7.50	1.120	4.61	5.41	6.12
CapEx	1,639	19.559	13.44	23.36	1.920	18.52	19.68	20.81
LEV	1,639	94.470	0.00	465.48	95.023	35.33	61.66	120.64
BTM	1,639	63.467	-7.87	329.37	61.042	24.22	43.49	82.28

Table III shows the correlation between Bloomberg's scores, Refinitiv's ESGC-scores and Refinitiv's ESG score. The correlation coefficients indicate a low linear correlation between Bloomberg's scores and Refinitiv's scores confirming the ultimate understanding of diverging ESG-scores from different rating providers.

Table III: Correlations ESG-scores.

Correlations of ESG Scores			
	Bloomberg ESG	Refinitiv ESGC	Refinitiv ESG
Bloomberg ESG	1.000		
Refinitiv ESGC	0.386	1.000	
Refinitiv ESG	0.501	0.724	1.000

Pairwise correlations between the main variables included in regression Model 1 are presented in Table IV. The correlation coefficient of more than 0.6 between the $SIZE1$ and $SIZE2$ variables is not an issue, as they are used interchangeably in the regression. Regarding the stated hypotheses, the $SIZE1$ correlation coefficient with ESG_DIFF indicates a positive linear connection between total assets and the extent of deviation between the ratings. $SIZE2$, however, indicates a negative linear relationship. This will be subject to further analyses in the regression analysis. Firm profitability shows a negative linear correlation with rating divergence. ESG_PERF, by contrast, has a positive correlation coefficient indicating a positive association with the degree of ESG-rating divergence.

Table IV: Correlations Model 1.

Correlations of regression variables								
	ESG_DIFF	*SIZE1*	*SIZE2*	*PROF*	*ESG_PERF*	*CapEx*	*LEV*	*BTM*
ESG_DIFF	1.000							
SIZE1	0.010	1.000						
SIZE2	-0.011	0.631	1.000					
PROF	-0.066	-0.408	0.042	1.000				
ESG_PERF	0.278	0.063	0.202	0.037	1.000			
CapEx	0.024	0.582	0.593	-0.198	0.249	1.000		
LEV	-0.056	0.367	0.028	-0.312	-0.055	0.234	1.000	
BTM	-0.057	0.494	-0.129	-0.317	-0.163	0.090	0.121	1.000

Table V outlines an excerpt of the correlations between the variables included in the regression broken down on the level of pillar score differences for the environmental (E_DIFF), social (S_DIFF) and governance pillar (G_DIFF). Interestingly, opposed to the overall score, $SIZE2$ shows a positive correlation coefficient with differences in scores for the environmental and social pillar. Same applies for $SIZE1$. Both size proxies, however, show a negative correlation coefficient for the governance pillar. $PROF$ and

ESG_PERF also, notably, show heterogenous signs in correlation coefficients throughout the three pillars. This will be subject to further analyses.

Table V: Correlations Model 1 pillar-level.

	E_DIFF	*S_DIFF*	*G_DIFF*
E/S/G_DIFF	1.000	1.000	1.000
SIZE1	0.316	0.229	-0.019
SIZE2	0.056	0.284	-0.032
PROF	-0.174	-0.119	0.019
ESG_PERF	-0.041	0.102	-0.156
CapEx	0.141	0.294	-0.022
LEV	0.185	0.052	-0.018
BTM	0.236	-0.042	-0.010

For the analysis of H3.2, a new sample is constructed in which ESG_PERF is replaced by a lagged version of this variable. The remaining variables are constructed analogue to Model 1. In this sample only firm-years of companies with ESG-ratings from the current and previous year from both Refinitiv and Bloomberg are considered. The data also is an imbalanced panel with 202 individuals for which information for every period t is available. The sample is depicted in Appendix C which includes 1,403 firm-year observations of 278 companies from 13 countries and 11 industries covering the period 2017-2022.

Table VI outlines the summarized statistics for Model 2. Similarly to Model 1, the social pillar score shows the highest mean in terms of differences between the rating agencies' scores. Moreover, the ranges of the (pillar) scores show similar tendencies as in Model 1. The average ESG_PERF of the previous year ranges from 0.98-7.33. The variable is a function of the Bloomberg and Refinitiv scores in year $(t-1)$. On average, the lagged variable exhibits lower ESG-performance compared to the non-lagged variable, as the mean, median, 25 and 75 percentile and the maximum and minimum values are below those of the non-lagged variable's statistics.

Table VI: Descriptive statistics Model 2.

Variable	N	Mean	Min	Max	SD	Prc. 25	Median	Prc. 75
Bloomberg ESG Score	1,403	4.252	1.02	8.27	1.163	3.47	4.28	5.07
Bloomberg E-Pillar	1,403	3.992	0.00	10.00	2.140	2.50	4.20	5.47
Bloomberg S-Pillar	1,403	3.241	0.00	9.29	1.622	2.02	2.98	4.20
Bloomberg G-Pillar	1,403	6.456	3.62	8.91	1.098	5.66	6.45	7.30
Refinitiv ESGC Score	1,403	6.519	0.39	9.51	1.532	5.51	6.69	7.70
Refinitiv E-Pillar	1,403	7.056	0.00	9.90	2.110	6.13	7.59	8.59
Refinitiv S-Pillar	1,403	7.580	0.24	9.82	1.667	6.70	7.92	8.84
Refinitiv G-Pillar	1,403	6.538	0.52	9.86	2.027	5.17	6.93	8.11
ESG_DIFF	1,403	2.408	0.00	6.86	1.303	1.43	2.45	3.34
E_DIFF	1,403	3.285	0.02	8.59	2.015	1.79	3.14	4.50
S_DIFF	1,403	4.415	0.00	8.59	1.902	3.11	4.57	5.78
G_DIFF	1,403	1.523	0.03	4.40	1.036	0.69	1.38	2.21
SIZE1	1,403	23.828	20.59	28.02	1.647	22.59	23.56	24.84
SIZE2	1,403	23.439	20.93	26.03	1.096	22.68	23.40	24.18
PROF	1,403	5.336	-11.13	29.08	6.104	1.51	4.49	7.94
ESG_PERF (t-1)	1,403	4.891	0.98	7.33	1.481	4.18	5.15	5.96
CapEx	1,403	19.585	13.58	23.35	1.891	18.53	19.70	20.81
LEV	1,403	94.505	0.00	454.20	93.092	35.70	63.03	123.25
BTM	1,403	64.153	-7.26	330.25	61.683	24.24	44.25	82.35

Table VII displays the correlation between the variables included in the regression of Model 2. Similar to Model 1, $SIZE1$ and $SIZE2$ show a positive correlation exceeding 0.6. Both proxies for firm size demonstrate a negative correlation with rating disagreement. Firm profitability similarly shows a negative linear connection to the extent of deviation between rating agencies. The correlation coefficient of $ESG\_PERF\ (t-1)$ with ESG_DIFF indicates a positive linear connection to the degree of rating divergence which, however, is not as high as for the non-lagged variable in Model 1 (Table IV).

Table VII: Correlations Model 2.

Correlations of regression variables								
	ESG_DIFF	*SIZE1*	*SIZE2*	*PROF*	*ESG_PERF (t-1)*	*CapEx*	*LEV*	*BTM*
ESG_DIFF	1.000							
SIZE1	-0.034	1.000						
SIZE2	-0.049	0.615	1.000					
PROF	-0.055	-0.399	0.048	1.000				
ESG_PERF (t-1)	0.182	0.118	0.268	0.078	1.000			
CapEx	-0.027	0.567	0.582	-0.181	0.278	1.000		
LEV	-0.083	0.371	0.023	-0.306	-0.017	0.225	1.000	
BTM	-0.071	0.501	-0.138	-0.309	-0.183	0.091	0.127	1.000

4. ANALYSES AND RESULTS

4.1. Analyses and Regression Results

Before beginning with the regression analyses, the rating agency scores were tested for equality of means in addition to the above shown correlation coefficients (Table III). As a result of the tests performed, the hypothesis of equality of means was rejected at the 1%-level which indicates that there indeed is a statistically significant difference between the scores from Bloomberg and Refinitiv. The test was performed for both ESGC- and ESG-scores from Refinitiv leading to the same conclusion. Moreover, the size proxies' total assets and market capitalization were tested for equality of means to verify that a difference between both proxies exists. The results of the tests allowed to reject the hypothesis of equality of means at the 1%-level. Therefore, it makes sense to analyse the effect of both size proxies individually.

Moreover, to identify multicollinearity the VIF test was applied to all models. Due to noticeable VIF results for the natural logarithm of total assets and CapEx the latter variable was excluded from the models in which total assets are featured as independent variable (Model 1.1, Model 2.1).

All models including the models broken down on a pillar level were tested for global significance applying F-Tests at a probability level of 5%. While all models are globally significant meaning all regressors jointly are significant, their explanatory value varies. For both variations of Model 1 the explanatory value was highest at an overall ESG-score level. R^2 was slightly lower for environmental and social score divergence as dependent variable. For both variations of Model 2 the regression on an environmental and social pillar score basis demonstrated higher explanatory values than the overall ESG-score. An overview of R^2 per model can be redeemed from Table VIII and Table XI. The average ESG-performance of the past period may, therefore, be especially relevant for the measurement of environmental and social metrics. The model is the least suitable in explaining differences in governance pillar ratings being only slightly globally significant. This implies that there may be other factors more relevant for the disagreement on governance-related topics, as for instance state ownership (Christensen

et al., 2022; Liu, 2022). Moreover, as demonstrated by the descriptive statistics, the consensus in general is higher on governance ratings.

Table VIII shows the regression results of Model 1.1 and Model 1.2 at an overall ESG-rating level as well as for each pillar score individually. The results of Model 1.1 and Model 1.2 are each displayed in the columns (1.1) and (1.2), respectively. In the primary model, where the overall ESG-rating divergence is the dependent variable, both size variables are statistically significant at the 1%-level. However, they both are negative implying that a 1% increase of total assets or market capitalization, respectively, will lead to a reduction of ESG-rating divergence by 0.001495 or 0.002161, respectively. H1, therefore, is rejected for the overall ESG-rating. Even the reverse effect of size on ESG-rating divergence is indicated. One reason may be, that larger companies can provide more clarifying information which may reduce disagreement as Kimbrough et al. (2024) suggest in their research. On an individual pillar score level, the size variables are only statistically significant for the social dimension of ratings. In this case, $SIZE1_{i,t}$ is statistically significant at a 1%-level, $SIZE2_{i,t}$ is statistically significant at a 10% level, implying that, for instance, a 1% increase of total assets yields a 0.004458 increase in social pillar rating disagreement. Thus, H1 is not rejected for the social pillar. Increasing size of companies, therefore, seems to especially lead to more disagreement of rating agencies in the conceptualization and measurement of social metrics.

Regarding H2, $PROF_{i,t}$ is statistically significant within both models at the overall ESG-score level. H2 is not rejected as the sign of the coefficient is negative aligning with the hypothesis. On an individual pillar score level, $PROF_{i,t}$ only displays statistically significant values for social pillar score differences. For S_DIFF H2 is also not rejected. The average ESG-rating of a company is statistically significant for the overall ESG-rating divergence as well as for the social and governance dimensions. Contrary to H3.1, however, the coefficients indicate that an increase in average ESG-ratings yield an increase of ESG-rating divergence (for Model 1.1 and Model 1.2). This leads to H3.1 being rejected for overall ESG-rating divergence. The same conclusion is reached on the social pillar score level. It should be mentioned that the average ESG-performance in this regression is only a function of Bloomberg and Refinitiv ratings and not a complete market-based reflection of the overall ESG-performance. The analyses could be repeated with an alternative, broadened calculation of performance. Interestingly, the coefficient

shows reversed signs for the governance dimension. H3.1, consequently, is not rejected for the governance pillar. This implies that a decrease in divergence of the governance pillar is associated with an on average higher ESG-score. In total, the results demonstrate that there are, indeed, differences between ratings in environmental, social and governance matters. The statistical significance of the independent variables is highest on an overall ESG-score basis in which the effects are accumulated. The effects of weightings of field and pillar ratings cannot be out ruled. On the overall ESG-score level apart from industry, country and year fixed effects only $LEV_{i,t}$ as control variable is statistically significant (Model 1.2). The control variable $CAPEX_{i,t}$ is statistically significant in the environmental and social dimension (Model 1.2). Leverage appears to be relevant for divergence in the environmental pillar, $BTM_{i,t}$ in the governance pillar. It must be highlighted, that the hypotheses suit the social pillar score divergences most.

Table VIII: Regression Results Model 1.

Dependent Variable	*ESG_DIFF*		*E_DIFF*		*S_DIFF*		*G_DIFF*	
Model	*(1.1)*	*(1.2)*	*(1.1)*	*(1.2)*	*(1.1)*	*(1.2)*	*(1.1)*	*(1.2)*
SIZE1	**-0.1495*** **	-	**0.1563**	-	**0.4458*** **	-	**-0.0013**	-
SIZE2	-	**-0.2161*** **	-	**-0.1320**	-	**0.2362***	-	**-0.0229**
PROF	**-0.0250*** **	**-0.0156** **	**-0.0196**	**-0.0197**	**-0.0207***	**-0.0387*** **	**0.0012**	**0.0021**
ESG_PERF	**0.6135*** **	**0.6077*** **	**0.1534**	**0.1377**	**0.1615***	**0.1825** **	**-0.1258*** **	**-0.1336*** **
CapEx	-	0.0092	-	0.2141**	-	0.1539**	-	0.0298
LEV	-0.0009	-0.0014**	0.0020*	0.0017	-0.0003	0.0003	-0.0002	-0.0003
BTM	0.0003	-0.0015	0.0022	0.0019	-0.0021	0.0012	-0.0014*	-0.0016**
Constant	3.5486***	4.989***	-1.2239	1.2427	-6.9872***	-5.4770**	2.8613**	2.8194**
Observations	1,639	1,639	1,639	1,639	1,639	1,639	1,639	1,639
R^2	0.3098	0.3157	0.2468	0.2575	0.2896	0.2783	0.0904	0.0915

This table presents the regression results of Model 1.1 and Model 1.2 on an overall ESG score level as well as for each pillar dimension. The dependent variable *ESG_DIFF* is the absolute difference between both ratings. Definitions of the variables are provided in Appendix A. Country, industry, and year fixed effects are excluded from this table and are shown separately. Bold text denotes the key regression variables and their estimated coefficients. The standard deviations are clustered at the firm year level. Statistical significance at the 1% ,5% and 10% levels are shown by ***, ** and *, respectively.

Table IX outlines the effects of the industry and country dummy variables controlled for in Model 1.1 and Model 1.2. There are no major differences in significance of parameters between the two sub-models. In general, most countries do have a statistically significant impact (in relation to the omitted country Belgium) on the dependent variable.

Country fixed effects, correspondingly, do have influence on the extent of disagreement between Bloomberg and Refinitiv ratings. This may be attributable to the fact that ESG disclosures are affected by country-level characteristics, such as national regulations (Christensen et al., 2022) and the political or cultural system (Baldini et al., 2018). This topic could be subject to future research. The industries Consumer Discretionary, Consumer Staples, Financials, Health Care and Industrials are statistically significant (in relation to the omitted industry Basic Materials). Industry at least to some extent is relevant for the degree of ESG-rating divergence. It has already been demonstrated that industry-specific risk profiles and stakeholder needs influence sustainability disclosures (Arkoh et al., 2024). Moreover, depending on the specific scoring methodology the ratings are tailored to the specific industry and industry-specific topics (Bloomberg, 2023; Refinitiv, 2022) which can be a further source of discrepancies.

Table IX: Regression Results Model 1: country and industry fixed effects.

Dependent Variable	ESG_DIFF		**Dependent Variable**	ESG_DIFF	
Model	(1.1)	(1.2)	*Model*	(1.1)	(1.2)
Country			*Industry*		
Denmark	-0.5434**	-0.5493**	*Consumer Discretionary*	0.7151***	0.7426***
Finland	-1.0325***	-1.0774***	*Consumer Staples*	0.8459***	0.8872***
France	-0.0469	-0.0071	*Energy*	-0.3095	-0.2355
Germany	-0.5615***	-0.5292**	*Financials*	1.3561***	1.2235***
Ireland; Republic of	-1.5831***	-1.6000***	*Health Care*	1.2451***	1.3050***
Luxembourg	-0.6582***	-0.6680***	*Industrials*	0.4768**	0.4677**
Mexico	-0.2896	-0.2788	*Real Estate*	0.0745	0.0675
Netherlands	-0.6417***	-0.5978***	*Technology*	0.3000	0.3245
Spain	-0.2553	-0.2930	*Telecommunications*	0.3992	0.4193
Sweden	-0.1537	-0.1472	*Utilities*	0.2072	0.1816
Switzerland	-0.9517***	-0.8754***			
United Kingdom	-0.6726***	-0.6783***			

The table shows the regression results for industry and country fixed effects of Model 1.1 and Model 1.2 for the overall ESG score. The industry dummy variables are shown in relation to the the omitted industry Basic Materials. The country dummy variables are shown in relation to the omitted country Belgium. The dependent variable ESG_DIFF is the absolute difference between both ratings. The standard deviations are clustered at the firm level. Statistical significance the the 1%, 5%, and 10% levels are shown by ***, ** and *, respectively.

Time fixed effects are shown in Table X. The yearly coefficients are all statistically significant at a 1%-level in relation to the omitted year 2016. The coefficients are negative and increase over time indicating that, firstly, year fixed effects are relevant to rating disagreement, and, secondly, from 2016-2022 year fixed effects continuously reduced rating disagreement. Time series analyses would be an option for future research to shed further light on this development.

Table X: Regression Results Model 1: year fixed effects.

Dependent Variable	*ESG_DIFF*	
Model	*(1.1)*	*(1.2)*
Year		
2017	-0.3192***	-0.3187***
2018	-0.4326***	-0.4376***
2019	-0.6494***	-0.6252***
2020	-1.0326***	-0.9854***
2021	-1.1800***	-1.1513***
2022	-1.2031***	-1.1900***

This table presents the regression results for year fixed effects of Model 1.1 and Model 1.2 for the overall ES 2016 is the omitted year within the yearly dummies. The dependent variable ESG_DIFF is the absolute differe both ratings. The standard deviations are clustered at the firm year level. Statistical significance at the 1% ,5% levels are shown by ***, ** and *, respectively.

In the following, Table XI provides the regression results for Model 2.1 and Model 2.2 on an overall ESG-score level as well as for each pillar score. The results lead to similar conclusions as reached for Model 1; however, the level of significance and the values of the coefficients differ. The coefficients for both size variables are negative and statistically significant at least at the 5% level. Analogue to Model 1, this suggests that H1 is rejected, and the linear causality is even negative instead of positive as assumed by H1. Only for the social pillar score H1 is not rejected, the coefficients being positive and statistically significant at least at the 5% level for both size variables which indicates that an increase in size will lead to an increase in rating differences for the social pillar.

The reducing impact of financial profitability on ESG-rating divergence is not rejected for the overall ESG-score as well as for the social pillar. In lack of statistically significant results, no conclusion on the impact on environmental and governance rating differences can be drawn.

The average ESG-performance of a firm in the previous period is highly statistically significant for the overall ESG-rating divergence. However, the positive sign indicates that analogous to current ESG-performance (Model 1) an increase in previous ESG performance yields an increase in ESG-rating divergence. Yet, the coefficients in absolute terms are smaller for the performance of the previous period compared to the current period. It may be possible, that analysts are led more by their current impressions than by their image of the previous period. The impact of the ESG-rating of the previous period appears to be relevant but on a smaller scale. Only for governance pillar divergence

a negative linear relationship between lagged average ESG-performance as independent variable on the one side and rating divergence as dependent variable on the other side is implied. H3.2 is, thus, only not rejected on the governance pillar level. It must, however, be pointed out that the overall explanatory value of Model 2.1 and Model 2.2 for governance rating divergence is rather low. For the divergence of the overall ESG-score, social and environmental score the reverse impact of what was originally assumed by H3.2 is displayed.

In conclusion, Model 2 does not generate noticeable differences compared to Model 1. The impact of prior year ESG-performance on current year rating divergence can overall be considered lower than divergences caused by current year ESG-performance. Only for the social pillar the coefficients of $ESG_PERF_{i,(t-1)}$slightly exceed the corresponding coefficients for $ESG\_PERF_{i,t}$. The control variable $LEV_{i,t}$ is statistically significant for overall ESG-rating divergence and has a reducing impact. $BTM_{i,t}$ is only statistically significant within Model 2.1 As in Model 1, the hypotheses are most true for the social pillar scores..

Table XI: Regression Results Model 2.

Dependent Variable	ESG_DIFF		E_DIFF		S_DIFF		G_DIFF	
Model	(2.1)	(2.2)	(2.1)	(2.2)	(2.1)	(2.2)	(2.1)	(2.2)
SIZE1	**-0.1153****	-	**0.1317**	-	**0.4086*****	-	**-0.0129**	-
SIZE2	-	**-0.2035*****	-	**-0.1589**	-	**0.2406****	-	**-0.0336**
PROF	**-0.0279*****	**-0.0201*****	**-0.0218**	**-0.0201**	**-0.0250****	**-0.0419*****	**0.0014**	**0.0031**
ESG_PERF (t-1)	**0.2556*****	**0.2532*****	**0.1472****	**0.1248***	**0.1893*****	**0.2033*****	**-.0611***	**-0.0700****
CapEx	-	0.0182	-	0.2189**	-	0.1106	-	0.0347
LEV	-0.0015***	-0.0019***	0.0018	0.0014	-0.0005	0.0003	0.0000	-0.0002
BTM	-0.0008	-0.0024**	0.0027	0.0021	-0.0019	0.0014	-0.0011	-0.0015*
Constant	4.6484***	6.3701***	-0.4726	1.9704	-5.9090***	-4.4302**	2.6847**	2.5261**
Industry FE	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Year	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.1832	0.1910	0.2361	0.2486	0.2937	0.2802	0.0812	0.0826
Observations	1,403	1,403	1,403	1,403	1,403	1,403	1,403	1,403

This table presents the regression results of Model 2.1 and Model 2.2 on an overall ESG score level as well as for each pillar dimension. The dependent variable *ESG_DIFF* is the absolute difference between both ratings. Definitions of the variables are provided in Appendix A. Country, industry, and year fixed effects are not shown separately. Bold text denotes the key regression variables and their estimated coefficients. The standard deviations are clustered at the firm year level. Statistical significance at the 1% ,5% and 10% levels are shown by ***, ** and *, respectively.

4.2.Robustness Tests and Additional Tests

In the main analyses, for both, the calculation of ESG-rating disagreement and average ESG-performance, Refinitiv's ESGC-rating which accounts for controversial events (Refinitiv, 2022) has been applied. According to Bloomberg (2023), Bloomberg's ESG-scores include negative events in their rating methodology. As it is, however, not fully transparent on how exactly controversies are considered the decision on using Refinitiv's ESGC-score for the analyses can be challenged and was subject to the following robustness checks. Table XII shows the regression results of Model 1.1 and Model 1.2 with ESG_DIFF_RO being measured as the absolute difference between Bloomberg's ESG and Refinitiv's ESG-rating instead of the ESGC-score. Moreover, $ESG\_PERF_{i,t}\ RO$ is calculated as the average ESG-rating of Bloomberg's ESG and Refinitiv's ESG-score. The definitions of variables applied in the robustness tests are listed in Appendix B. For overall ESG-rating divergence using Refinitiv's ESG-score instead of the ESGC-score regarding H2 and H3.1 the same conclusions are reached: H2 is not rejected, H3.1 is rejected. Both size proxies show noticeable differences to the original model. $SIZE1_{i,t}$ is statistically significant at the 1%-level with the coefficient being positive. In this case H1 would not be rejected. $SIZE2_{i,t}$, however, is not statistically significant. This leads to the conclusion, that it is meaningful which scores exactly are compared for the relevance of firm size for ESG-rating divergence. An explanation for firm size having a decreasing effect on rating divergence in the original model may be that for Refinitiv and Bloomberg specifically the degree of disagreement is lower when Refinitiv has identified controversial events (see the following analysis in Table XIII). For larger companies more information and news are available (Drempetic et al., 2020) and the exposure to public pressure on disclosing information is higher (Hackston & Milne, 1996; Michelon, 2011) which makes it easier to identify controversial events. However, this would need to be subject to further analyses before being able to confirm the hypothesis stated. On a pillar score level, mostly, the original conclusions reached are confirmed. On the environmental pillar score level, compared to the original model, $ESG\_PERF_{i,t}\ RO$ becomes statistically significant. The signs of the coefficients, however, are contradicting H3.1, similarly to the original model.

Table XII: Robustness Test Refinitiv ESG-score.

Dependent Variable	**ESG_DIFF RO**		E_DIFF		S_DIFF		G_DIFF	
Model	(1.1)	(1.2)	(1.1)	(1.2)	(1.1)	(1.2)	(1.1)	(1.2)
SIZE1	0.1698***	-	0.0623	-	0.3642***	-	0.0344	-
SIZE2	-	0.0155	-	-0.1857	-	0.1771	-	0.0049
PROF	-0.0212***	-0.0261***	-0.0219	-0.0182	-0.0227*	-0.0370***	0.0021	0.0013
ESG_PERF RO	0.3940***	0.4086***	0.2888**	0.2666**	0.2665**	0.3045***	-0.1471***	-0.1597***
CapEx	-	0.0975**	-	0.1820*	-	0.1205*	-	0.0424
LEV	0.0000	0.0001	0.0020*	0.0015	-0.0004	0.0001	-0.0001	-0.0002
BTM	0.0008	0.0017	0.0024	0.0015	-0.0020	0.0006	-0.0013*	-0.0012
Constant	-2.8975**	-1.4004	0.2968	2.5310	-5.6022***	-3.9970*	2.1300*	2.0310*
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Industry FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Observations	1,639	1,639	1,639	1,639	1,639	1,639	1,639	1,639
R^2	0.3636	0.3618	0.2562	0.266	0.2971	0.2887	0.0922	0.0945

This table presents the regression results of Model 1.1 and Model 1.2 on an overall ESG score level as well as for each pillar dimension with the variation that ESG_DIFF and ESG_PERF are calculated using Refinitiv's ESG score instead of the ESGC score. The dependent variable ESG_DIFF_RO is the absolute difference between both ratings. Definitions of the variables are provided in Appendix A and B. Country, industry, and year fixed effects are not shown separately. Bold text denotes the regression variables and their estimated coefficients changed with respect to the originalal model. The standard deviations are clustered at the firm year level. Statistical significance at the 1% ,5% and 10% levels are shown by ***, ** and *, respectively.

Original Model 2.1 and 2.2 were also re-run with Refinitiv's ESG-score. The results aligned with those for Model 1 and are attached in Appendix D. As demonstrated by the original model, the coefficients of the non-lagged variable exceed the lagged ESG-performance variable. The coefficients, nevertheless, are statistically significant at the 1%-level.

To further investigate the relevance of controversies on ESG-rating divergence in the original model, the model was supplemented by the dummy variable $CONTROV_{i,t}$. The variable takes the value 1 in case controversies exist and 0 in case no controversies exist. In case Refinitiv's ESGC-score equals the ESG-score no controversies are recorded. It must be noted that $CONTROV_{i,t}$ was constructed based on Refinitiv's evaluation of controversies. Therefore, completeness and accuracy of $CONTROV_{i,t}$ depend on Refinitiv's assessment. The analysis is only conducted on an overall pillar score level as it cannot be said to which pillar the controversies refer to. Table XIII provides the results of the altered regression model including variable $CONTROV_{i,t}$. The results provide evidence that controversial events are statistically significant for ESG-rating divergence

between Bloomberg and Refinitiv. The coefficient is negative which implies that ESG-rating divergence is on average 0.9973 lower in case Refinitiv has identified a controversial event in relation to when Refinitiv has not identified a controversial event. This is not surprising as the descriptive statistics already showed that the mean of Refinitiv's ESGC-ratings is significantly higher than Bloomberg's ESG-ratings. Refinitiv's ESG-ratings have an even higher mean as they are not reduced due to controversies. The absolute difference between both ratings, therefore, on average is lower when Refinitiv's ESGC-score is used. This demonstrates how rating agencies have different methodologies of measuring positive and negative aspects of ESG. It must, however, be noted that this analysis is specific to Bloomberg and Refinitiv scores and cannot simply be transferred to any other rating agency. Regarding H2 and H3.1 the same conclusions are reached as in the original model.

Table XIII: Regression Results for Model 1.1 and 1.2 supplemented with CONTROV variable.

Dependent Variable	ESG_DIFF	
Model	(1.1)	(1.2)
SIZE1	0.0249	-
SIZE2	-	-0.1040
PROF	-0.0203***	-0.0184***
ESG_PERF	0.5005***	0.5047***
CONTROV (=1)	-0.9973***	-0.9652***
CapEx	-	0.0612*
LEV	-0.0006	-0.0008
BTM	0.0005	0.0000
Constant	0.0611	1.7985
Observations	1639	1639
R^2	0.3784	0.3818

This table presents the regression results of Model 1.1 and Model 1.2 on an overall ESG score level supplemented by the variable CONTROV. Definitions of the variables are provided in Appendix A and B. The dependent variable ESG_DIFF is the absolute difference between both ratings. Country, industry, and year fixed effects are not shown separately. Bold text denotes the regression variables and their estimated coefficients changed with respect to the original model. The standard deviations are clustered at the firm year level. Statistical significance at the 1% ,5% and 10% levels are shown by ***, ** and *, respectively.

As a further robustness test, the size proxy of Model 1 was substituted by an alternative calculation $SIZE3_{i,t}$ measured as natural logarithm of total revenue as revenue has regularly been applied as size proxy in research (see Gallo & Christensen, 2011; Orlitzky, 2001). Due to their business model and the underlying accounting rules, companies in the financial sector show their main earnings differently. For this, this industry was excluded from the following analysis. The results of the regression come to

the same conclusion as reached for the original analyses of Model 1 shown in Table VIII. The results of the regression with $SIZE3_{i,t}$ are displayed in Appendix E. Moreover, a robustness test was performed on the $PROF_{i,t}$ variable. Instead of ROA, the operating margin was utilised in the regression ($PROF\_RO_{i,t}$). Again, the financial sector was excluded from this analysis. The results of the regression analysis are attached in Appendix F. They lead to the same conclusions as reached for Model 1.1. An overview of all regression results obtained from the original analyses and the robustness tests can be retrieved from Appendix G.

Finally, depending on the outcome of the Hausman test, random or fixed effects estimators were applied to Model 1 to test the robustness of OLS regression. Regarding the size variable, in case of being significant, the sign of the coefficients confirmed H1 which contradicts the results of the original analyses but aligns with the robustness tests performed on Refinitiv's ESGC-score. The profitability and ESG-performance variable, when being statistically significant in the original and the alternative model, aligned with the original conclusions drawn based on the OLS regression.

5. CONCLUSION

To conclude, the analyses performed contribute to the knowledge on ESG-rating divergence and relevant firm-level factors and, thus, set a starting point for further, more detailed analysis. The results add to existing literature on the evidence of the existence of ESG rating divergence and alarm researchers and investors to be careful blindly relying on individual ESG-ratings. By understanding determinants hindering or favouring rating convergence a more profound understanding of the deeper root of rating divergence can be generated. It is noticeable that the hypotheses were most suited to explain social pillar score divergence. The results of the analyses performed confirm a decreasing effect of financial performance on ESG-rating divergence. Statistical significance allowed to come to this conclusion for the overall ESG- and social pillar score. On average higher ESG-performance was shown to lead to higher ESG-rating disagreement. This is true for both, ESG-performance of the current and previous period. The impact of the current period, however, is stronger. Only for disagreement on governance topics a reducing tendency of higher average ESG-performance could be found.

The analyses highlighted that in cannot generally be confirmed that an increase in firm size yields an increase in ESG-rating divergence. The results depend on the compared scores (Refinitiv's ESG/ESGC-score) and the dimension considered (overall ESG-score/individual pillar score). Firm size contributes consistently to divergence for the social pillar throughout all models. The direction of the impact of firm size for the overall ESG-score varies depending on the details of the calculation while in total remaining statistically significant. For the original model even, a reverse impact than originally stated in the hypothesis was shown. The differences in the results depending on the exact ESG-scores used outline the relevance of individual rating agencies' rating methodology. Thus, this work demonstrates the sensitivity of ratings to small variations and the construction of the ratings. Moreover, the reduction of divergence in the case of existing controversial events contributes a new aspect to the research in this area and sets a possibility for future studies. For future research it is, beyond that, suggested to further understand the differences in methodology to, based on that, have an even more profound base for the analyses of ESG-rating divergence. For this, however, the transparency of rating methodologies of ESG-rating providers and the use of common ESG definitions must be increased.

The lack of detailed insight into scope definitions and procedures of rating agencies hinders the analyses of the impact of these aspects on rating divergence. In this context. Berg et al. (2022) who specifically investigated the impact of scope, measurement and weight differences on rating divergence already called for more transparency from rating agencies. This is especially relevant due to the high relevance of ESG-performance measurements (Berg et al., 2022; Drempetic et al., 2020; Eccles & Stroehle, 2018; Sultana et al., 2018) for research (see, e.g. Albuquerque et al., 2019; Drempetic et al., 2020; Lins et al., 2017; Sharma et al., 2024) and investors (Hartzmark & Sussman, 2019) especially in the light of greenwashing concerns (Biju et al., 2023; Wu et al., 2020). The results call for action as ESG-rating divergence can adversely impact their trustworthiness and, thus, hinder their use. Future research may shed further light on how attempts of standardization of ESG disclosures can help to reduce the extent of disagreement. Similarly to Dumrose et al. (2022), who demonstrated the impact the EU taxonomy can have on reducing rating disagreement on environmental ratings, the impact of recent regulatory concepts can be evaluated. Finally, future research could focus on the

development of rating divergence over time. A time series analyses could be conducted to further investigate if specific factors significantly contributed to the development over time. This could be conducted in conjunction with the investigation of the impact of standardization and regulations.

This study has limitations which must be noted. Firstly, the analyses are solely based on the disagreement of two rating agencies and cannot be generalized for all rating providers. For future analyses more rating agencies could be included. KLD being the most widely used ESG data vendor in research since the 1990s (Tang et al., 2022) is not represented in these analyses. Moreover, to increase the comparability between ratings the divergence measurement can be based on percentile rank scores (Kimbrough et al., 2024). Further, the analyses are limited to hand-selected European indices leading to a convenient sample. In further analyses the sample size could be increased and broadened. Due to limited transparency the analyses are limited with respect to the consideration of specific scope and measurement differences of the rating agencies. Moreover, the calculation of average ESG-performance is based on Bloomberg and Refinitiv's ratings rather than a composite broader market-based ESG-performance indicator.

REFERENCES

Aabo, T., & Giorici, I. C. (2023). Do female CEOs matter for ESG scores? Global Finance Journal, 56. https://doi.org/10.1016/j.gfj.2022.100722

Albuquerque, R., Koskinen, Y., & Zhang, C. (2019). Corporate social responsibility and firm risk: Theory and empirical evidence. *Management Science*, *65*(10), 4451–4469. https://doi.org/10.1287/mnsc.2018.3043

Ameer, R., & Othman, R. (2012). Sustainability practices and corporate financial performance: A study based on the top global corporations. *Journal of Business Ethics*, *108*(1), 61–79. https://doi.org/10.1007/s10551-011-1063-y

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. https://doi.org/10.2469/faj.v74.n3.2

Arkoh, P., Costantini, A., & Scarpa, F. (2024). Determinants of sustainability reporting: A systematic literature review. *Corporate Social Responsibility and Environmental Management* 31(3) 1578–1597. https://doi.org/10.1002/csr.2645

Avramov, D., Cheng, S., Lioui, A., & Tarelli, A. (2022). Sustainable Investing with ESG Rating Uncertainty * Sustainable Investing with ESG Rating Uncertainty. *Journal of Financial Economics*, *145*(2), 642–664. https://doi.org/https://doi.org/10.1016/j.jfineco.2021.09.009

Baldini, M., Maso, L. D., Liberatore, G., Mazzi, F., & Terzani, S. (2018). Role of Country- and Firm-Level Determinants in Environmental, Social, and Governance Disclosure. *Journal of Business Ethics*, *150*(1), 79–98. https://doi.org/10.1007/s10551-016-3139-1

Balzer, W. K., & Sulsky, L. M. (1992). Halo and Performance Appraisal Research: A Critical Examination. *Journal of Applied Psychology*, *77*(6), 975–985. https://doi.org/https://doi.org/10.1037/0021-9010.77.6.975

Bartolacci, F., Caputo, A., & Soverchia, M. (2020). Sustainability and financial performance of small and medium sized enterprises: A bibliometric and systematic

literature review. *Business Strategy and the Environment*, *29*(3), 1297–1309. https://doi.org/10.1002/bse.2434

Berg, F., Fabisik, K., & Sautner, Z. (2021). Rewriting History II: The (Un)Predictable Past of ESG Ratings. *European Corporate Governance Institute–Finance Working Paper*, 708(2020). https://ssrn.com/abstract=3722087

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

Biju, A. V. N., Kodiyatt, S. J., Krishna, P. P. N., & Sreelekshmi, G. (2023). ESG sentiments and divergent ESG scores: suggesting a framework for ESG rating. *SN Business & Economics*, *3*(12). https://doi.org/10.1007/s43546-023-00592-4

Billio, M., Costola, M., Hristova, I., Latino, C., & Pelizzon, L. (2021). Inside the ESG ratings: (Dis)agreement and performance. *Corporate Social Responsibility and Environmental Management*, *28*(5), 1426–1445. https://doi.org/10.1002/csr.2177

Bloomberg (2020). *Environmental & Social Scores.* Bloomberg. https://data.bloomberglp.com/professional/sites/10/ESG_Environmental-Social-Scores.pdf

Bloomberg (2023). *Bloomberg ESG Scores*. Bloomberg. https://hr.bloombergadria.com/data/files/Pitanja%20i%20odgovori%20o%20Bloomberg%20ESG%20Scoreu.pdf

Boffo, R., & Patalano, R. (2020). *ESG Investing: Practices, Progress and Challenges*. OECD. www.oecd.org/finance/ESG-Investing-Practices-Progress-and-Challenges.pdf

Bonsall, S. B., & Miller, B. P. (2017). The impact of narrative disclosure readability on bond ratings and the cost of debt. *Review of Accounting Studies*, *22*(2), 608–643. https://doi.org/10.1007/s11142-017-9388-0

Boso, N., Danso, A., Leonidou, C., Uddin, M., Adeola, O., & Hultman, M. (2017). Does financial resource slack drive sustainability expenditure in developing economy small and medium-sized enterprises? *Journal of Business Research*, *80*, 247–256. https://doi.org/10.1016/j.jbusres.2017.06.016

Campbell, D., Craven, B. and Shrives, P. (2003). Voluntary social reporting in three FTSE sectors: a comment on perception and legitimacy. *Accounting, Auditing & Accountability Journal*, 16 (4), 558-581. https://doi.org/10.1108/09513570310492308

Cantor, R., & Packer, F. (1994). The Credit Rating Industry. *Quarterly Review*, *19*, 1–26. https://doi.org/10.3905/jfi.1995.408153

Capizzi, V., Gioia, E., Giudici, G., & Tenca, F. (2021). THE DIVERGENCE OF ESG RATINGS: AN ANALYSIS OF ITALIAN LISTED COMPANIES. *Journal of Financial Management, Markets and Institutions*, *9*(2). https://doi.org/10.1142/S2282717X21500067

Chatterji, A. K., Durand, R., Levine, D. I., & Touboul, S. (2016). Do ratings of firms converge? Implications for managers, investors and strategy researchers. *Strategic Management Journal*, *37*(8), 1597–1614. https://doi.org/10.1002/smj.2407

Chatterji, A. K., & Levine, D. (2006). Breaking Down The Wall Of Codes: Evaluating Non-Financial Performance Measurement. *California Management Review*, *48*(2), 29–51. https://doi.org/https://doi.org/10.2307/41166337

Chen, R. C. Y., Hung, S. W., & Lee, C. H. (2017). Does corporate value affect the relationship between Corporate Social Responsibility and stock returns? *Journal of Sustainable Finance and Investment*, *7*(2), 188–196. https://doi.org/10.1080/20430795.2016.1272947

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

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

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of Management*, *37*(1), 39–67. https://doi.org/10.1177/0149206310388419

Cookson, J. A., & Niessner, M. (2020). Why Don't We Agree? Evidence from a Social Network of Investors. *The Journal of Finance*, *75*(1), 173–228. https://doi.org/https://doi.org/10.1111/jofi.12852

Crespi, F., & Migliavacca, M. (2020). The determinants of ESG rating in the financial industry: The same old story or a different tale? *Sustainability (Switzerland)*, *12*(16). https://doi.org/10.3390/SU12166398

Dathe, T., Dathe, R., Dathe, I., & Helmold, M. (2022). *Corporate Social Responsibility (CSR), Sustainability and Environmental Social Governance (ESG)*. Springer Nature Switzerland. https://doi.org/https://doi.org/10.1007/978-3-030-92357-0_2

De Villiers, C., & Alexander, D. (2014). The institutionalisation of corporate social responsibility reporting. *British Accounting Review*, *46*(2), 198–212. https://doi.org/10.1016/j.bar.2014.03.001

DiMaggio, P., & Powell, W. W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields (translated by G. Yudin). *American Sociological Review*, *48*(2), 147–160. https://doi.org/10.17323/1726-3247-2010-1-34-56

Dimson, E., Marsh, P., & Staunton, M. (2020). *Summary Edition Credit Suisse Global Investment Returns Yearbook 2020*. Credit Suisse. https://journal.tinkoff.ru/media/credit-suisse-global-investment-returns-yearbook-2020-summary-edition.pdf

Döpfner, C. (2016). Wie nützlich sind Nachhaltigkeitsratings für eine nachhaltige Entwicklung von Unternehmen?. In: Kopp, H. (Eds.) *CSR und Finanzratings. Management-Reihe Corporate Social Responsibility*, Berlin: Springer Gabler, 53-66. https://doi.org/10.1007/978-3-662-47461-7_6

Dowling, J., & Pfeffer, J. (1975). Organizational legitimacy: social values and organizational behaviour. *The Pacific Sociological Review*, *18*, 122–136. https://doi.org/https://doi.org/10.2307/1388226

Drempetic, S., Klein, C., & Zwergel, B. (2020). The Influence of Firm Size on the ESG Score: Corporate Sustainability Ratings Under Review. *Journal of Business Ethics*, *167*(2), 333–360. https://doi.org/10.1007/s10551-019-04164-1

Dumrose, M., Rink, S., & Eckert, J. (2022). Disaggregating confusion? The EU Taxonomy and its relation to ESG rating. *Finance Research Letters*, *48*, 102928. https://doi.org/10.1016/j.frl.2022.102928

Eccles, R. G., & Stroehle, J. C. (2018). *Exploring Social Origins in the Construction of Environmental, Social and Governance Measures*. SSRN. https://doi.org/http://dx.doi.org/10.2139/ssrn.3212685

El Ghoul, S., Guedhami, O., Kwok, C. C. Y., & Mishra, D. R. (2011). Does corporate social responsibility affect the cost of capital? *Journal of Banking and Finance*, *35*(9), 2388–2406. https://doi.org/10.1016/j.jbankfin.2011.02.007

Elkington, J. (1997). *Cannibals with Forks: The Triple Bottom Line of 21st Century Business*. Capstone.

Escrig-Olmedo, E., Fernández-Izquierdo, M. ángeles, Ferrero-Ferrero, I., Rivera-Lirio, J. M., & Muñoz-Torres, M. J. (2019). Rating the raters: Evaluating how ESG rating agencies integrate sustainability principles. *Sustainability*, *11*(3), 915. https://doi.org/10.3390/su11030915

European Commission. (2020). *Study on Sustainability-Related Ratings, Data and Research*. European Commission. https://a.storyblok.com/f/216981/x/3d3b57bb4f/studie-der-eu-kommission-2020.pdf

Freeman, R. E., Harrison, J., & Wicks, A. C. (2007). *Managing for Stakeholders: Survival, reputation and success*. Caravan. https://doi.org/10.12987/9780300138498

Freeman, R. E. (2010). *Strategic Management: A Stakeholder Approach*. Cambridge: Cambridge University Press**.** https://doi.org/10.1017/CBO9781139192675

Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance and Investment*, *5*(4), 210–233. https://doi.org/10.1080/20430795.2015.1118917

Friedman, M. (1970). The Social Responsibility of Business Is to Increase Its Profits. In: Zimmerli, W.C., Holzinger, M., Richter, K. (Eds) *Corporate Ethics and Corporate*

Governance, Berlin: Springer, 173-178. https://doi.org/10.1007/978-3-540-70818-6_14

Gallo, P. J., & Christensen, L. J. (2011). Firm size matters: An empirical investigation of organizational size and ownership on sustainability-related behaviors. *Business and Society, 50*(2), 315–349. https://doi.org/10.1177/0007650311398784

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

Gregory, R. P. (2024). The influence of firm size on ESG score controlling for ratings agency and industrial sector. *Journal of Sustainable Finance and Investment*, *14*(1), 86–99. https://doi.org/10.1080/20430795.2022.2069079

GSIA. (2022). *Global Sustainable Investment Review 2022*. Global Sustainable Investment Alliance. https://www.gsi-alliance.org/wp-content/uploads/2023/12/GSIA-Report-2022.pdf

Hackston, D., & Milne, M. J. (1996). Some determinants of social and environmental disclosures in New Zealand companies. *Accounting, Auditing and Accountability Journal*, *9*(1), 77–108. https://doi.org/http://dx.doi.org/10.1108/09513579610109987

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

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

Hasse, R., & Krücken, G. (2009). Neo-institutionalistische Theorie. In G. Kneer & M. Schroer (Eds.), *Handbuch Soziologische Theorien* (237–251). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91600-2_12

Hedesström, M., Lundqvist, U., & Biel, A. (2011). Investigating consistency of judgement across sustainability analyst organizations. *Sustainable Development*, *19*(2), 119–134. https://doi.org/10.1002/sd.511

Hörisch, J., Johnson, M. P., & Schaltegger, S. (2015). Implementation of Sustainability Management and Company Size: A Knowledge-Based View. *Business Strategy and the Environment*, *24*(8), 765–779. https://doi.org/10.1002/bse.1844

In, S. Y., Rook, D., & Monk, A. (2019). Integrating Alternative Data (Also Known as ESG Data) in Investment Decision Making. *Global Economic Review*, *48*(3), 237–260. https://doi.org/10.1080/1226508X.2019.1643059

Jámbor, A., & Zanócz, A. (2023). The Diversity of Environmental, Social, and Governance Aspects in Sustainability: A Systematic Literature Review. *Sustainability*, *15*(18). https://doi.org/10.3390/su151813958

Jensen, M. C., & Meckling, W. H. (1976). Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. *Journal of Financial Economics*, *3*(4), 305–360. https://doi.org/https://doi.org/10.1016/0304-405X(76)90026-X

Jonsdottir, B., Sigurjonsson, T. O., Johannsdottir, L., & Wendt, S. (2022). Barriers to Using ESG Data for Investment Decisions. *Sustainability*, *14*(9). https://doi.org/10.3390/su14095157

Kimbrough, M. D., Wang, X., Wei, S., & Zhang, J. (2024). Does Voluntary ESG Reporting Resolve Disagreement among ESG Rating Agencies? *European Accounting Review*, *33*(1), 15–47. https://doi.org/10.1080/09638180.2022.2088588

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

Lee, M. T., & Raschke, R. L. (2023). Stakeholder legitimacy in firm greening and financial performance: What about greenwashing temptations?. *Journal of Business Research*, *155*. https://doi.org/10.1016/j.jbusres.2022.113393

Lee, M. T., Raschke, R. L., & Krishen, A. S. (2023). Understanding ESG scores and firm performance: Are high-performing firms E, S, and G-balanced? *Technological Forecasting and Social Change*, *195*. https://doi.org/10.1016/j.techfore.2023.122779

Lins, K. V., Servaes, H., & Tamayo, A. (2017). Social Capital, Trust, and Firm Performance: The Value of Corporate Social Responsibility during the Financial Crisis. *Journal of Finance*, *72*(4), 1785–1824. https://doi.org/10.1111/jofi.12505

Liu, M. (2022). Quantitative ESG disclosure and divergence of ESG ratings. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.936798

McGuire, J. B., Sundgren, A., & Schneeweis, T. (1988). Corporate Social Responsibility and Firm Financial Performance. *The Academy of Management Journal*, *31*(4), 854–872. https://doi.org/http://dx.doi.org/10.2307/256342

Meyer, J.W., Rowan, B. (1977). Institutionalized Organizations: Formal Structure as Myth and Ceremony. *American Journal of Sociology*, 83(2), 340-363. http://dx.doi.org/10.1086/226550

Michelon, G. (2011). Sustainability disclosure and reputation: a comparative study. *Corporate Reputation Review*, *14*(2), 79–96. https://doi.org/https://doi.org/10.1057/crr.2011.10

Murphy, K. R., Jako, R. A., & Anhalt, R. L. (1993). Nature and Consequences of Halo Error: A Critical Analysis. *Journal of Applied Psychology*, *78*(2), 218–225. https://doi.org/https://doi.org/10.1037/0021-9010.78.2.218

Naser, K., Al-Hussaini, A., Al-Kwari, D., & Nuseibeh, R. (2006). Determinants of Corporate Social Disclosure in Developing Countries: The Case of Qatar. In *Advances in International Accounting* 19, 1–23. https://doi.org/10.1016/S0897-3660(06)19001-7

Nohria, N., & Gulati, R. (1996). IS SLACK GOOD OR BAD FOR INNOVATION? *Academy of Management Journal*, *39*(5), 1245–1264. https://doi.org/https://doi.org/10.2307/256998

O'Donnell, E., & Schultz, J. J. (2005). The Halo Effect in Business Risk Audits: Can Strategic Risk Assessment Bias Auditor Judgment about Accounting Details? *THE ACCOUNTING REVIEW*, *80*(3), 921–939. https://doi.org/https://doi.org/10.2308/accr.2005.80.3.921

Oprean-Stan, C., Oncioiu, I., Iuga, I. C., & Stan, S. (2020). Impact of sustainability reporting and inadequate management of esg factors on corporate performance and

sustainable growth. *Sustainability*, *12*(20), 1–31. https://doi.org/10.3390/su12208536

Orlitzky, M. (2001). Does Firm Size Confound the Relationship Between Corporate Social Performance and Firm Financial Performance. *Journal of Business Ethics*, *33*(2), 167–180. https://doi.org/https://doi.org/10.1023/

Orlitzky, M., Schmidt, F. L., & Rynes, S. L. (2003). Corporate Social and Financial Performance: A Meta-analysis. *Organization Studies*, *24*(3), 403–441. https://doi.org/https://doi.org/10.1177/0170840603024003910

Perez-Batres, L. A., Doh, J. P., Miller, V. V., & Pisani, M. J. (2012). Stakeholder Pressures as Determinants of CSR Strategic Choice: Why do Firms Choose Symbolic Versus Substantive Self-Regulatory Codes of Conduct? *Journal of Business Ethics*, *110*(2), 157–172. https://doi.org/10.1007/s10551-012-1419-y

Porter, M. E., & Van Der Linde, C. (1995). Toward a New Conception of the Environment-Competitiveness Relationship. *Journal of Economic Perspectives*, *9*(4), 97–118. https://doi.org/http://dx.doi.org/10.1257/jep.9.4.97

Prado-Lorenzo, J. M., & Garcia-Sanchez, I. M. (2010). The Role of the Board of Directors in Disseminating Relevant Information on Greenhouse Gases. *Journal of Business Ethics*, *97*(3), 391–424. https://doi.org/10.1007/s10551-010-0515-0

Rahi, F., Johansson, J., Blomkvist, M., & Hartwig, F. (2024). Corporate sustainability and financial performance: A hybrid literature review. *Corporate Social Responsibility and Environmental Management*, *31*(2), 801–815. https://doi.org/10.1002/csr.2600

Refinitiv (2022). *ENVIRONMENTAL, SOCIAL AND GOVERNANCE SCORES FROM REFINITIV*. Refinitiv. https://www.lseg.com/content/dam/marketing/en_us/documents/methodology/refinitiv-esg-scores-methodology.pdf

Rischkowsky, F., & Döring, T. (2008). Consumer policy in a market economy considerations from the perspective of the economics of information, the new institutional economics as well as behavioural economics. *Journal of Consumer Policy*, *31*(3), 285–313. https://doi.org/10.1007/s10603-008-9076-3

Russo, M. V, & Fouts, P. A. (1997). A Resource-Based Perspective on Corporate Environmental Performance and Profitability. *The Academy of Management Journal*, *40*(3), 534–559. https://doi.org/http://dx.doi.org/10.2307/257052

Schaltegger, S., & Hörisch, J. (2017). In Search of the Dominant Rationale in Sustainability Management: Legitimacy- or Profit-Seeking? *Journal of Business Ethics*, *145*(2), 259–276. https://doi.org/10.1007/s10551-015-2854-3

Serafeim, G., & Yoon, A. (2023). Stock price reactions to ESG news: the role of ESG ratings and disagreement. *Review of Accounting Studies*, *28*(3), 1500–1530. https://doi.org/10.1007/s11142-022-09675-3

Servaes, H., & Tamayo, A. (2013). The impact of corporate social responsibility on firm value: The role of customer awareness. *Management Science*, *59*(5), 1045–1061. https://doi.org/10.1287/mnsc.1120.1630

Sharma, U., Gupta, A., & Gupta, S. K. (2024). The pertinence of incorporating ESG ratings to make investment decisions: a quantitative analysis using machine learning. *Journal of Sustainable Finance and Investment*, *14*(1), 184–198. https://doi.org/10.1080/20430795.2021.2013151

Soppe, A. (2004). Sustainable Corporate Finance. *Journal of Business Ethics*, *53*, 213–224. https://doi.org/https://doi.org/10.1023/B:BUSI.0000039410.18373.12

Spence, C. (2009). Social and environmental reporting and the corporate ego. *Business Strategy and the Environment*, *18*(4), 254–265. https://doi.org/10.1002/bse.600

Suchman, M. C. (1995). Managing legitimacy: strategic and institutional approaches. *Academy of Management Review*, *20*, 571–610. https://doi.org/https://doi.org/10.2307/258788

Sultana, S., Zulkifli, N., & Zainal, D. (2018). Environmental, social and governance (ESG) and investment decision in Bangladesh. *Sustainability*, *10*(6), 1831. https://doi.org/10.3390/su10061831

Tamimi, N., & Sebastianelli, R. (2017). Transparency among S&P 500 companies: an analysis of ESG disclosure scores. *Management Decision*, *55*(8), 1660–1680. https://doi.org/10.1108/MD-01-2017-0018

Tang, D. Y., Yan, J., & Yao, C. Y. (2022). The determinants of ESG ratings: Rater ownership matters. *Paris December 2021 Finance Meeting EUROFIDAI-ESSEC*. https://doi.org/http://dx.doi.org/10.2139/ssrn.3889395

Tarquinio, L., & Posadas, S. C. (2020). Exploring the term "non-financial information": an academics' view. *Meditari Accountancy Research*, *28*(5), 727–749. https://doi.org/10.1108/MEDAR-11-2019-0602

Thorndike, E. L. (1920). A CONSTANT ERROR IN PSYCHOLOGICAL RATINGS. *Journal of Applied Psychology*, *4*(1), 25–29. https://doi.org/https://doi.org/10.1037/h0071663

Tsang, A., Frost, T., & Cao, H. (2023). Environmental, Social, and Governance (ESG) disclosure: A literature review. *British Accounting Review*, *55*(1). https://doi.org/10.1016/j.bar.2022.101149

Veronica Siregar, S., & Bachtiar, Y. (2010). Corporate social reporting: empirical evidence from Indonesia Stock Exchange. *International Journal of Islamic and Middle Eastern Finance and Management*, *3*(3), 241–252. https://doi.org/10.1108/17538391011072435

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

Walley, N., & Whitehead, B. W. (1994). It's not easy being green. *Harvard Business Review*, *72*, 46–52.

Windolph, S. E. (2011). Assessing Corporate Sustainability Through Ratings: Challenges and Their Causes. *Journal of Environmental Sustainability*, *1*(1), 1–22. https://doi.org/10.14448/jes.01.0005

Wu, Y., Zhang, K., & Xie, J. (2020). Bad greenwashing, good greenwashing: Corporate social responsibility and information transparency. *Management Science*, *66*(7), 3095–3112. https://doi.org/10.1287/mnsc.2019.3340

Xiao, C., Wang, Q., van Donk, D. P., & van der Vaart, T. (2018). When are stakeholder pressures effective? An extension of slack resources theory. *International Journal of Production Economics*, *199*, 138–149. https://doi.org/10.1016/j.ijpe.2018.03.002

APPENDICES

Appendix A – Variable definitions.

Variable	Variable definition
$ESG\_Dis_{i,t}$	Absolute difference between Bloomberg and re-scaled Refinitiv ESGC rating for firm i in year t
$E\_Dis_{i,t}$	Absolute difference between Bloomberg and re-scaled Refinitiv environmental pillar rating for firm i in year t
$S\_Dis_{i,t}$	Absolute difference between Bloomberg and re-scaled Refinitiv social pillar rating for firm i in year t
$G\_Dis_{i,t}$	Absolute difference between Bloomberg and re-scaled Refinitiv governance pillar rating for firm i in year t
$SIZE1_{i,t}$	Natural logarithm of total assets of firm i in year t
$SIZE2_{i,t}$	Natural logarithm of the market capitalization of firm i in year t
$PROF_{i,t}$	Return on assets of firm i in year t calculated by dividing net income by total assets
$ESG\_PERF_{i,t}$	Average ESG-Rating of Bloomberg and Refinitiv ESGC rating of firm i in year t
$ESG\_PERF_{i,(t-1)}$	Average ESG-Rating of Bloomberg and Refinitiv ESGC rating of firm i in year $t-1$
$BTM_{i,t}$	Book-to-market ratio of firm i in year t calculated by dividing the common shareholders equity by the market capitalization
$CAPEX_{i,t}$	Natural logarithm of the CapEx of firm i in year t as in the funds used during year t to acquire or improve long term assets such as property, plant and equipment
$LEV_{i,t}$	Leverage of firm i in year t in % calculated by dividing total debt by total shareholders' equity *100 (debt-to-equity ratio)

Appendix B – Definitions of variables applied in the robustness checks.

Variable	Variable Definition
$ESG\_Dis_{i,t}\ RO$	Absolute difference between Bloomberg and re-scaled Refinitiv ESG rating for firm i in year t
$ESG\_PERF_{i,t}\ RO$	Average ESG Rating of Bloomberg and Refinitiv rating of firm i in year t
$ESG_{PERF_{i,(t-1)}}RO$	Average ESG-Rating of Bloomberg and Refinitiv rating of firm i in year $t-1$
$CONTROV_{i,t}$	Dummy variable taking the value 1 in case there are controversies accounted for by Refinitiv and 0 in all other cases. When Refinitiv's ESGC and ESG-score are equal no controversies have been identified by Refinitiv for company *i* in year *t*.
$SIZE3_{i,t}$	Natural logarithm of the total revenue of firm i in year t
$PROF\_RO_{i,t}$	Operating Margin in % measuring operating income divided by total revenue

Appendix C –: Sample by industry and country.

Companies by industry

Industry	No. of companies
Basic Materials	25
Consumer Discretionary	38
Consumer Staples	17
Energy	7
Financials	53
Health Care	30
Industrials	58
Real Estate	13
Technology	13
Telecommunications	11
Utilities	13
Total	278

Companies by country of headquarter

Country	No. of companies
Belgium	18
Denmark	19
Finland	22
France	32
Germany	33
Ireland; Republic of	8
Luxembourg	2
Mexico	1
Netherlands	24
Spain	27
Sweden	23
Switzerland	17
United Kingdom	52
Total	278

Appendix D – Regression results Model 2 using Refinitiv's ESG rating.

Dependent Var.	**ESG_DIFF RO**		E_DIFF		S_DIFF		G_DIFF	
Model	(2.1)	(2.2)	(2.1)	(2.2)	(2.1)	(2.2)	(2.1)	(2.2)
SIZE1	0.2269***	-	0.0875	-	0.3599***	-	0.0000	-
SIZE2	-	0.0682	-	-0.1860	-	0.2008*	-	-0.0211
PROF	-0.0243***	-0.0320***	-0.0235	-0.0201	-0.0268**	-0.0417***	0.0018	0.0029
ESG_PERF (t-1) RO	0.2242***	0.2321***	0.1765**	0.1549**	0.2117***	0.2330***	-0.0632**	-0.0748**
CapEx	-	0.1005**	-	0.2044**	-	0.0923	-	0.0395
LEV	-0.0003	0.0000	0.0017	0.0012	-0.0006	0.0001	0.0000	-0.0002
BTM	0.0007	0.0021	0.0027	0.0018	-0.0019	0.0009	-0.0010	-0.0013
Constant	-3.1144**	-1.5502	0.4700	2.8120	-4.8334**	-3.2015	2.3836**	2.1430*
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Industry FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Obsertaons	1,403	1,403	1,403	1,403	1,403	1,403	1,403	1,403
R^2	0.3465	0.3384	0.2401	0.2521	0.298	0.2863	0.0816	0.0837

This table presents the regression results of Model 2.1 and Model 2.2 on an overall ESG score level as well as for each pillar dimension with the variation that ESG_DIFF and ESG_PERF (t-1) are calculated using Refinitiv's ESG score instead of the ESGC score. The dependent variable ESG_DIFF_RO is the absolute difference between both ratings. Definitions of the variables are provided in Appendix A and B. Country, industry, and year fixed effects are not shown separately. Bold text denotes the regression variables and their estimated coefficients changed with respect to the originalal model. The standard deviations are clustered at the firm year level. Statistical significance at the 1% ,5% and 10% levels are shown by ***, ** and *, respectively.

Appendix E – Regression results Model 1 with $SIZE3_{i,t}$ as size proxy.

Dependent Variable	ESG_DIFF	E_DIFF	S_DIFF	G_DIFF
Model	(1)	(1)	(1)	(1)
SIZE3	**-0.1328*** **	**0.1789***	**0.3735*** **	**0.0339**
PROF	-0.0190***	-0.0189	-0.0368***	-0.0080
ESG_PERF	0.6186***	0.0661	0.2026**	-0.1230**
LEV	-0.0005	0.0015	0.0007	0.0000
BTM	-0.0012	0.0005	-0.0037	-0.0008
Constant	3.1586***	-1.0586	-5.1526***	1.8421**
Country FE	✓	✓	✓	✓
Industry FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Observations	1,405	1,405	1,405	1,405
R^2	0.3164	0.1547	0.3036	0.0885

This table presents the regression results of Model 1 on an overall ESG score level as well as for each pillar dimension. The dependent variable ESG_DIFF is the absolute difference between both ratings. Definitions of the variables are provided in Appendix A and B. Country, industry, and year fixed effects are not shown separately. Bold text denotes the variable and their estimated coefficients altered for the robustness tests. CapEx was excluded from the regression due to multicollinearity issues. The standard deviations are clustered at the firm year level. Statistical significance at the 1% ,5% and 10% levels are shown by ***, ** and *, respectively.

Appendix F – Regression Results Model 1.1 with $PROF2_{i,t}$as profitability proxy.

Dependent Variable	ESG_DIFF	E_DIFF	S_DIFF	G_DIFF
Model	(1.1)	(1.1)	(1.1)	(1.1)
SIZE1	-0.1739***	0.1008	0.3830***	0.0324
PROF_RO	**-0.0059***	**-0.0099**	**-0.0146** **	**-0.0021**
ESG_PERF	0.6147***	0.0815	0.20082**	-0.1240**
LEV	-0.0001	0.0019	0.0008	0.0001
BTM	0.0001	0.0010	-0.0044*	-0.0007
Constant	4.0571***	0.5476	-5.5064***	1.8360*
Country FE	✓	✓	✓	✓
Industry FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Observations	1,405	1,405	1,405	1,405
R^2	0.3179	0.1497	0.3007	0.0874

This table presents the regression results of Model 1.1 on an overall ESG score level as well as for each pillar dimension. The dependent variable ESG_DIFF is the absolute difference between both ratings. Definitions of the variables are provided in Appendix A and B. PROF2 was winsorized at the 3% level to account for outliers. Country, industry, and year fixed effects are not shown separately. Bold text denotes the variable and their estimated coefficients altered for the robustness tests. The standard deviations are clustered at the firm year level. Statistical significance at the 1% ,5% and 10% levels are shown by ***, ** and *, respectively.

Appendix G – Overview Regression Results

	ESG-score		E-score		S-score		G-score	
Analyses								
(Sub-)Model	*(1.1)*	*(1.2)*	*(1.1)*	*(1.2)*	*(1.1)*	*(1.2)*	*(1.1)*	*(1.2)*
H1	☒	☒	n.s.	n.s.	✓	✓	n.s.	n.s.
H2	✓	✓	n.s.	n.s.	✓	✓	n.s.	n.s.
H3.1	☒	☒	n.s.	n.s.	☒	☒	✓	✓
(Sub-)Model	*(2.1)*	*(2.2)*	*(2.1)*	*(2.2)*	*(2.1)*	*(2.2)*	*(2.1)*	*(2.2)*
H1	☒	☒	n.s.	n.s.	✓	✓	n.s.	n.s.
H2	✓	✓	n.s.	n.s.	✓	✓	n.s.	n.s.
H3.2	☒	☒	☒	☒	☒	☒	✓	✓
Robustness Tests								
Refinitiv ESG-score								
(Sub-)Model	*(1.1)*	*(1.2)*	*(1.1)*	*(1.2)*	*(1.1)*	*(1.2)*	*(1.1)*	*(1.2)*
H1	✓	n.s.	n.s.	n.s.	✓	n.s.	n.s.	n.s.
H2	✓	✓	n.s.	n.s.	✓	✓	n.s.	n.s.
H3.1	☒	☒	☒	☒	☒	☒	✓	✓
CONTROV variable								
(Sub-)Model	*(1.1)*	*(1.2)*						
H1	n.s.	n.s.						
H2	✓	✓						
H3.1	☒	☒						
SIZE3								
(Sub-)Model	*(1)*		*(1)*		*(1)*		*(1)*	
H1	☒		✓		✓		n.s.	
H2	✓		n.s.		✓		n.s.	
H3.1	☒		n.s.		☒		✓	
PROF2								
(Sub-)Model	*(1.1)*		*(1.1)*		*(1.1)*		*(1.1)*	
H1	☒		n.s.		✓		n.s.	
H2	✓		n.s.		✓		n.s.	
H3.1	☒		n.s.		☒		✓	

This table presents a summary of the regression results obtained from the original analyses and the robustness checks. The results marked with a checkmark led to the hypothesis not being rejected. The results outlined with a cross highlight coefficients which are statistically significant but indicate a reverse effect as stated in the hypotheses. From the remaining results (marked with "n.s.") no conclusion can be drawn in lack of statistically significant results at the 1%, 5% or 10% level.