BERGISCHE
UNIVERSITÄT
WUPPERTAL

Bergische Universität Wuppertal
Fakultät für Wirtschaftswissenschaft
Schumpeter School of Business and Economics

Dissertationsschrift
zur Erlangung des akademischen Grades

Doktor der Wirtschaftswissenschaft
(Dr. rer. oec.)

Theory and Empirical Evidence on the Accuracy of ESG Metrics: An Empirical Investigation of Refinitiv ESG Performance Scores

von

Simon Funken, M.Sc.

vorgelegt am:

26.10.2022

Dekan:

Univ.-Prof. Dr. Nils Crasselt
Schumpeter School of Business and Economics

Vorsitzender des Promotionsausschusses:

Univ.-Prof. Dr. Stefan Bock
Lehrstuhl für Wirtschaftsinformatik und Operations Research

Gutachter:

Univ.-Prof. Dr. André Betzer
Lehrstuhl für Finanzwirtschaft und Corporate Governance

Univ. Prof. Dr. Nils Crasselt
Lehrstuhl für Controlling

Theory and Empirical Evidence on the Accuracy of ESG Metrics: An Empirical Investigation of Refinitiv ESG Performance Scores

ABSTRACT

ESG data providers and their metrics serve an ever-growing important role in informing capital markets on the non-financial performance of single corporate entities, investment funds, and equity indices. They build upon the idea of quantifying the potential costs and benefits of a firm's engagement in the three behavioral dimensions: Environment (E), Social (S), and Corporate Governance (G).

Analyzing the entire global database of ESG data provider "Refinitiv" (as of April 2022), this study gathers compelling evidence that Refinitiv ESG performance scores systematically fall short of providing an objective, persisting, and accurate indication of a firm's actual ESG performance. The results draw from three fields of concern: 1) The nature of publicly available ESG information, 2) Refinitiv's proprietary scoring methodology, and 3) the ability of Refinitiv's ESG metrics to reflect information that indicates the effectiveness of a firm's actual, yet unobservable level of commitment to ESG issues.

The empirical investigation builds upon 83,827 firm-year observations of 13 ESG performance metrics for up to 11,792 firms from 2002 to 2021. Additionally, Refinitiv's controversy database is analyzed, which reflects 31,963 separate appearances in news articles due to controversial business conduct. The study finds no persisting link between Refinitiv's ESG metrics and proxies for the materialization of ESG risks induced by the effectiveness of actual ESG commitment (i.e., market-based idiosyncratic volatility and the occurrence of ESG controversies).

As ESG metrics are primarily based on corporate disclosure, the available information on which to base an ESG scoring methodology is naturally affected by mandatory and customary reporting practices, as well as firm visibility. These tend to vary especially based on factors that heavily correlate with firm size. Refinitiv's proprietary ESG scoring methodology adds to the resulting size bias by penalizing non-reported data and intentionally assigning the worst possible performance. As a result, the study finds at least 75.1% of all yearly Refinitiv ESG performance scores to be affected by severe size and transparency biases that put larger firms at a systematic advantage over their competitors.

Scoring ESG information from corporate disclosure is known to reflect inputs (e.g., policies, programs, and processes) rather than actual outcomes. The study proposes a simple and data-driven approach to bridge input-based ESG measurement and proxies that reflect variation induced by the actual outcome of such actions. Based on 1,300 rolling window panel regression models, which account for the industry-, region-, and time-specific heterogeneity, ESG category scores are linked to the dependent variable idiosyncratic volatility (IV). The results indicate whether the score captures value relevant (i.e., material) information on a firm's specific risk induced by the effectiveness of its ESG commitment. Ex-post reweighting of ESG category scores according to their specific materiality (i.e., historical accuracy) enhances the ability of Refinitiv ESG scores to reflect future materializations of ESG risks. By additionally accounting for the methodology-induced fraction of size and transparency biases, the ability of Refinitiv's ESG metrics to indicate the occurrence of ESG controversies is consistently enhanced.

In conclusion, the results of this dissertation imply an analytical use of ESG performance metrics reflective of the individual purpose (i.e., risk mitigation, value alignment, or the intention of impact). The ambiguous nature of ESG performance and the interest of metrics to boil down such non-financial corporate performance to a simple and easy-to-interpret indication of a firm's commitment to a multitude of issues allows for a similar extent of design choices. Constructing metrics inherently requires to compromise on a qualitative and interpretative level. Simultaneously, the very nature of ESG information and its acquisition imposes structural burdens that affect the quantitative database on which ESG metrics are based imperatively.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

_adj	adjusted (limited to fully covered firms)
AUM	assets under management
BLS	U.S. Bureau of Labor Statistics
CAPM	Capital Asset Pricing Model
CDSB	Climate Disclosure Standards Board
CEO	chief executive officer
CFP	corporate financial performance
CMI	conditional mean independence
COM	Refinitiv ESG "Community" category score
Cycl.	cyclicals
CSP	corporate social performance
CSR	Corporate Social Responsibility; Refinitiv ESG "CSR strategy" score
DAX	Deutscher Aktienindex
DJSI	Dow Jones Sustainability Indices
E	Environmental; Refinitiv “Environmental” pillar score
EBA	European Banking Authority
EC	European Commission
EMM	Refinitiv ESG "Emission" category score
ESG	Environmental, Social, Corporate Governance; Refinitiv ESG score
EU	Europe; European Union
FASB	Financial Accounting Standards Board
FDA	U.S. Food and Drug Administration
FE	fixed effects
FTSE4Good	FTSE ESG investment indices
G	Corporate Governance; Refinitiv "Governance" pillar score
GHG	greenhouse gas
GRI	Global Reporting Initiative
GSIA	Global Sustainable Investment Alliance
HR	Refinitiv ESG "Human rights" category score
IA	intangible assets
IIRC	International Integrated Reporting Council
INO	Refinitiv ESG "Innovation" category score

IPCC	Intergovernmental Panel on Climate Change
IR	Integrated Reporting
IV	idiosyncratic volatility
KLD	Kyle, Lydenburg, and Domini Research & Analytics
mat	materiality-adjusted
MDG	Millennium Development Goals
MNG	Refinitiv ESG "Management" category score
NASDAQ	National Association of Securities Dealers Automated Quotation
NGO	non-governmental organization
obs.	observations
OLS	ordinary least squares
PRD	Refinitiv ESG "Product responsibility" category score
R&D	research and development
RES	Refinitiv ESG "Resource use" category score
ROA	return on assets
S	Social; Refinitiv "Social" pillar score
S&P	Standard & Poor’s
SASB	Sustainable Accounting Standards Board
SDG	Sustainable Development Goals
SDR	Sustainability Disclosure Requirements
SFDR	EU sustainable finance disclosure regulation
SH	Refinitiv ESG "Shareholders" category score
TRBC	The Refinitiv Business Classification
UN PRI	UN Principles of Responsible Investment
UNEP	UN Environmental Program
UNGC	UN Global Compact
UNGP	UN Guiding Principles on Business and Human Rights
UN-M49	UN Standard Country and Area Codes for Statistical Use
USD / $	US Dollar
VIF	variance inflation factor
VRF	Value Reporting Foundation
WEF	World Economic Forum
WF	Refinitiv ESG "Workforce" category score
z / zadj	standardized / standardized (by firm size deciles)

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS AND FORMULAS

1 INTRODUCTION

1.1 Motivation

Sustaining the efficient allocation of resources in modern capital markets requires information communicated via financial reporting and corporate disclosure. Markets implement institutions that process such information and provide entity-specific performance evaluations, like analyst forecasts, investment recommendations, and credit ratings.[1] A relatively new set of performance measures are *ESG scores*. They build upon the idea of quantifying the potential costs and benefits of a firm's engagement in the three behavioral dimensions: Environment (E), Social (S), and Corporate Governance (G).[2]

The demand for ESG performance assessment is driven by economic, societal, and political expectations that push forward a normative framework against which corporate behavior is judged. Around the idea of putting such non-financial corporate performance into numbers, numerous recipients have evolved. ESG information finds widespread usage among regulators, media, NGOs, business professionals, and foremost in the investment community.[3] Socially responsible investing, sustainable investing, ethical and philanthropic investing, ESG and green investing, or impact investing are examples of strategies that, to some extent, rely upon the same principle: Incorporating information on at least one of the three pillars of ESG into investment decision-making.[4]

Current estimations of assets being managed alongside ESG information range up to US$ 35.4 trillion in 2020, making up roughly a quarter to a third of all global assets under management (AUM).[5] The UN Principles of Responsible Investing (UN PRI), which seek to encourage investors to incorporate ESG information in their investment decisions, report a consistent growth of signatories since their introduction in 2006. By 2021, 3,826 Investment Managers, Asset Owners,

[1] cf. Healy/Palepu (2001, p. 407)
[2] Throughout the dissertation, the terms "ESG scores", "ESG metrics", and "ESG ratings" will interchangeably refer to this idea.
[3] cf. Windolph (2011, p. 37-37)
[4] cf. Gillan/Koch/Starks (2021, pp. 1-2), cf. Matos (2020, p. 1)
[5] cf. GSIA (2021, p. 10). The quantification of global AUM relies upon assumptions and requires the aggregation of data from numerous data sources. Different estimation methodologies impose additional inaccuracies. The GSIA sets global AUM at the beginning of 2020 to be around US$ 98.4 trillion (GSIA 2020, p. 9). Other estimations range from US$ 92.3 trillion (source: BCG, Global Asset Management report, July 2021, https://www.bcg.com/de-de/publications/2021/global-asset-management-industry-report, accessed: 02/04/2022) to around US$ 115 trillion (source: *Crossing the Horizon: North American Asset Management in the 2020s report* by McKinsey & Co., https://www.mckinsey.com/industries/financial-services/our-insights/crossing-the-horizon-north-american-asset-management-in-the-2020s, accessed: 02/04/2022).

and Financial Service Providers, accounting for US$ 29.2 trillion AUM (see **Figure 1**), engaged in following the UN PRI (**Appendix 1**).[6]

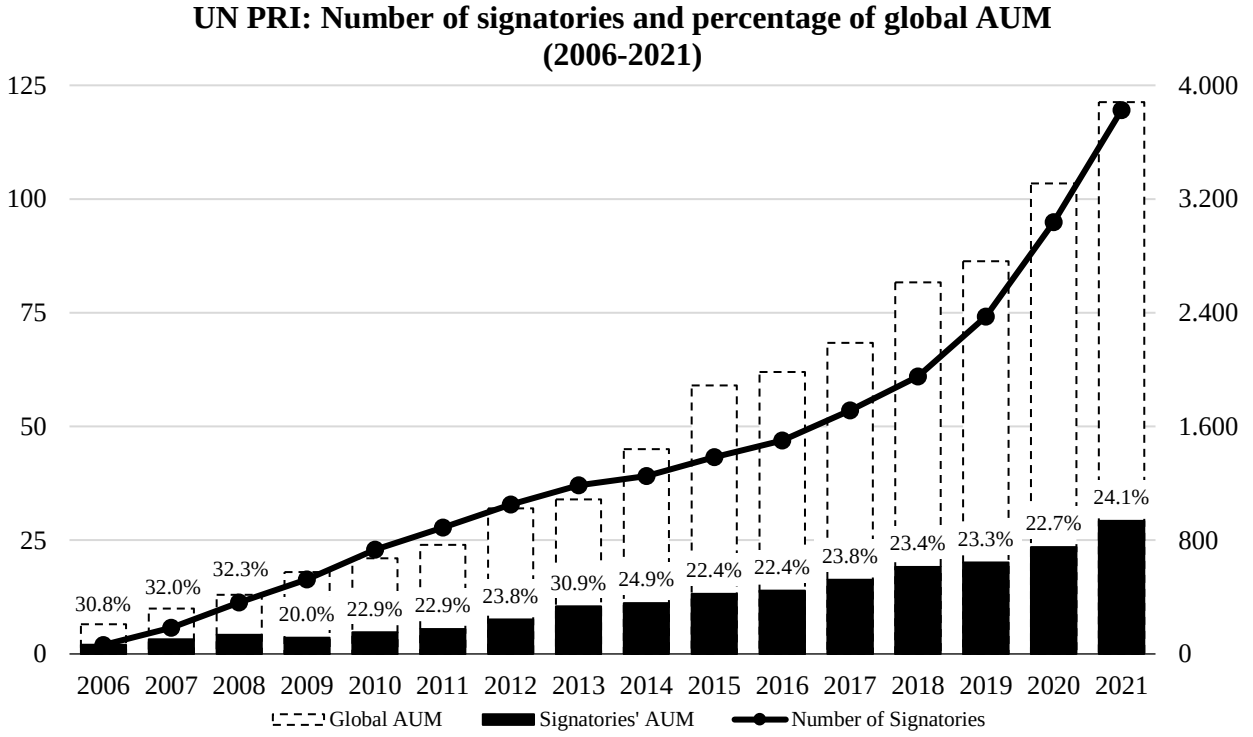


Figure 1 UN PRI: Number of signatories and percentage of global AUM (2006-2021), source: https://www.unpri.org/about-us/about-the-pri, accessed: 08/09/2022.

Being such an integral part of the financial industry, the concept of ESG measurement finds itself in a lively debate over its motivations, environmental and societal benefits, economic merits, and methodological accuracy. Decades of political discourse and scholarly debate have argued on behalf of or against corporate engagement on issues related to ESG performance and its disclosure. Corporate Social Responsibility (CSR) and Philanthropy, Stakeholder Management, Sustainability, Intergenerational Equity, or Environmental and Climate Change exemplify the broad range of public debates that are heavily influenced by the implications of ESG performance measurement.

Despite an ongoing process of linking the dimensions of ESG to overarching objectives, the impact of ESG information and metrics to modern capital markets and the corporate world is undisputed. Serafeim and Yoon (2022) find that stock prices systematically react to news related to ESG issues.[7] Hartzmark and Sussman (2019) show how the introduction of market-wide sustainability metrics by Morningstar in 2016 led to massive inflows (outflows) in high-rated (low-rated) US mutual funds.[8] As higher-rated CSR funds are found to systematically underperform their lower-rated counterparts, Riedl/Smeets (2017) and El Ghoul/Karoui (2017) argue that utility functions

[6] https://www.unpri.org/about-us/about-the-pri, accessed: 09/08/2022.
[7] cf. Serafeim/Yoon (2022, p. 59)
[8] cf. Hartzmark/Sussman (2019, pp. 2790-2791)

of investors are likely driven by a set of non-financial performance characteristics (e.g., based on social preferences) that represent a respective investment rationale.[9] Further, Amel-Zadeh and Serafeim (2018) provide evidence based on an investor survey that ESG information is considered to convey insights into the potential to mitigate investment risks.[10]

Surveys among investment professionals, however, show that the comparability of ESG information among different sources is perceived as relatively low.[11] In fact, a thriving industry gathers ESG information on companies, assigns proprietary ratings, and ranks them among their peers. Current estimations assume that about 600 different ESG metrics exist. A recent study suggests that among those, about a dozen provided by third-party rating agencies are seen as most favorable by practitioners and academia regarding their data quality and credibility.[12] Each of those ratings utilizes different ranking and standardization methodologies.[13] The variety of proprietary scoring methodologies, however, leads to a severe lack of comparability among the ESG scores of different providers.[14]

Berg, Kölbel, and Rigobon (2019), for example, analyze ESG scores from six leading third-party ESG data providers widely used in research and investment practice.[15] They state ESG metrics tend to differ by three characteristics: 1) The scope to which particular aspects of ESG performance are considered, 2) the measurement of ESG aspects that potentially considers different data points for the same issue, and 3) the weighting of performance on specific ESG issues when aggregating a rating.[16] Given that divergence among those dimensions might very well be driven by proprietary but in itself conclusive approaches toward evaluating ESG performance, the question on the usefulness of ESG scores arises.[17] Recent regulatory advances and expected changes, like the 2022 SEC ESG disclosure proposal or the 2022 EU Regulation on sustainability-related disclosure in the financial services sector (SFDR), further enhance the pressure on the quality and reliability of ESG performance metrics.[18]

[9] cf. Riedl/Smeets (2017, p. 2533), cf. El Ghoul/Karoui (2017, pp. 53-54)
[10] cf. Amel-Zadeh/Serafeim (2018, p. 87)
[11] cf. ibid.
[12] cf. Wong/Brackley/Petroy (2019, pp. 4, 15-17)
[13] cf. Saadaoui/Soobaroyen (2018, p. 26)
[14] cf. Escrig-Olmedo et al. (2019, pp. 15-16), cf. Berg/Kölbel/Rigobon (2019, p. 3), cf. Serafeim (2021, pp. 18-19)
[15] The following ESG data providers are analyzed: MSCI, KLD, Sustainalytics, Moody's (former Vigeo-Eiris), S&P Global (former RobecoSAM), and Refinitiv (former Asset4).
[16] cf. Berg/Kölbel/Rigobon (2019, pp. 29-30), Refinitiv ESG ratings are based on the ASSET4 database.
[17] cf. Serafeim (2021, pp. 18-19)
[18] https://www.sec.gov/news/press-release/2022-92, accessed: 05/27/2022, https://finance.ec.europa.eu/sustainable-finance/disclosures/sustainability-related-disclosure-financial-services-sector_en, accessed: 05/01/2022.

1.2 Aim and contribution

In order to shed light on the accuracy of ESG performance metrics, as such an important yet fairly distorted source of information within the financial industry, this study analyses whether ESG metrics actually measure what they are supposed to measure. In doing so, this study turns toward the often-proclaimed capability of ESG metrics to identify the exposure to ESG risks and deals with the question whether ESG metrics can convey such information in a predictive manner.

Following the notion of the European Banking Authority's (EBA) definition of ESG risks in its recent report *On Management and Supervision of ESG Risks for Credit Institutions and Investment Firms*, this study defines ESG risks as the negative materialization of the actual, yet unobservable effectiveness of commitment to ESG issues (i.e., ESG performance).[19] Like numerous previous articles, this study implements future idiosyncratic volatility (IV) as an ex-post measure of a firm's exposure to events and circumstances posing firm-specific effects on market valuation.[20] By investigating the capability of ESG metrics to reflect the occurrence of ESG-related controversies in the future, this study gathers additional evidence on the accuracy of ESG performance metrics in reflecting the probability of adverse reputational effects induced by ESG-related misconduct.

Deviating from Berg, Kölbel, and Rigobon (2019), who prominently seek to "Aggregate Confusion" among widely used ESG metrics by finding their commonalities,[21] this study focuses on one ESG metric in particular: The *Refinitiv ESG company scores*.[22] Founded in 2003 as ASSET4 and later acquired by Thomson Reuters (2009) and Refinitiv (2018), Refinitiv ESG scores are widely used in research applications.[23] Until 2021, roughly over 1.800 academic articles mentioned the underlying ASSET4 database or one of its successors (**Figure 2**).

[19] cf. EBA (2021, pp. 32-33)
[20] Among others: cf. Luo/Bhattacharya (2009, p. 200), cf. Sassen et al. (2016, pp. 874-875), cf. Kaiser (2020, p. 39), cf. Reber/Gold/Gold (2021, p. 4), cf. Boucher et al. (2022, pp. 4-5), Idiosyncratic volatility is the annualized standard deviation of the entity-specific residual risk derived from a Carhart four-factor model.
[21] Berg/Kölbel/Rigobon (2019, p. 1)
[22] https://www.refinitiv.com/en/sustainable-finance/esg-scores#company-esg-scores, accessed: 09/13/2022.
[23] cf. Escrig-Olmedo et al (2019, p. 5), cf. Berg/Fabisik/Sautner (2021, pp. 33, 47)

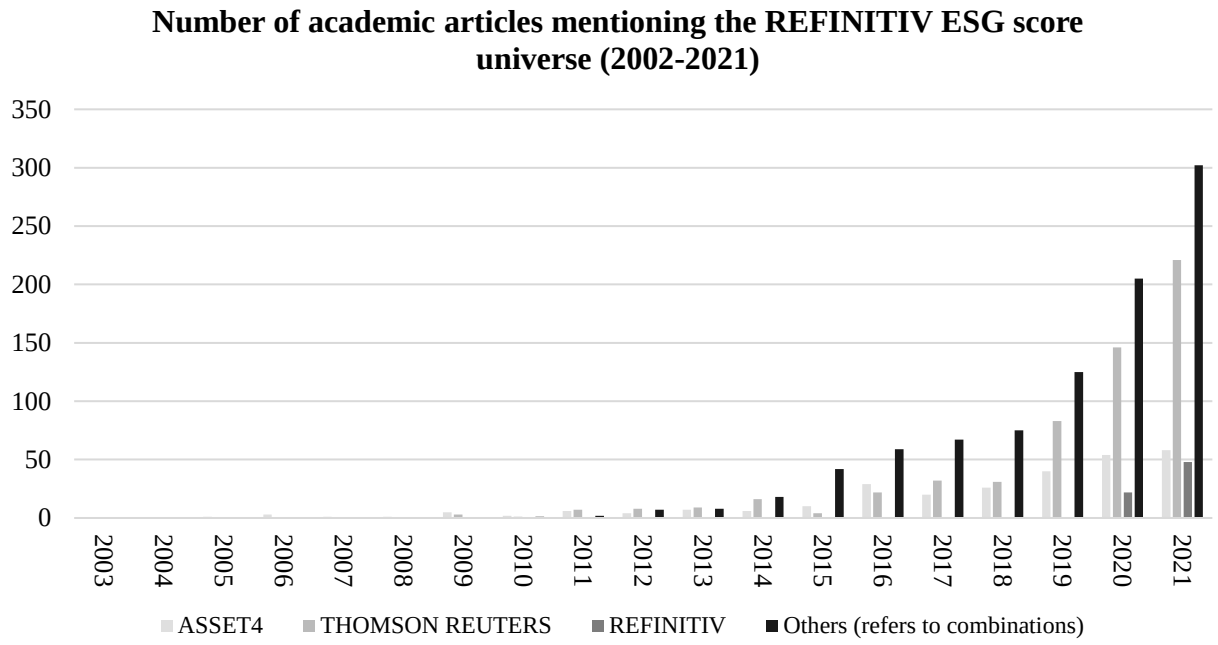


Figure 2 Cumulative number of publications mentioning the Refinitiv ESG score universe (2002-2021), source: https://app.dimensions.ai/discover/publication, accessed: 15.08.2022.

According to Refinitiv's most recent methodology paper, they are "…designed to transparently and objectively measure a company's relative ESG performance" within "… 10 main themes including emissions, environmental product innovation, human rights, shareholders and so on."[24] This study, however, investigates whether and to what extent Refinitiv's ESG metrics live up to that promise as well as the self-imposed claim to account "…for the most material industry metrics, with minimal company size and transparency biases."[25]

In doing so, this dissertation contributes to a recent body of literature focusing on methodological peculiarities of ESG performance measurement. Among others, Drempetic, Klein, and Zwergel (2020) claim that the capabilities of larger firms to prominently communicate their engagement on ESG issues is disproportionately reflected in higher ESG scores (size biases).[26] Sahin et al. (2022a) point toward the practice of ESG data providers to assign the worst possible performance evaluation to firms not reporting on specific data points within the scoring methodology. This treatment potentially induces even stronger transparency biases heavily tied to the already-mentioned size bias.[27] Serafeim, Zochowski, and Downing (2019) state that ESG metrics based on corporate disclosure fall short of observing the outcome of corporate actions, as they primarily reflect policies, programs, or resources in place to reach a particular objective (input orientation).[28] Rogers and

[24] Refinitiv (2022a, p. 3)
[25] ibid.
[26] cf. Drempetic/Klein/Zwergel (2020, pp. 354-355)
[27] cf. Sahin et al. (2022a, p. 3)
[28] cf. Serafeim/Zochowski/Downing (2019, p. 16)

Serafeim (2019) point toward shifts in materiality, i.e., how the relevancy of ESG issues to the business models of firms within certain industries might differ over time (dynamic materiality).[29]

This study relies on two distinct bodies of scholarly literature. One of which builds upon an excessive strand of asset pricing research, as it investigates how firm-specific characteristics induce stock price volatility not explained by broader movements in the market (i.e., idiosyncratic volatility). In a secondary approach, this study tries to directly assess relationships between ex-ante ESG performance metrics and the occurrence of ESG controversies as an embodiment of potential misconduct related to ESG issues. The two approaches address the aforementioned challenges in ESG performance measurement as they specifically focus on the link between input-based ESG performance metrics and outcome-based proxies of actual ESG performance. Further, the study proposes a data-driven correction methodology that addresses quality concerns by resting upon historical accuracy and methodological peculiarities that systematically induce inaccuracies. Thereby, this study combines recent findings in the literature on methodological peculiarities of ESG performance measurement and investigates descriptively and inferentially how common, and methodology-specific sources of inaccuracies strike down to measurable divergence in the rating's capability to accurately reflect future ESG risks. Adding to the literature mentioned above, this study introduces a novel and data-driven approach to assess the historical accuracy of ESG metrics and an easy-to-replicate procedure to enhance the predictive quality of Refinitiv's ESG metrics ex-ante.

As ESG scores are based on relative rankings among industry peers,[30] stripping the sample to regional sub-samples potentially induces unintended tilts in the rating distribution that are not based on the actual methodology. Hence, this dissertation is based on the entire global Refinitiv ESG rating universe and makes use of all scores being available as of the 12th and 13th of April, 2022. Further, Refinitiv adjusts ESG ratings on a weekly basis when new information requires the underlying data to be updated.[31] Berg, Fabisik, and Sautner (2020) question whether Refinitiv ESG scores are suitable for ex-post empirical investigations, as only adjusted metrics are available and not the ones available at a past point in time.[32] This study, however, focuses on the capability of Refinitiv's ESG rating methodology to reflect future outcomes. As the logic of the methodology does not change, adjustments in the ESG scores derived from it do not impose an unjust restatement of ESG performance. It rather reflects how well the deployed methodology is able to retrieve

[29] cf. Rogers/Serafeim (2019, p. 7)
[30] cf. Refinitiv (2022a, p. 6)
[31] cf. ibid (p. 5)
[32] cf. Berg/Fabisik/Sautner (2020, p. 25)

relevant informational content from the same data points. Hence, this study empirically focuses on data that an investor would have at hand when gaining access to Refinitiv's ESG database at any given point in time.

1.3 Outline

The remainder of this dissertation is organized as follows: Chapter 2 deals with the fundamentals of ESG performance measurement. It contextualizes managerial action on ESG as a classic Corporate Governance problem potentially conflicting with the principle of shareholder primacy. It describes the socioeconomic debate on the social responsibility of the firm and the normative and behavioral expectations that society disposes to the corporate world. In doing so, the chapter summarizes the concept of Corporate Social Responsibility (CSR) and its implications for management and reporting. Further, the chapter shows how ESG metrics serve as instruments to supplement investment decision-making. Chapter 3 focuses on particular challenges that go along with ESG performance measurement. The chapter defines the three pillars, environmental, social, and governance, describes the implications of materiality considerations for ESG performance measurement, and shows how ESG metrics might be theoretically and methodologically limited in their ability to measure the actual impact of ESG performance. Chapter 4 gives an extensive literature review on the accuracy of ESG performance metrics. It follows Larcker and Tayan (2021) in defining the premises for reliable and valid ESG metrics and structures the review based on the credibility and independence of ESG data providers, the explanatory power of ESG metrics, and disagreement among different sources of ESG information.[33] Finally, chapter 5 turns toward Refinitiv ESG performance scores. It formulates research hypotheses based on the previous chapters, describes the research methodology, and analyses the ESG scores retrieved from the global ESG rating database descriptively and inferentially based on global and regionally sub-divided panel data. Finally, chapter 6 derives implications from the results and summarizes the major findings of this dissertation.

[33] cf. Larcker/Tayan (2021, p. 439)

2 FUNDAMENTALS OF ESG PERFORMANCE MEASUREMENT

2.1 Non-financial corporate performance and the social responsibility of the firm

> "There is only one social responsibility of a business – to use its resources and engage in activities designed to increase its profits as long as it stays within the rules of the game, which is to say, engages in open and free competition without deception or fraud."[34] Friedman (1970, p. 32).

Fifty years have gone by, and this highly influential claim by Nobel laureate Milton Friedman still marks a natural starting point when asking whether non-financial aspects should be considered when doing business. Friedman argues that firms should maximize profits rather than engage in activities associated with outcomes other than mere financial output. While firms generate profits, the government and each individual take care of externalities.[35]

The doctrine of shareholder primacy builds upon Adam Smith's (1776) basic economic principle of the "invisible hand" that leads market participants to follow their own interests, which results in the most efficient allocation of resources.[36] During the Industrial Revolution, the structure of businesses changed significantly. Innovative manufacturing businesses required financing on larger scales. As a consequence, enterprises grew into publicly-held corporations with dispersed ownership and professional managers. The owners, or shareholders, are in control of the firm but transfer control to the management board. As the owners of a company, however, they hold the residual claim on the firm's profits, which creates a strong focus on the obligations of managers and strengthens the implication of shareholder primacy.[37]

Further, when there are two parties, the suppliers of money and the ones that decide what is done with it (corporate executives), the separation of financing and control imposes potential conflicts of interest. Jensen and Meckling (1976), in their highly influential paper, "Theory of the firm: managerial behavior, agency costs and ownership structure", provide a framework in which such conflicts of interest bear costs. These costs might ultimately result in the corporate executive's (agent) inability to maximize the value of a company's owners (principals).[38]

[34] Friedman (1970, p. 32)
[35] cf. Hart/Zingales (2017, p. 249)
[36] Smith (1937, p. 117)
[37] cf. O'Connell/Ward (2020, p. 2)
[38] cf. Jensen/Meckling (1976, p. 308)

Shareholder primacy assumes that shareholder value is attributable to measurable outcomes, namely share price and dividends. Therefore, one could argue that it holds true for only a very narrow set of conditions: value-generating and value-damaging corporate activities are entirely separable, governments make sure externalities are fully internalized, and shareholders are completely non-social. Consequently, shareholder value might not be the only maxim guiding corporate acting.[39]

2.1.1 Historic antecedents

From a legal point of view, the idea that firms might serve a social purpose is rooted in the actual corporate form itself. Corporation derives from the Latin word "corpus", in English, referring to a body of people. Ancient Roman law defined groups of individuals that were organized for a mutual purpose as separate entities. Many of such organizations served social purposes as they maintained asylums, hospitals, orphanages, or political clubs.[40] Later adopted in English law, the corporation was deployed globally with the rise of the English Empire. The corporate status was granted through a legislative process. Corporations — even those doing rather non-charitable business — were mainly undertaking businesses that served social development, e.g., building and maintaining infrastructure or running banks and insurance companies.[41] In the early 1800s, the increasing number of requests for corporate status sparked the introduction of modern corporate law as we know of today. In the process, the idea of a social responsibility contributed significantly but not entirely to the modern distinction and self-selected purpose of for-profit and non-profit organizations.[42]

From a philosophical standpoint, the question whether the accumulation of wealth is an end in itself or serves a societal purpose already concerned famous writers of antiquity. The Greek philosopher Aristotle distinguished between an honorable (gr., oeconomia) and a non-honorable way of creating wealth (gr., chrematistics). The honorable, or natural way, is associated with creating as much wealth as is needed to be sufficient for a family, community, or state. The non-natural way, chrematistics, describes the limitless accumulation of wealth, which culminates in the exploitation of others.[43] Thus, the notion of wealth creation was not questioning the legitimacy of pursuing personal gains. It was rather characterized by social awareness within economic action.

[39] cf. Hart/Zingales (2017, p. 270-271)
[40] cf. Chaffee (2017, pp. 351-352)
[41] cf. ibid. (p. 353)
[42] cf. ibid. (p. 354)
[43] cf. Sison (2011, pp. 194-195)

Especially on the European continent, this notion of a virtuous way of creating wealth grew out to become the ideal of the "honorable merchant" (ger., ehrbarer Kaufmann).[44] The concept of honor was historically influenced by religious beliefs. The honorable merchant respects and sustains God's creation in the sense that he is aware of his responsibility for the well-being of society and the prosperity of the community.[45] In a more secularized era today, the idea of the responsibility of the honorable merchant set the ground for the understanding of what should be the social responsibility of firms, referred to as Corporate Social Responsibility. In CSR literature, the pursuit of gains and embracing a social purpose are often seen as contradictory. But following the model of the honorable merchant, these two objectives go hand-in-hand when doing business.[46]

2.1.2 A managerial decision

The term CSR was first discussed in research literature in the 1930s as an exchange of legal views sought to justify whether corporate executives are responsible only to the company's owners or the society as well.[47] The quintessence of this debate is that without the legal obligation, managers most likely do not fulfill any social responsibilities which may primarily benefit a firm's broad range of stakeholders and rather tend to feel responsible for the shareholders' wealth only.[48] The publication of Howard R. Bowen's book *Social Responsibilities of the Businessman* in 1953 is considered the starting point from which the term Social Responsibility of companies was established in management literature.[49] Because of the huge impact large companies have in the life of their stakeholders and the community, Bowen states that companies have a social responsibility that exceeds the mere notion of being profitable, advocating a rather philanthropic understanding of social responsibilities.[50] The philanthropic notion, however, sparked the discussion whether it is economically reasonable to spend corporate resources on social ends. Davis (1960) argues that corporate actions might very well go "…beyond the firm's direct economic or technical interest", but positive economic effects might only show up in the long run.[51] Walton (1967) adds to this by stating that the benefits of voluntary philanthropic actions might ultimately not even be measurable

[44] Schwalbach/Klink (2012, p. 224)
[45] cf. ibid. (p. 225)
[46] cf. ibid. (p. 233)
[47] cf. Macintosh (1999, p. 150)
[48] cf. Berle (1932, pp. 1369-1370)
[49] cf. Carroll (1999, p. 269)
[50] cf. ibid. (p. 270)
[51] Davis (1960, p. 70)

in economic terms.[52] Hence, in the early 1970s, most prominently, Nobel laureate Milton Friedman postulated shareholder primacy (see above), which finds economic outcomes being the central rationale of managerial decision-making.

During this debate, however, the notion of CSR relying on voluntary and philanthropic actions of single managers changed when the social responsibility of an organization as a whole came into focus.[53] In a prominent statement from 1971, the business-led Committee for Economic Development, for example, argues that it is crucial for companies to serve societal needs.[54] It states that "… business exists to serve society".[55] A more granular view is brought up by Carroll (1979), who defines the obligations an organization has to society as an embodiment of economic, legal, ethical, and discretionary expectations.[56] Meeting these expectations is a matter of corporate decision-making. Consequently, management literature picks up CSR as a rationale for a new perspective on Governance.[57] Freeman and Reed (1983) introduce the stakeholder concept. Managerial decision-making is happening on behalf of both shareholders and stakeholders (i.e., each interest group connected to the company) alike, as firms and society are mutually interdependent.[58] The understanding of CSR, mainly featuring ethical aspects, is gradually replaced by a broader concept. Carroll (1991) introduces the "Pyramid of Corporate Social Responsibility", which defines four major responsibilities of firms within society.[59] While being profitable and obeying the law are foundations for the existence of an organization, firms are obliged to act ethically correct and be philanthropic. All of this culminates in the picture of the firm as a "good corporate citizen" that contributes resources to the community and improves the quality of life within a society.[60]

2.1.3 A societal necessity

Societal expectations tend to change over time, and as such, responsibilities change and need to be addressed in a contemporary manner. Such changes tend to develop alongside major societal shifts. Consequently, CSR needs to be understood dynamically. [61] In 1972, Meadows et al. published a highly influential report on *The Limits to Growth* commissioned by the Club of Rome.

[52] cf. Walton (1967, p. 18)
[53] cf. Waßmann (2014, p. 4)
[54] cf. Committee of Economic Development (1971, p. 16)
[55] ibid.
[56] cf. Carroll (1979, p. 499)
[57] cf. Carroll/Hoy (1984, p. 48)
[58] cf. Freeman/Reed (1983, p. 88)
[59] Carroll (1991, p. 39)
[60] cf. ibid. (p. 42)
[61] cf. Bassen/Jastram/Meyer (2005, p. 235)

Based on a computer simulation, the report finds that the limits of global population growth, industrial output, pollution, food production, and natural resources could be reached within the next one hundred years. Altering these growth trends would require an ecological and economic equilibrium in which each person has an equal opportunity to satisfy their basic material needs.[62] The report falls into a time when many of the challenges of an increasingly globalized world, namely environmental and humanitarian, were not yet adequately identified nor addressed.[63]

The United Nations Conference on the Human Environment in Stockholm (1972) sparked a variety of multilateral activities evolving around ecological protection efforts (e.g., the United Nations Environmental Program).[64] One major landmark subsequent to the Stockholm Conference was the creation of the UN Brundtland Commission in 1981, formerly known as World Commission on Environment and Development. The Brundtland Commission's goal was to coordinate and foster international collaboration on behalf of sustainable development.[65] In 1987 the Brundtland Commission published its final report, *Our Common Future*. The report introduced a definition of sustainable development that is still the guiding principle shaping efforts designed to address challenges associated with economic, ecological, and humanitarian development of the world: intergenerational equity.[66]

> "Humanity has the ability to make development sustainable to ensure that it meets the need of the present without compromising the ability of future generations to meet their own needs."[67] *Our Common Future*, 1987, I.3. (27).

In 1989 the United Nations adopted the Montreal Protocol. The treaty was designed to acknowledge the human role in ozone depletion and to find measures to protect the ozone layer. The Montreal Protocol initiated significant shifts in the understanding of the role of industrial production in global warming. Based on the precautionary principle, the treaty introduced measures to gradually eliminate the production of several climate-wrecking gases and industrial chemicals. The Montreal Protocol is the first UN Protocol ratified by all member countries of the UN and is considered the most successful multilateral agreement on environmental policy.[68]

Intergenerational equity and the precautionary principle remained the main characteristics of all other subsequent efforts of the world community to foster sustainable development. Among those,

[62] cf. Meadows et al. (1972, pp. 23-24)
[63] cf. Handl (2012, p. 1)
[64] cf. Linnér/Selin (2013, p. 972)
[65] cf. ibid. (p. 976)
[66] cf. Schneider (2012, p. 24)
[67] Brundtland (1987, I.3.(27))
[68] cf. European Commission (2009), cf. Gonzales/Taddonio/Sherman (2015, p. 122)

significant events were the UN Conference on Environment and Development in Rio de Janeiro (1992), the UN Framework Convention on Climate Change in Kyoto (1997), the World Summit on Sustainable Development in Johannesburg (2002), the UN Conference on Sustainable Development in Rio de Janeiro (2012), and the UN Climate Change Conference in Paris (2015).[69]

The outcomes of all the events above express diplomatic efforts to conquer environmental and humanitarian challenges that the world community has become increasingly aware of. The degree to which states engage in the process of putting measures into actual law might differ immensely. However, firms face an increasing obligation to adapt as expectations — either mandated by law or demanded by society — change alongside trends in social development.[70]

2.1.4 A political postulation

During the 1990s, the already-mentioned notion of good corporate citizenship continued to gain prominence. However, the definition of CSR remained relatively undefined when it came to actual implementation in the form of business practices. This lack of mutual understanding led to a rather low level of acceptance within global markets.[71] In 1999 then Secretary General of the United Nations, Kofi Annan, raised global attention to this topic, as he proposed that "… the business leaders …, and we, the United Nations, initiate a global compact of shared values and principles, which will give a human face to the global market" at the World Economic Forum (WEF) in Davos.[72] Subsequent to the offer, the International Chamber of Commerce gathered about 50 globally active companies and business associations, and civil and labor society organizations to create the United Nations Global Compact (UNGC) in 2000.[73] Even though the UNGC did not explicitly promote CSR, it most notably provided universal principles that UNGC participants should use to guide business strategies and processes. These principles refer to the protection of human rights, labor policies that address the freedom of association, the abolishment of child and forced labor, precautionary measures to protect the environment, and the opposition to corruption and discrimination.[74] As of today, the number of signatories has grown to over 12,000 companies and over 7,000 other related organizations and associations.[75]

The momentum of global efforts to address economic and humanitarian challenges led to the international commitment of the UN member states to the Millennium Development Goals (MDG)

[69] cf. Linnér/Selin (2013, p. 972)
[70] cf. Lantos (2001, pp. 599-600)
[71] cf. ibid. (p. 595)
[72] Annan (1999, https://press.un.org/en/1999/19990201.sgsm6881.html, accessed: 01/10/2022)
[73] cf. Gonzales-Perez/Leonard (2017, pp. 118-119)
[74] cf. ibid. (p. 120)
[75] https://www.unglobalcompact.org/what-is-gc/participants, accessed: 12/07/2021.

in 2000. The idea was to create universal and measurable outcomes that should shape political strategies to promote environmental sustainability and address poverty, hunger, education, gender equality, child mortality, and diseases like HIV, AIDS, and malaria.[76] Although the MDGs did not directly express any guidance on behalf of CSR practices within companies, they represented universal goals the world community agreed upon in order to conquer the most prevailing issues.[77] It is reasonable to assume that such issues shape the expectations upon business leaders around the globe. As such, the United Nations explicitly called for the engagement of business enterprises within the Sustainable Development Goals (SDG), which followed the MDGs in 2015. The SDGs primarily focus on governmental action but explicitly invite businesses to "do more good".[78]

The question of how to do such good is subject to a debate on political measures that set standards for corporations and their business practices. Starting with the foundation of the Commission on Transnational Corporations in 1973, the UN was trying to establish standards and a code of conduct for the business-making of multinational firms.[79] The first attempt failed as the interest of various groups involved in the processes could not be moderated. In 2005 the UN appointed John Ruggie, Harvard professor of Human Rights and International affairs and one of the architects of the UNGC and the MDGs, as the Secretary-General's Special Representative for Business and Human Rights. Based on his work, the UN Human Rights Council presented the "Protect, Respect and Remedy" framework in 2008, which was designed to guide the debate on CSR and what regulators around the globe should do in order to anchor the implications of the International Bill of Human Rights in the corporate world.[80] In 2011, the United Nations Guiding Principles on Business and Human Rights (UNGP) operationalized this normative framework. The UNGP provided actual advice on what public authorities, business organizations, and states should do to analyze and manage the impact of corporate action on human rights and to ensure that business organizations keep a "social license to operate".[81]

Especially the European Union adopted strategies to explicitly promote CSR starting in the early 2000s. In 2001, the European Commission (EC) presented a Green Paper on CSR. Most notably, the EC defined CSR as a "… concept whereby companies integrate social and environmental concerns in their business operations … on a voluntary basis".[82] The notion of voluntary CSR was

[76] cf. Hulme (2009, pp. 49-50)
[77] cf. ibid. (pp. 47-48)
[78] Buhmann/Jonsson/Fisker (2019, p. 394)
[79] cf. Weilert (2010, p. 463)
[80] cf. Buhmann (2016, p. 701)
[81] cf. UN (2011, p. 1), Buhmann (2016, pp. 701-702)
[82] European Commission (2001, p. 6)

heavily endorsed by the lobby of European enterprises within the subsequent EU Multi-Stakeholder Forums in 2002 and 2004. These events ultimately shaped the early approach of the EC to encourage companies to exceed legal obligations in order to promote social and sustainable developments.[83]

In 2011 the EC revised its definition of CSR and promoted regulations that no longer let CSR be a voluntary engagement of firms. It follows the notion that enterprises bear responsibility "for their impacts on society".[84] As such, the EC opted for regulations that create a level playing field on behalf of CSR-related policies and practices. Examples of the intended sort of regulations promoted sincere approaches to address multi-stakeholder needs, enhancing attention toward human rights, promoting market rewards for strong CSR conduct, and enhancing transparency via mandatory non-financial disclosure. The latter is considered the most important outcome of the new CSR definition. In 2014, the EU introduced mandatory non-financial reporting via the accounting Directive 2014/95/EU. From 2017 on, public interest entities were urged to disclose information on the environmental, social, and human rights-related impacts of their corporate actions. Although this is considered a major leap in the enforcement of CSR as an integrated business objective, firms, on the one hand, and civil society organizations, on the other hand, were arguing about the effectiveness, adequacy, and efficiency of the new regulations.[85]

More pressure on the corporate world and financial markets is expected from the most recent multinational efforts to conquer global warming, starting with the Paris Agreement resulting from the UN Climate Change Conference in 2015.[86] It resembles the multi-lateral agreement on limiting the rise in mean global temperature and reaching net-zero emissions by the middle of the 21st century.[87] The European Union is trying to become the world's leading economy and foster a politically driven transformation toward sustainable growth.[88] Major efforts that affect European corporations to deliver that goal were implemented within the "European Green Deal" introduced in 2019.[89] The European Green Deal involves measures that put new regulations on financial market participants in order to embed the concept of sustainable finance, a new action plan to promote a

[83] cf. Kinderman (2018, pp. 108-109), cf. European Commission (2006)
[84] European Commission (2011, p. 6)
[85] cf. Kinderman (2018, pp. 109-111)
[86] cf. Koundouri/Devves/Plataniotis (2021, p. 744)
[87] cf. UN (2015, pp. 3-4)
[88] cf. Eckert/Kovalevska (2021, p. 19)
[89] cf. Koundouri/Devves/Plataniotis (2021, p. 746)

circular economy, and a classification system to identify sustainable and eco-friendly business practices.[90]

As a major example of the efforts to implement sustainable finance, the European Commission in 2016 pushed forward a plan to guide future capital market development, promoting a financial system aware of the current and future challenges associated with climate change. The High-Level Expert Group on Sustainable Finance was established to provide advice on how to stimulate public and private capital flows toward sustainable investments as well as the measures to protect the financial system's stability from environmental risks.[91] The final report states that the essential prerequisite is the introduction of a mutual understanding of what is meant by the term "sustainability".[92] In 2020 the Technical Expert Group on Sustainable Finance finally published a report and introduced a universal taxonomy for European capital markets (so-called "EU taxonomy"). It contains criteria to screen companies and financial institutions for actions that foster climate change mitigation and adaptation as well as activities that harm environmental objectives.[93]

2.1.5 Integrating business interests, societal necessities, and political expectations

Policymakers are trying to generate actionable objectives that address potentially fuzzy definitions within several fields of CSR. While the measures of the European Union certainly influence policy-making within other parts of the world, it is not clear how effective the new measures actually are and how well they oppose non-responsible business conduct. Eckert and Kovalevska (2021) state that there are advocates of the view that a sustainable transformation of the economy might only be possible by limiting growth opportunities. As a consequence, especially corporations will hold on to opposing measures that are politically designed to foster sustainable development.[94]

From a scientific view, this might already be addressed with the concept of 'shared value' introduced by Porter and Kramer (2006). As corporations and the societies they act in are mutually interdependent, a firm's responsibility is to identify those areas that reflect the greatest opportunity to create benefits for both society and the competitiveness of the firm.[95] The implementation of measures to create such shared value could be taken at any given point within the value chain.[96] Trapp (2012) puts the idea further as he states that firms might not be able to address specific

[90] cf. Koundouri/Devves/Plataniotis (2021, pp. 747-748)
[91] cf. European Commission (2016)
[92] cf. European Commission (2018, p. 13)
[93] cf. European Commission (2020, p. 2)
[94] cf. Eckert/Kovalevska (2021, pp. 19-20)
[95] cf. Porter/Kramer (2006, p. 5)
[96] cf. ibid. (p. 5)

issues by themselves. They are rather obliged to address them collaboratively, driven by their ethically motivated concern for global challenges. Trapp states that when firms obey the law (first generation), address the needs of their stakeholders (second generation), and start to feel responsible for the development of all humankind, this could be seen as the "third generation" of CSR.[97]

Carroll (2015) summarizes that CSR has contributed to mainstream responsible business practices over the last decades. Several concepts have evolved around the idea of fostering societal improvements while sustaining profitability and competitiveness (e.g., corporate citizenship, business ethics, stakeholder management, creating shared value, and sustainability). Carroll attributes this development to four major trends. Firstly, firms have been competing within markets that have become increasingly aware of the corporate impact on ecological and social issues. Simultaneously, firms faced growing pressure to address societal expectations because of the increasing visibility of their actions, gradually enhancing their exposure to reputational risks. As a result, CSR is continually institutionalized within the internal organizations of firms and becoming an integral part of decision-making, policies, and organizational structures. The third trend Carroll identifies is the acceptance within the research community, as scholars of several disciplines (e.g., economics, law, engineering, and journalism) increasingly focused on questions regarding CSR.[98] The fourth, and most important trend, however, might have been and will be most deterministic to the development of CSR. Combining profitability and CSR engagement will hold on to drive the actual application of CSR-related concepts. Nevertheless, Carroll argues that in a realistic projection of the future, a social purpose and respective practices will always find a baseline kind of legitimization and appreciation.[99]

In summary, the social responsibility of a firm ever since has been the result of societal expectations toward the corporate world and political, diplomatic, and regulatory efforts to conquer prevailing issues that potentially harm the prosperity and well-being of society. The short revision of the development of CSR shows which dynamics and interdependencies drive CSR and how they contribute to non-financial dimensions in corporate performance primarily concerned with a firm's impact on the Environment and Society. The major driver of CSR, however, will likely remain its economic reasoning. Following the notion of shareholder primacy, the question arises of how CSR measured by entity-specific ESG performance aligns with the motivations of capital market participants as the suppliers of financing.

[97] cf. Trapp (2012, p. 459)
[98] cf. Carroll (2015, p. 1)
[99] cf. ibid. (p. 2)

2.2 ESG as performance parameters in investing

2.2.1 Investor motives

As pointed out in the introduction of chapter 2.1, a major concern in the debate over CSR (i.e., corporate engagement in ESG issues) is how it complies with the notion of shareholder primacy. Shareholder primacy is — especially in Anglo-Saxon countries — the central principle that informs Corporate Governance and how potential conflicts of interests between shareholders and the management are defined.[100] Applied to ESG investing, conflicts of interest would arise when the management of the firm commits to ESG goals that the owners (shareholders) do not internalize as their own goals.[101] Literature, however, suggests that shareholder value (i.e., measurable financial outcomes) might not necessarily be the sole consideration of investors.[102] In fact, the internalization of ESG issues is driven by a wide range of altruistic, material (financial or regulatory), and self-serving motivations.[103] These motivations constitute three cases in which private investors predominately use ESG information: 1) to align investments with their own values and beliefs, 2) to foster a positive impact, or 3) to benefit financially (i.e., profits or risk).[104] Institutional investors are additionally concerned with meeting client demand (based on the above-stated motivations) and regulatory obligations.

Especially the latter is increasingly fostering the integration of ESG information into investment processes. Recent regulatory advances, like the 2022 SEC ESG disclosure proposal,[105] or the 2022 EU Regulation on sustainability-related disclosure in the financial services sector (SFDR),[106] are already or are expected to put pressure on institutional investors to prioritize ESG factors. As of 2022, for example, EU SFDR requires institutions to disclose the extent to which certain investment products are aligned with the EU Taxonomy for sustainable activities.[107] Not entirely as specified as in the European regulation, but just recently, the SEC announced plans on the introduction of disclosure allowing for "… consistent, comparable, and reliable information for investors concerning funds' and advisors' incorporation of … ESG … factors."[108]

[100] cf. Goergen (2018, pp. 3-5)
[101] cf. Tirole (2001, p. 4), cf. Larcker/Tayan (2021, p. 398)
[102] cf. Hart/Zingales (2017, p. 270)
[103] cf. Bénabou/Tirole (2010, pp. 15-16)
[104] cf. Matos (2020, p. 39), cf. Amel-Zadeh/Serafeim (2018, pp. 11-12)
[105] https://www.sec.gov/news/press-release/2022-92, accessed: 05/27/2022.
[106] https://finance.ec.europa.eu/sustainable-finance/disclosures/sustainability-related-disclosure-financial-services-sector_en, accessed: 05/01/2022.
[107] ibid.
[108] https://www.sec.gov/news/press-release/2022-92, accessed: 05/27/2022.

Regulation on the disclosure of ESG performance is coined to serve two distinct goals. On the one hand, European regulation is primarily designed to make the financial industry the key player financing a sustainable transformation of European economies.[109] On the other hand, the notion of the SEC's proposal is rather driven by the need for investor protection, intending to allow investors to "… drill down to see what's under the hood" of ESG investing strategies and products.[110] Both regulatory advances build upon the recognition of a growing demand for investment practices aligned with ESG issues. However, one can derive differences in the presumed investment goals motivating the regulation. The SEC's proposal is based on the SEC's main task to administer laws that are "…based on the notion that disclosure of financial results and business risks … must be sufficient for investors to make informed decisions about how to allocate scarce capital." (i.e., investor-driven).[111] Hence, the SEC proposal assumes primarily economically driven use of ESG metrics. The SFDR, however, rather builds upon the idea of supporting those causes that create positive externalities (i.e., impact-driven).

Survey studies confirm the presumption of primarily economically motivated investors. Amel-Zadeh/Serafeim (2018) show that among institutional investors, the belief in the value-relevance of ESG performance is the major driver to invest alongside ESG information.[112] Investors try to lower investment risks (e.g., reputational damage or long-term risk associated with ESG issues) in their portfolios or focus on specific business opportunities and future market conditions that go in hand with changes in the overall perception of ESG issues.[113] The belief in value-relevant features of ESG performance is backed up by a wide body of research literature. Studies from various disciplines in management and economics investigate the link between ESG performance and financial or organizational performance.[114]

[109] https://finance.ec.europa.eu/publications/renewed-sustainable-finance-strategy-and-implementation-action-plan-financing-sustainable-growth_en, accessed: 05/01/2022.
[110] Gensler (2022, https://www.sec.gov/news/press-release/2022-92, accessed: 05/27/2022)
[111] Eccles/Crowley (2022)
[112] cf. Amel-Zadeh/Serafeim (2018, p. 34)
[113] cf. Matos (2020, p. 39)
[114] See chapter 3.3 for a more extensive overview of research fields investigating the outcome of corporate commitment to ESG goals.

The debate is primarily concerned with two questions: Does commitment to higher ESG performance make firms more profitable and less risky ("doing well by doing good")? Or, do successful firms have more resources to engage in ESG issues ("doing good by doing well")?[115] To shortly reflect on the general notion: There seems to be no clear-cut evidence for either of those relationships. The primary concern not allowing to draw such an overarching conclusion lies in the vast amount of different definitions and various conceptualizations underlying the term ESG performance.[116] Chapter 3.1 deals with the wide range of potential issues informing the understanding of ESG performance and how these pose a challenge to the very nature of ESG performance measurement. Focusing on single aspects within the realm of ESG issues instead makes it possible to gather more selective evidence.[117] Chapter 3.2 briefly discusses how the concept of materiality, which refers to ESG aspects that significantly impact firm performance, differs on various dimensions (i.e., time, region, and industry), inducing a dynamic nature to the relationship between ESG and firm performance.[118] Chapter 3.3 is turning toward the third challenge in ESG performance measurement. It refers to the distinctive theoretical underpinning of any link between commitment to ESG goals and corporate performance. Different activities considered ESG-relevant might be associated with various underlying assumptions and resources. Hence, any relationship necessarily needs to rest upon presumptions of causality or actual evaluations of the achievement of intended goals, which are certainly not trivial (as described above).[119]

This challenge also holds true for the second major motivation of ESG investing. Investors utilize ESG information as a measure to identify those investment opportunities that promote the solution of a real-world problem in the realm of ESG issues. Although such a real-world impact and the already stated question on the financial benefits of implementing ESG information in investment contexts are somewhat vague and hard to assess, it does not necessarily oppose the motivation to use ESG information for such purposes. In terms of real-world impact driven by the investment in ESG-committed firms, Heeb et al. (2022) suggest that the impact is of subsequent meaning. Investors instead tend to follow a certain belief that does not necessarily need to match actual real-world outcomes.[120]

As intended outcomes and the impact of investing alongside ESG information are based on certain beliefs, another primary motive for ESG investing becomes apparent. Investors tend to follow ESG

[115] Matos (2020, pp. 30-31)
[116] cf. Friede/Busch/Bassen (2015, pp. 225-226), cf. Matos (2020, p. 27), cf. Whelan et al. (2021, p. 3)
[117] cf. Matos (2020, pp. 31-32)
[118] cf. Khan/Serafeim/Yoon (2016, p. 1716)
[119] cf. Howard-Grenville (2021, p. 1), cf. Crace/Gehman (2022, pp. 5-6)
[120] cf. Heeb et al. (2022, pp. 39-40)

investing strategies to align with personal values. Many of which have long-standing ethical and religious roots. The Christian church, for example, is a prominent origin of ethical principles underlying most western societies.[121] Those principles largely contributed to what is seen as the first modern considerations of ESG-based investment strategies, especially in Northern America and Europe. The Quakers, since the 17th century and, more recently, churches in the early 1900s, refused to engage in controversial business fields. They strictly advocate against investments in slavery, tobacco, alcohol, weapons, and gambling, the so-called "sin stocks".[122] Another example is the Pax World Fund, which is considered the first ethical investment fund. It was launched as an alternative investment option, distinctively excluding firms profiting from producing nuclear and military arms.[123]

Restricting investment opportunities to stocks that conform to a particular religious or ethical belief is an example of investment strategies that play a prominent role in the field of ESG-based investment products. The next chapter describes and categorizes corresponding investment strategies and their prominence, especially among institutional investors.

2.2.2 Investing strategies

Research literature and investment practice show a variety of methods to incorporate ESG information in investment decision-making. There is no official taxonomy of the various investment types. However, the CFA Institute developed a categorization scheme that summarizes the most commonly applied ESG-based investment strategies found in scholarly publications and reports of institutions like the UN PRI or the Global Sustainable Investment Alliance.[124] ESG investment strategies can be broadly divided into three categories: Screening, integration, and engagement (**Figure 3**).[125]

Screening

Screening strategies incorporate ESG information to directly identify companies to invest in. The most straightforward approach to address certain ESG concerns is *negative screening*. Based on specific ESG traits, negative screening enables companies or certain sectors to be excluded from an investment universe. Companies are excluded even when the prospects of investing in them

[121] One example is the opposition of interests on loans. „If you lend money to one of my people among you who is needy, do not treat it like a business deal; charge no interest", New International Version of the Protestant Bible (1978, Old Testament, Exodus 22:25). Similar guidelines also contribute a major pillar of Islamic finance ("Riba"), Warde (2010, p. 55).
[122] Cf. Louche/Arenas/van Cranburgh (2012, p. 302)
[123] Cf. Renneboog/Terhorst/Zhang (2008, pp. 1725-1726)
[124] cf. Hayat/Orsagh (2015, p. 29), cf. GSIA (2020, p. 7), cf. Amel-Zadeh (2018, p. 89)
[125] cf. Matos (2020, p. 39)

otherwise appear profitable. Characteristics or practices that lead to the exclusion are often grounded in moral standards and religious beliefs (*value-based*) or certain internationally accepted standards on human rights, labor practices, or environmental protection (*norms-based*).[126]

As stated in the previous chapter, prime examples of value-based investment follow religious beliefs, for example, by avoiding investments in "sin stocks" (i.e., firms engaged in slavery, tobacco, alcohol, weapons, or gambling).[127] Similar considerations can also be found in Islamic finance, which identifies specific business activities as impermissible. Prominent examples of companies deemed ineligible for investment are financial institutions that rely upon interest payments or restaurants, hotels, and airlines that sell alcohol.[128]

Apart from value-based exclusion, negative screening is also applied based on widely accepted international norms, e.g., the Principles of the UN Global Compact (**Appendix 2**). Actions that violate such norms primarily concern human rights and working conditions, e.g., forced and child labor, discrimination, or censorship and impeding free speech.[129] A historical example is the massive divestments in South African companies during apartheid in the 1970s.[130] Considering environmental protection norms, one might see the Volkswagen emissions scandal in 2015 as an appropriate example. As systematic misconduct regarding the emissions of specific models became public, Volkswagen lost almost a fourth of its market capitalization in one day. The immense market reaction was primarily due to the expected litigation costs and reputational damage, which underlines the implications of economic motivations of ESG investing.[131]

Positive or *best-in-class* screening refers to the distinct selection of individual companies or sectors that are preferred due to values, norms, or economic reasons. Instead of excluding companies to penalize perceived misconduct on ESG aspects or to hedge against ESG risks, positive or best-in-class approaches aim directly for companies addressing desired outcomes. Positive screening is often applied to entire sectors, which is then referred to as *thematic* investing, e.g., investment products targeting emission-free energy production or sustainable farming.[132] Best-in-class approaches seek to identify companies that show the best ESG performance or the most significant improvements on certain ESG aspects relative to their peers (i.e., competitors, industry, or other

[126] cf. Hayat/Orsagh (2015, p. 17)
[127] cf. Louche/Arenas/van Cranburgh (2012, p. 302)
[128] cf. Warde (2010, p. 138)
[129] cf. Hayat/Orsagh (2015, p. 18)
[130] cf. Johnston/Morrow (2016, p. 4)
[131] cf. ibid. (pp. 3-4)
[132] cf. Matos (2020, p. 39)

peer groups that investors deem sufficient).[133] Security selection then follows the idea of over-weighting ESG leaders and underweighting ESG laggards.[134] Most often, this approach is used in ESG rating methodologies. It accounts for sector-specific differences in the relevancy of ESG aspects, e.g., greenhouse gas emissions being more important for automobile manufacturers than insurance companies.[135]

Integration

The systematic incorporation of ESG information in financial analyses is referred to as *ESG integration*. ESG integration is a holistic approach, which does not require the investor to dedicate to predetermined ESG goals.[136] ESG information becomes an integral part of each step in the analysis of investment opportunities. The integration includes market inquiry, security selection, portfolio formation processes, risk monitoring, and rebalancing.[137]

Risks and opportunities associated with specific ESG issues are examined in more detail than screening approaches, as the latter tend to rely on rather broad and fundamental perceptions of ESG issues.[138] ESG integration approaches focus on two particular dimensions. On the one hand, outcomes of specific corporate conduct on certain ESG issues are analyzed. The analysis relates to questions on how specific ESG behavior (e.g., a stronger engagement in lowering climate-gas emissions) affects the prospects of a business in comparison to firms that act differently. On the other hand, ESG integration builds on evaluating how certain circumstances affect a company's business model from the outside. Hence, it requires, for example, the analysis of changing weather conditions related to climate change and how they affect the current business model. In doing so, ESG integration regularly depends on projections of such relationships into the future that are only possible in an assumptive manner.[139]

Engagement

The third major type of investment strategies targets a single company's attitude toward a particular ESG issue. By using ownership rights, shareholders try to effect change by engaging with the management of the firm.[140] Unlike the other approaches listed above, it requires a prior investment into the targeted firm. The motivation to do so builds upon the idea of gaining the possibility to

[133] cf. Hayat/Orsagh (2015, pp. 18-19)
[134] cf. ibid. (p. 25)
[135] cf. Wong/Petroy (2020, p. 25), cf. Eccles et al. (2012, pp. 66-67)
[136] cf. Hayat/Orsagh (2015, p. 25)
[137] cf. Matos (2020, p. 40), cf. van Duuren/Platinga/Scholtens (2016, p. 526)
[138] cf. van Duuren/Platinga/Scholtens (2016, pp. 526-527)
[139] cf. ibid. (pp. 532-533)
[140] cf. Hayat/Orsagh (2015, pp. 20-21)

impact the ESG performance of a company directly. Such investment approaches are referred to as *active ownership*, *shareholder engagement*, or *stewardship*. Shareholder engagement happens individually, in collaboration with other shareholders, or by the use of proxy votes.[141] Actions referred to as active ownership either occur in private consultation of the targeted company, for example, by meeting with senior management or writing open letters to the company's executive board. Rather public shareholder engagement involves the execution of voting rights or raising questions at a general meeting, filing a shareholder resolution, trying to gain a seat on the board of directors, or issuing statements to news media.[142]

ESG investment strategies

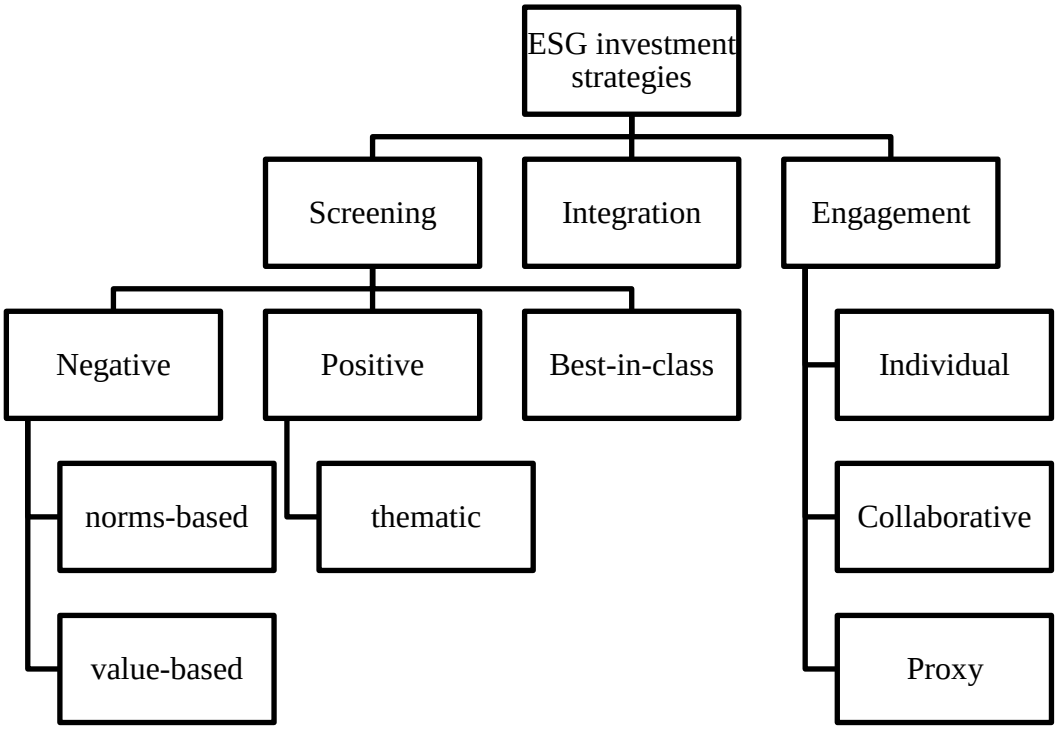


Figure 3 ESG investment strategies, source: Matos (2020, p. 39).

The Global Sustainable Investment Alliance (GSIA) gathers data on the usage of ESG investment strategies around the globe. Within its biennial *Investment Review*, the GSIA summarizes data for major financial markets in the United States, Europe, Japan, Canada, Australia, and New Zealand. The data is based on surveys of institutional investors (e.g., by the United States Sustainable Investment Forum, the European Fund and Asset Management Association, or the Japan Sustainable Investment Forum) and other publicly available third-party data sources.[143]

[141] cf. Matos (2020, p. 40)
[142] cf. Hayat/Orsagh (2015, pp. 20-21)
[143] For a detailed view on the methodology: GSIA (2021, pp. 28-29)

As already pointed out in the introduction, within its latest report, the GSIA estimates that at the beginning of 2020, US$ 35.4 trillion of assets were professionally managed following ESG investment approaches.[144] Measured in AUM, about 41.6% of institutional investments in Europe and 33.2% of investments from the United States rely to some degree upon the abovementioned considerations.[145] In detail, ESG integration and negative screening are the most utilized investment approaches. Interestingly, there are disparities within different regional groupings.[146] **Figure 4** depicts the estimations of the GSIA for the year 2020.

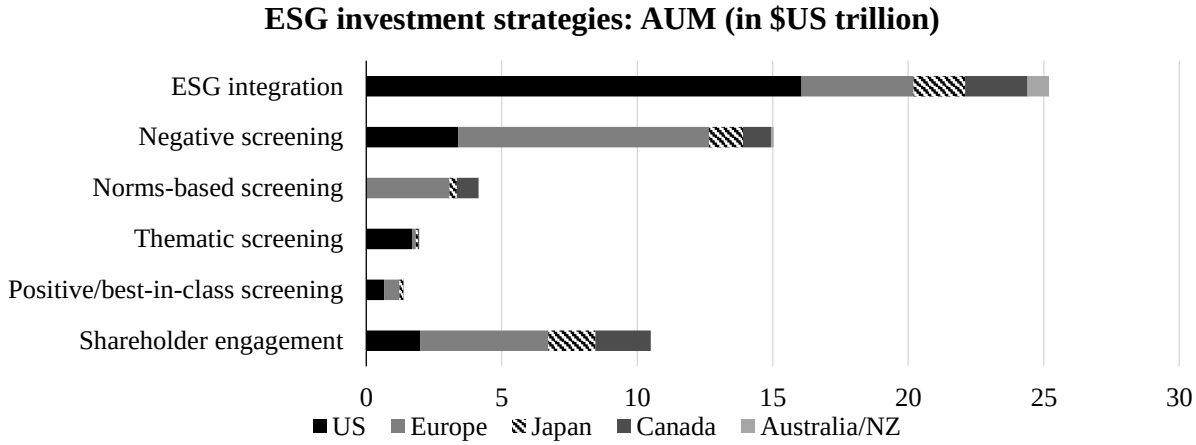


Figure 4 ESG investment strategies: AUM (in $US trillion), source: GSIA (2021, p. 11).

ESG integration is the most common investment approach (US$ 25.2 trillion) among institutional investors. The second most prevailing approach is negative screening (US$ 15.9 trillion).[147] Historically norms-based and negative screening has been the prevailing ESG investment approach in Europe. The 2019 EU Sustainable Finance Disclosure Regulation promotes practices like negative and norms-based screening as well as ESG integration even stronger, as it requires institutional investors to incorporate ESG risks into their investment decisions.[148] Consequently, negative and norms-based screening approaches are especially prevailing among European investors. Shareholder engagement (US$ 10.5 trillion) is the third most prominent investment approach.[149]

Similar results are shown when looking at the investment approaches UN PRI signatories apply. The UN PRI is the world's largest multinational investor initiative that promotes ESG incorporation in the investment industry. Investors can list as signatories and thereby commit to the six UN principles of responsible Investment (**Appendix 1**).[150] Based on PRI signatory data Matos (2020)

[144] See **Footnote 5** for further information on the estimation of global AUM.
[145] cf. GSIA (2021, p. 10)
[146] cf. ibid. (p. 11)
[147] cf. ibid. (p. 10)
[148] cf. ibid. (p. 13)
[149] cf. ibid. (p. 11)
[150] cf. Matos (2020, pp. 34-35)

confirms the findings of the GSIA. However, differentiating between equity investments- and debt-based instruments reveals further peculiarities in the usage of ESG investment approaches. Measured in the percentage of PRI signatories that report on using a particular investment approach, ESG integration (87%) and negative screening (73%) still mark the prevailing strategies. The majority of signatories also report they engage with the firms they invested in (86%). As active ownership requires voting rights obtained through an equity investment, engagement is not reported for fixed-income instruments (e.g., sovereign, corporate, or securitized bonds).[151]

Figure 5 shows the aforementioned application of ESG investment strategies among UN PRI sig-

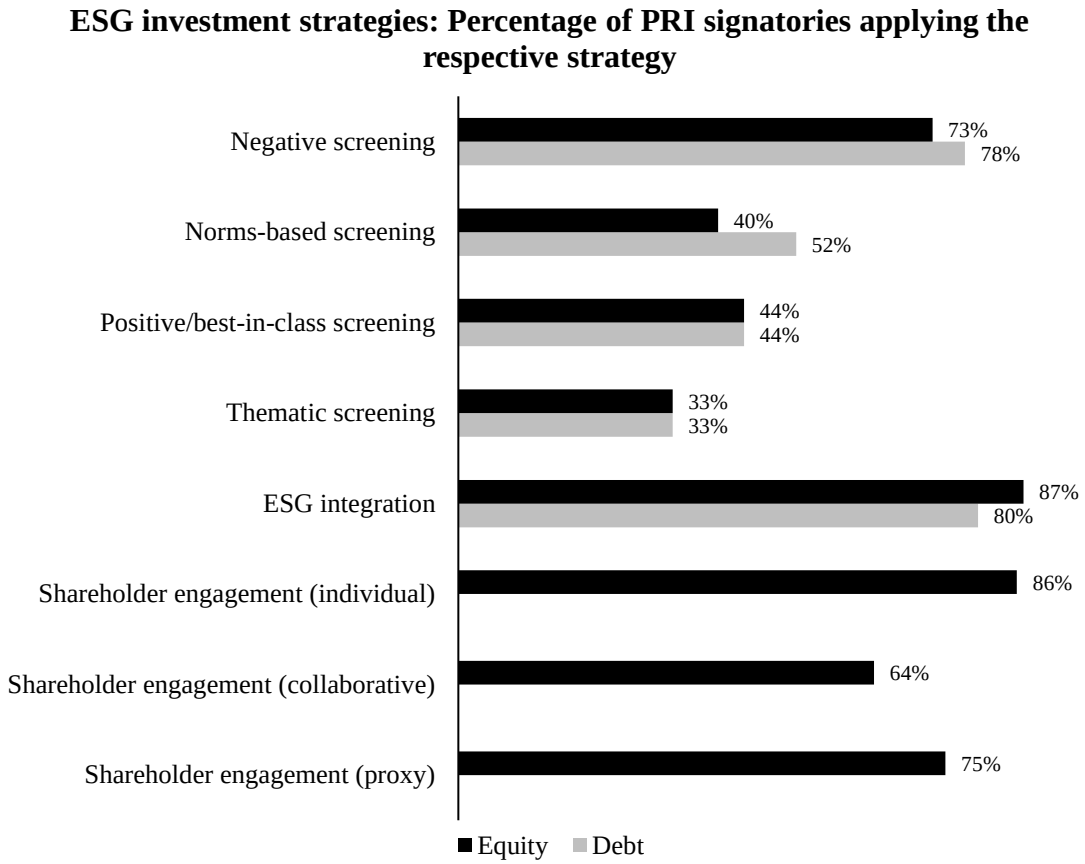


Figure 5 ESG investment strategies: Percentage of PRI signatories applying the respective strategy, source: Matos (2020, p. 42).

natories. It is important to note that ESG investment strategies are not mutually exclusive. In most cases, the different approaches share commonalities and impose similar steps in investment decision-making.[152] One of which is the use of ESG performance metrics in order to streamline ESG research efforts. Third-party data providers, e.g., RobecoSAM, Sustainalytics, MSCI, or Refinitiv, systematically analyze a company's CSR commitments and strategies in place to reach certain

[151] cf. ibid. (pp. 41-42)
[152] cf. ibid. (p. 40)

ESG goals.[153] Such so-called ESG ratings usually provide information on a variety of different firms and industries. Analyses are primarily based on financial and sustainability reporting, media coverage, and other non-financial information. Users of the ratings then get an indication of how single firms and industries perform on certain ESG issues.[154]

As already pointed out in the introduction, throughout the dissertation, the terms "ESG scores", "ESG metric", and "ESG ratings" interchangeably refer to aggregates of ESG information provided by third-party ESG data sources in order to measure ESG performance. The next chapter briefly describes the market for third-party ESG performance metrics and how they potentially differ.

2.2.3 Market for ESG performance metrics

Given the prominence of ESG investing strategies (**Figure 1**) and recent regulatory efforts described above, ESG information is an increasingly integral part of investment decision-making and professional financial advisory. Based on a survey among institutional investors and financial industry experts, mainly from Europe and the United States, the SustainAbility Institute gathers data on which and how ESG information is used.[155] According to their latest *Rate the Raters Report* (2020), ESG metrics by third-party providers rank among the primary sources of institutional investors to gain information on corporations and their ESG conduct.[156] The majority of survey participants emphasize the crucial role of ESG metrics in supplementing research efforts and risk assessment on corporate ESG performance. Survey participants also state the integration of ESG metrics into investment decision-making as a direct translation of customer demand and legal requirements.[157]

Not surprisingly, there is a thriving industry that gathers ESG information on companies, assigns scores and ratings, and ranks firms among their peers. In a recent report, management consultancy Opimas estimates the volume of the market of ESG data providers (based on the revenues of its participants) at around US$ 1 billion in 2021, following a decade of constant growth.[158] Although Opimas finds three leading players (i.e., MSCI, ISS ESG, and Sustainalytics) making up almost

[153] cf. Munoz-Torres et al. (2018, p. 8)
[154] cf. Escrig-Olmedo et al. (2019, p. 3)
[155] The SustainAbility Institute is a private advisory firm within the multinational consultancy Environmental Resources Management (https://www.sustainability.com/about/), cf. Wong/Petroy (2020, pp. 9-10)
[156] cf. Wong/Petroy (2020, p. 17)
[157] cf. ibid. (p. 22)
[158] cf. Foubert (2022; https://www.opimas.com/research/742/detail/, accessed: 06/26/2022)

60% of the market volume,[159] it is estimated that about 600 different ESG scores and ratings exist.[160] **Figure 6** is based on an expert survey from 2019 described in the *Rate the Raters Report* of 2020. It shows the key data providers SustainAbility identifies based on the answers of 319 financial industry experts, business professionals, and academic scholars.[161] The graph displays the differences between the fraction of answers indicating a high and a low assessment of the respective metrics. The results show that most of the respondents appreciate the quality and usefulness of metrics provided by Sustainalytics, CDP, RobecoSAM, and MSCI. Although almost all of the metrics under consideration are rated rather positively, the magnitude of the positive assessment substantially differs.

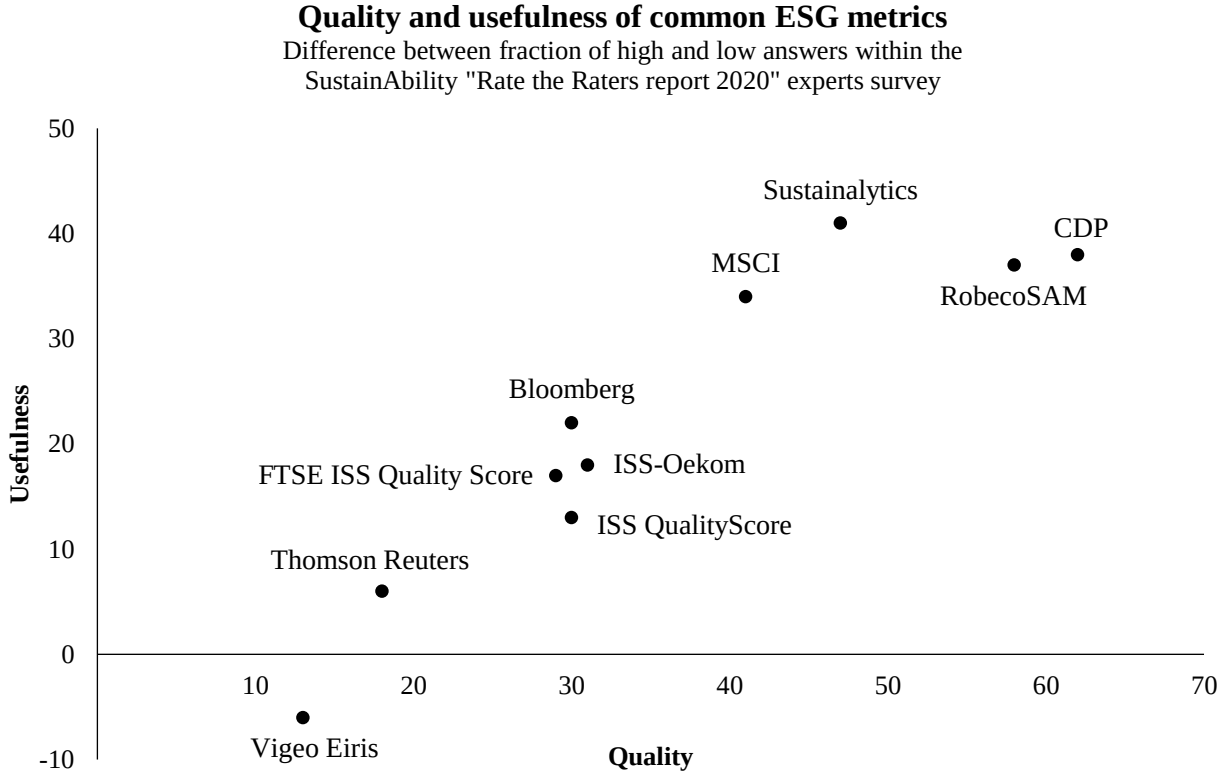


Figure 6 Quality and usefulness of common ESG metrics, based on data from: Wong/Petroy (2020, pp. 37-38).

The assessment of ESG performance relies on the ability to validate firms based on their engagement in environmental, social, and governance issues. It requires a systematic approach to each of those pillars, which builds upon aggregating the performance and (ultimately) yields an easy-to-interpret indication of the extent of commitment to ESG-related issues and the exposure to ESG risks. Each of the metrics shown above utilizes proprietary ranking and standardization methodologies.[162] Recent research literature on ESG performance metrics provides evidence on the extent

[159] cf. Foubert (2022; https://www.opimas.com/research/742/detail/, accessed: 06/26/2022)
[160] cf. Wong/Brackley/Petroy (2019, p. 4)
[161] https://www.sustainability.com/thinking/rate-raters-2019/, accessed: 06/27/2022.
[162] cf. Saadaoui/Soobaroyen (2018, p. 26)

and sources of divergence induced by different methodological frameworks. Berg, Kölbel, and Rigobon (2019) state that ESG data providers tend to differ by three characteristics: 1) The scope to which certain aspects of ESG performance are considered, 2) the measurement of ESG aspects that potentially considers different data points for the same issue, and 3) the weighting of performance on specific ESG issues when aggregating a rating.[163]

Based on the most recent methodology papers and websites on ESG performance metrics, **Table 1** provides a short overview of the central characteristics of some of the largest (by the scope of covered firms) ESG scoring universes offered by third-party ESG data providers. The number of firms covered within the data universe ranges from about 8,500 (MSCI) to 13,000 (CDP). Most ESG metrics are based on publicly available corporate disclosure, with additional overlays of media coverage or data from analysts and NGOs (so-called "passive raters"). ESG data provided by CDP relies upon questionnaires ("active rating") on climate change, forests, and water security. The active approach, on the one hand, completely alters the information the CDP assessment is based on relative to the other metrics; on the other hand, it induces a specialized focus on "the environment and natural resources".[164] RobecoSAM, which was acquired by S&P Global in 2019,[165] is based on both a set of data primarily derived from publicly available corporate disclosure and data from industry-specific questionnaires. The distinct approaches to collect data lead to severe differences in the scope of data points underlying each metric (e.g., 630 within the Refinitiv ESG metric to up to 1,300 data points within the Sustainalytics universe). Based on these data points, each metric comprises one (CDP) or up to five (MSCI) levels of ESG scores. Each level is designed to put a more granular focus on single aspects or aggregation levels of ESG performance. The resulting metrics most often represent an industry-benchmarked or -adjusted scoring, which accounts for differences in the respective relevance of different ESG issues between industries (see chapter 3.2). While the scale and aggregation methodologies are proprietary and individual in each case, it is worthwhile noticing that the underlying industry classifications substantially differ. The ESG metrics that focus on full ESG performance assessment rely on 54 (Refinitiv) to up to 138 (Sustainalytics) different industries. Other than the special focus of CDP, the other data providers intend to identify and measure "significant risks and/or opportunities" (MSCI), "exposure to and performance on key ESG risks and opportunities" (S&P), and "ESG performance, commitment

[163] cf. Berg/Kölbel/Rigobon (2019, pp. 29-30), Refinitiv ESG ratings are based on the ASSET4 database; The following ESG data providers are analyzed: MSCI, KLD, Sustainalytics, Moody's (former Vigeo-Eiris), S&P Global (former RobecoSAM), and Refinitiv (former Asset4).
[164] Wong/Petroy (2020, p. 7)
[165] https://www.spglobal.com/esg/podcasts/how-green-banks-aim-to-use-new-federal-funds-to-accelerate-low-carbon-transition, accessed: 07/01/2022.

and effectiveness" (Refinitiv).[166] Hence, the metrics put a clear focus on both the risk and opportunities presumed to accompany corporate action aware of ESG issues. The Sustainalytics ESG risk rating intends to identify "financially material ESG risks" and "how they might affect the long-term performance". Not only is the focus on ESG risks a commitment to the negative externalities of ESG issues it also frames the ESG metric as shareholder value-oriented (i.e., financially material ESG information). The CDP, for example, rather intends to "motivate companies to disclose … and take action to reduce negative impacts".

Given the apparent divergence in the methodologies of the above-stated examples, it is not surprising that the appreciation of ESG metrics from different providers (**Figure 6**) is diverse. As investor motivations and intentions behind the use of ESG metrics differ, the divergence in ESG scoring methodologies allows for additional choices and imposes the need to discuss each rating decisively. The empirical investigation of this study will focus on the Refinitiv ESG metric. Given its widespread usage within academic research and the similarity in the database, the weak appreciation of professional ESG data applicants creates a compelling case for a sorrow investigation.

As already discussed above, the market of ESG metric providers has grown substantially over the past decades. Simultaneously, the market goes through an ongoing phase of consolidation, with small and rather specialized ESG rating providers being incorporated into an elite group of large competitors. As one of the major players, MSCI uses its ESG research capabilities to market a wide range of ESG indices compiled with proprietary ESG metrics.[167] In 2010, MSCI acquired RiskMetrics. Along with it, MSCI acquired the three well-established ESG data providers and research agencies ISS Institutional Shareholder Services, Innovest Strategic Value Advisors, and KLD Research & Analytics Inc. Adding to the already acquired Barra (2004), Measurisk (2010), InvestorForce (2013), and GMI Ratings (2014) MSCI is now hosting the capabilities of year-long driving forces in the development of widely-used ESG metrics.[168] Another recent trend is traditional credit rating agencies entering the market of ESG data providers. As already shown, S&P Global acquired RobecoSAM in 2019, which included the proprietary Corporate Sustainability Assessment (CSA) informing the constituents of the Dow Jones Sustainability and S&P ESG indices.[169] S&P also houses Trucost ESG Analysis (since 2016), one of the major market players following a current trend of methodologies backed up by artificial intelligence.[170] Credit rating

[166] These and the following citations are taken from **Table 1**.
[167] https://www.msci.com/our-solutions/indexes/esg-indexes, accessed: 07/02/2022.
[168] cf. Escrig-Olmedo (2019, p. 4), cf. Wong/Petroy (2020, p. 34), cf. Eccles/Lee/Stroehle (2020, p. 582)
[169] cf. Wong/Petroy (2020, p. 6)
[170] cf. ibid. (p. 14)

agency Moody's entered the market in 2019 by acquiring the large European ESG data provider Vigeo Eiris.[171]

Since the introduction of KLD metrics as a measure of Environmental and Social performance in 1990, they have enjoyed a status of immense popularity within the academic research literature.[172] Another widely used ESG metric incorporating the additional Governance pillar are Refinitiv ESG performance scores (**Figure 2**).[173] They build upon the original ASSET4 database from 2003, which was later acquired by Thomson Reuters in 2009 and Refinitiv in 2018.[174] Unlike other ESG rating providers, Refinitiv offers access to the raw data underlying its ESG metrics, which is appreciated both among investment professionals and especially in the research community.[175]

The short review of the market of ESG performance metrics shows how proprietary approaches of single data providers induce a wide range of potential ESG performance metrics. Each of the scoring methodologies follows a distinct theorization of ESG performance and what information contributes to its accurate assessment. The next chapter, however, turns toward the inherent challenges in ESG performance measurement. It summarizes the major burdens that lie in the very nature of ESG performance, namely the definition of ESG, the importance to a firm and business model (i.e., materiality), and what ESG performance actually intends to measure (i.e., commitment vs. impact).

[171] cf. Wong/Petroy (2020, p. 7)
[172] cf. Eccles/Lee/Stroehle (2020, p. 584)
[173] cf. Berg/Fabisik/Sautner (2021, p. 33, 47)
[174] cf. Escrig-Olmedo et al (2019, p. 5),
[175] cf. Wong/Petroy (2020, p. 35), cf. De Villiers/Jia/Li (2022, p. 2)

Selected ESG data providers and central characteristics of their ESG scoring methodology

Data provider	Firms	Industries	Purpose	Focus	(primarily) based on	Data points	Granularity (number of metrics provided)	Scale
Carbon Disclosure Project (CDP)[176]	~13,000	16 (high-impact) sectors + remaining sectors as one	"CDP works with market forces to motivate companies to disclose their impacts on the environment and natural resources and take action to reduce negative impacts." CDP (2022, p. 5).	"The scoring methodology is a means to assess the responder's progress toward environmental stewardship as communicated through the company's CDP response." CDP (2022, p. 5).	questionnaires	130 questions + 202 sector-specific questions	1 level assessment	relative benchmark scoring / A-D- (6-scale)
MSCI[177]	~8,500	61 industries (GICS)	"MSCI ESG Ratings are designed to help investors understand ESG risks and opportunities and integrate these factors into their portfolio construction and management process." MSCI (2022, p. 3).	"… the intersection between a company's core business and the industry issues that can create significant risks and/or opportunities for a company." MSCI (2022, p. 3).	corporate disclosure	1,000	5 level assessment (300, 35, 10, 3, 1)	relative benchmark scoring / AAA-CCC (7-scale)
Refinitiv[178]	~11,800	54 industries (TRBC)	"We strive to be the trusted and preferred partner in the transition to sustainable finance and are committed to bringing to the market an array of best in-class data, analytics and workflow solutions, which allow customers to use Refinitiv data as the backbone of their investment processes." Refinitiv (2022a, p. 3).	"ESG scores from Refinitiv are designed to transparently and objectively measure a company's relative ESG performance, commitment and effectiveness, based on company-reported data." Refinitiv (2022a, p. 3).	corporate disclosure	630	3 level assessment (10, 3, 1)	relative benchmark scoring / 0-100 (percentile)
S&P / RobecoSAM[179]	~11,000	61 industries (GICS)	"S&P Global ESG Scores – and the CSA research process that underpins them – form the basis of a unique ecosystem that actively drives corporate disclosures and raises the bar on sustainability standards over time."[180]	"S&P Global ESG Scores ('ESG Scores') measure companies' exposure to and performance on key ESG risks and opportunities, the quality and completeness of their public disclosures, and their awareness of emerging but underreported ESG issues." S&P (2022, p. 3).	corporate disclosure, questionnaires	1,000 + industry-specific questionnaires	4 level assessment (130, 30, 3, 1)	relative benchmark scoring / 0-100
Sustainalytics[181]	~12,400	138 industries	"Sustainalytics' ESG Risk Ratings are designed to help investors identify and understand financially material ESG risks at the security and portfolio level and how they might affect the long-term performance for equity and fixed income investments." Sustainalytics (2020, p. 1).	"A material ESG issue (MEI) is the core building block of the ESG Risk Rating. For Sustainalytics, an ESG issue is material if it is likely to have a significant effect on the enterprise value of a typical company within a subindustry, and if the presence or absence of an [sic] MEI in financial reporting is likely to influence the decisions made by a reasonable investor." Sustainalytics (2020, p. 2).	corporate disclosure	1,300	4 level assessment aggregating the risk exposure as sum of unmanaged manageable risk and unmanageable risks	risk score / 0-40

Table 1 Selected ESG data providers and central characteristics of their ESG scoring methodology

[176] https://www.cdp.net/en/guidance/guidance-for-companies, accessed: 06/30/2022.
[177] https://www.msci.com/our-solutions/esg-investing/esg-ratings, accessed: 06/30/2022.
[178] https://www.refinitiv.com/en/sustainable-finance/esg-scores, accessed: 06/30/2022.
[179] https://www.spglobal.com/esg/solutions/data-intelligence-esg-scores, accessed: 06/30/2022.
[180] ibid.
[181] https://www.sustainalytics.com/esg-ratings, accessed: 06/30/2022.

3 CHALLENGES IN ESG PERFORMANCE MEASUREMENT

Measuring ESG performance intends to quantify how firms allocate resources and capital within their business processes and what externalities associated with these actions affect their environment.[182] All efforts to measure ESG performance rely on the idea of quantifying corporate behavior that can hardly be expressed in numeric terms. As a result, ESG scoring methodologies inherently face a variety of challenges.[183]

3.1 Defining ESG: Issues, factors, and risks

The 2004 UN Global Compact report *Who Cares Wins – Connecting Financial Markets to a Changing World* is widely considered the first publication that formalized the acronym "ESG".[184] The report summarizes the joint work of large, multi-national financial institutions led by the United Nations and seeks to describe "… guidelines and recommendations on how to better integrate environmental, social and corporate governance issues in asset management, securities brokerage services, and associated research functions."[185] The report builds upon the premise that ESG integration creates positive long-term value for business organizations and investors and invites the financial industry to adopt ESG issues as a major business rationale within their investment processes and financial analyses.[186] The report was especially motivated by a 2003 executive survey conducted by the WEF, which found severe differences in the understanding of topics related to ESG.[187] Due to the intangible nature of many such topics, the UN Global Compact states the necessity to find a common understanding of ESG issues.[188]

Based on the ten UN Global Compact Principles (**Appendix 2**) that the UN Secretary-General introduced the same year, the UN seeks to promote sustainable and socially responsible business conduct within the areas of human rights, labor, environment, and anti-corruption.[189] Building upon those broad goals, the UN Global Compact report describes three dimensions by that *ESG issues* are characterized. ESG issues: 1) Can pose distinct risks to the operation of a business, the society, or the environment, 2) address fields of action in which regulation falls short of advocating business' sustainability and social responsibility, and 3) allow to meet consumer trends and access

[182] cf. Kotsantonis/Serafeim (2019, p. 50)
[183] cf. Windolph (2011, p. 42), cf. Berg/Kölbel/Rigobon (2019, p. 3)
[184] cf. Billio (2020, p. 1427), cf. Lee et al (2021, p. 2)
[185] UN Global Compact (2004, p. i)
[186] cf. ibid. (pp. i-ii)
[187] cf. WEF (2003, p. 17)
[188] cf. UN Global Compact (2004, pp. 3-4)
[189] cf. ibid. (p. ix)

new market opportunities.[190] Within these broad boundaries, the report states several exemplary issues "impacting company and investment value".[191]

Following the initial definatory notion, the concept of ESG is clearly coined toward serving capital markets as a catchall term for non-financial corporate performance that might be relevant for investment purposes. As such, ESG acts more as a cognitive system than a clear-cut and predetermined definition of specific fields of action.[192] A recent report by the European Banking Authority (EBA) summarizes survey results on leading European financial institutions. The report lists roughly twenty international frameworks that are the basis for organization-specific approaches toward ESG definitions.[193] Following the results, the most common frameworks shaping ESG definitions are sustainable reporting guidelines (e.g., the Global Reporting Imitative, the Sustainable Accounting Standards Board, the International Integrated Reporting Council), investor initiatives (e.g., UN PRI), and various internationally accepted conventions and treaties introduced as multinational efforts to respond to global challenges (e.g., UN SDG).[194]

Given its close tie to capital markets, the notion of ESG becomes especially relevant when it is linked with real-world economic outcomes. In an effort to derive commonalities from various definitional perspectives, the EBA emphasizes the understanding of *ESG factors*. ESG factors consist of those ESG issues that either impact (outside-in perspective) or are impacted (inside-out perspective) by an organization and its actions.[195] Impacts associated with ESG issues could arise during all parts of an organization's value chain. They might be uncertain on behalf of their magnitude and their temporal occurrence (i.e., short-, medium-, or long-term).[196] However, prioritizing ESG issues based on their association with the prosperity of a firm translates ESG issues into economic outcomes relevant to capital markets. Simultaneously, ESG factors, as opposed to factors traditionally considered financial (e.g., costs, revenues, profits), are still characterized as non-financial.

An essential feature of ESG factors is their dynamic nature. Some ESG issues impose impacts that are subject to shifts driven by, e.g., market conditions, public pressure, or regulatory frameworks.[197] As the magnitude of ESG factors varies, tracking ESG performance becomes inherently

[190] cf. UN Global Compact (2004, p. 9)
[191] ibid. (p. 6)
[192] cf. Li et al. (2021, p. 1)
[193] cf. EBA (2021, pp. 23-25)
[194] cf. Coleton et al. (2020, p. 9)
[195] cf. EBA (2021, p. 29)
[196] cf. ibid. (pp. 28-29)
[197] cf. ibid. (pp. 28-29)

important in managing so-called *ESG risks*. The EBA defines ESG risks as "… the negative materialization of ESG factors".[198] Although the concept of ESG is ambiguous in its nature, the differentiation between ESG issues, factors, and risks stated above guides this study's further elaborations. To get a detailed understanding of what organizational behavior might be considered when referring to ESG issues, factors, and risks, the following chapters list exemplary topics of common use when referring to ESG performance.

3.1.1 Environment

The Oxford English Dictionary defines the environment as "… the natural world or physical surroundings in general, either as a whole or within a particular geographical area, esp. as affected by human activity."[199] Applied to ESG performance measurement, the environmental pillar, therefore, aims to reflect possible interactions between an organization and the natural world.[200] These interactions can be divided into four distinct dimensions of nature: air, water, soil, and habitat. For each of those dimensions, significant shifts in their conditions can be observed over the last three centuries.

In its most recent report (2021), the Intergovernmental Panel on Climate Change (IPCC) summarizes contemporary findings on the physical science basis of climate change published by leading scholars in a variety of related research fields.[201] Following the report, the global mean surface temperature within each decade since 1970 has successively been higher than in the preceding decades.[202] This pattern of global warming is accompanied by a globally observed retreat of glaciers and arctic ice, an increase in global mean sea level, an increase in upper ocean temperature, an increase in average global precipitation over land, and a poleward shift of climate zones.[203] Climate change and the observed increase in the chances of extreme weather events (e.g., heatwaves, droughts, and flooding) since the 1950s are affecting global ecosystems and every inhabited region of the world.[204]

[198] EBA (2021, p. 32)
[199] https://www.oed.com/view/Entry/63089?redirectedFrom=environment#eid, accessed: 01/25/2022.
[200] cf. Li et al. (2021, p. 2)
[201] https://www.ipcc.ch/report/sixth-assessment-report-working-group-i/, accessed: 03/18/2022.
[202] cf. IPCC (2021, p. 6)
[203] cf. ibid. (p. 7)
[204] cf. ibid (pp. 12-13)

The main driver of global warming is considered to be a severe rise in Greenhouse gas (GHG) concentrations caused by human activities.[205] The IPCC states: "Observed increases in well-mixed greenhouse gas (GHG) concentrations since around 1750 are unequivocally caused by human activities."[206] Thus, human activities are regarded to have significantly contributed to severe shifts in the conditions of the environment since the 18th century and especially during the last five to seven decades.[207] Climate change is considered the leading environmental issue posing socioeconomic and political challenges on every geographical scale. Thus, many concerns labeled environmental issues can be subsumed under the broad implications of climate change and its causes.[208] Widening the perspective, however, the general environmental issues concern the boundaries that define a safe, stable, and sustainable biosphere for societal well-being. These boundaries are referred to as planetary boundaries and pose immediate or future constraints to sustainable actions.[209]

Research on planetary boundaries suggests several integrated and interdependent environmental processes posing boundaries. Each of these boundaries defines a state of "Holocene-like conditions" that describe "a safe operating space" that allowed, for example, for agricultural advancement and the growth of complex societies during roughly the last 10,000 years.[210] Rockström et al. (2009) suggest a set of nine planetary boundaries, namely: climate change, ocean acidification, stratospheric ozone depletion, atmospheric aerosol loading, biochemical flows (interference with phosphor and nitrogen cycles), global freshwater use, land-system change, the rate of biodiversity loss (biosphere integrity), and chemical pollution (or novel entities).[211] A recent study argues that five of these boundaries have likely already been crossed (i.e., the ones concerning biochemical flows, land-system use, biosphere integrity, chemical pollution, and climate change).[212] The heterogeneous nature of issues underlying the boundaries, however, makes it a complex ongoing task to identify operational measures to cope with them.[213] Nevertheless, the planetary boundaries framework lays out the ecological foundation to categorize the most prevailing environmental issues and assess corporate impact.[214]

[205] cf. IPCC (2021, p. 6), GHG especially refer to carbon dioxide, methane, and nitrous oxide.
[206] ibid.
[207] cf. ibid (p. 8)
[208] cf. Dryzek (2021, p. 5)
[209] cf. Rockström et al. (2009, p. 1), cf. Whiteman/Walker/Perego (2013, p. 313), cf. Steffen et al. (2015, p. 1)
[210] Persson et al. (2022, p. 1510), Rockström et al. (2009, p. 3)
[211] cf. Rockström et al. (2009, pp. 9-10), cf. Steffen et al. (2015, p. 1)
[212] cf. Persson et al. (2022, p. 1510)
[213] cf. ibid. (p. 1517)
[214] cf. Whiteman et al. (2013, p. 307)

Following the comprehensive yet extensive elaboration of the EBA (2021), the most common environmental factors are found in the following fields:[215]

- GHG emissions
- Water usage and consumption
- Waste management and production
- Consumption/Production of fossil fuels
- Energy consumption
- Pollution
- Biodiversity (e.g., land degradation, desertification, soil sealing, deforestation)
- Environmentally-friendly innovation in products and services

3.1.2 Social

The social pillar of ESG is generally defined by factors that serve the well-being of society.[216] In a broad sense, this pillar is directly resulting from the debate over the social responsibility of the firm (see chapter 2.1). Thus, one could argue that within this broad understanding, the social pillar might not necessarily be limited to factors directly affecting humans but also captures the environmental challenges humanity faces.[217] To sharpen the understanding, however, the social pillar within the ESG concept refers to efforts that directly constitute the realm of actions corporations can take regarding their societal stakeholders.[218]

In his seminal work Freeman (1983) defines "Any ... group or individual who can affect the achievement of an organization's objectives or who is affected by the achievement of an organization's objectives…" as stakeholders.[219] Stakeholders are either directly (e.g., investors, suppliers, employees, customers) or indirectly (e.g., government, trade associations, communities) involved in a firm's actions. The stakeholder theory conceptualizes the understanding of the corporation based on three dimensions: Descriptive, instrumental, and normative considerations of interrelationships between stakeholders, the firm, and their distinct objectives.[220] Especially the normative character of stakeholder theory goes beyond the mere portrayal of potential managerial

[215] EBA (2021, pp. 27-28)
[216] cf. ibid. (p. 43)
[217] cf. ibid. (p. 43)
[218] cf. ibid. (p. 43)
[219] Freeman/Reed (1983, p. 91)
[220] cf. Donaldson/Preston (1995, pp. 69-70)

implications of such interrelationships. It combines theories on the purpose of the firm with philosophical and moral questions in order to provide management guidelines.[221] As such, it goes beyond the instrumental approach of asking *what could be done* and instead focuses on the question *what should be done*.[222] It is in its nature that especially managerial action trying to find answers to such normative expectations will face a myriad of potential pathways.[223]

Business decisions and their impact on stakeholders, however, are especially prone to be the subject of public debate and regulatory action. Certain structures and actors put external pressure on a firm (e.g., norms and regulations, the degree of institutionalization of CSR within political decision-making, NGOs, or the competitive environment). Simultaneously, there are factors that internally motivate the implementation of practices considered sound and socially compatible business conduct (e.g., corporate culture, strategy, management style).[224] Hence, the wide range of normative considerations in managerial decision-making is structured by regulatory frameworks, social policy, and public sentiment.[225] The social pillar follows this notion, which leads to its primary focus on interrelationships between the firm and its workforce, customers, and community.[226] Following the EBA (2021), the most common social factors are found in the following fields:[227]

- Labor and workforce considerations (e.g., freedom of association, workplace health and safety)
- Community impacts (e.g., poverty)
- Customer protection (e.g., data privacy, product safety)
- Human rights (e.g., forced and child labor, human trafficking)
- Inequality and discrimination (e.g., gender equality)

[221] cf. Donaldson/Preston (1995, p. 71)
[222] cf. ibid. (p. 72)
[223] cf. Phillips/Freeman (2010, p. xvi)
[224] cf. Brown/Vetterlein/Roemer-Mahler (2010, p. 6)
[225] cf. EBA (2021, pp. 45-46)
[226] cf. Freeman/Reed (1983, p. 91), cf. Matos (2020, p. 7)
[227] EBA (2021, pp. 27-28)

3.1.3 Corporate Governance

According to Goergen (2018), "Corporate Governance … deals with conflicts of interest, and their prevention or mitigation, between: … shareholders and the managers; … shareholders and the debtholders; … shareholders and the non-financial stakeholders; [and] different types of shareholders."[228] Conflicts of interest arise from two specific types of conditions, which are characteristic of publicly-held corporations: A wide range of principle-agent relationships and the existence of incomplete contracts to resolve the accruing agency costs.[229]

The relationship between shareholders (principal) and corporate executives (agent) is a prime example of an agency problem. As shareholders hold the residual claim on the firm's assets, their sole objective will likely be the firm's profitability.[230] The separation of ownership and control within publicly-held corporations, however, allows the manager to engage in actions that benefit other (potentially self-serving) objectives.[231] Any measure taken to align both interests (e.g., monitoring, contractual agreements, incentives) bears additional transaction costs (agency costs), which makes it generally impossible for the manager to act in the best interest of the shareholders.[232] Corporate Governance then refers to the institutions and mechanisms that help to limit agency costs in order to maximize the outcome of the principal (i.e., shareholder value).[233]

Given that shareholder value is the main motivation of Corporate Governance mechanisms, often referred to as shareholder primacy, there are numerous other parties that impose potential conflicts of interest and agency costs in real-world settings. Based on the cascading distribution of corporate cash flows to the suppliers of financing (i.e., shareholders after creditors), agency costs arise from the incentive to engage in riskier projects when external debt financing is at stake.[234] Further, since voting rights are coupled with the stake in the firm, large shareholders face the possibility to influence corporate action in their own interest at the expense of minority shareholders.[235] Although often argued, it also stands to debate whether the wealth of other non-financial stakeholders (e.g., employees, customers, community) is perfectly internalized in shareholder value based on contracts and legal obligations.[236]

[228] Goergen (2018, p. 5)
[229] cf. ibid. (p. 6)
[230] cf. Shleifer/Vishny (1997, p. 764), cf. Fama/Jensen (1983, p. 328)
[231] cf. Larcker/Tayan (2021, p. 4)
[232] cf. Jensen/Meckling (1976, pp. 308-309), cf. Denis (2001, p. 195)
[233] cf. Goergen (2018, p. 7), cf. Larcker/Tayan (2021, p. 8)
[234] cf. Jensen/Meckling (1976, p. 334)
[235] cf. Goergen (2018, pp. 11-12)
[236] cf. Tirole (2001, p. 4)

Especially the latter informs the very foundation of this study. It is the crucial idea of ESG performance metrics to quantify the extent to which a firm's management takes actions that are designed to internalize stakeholder interests. Framing Corporate Governance mechanisms as instruments to enforce shareholder primacy presumes that any action taken to internalize (other) stakeholder interests deprives of the shareholders' residual claim.[237] As one can observe over the last decades, however, (institutional) investors seem to be at the forefront of promoting the consideration of environmental and social issues.[238] The idea that those induce inherent risks that eventually affect shareholder value in the long run enforces an ambivalent nature to the conflicts of interests that Corporate Governance mechanisms are intended to resolve.[239]

Since a shareholder-centric Corporate Governance might ultimately be aligned with stakeholder interests, the idea of the Governance pillar within ESG is to evaluate the efficiency by which these interests are met. Hence, Corporate Governance especially evolves around the role of the management board. Larcker and Tayan (2021) define "… corporate governance as the collection of control mechanisms that an organization adopts to prevent or dissuade potentially self-interested managers from engaging in activities detrimental to the welfare of the shareholders and stakeholders."[240] According to the EBA (2021), the most common Corporate Governance factors belong to areas like:[241]

- Codes of conduct and rules to define rights, responsibilities, and expectations
- Accountability
- Corruption and Bribery
- Transparency and Disclosure
- Remuneration
- Shareholder rights
- Board diversity and structure

The next chapter turns toward the question which ESG issues are potentially relevant to a distinct firm and its prospects. As referred to on numerous occasions above, the chapter turns toward the concept of *materiality* and how it informs corporate disclosure as the prime source of ESG information for third-party ESG scoring methodologies (see chapter 2.2.3).

[237] cf. Tirole (2004, p. 4)
[238] cf. Matos (2020, pp. 56-57)
[239] cf. Larcker/Tayan (2021, p. 398)
[240] ibid. (2021, p. 8)
[241] EBA (2021, pp. 27-28)

3.2 Assessing ESG materiality

ESG materiality refers to the question of which ESG issues from the broad range of potentially relevant ESG issues are actually relevant to a specific company and its prospects. The notion of *materiality* is borrowed from financial accounting. In Accounting, materiality determines the information that needs to be reported within corporate disclosure. There are several definitions of materiality, depending on the reporting framework corporate disclosure follows. However, materiality definitions usually focus on two dimensions: the relevance and magnitude of the content of specific information. The Financial Accounting Standards Board (FASB), for example, states:

> "Relevance and materiality are defined by what influences or makes a difference to an investor or other decision maker; however, the two concepts can be distinguished from each other. Relevance is a general notion about what type of information is useful to investors. Materiality is entity specific. The omission or misstatement of an item in a financial report is material if, in light of surrounding circumstances, the magnitude of the item is such that it is probable that the judgment of a reasonable person relying upon the report would have been changed or influenced by the inclusion or correction of the item."[242] FASB (2018, p. 17).

In other words: While much information might be relevant in general, not all information is considered material to a specific firm. Materiality is based on the magnitude to which this information affects decision-making with respect to a particular entity. Transferring this notion to the reporting of non-financial information, there is undoubtedly a wide range of ESG issues that could be considered relevant. But if an ESG issue is material and needs to be reported depends on the magnitude to which this issue affects decision relevant measures of a firm. Therefore, the materiality of ESG issues usually differs and is heavily influenced by a firm's context.

Consider one of the world's leading food manufacturers Nestlé. Within its latest sustainability report, Nestlé states: "We conduct a formal materiality analysis every other year to ensure we prioritize the issues that have the biggest impact on the economy, society and the environment, and that matter the most to our stakeholders".[243] Nestlé focusses its materiality assessment on the two dimensions "Importance to stakeholders" and "Impact on Nestlé's success".[244] Following this idea Nestlé finds "climate and decarbonization", "responsible sourcing", and "human rights" being

[242] FASB (2018, p. 17)
[243] Nestlé (2021, p. 8)
[244] cf. ibid.

the most important issues to report on.[245] Deutsche Bank as well analyses "stakeholder expectations" and the "Impacts" of its business activities.[246] However, issues Nestlé identifies as material are, if at all, captured by the materiality assessment of the Deutsche Bank with relatively low importance. Deutsche Bank, for example, states efforts to fight "financial crime" or to ensure "information security" as the most critical ESG factors.[247]

These examples show a major peculiarity of ESG materiality assessment. While materiality in financial reporting focuses primarily on shareholders, the assessment of materiality in non-financial reporting concerns a broader range of stakeholders. It emphasizes the importance of stakeholder perceptions and the impact of specific ESG issues on the business.[248] Chapter 3.1 focuses on the various aspects that could be considered when assessing ESG performance. As ESG issues concern various interest groups and stakeholders, standardization of material aspects is a significant challenge in ESG performance measurement.[249]

[245] Nestlé (2021, p. 8)
[246] Deutsche Bank (2021, p. 21)
[247] cf. ibid. (p. 22)
[248] cf. Eccles et al. (2012, p. 66)
[249] cf. Windolph (2011, p. 42)

3.2.1 Materiality in CSR reporting frameworks

ESG ratings heavily rely upon publicly available information retrieved from corporate disclosure.[250] Within non-financial reporting, the notion of materiality is the key driver of information corporations make publicly available. Various reporting frameworks have evolved that guide corporations in determining ESG materiality. Hence, some kind of standardization in ESG ratings is induced via a mutual understanding of materiality assessment.[251]

One of the most popular frameworks for sustainability reporting is provided by the Global Reporting Initiative (GRI). The international non-profit organization published the first version of its Guidelines in 2000. The 2020 KPMG survey of sustainability reporting finds that GRI is the most widely used framework among large multinational corporations.[252] In 2017, GRI reported almost 11,000 organizations following the GRI standards.[253] As the examples above show (Nestlé and Deutsche Bank both follow GRI guidelines), the materiality assessment is based on two steps. First, organizations need to identify and assess the impact of their activities. The GRI understands *impact* as the "… effect an organization has or could have on the economy, environment, and people …, as a result of the organization's activities or business relationships."[254] Secondly, firms determine the most significant impacts by consulting relevant stakeholders and testing their assessment against sector standards that GRI provides. Following this process, material ESG topics to report on are determined regularly for each reporting cycle.[255] The resulting materiality assessment is often depicted graphically to reflect the impact on stakeholders (y-axis) and the prospects of the firm (x-axis).[256] **Figure 7** and **Figure 8** depict the *materiality matrix* of Nestlé and Deutsche Bank directly taken from the respective annual (sustainability) reports.

[250] cf. Windolph (2011, p. 43), see chapter 2.2.3 for a more detailed description of ESG scoring methodologies.
[251] cf. Taubken/Feld (2018, p. 87)
[252] cf. KPMG (2020, p. 25)
[253] cf. Taubken/Feld (2018, p. 88), GRI discontinued to list all organizations utilizing the GRI framework on its website (https://www.globalreporting.org/how-to-use-the-gri-standards/register-your-report/, accessed: 02/16/2022).
[254] GRI (2022, p. 11)
[255] cf. ibid. (2022, p. 103)
[256] cf. Taubken/Feld (2018, p. 88)

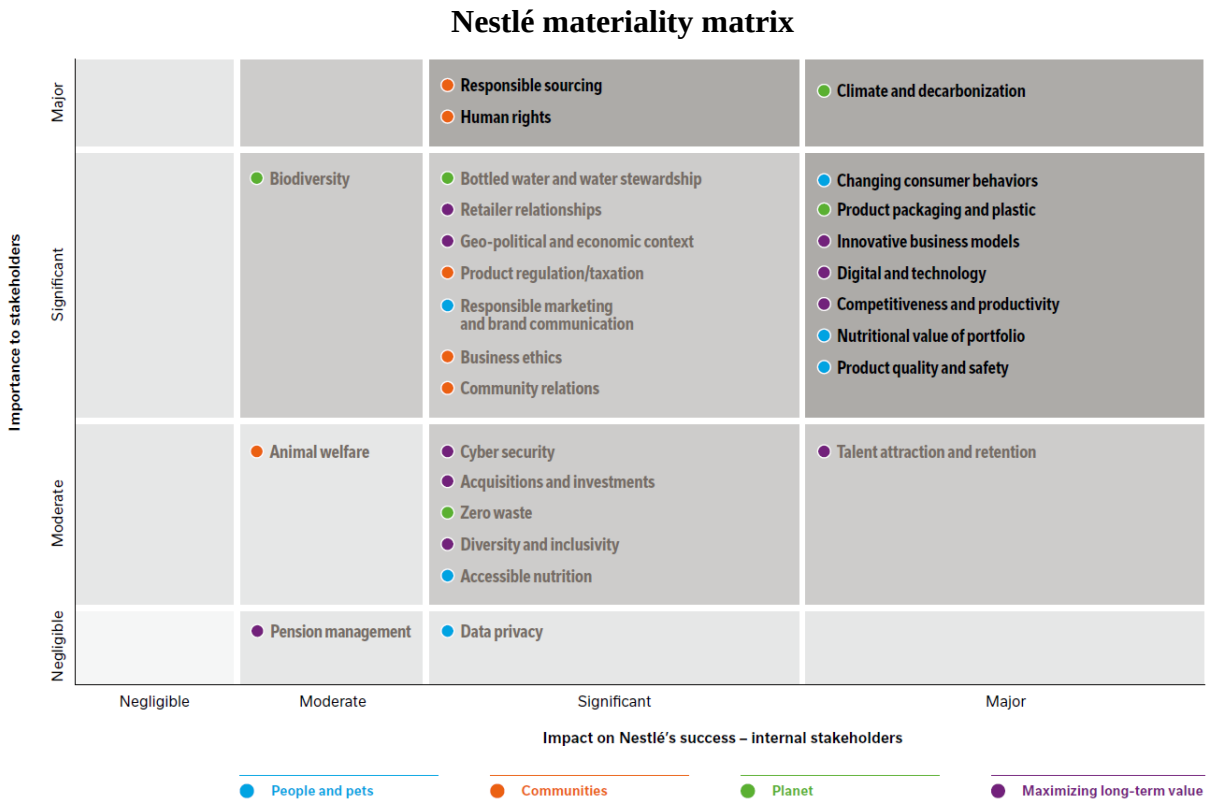


Figure 7 Nestlé materiality matrix, source: Nestlé (2021, p. 8).

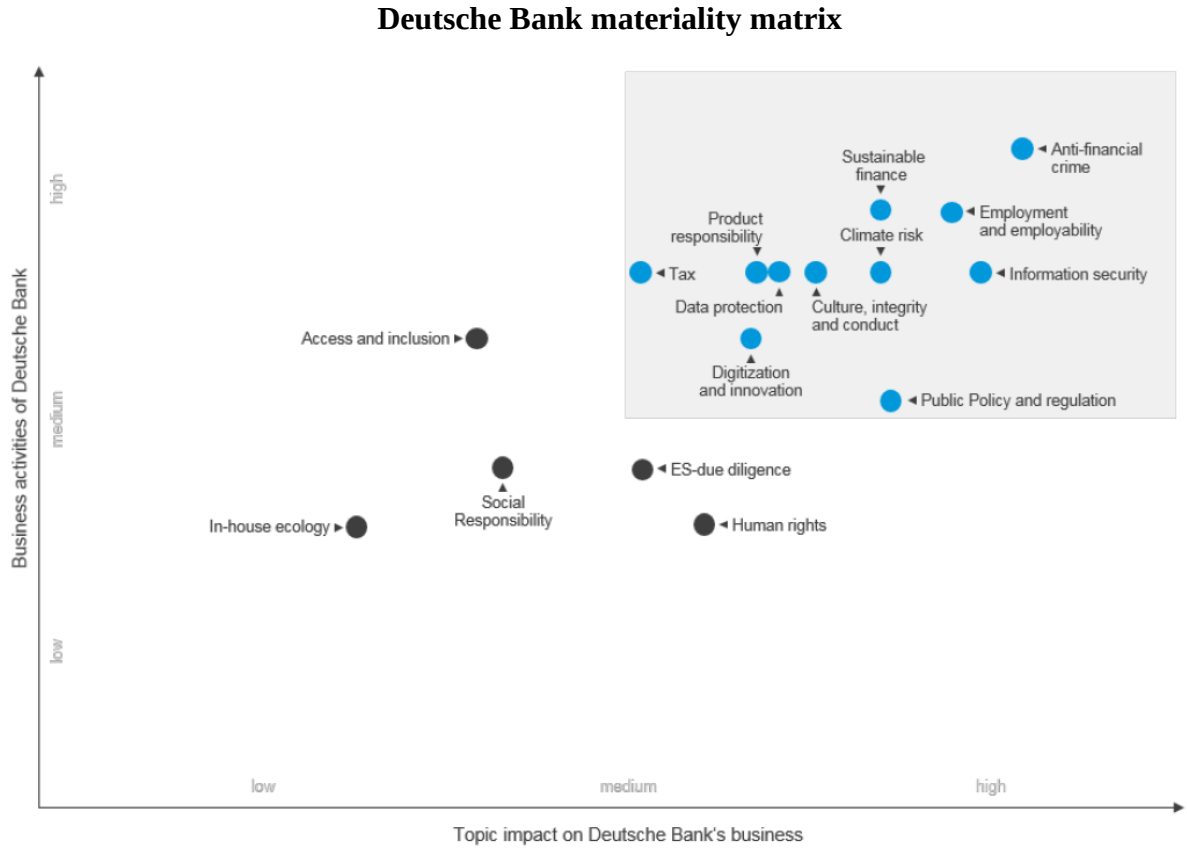


Figure 8 Deutsche Bank materiality matrix, source: Deutsche Bank (2021, p. 22).

Besides GRI, there are other major non-financial reporting initiatives laying out frameworks and a distinct understanding of ESG materiality.[257] One of which, the Sustainable Accounting Standards Board (SASB), focuses on the identification of material topics from a shareholder perspective.[258] Its main goal is to provide industry standards for corporations to disclose information on "…topics that are reasonably likely to have material impacts on the financial condition or operating performance of companies in an industry."[259]

SASB standards are specifically developed to fit SEC reporting requirements. Thus, the SASB standard-setting process involves an analysis of "… evidence of interest to a reasonable investor" (e.g., impact on revenues, costs, and risk profiles),[260] as well as the consultation of capital market participants (e.g., corporations and investors).[261] By doing so, issues that the SASB identifies as material potentially affect financial performance indicators that appeal to investors. Simultaneously, the SASB aims for issues that are, on the one hand, in control of corporations themselves, on the other hand, relevant to a whole industry rather than a single company.[262]

The SASB standards are designed to complement the materiality assessment of corporations (e.g., the one illustrated above following GRI guidelines).[263] In doing so, the SASB provides illustrations of 26 potentially material ESG topics and how they apply specifically to each of the 77 industries for which the SASB currently provides standards. Companies within an industry usually follow similar business models. As long as there are no significant differences in size or business strategies, it is reasonable to assume that issues being material to one company in an industry are also material to others within the same industry.[264]

An example of a so-called *materiality map* for nine industries within the sector of "Extractives & Minerals Processing" can be seen in **Figure 9**.[265] The materiality map shows that topics within the environmental pillar (e.g., greenhouse gas emissions, ecological impacts) show the highest potential to be material for companies within these industries (e.g., coal operations, oil and gas services).[266]

[257] cf. KPMG (2020, p. 8). The KPMG survey of Sustainability Reporting lists GRI, SASB, IIRC, CDSB, and CDP as institutions that provide the most frequently used guidelines on reporting.
[258] cf. Taubken/Feld (2018, p. 8)
[259] SASB (2017, p. 9)
[260] ibid. (p. 12)
[261] cf. ibid. (p. 15)
[262] cf. ibid. (p. 18)
[263] cf. ibid. (p. 9)
[264] cf. Eccles et al. (2012, p. 70)
[265] https://www.sasb.org/standards/materiality-map/, accessed: 02/17/2022.
[266] Grey cells indicate topics likely to be material.

SASB Materiality map: "Extractives & Minerals Processing"

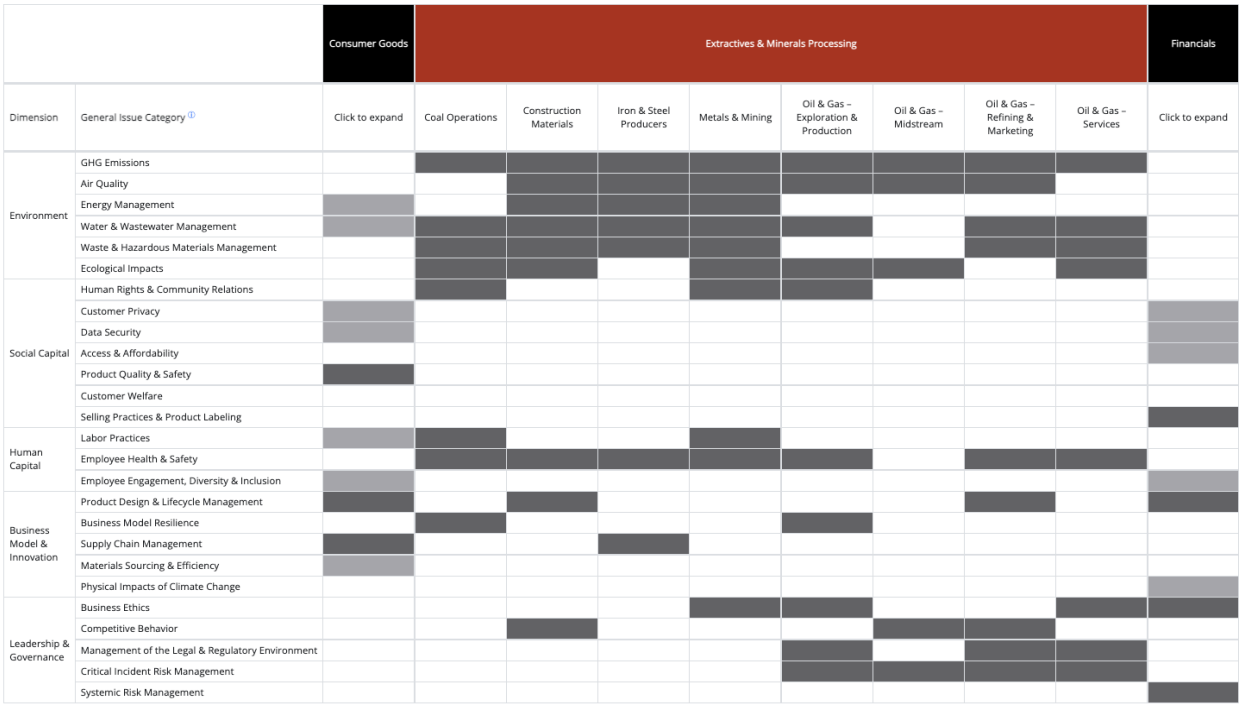


Figure 9 SASB Materiality map: "Extractives & Minerals Processing", source: https://www.sasb.org/standards/materiality-map/, accessed: 02/15/2022.

GRI and SASB help corporations to report on ESG issues. The reporting on non-financial performance, however, is often done within separate sustainability report. The International Integrated Reporting Council (IIRC) was established to promote an international framework to integrate sustainability reporting and financial reporting. The IIRC is a joint initiative of international policy-makers, investors, corporations, accounting standards organizations, and NGOs founded in 2010.[267] Within the integrated reporting (IR) framework, materiality is determined by issues that "… substantively affect the organization's ability to create value in the short, medium or long term."[268] Thus, IR relies upon the notion of value creation, which results from the specific business model a company chooses. Value is entity-specific and enables firms to identify material information on their own. However, the underlying premise is to assume that value consists of both financial and non-financial information that affect one another.[269]

The reporting frameworks described above share the same idea of guiding reporting on ESG issues. Differences primarily result from the perspective from which materiality is assessed. While SASB and IIRC focus on financial materiality from a shareholder perspective, the GRI follows a rather broad stakeholder approach. The commonalities, however, result in the compatibility of

[267] cf. Taubken/Feld (2018, p. 88), https://www.integratedreporting.org/10-years/10-years-summary/, accessed: 02/17/2022.
[268] IR (2015, p. 4)
[269] cf. ibid. (p. 7)

various standards. Consequently, the range of initiatives that promote sustainability reporting currently undergoes a consolidation process.

In June 2021, the IIRC and the SASB announced their merger and founded the Value Reporting Foundation (VRF). The VRF aims to provide a "… comprehensive suite of tools to assess, manage and communicate value."[270] Later that year, in November 2021, the major institution to set international accounting standards, the International Financial Reporting Standards Foundation, announced the development of "… a comprehensive global baseline of high-quality sustainability disclosure standards". The newly founded International Sustainability Standards Board (ISSB) consists of the VRF and another central investor-focused sustainability disclosure organization, the Climate Disclosure Standards Board (CDSB).[271] The GRI welcomed these steps and offered to collaborate in the standard-setting process. However, Eelco van der Enden, CEO of the GRI, pointed out, "It is important to recognize that, for reporting to illuminate the full range of impacts that companies have on society and the planet, it needs to take account of a multitude of stakeholders."[272]

ESG performance measurement relies upon the information companies disclose regarding their conduct on non-financial issues. The ongoing process of finding general standards for sustainability reporting makes materiality a primary driver of ESG performance measurement. As reporting is either focusing on broader stakeholder relationships (impact-materiality) or primarily focusing effects of certain ESG-related conduct on financial performance (financial-materiality), ESG ratings need to find a way to make these considerations comparable.[273]

[270] https://www.integratedreporting.org/news/iirc-and-sasb-form-the-value-reporting-foundation-providing-comprehensive-suite-of-tools-to-assess-manage-and-communicate-value/, accessed: 02/18/2022.
[271] https://www.ifrs.org/news-and-events/news/2021/11/ifrs-foundation-announces-issb-consolidation-with-cdsb-vrf-publication-of-prototypes/, accessed: 02/18/2022.
[272] https://www.globalreporting.org/about-gri/news-center/gri-looks-forward-to-working-with-new-issb-chair/, accessed: 02/18/2022.
[273] cf. Walter (2020, pp. 318-319)

3.2.2 Dynamic materiality

Another major aspect when assessing ESG materiality is its potentially time-variant nature. Literature on ESG materiality shows that the relevancy and urgency of ESG issues tend to differ over time. Rogers and Serafeim (2019) develop a model which seeks to identify crucial steps in ESG issues becoming financially material. They introduce a five-step model (**Figure 10**), depicting the "pathway" on which single ESG issues become material.[274]

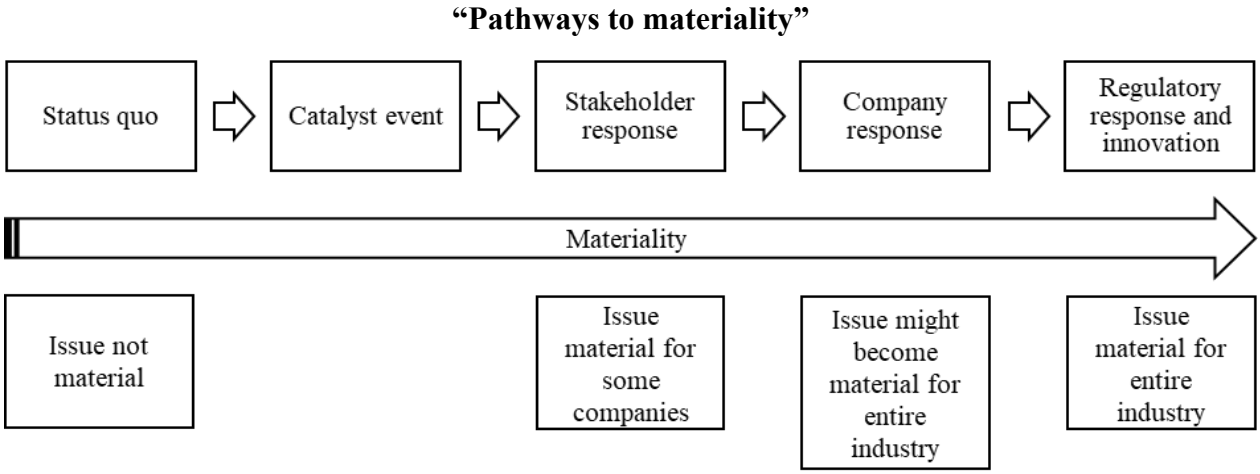


Figure 10 "Pathways to materiality", adapted from: Rogers/Serafeim (2019, p. 24).

According to Rogers and Serafeim (2019), the key driver of issues becoming material is the "misalignment" of corporate behavior with societal interests. In some cases (status quo), misalignment might be seen as acceptable externalities and part of the actual business model of a firm (e.g., adverse health effects of the consumption of tobacco products).[275] The status quo is eventually judged differently after a company, an industry, or the general public experiences a "catalyst" event.[276] Rogers and Serafeim point out the example of JUUL, the leading e-cigarette brand in the United States.[277] During the 2010s, the consumption of e-cigarettes among younger age groups in the U.S. grew extensively.[278] Studies showed that teenagers were 16-times more likely to consume JUUL products than older age groups.[279] JUUL's marketing practices were publicly scrutinized to specifically target young adults, while simultaneously not opposing the illegal sales to young people effectively enough.[280]

[274] Rogers/Serafeim (2019, p. 7)
[275] cf. ibid. (p. 8)
[276] cf. ibid. (p. 9)
[277] ibid. (p. 24)
[278] https://www.cdc.gov/tobacco/data_statistics/fact_sheets/youth_data/tobacco_use/index.htm, accessed: 02/15/2022.
[279] https://truthinitiative.org/press/press-release/new-study-reveals-teens-16-times-more-likely-use-juul-older-age-groups, accessed: 02/15/2022.
[280] https://www.fda.gov/news-events/press-announcements/statement-fda-commissioner-scott-gottlieb-md-new-steps-address-epidemic-youth-e-cigarette-use, accessed: 02/15/2022.

With growing recognition of the negative public health effects of e-cigarettes on younger adults, the misalignment of societal interest (i.e., good public health) and the potential impact of JUUL's marketing practices became more apparent. Consequently, the misalignment sparked a "stakeholder response". NGOs, media, and other stakeholders actively pushed the company to address the issue.[281] Most prominently, the U.S. Food and Drug Administration (FDA) labeled the use of e-cigarettes among young adults an "epidemic" and focused regulatory attention on JUUL's sales practices.[282]

With growing stakeholder pressure, the company is pushed to respond and tries to address the now material issue (company response). In some cases, misalignment has become so apparent that it affects an entire industry, e.g., if the business actions in question are considered industry standards.[283] In the case of e-cigarettes, the FDA sent letters to various other manufacturers and requested them to introduce measures that limit the appeal of their products to young age groups.[284] JUUL publicly apologized, deleted marketing on social media, and introduced retail regulations that limited the purchase of certain product lines that were explicitly attracting younger people to the age of 21 years.[285]

To further shrink the misalignment between corporate action and societal interests, Rogers and Serafeim (2019) identify two possible pathways. When stakeholder pressure is large enough to ignite political action, new regulations are enforced.[286] In other cases, firms might introduce innovative ways to address the misalignment and disrupt processes and business activities of their industry. When innovation comes with a competitive advantage, this even strengthens the materiality of the business conduct in question.[287] Simultaneously, the introduction of regulations can impose stronger materiality of certain issues. However, if regulations are loosened, penalties for misalignment (e.g., carbon-emission pricing) become less effective, making the materiality imposed by regulations weaker.[288] In the case of JUUL, regulations that put the legal purchasing age to 21 years old were enforced, and in some places, the sale of e-cigarettes was banned.[289]

[281] cf. Rogers/Serafeim (2019, pp. 14 f.)
[282] https://www.fda.gov/news-events/press-announcements/statement-fda-commissioner-scott-gottlieb-md-new-steps-address-epidemic-youth-e-cigarette-use, accessed: 02/15/2022.
[283] cf. Rogers/Serafeim (2019, p. 18)
[284] https://www.fda.gov/news-events/press-announcements/statement-fda-commissioner-scott-gottlieb-md-new-steps-address-epidemic-youth-e-cigarette-use, accessed: 02/15/2022.
[285] cf. Rogers/Serafeim (2019, p. 24), https://www.juullabs.com/statement-from-kevin-burns-juul-labs-chief-executive-officer-regarding-recent-fda-inspection/, accessed: 02/15/2022.
[286] cf. Rogers/Serafeim (2019, p. 19)
[287] cf. ibid. (p. 21)
[288] cf. ibid. (p. 20)
[289] cf. ibid. (p. 25)

Rogers and Serafeim (2019) conclude that materiality is evolving over time and resulting from a constant reevaluation of misalignment between corporate business conduct and societal interests. Materiality is particularly affected by the amount of information on the misalignment, as well as the power of media, NGOs, and politicians to drive changes and the effective enforcement of regulations.[290] Thus, materiality is driven by a complex and sometimes contradictory set of perceptions against which misalignment might be judged. On the one hand, firms are benchmarked against a traditional set of metrics judging their financial competitiveness. This focus is primarily the case when materiality is assessed to find the ESG issues that are particularly relevant for shareholders (financial-materiality). On the other hand, a particular set of conduct standards (impact-materiality) that might differ among stakeholder groups, jurisdictions, countries, industries, and ultimately even over time creates tension: Firms either follow a rather unrestrained pathway and risking regulatory and reputational pressure, or impose a level of ESG standards that might deter shareholder value.[291] As pointed out in the previous chapters (especially chapter 2.2.3), the managerial obligation arises to decide whether to compromise and to identify synergies between a rather financially or impact-driven business conduct. The next chapter shows how such synergies contribute to the concept of double materiality.

3.2.3 Double materiality

Currently, the European Commission is trying to bring the concepts of financial- and impact-materiality closer together. As central measures within its "action plan on financing sustainable growth", the EC introduced new sustainable disclosure regulations in 2019 and a detailed taxonomy to identify sustainable business activities in 2020. The action plan aims for "Reorienting capital flows toward a more sustainable economy", "Mainstreaming sustainability into risk management", and "Fostering transparency and long-termism".[292]

Within the *Guidelines on non-financial reporting* (2019), the EC introduced the concept of "double materiality" (**Figure 11**). It states, "… a company is required to disclose information … to the extent that such information is necessary for an understanding of the company's development, performance, position and impact of its activities."[293] The information that relates to the "development", "performance", and "position" of the firm puts the focus upon the financial-materiality. The additional need to highlight the "impact of its activities", however, formalizes the need to

[290] cf. Rogers/Serafeim (2019, pp. 23-26)
[291] cf. Walter (2020, p. 32)
[292] https://ec.europa.eu/info/publications/sustainable-finance-renewed-strategy_en#action-plan, accessed: 02/21/2022.
[293] European Commission (2019, p. 6)

assess materiality from two perspectives.[294] The two perspectives are often referred to as either an "outside-in" perspective, which examines whether current or expected circumstances related to certain ESG topics influence the business of the respective firm; Or an "inside-out" view, which relates to the question how a firm's business activities related to outcomes on particular ESG issues.[295]

The concept of "double materiality"

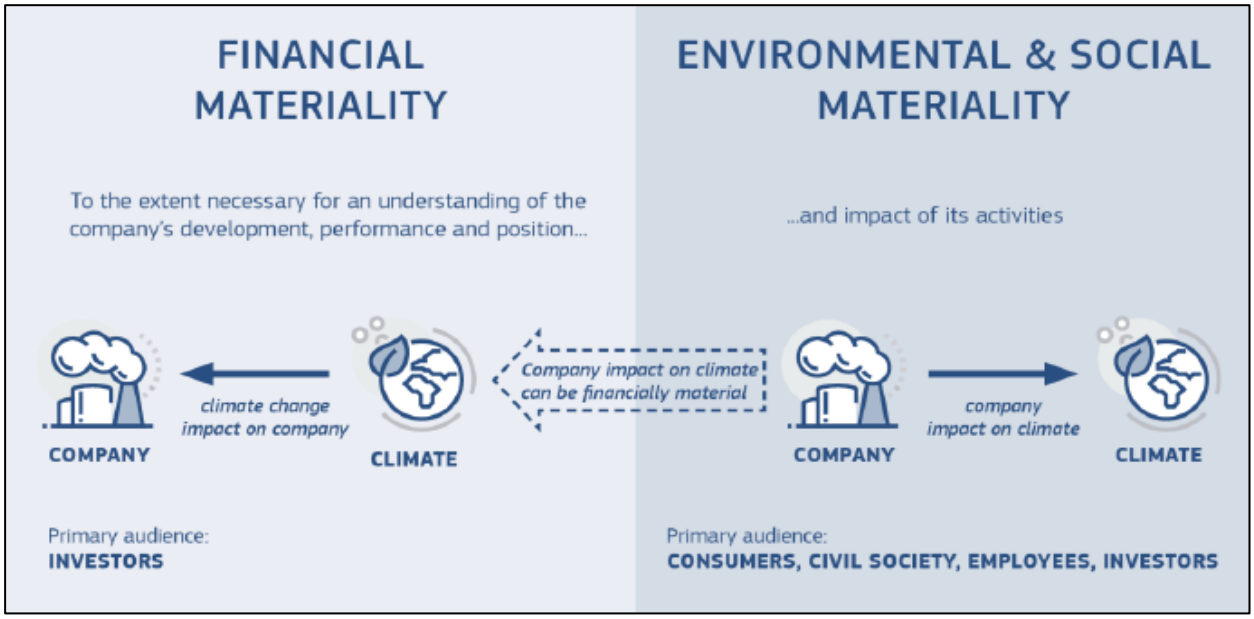


Figure 11 The concept of "double materiality", source: European Commission (2019, p. 7).

In order to implement the understanding of double materiality within capital markets, the EP introduced sustainability-related disclosure regulations for financial intermediaries (SDR). The regulations are designed to "… mobilize capital by the financial sector …" and urge financial market participants and advisers "… to disclose specific information regarding their approaches to the integration of sustainability risks and the consideration of adverse sustainability impacts."[296] The term "sustainability risks" refers to ESG issues that bear the potential to affect investment performance negatively. The additional analysis of adverse sustainability impacts requires reporting on negative effects on sustainability objectives and making these transparent on the level of single financial products.[297] Thereby, the SDR tries to facilitate a transparent and objective way of labeling financial products and their relationship with sustainability objectives.[298]

[294] European Commission (2019, pp. 6-7), cf. Adams et al. (2021, p. 5)
[295] cf. Taubken/Feld (2018, p. 92)
[296] SFDR (2019, 8)
[297] cf. Chiu (2022, pp. 6-7)
[298] cf. ibid. (p. 8)

In order to standardize the understanding of which corporate actions contribute to sustainable objectives, the EU's first attempt on focusing environmental effects was introduced within the EU Taxonomy Climate Delegated Act (as already reflected on in chapter 2.1).[299] The EU taxonomy tries to enable financial market participants and companies to identify those business activities that are contributing environmental sustainability.[300] "Climate change mitigation", "Climate change adaptation", "The sustainable use and protection of water and marine resources", "The transition to a circular economy", "Pollution prevention and control", and "The protection and restoration of biodiversity and ecosystems" are the six goals on environmental sustainability the EU declares.[301] To label a certain percentage of a firm's business activities (measured, e.g., in capital expenditures or revenues) as "EU Taxonomy-aligned", a company needs to contribute to at least one of the above-stated goals (e.g., via reforestation programs). Additionally, the assessment of environmentally sustainable business conduct follows the principle of "Do no significant harm", which requires any of the other business activities of the firm to be at least neutral to the remaining sustainability goals. In the sense of double materiality, the EU taxonomy is designed to 1) help companies disclose such information that follows the guidelines on non-financial reporting and 2) allow financial intermediaries to label certain products aligned with the EU taxonomy (i.e., to meet SDR requirements). [302]

The EU taxonomy marks a first step to standardizing performance measurement on a particular ESG issue, namely CO_2 reduction. Recent studies find that the taxonomy is able to do this exceptionally well in terms of its orientation on scientific evidence. Thereby, the taxonomy is a first step in helping scientific findings contribute to transition and climate gas reduction.[303] Additionally, the taxonomy and its classification guidelines are able to capture the majority of business activities that contribute to at least 80 percent of the EU's greenhouse gas emissions.[304] However, the taxonomy does not necessarily promote transition in the long run. The DIW finds that there is no consideration of business activities where there are already existing technological alternatives. Further, the classifications are based on objectives that do not necessarily foster climate neutrality by 2050 (as intended by the EC), which could be specifically relevant in capital-intensive industries (e.g., basic materials) with long-term investment cycles.[305] One major concern in this regard

[299] https://ec.europa.eu/commission/presscorner/detail/en/ip_21_1804, accessed: 02/22/2022.
[300] cf. Lucarelli et al (2020, p. 2)
[301] https://ec.europa.eu/info/business-economy-euro/banking-and-finance/sustainable-finance/eu-taxonomy-sustainable-activities_en, accessed: 02/24/2022.
[302] cf. Lucarelli et al. (2020, p. 4), cf. Schütze et al. (2020, p. 486)
[303] cf. Lucarelli et al. (2020, p. 16)
[304] cf. Schütze et al. (2020, p. 487)
[305] cf. ibid. (p. 492)

is the assessment of nuclear energy. At the beginning of 2022, the EC acknowledged some specific business activities in nuclear and gas energy production as being aligned with the taxonomy.[306] However, the decision did not meet the expectations of every EU member state alike. Especially in Germany, which decided to abolish nuclear energy production by the end of 2022 because of uncertain long-term effects on the environment and potential operating risks, the decision to label nuclear power as sustainable caused disapproval by the federal government.[307]

3.2.4 Materiality in ESG performance metrics

The short revision of ESG materiality shows that measuring ESG performance needs to take an entity-specific perspective. While some aspects might be relevant financially, others might be especially important due to their impact on the environment or the social sphere. However, the notion of materiality in non-financial reporting frameworks tends to guide reporting to the extent that it remains in the realm of the company itself to label certain ESG information as material and to report on it. Not surprisingly, ESG performance assessment is often criticized for the credibility of its data sources (i.e., mainly corporate disclosure) and a potential lack of material information (e.g., if companies do not dispose themselves to controversy).[308] Recent literature demonstrates that the ESG metrics of different providers vary systematically, especially in the manner they measure certain ESG aspects. The divergence makes a common and standardized understanding of what is material when assessing ESG performance one of the main drivers in ESG rating divergence.[309]

Ioannou and Serafeim (2019) find that ESG-related corporate actions have converged over the last decades. They show that the imitation of sustainable corporate action practices is mostly driven by industry-specific materiality considerations. Within industries, there is a tendency of companies to especially imitate those actions that are less unique and regulatory uncertain.[310] Regulatory uncertainty tends to differ systematically for corporations from different regions of the world. Liang and Renneboog (2017) show that common law traditionally overweight shareholder primacy and a rather market-oriented control of externalities. CSR practices are supported by civil law regimes, which can be explained by lower shareholder litigation risk, stronger labor regulations, and a

[306] https://ec.europa.eu/info/business-economy-euro/banking-and-finance/sustainable-finance/eu-taxonomy-sustainable-activities_en, accessed: 02/24/2022.
[307] https://www.reuters.com/business/sustainable-business/germany-cries-foul-over-nuclear-energy-eus-green-investment-rule-book-2022-01-22/, accessed: 02/24/2022.
[308] cf. Windolph (2011, p. 47)
[309] cf. Berg/Kölbel/Rigobon (2019, p. 31)
[310] cf. Ioannou/Serafeim (2019, p. 33)

higher degree of state involvement.[311] Corporate actions are increasingly imitated among industry peers but might be judged differently and incentivized on various levels depending on the respective legal background. Therefore, ESG metrics also need to look at materiality on a regional level.

One of the major aspects of ESG materiality, however, remains its objective assessment. As much of the research on financial materiality and the impacts on stakeholders is up to corporations themselves, the actual impacts are hard to identify. Further, while the information used to assess ESG performance is based on data deemed material and hence being reported, the actual impact of business activities might not even be measurable in all cases. The next chapter deals with the challenges arising thereby.

3.3 Assessing impact

ESG performance measurement is historically rooted in the literature on Corporate Social Performance (CSP), which primarily focuses on the social pillar of ESG. Despite the narrow focus, defining CSP gives important implications for challenges in measuring ESG performance. In a frequently-cited paper, Donna J. Wood (1991) surveys the literature on the measurement and understanding of CSP. She defines CSP as "… a business organization's confirmation of principles of social responsibility, processes of social responsiveness, and policies, programs, and observable outcomes as they relate to the firm's societal relationships".[312] In other words, CSP measures the degree to which the acknowledgment of the firm's social responsibility drives its actions and how the firm implements policies, programs, and processes (inputs) that generate intended outcomes that address this responsibility.[313]

3.3.1 Logic models

The idea of inputs generating outcomes is characteristic of so-called "logic models". Logic models are used to systematically evaluate the performance and effects of activities (e.g., programs or interventions) intended to serve a specific goal.[314] As such, logic models are a process- and theory-driven approach to assess how well this goal is actually met, which channels contribute to its achievement, and what externalities need to be considered or might be imposed.[315] Logic models are frequently used by governmental, non-governmental, or private organizations and within economic, social, or political contexts.[316] The assessment is done ex-ante or ex-post in a matter of

[311] cf. Liang/Renneboog (2017, p. 896)
[312] Wood (1991, p. 693)
[313] cf. Grewal/Serafeim (2020, p. 78)
[314] cf. Newcomer/Hatry/Wholey (2004, pp. xxxiii-xxxiv)
[315] cf. McLaughlin/Jordan (2004, pp. 11-12)
[316] cf. Anderson et al. (2011, p. 34)

planning, anticipating, and evaluating the results of a program or activity.[317] Thus, they are intended to contribute to decision-making in operating programs and assess whether, e.g., funds and resources provided are used effectively.[318]

Logic models are able to map the theoretical underpinnings of activities graphically. When utilizing logic models to analyze certain subjects, they are typically implemented as a flowchart following certain dimensions.[319] We refer to these graphic implementations as an "impact value chain" (**Figure 12**).[320]

Impact value chain

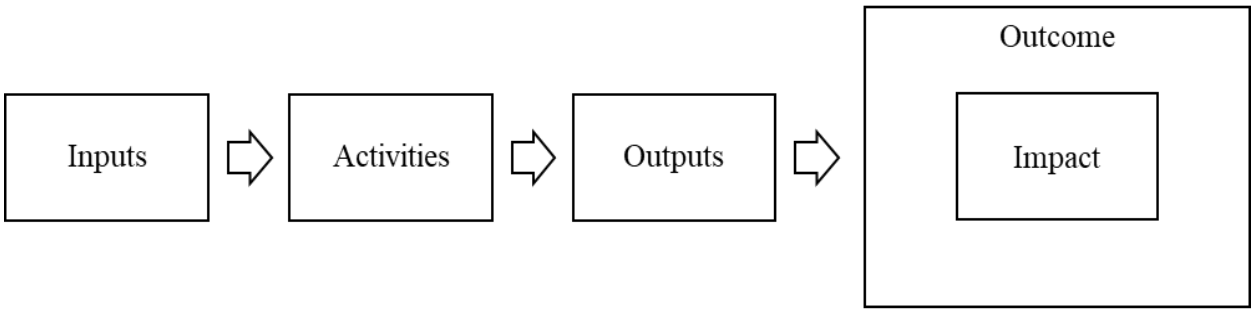


Figure 12 Impact value chain, adapted from: Clark et al. (2004, p. 7).

Within logic models, *inputs* motivate certain activities. Such inputs might incorporate human and financial resources deployed or needed to fulfill the intended activities. The activities result in *outputs*. Outputs might be products or services directly targeted to a group of beneficiaries. As their main characteristic, outputs are immediately observable and usually measurable if suitable indicators are defined. However, outputs do not necessarily be the main goals of a program. Programs are usually intended to serve goals that are indirectly caused by the actions within or the outputs of the program.[321]

These causal relationships are laid out when determining the *outcomes* of the program. Outcomes are the "… changes or benefits resulting from activities and outputs."[322] Such outcomes might be assessed over specific time horizons (i.e., short- and long-term) or dimensions of potentially affected individuals. Most importantly, outcomes differ from outputs as they are regularly subject to a discussion of causal relationships and need to be analyzed argumentatively.[323] Differentiating between outcomes and what would have happened without the actions of the program leads to an

[317] cf. Khandker/Koolwal/Samad (2009, pp. 7-8)
[318] cf. Newcomer/Hatry/Wholey (2004, p. xxxiii), cf. Anderson et al. (2011, p. 33)
[319] cf. Newcomer/Hatry/Wholey (2004, p. xxxv)
[320] Clark et al. (2004, p. 7)
[321] cf. McLaughlin/Jordan (2004, pp. 9-10)
[322] ibid. (p. 9)
[323] cf. ibid. (p. 9)

assessment of *impact*.[324] In an experimental sense, impact can be understood as the counterfactual of being treated by the program.[325] Most importantly, impacts are linked to the effects of the outputs on the beneficiary of the activities. Outputs and outcomes are rather associated with the one organization undertaking the action.[326]

Logic models help to structure and evaluate complex causal relationships of certain activities. Consequently, there is a broad range of literature on various topics relying on logic model depictions and analysis.[327] Development studies literature, for example, during the 1960s, introduced logic models to assess the impact of programs and interventions of the United States Agency for International Development.[328] Within the context of this thesis, CSR literature focused on the monetary and non-monetary benefits of CSR activities for in- and outside stakeholders.[329] Management and accounting literature regularly relies upon impact assessment on behalf of economic, environmental, or social dimensions of corporate actions.[330]

Considering the case of ESG performance metrics, the logic model implies critical steps in the development of such measures. ESG performance is complex and intends to depict various fields of corporate action with positive, negative, or interdependent relationships.[331] Different activities considered ESG-relevant might be associated with a diverse set of underlying assumptions and resources. Individual logic models could be applied to a broad range of potential ESG-related activities. Hence, a major challenge in ESG measurement is the question whether ESG metrics reflect different layers within the impact value chain and to what extent the resulting metrics rely upon presumptions of causality or actual evaluations of the achievement of intended goals.[332]

3.3.2 Impact assessment in CSR literature

Anderson et al. (2011) point out the potential to guide systematic literature reviews, as logic models help to systematize the central questions, causal links, and determinants subject to a specific field of research.[333] Following this approach, Barnett, Henriques, and Husted (2020) investigate almost six thousand articles dealing with CSR performance over the last five decades.[334] They

[324] cf. Ebrahim/Rangan (2014, p. 120), cf. Clark et al. (2004, p. 7)
[325] cf. Barnett/Henriques/Husted (2020, p. 951)
[326] cf. Kolodinsky/Stewart/Bullard (2006, p. 34)
[327] cf. Rogers (2008, p. 30)
[328] cf. Martinez/Cooper (2020, p. 1239), cf. Rogers (2008, pp. 30-31)
[329] cf. Weber (2008, pp. 250-251)
[330] cf. Maas/Liket (2011, pp. 171-172), cf. Kolodinsky/Stewart/Bullard (2006, p. 28), cf. Clark et al. (2004, p. 7), cf. Crace/Gehman (2022, p. 1)
[331] cf. Crace/Gehman (2022, p. 5)
[332] cf. Howard-Grenville (2021, p. 1), cf. Crace/Gehman (2022, pp. 5-6)
[333] cf. Anderson et al. (2011, p. 33)
[334] cf. Barnett/Henriques/Husted (2020, p. 938)

build upon the premise that CSR literature addresses the question whether CSR activities lead to improvements in societal and environmental ends. By introducing a logic model (**Figure 13**), they are able to categorize historical publications (i.e., the most frequently cited papers) within four levels of a CSR-themed value creation chain.[335]

Logic model on CSR literature

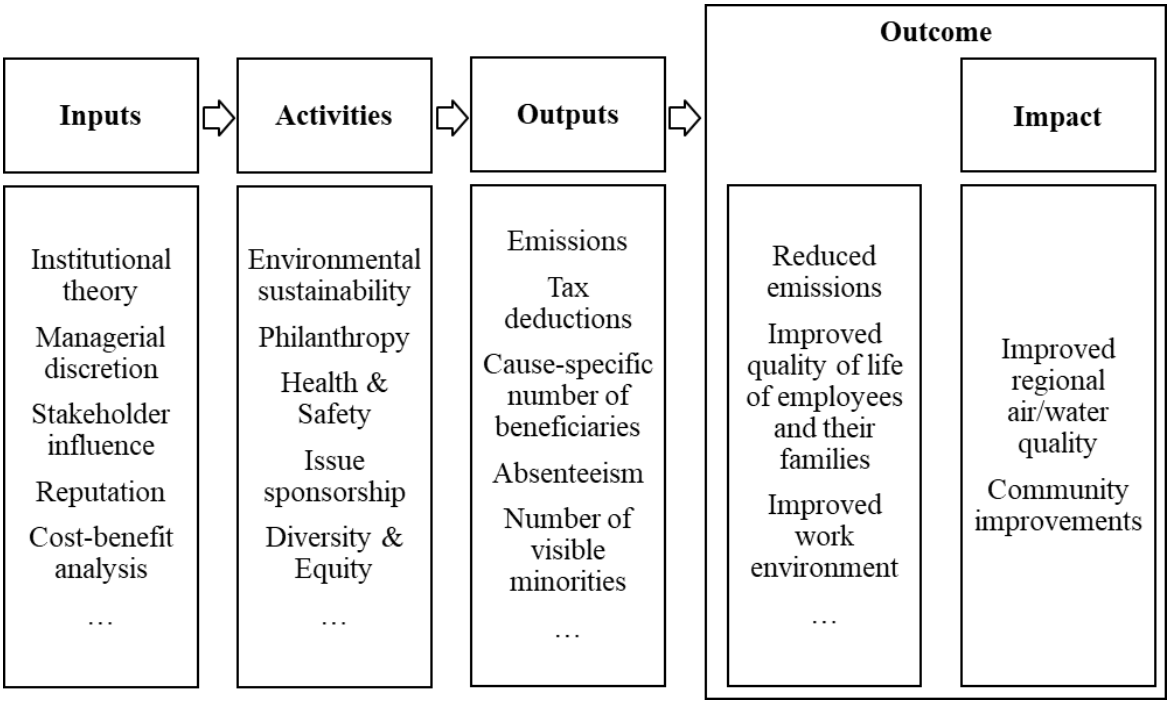


Figure 13 A CSR logic model, adapted from: Barnett/Henriques/Husted (2020, p. 948).

Studies focusing on *inputs* try to establish a link between CSR activities and what drives those activities (i.e., CSR is investigated as a dependent variable).[336] Publications, for example, show that the CEO's political association is driving CSR performance, as firms with CEOs associated liberal (as opposed to conservative CEOs) tend to have higher ESG performance.[337] Similar firm-specific examples include literature on the effect of gender diversity on corporate boards and how it affects ESG performance measures positively.[338] Further, literature seeks to establish links between broader and systematic factors outside the firm itself that explain CSR performance. Common law countries (e.g., UK or USA) traditionally emphasize shareholder primacy and a market-driven approach toward control over social factors. Liang and Renneboog (2017) infer that firms from civil law countries (e.g., central Europe or Scandinavia) therefore, tend to show higher ESG performance than firms in countries with a legal origin in the common law.[339]

[335] cf. Barnett/Henriques/Husted (2020, pp. 946-945)
[336] cf. ibid. (p. 947)
[337] cf. Chin/Hambrick/Trevino (2013, p. 219)
[338] cf. Bear/Rahman/Post (2010, p. 217)
[339] cf. Liang/Renneboog (2017, p. 896)

Studies in which ESG performance is used as an explaining variable on some sort of operational performance measure are deemed *output-oriented*. Barnett, Henriques, and Husted find the majority of literature investigates the correlation between CSR and certain output variables.[340] Publications deal with the effects specific dimensions of ESG performance have on financial performance and other outputs, e.g., metrics for the level of emissions or gender diversity.[341] There is a vast amount of literature exemplifying this category, as these studies try to find links between measures of financial performance and ESG (e.g., to prove the business case of ESG).[342] Recent meta-studies show that thousands of empirical papers have been published over the last decades.[343] This literature, however, is characterized by many differences. Research varies in the ESG measures and topics being considered, the time horizons, countries, industries, methodologies, and other aspects. Hence, there seems to be no striking evidence for actual outcomes that are based on causal relationships.[344]

Conclusively, studies investigating inputs and outputs cannot be assumed to show any effects caused by ESG performance. In finding factors influencing ESG performance or being influenced by it, they usually presume a kind of relationship or greater good in ESG performance.[345] Barnett, Henriques, and Husted find that studies going beyond this presumption are rare. They conclude that there is a striking lack of literature adequately addressing the outcomes, let alone the impacts of CSR activities. Literature falls short of investigating how such outputs actually support a desired outcome and how these can solely be attributed to the activities of a firm (impact).[346] Hence, the next chapter will describe how the lack of a proper theoretical underpinning for impact assessment and the nature of publicly available ESG information enforces ESG performance metrics to be input-oriented.

3.3.3 Input orientation of ESG metrics

ESG metrics are regularly based on data points retrieved from corporate disclosure (see chapter 2.2.3). For example, ESG ratings benchmark firms according to their contribution to Greenhouse gas emissions. Greenhouse gases, e.g., carbon dioxide, are widely considered to contribute to global warming.[347] ESG ratings incorporate firms that report on the amount of CO_2 accrued during

[340] cf. Barnett/Henriques/Husted (2020, p. 949)
[341] cf. ibid. (p. 949)
[342] cf. Matos (2020, p. 30)
[343] cf. Friede/Busch/Bassen (2015, p. 210)
[344] cf. Matos (2020, pp. 30-31)
[345] cf. Barnett/Henriques/Husted (2020, p. 954)
[346] cf. ibid. (pp. 949-950)
[347] cf. Lashof/Ahuja (1990, p. 529), cf. Ramanathan/Feng (2009, p. 37)

production processes. The calculation of emission levels, however, usually relies on multivariate modeling and broad assumptions. Thus, it is a complex task to observe emissions and quantify their actual amount adequately. Consequently, ESG metrics relying on such figures inherently incorporate outputs that are hardly comparable between different entities. A more comprehensible approach is gathering data on initiatives to reduce carbon emissions, like afforestation programs. Firms might report on the number of trees that were planted or the expanse of preserved wilderness.[348] An ESG metric picking up on this data would rely on the premise that those programs offset climate-damaging gases and that firms engaging in such programs discharge their ecological responsibility. Outcomes in the form of carbon-gas reduction, however, are still almost impossible to measure comparatively. Therefore, ESG rating methodologies often remain on the level of input factors, i.e., the existence of afforestation programs.[349]

Within the social pillar of ESG ratings, gender diversity is a prominent characteristic firm being benchmarked on. Ratings consider firms that, for example, implement programs or policies designed to enhance the percentage of female employees (e.g., female executive qualification programs or fixed gender quotas). Apart from simply stating the existence of such activities, firms could report their efforts in numeric terms, e.g., by stating monetary resources paid per employee. Consequently, ESG ratings base their assessment on the activities rather than the intended outcome of the programs and policies implemented or monetary resources spent.[350]

Although gender diversity is just one factor an ESG rating might consider, it is easy to exemplify the challenges in quantifying the outcomes of related firm activities. It is reasonable to assume that ability as a key trait of potential employees is evenly distributed among different genders. Companies that maintain a gender-balanced workforce could be considered to hire more appropriate candidates on average. Nevertheless, it is also fairly reasonable that structural circumstances (with regard to, for example, labor market conditions, different professions, or regional differences) affect the workforce's gender diversity to some extent that it is not manageable by a company. Thus, quantifying the actual outcomes of a firm's activities would require to consider a multitude of different dimensions that in themselves might affect one another.

With ESG metrics primarily focusing CSR activities and observable outputs, investigating ESG performance metrics inherently falls short of observing actual outcomes. The set of variables is

[348] cf. Serafeim/Zochowski/Downing (2019, p. 17)
[349] cf. Serafeim/Zochowski/Downing (2019, p. 17), cf. Grewal/Serafeim (2020, p. 78)
[350] cf. Grewal/Serafeim (2020, p. 79)

often limited to what firms are able to observe by themselves and communicate via corporate disclosure. As outcomes often need to be set in a broader context exceeding the usual reporting practices, one could not expect to see actual outcomes being measured sufficiently.[351]As a result, recent literature on ESG performance measurement finds that most ESG ratings reflect inputs rather than actual outcomes or impact, which potentially opposes the intention of investors using ESG metrics.[352] The empirical investigation within this study specifically addresses these concerns and proposes measures that mitigate the divergence of the purpose of ESG metrics and what they are actually able to measure. First of all, however, the next chapter briefly discusses how the challenges and practical and theoretical burdens in ESG performance measurement translate into measurable divergence in ESG performance measurement. It presents recent literature on the accuracy and predictive qualities of ESG performance metrics.

[351] cf. Ebrahim/Rangan (2014, p. 123)
[352] cf. Serafeim/Zochowski/Downing (2019, p. 16)

4 LITERATURE REVIEW ON THE ACCURACY OF ESG METRICS

There is a large body of literature implementing ESG metrics in various research fields.[353] Throughout the decades of scholarly research, however, several terminologies evolved. As "ESG", "CSP", "CSR ratings", and other terms share the common conceptualization of non-financial performance measures, this study does not deliberately distinguish between different terminologies.[354] Various terms are used interchangeably and particularly reflect the respective context of the described publication. Further, the ESG rating industry consistently undergoes consolidation processes (s. chapter 2.2.3). To provide a comprehensive overview of relevant literature featuring the main arguments on the quality and accuracy of ESG metrics, the review focusses on the large body of literature implementing the ASSET4 database. ASSET4, founded in 2003, and later acquired by Thomson Reuters (2009) and Refinitiv (2018), delivers ESG data for a significant portion of relevant literature in highly-cited finance and management publications (see **Figure 2**).[355] Further, it is the basis of the empirical analysis in the remainder of this study.

Given the increasing importance of ESG performance metrics and the aforementioned challenges in ESG performance measurement, scholars turn toward the quality and accuracy of ESG scores. Larcker and Tayan (2021) give a concise statement on the conditions that facilitate qualitative and accurate ESG performance metrics: "To be reliable and valid, ESG ratings must be based on data that another outside expert observer would similarly evaluate, predictive in that they are shown consistently to lead to future outcomes, and developed by a provider free from conflicts."[356] In that sense, the remainder of the chapter follows the three dimensions: Credibility and independency of ESG data, the explanatory power of ESG metrics, and disagreement among different ESG data providers.

4.1 Credibility and independency of ESG data

Windolph (2011) provides a systematic overview of theoretical burdens in the reliable assessment of ESG performance. Based on an analysis of potential sources of ESG information (e.g., ratings and indices, awards, NGOs, or internal research), she finds six major causes for variation in the

[353] Gillan/Koch/Starks (2021) provide a comprehensive overview of literature published in the field of Corporate Finance. Christensen/Hail/Leuz (2021) review literature on mandatory CSR and sustainability reporting. Renneboog/Ter Horst/Zhang (2008) and Liang/Renneboog (2020) summarize research on SRI and the field of sustainable finance. McWilliams/Siegel/Wright (2006), Margolis/Elfenbein/Walsh (2009), and Aguinis/Glavas (2012) are examples of comprehensive reviews of literature published in top management journals.
[354] cf. Escrig-Olmedo et al. (2019, p. 2)
[355] cf. Escrig-Olmedo et al (2019, p. 5), cf. Berg/Fabisik/Sautner (2021, p. 33, 47)
[356] Larcker/Tayan (2021, p. 439)

assessment of ESG performance.[357] In finding a "lack of standardization", she points to the main driver of divergence in assessments stemming from the underlying theoretical concept of ESG metrics. Meuer, Kölbel, and Hoffmann (2019) investigate the literature on corporate sustainability as a distinct concept describing corporate actions on ESG issues.[358] Based on two decades of scholarly research, they identify over 30 different definitions of what sustainable corporate actions might include.[359] Controlling for these definitions in the market of ESG data providers, they state that leading participants tend to follow the same variety of definitional clusters as in research publications.[360] However, research on the integration of sustainability principles as the main feature motivating a standardized approach in ESG performance measurement acknowledges the distinct limitation due to somewhat opaque information on the actual rating methodologies of leading data sources.[361] Hence, Windolph states the "lack of transparency" on the actual rating methodology, "biases" and "trade-offs" in the choice of relevant issues, and the weighting of particular dimensions to further induce variation. As a general impediment, she identifies a "lack of credible information" and a "lack of independence" due to idiosyncratic factors within firms and business relations among ESG data providers, further intensifying the potential sources of variation in ESG performance assessment.[362] Building upon these impediments, Escrig-Olmedo et al. (2019) investigate changes in the conceptualization underlying major ESG metrics between 2008 and 2018. Influenced by general growth in the ESG rating market, they find major consolidation patterns among different ESG data providers.[363] Although these patterns potentially alleviate the lack of standardization, the commercial use of ESG data, as well as the increased bargaining power of data providers in light of the market growth, do not resolve supply-side peculiarities on the credibility and independence of ESG data described by Windolph.[364] Hence, Walter (2020) compares ESG rating agencies with the market for credit ratings and states that "maturity" and "commonality" are the primary requirements for ESG ratings to gain credibility. Although imposing new impediments, further consolidation will eventually make ratings more credible over time, according to Walter.[365]

[357] cf. Windolph (2011, pp. 39, 42)
[358] cf. Meuer/Kölbel/Hoffmann (2019, p. 320)
[359] cf. ibid. (pp. 324-326)
[360] cf. Meuer/Kölbel/Hoffmann (2019, p. 337)
[361] cf. Escrig-Olmedo et al. (2019, p. 14)
[362] Windolph (2011, pp. 49-51)
[363] cf. Escrig-Olmedo et al. (2019, pp. 4-5)
[364] cf. ibid. (pp. 13-14)
[365] Walter (2020, p. 334)

4.2 The explanatory power of ESG metrics

Due to a far-reaching database starting in the early 1990s, ESG performance metrics provided by Kyle, Lydenburg, and Domini Research & Analytics (KLD) belong to the most influential metrics in research on ESG performance.[366] KLD data captures data points on various ESG issues (e.g., community, corporate governance, diversity, and environment) and assigns binary values to either "strengths" or "concerns" related to a specific firm.[367] In a first attempt to shed light on the question of how well the KLD scores actually measure past performance and current managerial decision-making, Chatterji, Levine, and Toffel (2009) investigate a sample of 588 US companies between 1991 and 2003.[368] As KLD does not provide overall firm ratings, which aggregate all strengths and concerns indications of one firm, they especially focus on data capturing environmental performance.[369] Chatterji, Levine, and Toffel find that the component "concerns" captures past environmental performance and correlates predictively with future outcomes (e.g., relative pollution levels or litigations and regulatory actions based on environmental performance). However, they find most of the KLD scores to be rather backward-looking and have no evidence of a correlation between strengths with positive outcome variables (e.g., prevention of pollution).[370]

Especially the question for the latter, namely whether firms that specifically conduct business in accordance with the implications of CSR and are aware of their impact on ESG issues outperform others that do not, informed previous literature from various research fields (e.g., corporate finance, accounting, business ethics, or strategic management).[371] Given the heterogeneous viewpoints from which this question can be approached, the theory implies positive, negative, neutral or even bidirectional relationships.[372] Margolis, Elfenbein, and Walsh (2009) review 251 studies published till 2009. They call to investigate the mechanisms that might lead to higher performance (or higher impact) individually and according to the conditions that facilitate their systematic relationship.[373] Based on roughly 700 scientific publications, Aguinis and Glavas (2012) identify several such conditions that either predict, mediate, or moderate to establish CSR outcomes and impacts.[374] They propose that any relationship builds upon various levels of interdependent contexts (i.e., individual performance is influenced by individual motives; organizational performance

[366] cf. Ioannou/Serafeim (2015, pp. 1061-1062), cf. Margolis/Elfenbein/Walsh (2009, p. 11), cf. Kempf/Osthoff (2007, p. 910)
[367] Ioannou/Serafeim (2015, p. 1062)
[368] cf. Chatterji/Levine/Toffel (2009, p. 127, 132)
[369] cf. Ioannou/Serafeim (2015, p. 1062), cf. Chatterji/Levine/Toffel (2009, p. 126)
[370] cf. Chatterji/Levine/Toffel (2009, pp. 162-163)
[371] cf. Gillan/Koch/Starks (2021, p. 2)
[372] cf. Preston/O'Bannon (1997, p. 422)
[373] cf. Margolis/Elfenbein/Walsh (2009, pp. 28-29)
[374] cf. Aguinis/Glavas (2012, p. 952)

is influenced by organizational level motives, etc.).[375] In that sense, Kalaitzoglou, Pan, and Niklewski (2021) analyze the marginal effect of ESG engagement in order to sharpen the understanding of the link between ESG performance and financial performance. Specifically, due to endogeneity concerns being illustrated in previous research,[376] they suggest the introduction of an additional factor capturing heterogeneous and firm-specific unobserved characteristics in the ESG-CFP relationship.[377] They confirm the results of, e.g., Aguinis and Glavas in finding an individual and firm-specific threshold from which ESG engagement has positive marginal effects.[378] Among many others, Gyönyörová, Stachoň, and Stašek (2021), Eccles et al. (2012), and Liang and Renneboog (2017) point out how the industry (e.g., sector membership) and regional peculiarities (e.g., legal origins) systematically affect the effectiveness and relevance of certain ESG dimensions.[379] In conclusion, ESG metrics tend to be, at best, a proxy of various potentially interrelated performance dimensions that differ systematically depending on the specific context.[380]

Given that the conceptualization of ESG metrics, their purpose, underlying data, and recipients potentially differ, research can easily be motivated by questions on the quality and usefulness of ESG metrics.[381] In that sense, it is worthwhile countering the question of the usefulness of ESG metrics. As the introduction in chapter 2 already shows, ESG performance metrics have a justification, as demand has grown considerably in recent decades, and scientific interest is sustainably high and empirically underpinned. Hence, George Serafeim (2021), Harvard scholar and author of many relevant publications in the field of ESG research, points out that the improvement of accuracy and predictive capabilities should lead the discussion on ESG metrics rather than a general questioning of the potential usefulness: "… while assessments could be improved and their connection to future realizations could improve this does not mean that already they do not perform an important function."[382]

[375] cf. Aguinis/Glavas (2012, p. 958)
[376] See for example Bénabou/Tirole (2010, pp. 12-13)
[377] cf. Kalaitzoglou/Pan/Niklewski (2021, p. 219)
[378] cf. ibid. (p. 240)
[379] cf. Gyönyörová/Stachoň/Stašek (2021, p. 1), cf. Eccles et al. (2012, p. 71), cf. Liang/Renneboog (2017, pp. 896-897)
[380] cf. Walter (2020, p. 332)
[381] cf. Christensen/Serafeim/Sikochi (2022, p. 150-151)
[382] Serafeim (2021, p. 19)

Not surprisingly, a recent stream of literature in ESG research focuses on the quality of ESG performance metrics and points out inaccuracies induced by proprietary rating methodologies.[383] Drempetic, Klein, and Zwergel (2020) analyze a global sample of 3,828 firms with ASSET4 ratings from 2004 to 2015.[384] Applying the sample within linear mixed-effects models and structural equation models, they state a significant large cap bias within the ASSET4 rating.[385] The main driver of larger firms gaining better ratings is the amount of disclosed information. As larger firms are able to allocate more resources to ESG reporting, the ASSET4 rating's large-cap bias tends to be based on the mere existence of more information for a firm.[386] Drempetic, Klein, and Zwergel argue that this is partly due to the demand for organizational legitimacy, which is a crucial reason for corporate CSR disclosure.[387] Further, previous literature has shown that non-disclosure is seen as a sign of purposely withholding information that otherwise would have been interpreted as "bad news".[388] As major rating providers follow this notion and assign the worst possible values to non-disclosure, this adds to the aforementioned bias.[389] A feature that has been empirically documented, for example, by Hughey and Sulkowski (2012) for a sample of CSR Hub performance ratings; [390] Or Chen, von Behren, and Mussalli (2021) for a sample of MSCI ESG data.[391] Kotsantonis and Serafeim (2019), however, point out that there might be as many examples of the opposite relationship. They cite the example of "fatalities in the workplace" and argue that depending on the provider, firms that do not disclose corresponding information might also be assessed well (e.g., by assigning zero fatalities) when the methodology assumes that fatalities would otherwise have been disclosed if they occurred.[392] In an attempt to capture the effect of such non-reported information or "not yet-reported" data, Sahin et al. (2022a) introduce a customized ESG score accounting for zero values in the Refinitiv ESG database. Refinitiv assigns values of zero in their category scores when there is no information contributing information to the underlying data points. Sahin et al. (2022a) assume that this information is "not yet reported" by the respective companies and therefore reflects missing information rather than bad ESG performance.[393] Based on a sample of S&P 500 and Euro Stoxx 600 constituents from 2017 to 2019, they find up to 77%

[383] cf. Sahin et al. (2022a, p. 2)
[384] cf. Drempetic/Klein/Zwergel (2020, p. 340)
[385] cf. ibid. (p. 348)
[386] cf. ibid. (pp. 354-355)
[387] cf. ibid. (p. 353)
[388] Verrecchia (2001, p. 155)
[389] cf. Drempetic/Klein/Zwergel (2020, p. 339)
[390] cf. Hughey/Sulkowski (2012, p. 29)
[391] cf. Chen/von Behren/Mussalli (2021, p. 12)
[392] cf. Kotsantinos/Serafeim (2019, p. 54)
[393] Sahin et al. (2022a, p. 3)

of the companies in a specific sector are characterized by zero values in their ESG assessments.[394] Similar to the Refinitiv ESG category scores, they introduce an additional fourth "M" pillar score (for "missing" values) and optimize the weighting with an extended ESGM score based on risk characteristics.[395] Sahin et al. (2022a) find that firms being rated better in ESGM scores tend to show a better risk profile indicating the not yet reported data points bearing valuation-relevant information.[396]

Eccles et al. (2012) demonstrate how the concept of materiality creates a strong case for sector-specific considerations of what is relevant in creating financial or operational value for a company.[397] Khan, Serafeim, and Yoon (2016) pick up the concept of ESG materiality and adjust former non-sector-specific KLD metrics according to industry-specific materiality classifications by the SASB. They find that investing in firms with higher performance on material issues and low performance on immaterial issues tends to create a significant outperformance, as screening only focuses on value-relevant ESG issues.[398] The insight creates a solid case to adjust ESG performance metrics accordingly.

In a recent working paper, Berg, Fabisik, and Sautner (2021) exploit materiality-induced changes in the rating methodology of Refinitiv ESG ratings (former Thomson Reuters ASSET4 database). As of April 2020, Refinitiv introduced two significant changes to its ESG rating methodology. First and foremost, industry-specific weights for single category scores were introduced based on a proprietary materiality assessment. Hence, ESG performance on various issues was no longer weighted equally but rather based on the materiality within the industry a company is conducting most of its business in.[399] A second change was introduced to the scoring of Boolean data points. As described above, Refinitiv adopted a scoring that penalizes companies for non-disclosure of relevant information. Instead of assigning 0.5 on non-reported values (i.e., a slightly positive indication in the "benefit of the doubt"), the new methodology assigns a value of zero.[400] Upon introduction, Refinitiv recalculated each score within the database, making it possible to compare the scores of the old with the ones the new methodology produces. Based on two samples, retrieved in September 2018 (old) and September 2020 (new), of constituents of the S&P 500 from 2011 to

[394] cf. Sahin et al. (2022a, p. 6)
[395] ibid. (pp. 3-5)
[396] cf. ibid. (p. 11)
[397] cf. Eccles et al (2012, p. 67)
[398] cf. Khan/Serafeim/Yoon (2016, p. 1716)
[399] In the remainder of this study the empirical analysis makes use of Refinitiv ESG data. A detailed description of the rating methodology can be found in chapter 5.3 of this dissertation.
[400] Berg/Fabisik/Sautner (2021, p. 11)

2017, Berg, Fabisik, and Sautner (2021), find substantial changes in the scores. Within their samples, 87% scored lower according to the new materiality-adjusted ESG scores, as opposed to the old equally-weighted scores.[401] Hence, Berg, Fabisik, and Sautner state a systematic effect of the new methodology on the outcome of the score distribution.[402] As Refinitiv calculates relative ESG scores, additions to the dataset potentially affect scores even when the underlying data of the respective firm does not change at all. These changes happen on a regular basis as part of a weekly recalculation.[403] Berg, Fabisik, and Sautner (2021) find these changes to potentially have drastic effects on the scoring of certain firms. Comparing a third dataset from November 2020, they find an overlap in the top-ranked decile of around 95%.[404] They compare stock price performances of portfolios formed using the "old" and "new" samples. As they find significant outperformance of highly-scored stocks according to the "new" methodology, they argue Refinitiv might have adjusted for factors that are positively correlated with stock performance.[405] However, considering the results of Khan, Serafeim, and Yoon (2016) described above, one could easily argue that the new methodology of Refinitiv is more efficient in distinguishing between financially material and immaterial ESG data points, thereby imposing an outperformance of highly-rated stocks.[406] This argument is additionally backed up by Kaiser (2020), who analyzes samples of US and European firms ranked with ASSET4 ratings following the "old" methodology.[407] Based on fixed effect panel regressions, Kaiser investigates the correlations between single category scores and future idiosyncratic risk. Assuming category scores, negatively related to future idiosyncratic risk, capture important risk-mitigating information, he argues for distinguishing between material and immaterial categories within the ASSET4 rating.[408] Calculating materiality-adjusted ESG metrics accordingly, he finds portfolios of higher-ranked firms experiencing improvements in risk-adjusted returns and exposure to common systematic risk factors. On the contrary, however, Sahin et al. (2022b) cast doubt on the comparability of findings based on the Refinitiv (ASSET4) database, as the effect on non-definitive ESG metrics (i.e., those that belong to the last five fiscal years, which Refinitiv potentially adjusts on a weekly basis) could significantly affect the findings of previous research.[409] They encourage researchers to exactly report on the extraction date of their

[401] cf. Berg/Fabisik/Sautner (2021, p. 2)
[402] cf. ibid. (p. 25)
[403] cf. Refinitiv (2022a, p. 4)
[404] cf. Berg/Fabisik/Sautner (2021, p. 24)
[405] cf. ibid. (p. 25)
[406] cf. Khan/Serafeim/Yoon (2016, pp. 1697-1698, 1716)
[407] cf. Kaiser (2020, pp. 33-34)
[408] cf. ibid. (p. 39)
[409] cf. Refinitiv (2022a, p. 5), cf. Sahin et al. (2022b, p. 19)

ESG metrics. Information on the time of retrieval from Refinitiv's database would allow to evaluate the comparability of varying results in similar methodological applications. [410]

In a recent working paper Boucher, Le Lann, Matton, and Tokpavi (2022) develop a model to back-test the predictive power of ESG metrics on the example of firm-specific risks. They argue firm-specific risk to be an objective measure, which captures how adverse effects of potential misconduct materialize in market performance.[411] Based on monthly panel regressions, they predict values for idiosyncratic risk based on innovations in an array of accounting variables and an additional set of Refinitiv and Sustainalytics ESG metrics.[412] By comparing the mean squared deviation between idiosyncratic risk estimated in models including and excluding ESG metrics with the actual realized idiosyncratic risk, they assume to identify the informational content of the respective ESG metric.[413] Based on a sample of 781 firms rated between 2010 and 2018, they conclude a high informational content of ESG metrics, especially in Europe. For sub-samples of North American and Asian-Pacific firms, the predictive power of the investigated ESG metrics, however, is relatively low. By further sub-sampling from firms with the highest consensus in their Refinitiv and Sustainalytics ESG ratings, they eventually find the highest predictive power.[414]

4.3 Disagreement among ESG data providers

In addition to within-rating quality issues and sources of inaccuracies, research also compares results on the same set of firms provided by different rating agencies. Chatterji et al. (2016) analyze scores for firms within indices and top-rated percentiles of six major data providers (i.e., KLD, ASSET4, Calvert, FTSE4Good, DJSI, and Innovest). They assess whether the ratings follow a shared understanding of what is being measured ("theorization") and to what extent they rely on the same indicators and proxies ("commensurability").[415] Measurement errors induced by varying theorization and low commensurability have been previously shown to heavily influence the variation in results of empirical investigations. Orlitzky, Schmidt, and Rynes (2003), for example, find that varying scopes of environmental and social issues being incorporated explain 15 to 100 percent of the variation in opposing results of 52 studies on the link between CSP and CFP.[416] Not surprisingly, Chatterji et al. (2016) gather evidence that major ESG data providers and their rating methodologies induce divergence in ESG performance assessment. Although they are not able to

[410] cf. Sahin et al. (2022b, p. 20)
[411] cf. Boucher et al. (2022, p. 5)
[412] cf. ibid. (p. 21)
[413] cf. ibid. (p. 22)
[414] cf. Boucher et al. (2022, p. 29)
[415] Chatterji et al. (2016, p. 1599)
[416] cf. Orlitzky/Schmidt/Rynes (2003, p. 427)

infer a lack of theorization, they state a low agreement on how certain ESG aspects are being measured (commensurability). Hence, they question the validity of ESG ratings and suggest a careful reassessment whether utilized ratings meet the purpose as a valid instrument to assess actual ESG performance.[417]

Semenova and Hassel (2015) focus on the environmental pillar of three major ESG data providers, namely KLD, ASSET4, and GES, for a sample of US firms from 2003 to 2011.[418] They distinguish components of the ratings to reflect either environmental performance (positive impact on the environment) or environmental risks (negative impact on the environment).[419] They find strong correlations between the three ratings in measuring environmental performance. The KLD component concerns is found to have strong convergence with other risk-related dimensions of the ASSET4 and the GES rating, especially as the metrics tend to feature a similar ex-ante measurement approach of the same data points (e.g., amount of emissions and waste). However, when measuring strengths instead, the ratings are found to reflect a rather forward-looking (opportunity) perspective.[420] One major distinction between the measurement of KLD strengths and concerns is the underlying data. Strengths tend to be more forward-looking as they inherit information reflecting input-level activities (e.g., environmental policies and programs). Concerns are back-looking as they reflect output-oriented data like emission levels or water consumption.[421] In that sense, Christensen, Serafeim, and Sikochi (2022) distinguish between input and outcome data points in ESG metrics in order to identify drivers of ESG metric disagreement. They analyze a sample of 5,637 firms from 69 countries with ratings of MSCI, ASSET4, and Sustainalytics from 2004 to 2016.[422] Based on fixed effect regression models and difference-in-differences approaches, they find that especially output-level data points induce rating divergence.[423] Just like Semenova and Hassel (2015), they state that input-level metrics tend to show the lowest disagreement. More importantly, however, Christensen, Serafeim, and Sikochi (2022) find two central patterns in ESG rating disagreement. Divergence among different ratings becomes greater the more information an underlying firm discloses. However, during the sample period, they observe disagreement becoming smaller, which they trace back to the growing consensus on the measurement of certain ESG data

[417] cf. Chatterji et al. (2016, pp. 1607-1608)
[418] cf. Semenova/Hassel (2015, p. 249)
[419] cf. ibid. (p. 254)
[420] cf. ibid. (pp. 256-257)
[421] cf. ibid. (p. 254)
[422] cf. Christensen/Serafeim/Sikochi (2022, p. 149)
[423] They match the sub-samples of firms rated with ASSET4 and Sustainalytics based on the underlying data points and assign manually whether certain data points are input-level or output level. For more on this see Christensen/Serafeim/Sikochi (2022, p. 163).

points over time.[424] Considering the effect of zero-values induced by non-reported information, as Sahin et al. (2022a) show, this pattern might be systematically driven by methodological peculiarities within the ratings.[425] Christensen, Serafeim, and Sikochi (2022), however, find lower disagreement among governance pillar scores as opposed to the environmental and social pillar. They assume that governance-related issues tend to be defined more precisely and that outcomes of certain conduct become more predictable over time.[426] Hence, a "learning effect" potentially opposes the methodological peculiarity described by Sahin et al. (2022a), creating a strong case for a non-stationary (i.e., dynamic) investigation of ESG metrics.

Liang and Renneboog (2017) create an additional case for the positive link between the general consensus on good ESG performance and ESG rating disagreement. They identify systematic differences in the ESG performance levels between firms in common law (e.g., USA, UK) and civil law countries (e.g., Germany, France, or Japan).[427] Applying datasets of various ESG rating providers (i.e., MSCI, ASSET4, Vigeo Eiris, and RiskMetrics) on a sample of firms from 114 countries, they find that firms in civil law countries systematically rank higher in ESG ratings. They explain the pattern with a higher preference for stakeholder orientation, which strongly influences regulatory frameworks and rules of making business.[428] Assuming such rules foster an environment of stronger regulations and state involvement,[429] one could argue that the underlying data of ESG ratings in civil law countries follows a higher degree of standardization due to a joint understanding of "good" ESG performance.[430] Higher reporting requirements in such countries further tend to facilitate such an environment.[431] Hence, not only the informational content or methodological peculiarities (e.g., the assignment of zero-values) but also regional factors bear the potential to drive the accuracy and informational content of ESG metrics.

Nevertheless, disagreement among different providers remains relatively robust. In a highly cited paper, Berg, Kölbel, and Rigobon (2019) analyze ESG scores from six ESG data providers widely used in research and investment practice.[432] They state a maximum level of correlation of scores for the same underlying firms of about 0.38 to 0.71.[433] Following Berg, Kölbel, and Rigobon

[424] cf. Christensen/Serafeim/Sikochi (2022, p. 169)
[425] cf. Sahin et al. (2022a, p. 11)
[426] cf. Christensen/Serafeim/Sikochi (2022, pp. 159-160)
[427] cf. Liang/Renneboog (2017, p. 853)
[428] cf. ibid. (pp. 863-864, 896-897)
[429] cf. ibid. (p. 897)
[430] cf. Christensen/Serafeim/Sikochi (2022, p. 169)
[431] cf. Liang/Renneboog (2020, p. 10)
[432] The following ESG data providers are analyzed: MSCI, KLD, Sustainalytics, Moody's (former Vigeo-Eiris), S&P Global (former RobecoSAM), and Refinitiv (former Asset4).
[433] cf. Berg/Kölbel/Rigobon (2019, p. 44)

(2019), ESG ratings tend to differ by three characteristics: Scope, measurement, and weighing.[434] In order to quantify the influence of each source of divergence, they deconstruct each rating by rearranging the respective data points of each rating into common categories and rating-specific categories. Thereby they acknowledge the proprietary differences in rating construction while simultaneously focusing on the common ground of the ratings.[435] They argue that predominately measurement divergence, meaning the assessment of information on the same issue, implies divergence in the resulting ESG metrics.[436] Aside from qualitative considerations which drive measurement divergence, they also find a systematic effect that might reflect the internal organizational processes of the rating providers themselves. The so-called "rater effect" can be observed when firms that have been rated comparatively well in one category also gain high assessments in other categories, which is likely driven by the same analyst working on the whole company.[437]

Building upon contemporary research identifying the sources for variation in ESG metrics, researchers recently turn toward the effects of disagreement. The empirical investigations primarily focus on stock performance as a representative output variable often investigated in the fields of finance and corporate governance.[438] Gibson, Brandon, Krueger, and Schmidt (2021) investigate how rating disagreement affects stock returns. They investigate S&P 500 constituents between 2010 and 2017 and their respective ratings of seven major data providers, namely Refinitiv, Sustainalytics, Inrate, Bloomberg, GTSE, KLD, and MSCI.[439] They find that the degree to which firms experience the divergence in their ESG assessment of different providers is positively associated with stock returns.[440] Hence, they conclude that uncertainty about actual ESG performance (measured in higher disagreement among raters) is considered an additional risk factor, which is compensated by the market.[441]

Billio et al. (2021) investigate rating disparities of a sample rated with Sustainalytics, RobecoSAM, Refinitiv, and MSCI ESG metrics.[442] Based on portfolio analysis, they find that using metrics of different providers leads to the selection of vastly different assets. Within a subset depicting the common ground of various ratings (i.e., the assets with the lowest disagreement), they are not able to infer significant performance implications. They conclude that ESG-conscious

[434] cf. ibid. (pp. 29-30), Refinitiv ESG ratings are based on the ASSET4 database.
[435] cf. Berg/Kölbel/Rigobon (2019, 3-4)
[436] cf. ibid. (p. 44)
[437] ibid. (p. 4)
[438] cf. Friede/Busch/Bassen (2015, pp. 211-212), cf. Liang/Renneboog (2020, p. 12)
[439] cf. Gibson/Krueger/Schmidt (2021, p. 8)
[440] cf. ibid. (p. 23)
[441] cf. ibid. (p. 4)
[442] cf. Billio et al. (2021, p. 1431)

investors systematically identify dispersed sets of highly-rated firms; therefore, the demand for ESG investments does not significantly add to the demand for a similar set of assets.[443]

Berg, Kölbel, Pavlova, and Rigobon (2021) investigate the effect of rating disagreement on individual stock levels.[444] They propose a method to cope with disagreement among rating providers. They assume disagreement is driven by noise stemming from measurement errors in ESG performance metrics. Hence, they instrument ESG performance based on ESG metrics of eight leading data providers (i.e., ISS ESG, MSCI, Refinitiv, RepRisk, Sustainalytics, S&P Global CSA, Truevalue Labs, and Vigeo-Eiris).[445] In order to quantify the effect of noise imposed by different providers, they estimate stock returns either based on the single ratings or two-stage least-squares regressions, which instrument the respective rating with ratings of other providers. According to their analysis, most ratings are subject to a measurement error imposing partly severe attenuation biases to the effect of ESG performance on stock returns. Contrary to Billio et al. (2021), however, almost all instrumented ratings show a higher and more positive effect on stock returns than in the original setting.[446] According to their analysis, especially those ratings that incorporate news data or complementary information from firm disclosure suffer from noisy performance metrics (i.e., Truevalue Labs and RepRisk).[447]

[443] cf. Billio et al. (2021, p. 1437)
[444] cf. Berg et al. (2021, p. 21)
[445] cf. ibid. (pp. 10, 19-20)
[446] cf. ibid. (pp. 21, 23)
[447] cf. ibid. (pp. 47-48)

5 EMPIRICAL ANALYSIS OF REFINITIV ESG PERFORMANCE SCORES

The following chapter turns toward the empirical investigation of this dissertation. Deviating from studies seeking to evaluate the extent of disagreement and its causes for several ESG data providers, this study investigates one of the most widely used ESG performance measures in particular: The "Refinitiv ESG company scores". In light of the challenges described in the previous chapters and recent empirical findings in research literature, the empirical part of this dissertation explores the scores as well the underlying Refinitiv ESG database in broad detail. Further, it introduces a data-driven approach to bridge input-oriented ESG performance measurement and the outcome of actual yet fairly unobservable ESG conduct in order to explore the quality and accuracy of Refinitiv's ESG performance metrics.

5.1 Hypotheses

The following chapter 5.1 derives hypotheses on the accuracy of ESG metrics. Explicitly investigating the Refinitiv ESG rating universe, the impact of inaccuracies in identifying actual ESG performance is addressed. Refinitiv ESG scores have been widely used in previous literature (**Figure 2**).[448] In order to infer the accuracy of Refinitiv's metrics, this study seeks to investigate the link between ESG performance measures and proxies for actual future ESG performance. Chapters 3 and 4 describe, theoretically and based on contemporary literature, what impediments go along with assessing actual ESG performance. An adequate proxy needs to rest upon a sound set of definitions for single ESG issues, an assessment of those issues considered material to a particular entity, a logic that illustrates how certain business actions create measurable outcomes, and an algorithm which assigns a set of indicators that make it possible to quantify such, potentially oppositional relationships.

This study relies on two distinct bodies of scholarly literature. One of which builds upon an excessive strand of asset pricing research, as it investigates how firm-specific characteristics induce stock price volatility that is not explained by broader movements in the market (i.e., idiosyncratic volatility). In a secondary approach, the study tries to directly assess relationships between ex-ante ESG performance metrics and the occurrence of ESG-related controversies as an embodiment of potential misconduct related to ESG issues. The two approaches address the aforementioned challenges in ESG performance measurement as they specifically address the link between input-based

[448] cf. Berg/Fabisik/Sautner (2020, p. 2)

ESG performance measurement metrics and outcome-based proxies of ESG performance. In doing so, the empirical part of this study gathers empirical evidence from market-based (i.e., ESG performance and IV) and reputation-based (i.e., ESG performance and the occurrence of actual controversies) tests on the variation and its effects in ESG performance metrics. Further, the chapter suggests a data-driven correction methodology that specifically addresses quality concerns by resting upon historical accuracy and methodological peculiarities that systematically induce inaccuracies.

5.1.1 ESG data and firm size

As the literature review in chapter 4 shows, potential sources of inaccuracies in ESG performance measurement have been shown on both qualitative and quantitative levels. Most often, potential biases can be explained based on economic reasoning but simultaneously by methodological burdens that lie in the very nature of ESG performance measurement. One of which is the relationship between ESG metrics and firm size. Based on a vast amount of previous literature, especially in the field of strategic management, Aguinis and Glavas (2012) identify firm size as one of the main moderators of ESG inputs creating actual outcomes. As the firm size is positively related to a firm's visibility as well as resources to engage in ESG issues, the potential to follow effective ESG strategies is enhanced.[449] The relationship is further amplified as higher visibility, on the one hand, creates a rise in stakeholder pressure to lead the way with stronger ESG engagement.[450] As a result, larger firms tend to invest more heavily in ESG performance.[451] On the other hand, the greater availability of resources to invest makes larger firms more responsive to stakeholder demands and enhances the ability to engage in ESG strategies.[452] Consequently, Drempetic, Klein, and Zwergel (2020) and Kaiser (2020) illustrate a substantial, large-cap tilt within the ASSET4 ESG scores.[453] However, they trace the positive association to methodological peculiarities. According to Drempetic, Klein, and Zwergel (2020), the main driver of larger firms gaining better ratings is the amount of disclosed information. As larger firms tend to be able to allocate more resources to ESG reporting, the ASSET4 rating's large-cap bias could be based on the mere existence of more information for a firm.[454] Drempetic, Klein, and Zwergel argue that this is partly due to the demand for organizational legitimacy, which is a key reason for corporate CSR disclosure.[455] As of April

[449] cf. Aguinis/Glavas (2012, p. 943)
[450] cf. Artiach et al. (2010, p. 47), cf. Aouadi/Marsat (2018, p. 1027)
[451] cf. Liang/Renneboog (2017, p. 875)
[452] cf. Waddock/Graves (1997, p. 308)
[453] cf. Drempetic/Klein/Zwergel (2020, p. 348), cf. Kaiser (2020, p. 37)
[454] cf. Drempetic/Klein/Zwergel (2020, pp. 354-355)
[455] cf. ibid. (p. 353)

2020, Refinitiv introduced substantial changes to its rating methodology.[456] One of the changes affected Boolean data points of not-reported data, which affected the results of the scoring methodology greatly, as Berg, Fabisik, and Sautner (2021) show. Considering Sahin et al. (2022a), who state that non-reported data is assigned a value of zero, indicating the worst possible performance (as compared to a value 0.5 within the former rating methodology),[457] the question arises whether the firm size bias is further prevailing under the new methodology as well. As visibility is assumed to be positively related to firm size, the number of ESG controversies published in the media should also be affected positively by firm size.

H1.1 *ESG performance metrics are positively related to firm size.*

H1.2 *Data on the occurrence of ESG controversies is positively related to firm size.*

5.1.2 ESG performance assessment and the availability of ESG information

Major rating providers assign the worst possible performance to firms that do not disclose information on certain data points within their rating methodology. Such providers follow the notion that non-disclosure is seen as a sign of purposely withholding information.[458] However, the assignment of zero values bears a great potential to bias ESG performance assessment when non-reported information is not actually reflecting lower engagement in ESG issues but instead is due to lower firm visibility or fewer resources to engage in extensive reporting. The rating methodology would unjustifiably affect the assessment negatively and, as this is rooted in structural dependencies to specific clusters of firms, induce a systematic flaw in the entire rating approach.[459] A rather qualitative perspective is brought up by Kotsantonis and Serafeim (2019). Their example of non-reported information on "fatalities in the workplace" illustrates how purposely assigning a value of 0 could eventually indicate positive ESG actions fostering employee health and safety.[460] Just recently, the empirical literature has turned toward the effect and potential impact of this methodological peculiarity. Sahin et al. (2022a) state that Refinitiv assigns values of zero in their category scores when there is no information reported in the data. They assume that this information is not actually actively withheld but rather "not yet reported" as the scope of reporting might not already be wide enough to cover all relevant data points of the Refinitiv ESG rating methodology.[461] Further, Berg, Fabisik, and Sautner (2021) point toward the new assessment of

[456] https://www.refinitiv.com/en/media-center/press-releases/2020/april/refinitiv-enhances-esg-scoring-methodology-to-reflect-sustainable-industry-developments-and-market-changes, accessed: 07/14/2022.
[457] cf. Berg/Fabisik/Sautner (2021, p. 11)
[458] Verrecchia (2001, p. 155)
[459] cf. Drempetic/Klein/Zwergel (2020, p. 339)
[460] Kotsantonis/Serafeim (2019, p. 54)
[461] cf. Sahin et al. (2022a, p. 3)

Boolean data points Refinitiv introduced in 2020. Each value not reported by the firm is assigned the worst possible performance and essentially treats non-disclosure as equal to ESG unconscious conduct.[462] The rating methodology till 2020 instead relied on assigning a value of 0.5 on non-reported data points in order to implement cases where there should be a "benefit of the doubt".[463] In conclusion, the empirical investigation of this study tries to find evidence for the following two hypotheses on the effect of non-reported ESG information and Boolean data points to the results of the Refinitiv rating methodology.

H2.1 *ESG performance metrics are negatively biased by the incorporation of non-reported ESG information.*

H2.2 *The fraction of Boolean data points that inform ESG metrics enhances the bias induced by non-reported ESG information.*

[462] cf. Refinitiv (2022a, p. 9)
[463] cf. Thomson Reuters (2018, p. 9), Berg/Fabisik/Sautner (2021, p. 11)

5.1.3 Measuring outcome-level ESG performance

Chapter 3.3 shows that a major concern in literature analyzing ESG performance is the level of ESG performance assessed. Most data points incorporated in ESG metrics are based on corporate reporting. The set of variables is often limited to what firms are able to observe by themselves and communicate via corporate disclosure. As a result, corporate action on ESG issues is often reported on an input-level, reflecting what programs and policies a corporation has introduced or how much money the company has spent to serve a specific ESG-conscious purpose. How well the actions of the company actually fulfill this purpose is rather opaque or hard to assess objectively. Hence, ESG performance measurement might fall short of observing actual outcomes or the impact of corporate actions, let alone comparing such impacts among a variety of firms.[464] Serafeim, Zochowski, and Downing (2019) argue that the resulting input orientation of ESG metrics opposes the intention of investors when using ESG metrics.[465] Amel-Zadeh and Serafeim (2018) and Hartzmark and Sussman (2019) find that investor assign features of financial materiality to the informational content of ESG information. Further, a central motivation to supplement investment decisions with ESG information is rooted in the belief that better performance on ESG issues identifies firms with lower risk profiles.[466] Hence, the question arises of how well input-oriented measurement of ESG performance actually indicates financial material information and whether this materializes in lower risk.

Identifying risk features of firms is characteristic of methodological approaches in asset pricing. Asset pricing models are based on stock prices and therefore reflect an aggregate of subjective perceptions of investors on a firm's valuation and future prospects.[467] As such, the market valuation can serve as a proxy of the objective and contemporary perception of the firm's intrinsic value.[468] Rooted in the seminal works of Harry Markowitz (1952), William Sharpe (1964), John Lintner (1965), and Jan Mossin (1966) in developing the Capital Asset Pricing Model (CAPM), the basic idea of asset pricing is the relationship of a market valuation with distinct firm-specific risk properties. Prices reflect the returns investors wish their choice of risk to be compensated for (so-called "risk premium").[469] The total risk of an asset comprises systematic and unsystematic risks. Systematic risks refer to a variety of risk premia empirically shown to affect the market valuation of all assets within a market. The most common systematic risk factors compensate for

[464] cf. Ebrahim/Rangan (2014, p. 123)
[465] cf. Serafeim/Zochowski/Downing (2019, p. 16)
[466] cf. Amel-Zadeh/Serafeim (2018, p. 28), cf. Hartzmark/Sussman (2019, pp. 2826-2827)
[467] cf. Markowitz (1952, p. 77)
[468] cf. Lee/Myers/Swaminathan (1999, p. 1693)
[469] cf. Sharpe (1964, p. 439), cf. Fama/French (2004, p. 25)

the effect of broad market movements on a specific asset (i.e., market risk premium),[470] the uncertainty about the future performance of small capitalized firms as opposed to large firms (i.e., size premium), and fewer growth opportunities of firms of which the book value exceeds the market valuation (i.e., value premium).[471] Hundreds of such systematic risk factors have been economically described and detected within historical stock prices. Nevertheless, finding the marginal effects of such factors in explaining market returns is an ongoing subject of asset pricing research.[472] Turning toward the actual modeling of asset prices, research literature suggests total risk also consists of an unsystematic or idiosyncratic part. It is rooted in firm-specific characteristics that cannot be explained by factors systematically affecting investors' perception of a firm.[473]

The empirical investigation builds upon the idea that corporate actions on ESG issues can impose adverse effects on the level of firm-specific risk. Numerous examples can be found in the past. Boucher et al. (2022) describe how corporate scandals like Volkswagen's fraud on emission tests induced strong effects on the market valuation.[474] Reber, Gold, and Gold (2021) point out the example of Monsanto during the late 2010s.[475] The Bayer brand faced several lawsuits regarding the adverse health effects of Monsanto's herbicide products, which summed up to litigation costs of several billion dollars. In response, stock prices varied greatly due to the development of the lawsuits that were eventually settled in 2020.[476]

These examples show that uncertainty about the firm's future financial performance is indicated by increased stock price volatility.[477] This study follows Luo and Bhattacharya (2009), Sassen et al. (2016), Kaiser (2020), Reber, Gold, and Gold (2021), and Boucher et al. (2022) and implements stock price volatility as a well-suited ex-post measure of a firm's exposure to events and circumstances posing significant effects on market valuation. Controlling for systematic risk allows to isolate the idiosyncratic risk component, which eventually hints at the firm's exposure to volatility that is rooted in firm-specific characteristics (i.e., idiosyncratic volatility).[478] Hence, this study implements idiosyncratic volatility (IV) as an outcome-oriented measure of ESG performance. If the ESG metric is able to capture the actual ESG performance, one can expect to see lower levels of

[470] cf. Sharpe (1964, pp. 436-437)
[471] cf. Fama/French (1993, p. 48)
[472] cf. Fend/Giglio/Xiu (2020, p. 1359)
[473] cf. Sharpe (1964, p. 439)
[474] cf. Boucher et al. (2022, p. 5)
[475] cf. Reber/Gold/Gold (2021, p. 16)
[476] https://www.nytimes.com/2020/06/24/business/roundup-settlement-lawsuits.html, accessed: 07/15/2022.
[477] cf. Luo/Bhattacharya (2009, p. 199)
[478] cf. Sharpe (1964, pp. 438-439), cf. Luo/Bhattacharya (2009, p. 200), cf. Sassen et al. (2016, pp. 874-875), cf. Kaiser (2020, p. 39), cf. Reber/Gold/Gold (2021, p. 4), cf. Boucher et al. (2022, pp. 4-5)

idiosyncratic volatility, as the outcome of ESG activities relevant for valuation is already captured in the performance assessment of the ESG metric.

Considering varying legal environments (e.g., on behalf of mandatory reporting) and litigation frameworks, this study assumes different levels of regulatory pressure within different regions.[479] Adding to the regional differences, Eccles et al. (2012) and Khan, Serafeim, and Yoon (2016) show that differences in the materiality of specific ESG issues are systematically affected by industry affiliations.[480] This study expects ESG metrics are able to address these systematic peculiarities (e.g., through intra-industry benchmarking and proprietary materiality weighting schemes). Hence, we do not assume to see significant differences in the association between ESG metrics and idiosyncratic volatility, reflecting the level of private information in stock prices,[481] based on regional or industry groupings.

H3.1 *ESG performance metrics are inversely related to idiosyncratic volatility.*

H3.2 *Regional peculiarities do not induce variation in the relation between ESG performance metrics and idiosyncratic volatility.*

H3.3 *Industry affiliation does not induce variation in the relation between ESG performance metrics and idiosyncratic volatility.*

5.1.4 Dynamic materiality considerations in ESG metrics

Rogers and Serafeim (2019) develop a model that identifies crucial steps in ESG issues becoming financially material. They conclude that materiality is evolving over time and resulting from a continuous reevaluation of misalignment between stakeholder expectations and corporate conduct.[482] Materiality is particularly motivated by the amount of information on the misalignment as well as the power of media, NGOs, and politicians to drive changes and the effective enforcement of regulations.[483] Thus, materiality is driven by a complex and sometimes contradictory set of perceptions against misalignment is judged. This circumstance ultimately imposes the necessity to assess ESG performance dynamically and in light of the contemporary perception of what is particularly material for a specific entity. Ioannou and Serafeim (2021) find that the sort of ESG-related corporate actions tends to converge among industry peers. Hence, they show that imitating

[479] cf. Liang/Renneboog (2017, p. 896)
[480] cf. Eccles et al. (2012, pp. 70-71), cf. Khan/Serafeim/Yoon (2016, p. 1697), cf. Kaiser (2020, p. 49)
[481] cf. Ferreira/Laux (2007, pp. 951-952)
[482] cf. Rogers/Serafeim (2019, p. 7)
[483] cf. ibid. (pp. 23-26)

sustainable corporate action practices is most evident within industry clusters of firms facing arguably similar market and production conditions. On the other hand, Gold and Heikkurinen (2018) argue that the increasing call for transparency on corporate conduct over the last decades does not necessarily alter the actions corporations take on behalf of ESG issues (so-called "transparency fallacy").[484] Considering that the scope of corporate disclosure might significantly affect the assessment of ESG performance, this potentially trickles down to the degree to which an ESG metric is able to capture the true underlying ESG performance. Hence, the materiality of ESG performance might differ over time, just as the ESG metric's ability to proxy actual ESG performance fluctuates over time. To shed light on this field of tension, this study investigates how the link between ESG performance metrics and idiosyncratic volatility eventually differs over time. Kaiser (2020) and Sahin et al. (2022a) suggest that dynamic materiality and the increasing availability of ESG information provided by firms significantly affect both risks and ESG performance metrics.[485] Hence, this study investigates the following Hypothesis:

H4 *Shifts in materiality over time significantly affect the ability of ESG performance metrics to indicate actual ESG performance.*

5.1.5 ESG metrics and future ESG controversies

ESG controversies reflect negative news coverage of single companies.[486] They can serve as a proxy for the ex-post materialization of actual ESG performance induced by a certain level of misconduct on ESG issues. Unlike idiosyncratic volatility, which is primarily assumed to reflect ESG risks within the market perception of future firm performance, ESG controversies reflect the direct external recognition of non-effective or negative ESG engagement. This study follows Champagne, Coggins, and Sodjahin (2021), who implement logistic regression models to investigate whether KLD metrics are able to indicate the future occurrence of adverse ESG events.[487] Krueger (2015) and Serafeim and Yoon (2021) find strong negative effects in stock prices based on event study approaches on dates of the occurrence of adverse ESG events and negative ESG news coverage.[488] In line with its risk-mitigation properties presumed by investors,[489] this study aims to test whether ESG metrics actually identify companies with the lowest involvement in ESG controversies.

[484] Gold/Heikkurinen (2018, pp. 318-319)
[485] cf. Kaiser (2020, p. 39), cf. Sahin et al. (2022a, p. 6)
[486] cf. Refinitiv (2022a, p. 4)
[487] cf. Champagne/Coggins/Sodjahin (2021, p. 2)
[488] cf. Krueger (2015, p. 304), cf. Serafeim/Yoon (2021, p. 2)
[489] cf. Amel-Zadeh/Serafeim (2018, p. 28), cf. Hartzmark/Sussman (2019, pp. 2826-2827)

H5.1 *The probability of ESG controversies and strikes is negatively related to ESG metrics.*

H5.2 *The number of ESG controversies is skewed toward the low end of the ESG metric distribution.*

5.1.6 ESG metrics and future firm-specific risk

In order to test the inferential accuracy of ESG performance metrics, the hypotheses above test whether ESG metrics help to identify firms with lower idiosyncratic risk profiles and less exposure to ESG controversies. The analysis reflects the most common sources of measurement error explained in the empirical literature and intends to shed light on how these affect the explanatory powers of Refinitiv's ESG metrics. Considering the results of the subsequent empirical analysis, however, the assessment also allows processing the original Refintiv ESG performance scores when a certain bias induced by the methodology can be adjusted for. This study aims to bridge the opaqueness between input-oriented ESG measurement and its predictive accuracy for risk management purposes in investment contexts. This study does not assume ex-ante any concept of value creation within a positive ESG performance. It allows for several directions of the relationship to hold and instead tries to identify conditions under which accuracy in the prediction of outcomes, which investors intend to aim for in investing alongside ESG metrics, can be enhanced.

This study finds methodological inspiration in the recent work of Kaiser (2020), Boucher et al. (2022), and Sahin et al. (2022b). Kaiser (2020) assumes differences in the relationship between disaggregated ESG metrics and future idiosyncratic volatility based on industry-specific materiality considerations. He reweights each individual ESG category score according to its significant association with a risk measure in a panel-regression setting.[490] As one would expect from conceptual evidence provided by Rogers and Serafeim (2019), materiality presumably underlies changes over time. Kaiser (2020) acknowledges evidence for shifting materiality profiles; however, he does not control for these within the adjusted ESG rating.[491] This study intends to address this gap by providing evidence on dynamic ESG profiles identified based on idiosyncratic volatility.

[490] cf. Kaiser (2020, p. 39)
[491] cf. ibid. (p. 48)

Boucher et al. (2022) predict values for idiosyncratic risk from innovations in an array of accounting variables and an additional set of ESG metrics.[492] By comparing the mean squared deviation between idiosyncratic risk estimated in models including and excluding ESG metrics with the actual realized idiosyncratic risk, they assume to identify the informational content of the respective ESG metric.[493] In line with this dissertation's central argumentation, Boucher et al. (2021) argue that idiosyncratic volatility is an objective measure of adverse effects of ESG performance that materializes in market performance.[494] However, the analysis only takes pillar scores into account. This study applies an adjustment based on single ESG category scores, thereby addressing biases from the compensation of opposing ESG signals that potentially converge in aggregate pillar scores. Further, Boucher et al. (2022) introduce an approach that forecasts idiosyncratic volatility. Hence, they test whether ESG performance predicts ESG risks is backward-looking and not representative of the database an investor at a specific point in time rests her decisions on. This study tests ESG metrics based on historical accuracy. It tries to establish an investor-like setting by resting the analysis of historical accuracy on the data already available at a certain point in time (see **Appendix 13** for an illustration of the timeline).

Berg, Fabisik, and Sautner (2021), and especially Sahin et al. (2022b) illustrate that Refinitiv ESG scores are subject to weekly reassessments. Refinitiv calculates percentile rank scores that after a reassessment, due to their relative nature, do not only alter because the underlying firm introduced innovations in ESG performance. Scores eventually differ after reassessment also because competitors on which performance is benchmarked might have disclosed new information on its actions. Sahin et al. (2022b) evaluate ESG scores by data-mining weighting schemes that specifically optimize the scores to their relationship with proxies for firm-specific risks.[495] However, the approach optimizes the ESG score ex-post. Therefore, an adoption into investment contexts to optimize predictive accuracy might be rather impracticable.

Further, frequent readjustment cycles and adjustments of the rating methodology could impose threats to the very idea of predicting risk exposure with ESG metrics. This study, however, argues that changes in the rating methodology or the rating itself are irrelevant. The main idea of an ESG rating is to reflect a fair representation of a firm's ESG performance at a certain point in time. This assessment rests upon a specific set of ESG information and data points that is gathered and ag-

[492] cf. Boucher et al. (2022, p. 21)
[493] cf. ibid. (p. 22)
[494] cf. ibid. (p. 5)
[495] cf. Sahin et al. (2022b, pp. 1-2)

gregated. Even though the rating at a certain point in time might have differed, the ESG performance assessment rests upon the methodological ability of an ESG metric to establish a consistent link between ESG assessment and outcomes proxying the effectiveness of genuine, yet unobservable commitment to ESG goals. Ex-post additions in the data pool might alter the ESG metric; however, the assessment still rests upon data that might have been available at the time but not have been incorporated in the rating. Hence, the empirical investigation of this study specifically focuses on the methodological accuracy of the Refinitiv ESG performance scores.

Kaiser (2020), Boucher et al. (2022), and Sahin et al. (2022b) base their results on subsets of the whole ESG rating universe. Because the benchmarking of ESG scores happens on the entire ESG rating universe, the analysis might not entirely reflect the ESG score distribution and therefore introduces additional biases in the inference. This study applies potential adjustments based on the inference that takes the benchmarking process on the whole sample into account. Hence, the associations identified within this study should paint an even more accurate picture of what the true relationship might be. Taking the previous deliberations together, this study evaluates the extent to which methodological peculiarities induce systematic biases in Refinitiv ESG performance scores. The study ultimately proposes adjustments to the original ESG metrics that intend to account for the inaccuracies established within the empirical investigation.

H6.1 *Adjusting ESG metrics for biases induced by firm-size, non-reported ESG information, and weighted according to an outcome-oriented and dynamic materiality profile helps to identify lower levels of idiosyncratic volatility.*

H6.2 *Adjusting ESG metrics for biases induced by firm-size, non-reported ESG information, and weighted according to an outcome-oriented and dynamic materiality profile helps to identify portfolios with lower exposure to ESG controversies.*

5.2 Methodology

5.2.1 Regression analysis

In order to test a variety of hypotheses (chapter 5.1), this study's primary methodological approach builds upon the regression analysis. Regression analysis tries to formalize the conditional relationship between an output variable *y* dependent on an explaining variable *x*.[496] Based on a sample of actual realizations of *x* and *y* observed in reality, regression analysis establishes econometric models that mathematically explore the causal effect of a one-unit change in *x* on the expected value of *y*.[497] A basic simple regression model follows the linear function:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

Equation 1 Simple regression model

Variable *y* denotes the dependent or predicted variable, β_0 is the intercept parameter, β_1 is the slope parameter depicting the relationship of *x* (also-called predictor variable) and the dependent variable, ε is the error term. In many contexts, more than one variable is expected to explain the dependent variable *y*. Within multiple regression models, additional realizations of explanatory variables are added to control for such effects. Accordingly, multiple regression models estimate several β_j. The parameter β_0 depicts the constant level of *y* (i.e., when each explaining variable is 0), and the error term ε captures the remaining unobservable parameters affecting *y* that are expected to be constant.[498]

It is rarely the case that econometric studies are able to identify causal relationships in a controlled experiment. However, regression models can indicate the linear effect of a predictor variable by holding all other potentially correlated variables constant.[499] Thus, the parameters β_j are of main interest in applied economics, as they show how the explanatory variables individually determine *y*.[500]

Within regression analysis, model parameters are estimated by an ordinary least square approach (OLS). It refers to the statistical method to minimize the sum of squared residuals. A residual is the difference between the actual value of *y*, e.g., observed in reality, and the fitted value of $\hat{y}$ predicted by the model. The smaller the difference (i.e., the sum of squared negative and positive

[496] cf. Wooldridge (2010, p. 3)
[497] cf. Wooldridge (2020, pp. 28-29)
[498] cf. ibid. (p. 69)
[499] cf. Wooldridge (2010, p. 3)
[500] cf. Wooldridge (2020, p. 21)

deviation), the more accurate the regression model is expected to capture the true (linear) relationship of *x* and *y*.[501] Within regression analysis, "linear" refers to the relationship between the parameters β_0 and β_j. Hence, there are no restrictions on the relationship between *x* and *y* as long as the resulting estimates are interpreted correctly.[502] In fact, throughout this empirical investigation of this thesis, there are several cases in which certain variables are adjusted (e.g., by taking the natural logarithm or squaring) to account for nonlinearity.

The parameters of a simple regression model are estimated as follows:[503]

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Equation 2 Regression coefficient - intercept

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\,(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{COV(x,y)}{VAR(x)}$$

Equation 3 Regression coefficient - explanatory variables

The variables $\bar{x}$ and $\bar{y}$ refer to the respective sample means of x and y. The estimate of the slope parameter $\hat{\beta}_1$ is calculated by dividing the covariance of x and y by the variance of x. The constant parameter $\hat{\beta}_0$ is estimated by rearranging the formula of the regression model based on the respective sample means and the estimated $\hat{\beta}_1$.[504]

Following the logic of OLS estimation, the parameters within a multiple regression setting are estimated simultaneously to minimize the sum of squared residuals (SSR) (i.e., the divergence from the observed sample value y_i and the sum of the fitted value given the estimated parameters) following the formula:[505]

$$SSR = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i,1} - \cdots - \hat{\beta}_J x_{i,N})^2$$

Equation 4 Sum of squared residuals (SSR)

Several assumptions need to be met to estimate the model parameters correctly. The main assumption is that the parameters of the model relate linearly to one another (assumption 1). Although variables could be adjusted for non-linearity (see above), the true model underlying the observed sample data needs to be linear in the estimated parameters. Otherwise, the parameters cannot be

[501] cf. Wooldridge (2020, p. 27)
[502] cf. ibid (p. 40)
[503] cf. ibid (p. 26)
[504] cf. ibid (pp. 25-26)
[505] cf. ibid (p. 70)

interpreted in a meaningful way.[506] Secondly, in non-experimental settings, it is often impossible to observe values for every possible realization of variables. Therefore, the sample data needs to be randomly selected from the true population (assumption 2).[507] The assumption of no perfect collinearity concerns the explanatory variables of the model. The assumption is met when none of the explanatory variables is perfectly correlated with one of the other explanatory variables (assumption 3). Collinearity does not allow to interpret parameters ceteris paribus and potentially adds a redundant variable.[508] A fourth assumption concerns the error term and potential outcomes of model misspecification. Misspecification occurs when the mean of the error given the model's explanatory variables is unequal to zero. In other words, the conditional mean of the error term needs to be zero and independent (CMI) of the explanatory variables (assumption 4). A violation of CMI is caused by endogenous variables, which lead to OLS estimates being biased.[509] Cases in which CMI is violated are presented in chapter 5.2.4. In addition to being independent of the explanatory variables, the variance of the error term needs to be the same for every possible realization of explanatory variables, which is referred to as homoskedasticity (assumption 5).[510] The assumption is not directly required in order to estimate OLS unbiased. However, assuming the error term's variance is constant, it simplifies the calculation of OLS estimates and makes it more efficient to find the combination of parameters that best describes the true model.[511] Following the Gauß-Markov theorem, the OLS estimator $\hat{\beta}_j$ represents the best linear unbiased estimator (BLUE) of the population parameters β_j. When the assumptions hold, no other empirical approach will lead to better estimators than OLS.[512]

This study analyzes binary indicators, such as *STRIKES*, which indicates whether a strike or industrial dispute occurred (“1”) or not (“0”) within a given year. Having dichotomous or binary data as a dependent variable imposes some concerns within OLS estimation. OLS estimation treats dependent variables as continuous, making it potentially exceed its natural boundaries of 0 and 1.[513] Additionally, the variance of the error term is not constant (violating assumption 5) as the error follows a binomial distribution.[514]

[506] cf. Wooldridge (2020, p. 40)
[507] cf. ibid. (p. 80)
[508] cf. ibid. (pp. 80-81)
[509] cf. ibid. (pp. 82-83)
[510] cf. ibid. (p. 88)
[511] cf. ibid. (p. 45)
[512] cf. ibid. (p. 96)
[513] cf. Hosmer/Lemeshow (2000, p. 1)
[514] cf. Hosmer/Lemeshow (2000, p. 7), cf. Wooldridge (2020, pp. 262-263)

A logistic regression calculates the odds of having a value of 1 based on its probability $P(X = 1)$ and the counter-probability of observing a value of 0.[515] By applying the natural logarithm on the odds (logit transformation), the dependent variable, now the log odds of observing a value of 1, remains within the boundaries but can be calculated as a continuous variable.[516] Regression coefficients are consequently be interpreted as the effect of a one-unit change in x_i (ceteris paribus) on the log odds of observing a value of 1 (e.g., the occurrence of a strike).[517] At first glance, this allows for inferring a general idea of the direction of the relationship (e.g., whether the occurrence of x is associated with a higher probability of observing a strike). Exponentiating the coefficient β_j, however, gives the actual odds ratio of observing a dependent variable of 1 over the probability of observing 0, holding the other explanatory factors constant. The odds ratio allows for a more stringent interpretation as it is able to directly tell the percentage change in the probability of observing a value of 1 influenced by a one-unit change in x_i.[518]

In general, a logistic regression follows the function:[519]

$$\ln\left[\frac{P(\boldsymbol{X}=1)}{1-P(\boldsymbol{X}=1)}\right] = \beta_0 + \beta_1 x_1 + \ldots \beta_J x_I$$

Equation 5 Logistic regression model

This study applies several different specifications of regression models. Chapter 5.3 will focus on the sample data, the dependent variables, and the explanatory variables utilized to calibrate these models. Functions of the models are provided as they are analyzed throughout the course of this thesis.

5.2.2 Statistical significance and goodness-of-fit

So far, regression analysis assumes the coefficients estimated by using sample data are representative of the true population. Testing for statistical significance allows qualifying this hypothesis, as it indicates whether or not and to what extent the parameters of the model are of explanatory value.[520] Statistical testing generally relies upon the idea of rejecting the so-called "null" hypothesis ($H_0: \beta_j = 0$) with sufficient certainty. Under the null hypothesis, the partial effect of β_j on y (i.e., after controlling for all other explanatory variables) is equal to zero. Rejecting this hypothesis allows to state the parameter β_j, underlying the true population, actually affects the dependent

[515] cf. Kleinbaum et al. (2010, p. 18)
[516] cf. ibid (p. 19)
[517] cf. ibid (p. 21)
[518] cf. ibid (p. 27)
[519] cf. ibid (p. 19), cf. Hosmer/Lemeshow (2000, p. 48)
[520] cf. Wooldridge (2020, pp. 121-122)

variable. Given that β_j is not observable, hypotheses testing tries to find out whether its estimate $\hat{\beta}_j$ is sufficiently different from zero. In doing so, a so-called t-statistic is calculated. The t-statistic follows the formula:

$$t_{\widehat{\beta_j}} \equiv \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$

Equation 6 t-statistics

Dividing the estimated parameter $\hat{\beta}_j$ by its standard error $se(\hat{\beta}_j)$ tells how many (estimated) standard deviations the parameter differs from zero. Thereby, the t-statistics recognizes the error that sampling from the true population imposes. The higher the t-statistics, the higher the probability that rejecting the null hypothesis is justified. This probability is referred to as the level of *significance*.[521] In practice, the t-statistics is compared with a critical value derived from the t-distribution. The t-distribution shows a value for the t-statistics that is needed to be sufficiently large enough to reject the null hypothesis, given a predetermined level of significance and the degrees of freedom of the model (i.e., the number of observations minus the number of estimated parameters).[522] Statistical significance is usually tested two-tailed to account for values lying on both sides of the distribution curve. As the t-distribution has a mean of zero, the two critical values (positive and negative) lay on the percentiles indicating the respective significance level. A 5% significance level rejects the null hypothesis if the t-statistic from the estimated parameter is higher than the 97.5th percentile of the distribution.[523]

Reading a critical value from the t-distribution requires a given significance level. Although there are quite common significance levels in applied econometrics (e.g., 10%, 5%, 1%), the *p-value* indicates statistical significance more efficiently. Wooldridge (2020) defines the p-value as "… the probability of observing a t statistics as extreme as we did if the null hypothesis is true. "[524] In other words, the p-value indicates "… the smallest significance level at which the null hypothesis would be rejected."[525] Thus, the smaller the p-value, the stronger the indication that the estimate differs systematically (i.e., significantly) from zero.[526] This study depicts both t-statistics and p-values for each estimate within the presentation of empirical results. To infer statistical significance, a level of at least 10% is assumed.

[521] cf. Wooldridge (2020, pp. 120-122)
[522] cf. ibid. (pp. 122, 799)
[523] cf. ibid. (p. 127)
[524] ibid. (p. 131)
[525] ibid. (p. 130)
[526] cf. ibid. (p. 131)

Turning to the model as a whole, one could question how well the prediction of the regression model based on a random sample fits the actual true population. The sample value of y_i is given by the model as the sum of the fitted value $\hat{y}_i$ and some deviation called residual, $\hat{u}_i$.[527] By comparing the variation in the sample values of y and the predicted values, it is possible to find how well the model fits the sample population (i.e., to determine the "Goodness-of-Fit"). The corresponding measure is given by the coefficient of determination, R^2. It is calculated by dividing the explained variation (SSE) by the total variation (SST):

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

Equation 7 Goodness-of-fit

In detail, *SST* is the squared sum of all positive and negative deviations of the sample values from the sample mean. *SSE* is calculated as the squared sum of all positive and negative deviations of the fitted values from the sample mean. The ratio of both, R^2, indicates what percentage of the variation in y is explained by the sample variables x. Given the ability to decompose SST, R^2 is also given by subtracting the ratio of SSR and SST from 1.[528] One specific peculiarity of R^2 is that it tends to become larger, the more explanatory variables are included in the model. Even though additional explanatory variables might actually be irrelevant, the *SSR* is expected to decrease. As a consequence, R^2 might indicate a model to be better than another model just because the number of explanatory variables is higher. Therefore, this study will also report the *adjusted* R^2, which scales *SSR* by the number of explanatory variables. Thereby, *adjusted* R^2 only increases if the additional variable adds explanatory qualities to the model.[529]

5.2.3 Common causes of statistical concern

Given that regression analysis relies upon certain assumptions, violations of these assumptions impose concern on the validity of the respective model and the inference drawn from it. One particular assumption concerns the variance of the error term, which is assumed to be constant for any given value of x (homoscedasticity). The absence of homoscedasticity imposes a bias on the variances of the estimated parameters of the model.[530] The variances of the estimated parameters are needed to calculate standard errors, which are the basis for calculating t-statistics. Consequently, t-values are no longer valid, and stating significance is no longer feasible.[531] White (1980)

[527] cf. cf. Wooldridge (2020, p. 33)
[528] cf. ibid. (pp. 34-35)
[529] cf. ibid. (pp. 196-197)
[530] cf. ibid. (p. 263)
[531] cf. ibid. (p. 264)

derives an estimator for the variance of the estimated parameter that is valid for all cases of heteroskedasticity.[532] In order to account for possible effects of heteroscedasticity, heteroskedasticity-robust standard errors are used throughout this study.

When investigating panel data, having a cross-section of individuals over time adds another layer of potential causes of correlated standard errors. In most empirical finance applications, it is fairly reasonable to assume that residuals are correlated among certain groups of observation (e.g., over time, within an industry).[533] CSR literature shows that such clusters can be observed at the portfolio and the firm level (e.g., a similar level of commitment in ESG topics).[534] To account for the within-cluster correlation of standard errors, this study relies on group-clustered standard errors at the firm level.[535] Cluster-robust standard errors allow for heteroskedasticity and correlation within a group while producing unbiased and consistent estimates in fixed effect panel regressions.[536]

Another common cause of concern is the presence of multicollinearity. It refers to linear models in which two explanatory variables or a linear combination of the explanatory variables are strongly correlated with one another. In a multiple regression setting, multicollinearity makes it harder to distinguish the partial effect of an individual explanatory variable on the dependent variable. As a result, linear relationships may be over- or underestimated. In the presence of multicollinearity, standard errors of the estimated parameters increase, which ultimately affects the correct inference of statistical significance.[537] To control for multicollinearity, this study uses pairwise correlation matrices (i.e., to check whether certain variables are highly correlated) and variance inflation factors (VIF). VIFs for each explanatory variable are calculated by regressing each explanatory variable on a set of the remaining explanatory variables. Following **Equation 8**, the R^2 of the corresponding model are used to calculate a VIF for each variable.[538]

$$VIF_{x_i} = \frac{1}{1 - R_{x_i}^2}$$

Equation 8 Variance inflation factor (VIF)

A common rule of thumb to infer multicollinearity is based on VIFs that exceed values of 4, 10, or even higher. O'Brien (2007) shows that such thresholds are somewhat arbitrary, as in many

[532] cf. White (1980, p. 818-820), cf. Wooldridge (2020, p. 264)
[533] cf. Petersen (2009, pp. 436-437)
[534] cf. Awaysheh et al. (2020, p. 978), cf. Choi et al. (2018, pp. 945-946)
[535] cf. Petersen (2009, p. 475)
[536] cf. Gormley/Matsa (2014, pp. 631-632)
[537] cf. Alin (2010, pp. 370-371)
[538] cf. ibid. (pp. 371-372)

cases, even high VIFs do not prevent drawing insights from regression results.[539] Nevertheless, this study controls for VIFs that exceed values of 10 and discusses whether inference drawn from the models is justified.

To alleviate the problems that outliers within the underlying dataset might impose, this study winsorizes the sample data on the 99th percentile. Winsorization is a common practice to reduce the impact of data points with extreme values. Measurement errors, false data entries within third-party databases, and calculation errors when deducing variables from other data points are examples that result in such extreme values.[540] The author of this study expects data derived from the Refinitiv database to be sufficiently genuine and plausible. Nevertheless, within the global sample, there is a slight possibility of extreme values imposing a bias on the estimated model parameters.[541] Hence, each value that exceeds the 99th percentile of its corresponding data point within the sample is put to the exact value of the 99th percentile.

The panel data design of this study allows for observing the same set of firms over several years. The data basis of Refinitiv grew consistently over the observed period of time. However, mergers or firms going out of business lead to the delisting of some firms from the data basis. It is crucial to consider all of these firms within the sample. This way, this study addresses potential biases imposed by firms characterized by traits that guarantee their survival over other firms. Such a "survivorship" bias is especially prevalent in portfolio or index studies.[542] The Refinitiv ESG database is already a portfolio of listed firms that are considered for the rating process of Refinitiv. As this study is based on the complete ESG rating universe since its introduction in 2022, it is unlikely to observe inferences being distorted by surviving firms. Firms are included within the sample data for each period Refinitiv provides their data.

[539] cf. O'Brien (2007, p. 381)
[540] cf. Ghosh/Vogt (2012, p. 3455)
[541] cf. ibid. (pp. 3456-3457)
[542] Brown et al. (1992, p. 576)

5.2.4 Endogeneity in empirical studies

In order to in establish causality in empirical studies, the applied models need to satisfy a particular set of assumptions. These assumptions ensure that OLS is able to identify relationships correctly and calculate consistent estimates of β_i. The most prominent source of methodological concern in empirical finance literature is endogeneity.[543] Endogeneity occurs when the following assumptions regarding the error term do not hold:

$$cov(x_i, \varepsilon) = 0 \; for \; i = 1, \dots, I$$

Equation 9 Endogeneity - assumption 1

$$E(\varepsilon|X) = 0$$

Equation 10 Endogeneity - assumption 2 (CMI)

The error term explains the fraction of y that is not explained by the dependent variables x_i. Thus, it captures unobserved factors other than the explanatory variables that might explain y. *Endogeneity* occurs when the error term is correlated with an explanatory variable (i.e., violation of assumption 1). As such, the conditional mean of the error term given the set of explanatory variables $E(\varepsilon|X)$ needs to be 0 (assumption 2). As the error term is unobservable, it is impossible to statistically test the correlation between the error term and each explanatory variable. Further, it is impossible to test whether the conditional mean of the error term is independent (CMI) of the set of explanatory variables. As a consequence, there is a solid need to justify the use of empirical models and identification strategies.[544]

There are three common sources of endogeneity. First and foremost, endogeneity results from the misspecification of a regression model.[545] When a model does not incorporate a relevant variable x, the effect of the omitted variable on y is captured by the error term ε. Suppose a true model explaining the dependent variable would require two variables, x_1 and x_2. The estimation within a simple regression model (**Equation 1**) would omit the effect of x_2 and capture it within the error term. For the estimates to be consistent, however, the conditional mean of the error term needs to be zero (see assumptions above), which is only the case, if x_1 and x_2 are uncorrelated. In case, x_1 and x_2 are correlated, the omission of one of the variables would bias the estimation within the misspecified simple regression model by the direction of the correlation between x_1 and x_2.[546]

[543] cf. Roberts/Whited (2013, p. 498)
[544] cf. ibid. (pp. 497-498)
[545] cf. ibid. (p. 498)
[546] cf. Wooldridge (2020, pp. 84-85)

If the omitted variable is observable, it can be added to the set of explanatory variables. McWilliams and Siegel (2000) challenged a then-common practice when assessing the link between a measure of CSR and firm profitability. They found that the estimator for CSR is upwardly biased if expenses in R&D are not incorporated in the explaining model.[547] They state that R&D expenses are strongly correlated with CSR because both ultimately promote innovation in processes and products. Consequently, both explain profitability (or, in the sense of finding evidence, are expected to have an effect) but are correlated to some extent. Excluding R&D performance would lead to an "omitted variable bias".[548]

In empirical investigations, there are oftentimes unobservable variables affecting y.[549] In such cases, there is a need to additionally argue the conclusiveness of causal inferences drawn from a regression model. Prominent examples can be found in the literature on corporate governance. Gompers, Ishii, and Metrick (2003), for example, construct a governance index reflecting the level of provisions that restrict shareholder rights. They find that firms with stronger shareholder rights experience a positive abnormal stock performance.[550] In order to justify their results, they discuss alternative explanations like the potential influence of unobservable variables like "corporate culture". They state corporate culture might cause an omitted variable bias, as it is most likely correlated with the level of shareholder protection (x) and simultaneously be a reason for higher firm valuation (y).[551] Consequently, they state that they cannot infer clear-cut conclusions about the causality of their findings and focus on the correlations they found.[552]

A particular case of omitted variable bias occurs when the underlying data used to calculate a model is not representative of the true population. A so-called "selection bias" results from a non-random selection of "treated" sample data. In such a case, the omitted variable is the unobserved counterfactual of being treated. As a result, the regression estimates of interest are biased as they partly reflect the probability of being selected for the sample.[553] Prominent examples of selection bias can be found in survey data. In survey data, the selection of potential respondents might lead to an underrepresentation of certain groups (e.g., web surveys require an internet connection). In addition, it is up to the potential respondent to participate in the survey (self-selection).[554] Amel-Zadeh and Serafeim (2018) surveyed institutional investors on their preferences regarding ESG

[547] cf. McWilliams/Siegel (2000, p. 605)
[548] McWilliams/Stiegel (2000, p. 608)
[549] cf. Wooldridge (2020, p. 84)
[550] cf. Gompers/Ishii/Metrick (2003, p. 107)
[551] ibid. (p. 142)
[552] cf. ibid. (p. 145)
[553] Heckman (1979, pp. 153-154)
[554] cf. Bethlehem (2010, p. 162)

investments. They state that interpreting the data needs to consider that a questionnaire on ESG investment practices appeals stronger to "more ESG-aware investors".[555]

The second cause of endogeneity concerns measurement errors in the model's variables. It arises when variables are difficult to measure or unobservable. In such cases, research relies on proxies of the unobservable variable. If the proxy measures the variable incorrectly, any divergence from the true value is reflected within the error term of the regression model. The error is certainly correlated with the outcome variable (as it captures the measurement error) and the explanatory variables (as the variables are expected to explain the dependent variable). Hence, conditional mean independence of the error term does not hold.[556] The measurement error in explanatory variables usually leads to inconsistent estimates. As some of the true effects are captured within the error term, the estimates derived from the proxy variable decrease and trend toward zero (attenuation bias).[557] ESG performance metrics are particularly prone to measurement error, as they are a proxy for the actual unobserved and mostly unquantifiable level of commitment to a multitude of ESG issues. Berg et al. (2021) construct a measure of noise that captures the divergence between different ESG rating providers and their assessments for the same set of firms. They find that the higher the noise, the smaller the relationship with stock prices.[558] They even decompose the measurement error and find that measurement errors happen on the level of each indicator (e.g., the level of greenhouse-gas emissions) and the divergence in weightings of the different indicators when aggregating the ESG proxy.[559] Measurement error does also affect ESG scores based on firm characteristics. In the literature on ESG scores, it is quite a common finding that larger firms tend to have higher ESG scores. It is reasonable to assume that larger firms possess more resources to report on ESG-related topics and experience higher visibility. These two characteristics of larger firms might give an advantage to them that is not directly based on the firm's actual ESG performance.[560] Being a small firm, however, does not cause low performance on ESG issues. It instead shows that the measure is not able to reflect a firm's information environment adequately.[561]

Another cause of endogeneity arises when it is possible to argue that the outcome variable affects at least one of the explanatory variables (reverse causality). In this case, variables on both handsides of the regression equation are expected to explain one another.[562] The so-called "simultaneity

[555] Amel-Zadeh/Serafeim (2018, p. 27)
[556] cf. Roberts/Whited (2013, p. 501)
[557] cf. ibid. (p. 503)
[558] cf. Berg et al. (2021, p. 2)
[559] cf. ibid. (p. 55)
[560] cf. Drempetic/Klein/Zwergel (2019, p. 333)
[561] cf. ibid. (pp. 354-355)
[562] cf. Wooldridge (2010, p. 51)

bias" then reflects the relative magnitudes of the opposing effects. Therefore, it is hard to distinguish the direction of the bias.[563] Examples can be found in the research literature on the link between ESG performance and financial performance. One could argue that firms with high profitability or lower financing costs have resources available to additionally engage in actions associated with social rather than financial outcomes. However, one could also expect firms taking the lead on certain ESG topics to experience a higher reputation, which eventually results in financial benefits.[564] Cheng, Ioannou, and Serafeim (2014) show that firms with higher ESG performance experience better access to financing. They argue that the correlation might be explained in both ways.[565] Countering the simultaneity methodologically, they compare effect sizes of both potential causal directions within separate model specifications to justify their results.[566] In this thesis's empirical investigation, endogeneity concerns are discussed as they may occur.

The dataset for this study consists of yearly ESG data for a cross-section of listed firms. Such data structures are generally referred to as panel data.[567] One primary cause of statistical concern within panel data is unobserved heterogeneity. It results from characteristics that are impossible to measure and hence can hardly be included by appropriate control variables.[568] For instance, when investigating ESG performance metrics, unobserved heterogeneity occurs on an industry level. The relevance of certain ESG issues for different firms tends to vary, especially on an industry basis (materiality).[569] Heterogeneity has also been shown to occur according to legal systems, as firms act in different markets and regulatory settings. Consequently, ESG performance is affected on a regional basis.[570] In addition, it is fairly common to have factors in panel data analysis affecting the cross-section of individuals differently during specific periods of time (e.g., within different phases of an economic cycle). To counter these peculiarities methodologically, the regression models within this thesis rely upon a variety of so-called "fixed effects" that are added to the set of explanatory variables. By introducing a dummy variable that equals one if the corresponding attribute is met, individual intercepts are introduced to the model. Thereby, the regression analysis compares observations likely to be characterized by the same unobservable peculiarities, eventually inducing heterogeneity otherwise not accounted for (e.g., industry, region, or time).[571]

[563] cf. Roberts/Whited (2013, p. 500)
[564] cf. Preston/O'Bannon (1997, p. 421)
[565] cf. Cheng/Ioannou/Serafeim (2014, p. 1)
[566] cf. ibid. (p. 15)
[567] cf. Wooldridge (2020, p. 9)
[568] cf. ibid. (p. 439)
[569] cf. Eccles et al. (2012, p. 65)
[570] cf. Liang/Renneboog (2017, p. 853)
[571] cf. Roberts/Whited (2013, pp. 558-559), cf. Wooldridge (2020, p. 440)

5.3 Sample data and proxies

5.3.1 ESG performance

The study builds upon the global Refinitiv ESG rating universe. Refinitiv ESG performance scores are one of the most widely used measures of non-financial corporate performance.[572] The rating allows judging an individual firm's ESG engagement compared to its industry peers and on a country-specific level. Refinitiv ESG data, also known by its former names, "Thomson Reuters ESG scores" or "ASSET4", form the basis for numerous previous studies in the field of ESG research and are well appreciated among investment professionals (see chapter 2.2.3).[573]

Refinitiv ESG performance scores are based on a set of 186 data points gathered for each individual firm from publicly available information. The data points are either qualitative, quantitative, or industry-specific, which makes the number of data points ultimately reflected in a firm's ESG rating vary slightly. Qualitative (or Boolean) data points are based on questions asking for company-specific information on a particular issue (e.g., "Has the company received an award for its social, ethical, community, or environmental activities or performance?"). Boolean data points are either 1 or 0, according to the answer that is derived from publicly available information on the firm (e.g., 1= "Yes"; 0 = "No"). Quantitative data points are adopted by their actual value (e.g., the percentage of employee satisfaction reported by the company or revenues generated by the sales of alcohol). Some metrics are relevant for specific industries only (e.g., "percentage of responsible asset management" in the financial sector). Such values are used only for those companies they apply to. In addition, each indicator has a particular polarity, distinguishing whether a high value is perceived to contribute to ESG performance positively or negatively. A default value of 0 is assigned when no information is available on a particular data point.[574]

The data points for each firm are aggregated within ten distinct ESG categories. Based on a best-in-class approach, percentile rankings are calculated for each data point within 54 industry groups.[575] As the Governance pillar tends to be driven by regulatory conditions rather than industry association, data points within the G pillar of the ESG metrics are benchmarked on a country-level.[576]

[572] cf. Wong/Petroy (2020, p. 5)
[573] cf. Escrig-Olmedo et al. (2019, p. 5), cf. Berg/Fabisik/Sautner (2021, pp. 33, 47)
[574] cf. Refinitiv (2022a, p. 9)
[575] The Refinitiv Business Classification (TRBC) consists of 13 economic sectors, 33 business sectors, 62 industry groups, and 154 industries, cf. Refintiv (2022b, p. 3). The ESG ratings are benchmarked on the industry group-level.
[576] cf. Refinitiv (2022a, p. 3)

Equation 11 shows the basic formula for calculating the percentile ranking used within the ratings. The percentile ranking represents the distribution value to which the other companies within the benchmark perform relatively worse or exactly as good as the observed company. First, an individual percentile ranking for each data point is assessed. The equation consists of the number of all companies that act worse, N_{worse}, all firms that act equally, N_{same}, and all firms that are included within the respective benchmark, N_{all}. The result of **Equation 11** is a percentile ranking for each firm i.[577]

$$percentile\ ranking_i = \frac{N_{worse} + \frac{N_{same} + 1}{2}}{N_{all}}$$

Equation 11 ESG category score calculation, source: Refinitiv (2022a, p. 9)

To calculate ESG category scores, the sum of all percentile rankings for firm *i*, belonging to a specific higher-level ESG category, is ranked by the same formula and benchmark group. The category scores "Emissions" (EMM), "Innovation" (INO), and "Resource use" (RES) belong to the Environmental pillar of Refinitiv's ESG metrics. Scores for "Community" (COM), "Human rights" (HR), "Product responsibility" (PRD), and "Workforce" (WF) form the S pillar. The Governance pillar consists of the "CSR strategy" (CSR), "Management" (MNG), and "Shareholders" (SH) category scores. **Table 2** describes each score as provided by Refinitiv in its most recent methodology paper.[578]

Description of Refinitiv ESG category scores

Pillar		Category score	Description
E	EMM	Emissions reduction	Measures a company's commitment and effectiveness toward reducing environmental emissions in its production and operational processes.
	INO	Innovation	Reflects a company's capacity to reduce the environmental costs and burdens for its customers, thereby creating new market opportunities through new environmental technologies and processes, or eco-designed products.
	RES	Resource use	Reflects a company's performance and capacity to reduce the use of materials, energy or water, and to find more eco-efficient solutions by improving supply chain management.
S	COM	Community	Measures the company's commitment to being a good citizen, protecting public health and respecting business ethics.
	HR	Human rights	Measures a company's effectiveness in terms of respecting fundamental human rights conventions.
	PRD	Product responsibility	Reflects a company's capacity to produce quality goods and services, integrating the customer's health and safety, integrity, and data privacy.
	WF	Workforce	Measures a company's effectiveness in terms of providing job satisfaction, a healthy and safe workplace, maintaining diversity and equal opportunities, and development opportunities for its workforce.
G	CSR	CSR strategy	Reflects a company's practices to communicate that it integrates economic (financial), social and environmental dimensions into its day-to-day decision-making processes.
	MNG	Management	Measures a company's commitment and effectiveness toward following best practice corporate governance principles.
	SH	Shareholders	Measures a company's effectiveness toward equal treatment of shareholders and the use of anti-takeover devices.

Table 2 Description of Refinitiv ESG category scores, source: Refinitiv (2022a, p. 22)

[577] cf. Refinitiv (2022a, p. 9)
[578] ibid. (p. 22)

In order to aggregate the ESG category scores within a single E, S, or G pillar score and, ultimately, the ESG overall rating, each category score gets an industry-specific weight. The weight of each category score reflects the relative importance (i.e., materiality) of the respective category to the overall ESG performance. To assess the importance of each category within an industry group, the relative rate of disclosure of representative data points is used. Based on the materiality considerations that guide corporate disclosure (see chapter 3.2), Refinitiv assumes that more companies within an industry disclose information that they deem specifically important. The magnitude of each category is assessed within decile ranks.[579] The weight of a category for firm *i* is then calculated per industry group or country *j* by the following formula:

$$weight_{i,j} = \frac{magnitude_{i,j}}{\sum magnitude_{i,j}}$$

Equation 12 Calculation of category weights, source: Refinitiv (2022a, p. 12)

The process of calculating the Refinitiv ESG performance score and all underlying category, E, S, and G pillar scores is depicted in **Figure 14**.

Refinitiv ESG data is continuously updated to incorporate new information becoming public. Scores are subsequently recalculated each week. Scores exceeding the five fiscal years prior to the most recent scores remain permanent, regardless of changes or restatements of the underlying data.[580] As most of the ESG data in this study is based on the information published within a firm's annual or quarterly reporting, one does not expect ESG data to change significantly over a year. Annual ESG performance scores, E, S, G, and all ESG category scores for this study were retrieved on the 12th and 13th of April, 2022.

[579] cf. Refinitiv (2022a, pp. 11-12)
[580] cf. ibid. (p. 5)

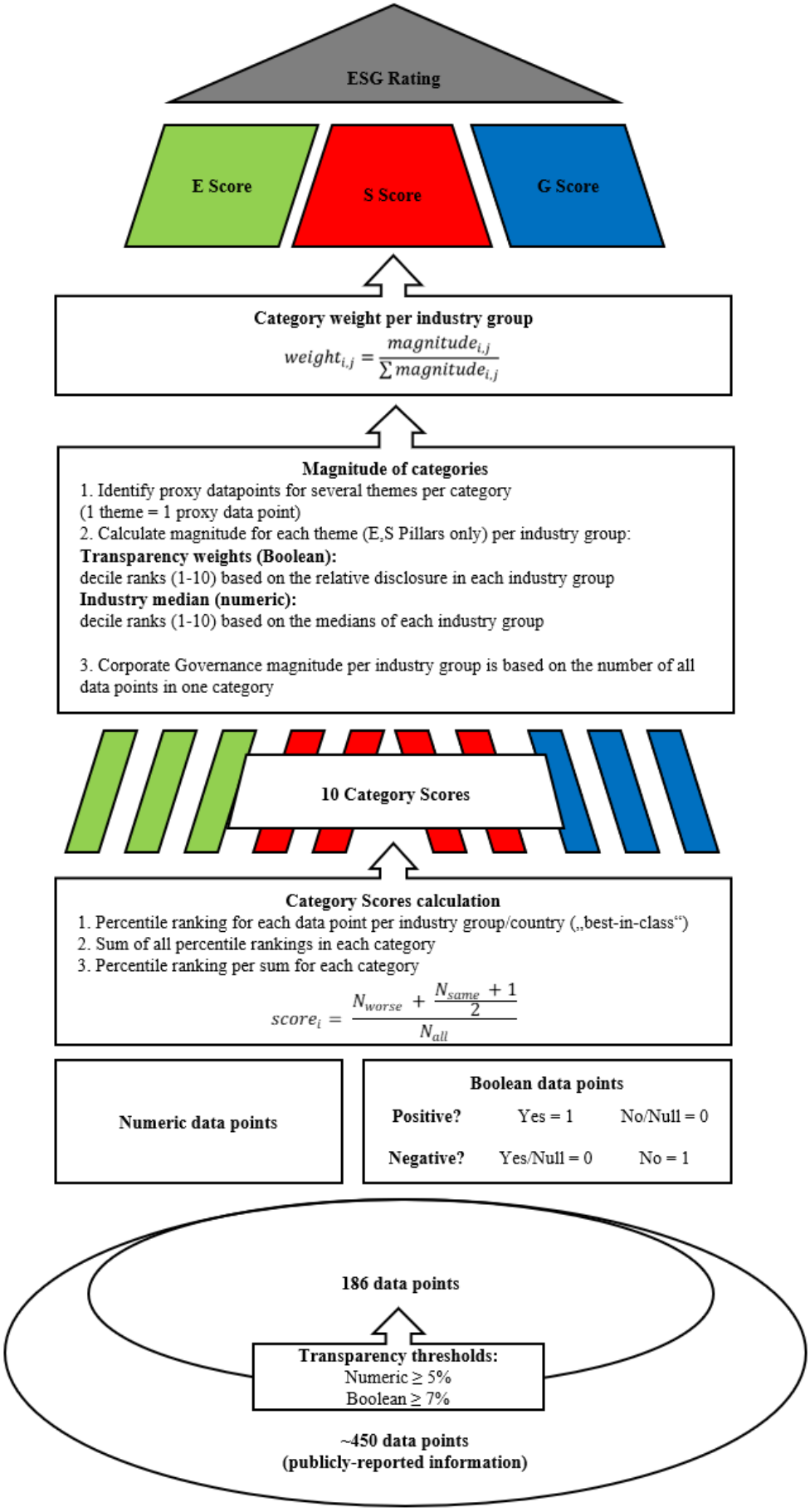


Figure 14 Refinitiv ESG performance score methodology, based on: Refinitiv (2022a, pp. 8-12).

5.3.2 ESG controversies

Refinitiv offers several ESG metrics based on two unique datasets. The primary one pools data mainly based on corporate disclosure. This data facilitates the construction of all major ESG metrics Refinitiv provides (see chapter 5.3.1). The second dataset includes news coverage on selected controversial issues related to firm-specific business conduct in one of the three dimensions, E, S, and G. The dataset is used to calculate Refinitiv's controversy score that measures a firm's current exposure to reputational ESG risks due to controversial business actions.[581] **Table 3** lists all 23 ESG controversy data points and their descriptions as published by Refinitiv in the most recent methodology paper.[582]

Description of Refinitiv ESG controversy data points

Pillar	Category	Controversy	Description
E	RES	Environmental	Controversies related to the environmental impact of the company's operations on natural resources or local communities.
S	COM	Critical countries	Controversies published in the media linked to activities in critical, undemocratic countries that do not respect fundamental human rights principles.
		Tax fraud	Controversies published in the media linked to tax fraud, parallel imports or money laundering.
		Business ethics	Controversies published in the media linked to business ethics in general, political contributions, or bribery and corruption.
		Public health	Controversies published in the media linked to public health or industrial accidents harming the health & safety of third parties (non-employees and non-customers).
		Intellectual property	Controversies published in the media linked to patents and intellectual property infringements.
		Anti-competition	Controversies published in the media linked to anti-competitive behavior (e.g., antitrust and monopoly), price-fixing, or kickbacks.
S	HR	Child labor	Controversies published in the media linked to use of child labor issues.
		Human rights	Controversies published in the media linked to human rights issues.
S	PRD	Product access	Controversies published in the media linked to product access.
		Responsible marketing	Controversies published in the media linked to the company's marketing practices, such as over marketing of unhealthy food to vulnerable consumers.
		Privacy	Controversies published in the media linked to employee or customer privacy and integrity.
		Responsible R&D	Controversies published in the media linked to responsible R&D.
		Customer health & safety	Controversies published in the media linked to customer health & safety.
		Consumer complaints	Controversies published in the media linked to consumer complaints or dissatisfaction directly linked to the company's products or services.
S	WF	Wages working condition	Controversies published in the media linked to the company's relations with employees or relating to wages or wage disputes.
		Diversity and opportunity	Controversies published in the media linked to workforce diversity and opportunity (e.g., wages, promotion, discrimination, and harassment).
		Employees health & safety	Controversies published in the media linked to workforce health and safety.
G	MNG	Mgt compensation	Controversies published in the media linked to high executive or board compensation.
G	SH	Shareholder rights	Controversies linked to shareholder rights infringements published in the media.
		Insider dealings	Controversies published in the media linked to insider dealings and other share price manipulations.
		Accounting	Controversies published in the media linked to aggressive or non-transparent accounting issues.

Table 3 Description of Refinitiv ESG controversy data points, source: Refinitiv (2022a, p. 23)

[581] cf. Refinitiv (2022a, pp. 3-4)
[582] ibid. (p. 23)

The data points described in **Table 3** indicate the number of times controversial business conduct has been reported on by a wide range of international news outlets (e.g., the number of controversies linked to business ethics, political contributions, or bribery and corruption). In addition to the numeric data points, Refinitiv captures the occurrence of strikes and work stoppages with a binary indicator. Binary data is used to indicate whether an incident happened ("1") or not ("0"). Thus, the data point *STRIKES* answers, "Has there been a strike or an industrial dispute that led to lost working days?"[583]

Refinitiv offers the most recent as well as historical values for each controversy data point separately. As controversies move to the previous fiscal year, the respective number of controversies or binary indications in the data points reflecting the most recent controversies is set to 0. Therefore, data on the most recent controversies usually covers only the last two years in the dataset for this study or the last two years of a company within the rating universe.[584] In order to capture both historical and the most recent controversy data points in one variable, both data items are summed up for each company and each year.

It is important to note that the two datasets of ESG scores and ESG controversies are independent of one another. ESG controversies do not affect the original ESG scores ex-post. Hence, no endogeneity concerns are imposed on the part of the empirical investigation seeking to explore the capability of Refintiv's ESG performance scores to indicate future ESG controversies.

5.3.3 Firm-specific risk and idiosyncratic volatility

Firm-specific (or idiosyncratic) risk is the residual component of total risk that cannot be explained by general movements in the market (systematic risk).[585] Following previous literature on the relationship between ESG performance and firm-specific risk, this study utilizes multifactor modeling to retrieve a measure of volatility in idiosyncratic risk.[586] Multifactor models are extensions of the seminal Capital-Asset-Pricing-Model (CAPM), which seeks to explain the expected returns of assets based on their exposure to several risk factors.[587]

The development of the CAPM in the 1960s is mainly associated with the individual work of William Sharpe (1964), John Lintner (1965), and Jan Mossin (1966).[588] It is based on the modern portfolio theory by Harry Markowitz (1992). Markowitz presumes risk-averse investors and states

[583] cf. Refinitiv (2022a, p. 23)
[584] cf. ibid. (p. 14)
[585] cf. Sharpe (1964, p. 439)
[586] cf. Sassen et al. (2016, pp. 878-879), cf. Kaiser (2020, p. 39), cf. Luo/Bhattacharya (2009, p. 205)
[587] cf. Feng/Giglio/Xiu (2020, p. 1332)
[588] cf. Fama/French (2004, p. 25)

that each investor optimizes a portfolio either based on the highest expected return for a given risk or the lowest risk for a given return.[589] The resulting portfolios are called "mean-variance-efficient" since they dominate any other combination of securities that neither yields a higher return nor a smaller risk.[590] The CAPM uses strict assumptions regarding the individual behavior of investors (i.e., rational, single-period mean-variance optimizers with homogeneous expectations), and the market structure (frictionless markets with a fixed supply of securities and all information publicly available).[591] Given these premises, all investors hold the same mean-variance-efficient portfolio in market equilibrium (market portfolio), independent of their individual risk preferences. This portfolio consists of all available assets, weighted by their market capitalization. Investors account for their risk preferences by mixing the market portfolio with a risk-free investment opportunity.[592] Analyzing expected returns across various individual securities, the CAPM postulates that those can only result from a systematic risk factor, the so-called beta (β).[593] The beta factor is defined as the covariance of a security's return and the return of the market portfolio relative to the variance of the market portfolio. Consequently, beta factors are a measure to assess the degree of risk individual securities are exposed to relative to the market (systematic risk).[594] In conclusion, the following equation highlights the way the CAPM explains expected returns on investments for an investment horizon comprising a single period:[595]

$$E(r_i) = r_f + \beta_i * [\, E(r_m) - \, r_f]$$

Equation 13 Expected returns following the CAPM

For security *i*, the expected return is defined as the sum of the risk-free rate of return and the individual systematic risk exposure of this security (β_i) multiplied by the market risk premium ($E(r_m) - \, r_f$). Hence, the market risk premium (i.e., the interest spread to which the market yields a higher return relative to the risk-free rate of return) is the primary factor that influences the expected return relative to the individual systematic risk.[596]

Empirical tests show that a model exhibiting a single risk factor loading does not sufficiently explain actually observed returns of securities.[597] For example, Banz (1981) shows that firms with

[589] cf. Markowitz (1952, p. 82)
[590] Fama/French (2004, p. 26)
[591] cf. Blitz/Falkenstein/van Vliet (2014, p. 62)
[592] cf. Fama/French (2004, p. 28)
[593] cf. ibid. (p. 28)
[594] cf. Fama/French (2004, p. 28), cf. Sharpe (1964, pp. 436-437)
[595] cf. Fama/French (2004, p. 29)
[596] cf. ibid. (pp. 29-30)
[597] cf. Fama/French (1992, p. 427), cf. Feng/Giglio/Xiu (2020, p. 1327)

smaller market capitalizations tend to have higher CAPM risk-adjusted returns.[598] Basu (1983) shows that firms with higher earnings-to-price ratios systematically show higher risk-adjusted returns.[599] Rosenberg, Reid, and Lanstein (1985) find that the same holds true for portfolios with high book-to-market ratios that generate unjustifiably high positive abnormal returns when measured by the CAPM.[600] Lakonishok, Shleifer, and Vishny (1994) find similar patterns associated with past sales growth and cash flows.[601]

As the CAPM fails to predict such systematic associations, they are usually referred to as "anomalies".[602] Fama and French (1996) argue that their extension of the CAPM is able to capture a variety of such anomalies and enhance the explaining power of the CAPM.[603] Fama and French (1993) introduced two additional factors based on firm size and book-to-market ratios, which is referred to as the Fama-French three-factor model.[604]

$$r_{i,t} - r_{f,t} = \alpha_i + \beta_{i,M} * \left[r_{M,t} - r_{f,t} \right] + \beta_{i,SMB} * SMB_t + \beta_{i,HML} * HML_t + \varepsilon_{i,t}$$

Equation 14 Fama-French three-factor model

The Fama-French three-factor model follows previous market models based on the CAPM, for example, introduced by Jensen (1968),[605] and proclaims a linear relationship regarding the excess return ($r_{i,t} - r_{f,t}$) of security *i* to the market in period *t*. Compared to the CAPM, the Fama/French model extends the model by including two additional risk premia. The model is extended for the betas regarding the size effect and the book-to-market ratios. The size beta stems from the empirically confirmed phenomenon that small-capitalized firms outperform their larger-capitalized peers due to higher volatility in stock price movements ("small-firm effect"). Thus, the size beta can be interpreted as a security's sensitivity to size-related price movements (size premium). Such price movements are captured in *SMB* ("small minus big"), which is calculated by deriving historic excess portfolio returns on small-cap stocks over large-cap stocks.[606] Former empirical research provides evidence, postulating significant excess returns of value stocks over growth stocks ("value effect"). Thus, the market-to-book beta (value premium) can be interpreted as a security's sensitivity to price movements, related to its characteristics as value stocks (high book-to-market ratio) and growth stocks (low book-to-market ratio). The return spread anomaly is captured in the

[598] cf. Banz (1981, p. 16)
[599] cf. Basu (1983, p. 129)
[600] cf. Rosenberg/Reid/Lanstein (1985, p. 9)
[601] cf. Lakonishok/Shleifer/Vishny (1994, pp. 1555-1556)
[602] Fama/French (1996, p. 55)
[603] cf. ibid. (p. 55)
[604] cf. Fama/French (1993, p. 48)
[605] cf. Jensen (1968, p. 393)
[606] cf. Fama/French (1993, p. 9)

HML factor loading ("high minus low"), which reflects excess portfolio returns on value stocks over growth stocks.[607] Because of its empirical foundation, the model finds great support in its explanatory power over the initial CAPM. Nevertheless, further empirical studies in this area of research find that the three-factor model remains insufficient to capture all price-relevant risks. Jegadeesh and Titman (1993) detect evidence that short-term well-performing stocks earn higher returns than the three-factor model estimates. The same holds for poorly performing stocks that tend to continue performing poorly in the short run.[608] This effect (so-called "momentum effect") arises due to the existence of market inefficiencies, given that capital markets, other than postulated in theory, react slowly to new information. This anomaly is robust over multiple periods.[609]

In light of the empirical shortcoming, Carhart (1997) proposes an additional extension of the Fama/French three-factor model. The Carhart four-factor model adds the factor momentum (MOM) to the initial three-factor model:

$$r_{i,t} - r_{f,t} = \alpha_i + \beta_{1,i} * \left[r_{m,t} - r_{f,t} \right] + \beta_{2,i} * SMB_t + \beta_{3,i} * HML_t + \beta_{4,i} * MOM_t + \varepsilon_{i,t}$$

Equation 15 Carhart four-factor model

The momentum beta $\beta_{4,i}$ captures the security's price sensitivity to its former one-year performance (momentum premium). The related factor MOM_t is based on the difference between the one-year historical returns of the best-performing over the worst-performing securities.[610]

To derive a measure of firm-specific risk, this study builds upon approaches of previous literature investigating idiosyncratic volatility (IV) within multifactor models. Following — among others — Campbell et al. (2001), Ang et al. (2006), and Cao et al. (2008), the error term of the market model captures the residual risk component after controlling for systematic risk captured by size, value, and momentum factors. Thus, the residual risk is expected to resemble the risk associated with firm-specific characteristics.[611] This study extends Kaiser's (2020) approach based on an Fama-French three-factor model by calculating yearly IV for single firms based on a Carhart four-factor model. To capture potential regional differences in capital markets, the four-factor models are run separately for a set of European and US-domiciled firms.

[607] Fama/French (1993 (p. 9)
[608] cf. Jegadeesh/Titman (1993, p. 89)
[609] cf. ibid. (p. 90)
[610] cf. Carhart (1997, p. 61)
[611] cf. Campbell et al. (2001, pp. 5-7), cf. Ang et al. (2006, p. 283), cf. Cao et al. (2008, pp. 2603-2604)

IV is defined as the standard deviation of the daily residuals of firm i ($\varepsilon_{i,t}$) from a four-factor model model based on daily stock returns for a single year.[612]

$$IV_{i,y} = std(\varepsilon_{i,t})$$

Equation 16 Idiosyncratic volatility

Daily returns are calculated based on the total return index (TRI) provided by Refinitiv and retrieved from the Worldscope database. In doing so, the return data within this study accounts for price deviations that might be caused by various causes, like dividend payments or stock splits. Daily factor data for capital markets in Europe and the U.S. is retrieved from the data library section on Kenneth R. French's website. The website offers a wide range of different multi-factor datasets following the most prominent configurations of multifactor models. Factor data following Fama/French (1993) and Carhart (1997) from April 2002 to March 2022 is used for this study. IV is calculated based on the respective dataset corresponding to a firm's domicile country. The US sample consists of all firms domiciled in the United States of America. The European sample captures all firms from Austria, Belgium, Switzerland, Germany, Denmark, Spain, Finland, France, Great Britain, Greece, Ireland, Italy, Netherlands, Norway, Portugal, and Sweden.[613]

5.3.4 Control variables

One major cause of endogeneity concerns is the omitted variable bias. It arises from explanatory variables that might be correlated with the dependent variable and other explanatory variables but are not included within a model to predict the dependent variable (see chapter 5.2.4). The dependent variables (e.g., idiosyncratic volatility) within this thesis may certainly be affected by other factors (e.g., firm size) than the mere ESG scores. Therefore, the regression models within this thesis account for various control variables. The idea is to control for variation in the dependent variable while holding the effect of other factors known to have a potential relationship with the dependent variable steady. The remaining variation in the dependent variable can be attributed more precisely to the variable of interest.[614] The following chapter explains the control variables and their potential impact on the main dependent variables within this thesis, namely IV and ESG controversies.

A common variable known to be associated with numerous other firm characteristics is *firm size*. Studies on ESG performance regularly control for variation systematically depending on firm

[612] cf. Kaiser (2020, p. 39)
[613] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/index.html, accessed: 01/02/2022.
[614] cf. Wooldridge (2010, pp. 3, 13)

size.[615] Empirical evidence suggests larger firms tend to show lower firm-specific risk levels.[616] Malkiel and Xu (1997) suggest that size effects in return patterns could be a proxy for idiosyncratic risk. Smaller firms are perceived as riskier; thus, markets demand a risk premium.[617] In the field of ESG research, the positive tilt in ESG performance toward larger firms is well documented.[618] This relationship is usually attributed to a rise in stakeholder pressure and firm visibility with increasing firm size.[619] Consequently, larger firms tend to invest more heavily in ESG performance.[620] Eccles et al. (2012) further point out that differences in the materiality of ESG issues within industries should be primarily driven by firm size if there are no substantial differences in business strategies.[621] Simultaneously, the ESG rating industry is challenged because of biases stemming from firm size. Rating methodologies, as they often rely on counting relevant data points, tend to give larger firms, with potentially more resources to engage in the beneficial preparation of ESG disclosure, a natural advantage.[622] Hence, one could expect to see firm size affecting the visibility of ESG controversies.

Investigating whether ESG performance indicates future IV and the probability of observing controversial media coverage can be expected to be correlated with firm size. Following (among numerous others) Liang and Renneboog (2017) and Kaiser (2020), this study measures firm size as the natural logarithm of total assets.[623] Data is retrieved from the Refinitiv Worldscope database. Total assets reported in local currencies are converted into US dollars.

It is reasonable to assume that another key driver of IV is *profitability*. Profitability is expected to contain information on future cashflows directly linked to volatility in future market valuation.[624] The impact of profitability on ESG performance builds upon the slack-resource hypothesis. As firms are more profitable, more funds are available to engage in ESG actions.[625] In literature, this relationship is often referred to as "doing good by doing well".[626] Consequently, this study expects the number of ESG controversies to be affected by profitability.[627] In order to account for such

[615] cf. Kaiser (2020, p. 36), cf. Sassen et al. (2016, p. 867)
[616] cf. Lee/Faff (2009, p. 229)
[617] cf. Malkiel/Xu (1997, p. 12)
[618] cf. Matos (2020, p. 51)
[619] cf. Artiach et al. (2010, p. 47), cf. Aouadi/Marsat (2018, p. 1027)
[620] cf. Liang/Renneboog (2017, p. 875)
[621] cf. Eccles et al. (2012, p. 70)
[622] cf. Drempetic/Klein/Zwergel (2020, p. 355)
[623] cf. Liang/Renneboog (2017, p. 866), cf. Kaiser (2020, p. 36), cf. Oikonomou/Brooks/Pavelin (2012, p. 497)
[624] cf. Luo/Bhattacharya (2009, p. 200)
[625] cf. Preston/O'Bannon (1997, p. 423)
[626] cf. Matos (2020, p. 31)
[627] cf. Giese et al. (2019, p. 4)

effects, this study controls for return on assets (ROA) as a measure of profitability.[628] To capture the effects of volatility in profits, this study additionally accounts for the standard deviation of quarterly ROA over three years prior to the observed year.[629]

ROA is retrieved from the Refinitiv Worldscope database and calculated as the ratio between net income of firm *i* and interest payments (adjusted for tax expenses) and the average amount of total assets at the end of the last two calendar or fiscal year-ends:

$$ROA_{i,t} = \frac{net\ income_{i,t} + [interest\ expenses_{i,t} * (1 - tax\ rate_t)]}{\frac{1}{2}\sum_{t=-1}^{0} total\ assets_{i,t}}$$

Equation 17 Calculation of control variable: Return on assets

In order to account for the potential effects of a firm's capital structure, this study uses *leverage* as an additional control variable. Literature on the leverage of publicly held firms primarily focuses on its ability to impose agency costs but enhance managerial efficiency. Agency theory suggests that in the presence of debtholders, managerial actions and potential conflicts of interest are limited while simultaneously being monitored more effectively.[630] Jensen (1986) builds upon the free cash flow hypothesis and states that continuous payments reduce excess cash that would otherwise fund non-positive investment projects.[631] On the other hand, higher leverage might impose a higher risk of financial distress and bankruptcy.[632] Consequently, empirical results on the relationship between leverage and its effects on return patterns remain contradictory.[633] Studies on ESG performance regularly control for a firm's level of debt.[634] It is reasonable to assume that debt financing bears the potential to significantly impact corporate decision-making concerning specific ESG issues and vice versa.[635] Leverage is obtained from the Refinitiv Worldscope database. It is the ratio of the sum of long-term and short-term debt of firm *i* divided by common equity at the calendar or fiscal year-end *t*.

$$Leverage_{i,t} = \frac{(long-term\ debt_{i,t} + short-term\ debt_{i,t})}{common\ equity_{i,t}}$$

Equation 18 Calculation of control variable: Leverage

[628] cf. Harjoto/Jo (2011, p. 50), cf. Eccles/Ioannou/Serafeim (2014, p. 2836), cf. Crace/Gehman (2022, p. 3)
[629] cf. Sassen et al (2016, p. 880), cf. Luo/Bhattacharya (2009, p. 200), cf. Ferreira/Laux (2007, p. 958)
[630] cf. Jensen/Meckling (1976, pp. 350 f.)
[631] cf. Jensen (1968, pp. 323-324)
[632] cf. DeAngelo/Masulis (1980, p. 4)
[633] cf. Korteweg (2010, p. 2137), Demarzo (2019, pp. 1590-1591)
[634] cf. Sassen et al. (2016, p. 880), cf. Luo/Bhattacharya (2009, p. 200)
[635] cf. Matos (2020, p. 53)

Other measures of firm prospects include *capital expenditures intensity* (CAPEX), *R&D expenditures intensity* (R&D), and *intangible assets intensity* (IA). McWilliams and Siegel (2000) find that empirical investigations on the relationship between CSR and financial performance often suffer misspecifications and omitted variables bias. They argue that CSR is strongly associated with intangible characteristics like reputation, quality, or reliability. Most prominently, R&D expenses strongly correlate with CSR and future firm prospects, as R&D promotes innovation in processes and products.[636] Moreover, recent literature finds that intangible assets (especially goodwill) drive stock price crash risk, as it is associated with high levels of uncertainty.[637]

Capital expenditures, R&D expenditures, and intangible assets are retrieved from the Refinitiv Worldscope database. Values reported in local currencies are converted into US dollars. To calculate intensity measures, each value is scaled for total assets at the corresponding calendar or fiscal year-end. To limit the effect of missing data on the sample size, R&D expenditures are set to zero in case they are not reported.

$$CAPEX\ intensity_{i,t} = \frac{capital\ expenditures_{i,t}}{total\ assets_{i,t}}$$

Equation 19 Calculation of control variable: CapEx intensity

$$R\&D\ intensity_{i,t} = \frac{research\ \&\ development\ expenditures_{i,t}}{total\ assets_{i,t}}$$

Equation 20 Calculation of control variable: R&D intensity

$$IA\ intensity_{i,t} = \frac{intangible\ assets\ _{i,t}}{total\ assets_{i,t}}$$

Equation 21 Calculation of control variable: IA intensity

Firm age is considered to contain information about different organizational cycles and the development stages of a firm.[638] Fink et al. (2004) find idiosyncratic risk is significantly driven by age characteristics of individual firms. Younger firms tend to bear higher risks, especially when issuing public equity in the early stages of their life cycle.[639] Concerning the engagement in CSR in different stages of a firm's development, no clear inferences can be drawn based on existing literature. Nevertheless, there seems to be evidence for firms showing stronger ESG performance with increasing age.[640]

[636] cf. McWilliams/Siegel (2000, p. 605)
[637] cf. Wu/Lai (2020, pp. 1-2)
[638] cf. Luo/Bhattacharya (2009, p. 200)
[639] cf. Fink et al. (2004, p. 23)
[640] cf. Jiraporn/Withisuphakorn (2016, p. 298)

Firm age (AGE) is retrieved from Refinitiv Eikon and calculated as the natural logarithm of the absolute years since a firm's initial public offering.

A common practice in panel data analysis is to account for fixed effects (see chapter 5.2.4). As panel data represents a time-series of cross-sectional data, there are certain effects that affect the cross-section within one period but might not affect the others. For market valuation, this could be investor sentiment, regulatory changes, or economic climate that affect all firms within the cross-section alike. Therefore, this study accounts for *year-fixed effects*. For each period, a dummy variable is implemented that takes a value of one if the observations were made in the respective year and zero otherwise. This thesis builds upon the idea that ESG materiality differs by a firm's respective context. Theoretical and empirical findings suggest that materiality differs systematically over time, by industry, and based on the regulatory framework within different countries.[641] Therefore, this study not only accounts for year-fixed effects but also considers *industry*- and *country-fixed effects*. The respective fixed effects are implemented by binary coding of industry membership and the country of domicile. Industry membership and the country of domicile are retrieved from the Refinitiv Worldscope database. The Refinitiv ESG scoring methodology (see chapter 5.3.1) benchmarks firms within their industry group (E, S) or country-wise (G). The industry grouping follows the Refinitiv Business Classification (TRBC), which consists of 136 industries within 54 industry groups that form 28 business sectors and ten economic sectors. Regression models estimated within the ten economic sectors. If the number of firms within different business sectors and industry groups is relatively steady over time, one cannot expect to see significant differences in materiality regardless of which granularity industry-fixed effects are implemented.[642] Whenever the number of observations is sufficient, however, this study relies on industry group-fixed effects to be as close as possible to the original rating methodology.

Appendix 12 gives a comprehensive overview of the variables, their definitions, and data sources used within this study's empirical analysis.

[641] cf. Rogers/Serafeim (2019, p. 1), cf. Eccles et al. (2012, p. 70), cf. Liang/Renneboog (2017, p. 853)
[642] cf. Eccles et al. (2012, p. 70)

5.4 Descriptive statistics and analysis

5.4.1 Data on ESG performance

The empirical investigation is based on the Refinitiv ESG data universe. Annual ESG performance metrics for this study were retrieved in April 2022. Refinitiv claims to cover over 9,000 active firms within its rating universe. Annual ESG ratings reach back to 2002 and cover firms listed in commonly known global (e.g., MSCI World, MSCI Emerging Markets) and regional equity indices (e.g., DAX, S&P 500, NASDAQ-100). The rating universe is expanding as constituents of new indices, and changes within the already covered indices are accounted for on a yearly basis. Historical data reaching back further than five years remains unchanged.[643] To avoid potential biases induced by the limitation to active firms only (i.e., survivorship bias), this study's historical panel dataset fully covers each firm Refinitiv ever offered ESG data for. The original dataset is comprised of 11,792 unique firms and 83,827 firm-year observations capturing ESG performance from 2002 to 2021.

Temporal distribution

Firm-year observations: ESG score (2002-2021)

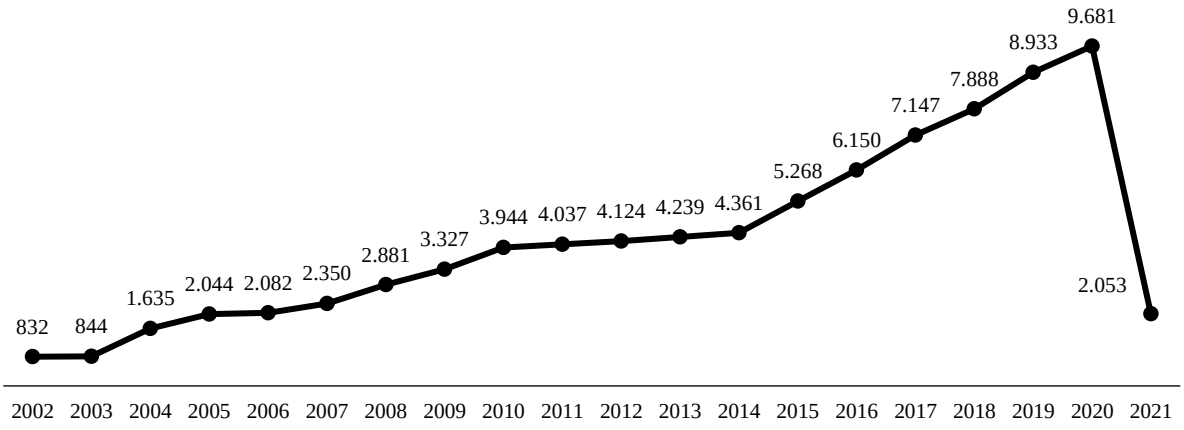


Figure 15 Number of firm-year observations: ESG scores (2002-2021)

Figure 15 shows the number of firm-year observations of yearly overall ESG scores captured within the dataset of this study. The Refinitiv rating universe constantly grew throughout the observed period. The dataset ranges from 832 firms in 2002 to 9,681 firms in 2020. Most of the ESG data points are retrieved from annual and CSR reports. Hence, the majority of ESG data points are

[643] cf. Refinitiv (2022a, p. 5)

updated in line with the annual reporting patterns of the underlying firms once a year.[644] As the data was retrieved in April 2022, the number of observations in 2021 (2,053) is relatively low. It can be expected that the rating universe will cover at least as many firms as in previous years and that the corresponding scores will be added as soon as the underlying information is disclosed. Each ESG score is based on ten individual ESG category scores. Due to the sample formation and the data collection process, not all firms are listed with each individual category score. The number of firm-year observations deviates on a scale from -1 to -28 from the total number of firms in the sample. The strongest deviations are observed between 2007 and 2012. The Innovation category score suffers the most missing firm-year observations when compared to the total number of rated firms in the sample. Category scores referring to the Governance pillar show the least missing scores. Deviation in the number of category scores per firm indicates that the ESG score methodology might not be capable of providing scores due to insufficient information in some years. Nevertheless, the small deviation impacts a maximum of 0.68% of the total number of observations in a specific year.

Most data points needed to calculate ESG scores stem from a firm's annual or sustainability reporting. Although scores are updated every week based on the most recent information, most data is actually added during specific periods throughout the year (i.e., a few months after a fiscal year has ended). Roughly 70 percent of the firms within the dataset publish annual reports following a fiscal year-end in line with the calendar year (i.e., December 31st). Therefore, ESG data relying on the annual reports of such firms captures the ESG performance within the corresponding calendar year. When fiscal year end dates fall in a previous month, however, some of the published information also refers to the previous calendar year. Within the yearly rating methodology of Refinitiv, this adds confusion about whether ESG performance partially refers to the previous year of the calendar year it is allocated to. As ESG scores are updated based on the underlying firms' reporting practice, this also implies ratings for a year being available far earlier than for other firms in the corresponding year. Attributing ESG ratings to the most recent year somehow makes sense from a data vendor's view. Considering new firms being added to the rating universe, one can assume that it is more appealing to users of such information to assume to use the most recent indication of ESG performance. However, this makes it harder to draw inferences from such numbers, especially when comparing through a cross-section of yearly ESG rating data. An exemplary look at two firms is described in **Appendix 4** and **Appendix 5**.

[644] cf. Refinitiv (2022a, p. 4)

This study analyzes future firm-specific risk and the possibility of ESG controversies on a calendar year basis. The ESG data of the previous year is expected to have a predictive value on future outcomes. Therefore, the data capturing the ESG performance needs to account for the corresponding ESG performance within a period prior to such outcomes being observed. This empirical investigation in chapter 5.5 runs panel regressions based on lagged explanatory variables. It is reasonable to assume that lagged explanatory variables alleviate the fact that potential outcomes observed in the dependent variable already affected the ESG scoring in the independent variables. Further, control variables retrieved from the Worldscope and Eikon database usually refer to a fiscal year. Hence, the risk of comparing values from different periods is accounted for. Nevertheless, the number of observations that are based on a fiscal year-end date prior (20,530) and after July 1st (63,297) makes up roughly a quarter (25.4%) of the observations in the dataset. One can assume ESG performance assessed based on data of firms that report on fiscal years ending in the first half of the year mostly refers to the previous year. Simultaneously, annual reports for fiscal years ending in the year's second half rather report on ESG performance in the corresponding year. This study relies on a sample of unadjusted data potentially being exposed to the described inconsistencies but matching the data an investor would have at hand.

Through most of this study, the dataset is limited to samples with a sufficient number of firm-year observations for each year (e.g., 2003 to 2020). Therefore, sample sizes differ depending on the corresponding model. Whenever needed, sample construction is described separately.

Regional distribution

The dataset contains firms from 89 different countries (**Appendix 6**). The countries of domicile are retrieved from Datastream to match the Refinitiv benchmark methodology of the Governance pillar score. It can be assumed that regional factors (e.g., legal framework) systematically affect a company's way of ESG engagement.[645] Following the UN M49 standard country codes for statistical use, the firms can be attributed to five distinct geographical regions: Americas, Europe, Asia, Oceania, and Africa (**Figure 16**).[646]

Number and fraction of firms per country

Country	Number of firms	Fraction of firms
United States	4,091	34.7%
United Kingdom^{EU}	831	7.0%
Sweden^{EU}	345	2.9%
Germany^{EU}	298	2.5%
Switzerland^{EU}	231	2.0%
France^{EU}	199	1.7%
Italy^{EU}	134	1.1%
Netherlands^{EU}	98	0.8%
Spain^{EU}	92	0.8%
Finland^{EU}	91	0.8%
Norway^{EU}	91	0.8%
Denmark^{EU}	69	0.6%
Belgium^{EU}	57	0.5%
Ireland^{EU}	57	0.5%
Austria^{EU}	40	0.3%
Greece^{EU}	35	0.3%
Portugal^{EU}	19	0.2%

Table 4 Number and fraction of firms per country

Almost half of the firms within the dataset are domiciled in the United States of America (34.7%), China (8.1%), and the United Kingdom (7.0%). Not surprisingly, the regional groupings of the "Americas", "Europe", and "Asia" make up 92.9% of all firms covered in the dataset. Country-fixed effects within panel data analysis are implemented using the corresponding countries of domicile throughout the remainder of this study.

The empirical investigation will focus on the ratings of firms domiciled in the United States (4,075) and on a sub-sample of European firms (2,684) matching the Fama French factor data retrieved from Kenneth French's data library (see chapter 5.3.3). The respective number of firms from each country is listed in **Table 4** (^{EU} denotes all countries constituting the European sub-sample investigated hereafter). Thus, the final scope of this study will cover roughly 57,5% of all firms Refinitiv ever offered ESG data for.

[645] cf. Liang/Renneboog (2017, p. 853)
[646] https://unstats.un.org/unsd/methodology/m49/, accessed: 04/20/2022.

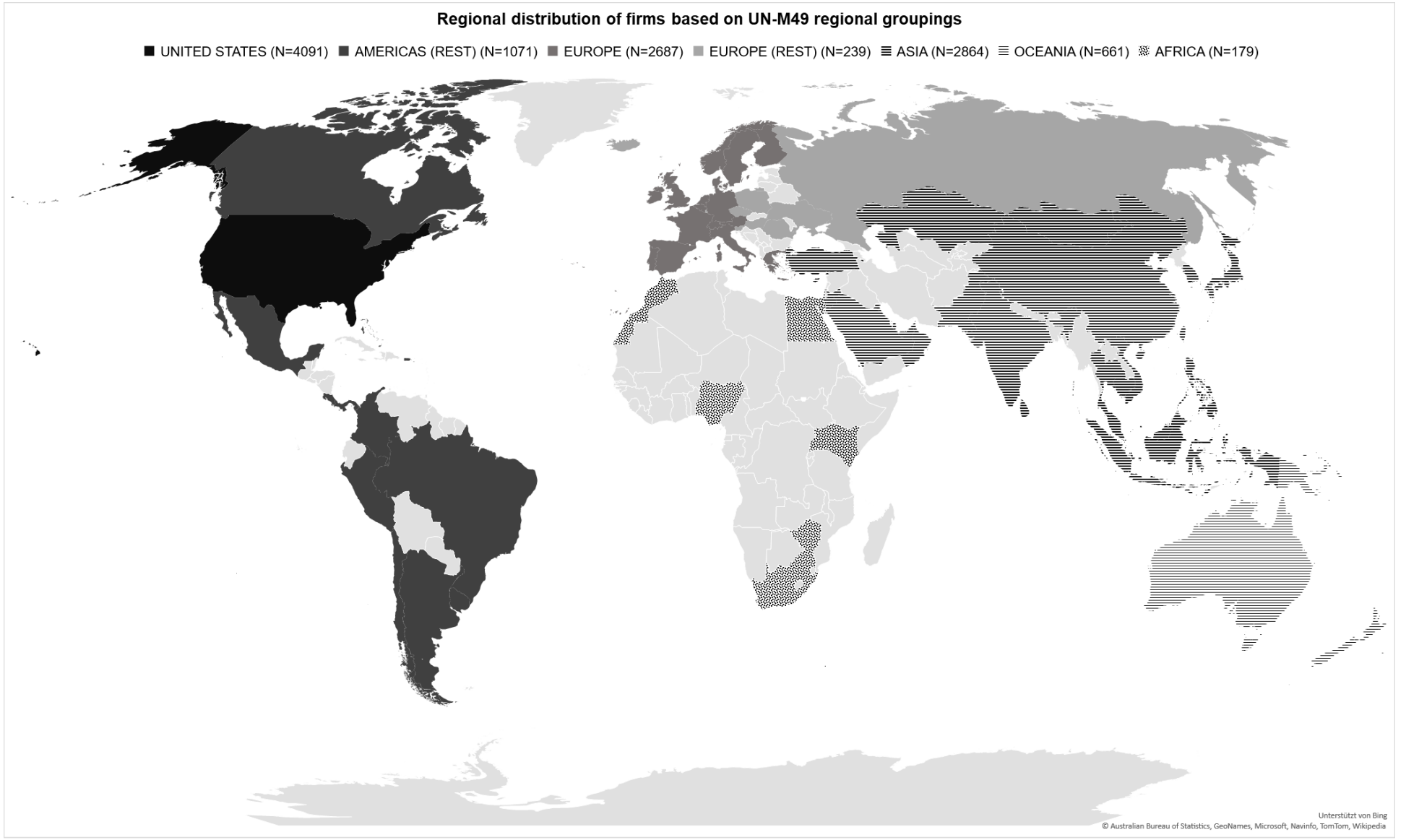


Figure 16 Regional distribution of firms based on UN-M49 regional groupings

Industry distribution

The Refinitiv ESG scoring methodology (i.e., best-in-class approach) for environmental and social category scores relies upon the TRBC industry classifications to benchmark firms among their industry peers.[647] Considering the ratings and scores are normalized on an industry basis, other industry classifications would lead to relative scores within one industry being compared to relative scores of other sub-samples. This disparity eventually induces measurement errors (e.g., when adding industry-fixed effects) by design and hinders the inference of the empirical investigation. Thus, following the same industry classification within this study allows the panel data analysis to be as closely in line with the rating methodology and compare scores of the same basis with one another.

The Refinitiv Business Classification (TRBC) classification consists of 13 economic sectors, 32 business sectors, and 61 industry groups. More detailed distinctions (i.e., into industries and activities) are not considered in the empirical investigation of this work, which is in line with the rating methodology. In case a firm could be assigned to two or more business segments, the dominant segment (i.e., in terms of revenues, assets, or operating profits) determines the company's classification as a whole.[648] Thereby, TRBC allows for a one-on-one mapping of firms and industry groups within the dataset.

Table 5 shows the number and fraction of firms in the dataset for each economic sector following TRBC. None of the firms in the dataset relates to the economic sectors “Institutions, Associations & Organizations” and “Government Activity”. Refinitiv does not consider firms belonging to these sectors for ESG rating calculation. Most of the firms rated in the dataset belong to the “Financial sector” (1,708). Among those, more than three quarters belong to the business sector “Banking & Investment Services” (1,294). The remaining firms spread among “Industrials” (14.2%), “Technology” (13.9%), “Consumer Cyclicals” (13.2%), and “Healthcare” (11.6%). Within the healthcare sector, firms belonging to the “Pharmaceuticals & Medical Research” (883) business group make up the second largest business group. This is predominantly driven by corresponding firms based in the U.S. (541). Firms belonging to the sectors “Basic Materials”, “Real Estate”, “Consumer Non-Cyclicals”, and “Energy” make up 9.1% to 6.4% of the dataset. Very few firms belong to the TRBC sectors “Utilities” (3.2%) and “Academic & Educational Services” (0.4%). Because of the small number of observations, the latter is dropped from most of the sub-samples

[647] cf. Refinitiv (2022a, p. 3)
[648] cf. Refinitiv (2022b, p. 4)

throughout the remainder of this study. The average number of firm-years covered within the dataset lies between 5.0 (Healthcare) and 8.9 years (Utilities).

Number of firms and firm-year observations grouped by TRBC economic sectors

TRBC - economic sector	Firms	Fraction	Firm-year obs.	Fraction	avg. firm-years
Financials	1,708	14.5%	12,775	15.2%	7.5
Industrials	1,680	14.2%	12,517	14.9%	7.5
Technology	1,641	13.9%	10,200	12.2%	6.2
Consumer Cyclicals	1,558	13.2%	11,932	14.2%	7.7
Healthcare	1,371	11.6%	6,835	8.2%	5.0
Basic Materials	1,069	9.1%	8,466	10.1%	7.9
Real Estate	812	6.9%	5,605	6.7%	6.9
Consumer Non-Cyclicals	772	6.5%	6,051	7.2%	7.8
Energy	751	6.4%	5,811	6.9%	7.7
Utilities	380	3.2%	3,379	4.0%	8.9
Academic & Educational Services	50	0.4%	256	0.3%	5.1

Table 5 Number of firms and firm-year observations within the dataset grouped by TRBC economic sectors

The empirical investigation mainly focuses on two samples of European and US firms. **Table 6** shows the number of firm-year observations per economic sector for each sub-sample. The industry distribution of both samples varies with the number of firms within each economic sector and the years covered in the ESG rating universe. While the US sample is dominated by firms in the financial sector (15.1%), Industrials make up the most firm-year observations in the European sample (14.8%). The US sample is dominated by observations of firms in the Financials, Technology (15.0%), and the Consumer Cyclicals (14.9%) sectors. Firm-year observations within the Technology sector in the European sample make up only half of the fraction (7.5%) compared to the US sample. The same holds true for the Healthcare sector, which makes up 14.1% of the US sample and only 5.2% of the firm-year observation in the European sample.

Number of firm-year observations grouped by TRBC economic sectors and regions

	United States		Europe	
TRBC - economic sector	Firm-year obs.	Fraction	Firm-year obs.	Fraction
Financials	4,038	15.1%	3,172	11.9%
Technology	3,994	15.0%	1,990	7.5%
Consumer Cyclicals	3,970	14.9%	3,293	12.3%
Healthcare	3,754	14.1%	1,400	5.2%
Industrials	3,307	12.4%	3,939	14.8%
Real Estate	1,854	6.9%	1,042	3.9%
Energy	1,795	6.7%	1,082	4.1%
Consumer Non-Cyclicals	1,526	5.7%	1,378	5.2%
Basic Materials	1,473	5.5%	1,745	6.5%
Utilities	983	3.7%	633	2.4%

Table 6 Number of firm-year observations grouped by TRBC economic sectors and regions

The differences in industry distribution underline the case to investigate the dataset on a sub-sample basis. Considering differences in the materiality of ESG issues, the perception of the same issues may vary according to the perceived number of potentially affected firms. If material ESG

issues in a large sector affect relatively large portions of firms in the market, media coverage will eventually be responsive to these topics.

Appendix 7 shows the number of yearly firm-year observations per economic, business sector, and industry group. The distribution of firm-year observations among the different business sectors and industry groups matches the number of firms listed above. Nevertheless, the observations for some industry groups lack a sufficient number to adequately reflect the comparative character of Refinitiv's ESG performance score. To mitigate the effect of unobserved heterogeneity between different industries, panel regressions implement industry-fixed effects. In light of the very small number of observations within some of the investigated industry groups, only regression models run over the full dataset implement fixed effects on the industry-group level.

The results on the relationship between ESG performance and future IV (chapter 5.5.1) are presented for individual economic sectors. In most of these specifications, especially as the sub-sample sizes increase with the scope of the Refinitiv rating universe over time, the number of observations is not sufficiently large enough to run regression models with industry group-fixed effects. Therefore, the regression models are either run with fixed effects on the business sector level ("Industrials", "Technology", "Consumer Cyclicals", "Consumer Non-Cyclicals", "Healthcare", and "Basic Materials", "Financials", "Energy") or do not implement industry fixed effects other than the subsample grouping within the economic sector ("Utilities" and "Real Estate").

For the results of the empirical investigation to hold, the distribution of business sector observations within different economic sectors needs to be relatively constant over time. Otherwise, the industry-fixed effects would not allow controlling for unobserved heterogeneity, as the relative impact of characteristics leading to heterogeneity varies over time. **Appendix 8** shows the year-to-year change in the fraction of observations in different business sectors belonging to a specific economic sector. The year-to-year change in fractions occasionally varies but does not exceed values of -9% to +11%. Further analysis (not reported in this thesis) shows that even within the business sectors consisting of various industry groups, the relative fractions of industry groups are almost even within all business sectors. Hence, aggregating findings on the level of economic sectors and applying business sector-fixed effects does not add any sources of empirical concern.

Appendix 9 shows the number of observations, mean, median, and standard deviation of the firm-year observations for each ESG metric per economic sector. The number of firm-year observations differs to some extent, as there are cases in which Refinitiv does not provide firm-specific scores for each ESG metric in a given year. In most of these cases, the rating only covers data points

belonging to one or two of the three ESG pillars and the overall ESG rating. Nevertheless, the rating methodology allows to calculate a weighted ESG overall rating comparable to other ESG ratings with a larger scope of underlying data. In addition, most of the panel regression specifications in this study incorporate only one category score at a time, hence, alleviating the potential of lacking category scores to impose empirical concern.

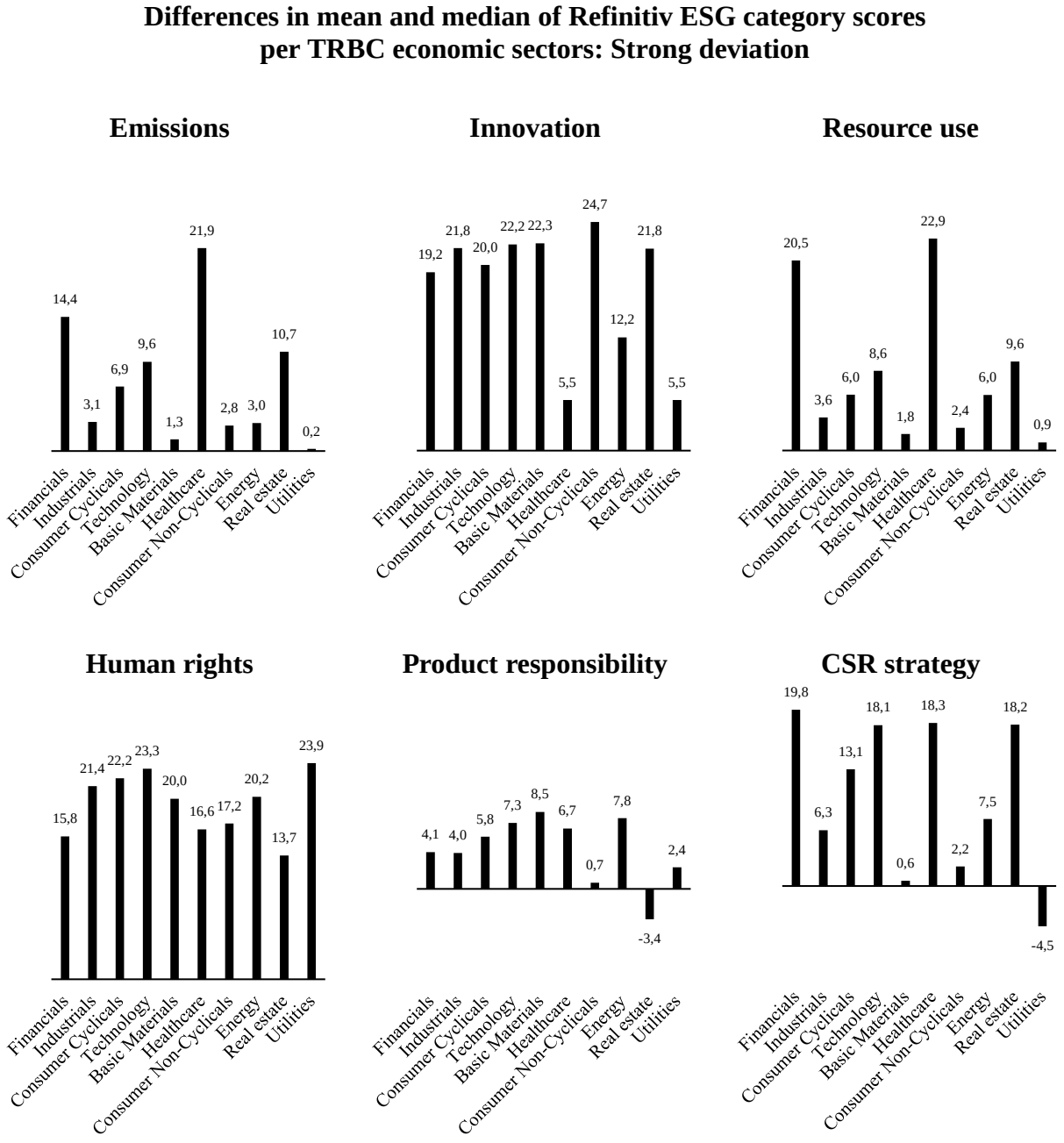


Figure 17 Differences in mean and median of Refinitiv ESG category scores per TRBC economic sectors: Strong deviation

Based on the mean and median depicted in **Appendix 9**, **Figure 17** and **Figure 18** show the differences of mean and median for category scores within each economic sector. In almost all cases, the sector mean of category scores is higher than the sector median. The differences reach up to 23.9, which is a quarter of the whole rating scale. Surprisingly, this pattern is relatively strong for

some category scores (**Figure 17**); for others, it is fairly weak or non-existent (**Figure 18**). This finding implies that the distributions of category scores tend to differ systematically. When the median category score is lower than the mean, it can be assumed that most of the observations are tilted toward the lower end of the distribution. In this case, **Figure 17** implies that most firms scored very low on EMM, INO, RES, HR, PRD, and CSR. Conversely, scores for the categories COM, WF, MNG, and SH tend to be evenly spread around the median value (**Figure 18**).

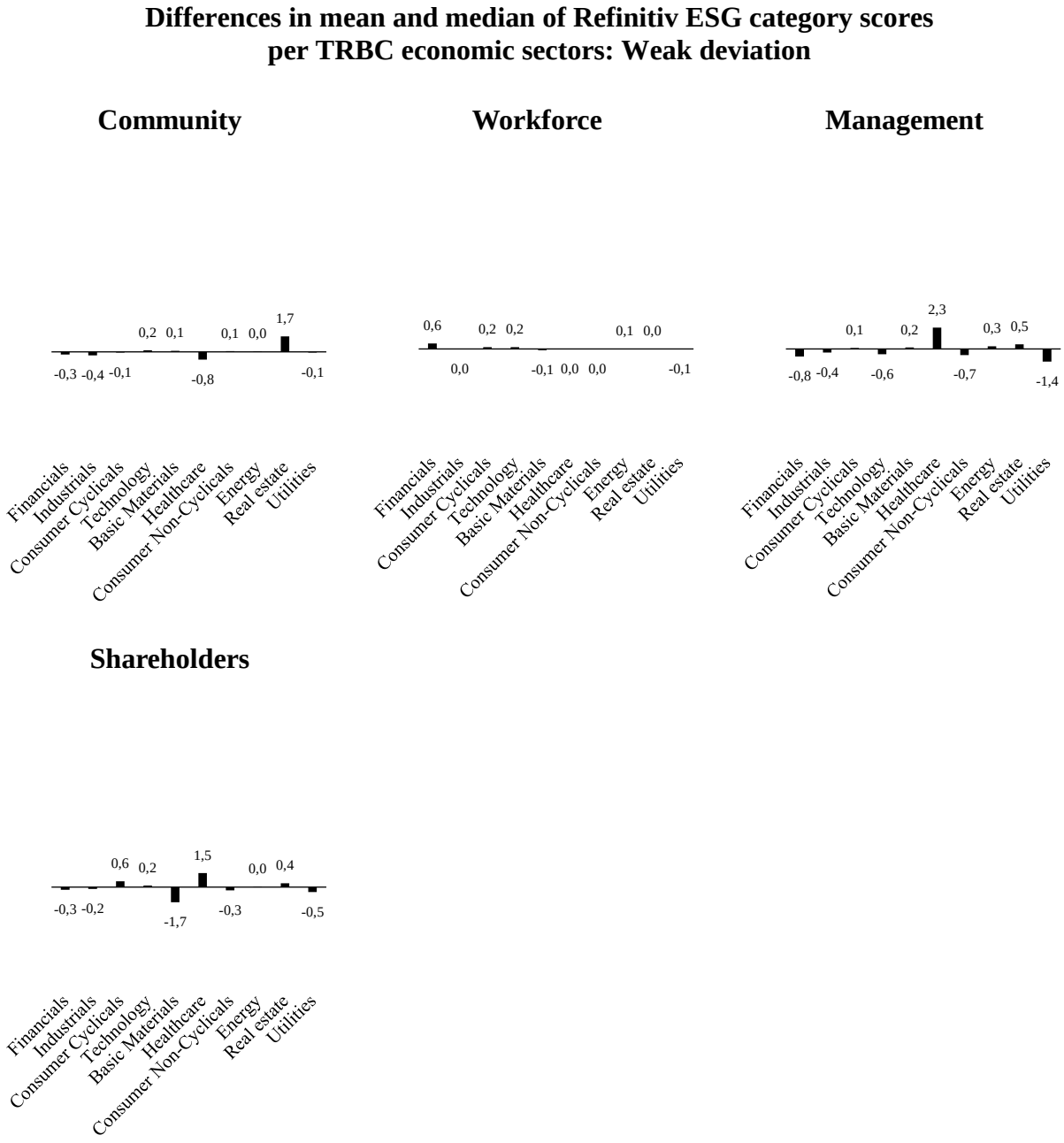


Figure 18 Differences in mean and median of Refinitiv ESG category scores per TRBC economic sectors: Weak deviation

ESG score distribution and standardization

In the previous chapter, the mean and median of the category scores within the dataset imply that there are tilts in the distribution of Refinitiv's ESG metrics. The tilt toward one end of the distribution comes as a surprise. As the scores are based on percentile rankings, the distribution is expected to be flat. **Figure 19** and **Figure 20** show the distribution and density plots of firm-year category scores within the dataset for each individual category score. COM, WF, MNG, and SH category scores show the expected uniform distribution. Each performance decile of the intervals indicates an almost identical distribution of ten percent of the observations. Correspondingly, the dashed black lines indicate the mean values of category scores to be almost 50.

Distribution, density plots, and means of category scores with a uniform distribution

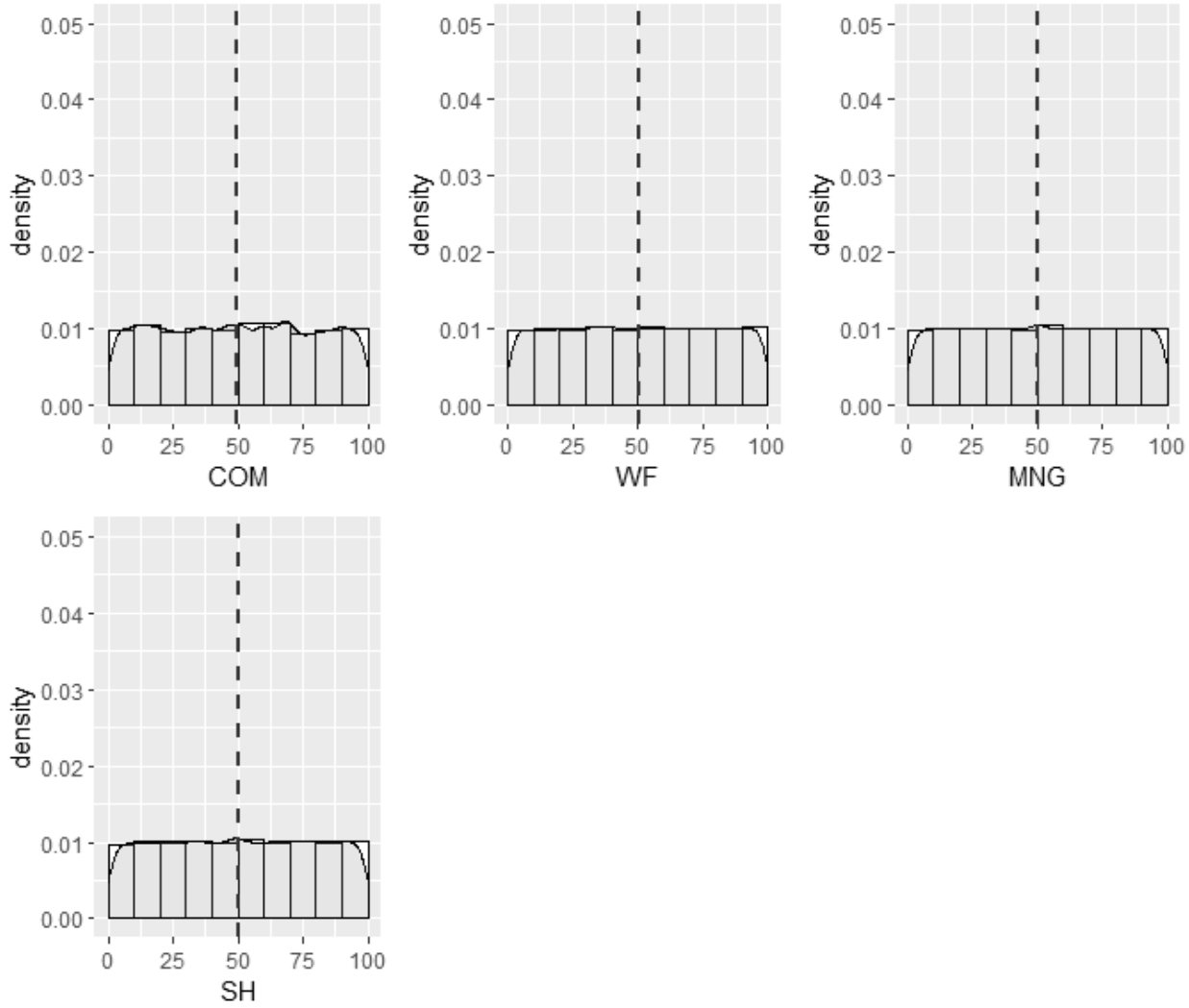


Figure 19 Distribution, density plots, and mean of category scores with a uniform distribution

The distributions of EMM, INO, RES, HR, PRD, and CSR category scores show a strong tilt toward low assessments. The number of firm-year observations indicating a score of 5 or lower

ranges from roughly a quarter (PRD) to more than 70 percent (INO) of all firms in the dataset. The question arises whether the rating is able to sufficiently address actual performance differences between the firms in the dataset in these categories.

Distribution, density plots, and means of category scores with tilted distribution

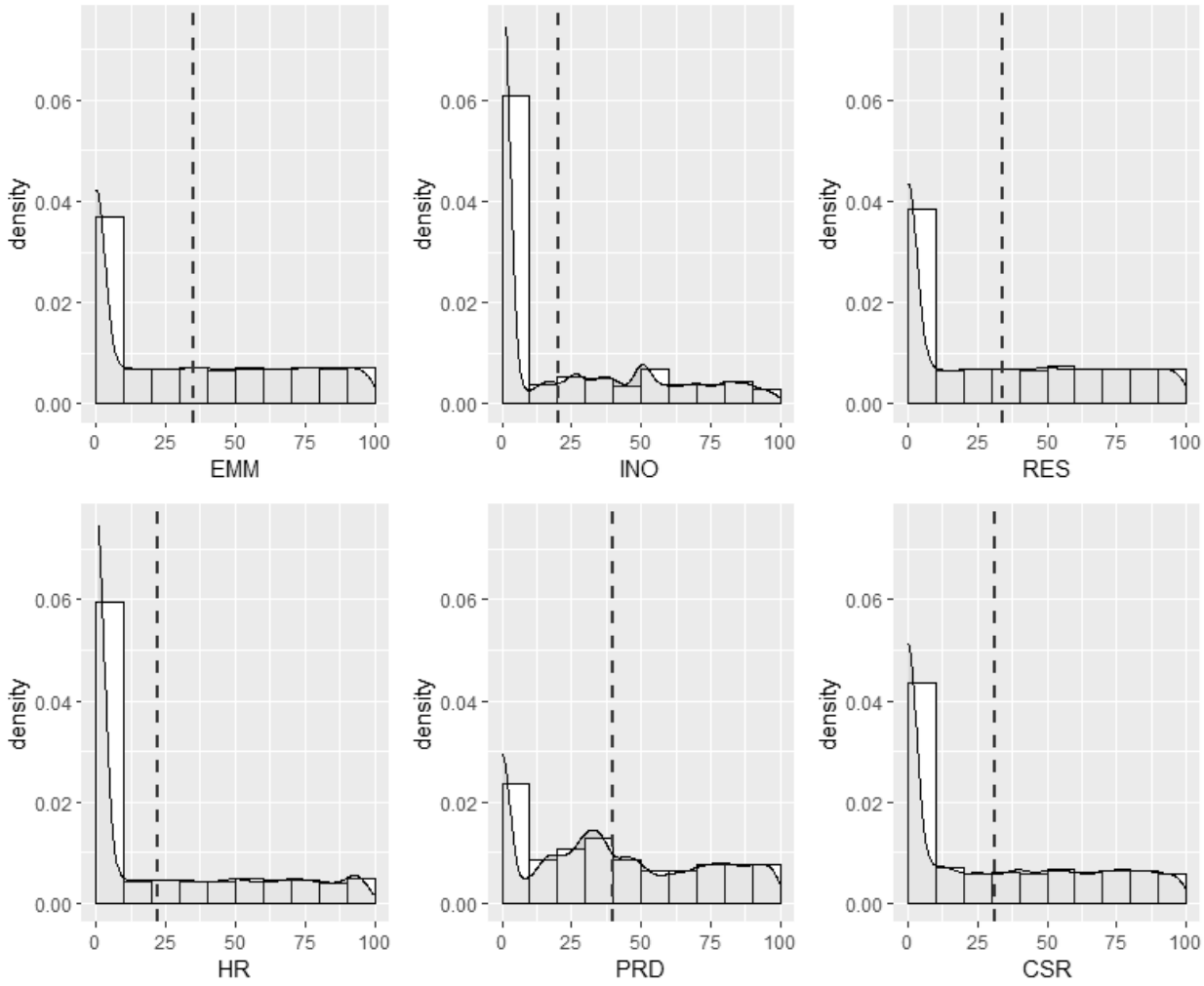


Figure 20 Distribution, density plots, and mean of category scores with tilted distribution

Mathematically the distribution of scores in that manner is implausible. Calculating a score of zero following Refinitiv's percentile ranking formula (see chapter 5.3.1) is impossible. However, the high fraction of firms being rated low results from Refinitiv's practice of assigning the worst possible performance if the company does not disclose respective information. The results suggest that firms are added to the database even when the available information is insufficient in providing data for all category scores. Missing data points are intentionally set to zero, indicating the worst possible performance.

The impact on the number of available scores is severe. **Table 7** shows the number of firm-year observations in the full sample per pillar and ESG category score affected by intentionally added

zeros. Recalling the ranking formula of Refinitiv, the number of firms that are performing worse (N_{worse}) is exaggerated by this practice and biases the performance of other firms positively. Due to the aggregation of percentile rankings in the pillar and ESG overall score, this bias also affects the scores' superior levels. As a result, at least 75% of all ESG overall scores in the Refinitiv ESG universe are affected by ESG information intentionally assigned the worst possible performance. Nevertheless, as percentile ranks are based on within-industry (or within-country) comparisons, the set of firms assigned zeros remains stable within a calculation cycle. Although the fraction of zeros might differ and impose biases of different magnitudes, the calculation of each score within a benchmark group should be relatively constant and bias each firm's score to the same extent.

One can easily spot that the number of zeros in the E and S pillar largely exceeds the affected firm-year observations in the G pillar (**Table 7**). It can be assumed that information on data points underlying the MNG and SH scores is fairly common in corporate disclosure (e.g., when prescribed by accounting regulation). It leads to firms being initially added to the ESG score universe. As the information on other ESG categories might not be that common to report, there is no other information disclosed that Refinitiv can use to calculate the additional scores. Hence, one can assume that the reporting patterns of firms and industries heavily influence the distribution of ESG scores.

Number and fraction of firm-year observations affected by intentionally added zeros

	Firm-year observations			
	N	**NA**	**0**	**Fraction of affected observations**
ESG	83,820	7	0	75.1%
E	83,688	139	16,744	63.6%
S	83,686	141	0	58.5%
G	83,801	26	0	38.0%
EMM	83,688	139	25,365	30.3%
INO	83,590	237	49,526	59.2%
RES	83,688	139	26,333	31.5%
COM	83,712	115	346	0.4%
HR	83,686	141	46,256	55.3%
PRD	83,686	141	17,536	21.0%
WF	83,712	115	0	0.0%
CSR	83,827	0	31,882	38.0%
MNG	83,827	0	0	0.0%
SH	83,827	0	0	0.0%

Table 7 Firm-year observations affected by intentionally added zeros (full dataset)

Given that the regression models conducted in this study rely on mean values, the potential of tilts toward one end of the distribution to drive regression parameters is relatively strong. To account for the empirical peculiarity, the ESG metrics provided by Refinitiv are standardized following a "z-score" transformation. Z-scoring is a common adjustment to ranking data and allows for the comparison of values from different distributions considering each sample's individual mean

value.[649] Z-scores are calculated (**Equation 22**) based on the mean and the empirical standard deviation of each ESG pillar and category score ($CategoryScore_t$). As the ESG metrics are based on yearly industry- or country benchmarks, this study follows Berg, Fabisik, and Sautner (2020) and standardizes the data on yearly sub-samples of the dataset. Thereby, different means and standard deviations in each year are addressed.[650]

$$zCategoryScore_t = \frac{CategoryScore_t - \overline{CategoryScore_t}}{std(CategoryScore)_t}$$

Equation 22 Calculation of standardized ESG scores (z-scores)

Scores of each pillar and category are set to a mean of 0 and a standard deviation of 1. Scores of different distributions are comparable, even though they still follow the same distribution as before the transformation.[651] Regression coefficients are interpreted as the effect of a one standard deviation change in the corresponding ESG score.

Size effects

Refinitiv ESG ratings are constructed based on percentile rankings. Each category score can be interpreted as the percentage of firms with the same or lower performance.[652] As a potential downside, one cannot expect actual ESG quality to be equally spaced within the rating scale (see the discussion in the previous chapters). In order to address this specific peculiarity, percentile rankings are based on the industry group (E, S) or the firm's country of domicile (G). Benchmarking avoids – by construction – unproportioned rating tilts toward less controversial industries or regulatory frameworks of different (potentially higher) ethical standards. The remaining differences in ratings are expected to originate solely from structural distinctions between various industries or countries, as the underlying information is comparable across the respective context.[653]

However, one primary concern in ESG literature is the effect of firm size on ESG ratings. As one expects larger firms to experience higher visibility, a certain degree of ESG engagement might be driven by stakeholder pressure rather than a genuine commitment to ESG objectives and beliefs. On the contrary, smaller firms might be restricted in financial resources to engage in higher levels of ESG performance.[654] Drempetic, Klein, and Zwergel (2019) focus on the ASSET 4 database (which is an older set of the same ESG rating universe used for this study) and find evidence for a

[649] Abdi (2007, p. 1055)
[650] cf. Berg/Fabisik/Sautner (2020, p. 21)
[651] cf. Abdi (2007, p. 1056)
[652] cf. Refinitiv (2022a, p. 3)
[653] cf. van Deth (2003, pp. 291-292)
[654] cf. Artiach et al. (2010, p. 32), cf. Windolph (2011, p. 42)

significant influence of firm size, a company's available resources, and the extent to which a company is publishing data on its sustainability performance.[655] Hence, they propose thoroughly investigating how ESG ratings work before using them.[656]

Figure 21 depicts the average ESG, E, S, and G pillar scores of the firms making up the respective firm size decile. The black line represents overall ESG performance scores, the green (E), red (S), and blue (G) the respective pillar scores. Throughout the entire sample (FULL SAMPLE), ESG ratings seem positively associated with firm size. Regardless of the more granular regional subsamples, firms within the last decile show the highest ESG scores. Further, the distance between the lines shrinks with increasing firm size. Especially environmental pillar scores within the European sample show this pattern. As firm size tends to be not equally related to ESG scores, the heterogeneity plots indicate a potential measurement bias.

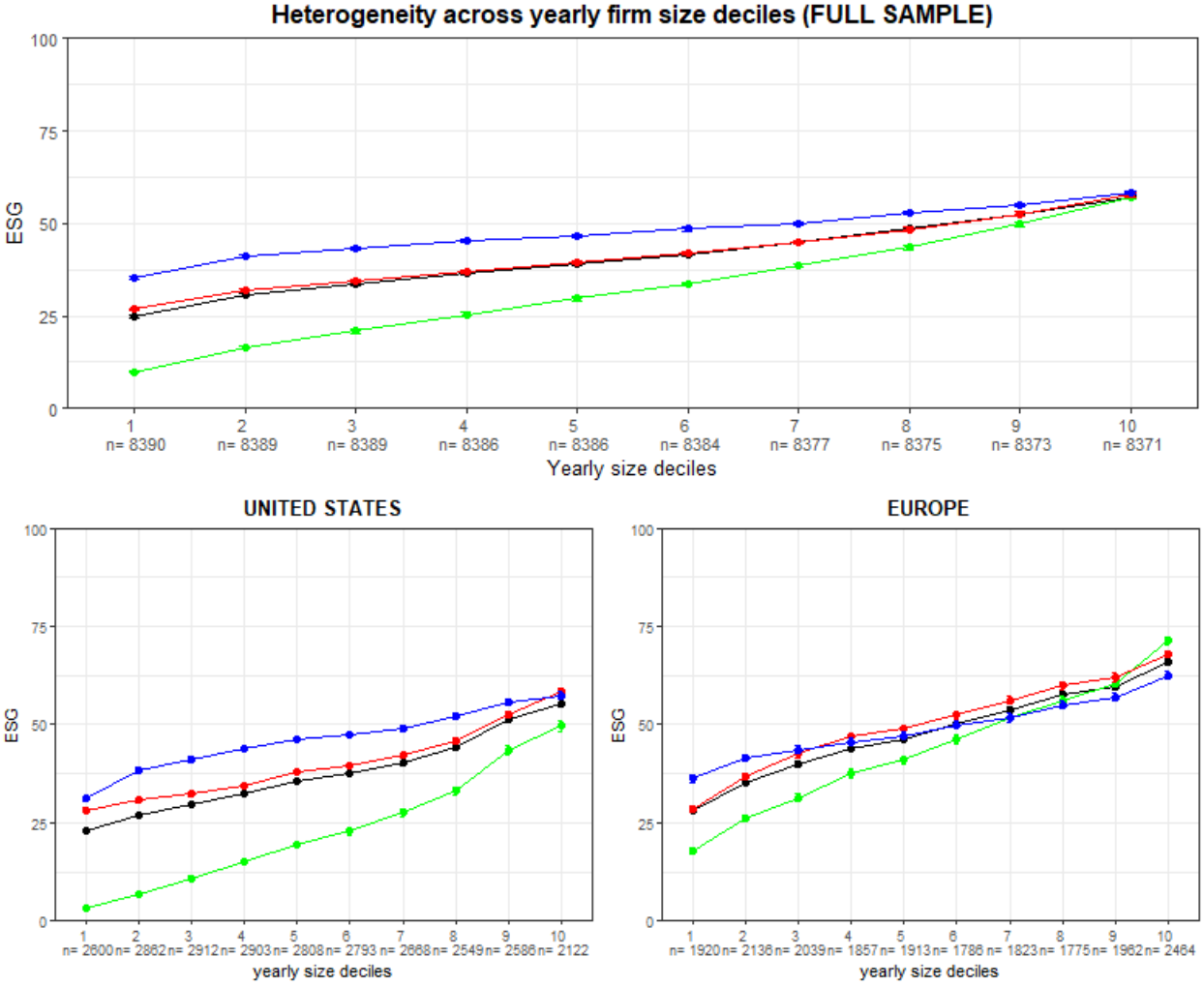


Figure 21 ESG, E, S, and G score heterogeneity across firm size deciles

[655] ASSET4 is the predecessor database of the ESG ratings used within this study.
[656] cf. Drempetic/Klein/Zwergel (2019, pp. 354-355)

In order to get a deeper understanding of whether firm size systematically affects Refinitiv ESG performance scores, this study follows Chemmanur et al. (2019) and runs the following fixed effect panel regression models on each ESG rating and pillar score for each firm *i* within year *t*:[657]

$$ESGscore_{i,t} = SIZE_{i,t} + {SIZE_{i,t}}^2 + IndustryFE_i + CountryFE_i + YearFE_t + \varepsilon_{i,t}$$

Equation 23 Regression model: Test of firm size-, industry-, country-, and year-fixed effects in Refinitiv ESG performance scores

Industry-fixed effects based on the TRBC industry group, country-fixed effects, and year-fixed effects are included in varying specifications.

The regression results in **Table 8** confirm that Refinitiv ESG measures strongly relate to firm size. The negative coefficients of *firm size²* indicate that the relationship becomes weaker with increasing firm size. Controlling for within-industry variation alleviates the relationship. ESG category scores are aggregated into pillar scores by assigning a specific weight to them (see **Figure 14**). The weight represents the relative importance of a particular issue to the firms within the respective industry (i.e., materiality). As chapter 5.3.1 shows, the importance is based on the relative rate of disclosure (for Boolean data points) or the industry median (for numeric data points). Both measures are empirically found to be strongly related to firm size. Larger firms tend to disclose more information that is captured by scoring-relevant data points. The higher rate of disclosure is either driven by their ability or the mere necessity, as disclosure provisions are often based on a firm's size or the number of employees. Consequently, larger firms within an industry are comparable in context, but the information underlying a rating differs systematically by variables correlated with firm size.

The results pertain when controlling for within-country variation. Governance scores within countries positively relate to firm size at a decreasing rate. The Governance score is primarily based on regulated corporate action that might differ at the country level. Therefore, data within the Governance scores is benchmarked with firms in the same country of domicile. As pointed out above, regulations often differ based on a firm's size, with larger firms facing stricter disclosure regulations. The rating methodology seems to be unable to account for this peculiarity. In the four specifications controlling for both industry-fixed and country-fixed effects, the relationships remain clear: ESG measures are positively correlated to firm size.

[657] cf. Chemmanur et. al. (2019, p. 2396)

To cope with the unproportioned and systematic large-cap tilt within the ESG measures utilized for this study, all ESG measures are adjusted based on a simple and comprehensible standardization approach.[658] In the first step, firms are sorted into yearly firm size deciles based on the total values of assets reported on their balance sheets at a calendar- or fiscal year-end. Then, standard scores are calculated (see **Equation 24**) based on the sub-sample mean and the empirical standard deviation of the ESG measures within the corresponding firm size decile (*I*).

$$zadjESG_{i,t} = \frac{ESG_{i,t} - \overline{ESG}_{I,t}}{std(ESG)_{I,t}}$$

Equation 24 Calculation of ESG scores standardized by firm size

[658] A similar approach can be found in Kaiser (2020, p. 37)

Refinitiv ESG scores and firm size-, industry-, country-, and year-fixed effects

	ESG SCORE				E PILLAR SCORE			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
firm size	**8.988*** **	**8.219*** **	**10.511*** **	**10.580*** **	**10.871*** **	**7.191*** **	**11.446*** **	**9.063*** **
	t = 17.947	t = 13.976	t = 22.079	t = 18.690	t = 15.567	t = 8.487	t = 17.501	t = 11.298
firm size²	**-0.228*** **	**-0.115** **	**-0.295*** **	**-0.217*** **	**-0.194*** **	0.097	**-0.222*** **	0.006
	t = -7.262	t = -3.087	t = -10.174	t = -6.166	t = -4.453	t = 1.798	t = -5.584	t = 0.112
CountryFE	No	No	Yes	Yes	No	No	Yes	Yes
IndustryFE	No	Yes	No	Yes	No	Yes	No	Yes
YearFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	83,820	83,820	83,820	83,820	83,688	83,688	83,688	83,688
R^2	0.224	0.273	0.317	0.373	0.250	0.325	0.368	0.434
Adjusted R^2	0.224	0.273	0.316	0.372	0.250	0.325	0.367	0.433

	S PILLAR SCORE				G PILLAR SCORE			
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
firm size	**7.482*** **	**7.904*** **	**9.633*** **	**10.860*** **	**7.392*** **	**6.429*** **	**9.081*** **	**8.517*** **
	t = 13.099	t = 12.172	t = 18.007	t = 17.560	t = 14.077	t = 11.141	t = 17.041	t = 14.591
firm size²	**-0.152*** **	**-0.086***	**-0.247*** **	**-0.209*** **	**-0.225*** **	**-0.147*** **	**-0.295*** **	**-0.231*** **
	t = -4.302	t = -2.111	t = -7.706	t = -5.516	t = -6.939	t = -4.145	t = -9.082	t = -6.497
CountryFE	No	No	Yes	Yes	No	No	Yes	Yes
IndustryFE	No	Yes	No	Yes	No	Yes	No	Yes
YearFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	83,686	83,686	83,686	83,686	83,801	83,801	83,801	83,801
R^2	0.160	0.212	0.291	0.356	0.087	0.115	0.105	0.136
Adjusted R^2	0.159	0.211	0.290	0.354	0.087	0.115	0.103	0.134

Notes: ESG, E, S, and G pillar scores are ESG metrics as provided by Refinitiv. Firm size (in $US million) is the natural logarithm of total assets reported on the balance sheet. Industry groups, years, and countries are fixed according to the notations for each specification. Values reported in local currencies are converted into USD. Variables are winsorized at the yearly 1%- and 99%-percentiles. Robust test statistics calculated with group-clustered standard errors are reported below coefficients. Significance levels are indicated as: * = 0.1, ** = 0.05, *** = 0.01.

Table 8 Regression results: Refinitiv ESG performance scores and firm size-, industry-, country-, and year-fixed effects

Final sample

Table 9 shows the number of firm-year observations, mean, standard deviation, 25th and 75th percentile, the median, and the minimum and maximum value of all ESG metrics in the dataset. The second and third parts of the table list the values after z-scoring and the corresponding adjustment on firm size deciles (see above).

Final sample of Refinitiv ESG performance scores and standardized scores

Full sample (2001-2021)

variable	N	mean	sd	25th-percentile	median	75th-percentile	min	max
ESG score	83,820	40.90	20.44	24.50	38.61	56.11	0.19	95.15
Controversy score	83,714	92.32	21.03	100.00	100.00	100.00	0.44	100.00
Combined ESG/controversy score	83,820	39.62	19.33	24.30	37.82	53.62	0.19	94.64
Environmental pillar score	83,688	32.48	28.55	4.23	26.70	55.78	0.00	99.14
Social pillar score	83,686	41.50	23.22	23.04	38.78	58.55	0.05	98.63
Governance pillar score	83,801	47.60	22.73	29.17	47.66	65.87	0.11	99.53
Emissions	83,688	35.00	33.34	0.00	28.65	64.39	0.00	99.91
Innovation	83,590	20.41	29.55	0.00	0.00	39.29	0.00	99.87
Resource use	83,688	34.34	33.30	0.00	27.23	63.48	0.00	99.90
Community	83,712	49.88	28.76	24.88	50.00	74.35	0.00	99.94
Human rights	83,686	22.40	31.40	0.00	0.00	44.59	0.00	99.52
Product responsibility	83,686	39.59	31.33	12.50	34.96	66.94	0.00	99.94
Workforce	83,712	50.22	28.79	25.39	50.00	75.12	0.10	99.94
CSR strategy	83,827	31.03	33.09	0.00	18.80	59.70	0.00	100.00
Management	83,827	50.14	28.75	25.16	50.00	75.00	0.02	99.99
Shareholders	83,827	50.17	28.74	25.24	50.00	75.00	0.01	99.99

Full sample (2001-2021): z-score

variable	N	mean	sd	25th-percentile	median	75th-percentile	min	max
ESG	83,820	0.000	1.000	-0.801	-0.106	0.745	-2.087	3.466
Environmental pillar score	83,688	0.000	1.000	-0.934	-0.195	0.813	-1.302	3.231
Social pillar score	83,686	0.000	1.000	-0.790	-0.111	0.736	-2.136	3.427
Governance pillar score	83,801	0.000	1.000	-0.811	0.004	0.805	-2.129	2.428
Emissions	83,688	0.000	1.000	-0.981	-0.204	0.875	-1.249	2.492
Innovation	83,590	0.000	1.000	-0.720	-0.603	0.614	-0.786	5.224
Resource use	83,688	0.000	1.000	-0.978	-0.229	0.868	-1.159	2.526
Community	83,712	0.000	1.000	-0.871	0.006	0.853	-1.853	1.827
Human rights	83,686	0.000	1.000	-0.724	-0.551	0.634	-1.076	4.873
Product responsibility	83,686	0.000	1.000	-0.747	-0.240	0.855	-1.864	3.923
Workforce	83,712	0.000	1.000	-0.862	-0.004	0.866	-1.890	1.859
CSR strategy	83,827	0.000	1.000	-0.877	-0.378	0.841	-1.131	3.578
Management	83,827	0.000	1.000	-0.867	0.001	0.866	-1.775	1.902
Shareholders	83,827	0.000	1.000	-0.865	-0.003	0.866	-1.835	1.950

Full sample (2001-2021): z-score based on yearly firm size deciles

variable	N	mean	sd	25th-percentile	median	75th-percentile	min	max
ESG	83,820	0.000	0.999	-0.757	-0.034	0.730	-3.145	4.926
Environmental pillar score	83,688	0.000	0.999	-0.802	-0.159	0.755	-2.263	8.359
Social pillar score	83,686	0.000	0.999	-0.762	-0.053	0.726	-2.797	4.336
Governance pillar score	83,801	0.000	0.999	-0.787	0.018	0.794	-2.747	3.149
Emissions	83,688	0.000	0.999	-0.791	-0.197	0.798	-2.535	6.932
Innovation	83,562	0.000	0.999	-0.673	-0.400	0.553	-1.663	9.853
Resource use	83,688	0.000	0.999	-0.816	-0.219	0.792	-2.123	7.388
Community	83,712	0.000	0.999	-0.860	0.058	0.833	-2.304	2.865
Human rights	83,672	0.000	0.999	-0.691	-0.398	0.582	-1.772	8.306
Product responsibility	83,672	0.000	0.999	-0.774	-0.176	0.812	-2.121	4.189
Workforce	83,712	0.000	0.999	-0.822	0.015	0.832	-2.950	2.910
CSR strategy	83,813	0.000	0.999	-0.759	-0.324	0.764	-2.320	6.929
Management	83,827	0.000	0.999	-0.847	0.007	0.850	-2.324	2.723
Shareholders	83,827	0.000	0.999	-0.861	0.001	0.861	-2.166	2.132

Table 9 Descriptive statistics: Final sample of Refinitiv ESG performance scores and standardized scores

Concluding remarks

Several features of Refinitiv's ESG database have been described and analyzed throughout this chapter. The most important findings and their implications for further analysis are listed below:

1. Most of Refintiv's ESG data is updated once a year, in line with the annual reporting patterns of the underlying firms. However, Refinitiv does not systematically account for the period ESG data refers to. When fiscal year end dates fall in a month prior to the end of the year, the published information might also refer to the previous calendar year. This peculiarity eventually introduces confusion within the yearly rating methodology, as ESG scores are only allocated to one specific year. Hence, it cannot be ruled out that the best-in-class approach benchmarks ESG performance, referring to different years.
2. ESG scores are recalculated every week to incorporate new information as soon as it becomes available. Due to the best-in-class approach, however, new data also affects the scores of firms not reporting new information. Yearly reporting patterns alleviate this concern to some extent.
3. There are severe differences in industry distribution. To alleviate the effect of unobserved heterogeneity across industries, panel regressions are run with industry-fixed effects. The relative fractions of observations per industry group within business and economic sectors are almost even, which allows for an aggregated analysis on the economic sector level.
4. Common reporting practices of firms provide information to assign MNG and SH category scores to each firm within the Refinitiv ESG database. Once a firm is added, however, all other data points that are not fed by any source of corporate disclosure are set to zero. As the relative benchmarks incorporate those data points, companies that disclose more information on relevant data points systematically perform better. As a result, *transparency biases* distort the assessment within benchmark groups. Further, when comparing companies rated in different groups, the number of data points intentionally assigned a value of 0 may differ and dilute comparability across scales of ratings over several industries. To mitigate the impact of transparency biases on distributions, this study standardizes the original ESG metrics following a "z-score" transformation. However, a slight bias stemming from systematically different reporting patterns (e.g., due to materiality and regulatory frameworks) is very likely to remain.
5. Larger firms within an industry are comparable in context, but the extent of information publicly available to feed a scoring methodology differs systematically by variables correlated with firm size (e.g., reporting regulation, financial resources, and media coverage). Refinitiv rating ESG scoring methodology seems unable to account for the resulting *size bias*. Hence, this study adjusts Refinitiv's original ESG scores by a z-transformation on yearly firm size deciles.

5.4.2 Data on ESG controversies

Data on the number and occurrence of ESG controversies is retrieved from the Refinitiv ESG database. Recent literature finds that ESG performance metrics of different providers tend to vary primarily based on measurement differences and scope divergence.[659] Hence, applying controversy data and ESG performance metrics of the same provider allows for the closest fit of conceptualization of ESG performance measures and ESG controversy categories.

Refinitiv ESG controversies are based on real-time and historical news analysis. Although Refinitiv only sparsely provides information about the methodology and sources examined to collect and assign ESG controversies, a significant part of news data likely stems from Reuters News.[660] Reuters News belongs to the most prominent news agencies on the globe.[661] As a result, ESG-related news are compiled from over 10,500 news sources (e.g., news wires, social media, and third-party agencies) in 48 languages.[662] News coverage considered material to firms within the ESG rating universe are screened for information either relevant to the ESG score calculation or whether they signal controversial business conduct.[663]

Refinitiv offers 23 individual data points, each referring to a specific group of controversial business conduct that leads to news coverage.[664] Twenty-two of the controversy data points (**Table 10**) reflect count data. For each firm, the database shows the number of controversies and the date when the underlying topic first became public. Hence, this study is able to assign a number of controversies to firms within the original dataset that occurred during a given year. Another data point, *STRIKES*, reflects the occurrence of a workers' dispute or strike during a given year. The corresponding data point is binary (i.e., it equals one if a strike occurred) and is analyzed separately throughout this study (see chapter 5.2 for further information on the applied logit models).

Based on controversy data points, Refinitiv also offers an ESG controversies score. As described in the previous chapters, ESG performance scores reflect business conduct predominately at the input level (see chapter 3.3.3). Most data points within the base ESG scoring model reflect policies and programs implemented by firms to ensure a particular type of ESG-conscious business conduct or serve a particular ESG objective. Investigating how accurate an input-level assessment reflects

[659] cf. Berg/Kölbel/Rigobon (2019, pp. 30-31)
[660] cf. Refinitiv (2022a, p. 4), https://wealthtechtoday.com/2022/07/14/ep-150-how-to-avoid-noise-in-esg-data-with-elena-philipova-refinitiv/, accessed: 08/06/2022.
[661] https://www.faz.net/aktuell/wirtschaft/thema/reuters, accessed: 05/11/2022.
[662] https://www.refinitiv.com/en/financial-data/financial-news-coverage/refinitiv-news-coverage#feature-and-benefits, accessed: 08/06/2022.
[663] cf. Refinitiv (2022a, p. 4)
[664] cf. ibid. (p. 23)

the occurrence of controversies ex-ante allows for evaluating how well the policies and programs in place actually prevent controversial business actions. As the study focuses primarily on this predictive capability of ESG metrics, the incorporation of ESG controversies, which alter the performance measure as soon as they occur, could foster drawing inconclusive inferences. Hence, controversy scores are not covered in greater detail within this study.[665] Instead, the empirical investigation focuses on the underlying number of controversies grouped by the same categories as the Refinitiv ESG performance scores described in the previous chapter. This matching allows for an integrated investigation throughout the remainder of this study. Simultaneously, the occurrence of controversies is signaling business conduct on the outcome level.

Table 10 shows the number of firms, firm-year observations, and controversies per controversy data point retrieved from Refinitiv Datastream. Each data point refers to one of the three ESG pillars and is assigned to a distinct ESG category matching the ESG category score scheme described in chapter 5.3.1. The number of firms covered per data point ranges from 80 ("Child Labor Controversies") to 1,313 ("Anti-Competition Controversies"). Given the scope of the Refinitiv rating universe, individual ESG controversy data points only cover a maximum of about a tenth of the firms within the dataset per year. Over all ESG controversy data points, the sample contains 3.169 unique firms experiencing controversial media coverage at least once. The number of controversies ranges from 67 ("Controversies Responsible R&D") to 6,015 ("Business Ethics Controversies"). The average number of controversies per firm ranges from 1.35 ("Controversies Product Access") to 4.65 ("Business Ethics Controversies"). The maximum number of controversies assigned to a specific firm in one year lies between 3 and 60. The latter, for example, refers to Anti-Competition controversies of "Alphabet Inc." (US02079K3059) in the year 2020.

[665] Refinitiv (2022a, p. 14) gives a detailed explanation of how Refinitiv uses controversy data points for score calculations.

Number of controversies per ESG category

Pillar	Category	Controversy data point	Firms	Number of firm-year observations	Number of controversies	Controversies/firm	Number of controversies (max)	Number of 0 controversies observations	Fraction of 0 controversies observations
E	RES	Environmental	438	587	1,030	2.35	16	8	1.4%
S	COM	Anti-Competition	1,313	3,344	5,702	4.34	60	53	1.6%
		Business Ethics	1,301	3,397	6,015	4.62	46	92	2.7%
		Critical Countries	122	186	193	1.58	3	14	7.5%
		Intellectual Property	642	1,508	2,389	3.72	21	56	3.7%
		Public Health	349	621	741	2.12	8	34	5.5%
		Tax Fraud	343	617	862	2.51	19	10	1.6%
S	HR	Child Labor	80	119	107	1.34	3	19	16.0%
		Human Rights	202	370	372	1.84	8	47	12.7%
S	PRD	Consumer Complaints	611	1,391	1,934	3.17	17	34	2.4%
		Customer Health & Safety	474	1,144	2,033	4.29	65	39	3.4%
		Privacy	536	896	1,472	2.75	53	32	3.6%
		Product Access	98	126	132	1.35	3	12	9.5%
		Responsible Marketing	498	1,109	1,558	3.13	13	36	3.2%
		Responsible R&D	45	64	67	1.49	4	11	17.2%
S	WF	Diversity and Opportunity	418	821	1,005	2.40	7	43	5.2%
		Employees' Health & Safety	636	1,257	1,615	2.54	12	71	5.6%
		Wages Working Condition	661	1,410	1,946	2.94	11	9	0.6%
G	MNG	Management Compensation	281	428	453	1.61	5	25	5.8%
G	SH	Accounting	318	435	480	1.51	8	36	8.3%
		Insider Dealings	302	437	466	1.54	4	16	3.7%
		Shareholder Rights	601	916	1,345	2.24	26	1	0.1%

Table 10 Number of controversies per ESG category

In 2020, Alphabet was being investigated and facing trials in various countries regarding the abuse of Google's dominant market position. For example, the EU Commission filed a full investigation of Google's planned acquisition of "Fitbit", a producer of fitness-tracking wristbands.[666] In the US, Alphabet faced several lawsuits filed by the U.S. Department of Justice and over 30 different states.[667] Causes for complaint were potential monopoly concerns in web searching and online advertising,[668] as Google was accused of blocking other web search engines from applications in automobiles, smart TVs, smartphones, and other electrical appliances.[669] Refinitiv claims that "… no controversy is double-counted".[670] Considering the scope and variety of causes for Anti-Competition concerns, the number of ESG controversies (here, 60 in one year) seems reasonable. In fact, a detailed control of the sample data shows that the dates and content of Alphabet's Anti-Competition controversies in 2020 referred to considerably different cases. Nevertheless, the possibility of technically double-counting certain controversies cannot be ruled out, given how most cases relate to the same base situation (e.g., Google's dominant market position).

Appendix 10 shows potential cases of double-counting in the example of Google's acquisition of Fitbit. Although one observes the same incident (i.e., "Fitbit-deal"), the controversies refer to the legal actions in different countries or general responses of different actors (e.g., interest groups like the European Consumer Organisation). Refinitiv is likely trying to account for the regional scope of certain controversies by counting respective events separately. For example, three controversies account for legal action within the EU. Since the three EU controversies reflect individual responses of the European Consumer Organisation and the European Commission, the underlying controversial business conduct (i.e., Fitbit-deal) is the same; however, the actors exposing the controversies are considerably different. Thus, the degree of granularity is a question of design, which is eventually debatable in many cases.

The minimum number of controversies per firm is 0. This applies to 711 firm-year observations, which make up 0.1% to 17.2% of the firm-year observations within an individual data point. Refinitiv claims to set default values of zero when calculating the ESG controversy score.[671] As the data points refer to the number of controversies and are considered negatively for the ESG controversy scoring model, these values come as a surprise. The ESG controversy score, as it relies on a

[666] https://ec.europa.eu/commission/presscorner/detail/en/ip_20_2484, accessed: 05/11/2022.
[667] https://www.cnbc.com/2020/12/17/google-faces-a-third-government-antitrust-lawsuit.html, accessed: 05/11/2022.
[668] https://www.bbc.com/news/business-54619148, accessed: 05/11/2022.
[669] https://www.bbc.com/news/business-55357340, accessed: 05/11/2022.
[670] Refinitiv (2022a, p. 14)
[671] cf. ibid.

best-in-class screening, could technically be biased by observations indicating zero controversies, especially considering how strong the scope of media coverage and the count of controversies are supposedly interrelated. Observing no controversy in a given year might also be caused by lower media coverage of some firms and topics. Refinitiv acknowledges the potential size bias, as larger firms tend to attract more media coverage and puts a higher weight on controversies of smaller firms.[672] Nevertheless, one cannot assume that all firms that actually experienced no controversies in a given year are entirely observed. Hence, it is unclear how the data points ended up in the dataset and what they intend to identify. It might be the case that values were restated after being tracked with other values. The values might also be based on unintended or accidental data entries to the database in general.

The data point strike is a binary data point indicating whether or not a strike occurred during a given year. Unlike the other controversy data points, the number of firms (11,971) and firm-year observations (84,699) almost exactly match the dataset of ESG performance measures described in the previous chapter. However, the number of firm-year observations indicating no occurrence ("N") of a strike vastly exceeds the number of strikes ("Y") identified by Refinitiv. The sample consists of 1,462 firm-year observations referring to the occurrence of a strike within 644 individual firms. 83,237 firm-year observations indicate no strike within a given year. Thus, strikes seem to be a relatively rare event.

The U.S. Bureau of Labor Statistics (BLS) provides data for the United States on "Work Stoppages" that involve at least 1,000 workers for at least one shift. Within the observed time period of this study (2002-2021), the BLS lists 311 such incidences in 169 individual organizations.[673] The number of strikes tracked by the BLS almost exactly matches the 300 US firm-year observations and 143 firms in the sample retrieved for this study. Because the BLS also tracks governmental organizations (like school districts or county administrations), the number of firms in the sample becomes even more reasonable. The empirical investigation focuses on the relationship between past ESG performance and the probability of strikes occurring in the future. A fraction of the strikes in the dataset, however, might be caused by bargaining on wages that are part of periodical efforts to set new arrangements. These strikes are fairly predictable.[674] Nevertheless, this study does not further investigate the causes of strikes. As strikes are a negotiation measure of last resort, firms can alleviate the chance of strikes by engaging in efforts that address a timely and

[672] cf. Refinitiv (2022a, p. 14)
[673] https://www.bls.gov/wsp/, accessed: 05/12/2022.
[674] https://www.destatis.de/EN/Themes/Labour/Labour-Market/Quality-Employment/Dimension5/5_2_WorkingDaysLostThroughStrikesLockouts.html, accessed: 05/12/2022.

efficient agreement before strikes occur. An ESG rating capturing the employee relationships should be able to identify firms that might be interested in such timely solutions, which is why it can be assumed to see no predictable strike biasing the results of the empirical investigation.

Several peculiarities fall into sight when analyzing the sample dataset. On the one hand, values within certain controversy topics largely exceed the number of controversies in other categories. One can assume that the preferences of different news outlets partly drive the number of controversies within the dataset. Hence, an important finding seems to be that controversies related to Anti-competition and Business Ethics are prime concerns in the news coverage of ESG. On the other hand, it might be possible that the mapping of news items to certain data points also relates to the scope of potential business actions relating to the corresponding controversy category. Considering what might fall under controversial conduct in the area of business ethics, the data point might reflect a broader range of business actions than controversies relating to more minor scope actions in "Product Access" or "Child Labor". **Appendix 11** shows the "Business Ethics Controversies" of Volkswagen AG in 2017. Of 23 controversies, 20 refer to the Volkswagen emission scandal starting in 2015. It is unclear why these controversies are not accounted for, for example, in the "Environmental Controversies" data point. Nevertheless, the mapping seems to be systematic. An important consequence of this peculiarity is that the sample of ESG controversies is potentially biased by specific topics (e.g., "Business ethics") that serve as an umbrella category for a variety of different business actions.

In conclusion, the analysis shows potentially incorrect and methodically illogical values when indicating a number of "0" controversies. To alleviate the influence of zero-observations, values indicating zero controversies are dropped from the sample. Furthermore, in order to prevent the study from unproportioned tilts toward some controversy categories, the data points are grouped due to the corresponding ESG category and ESG pillar (see **Table 10**). The aggregation allows to get a sufficient number of observations within each category that matches the ESG performance measurement applied in this study.

Temporal distribution

The original data points described in the previous chapter cover the years 2002 to 2021. Refinitiv assigns controversial media coverage based on the date of publication and according to the respective fiscal year of the firm.[675] Hence, each controversy count matches the temporal distribution of the original ESG scores. This pattern, however, can only apply to historical data from already completed fiscal years. To account for the latest controversies linked to a firm Refinitiv introduced "recent controversies" data points that match the topics of the ("normal") data points listed above (see chapter 5.3.2 for further details). Controversies accounted for in the recent controversy data point are moved to the latest fiscal year as soon as the corresponding year is added to the rating universe. Their primary use, however, is capturing controversies that happened after the latest completed fiscal year within the calculation of the controversy score.[676] As mentioned above, this study does not consider the Refintiv ESG controversy score within its empirical analysis. However, the underlying recent data points are added to the original ("normal") data points described above. Thereby, the resulting dataset covers the original data for 2002 to 2021 and additional data on recent controversies assigned to the latest year, 2021. Further, this practice facilitates a timely separation of ESG performance measures and ESG controversy data. The induced lag is central to the empirical investigation, as the ESG performance measures are tested on how well they predict the occurrence of controversies happening in the future.

Temporal distribution: ESG controversies (2002-2021)

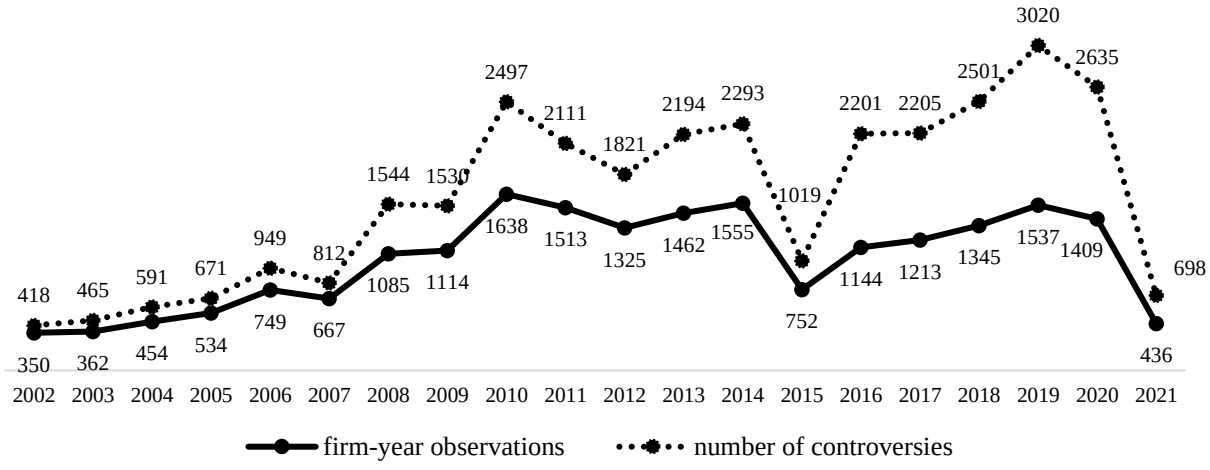


Figure 22 Temporal distribution: ESG controversies (2002-2021)

Figure 22 shows the temporal distribution of firm-year observations and the number of controversies over the observed time period (2002-2021). The sample consists of 3.221 individual firms,

[675] cf. Refinitiv (2022a, p. 14)
[676] cf. ibid. (pp. 14, 23)

indicating that most firms are observed over several years. The numbers of firm-year observations and controversies are moving almost synchronously. The growing gap between firm-year observations and the number of controversies, however, indicates that the average number of controversies per firm grew drastically every year starting in 2010. The dataset covers a period in which computational power and technical possibilities to evaluate news data evolved extensively. Thus, the growing number of controversies observed for each firm could partly be explained by a greater identification capability of the underlying ESG data-generating processes. As controversy data points are based on media coverage, the growing number of controversies in general and per firm also points toward a general increase in interest and awareness for ESG-related issues, which eventually attracts even more media coverage.

In 2015 and 2021, there are noticeable decreases in the number of firms and the number of controversies. Given the small number of observations already captured by the "normal" controversy data points (i.e., for each firm that has already completed the fiscal year) and the recent controversy data points, the number seems to be relatively low. Considering that fiscal years and calendar years might not be the same, one can assume that some controversies of the calendar year 2021 are accounted for in the previous year. Some firms might experience controversial news coverage in 2021 but before the end of their fiscal year. As a result, controversies are assigned to the previous year. One can assume the number of controversies for 2021 might grow as fiscal years end in 2022, and additional controversies are still being assigned to the previous year.

Refinitiv does not provide any information on the dip in 2015 and its potential causes. Hence, one can only try to infer circumstantial evidence. One of which might relate to major shifts to the product of Thomson Reuters ESG scores (later acquired by Refinitiv) itself. The controversy data points had already been part of the former ASSET4 rating universe, which Thomson Reuters acquired in 2009.[677] Since then, the dataset has grown extensively every year, as described in the previous chapter. In 2017, Thomson Reuters launched a new methodology to calculate ESG scores. The new scores, including a controversy score, were set to replace the former ASSET4 ESG scores by 2018.[678] When introducing the new scores in 2017, the fiscal year 2015 might have been the latest completed fiscal year for a significant fraction of the firms within the rating universe. It is unclear whether methodological and technical adjustments were necessary to introduce the new

[677] https://www.reuters.com/article/us-asset4-idUSTRE5AT0OW20091130, accessed: 05/12/2022.
[678] cf. Thomson Reuters (2017, p. 3), https://bizlib247.wordpress.com/2017/10/10/13-new-esg-scores-released-on-datastream/, accessed: 05/12/2022.

scores. Hence, these might have been the reason for restatements or critical adjustments to the data-generating processes.

Figure 23 shows the number of global and US strikes retrieved from the Refinitiv ESG rating universe and the number of strikes the BLS identifies in its data on work stoppages in the United States.[679] The global number of strikes fluctuates in line with the other controversy data points (as

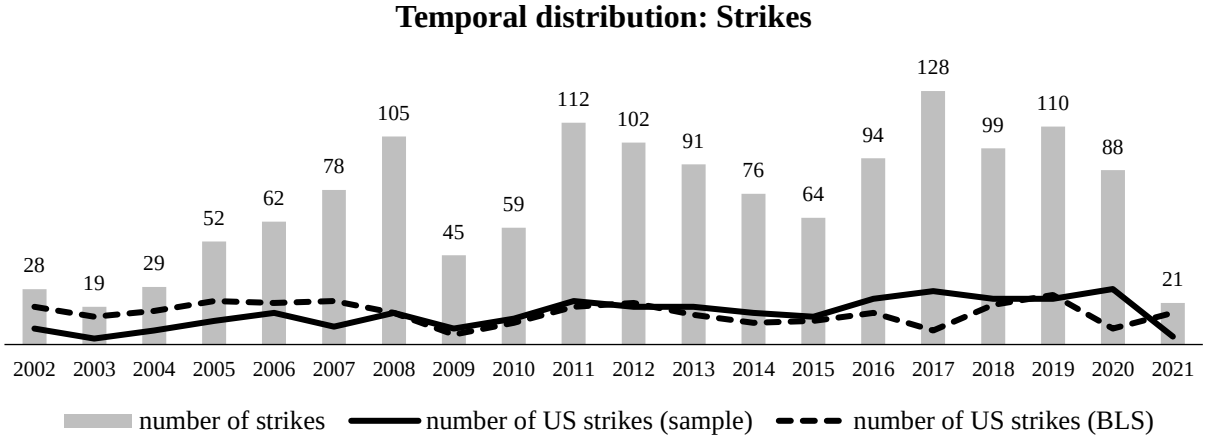


Figure 23 Temporal distribution: Strikes (2002-2021)

already described in the introduction of chapter 5.4.2). Noticeable dips can be observed for 2015 and 2021 and are in the previous paragraph. To further check on the robustness of the data, the number of US strikes as identified in the sample and the number of work stoppages the BLS reports are plotted separately (black and dashed lines). The number of strikes ranges roughly in the same scope. However, there are differences, especially in the years 2017 and before 2009.

Table 11 shows the US firms that experienced a strike in 2017 according to the Refinitiv sample and the BLS data. Only one of the firms (AT&T Inc.) is found in both datasets. The type of organizations the BLS lists, however, is different from the majority of the Refinitiv sample, which is solely based on listed firms. It is reasonable that the differences are rooted in the sensitivity to strikes, as the BLS only lists work stoppages of at least 1,000 workers per organization. Thus, the relatively high number of strikes in the Refinitiv sample might indicate a stronger sensitivity to even smaller strikes and reassure the usage of the corresponding data within this study. However, as the number of strikes in the Refinitiv sample is not consistently higher than the number of strikes identified by the BLS, other differences might cause the data to be incomparable. One of which might be the assignment to fiscal years, which do not necessarily match the calendar years. An-

[679] https://www.bls.gov/wsp/, accessed: 05/12/2022.

other reason might be a general difference in the definition of strikes identified by Refinitiv. Refinitiv does not publish any further information on the definition of strikes as the one described in chapter 5.3.2.

US strikes in 2017: Refinitiv and U.S. BLS data

Refinitiv	BLS
Alcoa	AT&T Inc.
Amazon	University of California
American Airlines	Charter Communications Inc.
AT&T	Tufts Medical Center
Bank of America	Chicago New Car Dealer Committee
Boeing	Riverside County Municipal Government
Chevron	City of Oakland
CNX Resources	
Comcast	
Electronic Arts	
Exxon Mobil	
Ford Motor	
General Motors	
Goodyear	
Hormel Foods	
Lowe's Company	
Macy's	
McDonald's	
Sanderson Farms	
Southern Copper	
Spirit Airlines	
Tesla	
Triumph Group	
United Parcel	
United States Steel	
Walmart	
Freeport-McMoRan	

Table 11 US strikes in 2017: Refinitiv and U.S. BLS data

Regional distribution

Figure 24 shows the distribution of firm-year observations grouped by the underlying regions of the firms' countries of domicile. Firms from the Americas (primarily the United States), Europe, and Asia make up almost all observations in the dataset (94.5%). The fraction of firm-year observation within each region is relatively constant. A slight decrease, however, can be seen for firms from the Americas and Europe. Simultaneously, firm-year observations from Asia, Oceania, and Africa gain a stronger representation over the observed period. The pattern underlines the generally expanding nature of the Refinitiv ESG rating universe. While the coverage of markets in the United States and Europe was the prime focus when introducing the ASSET4 scores in 2002, more indices were added year by year.[680]

[680] cf. Refinitiv (2022a, p. 5)

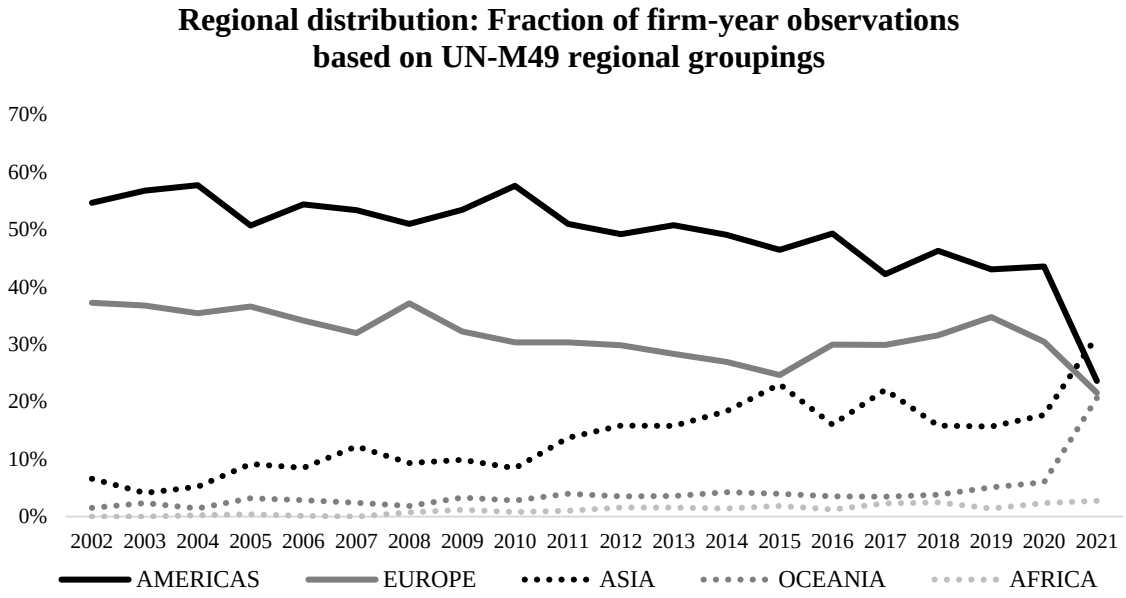


Figure 24 Regional distribution: Fraction of firm-year observations based on UN-M49 regional groupings

Similar to the ESG scores, controversy data points are back looking. Controversies, however, are directly added to the scoring methodology (as "recent controversy" data points when they occur). This facilitates a timely caption of adverse reputational effects eventually resulting from negative backlash following a controversy. Refinitiv claims to capture controversies over several years when there is an ongoing cause for the controversy.[681] However, the data cannot tell whether and to what extent historical controversies are added. Comparing the relative representation of firms in the ESG rating universe (i.e., the dataset described in the previous chapter) and the controversy data, however, exposes differences in the focus of the controversy caption. One can assume that firms of a region are "underrepresented" when the difference between the fractions within the ESG score dataset and the controversy dataset is negative; "overrepresented" when the difference is positive. As expected, **Figure 25** shows a strong presentation of US observations. From 2015, however, European firms tend to become more prevalent than US firms. Firm-year observations of Asian firms are consistently underrepresented.

[681] cf. Refinitiv (2022a, p. 7)

Difference in representation: ESG score sample vs. controversy dataset
Difference of fraction of firm-year observations within the ESG score sample (chapter 5.4.1) and the ESG controversy dataset

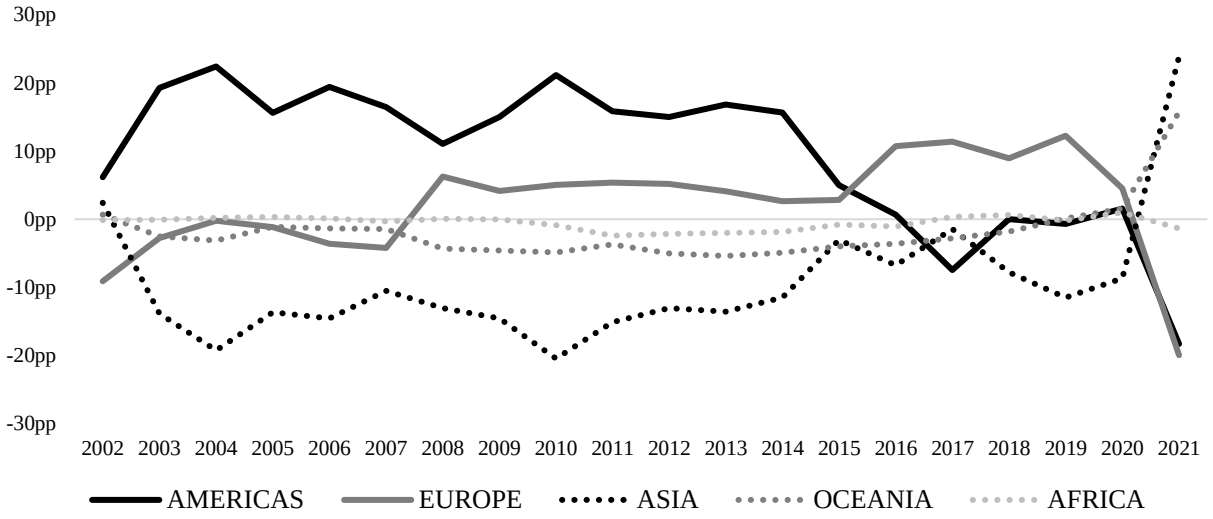


Figure 25 Difference in representation: ESG score sample vs. controversy dataset

As the empirical investigation focuses on two sub-samples of the ESG score dataset (i.e., United States and Europe), the final sample of controversy data points is slightly smaller than the dataset described up to this point. **Table 12** shows the firm-year observations and fraction of controversy data points considered for the final sample. As one can see, the regionally defined sample will cover almost three-quarters of all controversies Refinitiv provides in its controversy database.

Firm-year observations per country: ESG Controversies

Country	Firm-year obs.	Fraction of firm-year obs.
UNITED STATES	9,026	44.0%
UNITED KINGDOM	2,037	9.9%
GERMANY	936	4.6%
FRANCE	846	4.1%
SWITZERLAND	574	2.8%
NETHERLANDS	317	1.5%
IRELAND	276	1.3%
ITALY	260	1.3%
SWEDEN	210	1.0%
SPAIN	174	0.8%
FINLAND	111	0.5%
DENMARK	110	0.5%
NORWAY	85	0.4%
BELGIUM	82	0.4%
AUSTRIA	48	0.2%
GREECE	27	0.1%
REST	5,390	26.3%

Table 12 Firm-year observations per country: ESG Controversies

A fairly different pattern can be observed when looking at strike data within the dataset (**Table 13**). Although the two sub-samples still make up roughly two-thirds of all firm-year observations, the fraction of European firms is slightly higher. The most substantial difference, however, is observed for strike data from the US. The fraction of strikes compared to all firm-year observations

is less than half as big as the fraction over all controversies. Strikes have been very uncommon in the United States, especially during the last three decades.[682] Hence, it is pretty reasonable to see a supposedly low number of strikes observed in the US sample.

Firm-year observations per country: Strikes

Country	Firm-year obs.	Fraction of firm-year obs.
UNITED STATES	300	20.5%
UNITED KINGDOM	178	12.2%
FRANCE	117	8.0%
GERMANY	76	5.2%
FINLAND	51	3.5%
SWITZERLAND	32	2.2%
SWEDEN	32	2.2%
ITALY	31	2.1%
NETHERLANDS	30	2.1%
SPAIN	23	1.6%
GREECE	19	1.3%
IRELAND	14	1.0%
NORWAY	11	0.8%
BELGIUM	9	0.6%
DENMARK	5	0.3%
AUSTRIA	3	0.2%
REST	531	36.3%

Table 13 Firm-year observations per country: STRIKES

[682] https://www.epi.org/publication/2021-work-stoppages/, accessed: 05/13/2022, https://www.theguardian.com/commentisfree/2021/oct/19/america-strike-wave-john-deere-iatse, accessed: 05/13/2022.

Industry distribution

Figure 26 shows the number of controversies and firm-year observations grouped by TRBC economic sectors. Most controversies are recorded for the sectors "Financials" (5,851), "Technology" (5,748), and "Consumer-Cyclicals" (4,882). The smallest number of controversies is recorded for the sectors "Utilities" (951), "Real Estate" (185), and "Academic & Educational Services" (52). The number of controversies proportionally relates to the number of firm-year observations. The average of controversies per firm-year ranges from 1.18 ("Real Estate") to 1.82 ("Technology").

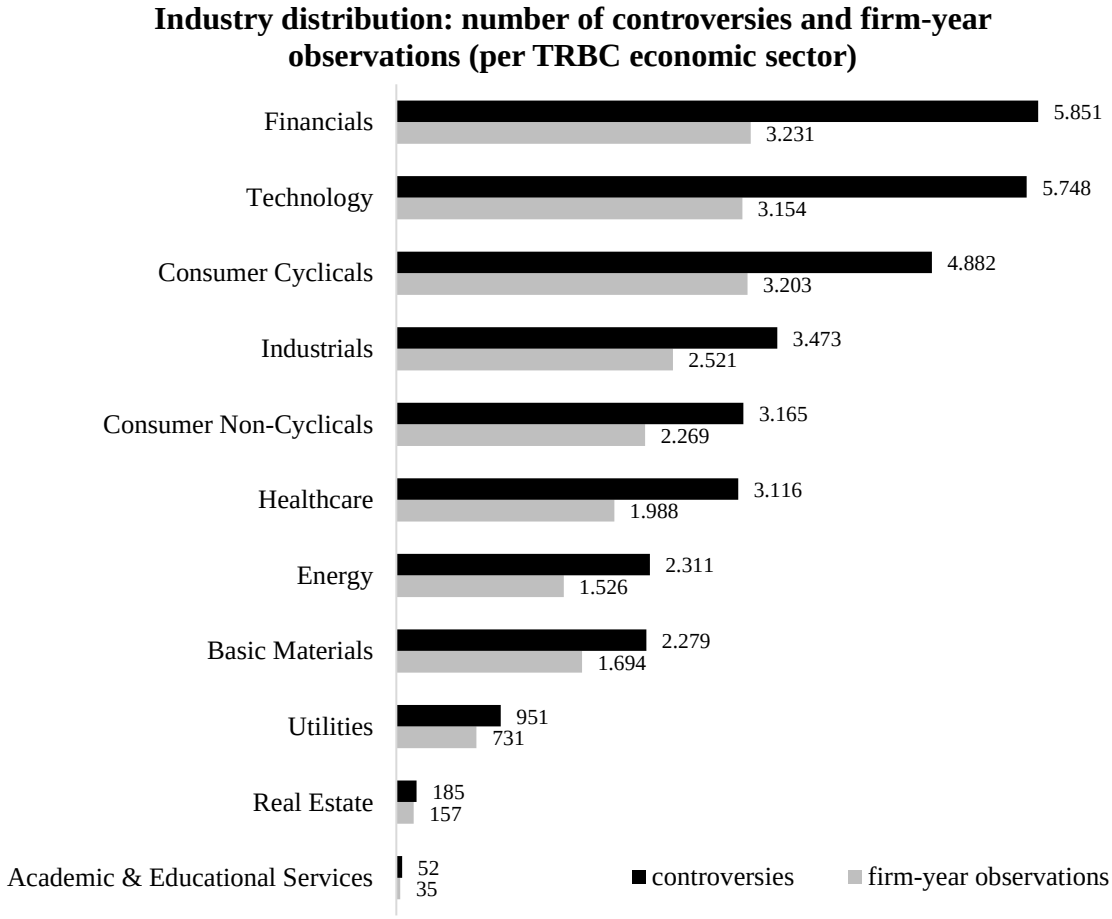


Figure 26 Industry distribution: number of controversies and firm-year observations (per TRBC economic sector)

The number of controversies, however, underlies significant changes over the years observed in the sample. "Academic & Educational Services" lacks a sufficient number of observations in several years. Consequently, the average number of controversies per firm ranges from 0 (in most years) to 6.5 in 2021. The high value in 2021 traces back to "TAL Education Group", which offers tutoring services for school children in China. In the summer of 2021, the Chinese Government announced plans for new regulations in the private schooling sector. As a result, TAL Education Group faces significant burdens, which will eventually render its business model unfeasible upon introduction. Despite its constant claims to meet such regulations, US investors filed a class action

lawsuit at the end of 2021. The litigation is motivated by claims of illicit business conduct, which the company strongly opposed being involved in (e.g., false positive course reviews or the enforcement of high advance payments and payment plans that violate Chinese regulations).[683] The twelve controversies are reflected in the data point "Recent Shareholders Rights Controversies", indicating that the company is linked to "shareholder rights infringements" in the period following the latest completed fiscal year in the Refinitiv ESG rating universe.[684]

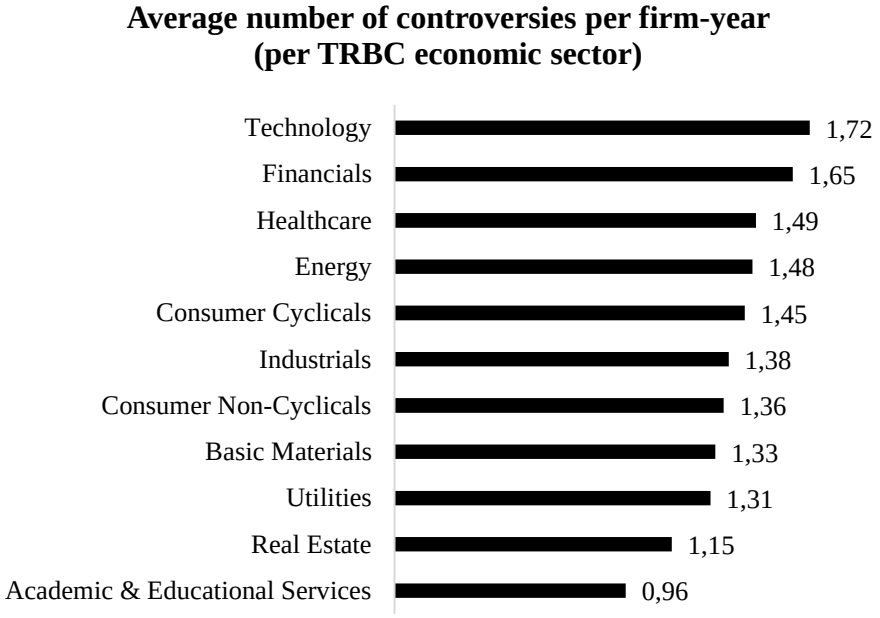


Figure 27 Average number of controversies per firm-year

To alleviate the effect of strongly controversial firms that reflect a large proportion of controversies in a small set of firm-year observations, one can look at the average number of controversies over several individual years. **Figure 27** shows the average number of controversies per year for each economic sector in the sample. Academic & Educational Services (0.96) show less than one controversy per firm-year observation, which is explained by the strong influence of TAL in 2021. Technology (1.72) and Financials (1.65) show the highest average number of controversies per firm-year.

To gain further insights on the number of controversies related to each industry captured within the different economic sectors, **Table 14** shows the ten TRBC business sectors with the most controversies recorded in the dataset. Not surprisingly, the reflected business sectors match the relative

[683] https://seekingalpha.com/article/4475535-tal-education-group-the-beginning-of-the-end, accessed: 05/23/2022.
[684] Refinitiv (2022a, pp. 14, 23)

proportion of superordinate economic sectors listed in **Figure 26**. “Banking & Investment Services” (5,045), “Software & IT Services” (2,294), and “Energy – Fossil Fuels” (2,254) show the most controversies.

TOP 10: Firm-year observations and controversies per TRBC business sector

TRBC - Business sector	Firm-year observations	Controversies
Banking & Investment Services	2,668	5,045
Software & IT Services	1,027	2,294
Energy - Fossil Fuels	1,478	2,254
Pharmaceuticals & Medical Research	1,268	2,168
Technology Equipment	1,173	2,109
Automobiles & Auto Parts	984	1,854
Mineral Resources	1,116	1,577
Cyclical Consumer Services	1,066	1,432
Telecommunications Services	952	1,343
Food & Beverages	993	1,330

Table 14 Firm-year observations and controversies per TRBC business sector

Figure 28 shows additional information on the sort of ESG controversies being captured in each of the business sectors. E, S, and G controversies are the sum of all controversies recorded for each of those controversy data points that match the respective ESG category. For example, Governance controversies (G) reflect all controversies from the data points “Management Compensation”, “Shareholder Rights”, “Insider Dealings”, and “Accounting”. As **Table 10** shows, the number of data points that are considered Social (S) controversies vastly exceeds the other two pillars. Therefore, **Figure 28** shows the share of the total number of controversies within each ESG pillar being recorded for individual business sectors. Thereby, the depiction allows seeing whether there are

TOP 10: Share of total number of controversies within each ESG pillar

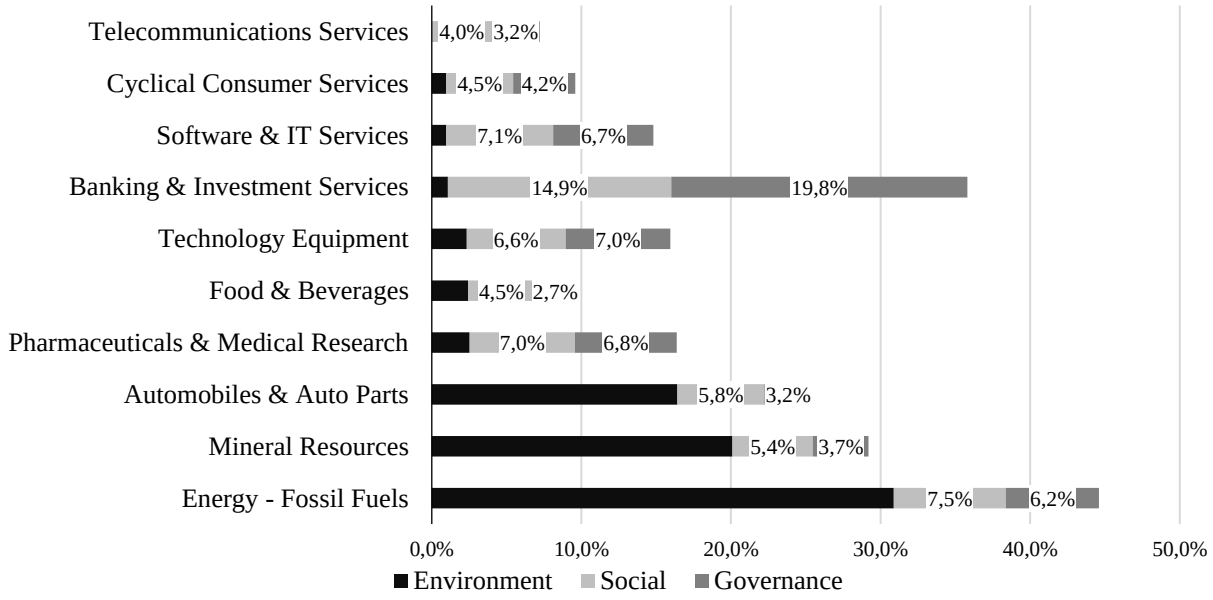


Figure 28 Share of total number of controversies within each ESG pillar

significant differences in the industries that are the main actors involved in controversies of a specific category.

Observations in the business sectors “Energy – Fossil Fuels” (30.9%), “Automobiles & Auto Parts” (16.4%), and “Mineral Resources” (20.1%) make up over two third of all environmental controversies in the dataset. These industries are usually considered emission-intensive, as their production processes or the consumption of their products adds a significant portion to the total mix of emissions. Observing most of the controversies in these industries, on the one hand, unveils for which industries controversial business conduct in a specific pillar is monitored the closest. On the other hand, it eventually shows that environmental issues are more material to these industries. One can assume that it shows how differences in the materiality of specific ESG issues influence the general public's perception and media coverage of different industries. Not surprisingly, service-oriented industries like “Telecommunication Services” (0.0%), “Cyclical Consumer Services” (1.0%), “Software & IT Services” (1.0%), “Banking & Investment Services” (1.1%) make up the lowest proportion of environmental controversies.

Similar patterns can be observed for Governance controversies. Although Governance refers to issues considered material to various industries, the Refinitiv data points capture almost a fifth of all Governance controversies in the business sector of “Banking & Investment Services” (19.8%). More specifically, “Management Compensation” (22.0%) and “Insider Dealings” (28.2%) data points in which the dataset captures a significant proportion of all controversies. Yet again, there are two major explanations for this. Media coverage of Banking & Investment services tends to focus on the business conduct of management and insiders primarily. Simultaneously, Governance issues are predominately material for financial services as such controversies tend to signal negative implications of a business’ trustworthiness and upright business conduct. Given that mainly controversies related to “Business ethics” drive a significant representation of Banking & Investment Services in the number of social controversies, this notion is further underlined.

Nevertheless, the proportion of social controversies in the dataset remains evenly distributed among the different business sectors. Social controversies are captured via 17 data points (as opposed to 1 in E and 4 in G) and constitute almost 90% of the controversies in the dataset. One can assume that Refinitiv news processing is especially powerful in capturing relevant controversies on social data points. As already stated, social controversy data points tend to feature many issues that could capture controversial or fraudulent business conduct in a more general way than the other pillar categories (i.e., act as umbrella categories).

Figure 29 shows the number of strikes grouped by economic sector. Most strikes are recorded for companies in the sectors “Basic Materials” (346), “Industrials” (290), and “Consumer Cyclicals” (251). More precisely, the majority of strikes occurred in the industry groups “Metals & Mining”

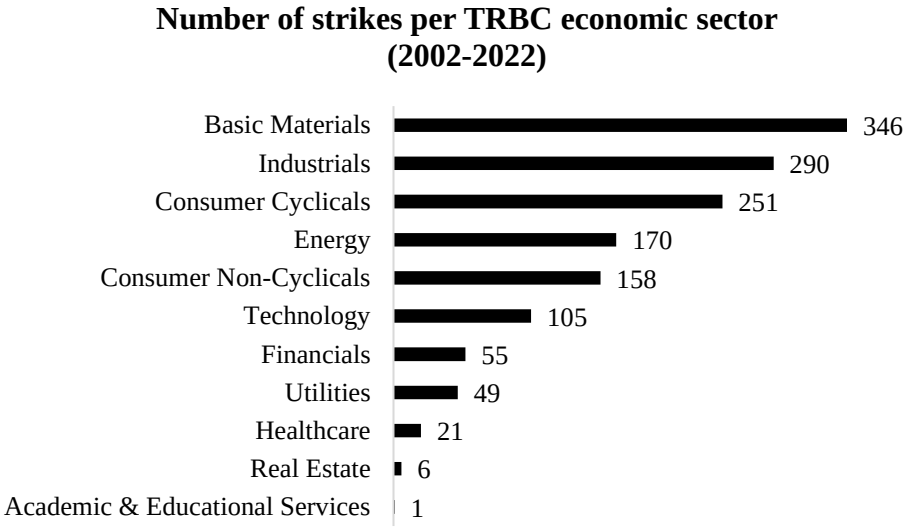


Figure 29 Number of strikes per TRBC economic sector (2002-2022)

(279), “Automobiles & Auto Parts (153), “Oil & Gas” (130), and “Passenger Transportation Services” (101). These values are partly driven by reoccurring strike events and work stoppages in a small set of firms. Most strikes, for example, are observed for the Australian mining company “BHP Group” (17). Followed by the French Oil and Gas Company “Total Energies” (16) and the Dutch automobile company “Stellantis” (14).

The patterns contribute several peculiarities. Stellantis was formed in 2021 after the merger of Italian-American “Fiat Chrysler Automobiles” and the French “PSA Group”. As each strike data point refers to one specific firm-year, the years in which a strike occurred exceeds the years Stellantis exists. Hence, the data point also captures strikes of the formerly individual entities that merged in 2021. The number of strikes observed happening at the “BHP Group” is even larger if one considers that the company is double-listed as two separate entities in the UK and Australia. The dataset captures the BHP Group based in the UK and counts 14 strikes. Strikes observed for either of the two companies, however, are not double counted as each strike captured by Refinitiv refers to a different work stoppage, which an analysis of the sources (not reported in the study at hand) unveils.

Size effects

Chapter 5.4.1 identified severe size biases affecting the sample of Refinitiv ESG performance scores. Larger firms tend to score higher performance scores than their smaller competitors systematically. It is empirically documented that larger firms tend to be covered more frequently by the media and face stronger visibility of their actions.[685] Consequently, one could also expect such patterns in the distribution of controversy observations. Refinitiv itself acknowledges potential size effects in their information processing of news data. To account for this fact, Refinitiv assigns different weights according to the market capitalization when calculating the ESG controversies score. As controversial business conduct of smaller firms might be subject to lower visibility and less media coverage, the corresponding data points are weighted up to three times as strong as controversies of mid- and large-cap companies.[686]

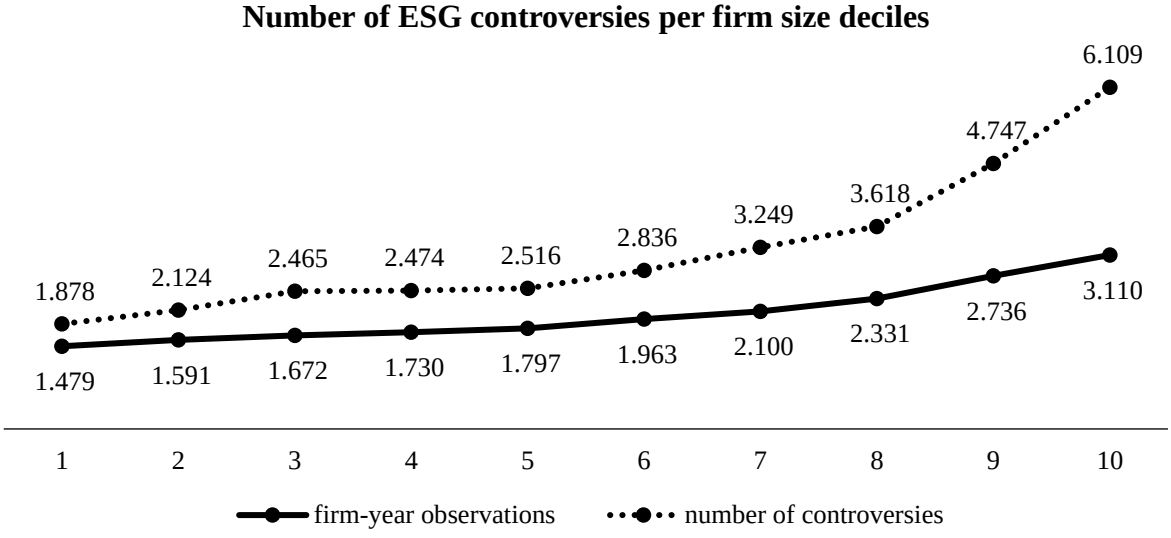


Figure 30 Number of ESG controversies per firm size deciles

Figure 30 shows the distribution of firm-year observations and the number of controversies per firm size decile (based on total assets as reported on the firms' balance sheets). The number of firm-year observations in the highest firm size decile is more than twice as big as the number of controversies of firms in the lowest. Especially when looking at the number of controversies per firm size decile, the positive relationship between firm size and controversies becomes apparent. The dataset captures more than three times as many controversies for the largest firms (10th percentile) as for the smallest firms (1st percentile).

One can assume that smaller firms, on average, do not necessarily conduct their business significantly less controversial than larger firms. However, it is fairly reasonable that the controversial business conduct of larger firms attracts more rigorous attention as the impact of

[685] cf. Drempetic et al. (2020, p. 354), cf. Schreck/Raithel (2015, p. 769)
[686] cf. Refinitiv (2022a, p. 14)

business actions potentially creates larger-scale externalities (e.g., a more significant impact on the environment). Hence, observing such strong patterns underlines the fact that larger firms are subject to greater media coverage, and the likelihood of observing controversies in the Refinitiv ESG database can be considered positively associated with a firm's size.

The ten firms with the highest number of controversies (**Table 15**) rank among the largest global companies in the world. Almost all of them are found in the top 30 of the latest Forbes "Global 2000" list, which ranks the largest firms in the world by several fundamentals (i.e., total assets, sales, and market capitalization).[687] Again some of the controversies stem from single firm-year observations that vastly exceed the number of controversies in other periods.

Top 10: Firms with the most ESG controversies

Company	Firm-year observations	Controversies
ALPHABET A	111	550
APPLE	116	408
META PLATFORMS A	57	345
VOLKSWAGEN	72	335
BANK OF AMERICA	106	322
WALMART	135	308
JP MORGAN CHASE & CO.	82	295
DEUTSCHE BANK	79	259
CITIGROUP	74	247
WELLS FARGO & CO	78	246

Table 15 Top 10: Firms with the most ESG controversies

According to the dataset, for example, "Meta Platforms" (formerly known as "Facebook") experienced 109 controversies in 2018. For the previous year, the dataset contains only 25 controversies. A large fraction of the controversies in 2018 is explained by "data privacy" issues (data point tracks 53 controversies for 2018), which were ignited after the general public became aware of facebook's business connections to "Cambridge Analytica". Among others, the New York Times revealed that Cambridge Analytica analyzed user data to profile potential voters in order to influence their behavior in upcoming US elections. The use and analysis of user data without their consent led to international uproar and general awareness of data privacy issues in social media services.[688]

As the examples above show, the scope of controversies in larger firms can differ immensely. The previous example of TAL shows that it was accused of internal wrongdoing, which primarily affected its shareholders and their perception of the firm's future prospects. Facebook's controversial business connections to Cambridge Analytica, in contrast, sparked a general debate

[687] https://www.forbes.com/lists/global2000/, accessed: 05/24/2022.
[688] https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html, accessed: 05/24/2022.

over data privacy issues for Facebook and the industry as a whole. Through the global extent of facebook's user base, the number of potentially affected customers vastly exceeded the somewhat limited portion of people affected by TAL's controversies. Assuming that the impact of controversies is stronger the larger the causative company, a higher number of controversies might certainly account for the relatively higher impact of such controversies. Nevertheless, it cannot be ruled out that controversy data counts several individual controversies, even though the cause of the controversy is similar or even the same as already described in the introduction of chapter 5.4.2.

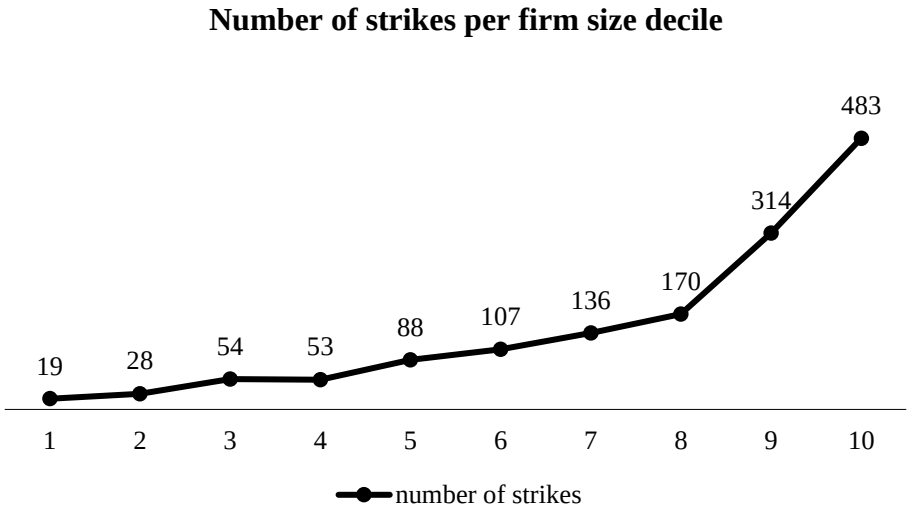


Figure 31 Number of strikes per firm size deciles

Strikes only refer to single firm-year observations. The data point describes whether or not a strike has occurred during a given year. **Figure 31** shows the distribution of strikes over several firm size deciles. Just as analyzed for count data on controversies, there seems to be a clear association between strikes and firm size. Larger firms tend to experience more strikes, eventually captured more accurately by Refinitiv's data processing.

Given the peculiarities of Refinitiv's controversy data, this study does not only rely upon the original data points retrieved from Refinitiv. In order to account for miss counting and double counting of controversies caused by the same business actions, this study also applies binary data points reflecting if there have been controversies recorded in a given firm year. The corresponding specifications are described as they are implemented in particular model specifications.

Final sample

Table 16 describes the final sample formation process. Firm-year observations and the number of individual firms are presented for count data on controversies and binary data on the occurrence of strikes. The dataset originally retrieved from the Refinitiv database consists of a large amount of "NA" values. 89.5% of all firm-year observations on controversies and 63.5% of all firm-year observations on strikes were retrieved as "NA" values. The corresponding observations do not add explanatory capacities to the statistical analysis in the upcoming chapter. Hence, the observations are dropped from the final sample. Some firm-year observations on controversies signaled a value of 0. As the idea of the data points is to count controversies that happened during a specific firm year, values of 0 seem illogical. Especially the small amount of such observations tends to point toward mistakenly added data points or potentially falsely adjusted values. Hence, these observations are dropped as well. In order to analyze ESG data with the closest fit in conceptualization, controversy data and ESG scores are both retrieved from the Refinitiv ESG rating universe. A small amount of firm-year observations, however, reflects firms that could not be matched to the sample of ESG scores. The corresponding observations were dropped from the final sample. To further match both samples, observations of firms from the economic sector "Academic & Educational Services" were dropped. The final sample consists of 20,474 (83,785) firm-year observations on controversies (strikes) captured for 3,171 (11,792) individual firms.

Sample formation process: ESG controversies and STRIKES

	Controversies		Strikes	
	Firm-year observations	**Firms**	**Firm-year observations**	**Firms**
	203,763	3,248	232,081	11,971
NA	-182,408	(3,213)	-147,346	(11,971)
"0"-observation	-711	(401)		
	20,644	3,221	84,735	11,971
No ESG score data	-135	-50	-693	-179
	20,509	3,171	84,042	11,792
Academic & Educational Services	-35	-11	-257	-50
	20,474	3,160	83,785	11,742

Table 16 Sample formation process: ESG controversies and STRIKES

The analysis within this chapter shows severe differences in the number of observations for each controversy data point. In order to get a sufficient number to conduct further analyses, observations that match an ESG score category are aggregated based on the corresponding category score, which can be seen in the "Category" pillar. In doing so, the sample allows analyzing controversies on the corresponding category scores "Resource Use", "Community", "Product responsibility", "Human rights", "Workforce", "Management", and "Shareholders". **Table 17** shows the different

data points and how the observations are aggregated in broader ESG categories. Additional aggregation is also conducted on the level of the three ESG pillars.

As the observations are not evenly distributed among each category, the number of observations ranges from 392 (Management) to 9,392 (Community). The firm-year observations identify 1,043 controversies related to the Resource use category, 15,875 in Community, 7,197 in Product responsibility, 483 in Human rights, 4,576 in Workforce, and 2,352 controversies in the Shareholder categories. Due to the consolidation of controversy data points within different categories, the number of firm-year observations does not equal the sum of firm-year observations of the underlying data points.

Final sample: Firm-year observations and controversies per controversy category

Pillar	Category	Controversy data point	Firm-year obs.	Controversies
Environmental	RES	Environmental	586	1,043
		Resource use controversies	**586**	**1,043**
		Environmental controversies	**586**	**1,043**
Social	COM	Anti-Competition	3,292	5,699
Social	COM	Business Ethics	3,285	5,994
Social	COM	Critical Countries	173	194
Social	COM	Intellectual Property	1,449	2,386
Social	COM	Public Health	583	736
Social	COM	Tax Fraud	610	866
		Community controversies	**7,187**	**15,875**
Social	HR	Child Labor	99	106
Social	HR	Human Rights	326	377
		Human rights controversies	**413**	**483**
Social	PRD	Consumer Complaints	1,351	1,930
Social	PRD	Customer Health & Safety	1,105	2,033
Social	PRD	Privacy	872	1,478
Social	PRD	Product Access	115	133
Social	PRD	Responsible Marketing	1,071	1,556
Social	PRD	Responsible R&D	53	67
		Product responsibility controversies	**3,495**	**7,197**
Social	WF	Diversity and Opportunity	776	1,002
Social	WF	Employees' Health & Safety	1,187	1,614
Social	WF	Wages Working Condition	1,412	1,960
		Workforce controversies	**2,900**	**4,576**
		Social controversies	**10,477**	**28,131**
Governance	MNG	Management Compensation	392	437
		Management controversies	**392**	**437**
Governance	SH	Accounting	396	477
Governance	SH	Insider Dealings	419	463
Governance	SH	Shareholder Rights	918	1,413
		Shareholder controversies	**1,627**	**2,352**
		Governance controversies	**1,930**	**2,789**
		ESG controversies	**11,555**	**31,963**

Table 17 Final sample: Firm-year observations and controversies per controversy category matched with ESG category and pillar scores

Concluding remarks

Several features of Refinitiv's Controversy database have been described and analyzed throughout this chapter. The most important findings and their implications for further analysis are listed below:

1. There are severe differences in the occurrence of controversies among the observed controversy categories. Some categories (e.g., "Business ethics controversies") likely capture actions of a broad range of controversial business conduct by definition. Hence, variation is eventually caused by the proprietary data processing of Refinitiv.
2. Firm-year observations and the number of controversies experience a drop in 2015. Refinitiv does not provide information on the relatively small number of observations in 2015 and the years after. However, one can observe that in 2015 overrepresentation of US firms shrinks compared to the distribution of observations in the ESG score sample.
3. Refinitiv controversy data has a high degree of granularity. Several examples (e.g., data privacy issues at "Meta platforms" in 2015, class action lawsuits against "TAL" in 2021) show that the number of controversies is based on similar incidences that differ, for example, based on the impacted region or interest groups. This design choice, however, might also facilitate double-counting when interpretational boundaries are set differently.
4. Mergers in the sample (e.g., Stellantis in 2021) are not necessarily being observed as a new company. Thus, controversies eventually affect previous legal entities. As the dataset matches the logic of the ESG score sample, the comparison is justifiable even when different legal entities are involved. Further, this peculiarity alleviates potential survivorship biases.
5. Strikes in US firms are less likely than in European firms. Comparing the fraction of firm-year observations from different regions, however, unveils controversies affecting US-domiciled firms twice as often as strikes.
6. The small frequency of strikes is reasonable, given their uncommon nature. Nevertheless, the definition and scope of strikes being captured are unclear. Several firms and industries are subject to periodical work stoppages when, for example, wages are up for debate. However, the data does not explicitly identify strikes that point toward structural problems of potential misconduct. Regardless of their motivation, worker strikes might affect productivity, which renders the data points useful for further analysis.

5.4.3 Other variables

Table 18 provides descriptive statistics for all firms in the sample and its subsamples of firms domiciled in the United States and selected European countries (see chapter 5.3.3). Each table contains dependent and independent variables used to specify the models in this study (except for the aforementioned ESG data). For each variable, the table lists the number of observations (N), mean, standard deviation, and the 25^{th}, 50^{th} (median), and 75^{th} percentile of the corresponding sample values.

The maximum number of observations in the full sample is 80,887. The US (European) sample consists of up to 25,649 (19,093) observations. The investigation of idiosyncratic volatility primarily focuses on European (17,386) and US (24,625) firms, which leads to a total number of 42,011 observations in the full sample. Mean (0.273) and median (0.238) values in the European sample indicate a slightly lower idiosyncratic volatility for European firms as opposed to US firms (0.298 and 0.239). Mean (0.051) and median (0.050) values of ROA in the European sample indicate that European firms are more profitable than firms in the US sample (0.013; 0.043). Over the full sample, firms earn an average ROA of 0.037, which is slightly skewed to higher profitability in the upper percentiles (median of 0.046). The same pattern is observed for the US sample, indicating that high profitability is concentrated on fewer firms than in the European sample. This finding is backed by the standard deviation of ROA in the European sample (0.099), which is lower than for the full sample (0.127) and US sample (0.167). Not surprisingly, the volatility of profits (measured as the quarterly standard deviation in the three years prior to the observed year) shows the same pattern. The average firm size (26,639.77 million USD) is way higher than its median (3,956.70 million USD). Hence, the sample is dominated by a majority of comparably small firms. As firm size is measured in total assets, the variable is eventually driven by firms in asset-intensive industries. Large differences in the average firm size of European firms (40,867.49 million USD) and US firms (20,670.20 million USD) point toward structural differences in the industry distribution of the two sub-samples (as already shown in the previous section on the industry distribution of ESG data). Given the larger scale of European firms, capital expenditures (0.039; 0.028) and R&D expenditures (0.015; 0.000) measured as the fractions of total assets are lower than in the full sample average (0.044; 0.029 and 0.021; 0.000). Average R&D expenditures in the US sample (0.041) are almost twice as high as in the full sample (0.021), potentially indicating future growth opportunities. Intangible assets (IA) intensity is highest in the European sample with only slight differences to the US sample. However, both sub-samples exceed the full sample average and median, indicating a high concentration of intangible assets in these markets.

Major differences in the financial systems of the sub-samples become apparent when looking at the values of leverage. Leverage is measured as the ratio of long- and short-term debt to shareholder equity. The mean value of the European sample (1.223) exceeds the average of the full sample (1.016), signaling rather debt-based financing. The mean value in the US sample (0.968) indicates equity-based financing prevails. The average age of the firms in the full sample is 35 years. On average, firms in the US sample (31.5 years) are slightly younger than in the European sample (41 years). The median values (23; 25 years), however, indicate that there should be no substantial differences in the development stages of the majority of the firms.

Descriptive statistics: Idiosyncratic volatility and control variables

Full sample (2001-2021)						
variable	**N**	**mean**	**sd**	**25th-perc.**	**median**	**75th-perc.**
CapEx intensity	80,887	0.044	0.056	0.010	0.029	0.059
Firm age	80,887	35.19	33.06	13.00	24.00	48.00
Firm size (in $US million)	80,887	26,639.77	109,073.20	1,309.11	3,956.70	12,752.87
IA intensity	80,887	0.151	0.197	0.007	0.054	0.240
Idiosyncratic volatility	42,011	0.288	0.167	0.176	0.241	0.346
ROA	80,887	0.037	0.127	0.013	0.046	0.086
ROA volatility	80,887	0.040	0.068	0.007	0.018	0.041
R&D intensity	80,887	0.021	0.074	0.000	0.000	0.012
Leverage	80,887	1.016	1.921	0.178	0.555	1.187
Subsample: United States						
variable	**N**	**mean**	**sd**	**25th-perc.**	**median**	**75th-perc.**
CapEx intensity	25,649	0.041	0.055	0.009	0.025	0.052
Firm age	25,649	31.46	28.84	12	23	39
Firm size (in $US million)	25,649	20,670.20	88,983.41	1,111.20	3,478.70	10,781.00
IA intensity	25,649	0.193	0.215	0.012	0.104	0.329
Idiosyncratic volatility	24,625	0.298	0.186	0.169	0.239	0.362
ROA	25,649	0.013	0.167	0.009	0.043	0.084
ROA volatility	25,649	0.049	0.085	0.008	0.020	0.050
R&D intensity	25,649	0.041	0.115	0.000	0.000	0.028
Leverage	25,649	0.968	2.152	0.141	0.563	1.199
Subsample: Europe						
variable	**N**	**mean**	**sd**	**25th-perc.**	**median**	**75th-perc.**
CapEx intensity	19,093	0.039	0.044	0.01	0.028	0.053
Firm age	19,093	41.23	43.63	13	25	56
Firm size (in $US million)	19,093	40,867.49	160,329.90	1,370.39	4,181.44	15,418.13
IA intensity	19,093	0.203	0.206	0.023	0.138	0.333
Idiosyncratic volatility	17,386	0.273	0.137	0.182	0.238	0.324
ROA	19,093	0.051	0.099	0.016	0.050	0.088
ROA volatility	19,093	0.034	0.052	0.008	0.017	0.039
R&D intensity	19,093	0.015	0.049	0.000	0.000	0.011
Leverage	19,093	1.223	2.192	0.250	0.635	1.296

Notes: CapEx intensity is capital expenditures divided by total assets. Firm age is the number of years since the firm's IPO. Firm size (in $US million) is the natural logarithm of total assets reported on the balance sheet. IA intensity is the ratio of intangible assets and total assets. Idiosyncratic volatility is the yearly standard deviation of the residual risk derived from a Carhart four-factor model. Return on assets is net income divided by the book value of total assets. ROA volatility is the standard deviation of ROA over the last three fiscal years. R&D intensity is R&D expenditures divided by total assets. Leverage is the ratio of long- and short-term debt and shareholders' equity. Values reported in local currencies are converted into USD. Variables are winsorized at the yearly 1%- and 99%-percentiles.

Table 18 Descriptive statistics: Idiosyncratic volatility and control variables

5.5 Inferential statistics and analysis

5.5.1 ESG metrics and firm-specific risk

Results

Figure 32 depicts the average annualized IV for firms sorted into ESG performance deciles based on four different ESG rating modifications. The black line shows the level of IV for deciles based on the raw Refinitiv ESG score. The grey line is based on the size-adjusted ESG score, which standardizes the raw Refinitiv ESG score on firm size deciles. The dotted lines show the average IV based on the respective rating modifications additionally adjusted for observations intentionally scored 0 in the original Refinitiv ESG scoring model. Hence, it shows the ratings exclusively for firms fully covered in each category by the Refinitiv ESG scoring model.

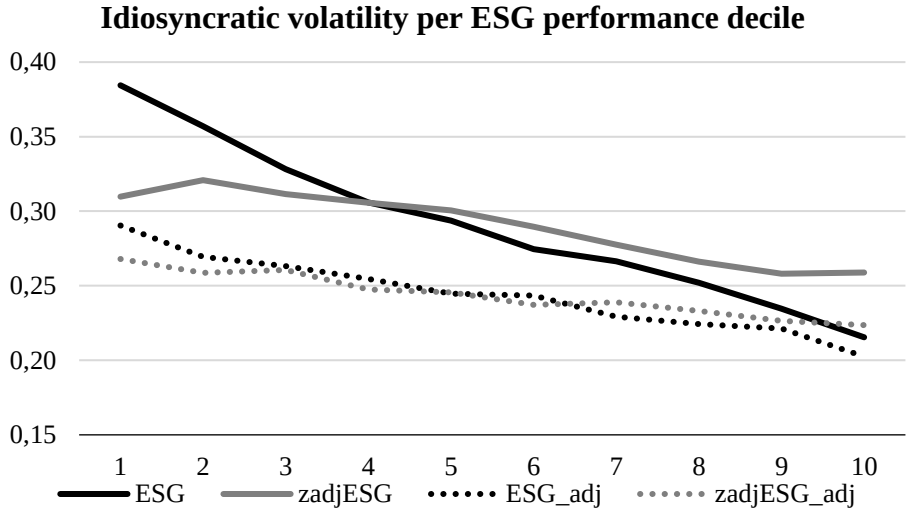


Figure 32 Idiosyncratic volatility per ESG performance decile

The graphs show that the average IV is inversely related to the ESG performance decile. The original ESG score (ESG) shows the highest average IV in the first decile, indicating that firms that scored the lowest experience higher IV than firms that scored best. The relationship follows a monotonic trend from roughly 39% in the lowest to 22% in the highest ESG performance decile. Adjusting for biases induced by firm size and the assignment of 0 to non-reported data (ESG_adj) reduces the spread between low and high ESG performance deciles. Nevertheless, the negative relationship is still apparent. The graphs based on size-adjusted ratings (zadjESG and zadj-ESG_adj) intersect their respective unadjusted counterparts at the fourth and fifth deciles. This indicates that adjusting for firm size leads to a more evenly spread distribution of IV among different ESG performance deciles. The ratings adjusted for non-reported data show lower levels of average annualized IV throughout all ESG performance deciles. Hence, on average, the subset of fully covered firms experiences lower levels of IV.

Table 19 depicts regression results for specifications regressing past ESG performance metrics on future IV. As opposed to **Figure 32**, the inferential approach helps to identify the predictive indication of the ESG performance metrics as it is based on lagged ESG performance data from the year prior to the year IV is calculated for. As shown in **Figure 32**, the negative relationship between ESG performance and IV is confirmed. A one standard deviation increase in the ESG scores is associated with a 0.95% decrease in IV per year.

Regression results: ESG performance and IV (full sample)

	IDIOSYNCRATIC VOLATILITY			
	(1)	(2)	(3)	(4)
ESG	**-0.0095*** **			
	t = -6.8226			
ESG_adj		**-0.0059*** **		
		t = -3.8544		
zadjESG			**-0.0081*** **	
			t = -6.6522	
zadjESG_adj				**-0.0053*** **
				t = -3.6997
CapEx intensity	**0.0643** **	0.0356	**0.0644** **	0.0343
	t = 2.5468	t = 0.6774	**t = 2.5561**	t = 0.6524
Firm age	**-0.0001*** **	**-0.0001** **	**-0.0001*** **	**-0.0001** **
	t = -3.1395	**t = -2.1332**	**t = -3.1673**	**t = -2.1414**
Firm size	**-0.0224*** **	**-0.0168*** **	**-0.0250*** **	**-0.0183*** **
	t = -22.1972	**t = -12.4169**	**t = -30.3993**	**t = -15.3029**
IA intensity	**-0.0339*** **	**-0.0446*** **	**-0.0342*** **	**-0.0452*** **
	t = -5.5198	**t = -4.9571**	**t = -5.5593**	**t = -5.0073**
ROA	**-0.4005*** **	**-0.4715*** **	**-0.4000*** **	**-0.4721*** **
	t = -36.1982	**t = -15.5196**	**t = -36.1630**	**t = -15.5094**
ROA volatility	0.0022	**0.4848*** **	0.0022	**0.4851*** **
	t = 1.1236	**t = 7.2566**	t = 1.1249	**t = 7.2430**
R&D intensity	**-0.0383** **	-0.0022	**-0.0389** **	-0.0034
	t = -2.2535	t = -0.0329	**t = -2.3001**	t = -0.0496
leverage	**0.0067*** **	**0.0051*** **	**0.0068*** **	**0.0051*** **
	t = 10.1481	**t = 6.2459**	**t = 10.2223**	**t = 6.2855**
BusinessSectorFE	Yes	Yes	Yes	Yes
CountryFE	Yes	Yes	Yes	Yes
YearFE	Yes	Yes	Yes	Yes
Observations	35,299	8,919	35,299	8,914
R^2	0.4136	0.3933	0.4135	0.3932
Adjusted R^2	0.4125	0.3887	0.4123	0.3886

Note: Idiosyncratic volatility is the yearly standard deviation of the residual risk derived from a Carhart four-factor model. ESG are (standardized) Refinitiv ESG performance scores as provided by Refinitiv. The prefix "zadj" identifies ESG scores standardized based on yearly firm size deciles. The suffix "_adj" indicates whether only fully covered firm-year observations are considered. Idiosyncratic volatility is the yearly standard deviation of the residual risk derived from a Carhart four-factor model. CapEx intensity is capital expenditures divided by total assets. Firm age is the number of years since the firm's IPO. Firm size is the natural logarithm of total assets reported on the balance sheet. IA intensity is the ratio of intangible assets and total assets. Return on assets is net income divided by the book value of total assets. ROA volatility is the standard deviation of ROA over the last three fiscal years. R&D intensity is R&D expenditures divided by total assets. Leverage is the ratio of long- and short-term debt and shareholders' equity. Business sectors, years, and countries are fixed. Values reported in local currencies are converted into USD. Variables are winsorized at the yearly 1%- and 99%-percentiles. All explanatory variables are lagged by one year. Robust test statistics calculated with group-clustered standard errors are reported below coefficients. Significance levels are indicated as: * = 0.1, ** = 0.05, *** = 0.01.

Table 19 Regression results: ESG performance and IV (full sample)

The relationship holds even after adjusting for non-reported data ("ESG_adj") and potential size biases by standardizing scores within yearly firm size deciles ("zadjESG"). The number of observations significantly drops after excluding firms only partially covered by the database underlying the Refinitiv ESG scores. This approach is exclusively based on firm-year observation in which all ten ESG category scores were not intentionally coded 0 otherwise applied by Refinitiv to account for non-reported ESG information. The effect sizes are smaller for the specifications in which the ESG score is adjusted for non-reported ESG data (2 and 4) than for the unadjusted (1) or size-adjusted (3) ESG scores. Similar to **Figure 32**, the regression coefficients for average annualized IV are lower within the smaller sub-samples of fully covered firms (regression models 2 and 4). The explanatory variables indicate that profitability and firm size are negatively associated with the level of future IV. The uncertainty of profits, measured as the three-year standard deviation of monthly ROA, is significantly affecting IV positively in the specifications that analyze the limited sub-samples (2 and 4). The positive relationship indicates that within the majority of firm-year observations, the uncertainty of profits does not add explanatory power to the indication of future IV. Limiting the sample to fully covered firms only, however, drastically enhances the significant positive association between uncertain prospects of a firm and IV. Within the samples of all firm-year observations (1 and 3), the effects of uncertainty in profits are insignificant. However, significant coefficients for CapEx and R&D intensity suggest that uncertainty and opportunities are identified rather indirectly in the form of resources deployed to certain projects, eventually indicating the firm's prospects. In particular, capital expenditures enhance uncertainty indicated by higher levels of IV in the next year. The intensity of intangible assets is associated with lower levels of future IV. The effect size is slightly larger in the limited sub-sample specifications (2 and 4). The coefficients for firm age indicate a significant but economically negligible association of older firms with lower IV. Each regression model implements year, business sector, and country fixed effects to capture otherwise unobserved heterogeneity among the investigated firms.

Regression results: ESG, E, S, and G scores and IV (US and European sub-samples)

	IDIOSYNCRATIC VOLATILITY (UNITED STATES)				IDIOSYNCRATIC VOLATILITY (EUROPE)			
	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ESG (lagged)	**-0.0065*** **				**-0.0069*** **			
	t = -3.8706				**t = -4.1432**			
E (lagged)		-0.0002				**-0.0057*** **		
		t = -0.1438				**t = -3.5473**		
S (lagged)			**-0.0052*** **				**-0.0070*** **	
			t = -2.9544				**t = -4.2249**	
G (lagged				**-0.0061*** **				**-0.0034** *
				t = -4.5330				**t = -2.3705**
Residual ESG		**-0.0137*** **	**-0.0072*** **	-0.0033		**-0.0060** *	-0.0033	**-0.0084*** **
		t = -5.7833	**t = -2.8586**	t = -1.4262		**t = -2.5576**	t = -1.2409	**t = -3.8841**
CapEx intensity	**0.0696** *	**0.0694** *	**0.0695** *	**0.0691** *	0.0369	0.0396	0.0418	0.0388
	t = 2.1191	**t = 2.1075**	**t = 2.1181**	**t = 2.1228**	t = 0.9757	t = 1.0481	t = 1.1105	t = 1.0245
Firm age	**-0.0003*** **	**-0.0003*** **	**-0.0003*** **	**-0.0003*** **	-0.00005	-0.00005	-0.00004	-0.00005
	t = -5.1829	**t = -5.3194**	**t = -5.1395**	**t = -5.2388**	t = -1.2197	t = -1.2173	t = -1.1529	t = -1.1691
Firm size	**-0.0275*** **	**-0.0278*** **	**-0.0276*** **	**-0.0277*** **	**-0.0196*** **	**-0.0197*** **	**-0.0195*** **	**-0.0197*** **
	t = -23.7138	**t = -24.0905**	**t = -23.5662**	**t = -23.7538**	**t = -16.5685**	**t = -16.5250**	**t = -16.4969**	**t = -16.6026**
IA intensity	**-0.0249*** **	**-0.0240*** **	**-0.0250*** **	**-0.0254*** **	**-0.0275*** **	**-0.0279*** **	**-0.0275*** **	**-0.0279*** **
	t = -3.2050	**t = -3.0946**	**t = -3.2284**	**t = -3.2784**	**t = -2.7463**	**t = -2.7730**	**t = -2.7481**	**t = -2.7887**
ROA	**-0.3785*** **	**-0.3778*** **	**-0.3783*** **	**-0.3778*** **	**-0.3634*** **	**-0.3622*** **	**-0.3619*** **	**-0.3626*** **
	t = -28.1590	**t = -28.2284**	**t = -28.1561**	**t = -28.1304**	**t = -17.9391**	**t = -17.8157**	**t = -17.8581**	**t = -17.8797**
ROA volatility	0.0012	0.001	0.0012	0.0011	**0.2067*** **	**0.2051*** **	**0.2054*** **	**0.2052*** **
	t = 0.9759	t = 0.8309	t = 0.9733	t = 0.9309	**t = 3.8976**	**t = 3.8925**	**t = 3.9030**	**t = 3.9014**
R&D intensity	**-0.0561*** **	**-0.0563*** **	**-0.0561*** **	**-0.0571*** **	-0.0705	-0.0686	-0.0692	-0.07
	t = -2.7918	**t = -2.8059**	**t = -2.7962**	**t = -2.8421**	t = -1.2994	t = -1.2599	t = -1.2704	t = -1.2858
Leverage	**0.0058*** **	**0.0059*** **	**0.0058*** **	**0.0058*** **	**0.0060*** **	**0.0061*** **	**0.0061*** **	**0.0061*** **
	t = 6.6223	**t = 6.6689**	**t = 6.6306**	**t = 6.6163**	**t = 6.1711**	**t = 6.1970**	**t = 6.1995**	**t = 6.2176**
BusinessSectorFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CountryFE	No	No	No	No	Yes	Yes	Yes	Yes
YearFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	20,520	20,508	20,508	20,516	14,779	14,755	14,754	14,776
R^2	0.4561	0.4575	0.4562	0.4565	0.3497	0.3498	0.3500	0.3500
Adjusted R^2	0.4546	0.4560	0.4547	0.4551	0.3467	0.3467	0.3470	0.3469

Note: Idiosyncratic volatility is the yearly standard deviation of the residual risk derived from a Carhart four-factor model. ESG, E, S, and G are Environmental, Social, and Governance pillar scores as provided by Refinitiv (standardized based on yearly firm size deciles). Residual ESG results from orthogonalizing single E, S, and G pillar scores on the overall ESG score. Idiosyncratic volatility is the yearly standard deviation of the residual risk derived from a Carhart four-factor model. CapEx intensity is capital expenditures divided by total assets. Firm age is the number of years since the firm's IPO. Firm size (in $US million) is the natural logarithm of total assets reported on the balance sheet. IA intensity is the ratio of intangible assets and total assets. Return on assets is net income divided by the book value of total assets. ROA volatility is the standard deviation of ROA over the last three fiscal years. R&D intensity is R&D expenditures divided by total assets. Leverage is the ratio of long- and short-term debt and shareholders' equity. Business sectors, years, and countries are fixed. Values reported in local currencies are converted into USD. Variables are winsorized at the yearly 1%- and 99%-percentiles. All explanatory variables are lagged by one year. Robust test statistics calculated with group-clustered standard errors are reported below coefficients. Significance levels are indicated as: * = 0.1, ** = 0.05, *** = 0.01.

Table 20 Regression results: ESG, E, S, and G scores and IV (US and European sub-samples)

Table 20 shows the results of regression models analyzing the US and European sub-samples of firm-year observations. All models are based on the size-adjusted ESG ratings ("zadj") consisting of all scores standardized based on yearly firm size deciles. In addition to the already investigated overall ESG rating, the models further distinguish ESG performance by regressing IV on E, S, and G pillar scores. All models confirm the negative association between high ESG scores and lower IV in the subsequent year. The insignificant negative relationship of E pillar scores and IV in the US sub-sample, however, also confirms that associations tend to differ on a regional basis. Effect sizes of separate E, S, and G pillar scores roughly match the coefficient of the overall ESG score. As the overall ESG score in the raw Refinitiv scoring model is a linear combination of the separate pillar scores, one could expect the coefficient to add up to the effect size of the overall ESG score to some extent. Since this is not the case, the results indicate that adverse effects stemming from the aggregation of different ESG categories potentially induce different effect sizes.

Further, the separate analysis of ESG pillar scores as components of the overall ESG scores does not account for the effects of the left-out components of overall ESG performance. Hence, these left-out components are omitted and could induce biases in the estimated models, which imply endogeneity. In order to address endogeneity concerns stemming from omitted variables, the models additionally incorporate a residual ESG variable. *Residual ESG performance* is assessed by orthogonalizing the respective pillar score on the overall ESG score.[689] Hence, the specifications incorporate the firm-specific residuals from a regression model of the pillar or (in the remainder of this study) category scores on the overall ESG score. As the residual ESG scores capture the information that the single pillar score does not incorporate and therefore indicate whether the other, left-out pillar scores add explanatory power to the model. All models confirm the explanatory capabilities of the ESG pillar scores. Within the US sub-sample, however, regression model 8 shows an insignificant residual ESG coefficient when implementing the Governance pillar score. The European sub-sample shows an insignificant coefficient in regression model 11 based on the Social pillar score. These results indicate that Governance pillar scores in the US sub-sample and Social pillar scores in the European sub-sample roughly capture the entire negative association between ESG performance and future IV.

The other explanatory variables show the same associations as in the previous model. Interestingly, the significance of the coefficients shows the same distinctive patterns as in **Table 19**. However, the significance of the coefficients in the US subsample matches the one of the unrestricted sample; the European sub-sample matches the significance pattern of the sample limited to fully covered firm-

[689] Orthogonalization refers to ESG metrics being regressed on the respective superordinate level of ESG scores (e.g., the independent variable ESG overall score is regressed on a single ESG category scores). The residuals of this model than constitute the variable *residual ESG* for each individual observation of the respective sample.

year observations (_adj). As the fraction of fully covered firms is substantially higher in the European sub-sample, the US sub-sample consists of more firm-year observations that were scored 0 because of non-reported information. Hence, the explanatory power of the dependent variables is substantially affected by the quantitative differences in ESG information underlying the different ESG metrics.

Table 21 shows the results of the same general regression model specification as in the previous tables. However, the ESG performance is further distinguished by incorporating all ten category scores separately. All category scores except for the SH score are significantly associated with lower IV in the subsequent year. Regressions coefficients range from -0,0076 (WF) to +0,0021 (SH). The significance pattern of the explanatory variables matches the one observed within the non-restricted sample, unadjusted for the potential effects of non-reported ESG information.

Regression results: ESG category scores and IV (full sample)

	IDIOSYNCRATIC VOLATILITY									
	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)
EMM (lagged)	**-0.0039*** ** t = -3.3830									
INO (lagged)		**-0.0017*** t = -1.6599								
RES (lagged)			**-0.0032*** ** t = -2.5897							
COM (lagged)				**-0.0044*** ** t = -3.8309						
HR (lagged)					**-0.0035*** ** t = -3.2356					
PRD (lagged)						**-0.0061*** ** t = -5.4734				
WF (lagged)							**-0.0076*** ** t = -6.0356			
CSR (lagged)								**-0.0042*** ** t = -3.5402		
MNG (lagged)									**-0.0063*** ** t = -6.6332	
SH (lagged)										**0.0021** ** t = 2.2754
Residual ESG	**-0.0099*** ** t = -6.3200	**-0.0091*** ** t = -6.7759	**-0.0114*** ** t = -7.3848	**-0.0078*** ** t = -5.5418	**-0.0090*** ** t = -6.8131	**-0.0068*** ** t = -5.0834	**-0.0062*** ** t = -4.1191	**-0.0091*** ** t = -6.3728	**-0.0060*** ** t = -4.0520	**-0.0094*** ** t = -7.5181
CapEx intensity	**0.0656*** ** t = 2.6069	**0.0647** ** t = 2.5727	**0.0648** ** t = 2.5753	**0.0652*** ** t = 2.5883	**0.0655*** ** t = 2.6076	**0.0657*** ** t = 2.5946	**0.0660*** ** t = 2.6262	**0.0644** ** t = 2.5502	**0.0638** ** t = 2.5417	**0.0641** ** t = 2.5173
Firm age	**-0.0001*** ** t = -3.1961	**-0.0001*** ** t = -3.1741	**-0.0001*** ** t = -3.2651	**-0.0001*** ** t = -3.1310	**-0.0001*** ** t = -3.1589	**-0.0001*** ** t = -3.1440	**-0.0001*** ** t = -3.1151	**-0.0001*** ** t = -3.1796	**-0.0001*** ** t = -3.2207	**-0.0001*** ** t = -3.1207
Firm size	**-0.0250*** ** t = -30.4953	**-0.0250*** ** t = -30.4488	**-0.0251*** ** t = -30.4285	**-0.0249*** ** t = -30.1036	**-0.0250*** ** t = -30.3580	**-0.0250*** ** t = -30.4572	**-0.0249*** ** t = -30.1456	**-0.0250*** ** t = -30.2919	**-0.0250*** ** t = -30.4439	**-0.0249*** ** t = -30.4077
IA intensity	**-0.0337*** ** t = -5.4866	**-0.0336*** ** t = -5.4504	**-0.0340*** ** t = -5.5241	**-0.0342*** ** t = -5.5584	**-0.0344*** ** t = -5.5943	**-0.0343*** ** t = -5.5823	**-0.0343*** ** t = -5.5730	**-0.0340*** ** t = -5.5249	**-0.0340*** ** t = -5.5341	**-0.0341*** ** t = -5.5508
ROA	**-0.3998*** ** t = -36.1589	**-0.3995*** ** t = -36.1782	**-0.3996*** ** t = -36.1521	**-0.3997*** ** t = -36.1319	**-0.3997*** ** t = -36.1491	**-0.3994*** ** t = -36.1390	**-0.3995*** ** t = -36.1328	**-0.3999*** ** t = -36.1501	**-0.4000*** ** t = -36.1665	**-0.4001*** ** t = -36.2041
ROA volatility	0.0021 t = 1.1081	0.0022 t = 1.1296	0.0021 t = 1.0772	0.0022 t = 1.1242	0.0022 t = 1.1284	0.0022 t = 1.1359	0.0022 t = 1.1447	0.0022 t = 1.1286	0.0021 t = 1.1060	0.0022 t = 1.1430
R&D intensity	**-0.0388** ** t = -2.2981	**-0.0392** ** t = -2.3125	**-0.0390** ** t = -2.3118	**-0.0385** ** t = -2.2773	**-0.0384** ** t = -2.2761	**-0.0389** ** t = -2.2860	**-0.0378** ** t = -2.2421	**-0.0388** ** t = -2.2958	**-0.0399** ** t = -2.3621	**-0.0379** ** t = -2.2444
Leverage	**0.0068*** ** t = 10.2377	**0.0068*** ** t = 10.2243	**0.0068*** ** t = 10.2092	**0.0068*** ** t = 10.2229	**0.0068*** ** t = 10.2450	**0.0068*** ** t = 10.2317	**0.0068*** ** t = 10.2346	**0.0068*** ** t = 10.2120	**0.0068*** ** t = 10.2178	**0.0068*** ** t = 10.2006
BusinessSectorFE | CountryFE | YearFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	35,263	35,251	35,263	35,269	35,262	35,262	35,269	35,299	35,299	35,299
R^2	0.4138	0.4124	0.4141	0.4134	0.4137	0.4138	0.4136	0.4136	0.4138	0.4144
adjusted R^2	0.4126	0.4112	0.4129	0.4123	0.4125	0.4126	0.4124	0.4124	0.4126	0.4132

Notes: Idiosyncratic volatility is the yearly standard deviation of the residual risk derived from a Carhart four-factor model. EMM, INO, RES, COM, HR, PRD, WF, CSR, MNG, and SH are category scores for Emissions, Innovation, Resource use, Community, Human rights, Product Responsibility, Workforce, CSR strategy, Management, and Shareholders provided by Refinitiv (standardized based on yearly firm size deciles). Idiosyncratic volatility is the yearly standard deviation of the residual risk derived from a Carhart four-factor model. CapEx intensity is capital expenditures divided by total assets. Firm age is the number of years since the firm's IPO. Firm size (in $US million) is the natural logarithm of total assets reported on the balance sheet. IA intensity is the ratio of intangible assets and total assets. Return on assets is net income divided by the book value of total assets. ROA volatility is the standard deviation of ROA over the last three fiscal years. R&D intensity is R&D expenditures divided by total assets. Leverage is the ratio of long- and short-term debt and shareholders' equity. Business sectors, years, and countries are fixed. Values reported in local currencies are converted into USD. Variables are winsorized at the yearly 1%- and 99%-percentiles. All explanatory variables are lagged by one year. Robust test statistics calculated with group-clustered standard errors are reported below coefficients. Significance levels are indicated as: * = 0.1, ** = 0.05, *** = 0.01.

Table 21 Regression results: ESG category scores and IV (full sample)

To further break down the results, **Table 23** shows the coefficients of interest within the respective specifications shown in **Table 21** applied on a US and a European sub-sample. The results confirm the general associations of the full sample models in **Table 19**. Differences occur primarily in the coefficients of the environmental category scores. While Emissions and Resource use scores are significantly associated with lower IV in the European sample, the coefficient for Innovation scores is insignificant. Whereas in the US sample, the Innovation score is associated with lower IV and the coefficients for the Emissions and Resource use scores are insignificant. The category scores for HR and CSR are significant in the European sample and insignificant in the US sample. Conversely, the only coefficient indicating an increase in next year's IV is the category score SH in the US sample. It is insignificant in the European sample. The effect sizes of significant coefficients in the US sample are slightly smaller for category scores belonging to the E and S pillar. The coefficient for the MNG category score in the G pillar is more than twice as high as in the European sample.

Regression results: ESG category scores and IV (US and European sub-samples)

	IDIOSYNCRATIC VOLATILITY	
	US	EUROPE
EMM	0.0001	**-0.0067*** **
	t = 0.0610	**t = -4.2719**
INO	**-0.0027***	-0.002
	t = -1.8406	t = -1.5538
RES	-0.0006	**-0.0046*** **
	t = -0.3142	**t = -2.9022**
COM	**-0.0049*** **	**-0.0043*** **
	t = -2.8580	**t = -2.7523**
HR	-0.0007	**-0.0038*** **
	t = -0.4248	**t = -3.0519**
PRD	**-0.0035** **	**-0.0045*** **
	t = -2.1777	**t = -3.1112**
WF	**-0.0060*** **	**-0.0070*** **
	t = -3.6796	**t = -3.5777**
CSR	-0.0003	**-0.0054*** **
	t = -0.1731	**t = -3.2716**
MNG	**-0.0078*** **	**-0.0033** **
	t = -6.2155	**t = -2.3734**
SH	**0.0036*** **	0.0009
	t = 2.8193	t = 0.7390

Notes: Idiosyncratic volatility is the yearly standard deviation of the residual risk derived from a Carhart four-factor model. EMM, INO, RES, COM, HR, PRD, WF, CSR, MNG, and SH are category scores for Emissions, Innovation, Resource use, Community, Human rights, Product Responsibility, Workforce, CSR strategy, Management, and Shareholders provided by Refinitiv (standardized based on yearly firm size deciles). Idiosyncratic volatility is the yearly standard deviation of the residual risk derived from a Carhart four-factor model. Control variables are: CapEx intensity is capital expenditures divided by total assets. Firm age is the number of years since the firm's IPO. Firm size (in $US million) is the natural logarithm of total assets reported on the balance sheet. IA intensity is the ratio of intangible assets and total assets. Return on assets is net income divided by the book value of total assets. ROA volatility is the standard deviation of ROA over the last three fiscal years. R&D intensity is R&D expenditures divided by total assets. Leverage is the ratio of long- and short-term debt and shareholders' equity. Business sectors, years, and countries are fixed. Values reported in local currencies are converted into USD. Variables are winsorized at the yearly 1%- and 99%-percentiles. All explanatory variables are lagged by one year. Robust test statistics calculated with group-clustered standard errors are reported below coefficients. Significance levels are indicated as: * = 0.1, ** = 0.05, *** = 0.01.

Table 22 Regression results: ESG category scores and IV (US and European sub-samples)

Considering the materiality of ESG issues might differ on an industry basis, **Table 23** shows the regression coefficients for ESG overall scores within specifications applied to different TRBC economic sectors. The values correspond to the coefficients of the size-adjusted overall ESG score applied in specification (3). Most results confirm the negative association between overall ESG performance and future IV. Except for the Consumer Cyclicals and Utilities sector in the US sub-sample and the Energy sector in both sub-samples, the coefficients are negative. However, not all coefficients are significant, indicating that overall ESG performance might not be a sufficient predictor of IV in every sector or region investigated.

Regression results: ESG scores and IV (TRBC economic sector sub-samples)

	IDIOSYNCRATIC VOLATILITY	
	US	EUROPE
ENERGY	0.0109	0.0035
	t = 1.0118	t = 0.6737
BASIC MATERIALS	**-0.0124***	-0.004
	t = -1.6777	t = -0.9842
INDUSTRIALS	**-0.0098***	**-0.0186*****
	t = -1.8102	**t = -3.6179**
CONSUMER CYCL.	0.0034	-0.0094
	t = 0.5105	t = -1.5895
CONSUMER NON-CYCL.	**-0.0107*****	**-0.0097****
	t = -3.2727	**t = -2.4713**
FINANCIALS	**-0.0116*****	**-0.0103****
	t = -3.5687	**t = -1.9951**
HEALTHCARE	**-0.0083****	-0.0034
	t = -2.2948	t = -0.7434
TECHNOLOGY	-0.0048	**-0.0112***
	t = -1.5812	**t = -1.8503**
UTILITIES	0.006	-0.0036
	t = 1.1611	t = -0.9857
REAL ESTATE	**-0.0100****	-0.0062
	t = -2.3789	t = -1.0949

Notes: Idiosyncratic volatility is the yearly standard deviation of the residual risk derived from a Carhart four-factor model. ESG scores are Refinitiv ESG performance scores as provided by Refinitiv (standardized based on yearly firm size deciles). Idiosyncratic volatility is the yearly standard deviation of the residual risk derived from a Carhart four-factor model. Control variables are: CapEx intensity is capital expenditures divided by total assets. Firm age is the number of years since the firm's IPO. Firm size (in $US million) is the natural logarithm of total assets reported on the balance sheet. IA intensity is the ratio of intangible assets and total assets. Return on assets is net income divided by the book value of total assets. ROA volatility is the standard deviation of ROA over the last three fiscal years. R&D intensity is R&D expenditures divided by total assets. Leverage is the ratio of long- and short-term debt and shareholders' equity. Business sectors, years, and countries are fixed. Values reported in local currencies are converted into USD. Variables are winsorized at the yearly 1%- and 99%-percentiles. All explanatory variables are lagged by one year. Robust test statistics calculated with group-clustered standard errors are reported below coefficients. Significance levels are indicated as: * = 0.1, ** = 0.05, *** = 0.01.

Table 23 Regression results: ESG scores and IV (TRBC economic sector sub-samples)

In order to investigate the effect of variation in the materiality of ESG performance, **Figure 33** depicts the value of the coefficient for the ESG overall scores over time. The results are based on model (3) and applied to different sub-samples of two- to six-year periods selected from the full sample of firm-year observations (see **Appendix 12** for detailed results). The coefficient slightly varies over the observed periods. It ranges from roughly -0.016 to -0.003. Hence, the coefficients indicate an average decrease in next year's IV from 0.3 to 1.6% per one standard deviation increase in size-adjusted ESG performance. The black line identifies yearly regression model results based

on 4-year rolling window sub-samples. The dotted lines result from the coefficients within 2, 3, 5, or 6-year-rolling window specifications. Because of the different lengths of the underlying rolling sub-samples, the number of models being calculated for each year is smaller the more years are covered in the rolling window. The coefficients roughly follow a similar path and indicate that the negative association between ESG performance and IV increased over time. The graphs show two substantial drops. One occurs in the years 2008 and 2009; the other one occurs after 2014.

ESG scores and IV: Coefficients of ESG score in rolling window panel regressions

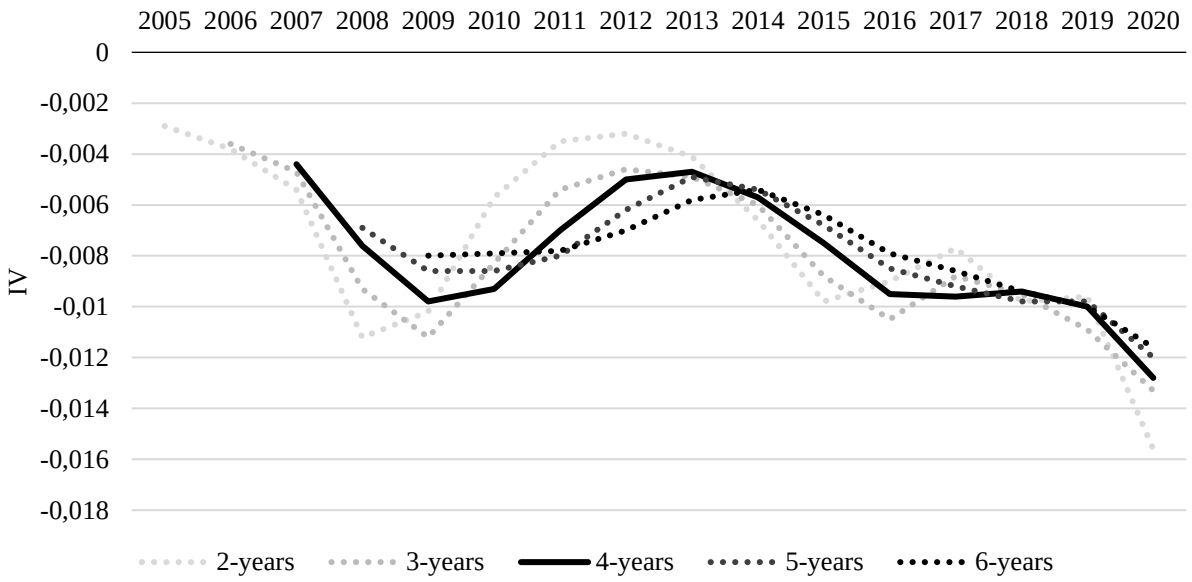


Figure 33 ESG scores and IV: Coefficient of ESG score in rolling window panel regressions

The regression models above analyze the association between IV, overall ESG scores, and single ESG category scores. The results show variation among different industries and over time. In combination of all these approaches, this study applies regression analysis on models incorporating single ESG category scores within rolling 4-year sub-samples for firms in each of the TRBC economic sectors individually. This approach allows conducting 1,300 regression models that depict the relationship of single ESG categories (10) within each economic sector (10) over thirteen rolling window sub-samples. The results indicate how certain ESG conduct measured in Refinitiv ESG scores is associated with future IV while accounting for industry-specific materiality that tends to differ over time. In sum, 465 of all 1,300 regression models (35.8%) calculate a negative association between ESG category scores and IV of the next year; another 119 (9.2%) of the regression models indicate a positive relationship. Consequently, a large fraction of 44.4% (584) of the regression models does not find a significant relationship.

Figure 34 depicts the number of regression models in which the coefficient for ESG category scores is significantly above 0 (bold lines) or below zero (dotted lines). The values indicate how

many of the coefficients are significant in each year based on the 4-year rolling window incorporating the respective year. As stated above, the number of regression models indicating a positive link between ESG category scores and future IV is lower than that of models indicating a negative relationship. Comparing the US and European sub-samples, the number of significant regression coefficients signaling a negative link is slightly higher based on US firms starting from 2011. Considering the total amount of models that are calculated for each year and regional sub-sample (100), the number of significant coefficients ranges rather low. Nevertheless, there seem to be systematic patterns in the industry distribution of significant coefficients when summed up over all 13 years.

ESG category scores and IV: Results of 4-year rolling window panel regressions: Number of significant negative and positive coefficients ($\beta_{ESG} < 0$ and $\beta_{ESG} > 0$)

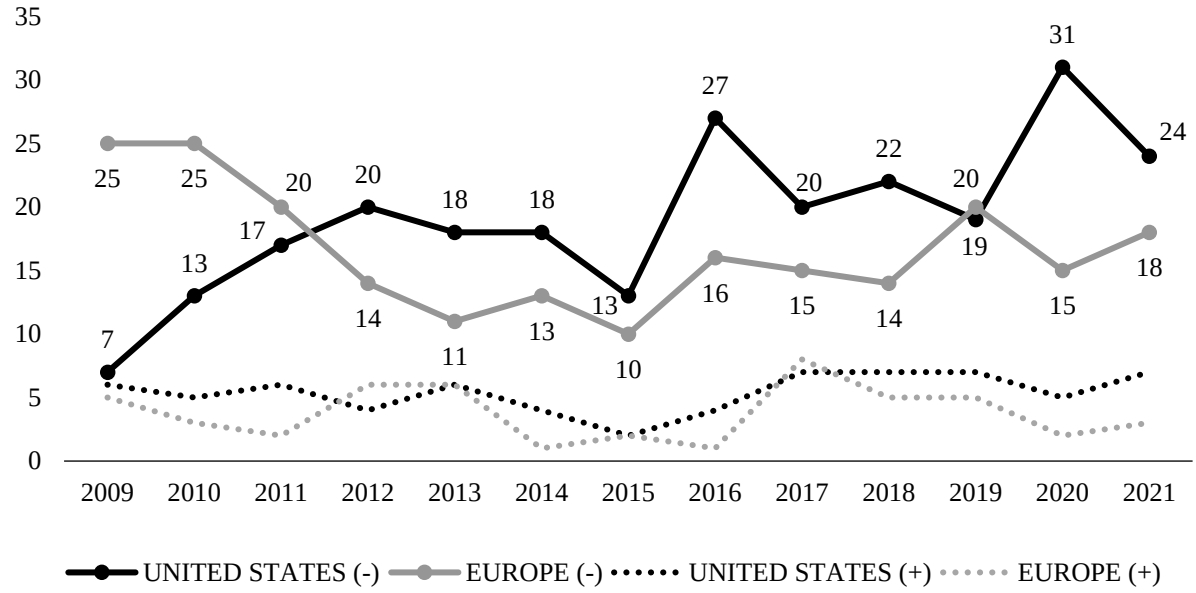


Figure 34 ESG category scores and IV: Results of 4-year rolling window panel regressions: Number of significant negative and positive coefficients ($\beta_{ESG} < 0$ and $\beta_{ESG} > 0$)

Table 24 and **Table 25** show the fraction of significant coefficients per category score, economic sector, and regional sub-sample. Each fraction is based on 13 four-year rolling window regression models from 2009 to 2021. The value indicates the number of significant negative (**Table 24**) and positive (**Table 25**) coefficients in the respective regression models. Hence, it shows how the direction of the link between ESG scoring and future IV most likely trends. The number of regression models indicating a negative link ranges from 0 to 12 (i.e., 91.7%). The number of regression models indicating a positive link range from 0 to 6 (i.e., 46.1%).

ESG category scores and IV: Results of 4-year rolling window panel regressions: Fraction of significant negative coefficients ($\beta_{ESG} < 0$) in US and European sub-samples

US	INDUSTRIALS	BASIC MATERIALS	CONSUMER CYCL.	CONSUMER NON-CYCL.	TECHNOLOGY	UTILITIES	ENERGY	HEALTHCARE	FINANCIALS	REAL ESTATE
EMM	-25%	0%	-42%	-33%	-25%	-33%	0%	0%	-8%	-17%
INO	-67%	-75%	-17%	0%	0%	-33%	-8%	-92%	-17%	-17%
RES	-33%	-58%	-17%	-25%	0%	-33%	0%	-17%	-8%	-17%
COM	-83%	0%	0%	0%	-58%	-42%	0%	-8%	-8%	0%
HR	-25%	-50%	0%	0%	-33%	-17%	0%	-75%	-8%	0%
PRD	-58%	-50%	-8%	-17%	-58%	-33%	0%	-42%	0%	0%
WF	-58%	-17%	-25%	0%	-50%	-33%	0%	-17%	0%	-17%
CSR	-25%	-17%	-33%	-17%	0%	-58%	0%	-8%	-8%	-17%
MNG	-33%	0%	-75%	0%	-58%	0%	0%	-42%	-17%	-8%
SH	0%	0%	-33%	0%	0%	-17%	0%	0%	0%	0%
EUROPE	**INDUSTRIALS**	**BASIC MATERIALS**	**CONSUMER CYCL.**	**CONSUMER NON-CYCL.**	**TECHNOLOGY**	**UTILITIES**	**ENERGY**	**HEALTHCARE**	**FINANCIALS**	**REAL ESTATE**
EMM	-38%	0%	-46%	-69%	-54%	0%	-15%	-15%	0%	-31%
INO	0%	-31%	-23%	-46%	0%	-23%	0%	0%	-8%	0%
RES	-31%	0%	-54%	-23%	-31%	-15%	0%	-23%	0%	-15%
COM	-62%	-8%	-38%	-15%	-31%	-23%	-31%	0%	0%	-15%
HR	-31%	-15%	-15%	-15%	0%	0%	-31%	-23%	0%	-23%
PRD	-8%	-8%	-54%	-23%	0%	0%	0%	0%	0%	-8%
WF	-46%	0%	-31%	-31%	-15%	-8%	-8%	-54%	0%	0%
CSR	-31%	-8%	-69%	-38%	0%	-8%	-15%	-31%	0%	-38%
MNG	0%	-8%	0%	0%	-38%	0%	-23%	0%	-46%	-15%
SH	0%	0%	0%	0%	0%	0%	-8%	0%	0%	-23%

Table 24 ESG category scores and IV: Results of 4-year rolling window panel regressions: Fraction of significant negative coefficients ($\beta_{ESG} < 0$) in US and European sub-samples

ESG category scores and IV: Results of 4-year rolling window panel regressions: Fraction of significant positive coefficients ($\beta_{ESG} > 0$) in US and European sub-samples

US	INDUSTRIALS	BASIC MATERIALS	CONSUMER CYCL.	CONSUMER NON-CYCL.	TECHNOLOGY	UTILITIES	ENERGY	HEALTHCARE	FINANCIALS	REAL ESTATE
EMM	0%	0%	0%	0%	0%	0%	0%	0%	15%	15%
INO	0%	0%	8%	0%	0%	0%	0%	0%	23%	0%
RES	0%	0%	0%	0%	0%	0%	23%	0%	0%	31%
COM	0%	0%	0%	0%	0%	0%	0%	8%	31%	31%
HR	0%	0%	15%	0%	0%	8%	46%	0%	23%	38%
PRD	0%	0%	0%	0%	0%	15%	23%	0%	38%	31%
WF	0%	0%	0%	0%	0%	0%	0%	0%	0%	23%
CSR	0%	0%	0%	0%	8%	15%	0%	0%	0%	38%
MNG	0%	0%	0%	8%	0%	0%	0%	0%	0%	0%
SH	0%	8%	0%	15%	0%	0%	0%	0%	0%	0%
EUROPE	**INDUSTRIALS**	**BASIC MATERIALS**	**CONSUMER CYCL.**	**CONSUMER NON-CYCL.**	**TECHNOLOGY**	**UTILITIES**	**ENERGY**	**HEALTHCARE**	**FINANCIALS**	**REAL ESTATE**
EMM	0%	15%	0%	0%	0%	15%	0%	0%	0%	0%
INO	0%	0%	0%	0%	0%	0%	0%	0%	0%	15%
RES	0%	31%	8%	0%	0%	8%	0%	0%	0%	0%
COM	0%	15%	0%	23%	0%	0%	0%	8%	0%	0%
HR	0%	8%	0%	0%	0%	15%	0%	0%	0%	0%
PRD	0%	0%	0%	0%	0%	0%	0%	15%	15%	0%
WF	0%	23%	0%	0%	0%	8%	0%	0%	0%	0%
CSR	0%	8%	0%	0%	0%	8%	0%	0%	15%	0%
MNG	0%	0%	31%	8%	0%	0%	0%	0%	0%	0%
SH	8%	0%	0%	38%	8%	0%	8%	0%	23%	0%

Table 25 ESG category scores and IV: Results of 4-year rolling window panel regressions: Fraction of significant positive coefficients ($\beta_{ESG} > 0$) in US and European sub-samples

It is reasonable to assume that the relationship between ESG categories and IV makes it possible to identify companies for which information on ESG performance is likely to influence stock pricing. Hence, the distinction between industry and regional levels over time is comparable to an assessment of material ESG information across industries and regions. As such, this study assumes the tables depict market-based materiality patterns of individual ESG categories over the observed period (2006 to 2022).

Most of the category scores show a negative association with future IV. Positive links are predominately found in the US sample, where firms with higher ratings in the Financials and Real Estate sector experience higher levels of future IV. For European firms, this can be seen in the sectors of Basic materials, Consumer Cyclicals, and Non-Cyclicals. The positive association in the US sample tends to occur slightly more frequently in E and S categories than in the Governance pillar. In the European sample, however, some models indicate a positive link for Management and Shareholder category scores. Further, robust materiality patterns for both regional samples are found for Management and Shareholder scores in the Consumer Non-Cyclicals sector. Nevertheless, the number of models that indicate a positive relationship remains relatively low compared to cases of a negative link.

Significant negative regression coefficients for several ESG category scores can be found in almost all economic sectors of both samples, except for the Energy sector in the US sample. In the Environmental pillar, more coefficients for the Emission score are significant, whereas, in the US sample, the fractions are slightly higher for the Innovations and Resource Use scores. For European firms of the Industrials, Consumer Cyclicals and Non-Cyclicals, and the Technology sector, Emission scores tend to provide robust indications of future IV. Especially US firms in the Healthcare, Basic Materials, and Industrials sectors seem to experience higher materiality of Environmental Innovations. In the social pillar, almost all category scores tend to provide material information. Some results stick out, as Community and Human rights show significant relationships to IV in the European sample for some sectors that are insignificant throughout all models within the US sample (e.g., Consumer Cyclicals, Consumer Non-Cyclicals). Community scores in the Industrials sector show a consistent capability to explain IV in both regional samples. Product Responsibility scores are more often negatively related to future IV in the US sample. Governance pillar scores show similar regionally divided associations. While CSR strategy scores explain IV in the European sample almost exclusively, the US sample shows more significant coefficients for the Management score. Management scores, however, tend to provide material information to European Financial firms, as it is almost the only

category score that shows significant relationships to IV over the observed time period within this sector.

Materiality adjustment

As already stated, one can assume a significant positive or negative coefficient provides period-specific information about the relative importance of ESG categories for firms in a specific region and economic sector. Hence, aggregation based on the varying yearly sector- and regional-specific significant results could provide a more concise picture of dynamic materiality patterns. Further, the information about a negative ($\beta_{ESG} < 0$) or positive ($\beta_{ESG} > 0$) link between category scores and future IV also allows for assessing the historical accuracy of ESG scores. This study applies the newly gained knowledge on the accuracy of each ESG category score to reweight category scores according to their yearly industry-based and regional materiality patterns. Hence, this study aggregates the category scores to new *materiality-adjusted ESG scores* based on the historical accuracy of the original Refinitiv ESG performance scores.

The approach is data-driven and follows an investor-like approach that: 1) assesses *current materiality patterns* and 2) balances current ESG category scores based on their *historical accuracy* in providing material information. Following this idea, this study applies even weights to each category score but adjusts those yearly scores of sectors in a region that showed a significant relationship in the analysis of the former year. **Appendix 13** displays the process and shows that it is not intended to optimize ESG scores based on their current relationship to an outcome variable (here IV) but instead seeks to work with data an investor would have at hand at a given point in time. The approach assumes that an investor would be interested in identifying firms with lower IV when using higher scores as a screening tool in decision-making. Therefore, each yearly category score experiencing a negative association with IV is adjusted to reflect better performance; positive associations lead to a worse score.

The reweighting is based on the original Refinitiv ESG category scores. For each score being adjusted upwards, an extra 100 points are added to the original score. Each score, which is adjusted negatively, is inversed. This allows matching the general preposition induced by Refinitiv's benchmarking approach via percentile rankings. Percentile rankings provide information on how many firms are worse or act exactly as good as the ranked firm. By adjusting the original scores based on the maximum level of 100, the materiality-adjusted scores still provide this information but account for firm performance in material ESG categories by either

inversing the score (i.e., $ESG_{mat} = 100 - ESG_{raw}$) or making the firm better than any other firm in the group of firms underlying the benchmark (i.e., $ESG_{mat} = 100 + ESG_{raw}$).

Figure 35 shows the relationship between the new materiality-adjusted ESG overall score ("ESGmat") and IV in different ESG performance deciles. It is a reproduction of **Figure 32** at the beginning of the chapter. However, it shows the difference of average IV based on portfolios sorted with the original score ("ESG") and the materiality-adjusted score ("ESGmat"). In addition, the dotted line shows the respective relationship based on a limited sample of firms that are fully covered by the Refinitiv ESG rating universe (i.e., firms that are not intentionally assigned category scores of 0 due to non-reported data). A difference of 0 would indicate that the portfolios assembled using the original ESG score and the materiality-adjusted score have the same average IV. However, that is not the case.

The differences indicate that the low ESG performance deciles are experiencing higher levels of IV, while the high ESG performance deciles indicate a relatively lower IV. The differences range from about -1 percentage point to almost +5 in average IV. Further, the conversion from a positive to a negative effect of the materiality adjustment happens between the third and fourth (full sample) and the fifth decile (limited sample). Hence, the materiality-adjusted score is feasible in assigning more firms that experience higher IV into low-performance deciles and more firms that experience lower IV into better-performing deciles.

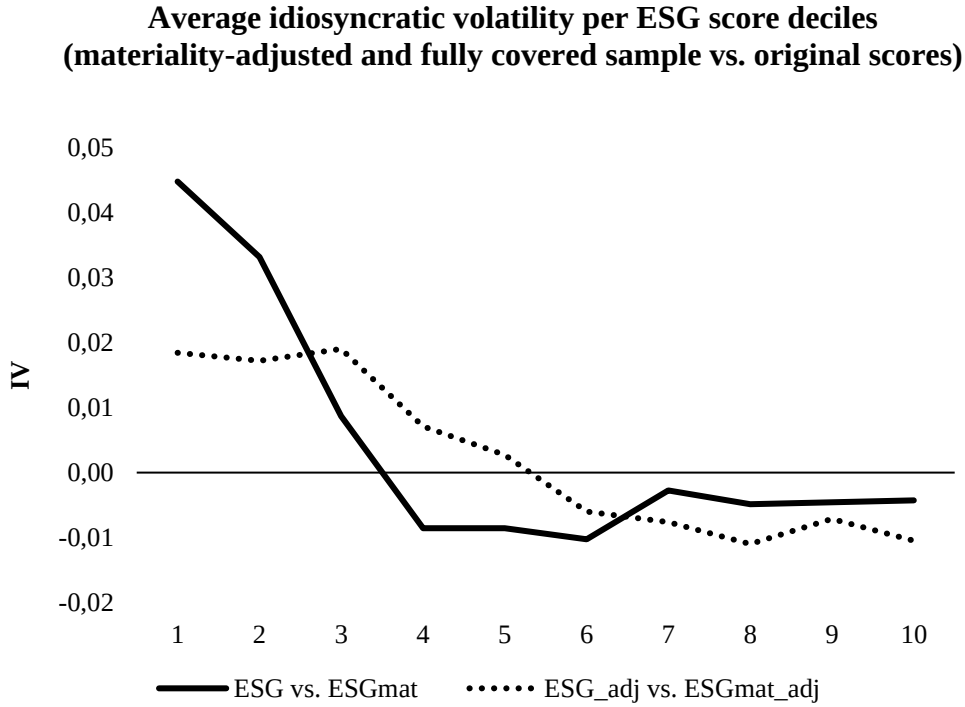


Figure 35 Average idiosyncratic volatility per ESG score deciles (materiality-adjusted and fully covered sample vs. original scores)

Discussion

Current research focuses on the ability of ESG metrics to measure ESG performance oriented on the actual outcome of corporate behavior on ESG issues.[690] ESG performance metrics are found to fall short of observing those outcomes as they are primarily based on input-level corporate behavior, like monetary units spent and policies implemented to address specific issues (see chapter 3.3.3).[691] The empirical investigation in the previous section is based on whether rather input-oriented ESG performance metrics are able to capture financial material information that materializes in lower risk properties in the future. Investors are found to consistently believe in ESG information's ability to identify those properties of lower risk.[692] Following Luo and Bhattacharya (2009), Sassen et al. (2016), Kaiser (2020), Reber, Gold, and Gold (2021), and Boucher et al. (2022), this study analyzes idiosyncratic volatility (IV) as an ex-post measure of a firm's exposure to events and circumstances posing firm-specific effects on market valuation.[693]

The results in the previous section show that original Refinitiv ESG performance scores are inversely related to IV. The top-performing firms tend to experience the lowest levels of IV. This is in line with previous literature utilizing IV as a measure of firm-specific risk.[694] However, Sahin et al. (2022a) and Drempetic, Klein, and Zwergel (2020) show that Refinitiv ESG scores face potential biases by Refinitiv's approach to missing ESG information and a firm's ability to report on ESG-related corporate conduct.[695] Refinitiv handles non-reported ESG information by assigning the worst possible performance, i.e., a value of 0, to such data points, under the assumption that positive measures would have been reported if undertaken. Thereby, Refinitiv seeks to promote ESG reporting and enhance the scope of firms' ESG disclosure. Considering larger firms, however, tend to face greater visibility while simultaneously being able to place more resources on reporting ESG information crucial to the rating methodology, there are two potential sources of biases posing a severe risk to the objective accuracy of Refinitiv's ESG performance metrics: Transparency and firm size.

It is reasonable to assume that firm visibility is a significant driver in Refinitiv's capability to capture ESG information and apply its rating methodology to it. Thus, it is no surprise that the

[690] cf. Serafeim/Zochowski/Downing (2019, p. 16)
[691] cf. Ebrahim/Rangan (2014, p. 123)
[692] cf. Amel-Zadeh/Serafeim (2018, p. 28), cf. Hartzmark/Sussman (2019, pp. 2826-2827)
[693] cf. Luo/Bhattacharya (2009, p. 208), cf. Sassen et al. (2016, pp. 886), cf. Reber/Gold/Gold (2021, p. 8), cf. Boucher et al. (2022, p. 23)
[694] cf. Sharpe (1964, pp. 438-439), cf. Luo/Bhattacharya (2009, p. 200), cf. Sassen et al. (2016, pp. 874-875), cf. Kaiser (2020, p. 39), cf. Reber/Gold/Gold (2021, p. 4), cf. Boucher et al. (2022, pp. 4-5)
[695] cf. Sahin et al. (2022a, p. 11), cf. Drempetic/Klein/Zwergel (2020, p. 353)

results of this study indicate that the negative relationship between ESG performance and IV is lower after adjusting Refinitiv's original ESG scores for firm size. The adjustment is rather naïve, as it standardizes the yearly benchmarked ESG metrics on firm size deciles. However, it adds a second layer to the benchmark as firms with supposedly the same reporting capabilities and information environments are being compared to one another.

In the same sense, this study analyzes the link between ESG performance and IV on a restricted sample of firms fully covered by the Refinitiv rating methodology. This way, only firms that have not been assigned the lowest possible performance in one of the ESG category scores due to missing data are included. The results indicate that an increase in ESG scores within this group is associated with smaller decreases in IV, alleviating the inverse relationship between ESG performance and IV even more.

Former literature suggests that IV is negatively related to firm size. Fu (2009) and Ang et al. (2006) document the highest levels of IV for firms having a relatively small stake in the total market capitalization of their datasets.[696] This study supports this finding, as the results suggest that firm size is negatively associated with IV. However, as smaller firms tend to be systematically rated lower, the negative effect of ESG performance on IV is enlarged by peculiarities of the rating methodology and the capability of larger firms to report on a larger scope. As **Figure 35** shows, the average level of IV per firm size decile is lower in the sample restricted to fully covered firms only. Hence, the actual negative relationship between ESG performance measured with the Refinitiv ESG metric and future IV is potentially even smaller. Nevertheless, controlling for firm size within models (1) to (4) alleviates the concern leaving space for some ability of the rating to convey information relevant to IV. In order to prevent a significant portion of the firm-year observations from being dropped and to focus on the actual Refinitiv ESG performance metric, the results rely on the size-adjusted modification of the rating ("zadjESG").

The results suggest that the negative relationship between ESG performance and IV is subject to severe differences taking into account the ESG category, region, or industry. Liang and Renneboog (2017) state that firm conduct on ESG issues, which results from the trade-off between the interests of shareholders and other stakeholders, tends to vary according to the legal system in which a company is primarily working.[697] Different legal origins substantially alter the rules under which companies conduct business. In the United States, for example, common law and its post-settling mechanisms tend to promote expectations of firm behavior ex-post, if

[696] cf. Fu (2009, p. 24), cf. Ang et al. (2006, p. 287)
[697] cf. Liang/Renneboog (2017, p. 855)

at all. Civil law countries, which can be found most over Europe, however, are based on rules which reflect the social preferences for firm behavior ex-ante. As a result, the basis to validate ESG performance differs systematically over different regions.[698] This study finds evidence for regional variation depicted in **Table 23**. Most of the ESG pillar scores are negatively associated with future IV. An exemption is environmental performance in the US sample (6). The results oppose former empirical findings of studies relying on Refinitiv ESG scores in panel data settings. Boucher et al. (2022) and Sassen et al. (2016) find that the environmental and social pillar scores are a better predictor of IV than the Governance pillar score.[699] However, in line with Liang and Renneboog (2017), the results show that in the US sample, the environmental engagement of firms is assessed against expectations that are not a priori associated with firm-specific risks in the future. Christensen, Serafeim, and Sikochi (2022) find that governance-related issues tend to be defined more precisely and outcomes of certain conduct are more predictable over time than in the environmental and social pillar.[700] Hence, one can assume a lower level of unambiguous social preferences to act in a specific way on environmental issues drives the negative association between environmental performance and IV. As the best-in-class approach underlying the environmental scores incorporates firms with stricter regulations in other legal frameworks, the scoring methodology imposes less predictive power of environmental performance metrics on future IV. Godfrey, Merrill, and Hansen (2009) build upon Mattingly and Berman (2006) and distinguish between two dimensions of CSR activities. Technical CSR reflects actions on behalf of primary stakeholders (e.g., shareholders, employees, customers); institutional CSR refers to secondary stakeholders (e.g., society and community).[701] Godfrey (2005) and Godfrey, Merrill, and Hansen (2009) find that institutional CSR activities provide moral capital, which translates to insurance-like protection from adverse perceptions of the company. As there tend to be fewer rules on the extent of ESG engagement, the effect of rather technical ESG engagement might easily outweigh the insurance-like protection of institutional ESG activities. As a result, the underlying ESG engagement might be 1) diverse in its scope and 2) not able to indicate performance in ex-post risk-relevant issues properly.

The residual ESG performance coefficient provides additional proof for this interpretation. This study finds evidence that especially Governance pillar scores carry information that is relevant to the prediction of future IV. By orthogonalizing ESG pillar scores on the overall ESG score,

[698] cf. Liang/Renneboog (2017, p. 896)
[699] cf. Boucher et al. (2022, p. 1), cf. Sassen et al. (2016, p. 886)
[700] cf. Christensen/Serafeim/Sikochi (2022, pp. 159-160)
[701] cf. Mattingly/Berman (2006, pp. 34-37), cf. Godfrey (2005, pp. 777-778), cf. Godfrey/Merrill/Hansen (2009, p. 429)

this study calculates a firm-specific measure of residual ESG performance. The residual ESG coefficient is significant in some modifications, indicating substantial information within the remaining two pillars. However, observing the coefficient being insignificant applying the G pillar score (8) in the US sample shows that the remaining E and S pillars do not add informational content to the prediction of future IV.

Corporate Governance is substantially different in the way it addresses stakeholders and rather technical in its nature, as it is predominately concerned with primary stakeholders. However, it aims to reflect how efficiently potential conflicts of interest among primary and secondary stakeholders are managed. Hence, there are conceptual overlaps with other issues within the environmental and social pillars.[702] Considering the significant negative association of social pillar scores and IV in the US sample, one can assume the overlap of primary stakeholders addressed within the social pillar (e.g., employees) induces materiality. Thus, the expectations incorporated within the rating methodology tend to provide relevant information within the Social pillar scores. However, the Governance pillar scores are superiorly reflective of material information on ESG performance and render the other category scores insignificant when applied instead. Within the European sample, specification (11) indicates an insignificant coefficient of residual ESG performance when the variable of interest is the Social pillar score. Hence, the informational content of the Social pillar score tends to outweigh the other ESG pillars and is able to solely reflect the material information on actual ESG performance and the perceived fulfillment of social preferences and rules. Additionally, both coefficients, the Social pillar score in the European and the Governance pillar score in the US sample, roughly match the effect of the overall ESG score (5 and 9).

The US sample shows significant coefficients for R&D and CapEx intensity, while the European sample does not. McWilliams and Siegel (2000) and Luo and Bhattacharya (2009) state that CSR is strongly associated with intangible characteristics like reputation, quality, or reliability but also reflects business uncertainty regarding growth opportunities. Hence, such expenses are strongly correlated with CSR and future firm prospects, as they intend to promote innovation in processes and products.[703] The results suggest that the informational content on the uncertainty of adverse effects stemming from ESG issues is rather captured by traditional book-based performance measures, like R&D and Capital expenditures. In the European sample, however, ESG performance measures tend to convey such information.

[702] cf. Liang/Renneboog (2017, p. 858)
[703] cf. McWilliams/Siegel (2000, p. 605)

Table 21 provides results for the association of single category scores and future IV. To prevent the omitted variable bias potentially induced by only incorporating one category of the overall ESG performance, the regression models rely on the same residual ESG performance variable as the previous models. The results state significant differences in the predictive power of ESG category scores on future IV. Especially the Environmental pillar shows severe differences in the significance of single ESG pillars. While Emissions and Resource Use scores tend to provide significant information in the European sample, the Innovation score remains insignificant. The US sample states the inverse relationship with Innovation scores being the sole significant environmental score providing information on future IV.

Yet again, the results indicate the importance of potential standards that contribute to the conceptualization of an ESG scoring methodology. The Emission and Resource Use scores tend to reflect actual measurement of the environmental impact of the firm and its processes. The EC is eager to promote such rules and a general understanding of corporate environmental impact (e.g., by publishing the EU taxonomy). The Innovation score, however, focuses on "a company's capacity to reduce environmental costs … and thereby creating new market opportunities."[704] Hence, the innovation score rather focuses on forward-looking opportunities to fulfill customer needs rather than regulatory or social preferences on "… a company's commitment … toward reducing environmental emission" and "… a company's … capacity to reduce the use of materials, energy or water."[705]

Surprisingly, the Shareholder score in the US sample identifies firms as good performers, which experience higher IV in the subsequent year. As already stated, this study assumes the goal of an investor would be to reduce the IV within her portfolio by relying on those firms the rating scores better. However, the observed positive relationship might be associated with the expectations and social preferences underlying the rating methodology. Ferreira and Laux (2007) investigate the relationship between antitakeover provisions and IV. They identify IV as a measure of stock price informativeness, indicating the fraction of private information within stock prices. They argue that governance provisions implying strong investor protection through openness to the market of corporate control motivate uninformed ownership, which in turn encourages trading on relatively more profitable private information.[706] The Shareholder score is designed to quantify "… a company's effectiveness toward equal treatment of shareholders and

[704] Refinitiv (2022a, p. 22)
[705] ibid. (p. 22)
[706] cf. Ferreira/Laux (2007, p. 952)

the use of anti-takeover devices."[707] Hence, one can assume that higher scores identify companies that may be more open to outside shareholders. However, IV is positively related to the score as the openness encourages altering the cost-benefit relation of gathering additional private information on the stock and consequently raising the fraction of idiosyncratic price components.

Observing that all other ESG category scores indicate a negative relationship between the metrics and IV implies that Governance scores substantially differ from the other two pillars to some extent. Confirming Christensen, Serafeim, and Sikochi (2022), the results suggest that governance-related issues tend to be defined more precisely.[708] The rating reflects Refinitiv's perception of social expectations of good Corporate Governance. However, such perceptions might ultimately convey information already reflected in stock prices. IV is associated positively with the Shareholder score, reflecting how easy it is to gather additional private information. In the alternative case, the negative association states that a higher rating is associated with a lesser fraction of private information in the stock price. Those firms being rated better are the ones for which there is no motivation to gather additional information. Significant negative associations then indicate that the scoring methodology already reflects the information material to prices.

Kaiser (2020) introduces a materiality map based on the relationship between single category scores and future IV. According to Eccles et al. (2012), Khan, Serafeim, and Yoon (2016), and Gyönyörová, Stachoň, and Stašek (2021), industry affiliations systematically affect the materiality of ESG issues reflected in stock prices.[709] **Table 23** shows the overall ESG score is differently associated (mostly negative) with IV over various economic sectors. However, Rogers and Serafeim (2019) point out materiality tends to differ over time, implying that the link between ESG performance metrics and future IV could be subject to changes over time. **Figure 33** confirms this idea by showing how the negative association follows wave-like changes over time. The rolling window regressions allow calculating coefficients over particular subsets of years, reflecting the average association of ESG metrics and future IV during a specific period. Regardless of the number of years contributing to the respective rolling window, **Figure 33** shows the relationship to become slightly more negative, indicating that the overall ESG met-

[707] Refinitiv (2022a, p. 22)
[708] cf. Christensen/Serafeim/Sikochi (2022, pp. 159-160)
[709] cf. Eccles et al. (2012, pp. 70-71), cf. Khan/Serafeim/Yoon (2016, pp.1697-1698), cf. Kaiser (2020, p. 49), cf. Gyönyörová/Stachoň/Stašek (2021, p. 1)

rics are increasingly reflective of information informing stock prices. Lins, Servaes, and Tamayo (2017) argue that the benefits of building social capital through CSR activities are especially strong when there is a loss of trust, as observed during the financial crisis.[710] Hence, additional proof for the ability of ESG metrics to convey information material to stock markets is based on coefficients being substantially lower for the years around 2009 and 2010.

This study applies the 4-year rolling window to break down industry-specific materiality considerations further. It combines this idea with the finding of regional variation and applies the regression model for each category score on individual sub-samples of economic sectors and regions. **Figure 34** shows that the number of models indicating significant negative coefficients of interest is consistently higher than those with a positive association. While the number of significant positive coefficients of interest is relatively flat, the number of significant negative coefficients is slightly growing. The results indicate additional back up for the claim that some ESG category scores could potentially identify stocks that systematically encourage the collection of private information. Simultaneously, it creates an additional point that a negative link between ESG performance assessment and future IV is driven by the score's ability to convey material information.

This study adds to Kaiser (2020) and Sahin et al. (2022a) by reweighting the category scores to follow the dynamic materiality patterns for each year. The adjustment process is relatively simple; however, it helps to enhance the ESG metric's ability to accurately identify those firms experiencing higher levels of IV (lower rating) and lower levels of IV (higher rating) in the future (**Figure 35**). Further, as Sahin et al. (2022a) argue, the negative scoring of non-reported data has the potential to alter empirical findings substantially.[711] **Figure 35** shows that the switch between a positive link in the low end of the rating distribution to a negative link of ESG performance and IV happens between the third and fourth ESG performance decile. Within the sub-sample limited to fully covered firms, which are just indirectly affected by the non-reported data, the switch occurs exactly in the middle of the ESG performance distribution.[712] Hence, the materiality adjustment bears the potential to enhance the predictive accuracy of ESG metrics. However, Refinitiv's handling of non-reported data dilutes this relationship by inducing a tilt toward lower-ranked firms.

[710] cf. Lins/Servaes/Tamayo (2017, p. 1788)
[711] cf. Sahin et al. (2022a, p. 10)
[712] Due to the relative best-in-class scoring applied over all firms in the dataset (including those that were intentionally scored 0) scores are affected positively (see chapters 5.3.1 and 5.4.2 for further details).

5.5.2 ESG metrics and ESG controversies

Results

The total number of ESG controversies is 22.462 and 18.810 within the limited sample of fully covered firms (i.e., excluding firms that were assigned scores of 0 because of non-reported information). **Figure 36** depicts the number of controversies that results from sorting firms into performance quintiles according to the raw Refinitiv ESG scores. Three of the four graphs show the number of controversies in a specific ESG pillar that results from using the respective original pillar scores. Refinitiv provides the number of controversies per firm on several different ESG-related topics. These topics are already matched to one of the ESG categories and a pillar score of the Refinitiv ESG rating universe. As such, this study adds up the number of controversies within each category and pillar to analyze the total number of E, S, and G controversies (as described in chapter 5.4.2).

Number of controversies per respective ESG, E, S, and G score quintile

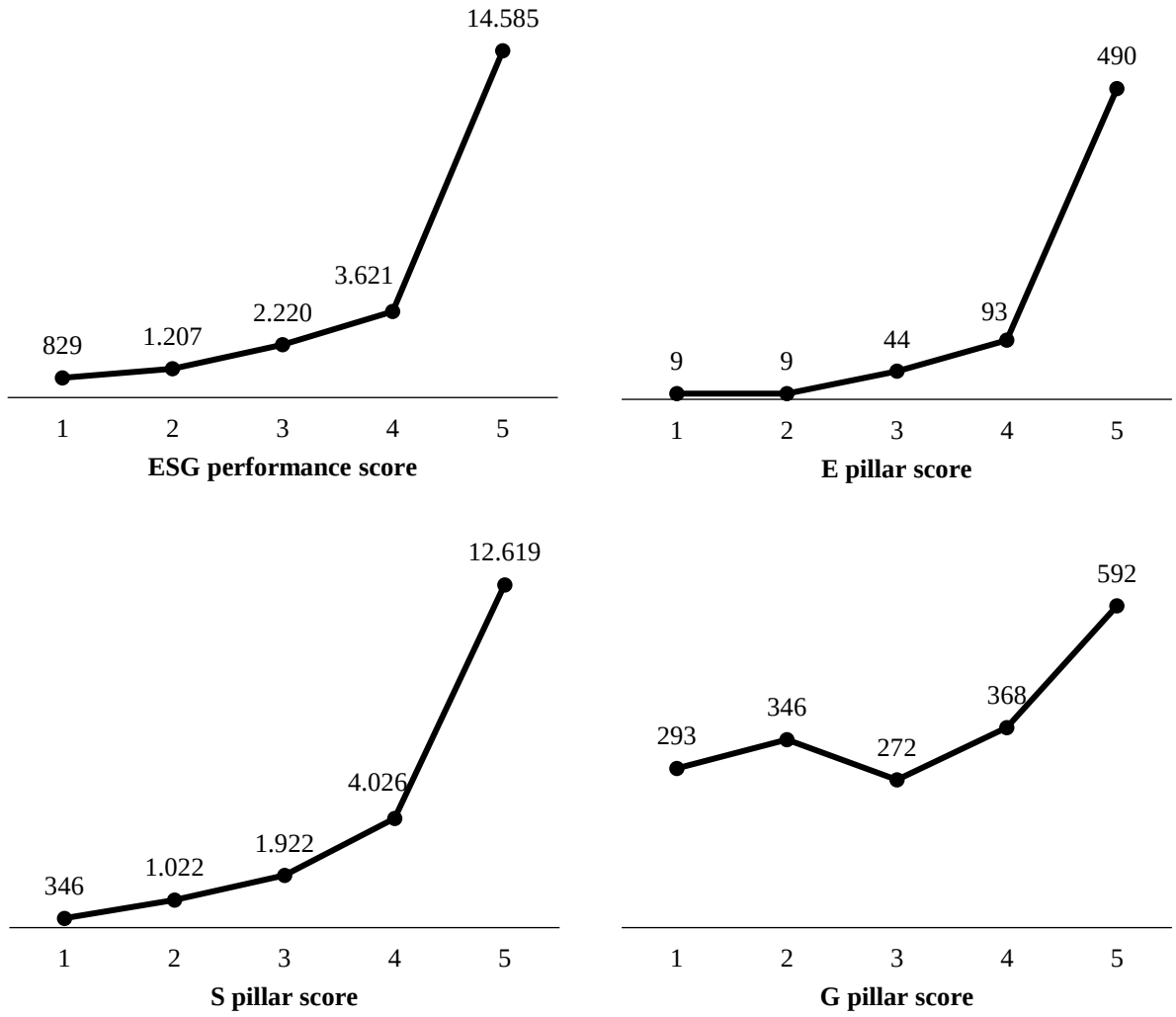


Figure 36 Number of controversies per respective ESG, E, S, and G score quintile

Each graph indicates that firms with the best ESG performance experience the highest number of controversies. Especially the overall ESG score, E, and S pillar scores identify firms as best performers that experience the highest number of controversies. There are drastic increases in the total number of controversies between the fourth and fifth score quintile. Creating a portfolio of the 20% best performing firms according to the raw Refinitiv Environmental pillar score leads to a five-fold exposure to environmental controversies as taking the best performing fourth quintile. Governance-related controversies are more evenly spread among the G pillar score quintiles. Nevertheless, there is a positive trend in the number of controversies indicating the best performers experience more controversies than their peers rated lower.

It is worthwhile remembering that the raw Refinitiv ESG score is strongly correlated to firm size, with larger firms tending to be rated better than smaller firms. Further, especially the ESG, E, and S scores are prone to the effect of the assignment of 0 to non-reported data affecting scores of firms with high coverage of relevant data points positively. Hence, one can assume the raw Refinitiv ESG score identifies large firms with high visibility, which puts the large number of controversies within the fifth performance quintile into perspective. The G pillar score consists of data points that are available for almost all firms in the ESG rating universe. As such, the effect of non-reported data is not that apparent. However, larger firms still tend to experience more controversies, which constitutes the observed distribution in **Figure 36**. In conclusion, one can hardly assume an inverse distribution of controversies among performance quintiles. Nevertheless, one can analyze whether the adjustment to size, materiality, and non-reported data leads to an enhancement in the detection of controversies. This study assumes an investor seeks to apply ESG metrics in order to mitigate the risk of exposure to potentially value-relevant ESG controversies. Therefore, a decrease in the positive association of ESG scores and controversies would indicate an improvement in the scoring of ESG performance.

Based on the total number of controversies experienced by each firm in the sample, **Figure 37** depicts the difference in the detection of controversial firms per ESG performance quintiles. In order to assess the effect of rating adjustments described above, the graphs depict the differences in the number of controversies being detected by using quintile portfolios based on the original ESG scores (ESG), the materiality-adjusted (“ESGmat”), the size-adjusted (“zadjESG”), and a size- and materiality-adjusted ESG score (“zadjESGmat”). Hence, the quintiles indicate the difference in the number of controversies of firms rated with the original Refinitiv ESG performance score compared to the firms rated equally but with an adjusted version of the rating accounting for features eventually inducing inaccuracies.

Using the adjusted scores over the original ESG scores leads to different results in all performance quintiles. Each specification indicates a significant drop in the number of controversies using the adjusted ratings in the fourth and fifth quintiles. The first through fourth quintile tend to identify portfolios of firms with a larger number of controversies. Most of the controversies being additionally detected are firms in the portfolios of the third and fourth ESG performance quintile. Using the size- and materiality-adjusted ESG score leads to a slight shift toward the third quintile, indicating that more controversies are detected for lower-performing firms than in the original ESG score.

Difference in the number of controversies per ESG performance quintile

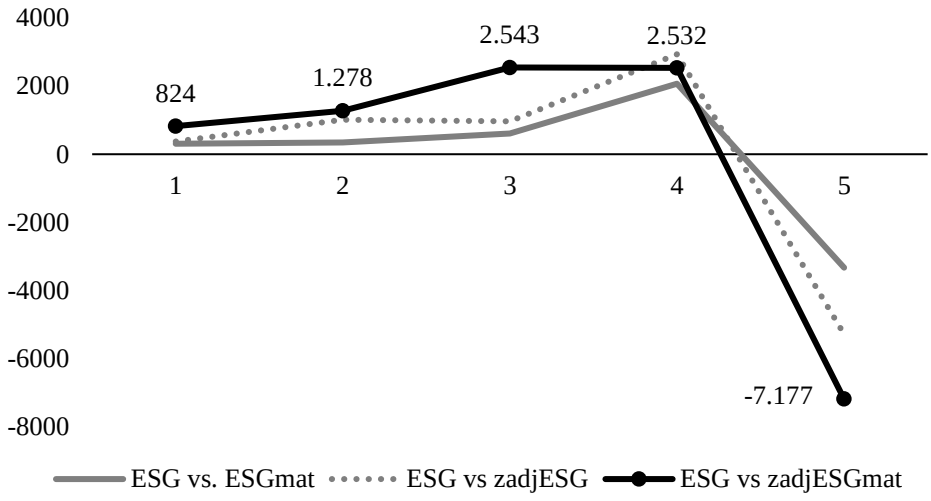


Figure 37 Difference in the number of controversies per ESG performance quintile

Figure 38 depicts the relationship already described above for the limited sample of fully covered firms. Each firm-year observation with at least one missing ESG pillar score (i.e., a value of 0) is excluded from the sample. The graphs follow roughly the same course as in the previous depiction of the full sample. However, the shift between more and fewer ESG controversies being detected is observed in the third quintile. Assuming an investor would prefer detecting most controversies in the lower quintiles and less in the higher performance quintiles, this indicates an improvement in the original ESG score by the adjustments introduced within this study. The materiality adjustment works better than the size adjustment in the full sample. It identifies more controversies in quintiles one, two, and three and less in the fourth and fifth quintiles. Hence, within the limited sample of fully covered firms, the materiality adjustment causes an improvement in comparison to the rather naïve size adjustment alone. Combining both adjustments leads to an even stronger improvement in the quantitative detection of firms involved in controversial business conduct.

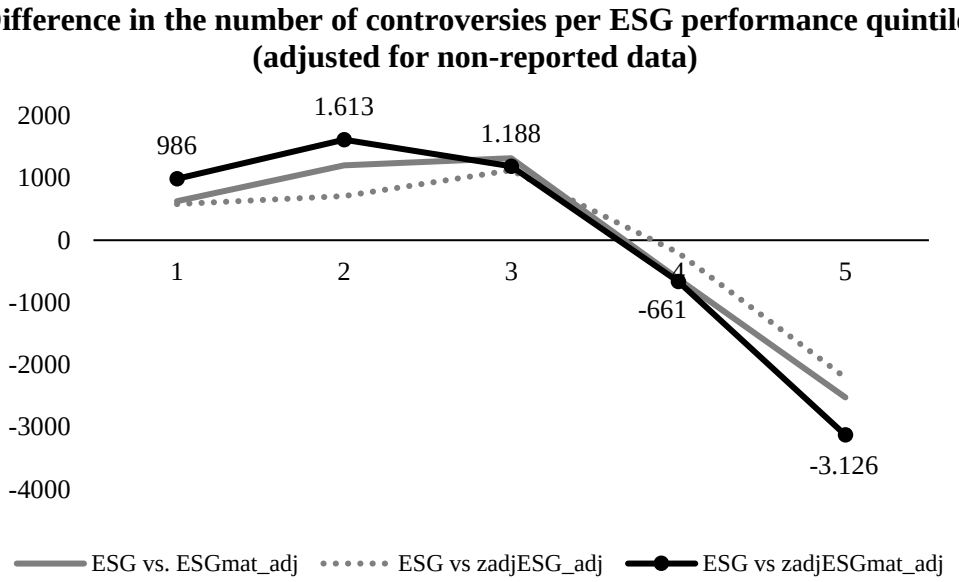


Figure 38 Difference in the number of controversies per ESG performance quintile (adjusted for non-reported data)

Figure 39 shows the percentage change in the detection of controversies using a materiality and a size-adjusted version of the original Refinitiv ESG rating. As seen in the two figures above, the number of controversies drastically changes after adjustments. The grey bars indicate that the adjustment enhances the number of controversies by up to 142% in the third ESG performance quintile as opposed to the portfolio of firms being ranked in the third quintile with the original ESG scores. The black bars are based on the sample restricted to fully covered firms. The number of controversies detected in the first quintile is up by 236% compared to the raw Refinitiv ESG score. Portfolios of the top-performing quintile experience 51% fewer controversies.

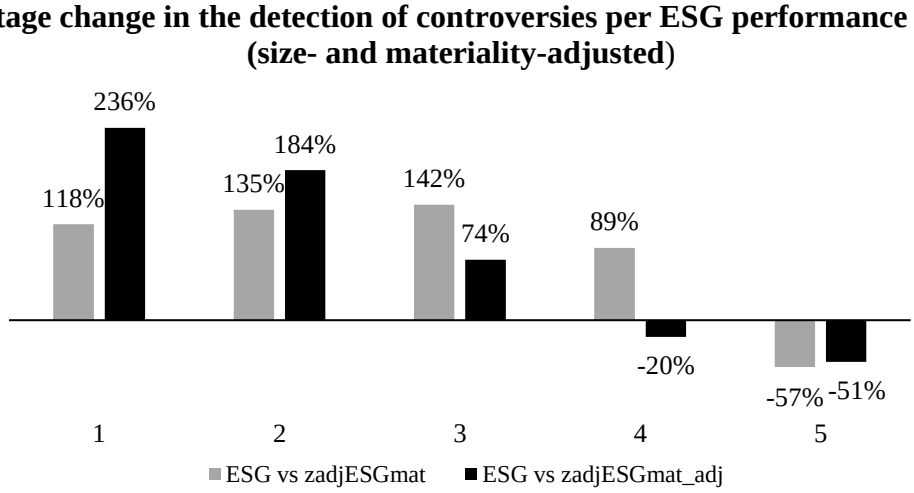


Figure 39 Percentage change in the detection of controversies per ESG performance quintile (size- and materiality-adjusted)

The scores above adjust for firm size and the historical accuracy in providing material ESG information reflected in future IV. Hence, the adjusted scores rely upon information an investor would have at hand when gaining access to the Refinitiv ESG database. The following models

analyze whether and how strongly the adjustments improve the ability of ESG metrics to accurately reflect a lower probability of being exposed to ESG controversies. **Table 26** provides results on logit regression models regressing several ESG rating specifications and control variables on the binary variable *STRIKE*. *STRIKE* takes a value of 1 if the firm experienced a strike during a given year. The ESG performance metrics are Refinitiv ESG scores, which are standardized on a yearly basis (ESG), standardized based on yearly firm size deciles (zadjESG), adjusted to its historical accuracy in providing material ESG information (ESGmat), and combinations of the two approaches (zadjESGmat). Each specification is also run on a subsample of fully covered firms, which are not subject to changes in their score due to Refinitiv's approach to handling non-reported data (ESG_adj). The regression coefficients are reported in odds ratios, reflecting the odds that a strike occurs given that the ESG performance score increases by one unit (see chapter 5.2.1 for further details).

The results confirm the findings from the figures above. An increase in the raw Refinitiv ESG score by one standard deviation is associated with a 46.3% increase in the odds of a strike (23). Hence, the higher the raw Refinitiv ESG score ranks a firm, the higher the probability of seeing this firm being involved in a strike in the next year. The adjustments are designed to enhance the accuracy and quality of the prediction of ESG controversies. As one would expect, the odds ratios relatively shrink with each step of adjustment. Especially the materiality-adjusted score (27) performs better in that sense, as a one standard deviation increase is associated with a lower increase in the odds of experiencing a strike of 12.5%.

A significant enhancement can be seen in the subset of fully covered firms (_adj). Within this sub-sample, an increase of one standard deviation in the materiality-adjusted ESG score (28) is associated with a 13.1% decrease in the odds of experiencing a strike. The effect is robust even when controlled based on a size-adjusted ESG metric (30). Similar results not reported within this study confirm these findings to a smaller extent for the Refinitiv Social pillar score, which conceptually relates most to factors eventually inducing a strike (e.g., employee satisfaction). The control variables indicate firm size, leverage, and CapEx intensity having a positive relationship with the log odds of experiencing a strike. In contrast, the log odds decrease in the case of higher profitability, R&D expenditures, and the intensity of intangible assets.

Logit regression results: ESG performance and STRIKES (full sample)

	STRIKES (1 = Yes)							
	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)
ESG	**1.4634*** **							
	t = 5.2156							
ESG_adj		1.0502						
		t = 0.6067						
zadjESG			**1.3950*** **					
			t = 4.7584					
zadjESG_adj				1.0723				
				t = 0.9469				
ESGmat					**1.1250** **			
					t = 2.0219			
ESGmat_adj						**0.8688** **		
						t = -2.0455		
zadjESGmat							1.0834	
							t = 1.3466	
zadjESGmat_adj								**0.8757** **
								t = -1.9782
CapEx intensity	5.7693	**10.7599*** **	6.0864	**11.1864*** **	5.1465	**9.8132*** **	4.9905	**9.2360*** **
	t = 1.5327	**t = 3.0330**	t = 1.5906	**t = 3.0504**	t = 1.4654	**t = 2.9984**	t = 1.4415	**t = 2.9673**
Firm age	1.0004	1.001	1.0006	1.001	1.0012	1.0013	1.0014	1.0013
	t = 0.3885	t = 0.8105	t = 0.5048	t = 0.7646	t = 1.1426	t = 1.0292	t = 1.2533	t = 1.0179
Firm size	**1.4333*** **	**1.7136*** **	**1.5768*** **	**1.7267*** **	**1.5621*** **	**1.7902*** **	**1.5991*** **	**1.7747*** **
	t = 7.9801	**t = 8.5454**	**t = 11.4607**	**t = 9.3443**	**t = 10.6604**	**t = 9.7439**	**t = 11.8650**	**t = 9.7211**
IA intensity	**0.0806*** **	**0.2303*** **	**0.0858*** **	**0.2313*** **	**0.0858*** **	**0.2343*** **	**0.0869*** **	**0.2307*** **
	t = -7.3765	**t = -3.2024**	**t = -7.2211**	**t = -3.1944**	**t = -7.2828**	**t = -3.1770**	**t = -7.2560**	**t = -3.2110**
ROA	**0.2737***	**0.1975***	**0.2790***	**0.1936***	0.3266	0.2289	0.3323	0.2226
	t = -1.8737	**t = -1.7257**	**t = -1.8507**	**t = -1.7483**	t = -1.6310	t = -1.5589	t = -1.6051	t = -1.5885
ROA volatility	0.2728	1.5501	0.2722	1.4916	0.3153	1.832	0.319	1.84
	t = -0.9618	t = 0.2370	t = -0.9695	t = 0.2164	t = -0.8699	t = 0.3298	t = -0.8636	t = 0.3318
R&D intensity	**0.0010** **	**0.0003** **	**0.0012** **	**0.0003** **	**0.0021** **	**0.0004***	**0.0022** **	**0.0004***
	t = -2.2781	**t = -1.9754**	**t = -2.2346**	**t = -1.9844**	**t = -2.0695**	**t = -1.9060**	**t = -2.0638**	**t = -1.9150**
Leverage	1.0363	**1.0579***	1.0361	**1.0583***	1.032	**1.0604***	1.0328	**1.0602***
	t = 1.4993	**t = 1.8061**	t = 1.4927	**t = 1.8166**	t = 1.3302	**t = 1.8700**	t = 1.3599	**t = 1.8679**
CountryFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
EconomicSectorFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
YearFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,920	3,053	5,920	3,053	5,920	3,053	5,920	3,053
Pseudo-R^2	0.2134	0.1821	0.2121	0.1823	0.2072	0.1837	0.2066	0.1836

Notes: STRIKES is a binary variable indicating whether the firm experienced a strikes or workers stoppage during a given year. ESG are (standardized) Refinitiv ESG performance scores as provided by Refinitiv. The suffix "_adj" indicates whether only fully covered firm-year observations are considered. The prefix "zadj" identifies ESG scores standardized based on yearly firm size deciles. The suffix "mat" identifies materiality-adjusted ESG scores based on the approach developed in chapter 5.5.1. CapEx intensity is capital expenditures divided by total assets. Firm age is the number of years since the firm's IPO. Firm size (in $US million) is the natural logarithm of total assets reported on the balance sheet. IA intensity is the ratio of intangible assets and total assets. Return on assets is net income divided by the book value of total assets. ROA volatility is the standard deviation of ROA over the last three fiscal years. R&D intensity is R&D expenditures divided by total assets. Leverage is the ratio of long- and short-term debt and shareholders' equity. Economic sectors, years, and countries are fixed. Values reported in local currencies are converted into USD. Variables are winsorized at the yearly 1%- and 99%-percentiles. All explanatory variables are lagged by one year. Robust test statistics calculated with group-clustered standard errors are reported below coefficients. Significance levels are indicated as: * = 0.1, ** = 0.05, *** = 0.01.

Table 26 Logit regression results: ESG performance and STRIKES (full sample)

Logit regression results: ESG performance and the occurrence of ESG controversies (full sample)

	CONTROVERSY ESG (1 = Yes)							
	(31)	(32)	(33)	(34)	(35)	(36)	(37)	(38)
ESG	**1.3255*****							
	t = 12.3633							
ESG_adj		**1.2086*****						
		t = 5.6480						
zadjESG			**1.3819*****					
			t = 15.1550					
zadjESG_adj				**1.2436*****				
				t = 7.0257				
ESGmat					**1.0499****			
					t = 2.3842			
ESGmat_adj						0.9845		
						t = -0.5155		
zadjESGmat							**1.1227*****	
							t = 5.6053	
zadjESGmat_adj								1.0159
								t = 0.5218
CapEx intensity	1.6545	2.8763	1.6228	2.6405	1.4604	3.5566	1.4735	3.5541
	t = 1.0538	t = 0.9500	t = 1.0095	t = 0.8728	t = 0.8031	t = 1.1394	t = 0.8193	t = 1.1400
Firm age	1.0003	1.0004	1.0001	1.0003	**1.0009****	1.0007	**1.0008***	1.0006
	t = 0.5870	t = 0.6183	t = 0.1692	t = 0.4692	**t = 2.0473**	t = 1.1205	**t = 1.6749**	t = 1.0729
Firm size	**2.3013*****	**2.5948*****	**2.4630*****	**2.6972*****	**2.5133*****	**2.7898*****	**2.5218*****	**2.7741*****
	t = 52.1416	**t = 33.1116**	**t = 64.1455**	**t = 37.2359**	**t = 60.5873**	**t = 37.5894**	**t = 65.7496**	**t = 38.1963**
IA intensity	**0.7140*****	**0.4395*****	**0.7403*****	**0.4454*****	**0.6843*****	**0.4509*****	**0.6954*****	**0.4515*****
	t = -3.3685	**t = -4.5217**	**t = -2.9984**	**t = -4.4429**	**t = -3.8210**	**t = -4.4065**	**t = -3.6531**	**t = -4.3979**
ROA	0.922	**0.3433****	0.8548	**0.3275****	1.1038	**0.4043****	1.0563	**0.3988****
	t = -0.4007	**t = -2.4314**	t = -0.7765	**t = -2.5423**	t = 0.4832	**t = -2.0608**	t = 0.2683	**t = -2.0929**
ROA volatility	1.0213	**2.6402*****	1.021	**2.6275*****	1.0231	**2.6676*****	1.0226	**2.6700*****
	t = 1.0531	**t = 3.3962**	t = 1.0508	**t = 3.4156**	t = 1.1272	**t = 3.2990**	t = 1.1052	**t = 3.2995**
R&D intensity	**2.8502*****	**25.8570*****	**2.6312*****	**21.3940*****	**3.6950*****	**47.3559*****	**3.5261*****	**45.2252*****
	t = 3.8983	**t = 2.9608**	**t = 3.6006**	**t = 2.7751**	**t = 4.5717**	**t = 3.5751**	**t = 4.4788**	**t = 3.5288**
Leverage	**0.9784*****	0.9885	**0.9784*****	0.9892	**0.9720*****	0.9833	**0.9724*****	0.9834
	t = -2.5962	t = -0.7987	**t = -2.5851**	t = -0.7444	**t = -3.3826**	t = -1.1650	**t = -3.3304**	t = -1.1536
CountryFE	No	No	Yes	No	No	Yes	No	No
BusinessSectorFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
YearFE	No	Yes	Yes	No	Yes	Yes	No	Yes
Observations	35,608	10,291	35,608	10,291	35,507	10,291	35,507	10,291
Pseudo-R^2	0.2895	0.2887	0.2925	0.2904	0.2855	0.2879	0.2860	0.2874

Notes: Controversy ESG is a binary variable indicating whether the firm experienced an ESG controversy during a given year. ESG are (standardized) Refinitiv ESG performance scores as provided by Refinitiv. The suffix "_adj" indicates whether only fully covered firm-year observations are considered. The prefix "zadj" identifies ESG scores standardized based on yearly firm size deciles. The suffix "mat" identifies materiality-adjusted ESG scores based on the approach developed in chapter 5.5.1. CapEx intensity is capital expenditures divided by total assets. Firm age is the number of years since the firm's IPO. Firm size (in $US million) is the natural logarithm of total assets reported on the balance sheet. IA intensity is the ratio of intangible assets and total assets. Return on assets is net income divided by the book value of total assets. ROA volatility is the standard deviation of ROA over the last three fiscal years. R&D intensity is R&D expenditures divided by total assets. Leverage is the ratio of long- and short-term debt and shareholders' equity. Business sectors, years, and countries are fixed. Values reported in local currencies are converted into USD. Variables are winsorized at the yearly 1%- and 99%-percentiles. All explanatory variables are lagged by one year. Robust test statistics calculated with group-clustered standard errors are reported below coefficients. Significance levels are indicated as: * = 0.1, ** = 0.05, *** = 0.01.

Table 27 Logit regression results: ESG performance and the occurrence of ESG controversies (full sample)

Table 27 shows the results of logit models on the odds of experiencing an ESG controversy. The variable *CONTROVERSY ESG* takes a value of 1 if the company is subject to controversial news coverage in a given year. All specifications (except for model 36) indicate that the applied ESG performance metric is positively associated with the odds of experiencing a controversy in the upcoming year. The increase in odds ranges from about 5% (35) to 38.2% (33) per one standard deviation increase in ESG performance. Unlike the results on *STRIKES*, the highest increase in odds is associated with size-adjusted ESG performance. Each specification (33, 34,

37, and 38) applying the size-adjustment indicates that an increase of one standard deviation within the corresponding firm size decile is associated with higher odds of experiencing an ESG controversy than within the rating standardizing firm-year observations over the entire sample. Using the materiality-adjusted ESG rating based on historical accuracy in predicting IV, the results still indicate a significant increase in the odds of experiencing controversial media coverage. However, the effect is much smaller. Although insignificant, Model 36 even indicates a negative association within the restricted sample of fully covered firms only. Firm size and R&D intensity are consistently positively associated with the occurrence of ESG controversies in the upcoming year. IA intensity is negatively associated with the odds of ESG controversies in all specifications. Within the restricted sample (32, 34, 36, and 38), the regression coefficients indicate a positive association of volatility in profits (ROA volatility) and negative effects of higher profitability. The unrestricted samples show a significant negative effect of a firm's debt on the odds of experiencing controversial news coverage.

Logit regression results: ESG scores and the occurrence of ESG controversies (US and European sub-samples)

	CONTROVERSY ESG (1 = Yes)	
	US	EUROPE
ESG	**1.2405*** **	**1.4457*** **
	t = 7.3805	**t = 9.4982**
ESG_adj	**1.1535*** **	**1.2503*** **
	t = 2.7414	**t = 4.9023**
zadjESG	**1.3330*** **	**1.4460*** **
	t = 10.5254	**t = 10.2082**
zadjESG_adj	**1.2127*** **	**1.2646*** **
	t = 3.9904	**t = 5.5753**
ESGmat	0.9988	**1.2448*** **
	t = -0.0450	**t = 5.9682**
ESGmat_adj	**0.9130***	**1.1665*** **
	t = -1.9434	**t = 3.3340**
zadjESGmat	**1.0934*** **	**1.2829*** **
	t = 3.2798	**t = 6.9649**
zadjESGmat_adj	0.9553	**1.1943*** **
	t = -0.9720	**t = 3.8837**

Notes: Controversy ESG is a binary variable indicating whether the firm experienced an ESG controversy during a given year. ESG are (standardized) Refinitiv ESG performance scores as provided by Refinitiv. The suffix "_adj" indicates whether only fully covered firm-year observations are considered. The prefix "zadj" identifies ESG scores standardized based on yearly firm size deciles. The suffix "mat" identifies materiality-adjusted ESG scores based on the approach developed in chapter 5.5.1. Control variables are: CapEx intensity is capital expenditures divided by total assets. Firm age is the number of years since the firm's IPO. Firm size (in $US million) is the natural logarithm of total assets reported on the balance sheet. IA intensity is the ratio of intangible assets and total assets. Return on assets is net income divided by the book value of total assets. ROA volatility is the standard deviation of ROA over the last three fiscal years. R&D intensity is R&D expenditures divided by total assets. Leverage is the ratio of long- and short-term debt and shareholders' equity. Business sectors, years, and countries are fixed. Values reported in local currencies are converted into USD. Variables are winsorized at the yearly 1%- and 99%-percentiles. All explanatory variables are lagged by one year. Robust test statistics calculated with group-clustered standard errors are reported below coefficients. Significance levels are indicated as: * = 0.1, ** = 0.05, *** = 0.01.

Table 28 Logit regression results: ESG scores and the occurrence of ESG controversies (US and European sub-samples)

Table 28 shows the coefficients of the ESG performance metrics based on the sample specifications in the aforementioned **Table 27** applied to US and European sub-samples of firms. The

coefficients indicate substantial differences. While all ESG performance metrics in the European sample are significantly associated with higher odds of experiencing ESG controversies, the US sample only partly indicates significant relationships. The effect sizes within the US sample are smaller, indicating a less positive effect of a one standard deviation increase in ESG performance on the odds of controversies in the upcoming year. Interestingly, the materiality-adjusted ESG ratings are associated with the lowest increase in the odds of ESG controversies. Within the restricted US sample of fully covered firms, a one standard deviation increase in materiality-adjusted ESG performance even leads to a significant decrease of 8.7%.

Regression results: ESG performance and number of ESG controversies (full sample)

	ESG CONTROVERSIES							
	(39)	(40)	(41)	(42)	(43)	(44)	(45)	(46)
ESG	0.0403							
	t = 0.2246							
ESG_adj		0.1102						
		t = 0.4388						
zadjESG			**0.2640***					
			t = 1.6894					
zadjESG_adj				0.3451				
				t = 1.6248				
ESGmat					0.0030			
					t = 0.0269			
ESGmat_adj						-0.1355		
						t = -0.7962		
zadjESGmat							0.1708	
							t = 1.5572	
zadjESGmat_adj								-0.0603
								t = -0.3597
CapEx intensity	-3.3666	**-20.1329****	-3.3036	**-20.3935****	-3.3569	**-19.7830****	-3.3071	**-19.8838****
	t = -1.3383	**t = -2.4232**	t = -1.3088	**t = -2.3923**	t = -1.3371	**t = -2.3243**	t = -1.3121	**t = -2.3237**
Firm age	-0.0029	-0.0065	-0.0036	-0.0068	-0.0028	-0.0062	-0.0034	-0.0063
	t = -0.9013	t = -1.5147	t = -1.1020	t = -1.5890	t = -0.8438	t = -1.4196	t = -1.0034	t = -1.4441
Firm size	**1.3523*****	**1.9053*****	**1.3374*****	**1.9006*****	**1.3658*****	**1.9693*****	**1.3497*****	**1.9491*****
	t = 6.6368	**t = 6.4421**	**t = 7.8321**	**t = 7.3201**	**t = 7.7875**	**t = 7.3809**	**t = 8.1142**	**t = 7.5009**
IA intensity	**-3.4834*****	**-5.9320*****	**-3.4208*****	**-5.9652*****	**-3.4864*****	**-5.8499*****	**-3.4456*****	**-5.8833*****
	t = -4.4800	**t = -3.7472**	**t = -4.3643**	**t = -3.6658**	**t = -4.5041**	**t = -3.5834**	**t = -4.4515**	**t = -3.5927**
ROA	1.1940	2.3275	1.0252	2.1608	1.2283	2.5305	1.1425	2.4633
	t = 0.7122	t = 0.5771	t = 0.6212	t = 0.5358	t = 0.7769	t = 0.6492	t = 0.7230	t = 0.6317
ROA volatility	0.2247	-0.1585	0.2329	-0.5557	0.2234	-0.0595	0.2250	-0.0205
	t = 0.9846	t = -0.0598	t = 1.0127	t = -0.2153	t = 0.9745	t = -0.0217	t = 0.9891	t = -0.0075
R&D intensity	2.0864	-13.3161	1.8338	**-14.1142***	**2.1279***	-12.6645	2.0043	-12.8418
	t = 1.6278	t = -1.6158	t = 1.4320	**t = -1.6835**	**t = 1.6474**	t = -1.4581	t = 1.5567	t = -1.4773
Leverage	0.0271	0.0717	0.0291	0.0800	0.0264	0.0669	0.0250	0.0675
	t = 0.4402	t = 0.6860	t = 0.4797	t = 0.7755	t = 0.4376	t = 0.6564	t = 0.4160	t = 0.6629
CountryFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
BusinessSectorFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
YearFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,297	2,538	5,297	2,538	5,297	2,538	5,297	2,538
R^2	0.1644	0.2164	0.1655	0.2178	0.1643	0.2165	0.1648	0.2163
Adjusted R^2	0.1574	0.2025	0.1585	0.204	0.1573	0.2026	0.1578	0.2024

Notes: ESG controversies is the number of incidents a firm is subject to controversial news coverage in a given year. ESG are (standardized) Refinitiv ESG performance scores as provided by Refinitiv. The suffix "_adj" indicates whether only fully covered firm-year observations are considered. The prefix "zadj" identifies ESG scores standardized based on yearly firm size deciles. The suffix "mat" identifies materiality-adjusted ESG scores based on the approach developed in chapter 5.5.1. CapEx intensity is capital expenditures divided by total assets. Firm age is the number of years since the firm's IPO. Firm size (in $US million) is the natural logarithm of total assets reported on the balance sheet. IA intensity is the ratio of intangible assets and total assets. Return on assets is net income divided by the book value of total assets. ROA volatility is the standard deviation of ROA over the last three fiscal years. R&D intensity is R&D expenditures divided by total assets. Leverage is the ratio of long- and short-term debt and shareholders' equity. Business sectors, years, and countries are fixed. Values reported in local currencies are converted into USD. Variables are winsorized at the yearly 1%- and 99%-percentiles. All explanatory variables are lagged by one year. Robust test statistics calculated with group-clustered standard errors are reported below coefficients. Significance levels are indicated as: * = 0.1, ** = 0.05, *** = 0.01.

Table 29 Regression results: ESG performance and number of ESG controversies (full sample)

Table 29 shows the results for a panel data regression of ESG performance metrics on the number of ESG controversies during a given year. The number of observations is fairly low compared to the models above, as it only covers those firm-year observations that experienced at least one controversy in a given year. The variables show that only the size-adjusted ESG performance (41) is significantly related to the number of ESG controversies. An increase in ESG performance of one standard deviation within a firm size decile is associated with 0.264 more ESG controversies in the next year. All other specifications show insignificant associations, although the coefficients indicate that at least the average association might turn negative in the restricted sample based on materiality-adjusted ESG metrics (44 and 46). Firm size is consistently associated positively with the number of ESG controversies, while the intensity of intangible assets shows a strong negative effect. CapEx intensity negatively affects the number of controversies in the restricted samples of fully covered firms (40, 42, 44, and 46).

In addition to taking the total number of ESG controversies, **Table 30** provides the regressions results of the ESG pillar scores on the respective number of ESG controversies within a particular pillar. The results are based on the same specifications as presented in **Table 29**.

Regression results: ESG, E, S, and G scores and the number of ESG controversies (full sample)

	ESG CONTROVERSIES		
	E	**S**	**G**
...	0.1118	0.0723	0.0449
	t = 0.6749	t = 0.4361	t = 0.9172
..._adj	-0.0564	0.2217	-0.0034
	t = -0.2299	t = 0.8565	t = -0.0642
zadj...	0.2237	**0.2670***	0.0558
	t = 10.4936	**t = 1.8868**	t = 10.1677
zadj..._adj	0.0989	**0.4597***	0.0097
	t = 0.5403	**t = 1.8787**	t = 0.1980
...mat	0.1347	**0.1657***	-0.0143
	t = 0.9139	**t = 1.8704**	t = -0.2386
...mat_adj	0.0221	0.1798	-0.0934
	t = 0.1169	t = 0.7739	t = -10.3395
zadj...mat	0.1886	**0.2870*****	-0.0207
	t = 10.3357	**t = 3.0015**	t = -0.3494
zadj...mat_adj	0.0164	0.1864	-0.0820
	t = 0.0936	t = 0.8622	t = -1.2549

Notes: ESG controversies is the number of incidents a firm is subject to controversial news coverage in a given year. ESG are (standardized) Refinitiv ESG performance scores as provided by Refinitiv. The suffix "_adj" indicates whether only fully covered firm-year observations are considered. The prefix "zadj" identifies ESG scores standardized based on yearly firm size deciles. The suffix "mat" identifies materiality-adjusted ESG scores based on the approach developed in chapter 5.5.1. Control variables are: CapEx intensity is capital expenditures divided by total assets. Firm age is the number of years since the firm's IPO. Firm size (in $US million) is the natural logarithm of total assets reported on the balance sheet. IA intensity is the ratio of intangible assets and total assets. Return on assets is net income divided by the book value of total assets. ROA volatility is the standard deviation of ROA over the last three fiscal years. R&D intensity is R&D expenditures divided by total assets. Leverage is the ratio of long- and short-term debt and shareholders' equity. Business sectors, years, and countries are fixed. Values reported in local currencies are converted into USD. Variables are winsorized at the yearly 1%- and 99%-percentiles. All explanatory variables are lagged by one year. Robust test statistics calculated with group-clustered standard errors are reported below coefficients. Significance levels are indicated as follows: * = 0.1, ** = 0.05, *** = 0.01.

Table 30 Regression results: ESG, E, S, and G scores and the number of ESG controversies (full sample)

Almost none of the coefficients indicates a significant relationship between ESG performance and the number of ESG controversies. However, the Social pillar score adjusted for size and materiality is significantly associated with higher numbers of Social controversies, with materiality-adjusted scores showing the smallest effect size. The number of observations is drastically lower in the Environmental and Governance pillar, as the scope of underlying categories in which ESG controversies are assessed is smaller than in the social pillar. Some coefficients depicting the association of Environmental and Governance pillar scores with the number of respective controversies indicate a negative but insignificant association. The negative association especially occurs in the restricted samples of fully covered firms and after adjusting the pillar scores for historical accuracy and materiality patterns.

Discussion

Analyzing the link between ESG metrics and the occurrence of ESG controversies seeks to evaluate the predictive power of Refinitiv's ESG performance scores. Rather than analyzing IV, which is primarily assumed to reflect ESG risks within the market perception of future firm performance (see chapter 5.1), ESG controversies represent the direct external recognition of non-effective or negative conduct on ESG issues. In line with its risk-mitigation properties assumed by investors,[713] this study tests whether ESG metrics actually identify companies with the lowest involvement in ESG controversies as the best performers. As chapter 5.3.2 shows, data on ESG controversies does not affect ESG scoring. Therefore, the results above can serve as an additional reputation-based test of the intended risk mitigation capabilities within investment applications.

In that light, it comes as a surprise that the number of ESG controversies is drastically higher in the highest-performing quintile of the ESG score distribution. Investors applying the raw Refinitiv ESG score could not expect their highly-rated portfolios to be less controversial. It resonates with the already stated severe impact of firm size on the capabilities of the Refinitiv ESG metric to capture a specific company to the full extent. Thus, this finding extends the literature on size biases in ESG metrics, e.g., Drempetic, Klein, and Zwergel (2020), by indicating that firm visibility leads to a higher chance of being captured by Refinitiv's process of assigning ESG controversies based on media coverage. The tilt toward higher-rated firms is slightly lower within the Governance pillar. As Christensen, Serafeim, and Sikochi (2022) state,

[713] cf. Amel-Zadeh/Serafeim (2018, p. 28), cf. Hartzmark/Sussman (2019, pp. 2826 f.)

Governance performance tends to underlie a higher level of standardization and mutual understanding.[714] Hence, the slightly more balanced number of controversies within different Governance score quintiles implies that governance controversies tend to be subject to higher visibility within a more informed information environment.

Nevertheless, additional testing unveils a severe potential of slight adjustments of the raw Refinitiv ESG performance score to allow for drastic enhancements in the intended predictive power of ESG performance metrics on controversies. Berg, Kölbel, and Rigobon (2019) identify the scope and measurement of ESG issues and their weighting within an overall rating as the drivers of divergence among various ESG rating providers.[715] The results of this study add to this notion as reasonable adjustments to the rating already designed to address these aspects imply severe divergence of the outcome within the same rating methodology. Adjusting for firm size and materiality patterns (i.e., the ones identified in chapter 5.5.1) lead to severe changes, especially in the high-rated quintile, which covers the most highly controversial firms. However, adjusting for non-reported data by only considering fully covered firms leads to portfolios following the intended distribution of controversial firms even stronger. Fully covered firms are those firms that disclose a sufficient amount of ESG information, contributing scores within all ESG categories.

Considering non-reported data as the worst possible performance alters the scope of the rating while simultaneously measuring aspects for which there is actually no information at hand. Adjusting the weight of each ESG category according to materiality patterns increases the chance of arriving at a distribution, placing more controversies on the low end and fewer controversies on the high end. Although completely altering the universe of available assets, these simple adjustments scratch the surface of potential scope, measurement, and weighting divergence among different raters. However, it underlines how empirical results might unjustifiably be biased based on the underlying motivations and data processing capabilities of the Refinitiv ESG scoring methodology.

This study adds to Champagne, Coggins, and Sodjahin (2021), who implement logistic regression models to investigate whether KLD metrics are able to indicate the future occurrence of adverse ESG events.[716] The regression models within this study suggest that Refinitiv ESG performance scores are unable to do so. After adjusting for systematic biases imposed by its

[714] cf. Christensen/Serafeim/Sikochi (2022, pp. 159 f.)
[715] cf. Berg/Kölbel/Rigobon (2019, pp. 29 f.)
[716] cf. Champagne/Coggins/Sodjahin (2021, p. 2)

proprietary rating methodology, however, the scores gain stronger capabilities in identifying less controversial firms. Especially when restricted to fully covered firms, the score is able to show the intended negative association of ESG scores and the odds of observing future controversies.

These results imply that Refinitiv's data collection processes capturing ESG controversies might be under-qualified to meet their intended purpose. Note that **Table 28** finds the negative association only within the restricted US sample. This circumstance suggests that processing news data to assign controversies might be over proportionally focusing on firms domiciled in the United States. In global equity indices such as the MSCI World, nearly 70% of companies are headquartered in the United States.[717] Considering global market capitalization being concentrated on US firms to such a high extent, Refinitiv's processing of ESG controversies most likely reflects this US-centric distribution. Further, the focus on US firms might also be motivated by a majority of controversial media coverage originating from English-speaking news sources.

It is worth noticing that Refinitiv offers an "ESG combined score" based on the controversy data investigated in this study. The "Controversy score", solely reflecting the percentile rank according to the yearly number of controversies, is aggregated with the usual ESG performance score and specifically marketed to "… verify companies' actions against commitments to magnify the impact of significant controversies on the overall ESG scoring".[718] Refinitiv accounts for market capitalization by assigning three different weights to the controversies, designed to overweight presumably fewer controversies of small firms and underweight the naturally higher amount of controversies of large firms.[719] However, the results of this study show that there are reasons to doubt the data processing capabilities of Refinitiv being sufficient to capture a viable approximation of firms conducting business controversially.

[717] https://www.msci.com/documents/10199/149ed7bc-316e-4b4c-8ea4-43fcb5bd6523, accessed: 08/15/2022.
[718] Refinitiv (2022a, p. 3)
[719] cf. ibid. (p. 14)

5.6 Summary

The empirical investigation builds on a sample of 83.827 firm-year observations depicting the ESG performance of 11.792 firms from 89 countries between 2002 and 2021. The study covers the entire Refinitiv ESG score universe (as of April 2022) and analyzes all available ESG metrics (i.e., ESG overall score, E, S, and G pillar scores, and 10 ESG category scores).

This study is motivated by the often-proclaimed capability of ESG metrics to identify and mitigate exposure to ESG risks. As those are expected to affect market valuation, the study utilizes idiosyncratic volatility (IV) as an ex-post measure of exposure to events and circumstances posing firm-specific effects on market valuation. Hence, IV serves as a proxy for the ex-post materialization of actual ESG performance. By linking the ex-post materialization with the ex-ante assessment in the ESG score provided by Refinitiv, the predictive accuracy of Refinitiv's proprietary ESG scoring methodology can be evaluated.

Additionally, Refinitiv provides firm-specific data on negative news coverage within a separate database. The empirical investigation builds on a sample of 11,555 firm-year observations representing a total number of 31,963 separate appearances in news outlets due to controversial business conduct. ESG controversies serve as another ex-post test of ESG scores, which is initially expected to identify firms with less controversial business conduct as better ESG performers.

In the first step, a descriptive analysis of the ESG score universe and the ESG controversy database gathers signs of severe biases and inconsistencies affecting the outcome of Refinitiv's ESG scoring methodology. The ESG scores experience a strong *firm size bias*, systematically identifying large firms as the best performers. Simultaneously, large firms experience the highest number of controversies contradicting the initial evaluation of ESG performance scores. The study supports previous findings in research literature that especially the scope of reporting affects the scoring of a firm. As the data points, underlying Refinitiv's ESG scores are drawn from corporate disclosure, firms reporting more transparently and on a larger scale enhance the available information that data points potentially capture. The same holds true for ESG controversies, as the magnitude of media coverage of larger firms is more intense and evidently more prone to be found by Refinitiv's data-capturing processes.

While the size bias in itself might be rather technical in origin, Refinitiv's proprietary scoring methodology enhances the bias to a large extent. By design, the methodology assigns the worst possible performance (i.e., a score of 0) to firms not reporting on all relevant ESG data points.

Trickling up to the higher levels of ESG scores, which are an aggregation of the underlying scores affected by this approach, at least 75.1% of all ESG overall scores are impacted. Given the fact two category scores in the Governance pillar ("Management" and "Shareholder") are calculated with information that is fairly common in regular reporting (i.e., not provided by an additional CSR/ESG reporting), the addition of firms to the database is triggered even for those firms that do not report a sufficient amount of information to feed other ESG category data points. Assigning the worst possible performance to such remaining ESG category scores could, on the one hand, impose an incentive for corporations to enhance the scope of ESG reporting. On the other hand, being transparent in general (i.e., regardless of the extent of engagement in ESG issues) already improves the ESG rating. Refinitiv calculates ESG scores based on percentile rankings, which renders category scores of zero mathematically implausible. Hence, assigning zeros to firms reporting on a smaller scale systematically improves all other firms automatically. Assuming the reporting scope is strongly affected by varying accounting regulations, the ESG scores are distorted based on regional and firm-specific factors that prominently correlate with firm size (e.g., number of employees or revenue). Hence, by tying the rather technical bias induced by the mere availability of ESG data with a penalty for lack of transparency, the magnitude of the observed size bias is substantially amplified (*transparency bias*).

Besides the scoring methodology's biases, the study also identifies data quality issues. Within the database, which is used to calculate ESG scores, inconsistencies are induced by matching yearly ESG scores with fiscal years. As most of the information on the data points is drawn from corporate disclosure, one can expect major changes (i.e., additions and restatements) in ESG scores to happen as soon as the annual (CSR) report becomes available. However, the study finds that ESG scores based on corporate disclosure referring to fiscal years that end prior to the actual end date of the year in the database also reflect the ESG performance of the previous year. Due to the relative scoring procedure drawing from yearly benchmark groups, there is a chance that ESG scores reflect ESG performance from varying periods of time. The empirical investigation unveiled examples that indicate that especially numeric data points (e.g., energy consumption) could be highly sensitive to the period they are capturing, ultimately (yet rather unsystematically) affecting the outcome of the ESG scores.

Refinitiv's ESG controversy data captures a wide range of controversy categories. The number of controversies captured in some categories (e.g., "business ethics") vastly exceeds those in other categories. Due to the thematic scope of some single controversy categories, some likely

serve as broader umbrella categories. Additional investigation of the underlying articles being captured as controversial news coverage unveils that some controversies are based on similar incidences that differ, for example, with regard to the impacted region or interest groups involved. It is debatable whether this design choice might also facilitate the double-counting of controversial business conduct. Again, the ESG controversy database can be expected to suffer from technical and transparency-induced inaccuracies and inconsistencies.

The connotation of controversies is initially expected to be negative. Controversies account for negative business conduct. One example of this notion being fairly debatable is the data point capturing the occurrences of strikes. The small frequency of strikes in the dataset is reasonable, given their uncommon nature. Nevertheless, the definition and scope of strikes being captured are unclear. Several firms and industries are subject to periodical work stoppages when, for instance, wages are up for debate. The data, however, does not specifically differentiate strikes that might rather point toward actual structural problems of potential misconduct.

Further, the empirical investigation finds indications of single controversies being disputable in their interpretation. For example, within the Facebook Controversy data point on "Business Ethics" in 2019, a fine for withholding Whatsapp user data being part of a drug-trafficking investigation in Brazil is captured.[720] The negative interpretation is valid as long as the controversial business conduct is motivated by the belief that the firm serves a societal purpose in opposing organized crime. However, if the rationale comes from a data privacy standpoint, withholding user data, regardless of origin, is a rather positive take on handling user data.

The results of this study show that there are reasons to highly doubt the data processing capabilities of Refinitiv being sufficient to capture a viable approximation of firms conducting business controversially. The number of controversies being captured by Refinitiv depends on: 1) Firm visibility heavily associated with firm size, 2) general standards and perceptions of controversial business conduct inducing a more informed and attentive information environment, and 3) data processing capabilities that tend to be more sensitive to news coverage in English-language and primarily focusing US markets.

In the second step, the study builds on an extensive inferential analysis, which tries to connect input-based ESG performance measurement with proxies for the materialization of actual ESG performance in the future. Analyzing the relationships between ex-ante ESG performance and ex-post IV allows testing whether the ESG metric reflects information affecting IV otherwise

[720] https://www.reuters.com/article/facebook-fine-brazil-idUKL2N23W1GE, accessed: 04/18/2022.

not explained by common firm-specific control variables. The analysis of the odds of experiencing future ESG controversies further allows to test how accurately the metric reflects actual ESG performance by indicating which firms are less likely to experience controversies.

In doing so, this study finds ESG metrics show a negative association with future IV. Hence, one can expect the scores to reflect value-relevant information that is highly correlated with the actual, unobserved ESG performance. While in the United States, this relationship is stronger in the S and G pillars, the European sample shows the strongest negative relationship in the E and S pillars. The results indicate that the E pillar score in the US sample does not provide additional information relevant to a firm's future growth opportunities or business uncertainty.

Further analyzing regional and sector-based sub-samples unveils that there are substantial differences in the ability of Refinitiv's ESG scores to indicate IV in the upcoming year. Some category scores are associated with higher levels of future IV. As the volatility of the idiosyncratic component of a stock price, IV indicates the fraction of private information reflected within the stock price. If the level of IV is high, this results from a previous incentive to gather private information, which was not yet reflected in stock prices but would pay out in positive returns. Hence, if the rating accurately provides information relevant to stock prices, the cost-benefit relation of gathering additional private information is negative.

Reassuring the empirical approach of this study, a positive relationship between IV and ESG metrics is predominately found in the Governance pillar. In confirming former empirical studies, the results suggest that governance-related issues tend to be defined more precisely.[721] The rating, therefore, rather reflects Refinitiv's perception of the notion of social expectations on good Corporate Governance. Such perceptions might ultimately convey information already reflected in stock prices. The negative association, however, implies that those firms being rated better are the ones for which there is no motivation to gather additional private information, as the ESG score is already reflecting relevant information accurately enough.

Further, analyzing a smaller set of sub-samples, the results indicate that the informational content, reflecting the accuracy of the ESG performance assessment, differs substantially over time. In addition to occasionally conveying non-relevant information, ESG scores tend to reflect a broad range of ESG issues that might not all be relevant to a specific firm at a certain point in time. Such value-relevant information is referred to as material in accounting. This study finds

[721] cf. Christensen/Serafeim/Sikochi (2022, pp. 159-160)

evidence for varying materiality patterns (*dynamic materiality*) on the level of economic sectors, regions, and throughout the observed period of time. By conducting 1,300 fixed effect regression models on future IV, the accuracy of each ESG category score in each of ten economic sectors over 13 rolling window sub-samples is assessed. Based on the significance and polarity of the coefficient indicating the relationship between a single ESG controversy score and the dependent variable IV, the approach finds that 35.8% of the models indicate a negative association. Consequently, in over a third of all cases, Refinitiv's ESG category scores provided relevant information material to a specific economic sector within a certain period. In almost two-thirds of the models, however, the ESG category scores provide non-significant informational content (44.4%) or even indicate the opposite of what the market would expect as material (i.e., a positive link between ESG category scores and IV).

In the third step, the empirical investigation tries to address the aforementioned issues by introducing several modifications to the original ESG scores. By standardizing ESG metrics on yearly firm size deciles, firms are being compared that face similar information environments alleviating the effect of the initial firm size bias. Additionally, the study introduces a data-driven approach to adjust for the historical accuracy of each ESG category score induced by dynamic materiality patterns. Based on 1,300 rolling window fixed effect panel regressions described above, each category score is assigned a specific weight within its pillar and the overall ESG score. Further, each of the size- and materiality-adjusted ESG scores are separately investigated based on a sample limited to firms that are fully covered in all ESG categories. This adjustment accounts for the effect of the assignment of the worst possible performance to non-reported information. The adjustment heavily decreases the number of firm-year observations. However, it allows evaluation of how well the ESG scoring methodology is able to accurately reflect ESG performance when all data points are fully covered.

The fairly simple adjustments lead to significant positive shifts in the ESG scores' ability to predict the occurrence of ESG controversies. Portfolios being constructed combining size and materiality adjustments identify significantly less controversial firms as best performers. The number of controversies in the lowest-performing quintile is enhanced by 118%, while the number of controversies in the highest-performing quintile is reduced by 57%. Limiting the sample to fully covered firms and thereby potentially alleviating size and transparency biases shows that the improvement is robust. Constructing portfolios from size- and materiality-adjusted scores in the fully covered sample improves the detection of controversies in the lowest-performing quintile by 236%.

Logistic regression results confirm these findings, indicating that size and methodology-induced transparency biases severely impact the predictive accuracy of Refinitiv's ESG metrics. ESG metrics are initially associated positively with the number of controversies in the future. Iteratively adjusting for firm size, materiality, and ultimately limiting the sample to fully covered firms only, progressively reduces the positive relationship. The positive relationship within the European sample is higher than in the US sample, implying that the database predominately captures news coverage in English and a more concise and extensive picture of actual controversial conduct in the US sample of firms.

Based on a detailed empirical investigation, the following indications regarding the hypotheses of chapter 5.1 can be drawn:

Hypothesis	**Result**
H1.1 *ESG performance metrics are positively related to firm size.*	confirmed
H1.2 *Data on the occurrence of ESG controversies is positively related to firm size.*	confirmed
H2.1 *ESG performance metrics are negatively biased by the incorporation of non-reported ESG information.*	confirmed
H2.2 *The fraction of Boolean data points that inform ESG metrics enhances the bias induced by non-reported ESG information.*	confirmed
H3.1 *ESG performance metrics are inversely related to idiosyncratic volatility.*	partly confirmed
H3.2 *Regional peculiarities do not induce variation in the relation between ESG performance metrics and idiosyncratic volatility.*	rejected
H3.3 *Industry affiliation does not induce variation in the relation between ESG performance metrics and idiosyncratic volatility.*	rejected
H4 *Shifts in materiality over time significantly affect the explanatory power of ESG performance metrics.*	confirmed
H5.1 *The probability of ESG controversies and strikes is negatively related to ESG metrics.*	rejected
H5.2 *ESG controversies are skewed toward the low end of the ESG metric distribution.*	rejected
H6.1 *Adjusting ESG metrics for biases induced by firm-size, non-reported ESG information and weighted according to an outcome-oriented and dynamic materiality profile helps to identify lower levels of idiosyncratic volatility.*	confirmed
H6.2 *Adjusting ESG metrics for biases induced by firm-size, non-reported ESG information and weighted according to an outcome-oriented and dynamic materiality profile helps to identify portfolios with lower exposure to ESG controversies.*	confirmed

Table 31 Conclusion: Hypotheses and results

6 CONCLUSION AND IMPLICATIONS

ESG data providers and their metrics serve an ever-growing important role in informing capital markets on the non-financial performance of single corporate entities, investment funds, and equity indices. They build upon the idea of quantifying the potential costs and benefits of a firm's engagement in the three behavioral dimensions: Environment (E), Social (S), and Corporate Governance (G). Investors expect ESG information to convey insights on a firm's exposure to relevant ESG issues, which in turn allows for the mitigation of related risks or the alignment with personal values and beliefs.[722] Proprietary data collection and processing, however, leads to severe divergences among different ESG data providers and their assessment of ESG performance.[723]

Despite ongoing debates over the motivations, societal benefits, and economic merits, especially in the realm of a politically advocated sustainable transformation of the global economy, ESG metrics are increasingly tackled for their accuracy in depicting actual ESG performance and accompanying risks.

The underlying dissertation deals with the theoretical and practical burdens of measuring ESG performance. Introducing the Social Responsibility of the firm (CSR) as a mixture of economic, societal, and political expectations helps to understand the diverse drivers that potentially affect a firm's position on ESG issues. As ESG metrics predominately serve as information sources for capital markets, the study describes the motivations, strategies, and prevalence of investment practices that rely on the implications of ESG performance metrics. Turning toward the practical challenges of ESG performance measurement, the study introduces the concept of materiality as the basis for weighting schemes that prioritize single ESG factors motivated by their firm-specific relevance to the business model. As ESG metrics predominately rely on corporate disclosure, the study extensively shows how ESG performance metrics almost compulsively fall short of observing the impact of ESG performance and remain on an input-level depiction of commitment to ESG objectives. An extensive literature review explains how contemporary research literature picks up on the divergence of ESG metrics that consequentially results from the above-stated peculiarities in ESG performance measurement.

[722] cf. Amel-Zadeh/Serafeim (2018, p. 87)
[723] cf. Berg/Kölbel/Rigobon (2019, pp. 29-30), Refinitiv ESG performance scores are based on the ASSET4 database.

Contrary to recent advances to find the lowest common denominator within the wide range of available ESG data sources,[724] this study empirically turns toward one ESG metric in particular: The "Refinitiv ESG company scores". Refinitiv's proprietary process of calculating ESG metrics relies on corporate disclosure and publicly-available data sources.[725] Hence, a thorough empirical investigation of the underlying data and metrics calculated thereof allows to shed light on the accuracy and predictive qualities of Refinitiv's input-oriented methodology to assess ESG performance.

The results of the empirical investigation of this dissertation cast great doubt on the ability of raw Refinitiv ESG performance scores to provide accurate information that renders beneficial in mitigating ESG-related risks. This conclusion is drawn from three fields of concern: 1) The nature of publicly available ESG information, 2) Refinitiv's proprietary scoring methodology and data processing, and 3) the ability of Refinitiv's ESG metrics to reflect information that indicates the materialization of firm-specific risks induced by ESG performance.

Based on the entire ESG rating universe as of April 2022,[726] the study identifies severe biases indicated by larger firms being systematically ranked higher throughout all of Refinitiv's ESG scores. Former literature suggests that larger firms are able to allocate more resources to informing the public and enforcing the *visibility* of their ESG activities. Further, reporting provisions systematically differ regarding the *scope* of information that is mandatory or common to communicate through corporate disclosure. Both visibility and scope as characteristics of the mere nature of publicly available ESG information are decisive on the extent of firm-specific ESG information available to base a scoring methodology on.

The study finds that Refinitiv's proprietary ESG scoring methodology is not able to account for the nature of publicly available ESG information. Even worse, in line with previous research findings on the accuracy of ESG metrics, the empirical investigation suggests Refinitiv's scoring methodology severely aggravates the firm size bias by design. The fraction of firm-year observations with a value of 0, indicating the worst possible performance, ranges up to 59.2% of firms being granted a score in one of Refinitiv's ESG categories. As the calculation of category scores is based on percentile rankings, values of 0 are mathematically implausible. Hence,

[724] See for example: Berg, Kölbel, and Rigobon (2019)
[725] cf. Refinitiv (2022a, p. 3)
[726] i.e., 83.827 firm-year observations, 11.792 firms, 89 countries, ESG overall, E, S, and G pillar, and ten ESG category scores between 2002 and 2021. Refinitiv's ESG controversy score is not a part of the empirical investigation.

the values are intentionally set by Refinitiv. As the Governance related category scores "Management" and "Shareholders" designate scores to all firms in the dataset, the results suggest Refinitiv follows a systematic take on processing missing information within its data points. Due to fairly common information on Management and Shareholders being provided via regular financial reporting, firms are added to the Refinitiv ESG rating universe. As no other information is provided on data points necessary to calculate other ESG category scores, firms are assigned a value of 0 on the remaining category scores. From a data vendor standpoint, this practice is pretty plausible. It incentivizes reporting on ESG issues while simultaneously amplifying the ESG rating universe to a larger number of firms. However, due to the relative ranking among firms in the dataset, each firm, regardless of the nature of its ESG conduct, gains an advantage over other firms simply by being transparent in any way. This type of data processing simultaneously entrenches the already described *size bias* and the additionally induced *transparency bias*.

To evaluate the predictive accuracy of Refinitiv's ESG performance metrics, this study links ex-ante and input-oriented Refinitiv ESG metrics with the actual, yet unobservable level of ex-post ESG performance. This study applies market-based idiosyncratic volatility (IV) and the occurrence of ESG controversies (i.e., drawn from controversial news coverage) as proxies for the materialization of firm-specific ESG risks induced by the genuine, yet unobservable level of commitment to ESG objectives. The results unveil no systematic and persisting link between the raw Refinitiv ESG metric and the aforementioned proxies for actual ESG performance.

Given the size and transparency biases in Refinitiv's scoring methodology and the theoretically described burdens in ESG performance measurement, the study introduces a simple, data-driven approach to account for common and methodology-specific inaccuracies. Standardizing ESG metrics on yearly firm size deciles allows to seize on the relative nature of Refinitiv's ESG performance metrics but instead benchmarks firms within more comparable information environments. By conducting 1,300 separate rolling window regressions models linking single ESG category scores with the dependent variable IV, a specific indication of whether the score captures value-relevant information on a firm's specific risk accounting for industry (i.e., ten economic sectors) and regional (i.e., US and European firms) associations is assessed. The industry- and region-specific indication whether scores carry value-relevant (i.e., material) information for a particular time period allows reweighting the category scores based on their historical accuracy.

Reweighting ESG overall, E, S, and G pillar scores based on their *historical accuracy* and introducing the *size adjustment* described above allows to severely enhance the ability to predict the occurrence of future ESG controversies. Limiting the investigation to the sample of firms, which are fully covered in all relevant data points (i.e., 24.9% of the original dataset), further enhances the accuracy of the ESG score's indication.

Nevertheless, the results are highly distorted by data quality issues in Refinitiv's controversy data. The results of this study suggest the number of controversies being captured depends on: 1) Firm visibility heavily associated with firm size, 2) general standards and perceptions of uncontroversial business conduct inducing a more informed and attentive information environment, and 3) data processing capabilities that tend to be more sensitive to news coverage in English-language and primarily focusing US markets. The results cast great doubt on Refinitiv's additionally marketed ESG controversies scores. As a blend of a firm's ESG score and an indication of the involvement in current controversial business conduct, the score most likely fails to meet its supposed purpose accurately.

In conclusion, the results of this dissertation implicate the analytical use of ESG performance metrics, reflective of the individual purpose (i.e., risk mitigation, value alignment, or the intention of impact). The ambiguous nature of ESG performance and the interest of metrics to boil down such non-financial corporate performance to a simple and easy-to-interpret indication of a firm's commitment to a multitude of issues allows for a similar extent of design choices. Constructing such metrics inherently requires compromises and trade-offs on a qualitative and interpretative level. Simultaneously, the mere nature of ESG information and its acquisition imposes structural burdens that affect the quantitative database every ESG metric imperatively is based on.

In light of common critiques that characterize ESG engagement as a strategic tool that makes use of the trending awareness for global environmental and societal challenges, incentives to exploit structural weaknesses of ESG performance measurement are high. The results of this dissertation cannot reassure users of ESG performance metrics that such practices (e.g., greenwashing) are effectively exposed. Hence, expecting the same rigor and consensus as observed among other common third-party firm-specific metrics, such as credit ratings, might be delusional at the very end. Demanding for mutual and uniform ESG scoring standards or frameworks, however, might be as equally naïve as fully trusting an indication of a single proprietary scoring methodology.

ESG metrics remain an important source of information for capital markets. However, their indication should rather be understood, for example, in the sense of investment recommendations. Based on proprietary calculation and individual reasoning, such information is also quite known to diverge among different analysts. Contrary to most of those, ESG score providers (e.g., Refinitiv) that allow for a granular investigation of their database and aggregation methodology can help to bridge information asymmetries between the corporate world and investor demands.

REFERENCES

Abdi, H. (2007). Z-scores. In: *Encyclopedia of measurement and statistics*, *3*, 1055-1058.

Adams, C. A., Alhamood, A., He, X., Tian, J., Wang, L., Wang, Y. (2021). The double-materiality concept: application and issues. Project Report. Global Reporting Initiative, Amsterdam.

Aguinis, H., Glavas, A. (2012). What we know and don't know about corporate social responsibility: A review and research agenda. In: *Journal of Management*, *38*(4), 932-968.

Alin, A. (2010). Multicollinearity. In: *Wiley interdisciplinary reviews: computational statistics*, *2*(3), 370-374.

Amel-Zadeh, A., Serafeim, G. (2018). Why and how investors use ESG information: Evidence from a global survey. In: *Financial Analysts Journal*, *74*(3), 87-103.

Anderson, L. M., Petticrew, M., Rehfuess, E., Armstrong, R., Ueffing, E., Baker, P., Francis, D., Tugwell, P. (2011). Using logic models to capture complexity in systematic reviews. In: *Research synthesis methods*, *2*(1), 33-42.

Ang, A., Hodrick, R. J., Xing, Y., Zhang, X. (2006). The cross-section of volatility and expected returns. In: *The Journal of Finance*, *61*(1), 259-299.

Aouadi, A., Marsat, S. (2018). Do ESG controversies matter for firm value? Evidence from international data. In: *Journal of Business Ethics*, *151*(4), 1027-1047.

Artiach, T., Lee, D., Nelson, D., Walker, J. (2010). The determinants of corporate sustainability performance. In: *Accounting & Finance*, *50*(1), 31-51.

Awaysheh, A., Heron, R. A., Perry, T., Wilson, J. I. (2020). On the relation between corporate social responsibility and financial performance. In: *Strategic Management Journal*, *41*(6), 965-987.

Banz, R. W. (1981). The relationship between return and market value of common stocks. In: *Journal of financial economics*, *9*(1), 3-18.

Barnett, M. L., Henriques, I., Husted, B. W. (2020). Beyond good intentions: Designing CSR initiatives for greater social impact. In: *Journal of Management*, *46*(6), 937-964.

Bassen, A., Jastram, S., Meyer, K. (2005). Corporate social responsibility: eine begriffserläuterung. In: *Zeitschrift für Wirtschafts- und Unternehmensethik*, *6*(2), 231-236.

Basu, S. (1983). The relationship between earnings' yield, market value and return for NYSE common stocks: Further evidence. In: *Journal of Financial Economics*, *12*(1), 129-156.

Bear, S., Rahman, N., Post, C. (2010). The impact of board diversity and gender composition on corporate social responsibility and firm reputation. In: *Journal of Business Ethics*, *97*(2), 207-221.

Bénabou, R., Tirole, J. (2010). Individual and corporate social responsibility. In: *Economica*, *77*(305), 1-19.

Berg, F., Fabisik, K., Sautner, Z. (2020). Rewriting history II: The (un) predictable past of ESG ratings. In: *European Corporate Governance Institute – Finance Working Paper*, *708*(2020), 10-2139.

Berg, F., Kölbel, J. F., Rigobon, R. (2019). Aggregate confusion: The divergence of ESG ratings. In: *Review of Finance (forthcoming)*.

Berg, F., Kölbel, J. F., Pavlova, A., Rigobon, R. (2021). ESG confusion and stock returns: Tackling the problem of noise. *Available at SSRN 3941514*.

Berle, A. A. (1932). For whom corporate managers are trustees: a note. In: *Harvard Law Review*, *45*(8), 1365-1372.

Bethlehem, J. (2010). Selection bias in web surveys. In: *International Statistical Review*, *78*(2), 161-188.

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

Blitz, D., Falkenstein, E., Van Vliet, P. (2014). Explanations for the volatility effect: An overview based on the CAPM assumptions. In: *The Journal of Portfolio Management*, *40*(3), 61-76.

Boucher, C., Le Lann, W., Matton, S., Tokpavi, S. (2022). Backtesting ESG Ratings (DR LEO 2021-09). In: *Working Paper Series, Economic Research Department of the University of Orleans (LEO)*.

Brown, D. L., Vetterlein, A., Roemer-Mahler, A. (2010). Theorizing transnational corporations as social actors: An analysis of corporate motivations. In: *Business and Politics*, *12*(1), 1-37.

Brown, S. J., Goetzmann, W., Ibbotson, R. G., Ross, S. A. (1992). Survivorship bias in performance studies. In: *The Review of Financial Studies*, *5*(4), 553-580.

Brundtland, G. H. (1987). Report of the World Commission on environment and development: "Our Common Future.". UN.

Buhmann, K. (2016). Public regulators and CSR: the 'social license to operate' in recent united nations instruments on business and human rights and the juridification of CSR. In: *Journal of Business Ethics, 136*(4), 699-714.

Buhmann, K., Jonsson, J., Fisker, M. (2018). Do no harm and do more good too: Connecting the SDGs with business and human rights and political CSR theory. In: *Corporate Governance, 19*(3), 389-403.

Campbell, J. Y., Lettau, M., Malkiel, B. G., Xu, Y. (2001). Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. In: *The Journal of Finance, 56*(1), 1-43.

Cao, C., Simin, T., Zhao, J. (2008). Can growth options explain the trend in idiosyncratic risk?. In: *The Review of Financial Studies*, *21*(6), 2599-2633.

Carhart, M. M. (1997). On persistence in mutual fund performance. In: *The Journal of Finance*, *52*(1), 57-82.

Carroll, A. B. (1979). A three-dimensional conceptual model of corporate performance. In: *Academy of Management Review*, *4*(4), 497-505.

Carroll, A. B. (1991). The pyramid of corporate social responsibility: Toward the moral management of organizational stakeholders. In: *Business Horizons*, *34*(4), 39-48.

Carroll, A. B. (1999). Corporate social responsibility: Evolution of a definitional construct. In: *Business & Society*, *38*(3), 268-295.

Carroll, A. B. (2015). Corporate social responsibility (CSR) is on a sustainable trajectory. In: *Journal of Defense Management*, *5*(2), 1-2.

Carroll, A. B., Hoy, F. (1984). Integrating corporate social policy into strategic management. In: *Journal of Business Strategy, 4* (3), 48-57.

CDP (2022). Scoring Introduction 2022. An introduction to 2022 scoring. Carbon Disclosure Project, London.

Chaffee, E. C. (2017). The origins of corporate social responsibility. In: *University of Cincinnati Law Review, 85*, 347–373.

Champagne, C., Coggins, F., Sodjahin, A. (2021). Can extra-financial ratings serve as an indicator of ESG risk?. In: *Global Finance Journal*, 100638.

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

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

Chemmanur, T. J., Kong, L., Krishnan, K., Yu, Q. (2019). Top management human capital, inventor mobility, and corporate innovation. In: *Journal of Financial and Quantitative Analysis*, *54*(6), 2383-2422.

Chen, M., von Behren, R., Mussalli, G. (2021). The Unreasonable Attractiveness of More ESG Data. In: *The Journal of Portfolio Management*, *48*(1), 147-162.

Cheng, B., Ioannou, I., Serafeim, G. (2014). Corporate social responsibility and access to finance. In: *Strategic Management Journal*, *35*(1), 1-23.

Chin, M. K., Hambrick, D. C., Treviño, L. K. (2013). Political ideologies of CEOs: The influence of executives' values on corporate social responsibility. In: *Administrative Science Quarterly*, *58*(2), 197-232.

Chiu, I. H. (2022). The EU Sustainable Finance Agenda: Developing Governance for Double Materiality in Sustainability Metrics. In: *European Business Organization Law Review*, *23*(1), 87-123.

Choi, J. J., Jo, H., Kim, J., Kim, M. S. (2018). Business groups and corporate social responsibility. In: *Journal of Business Ethics*, *153*(4), 931-954.

Christensen, H. B., Hail, L., Leuz, C. (2021). Mandatory CSR and sustainability reporting: economic analysis and literature review. In: *Review of Accounting Studies*, *26*(3), 1176-1248.

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

Clark, C., Rosenzweig, W., Long, D., Olsen, S. (2004). Double Bottom Line Project Report: Assessing Social Impact in Double Line Ventures. New York.

Clark, C., Rosenzweig, W., Long, D., Olsen, S. (2004). Double bottom line project: Assessing social impact in double bottom line ventures. In: *Working Paper Series*. Berkeley, CA: Center for Responsible Business, University of California Berkeley.

Coleton, A., Font Brucart, M., Gutierrez, P., Le Tennier, F., Moor, C. (2020). Sustainable Finance: Market Practices. In: *European Banking Authority Research Paper*, (6).

Committee for Economic Development. (1971). Social responsibilities of business corporations. New York: Committee for Economic Development.

Crace, L., Gehman, J. (2022). What Really Explains ESG Performance? Disentangling the Asymmetrical Drivers of the Triple Bottom Line. In: *Organization & Environment (forthcoming)*.

Davis, K. (1960). Can business afford to ignore social responsibilities?. In: *California Management Review, 2*(3), 70-76.

De Villiers, C., Jia, J., Li, Z. (2022). Corporate social responsibility: A review of empirical research using Thomson Reuters Asset4 data. In: *Accounting & Finance (forthcoming)*.

DeAngelo, H., Masulis, R. W. (1980). Optimal capital structure under corporate and personal taxation. In: *Journal of Financial Economics*, *8*(1), 3-29.

Demarzo, P. M. (2019). Presidential address: Collateral and commitment. In: *The Journal of Finance, 74*(4), 1587-1619.

Denis, D. K. (2001). Twenty-five years of corporate governance research… and counting. In: *Review of Financial Economics*, *10*(3), 191-212.

Deutsche Bank (2021). Annual report 2020. https://investor-relations.db.com/reports-and-events/annual-reports/, accessed: 02/16/2022.

Donaldson, T., Preston, L. E. (1995). The stakeholder theory of the corporation: Concepts, evidence, and implications. In: *Academy of Management Review, 20*(1), 65-91.

Drempetic, S., Klein, C., Zwergel, B. (2020). The influence of firm size on the ESG score: Corporate sustainability ratings under review. In: *Journal of Business Ethics, 167*(2), 333-360.

Dryzek, J. S. (2021). The politics of the earth. Oxford: Oxford university press.

Ebrahim, A., Rangan, V. K. (2014). What impact? A framework for measuring the scale and scope of social performance. In: *California Management Review, 56*(3), 118-141.

Eccles, R. G., Ioannou, I., Serafeim, G. (2014). The impact of corporate sustainability on organizational processes and performance. In: *Management Science, 60*(11), 2835-2857.

Eccles, R. G., Krzus, M. P., Rogers, J., Serafeim, G. (2012). The need for sector-specific materiality and sustainability reporting standards. In: *Journal of Applied Corporate Finance, 24*(2), 65-71.

Eccles, R. G., Lee, L. E., Stroehle, J. C. (2020). The social origins of ESG: An analysis of Innovest and KLD. In: *Organization & Environment, 33*(4), 575-596.

Eccles, R., Crowley, D. F. C. (2022). Turning down the heat on the ESG debate: separating material risk disclosures form salient political issues. Harvard Law School Forum on

Corporate Governance (September 01, 2022). https://corpgov.law.harvard.edu/2022/09/01/turning-down-the-heat-on-the-esg-debate-separating-material-risk-disclosures-from-salient-political-issues/, accessed: 09/05/2022.

Eckert, E., Kovalevska, O. (2021). Sustainability in the European Union: analyzing the discourse of the European green deal. In: *Journal of Risk and Financial Management*, *14*(2), 80.

El Ghoul, S., Karoui, A. (2017). Does corporate social responsibility affect mutual fund performance and flows?. In: *Journal of Banking & Finance*, *77*, 53-63.

Escrig-Olmedo, E., Fernández-Izquierdo, M. Á., Ferrero-Ferrero, I., Rivera-Lirio, J. M., Muñoz-Torres, M. J. (2019). Rating the raters: Evaluating how ESG rating agencies integrate sustainability principles. In: *Sustainability*, *11*(3), 915.

European Banking Authority (2021). EBA report on management and supervision of ESG risks for credit institutions and investment firms. EBA/REP/2021/18. European Banking Authority, Paris.

European Commission (2001). Green paper: promoting a European framework for corporate social responsibility. Commission of the European Communities. European Commission, Brussels.

European Commission (2006). Launch of "European Alliance for Corporate Social Responsibility". European Commission, Brussels.

European Commission (September 16, 2009). European union hails universal ratification of the Montreal protocol on protecting the ozone layer. IP/09/1328. European Commission, Brussels.

European Commission (2011). A renewed EU strategy 2011-14 for Corporate Social Responsibility. European Commission, Brussels.

European Commission (October 28, 2016). European Commission establishes an expert group to develop a comprehensive European strategy on sustainable finance. https://ec.europa.eu/info/sites/default/files/161028-press-release_en.pdf, accessed: 09/09/2020.

European Commission (2018). Financing a sustainable European economy: final report 2018. High-Level Expert Group on Sustainable Finance, European Commission, Brussels.

European Commission (2019). Guidelines on reporting climate-related information. European Commission, Brussels.

European Commission (2020). Technical Report - Taxonomy: Final report of the Technical Expert Group on Sustainable Finance. European Commission, Brussels.

Fama, E. F., French, K. R. (1992). The cross-section of expected stock returns. In: *The Journal of Finance*, *47*(2), 427-465.

Fama, E. F., French, K. R. (1993). Common risk factors in the returns on stocks and bonds. In: *Journal of Financial Economics*, *33*(1), 3-56.

Fama, E. F., French, K. R. (1996). Multifactor explanations of asset pricing anomalies. In: *The Journal of Finance*, *51*(1), 55-84.

Fama, E. F., French, K. R. (2004). The capital asset pricing model: Theory and evidence. In: *Journal of Economic Perspectives*, *18*(3), 25-46.

Fama, E. F., Jensen, M. C. (1983). Agency problems and residual claims. In: *The Journal of Law and Economics*, *26*(2), 327-349.

FASB (2018). Conceptual framework for financial reporting. Chapter 1, The objective of general purpose financial reporting, and chapter 3, qualitative characteristics of useful financial information. August 2018. Financial Accounting Standards Board, Norwalk.

Feng, G., Giglio, S., Xiu, D. (2020). Taming the factor zoo: A test of new factors. In: *The Journal of Finance*, *75*(3), 1327-1370.

Ferreira, M. A., Laux, P. A. (2007). Corporate governance, idiosyncratic risk, and information flow. In: *The Journal of Finance*, *62*(2), 951-989.

Fink, J., Grullon, G., Fink, K., Weston, J. (2004). Firm age and fluctuations in idiosyncratic risk. Available at SSRN 891173.

Fink, L. (2020). “Letter to CEOs”. https://www.blackrock.com/corporate/investor-relations/2020-larry-fink-ceo-letter, accessed: 09/09/2020

Freeman, R. E., Reed, D. L. (1983). Stockholders and stakeholders: A new perspective on corporate governance. In: *California Management Review*, *25*(3), 88-106.

Friede, G., Busch, T., Bassen, A. (2015). ESG and financial performance: aggregated evidence from more than 2000 empirical studies. In: Journal of sustainable finance & investment, 5(4), 210-233.

Friedman, M. (1970). A Friedman doctrine: The social responsibility of business is to increase its profits. In: *The New York Times Magazine*, *13*(1970), 32-33.

Fu, F. (2009). Idiosyncratic risk and the cross-section of expected stock returns. In: *Journal of Financial Economics*, *91*(1), 24-37.

Ghosh, D., Vogt, A. (2012). Outliers: An evaluation of methodologies. In: *Joint Statistical Meetings*, 3455–3460.

Gibson Brandon, R., Krueger, P., Schmidt, P. S. (2021). ESG rating disagreement and stock returns. In: *Swiss Finance Institute Research Paper Series, 19-67 (July 2021), Financial Analysts Journal (forthcoming)*.

Gillan, S. L., Koch, A., Starks, L. T. (2021). Firms and social responsibility: A review of ESG and CSR research in corporate finance. In: *Journal of Corporate Finance*, *66*, 101889.

Godfrey, P. C. (2005). The relationship between corporate philanthropy and shareholder wealth: A risk management perspective. In: *Academy of Management Review*, *30*(4), 777-798.

Godfrey, P. C., Merrill, C. B., Hansen, J. M. (2009). The relationship between corporate social responsibility and shareholder value: An empirical test of the risk management hypothesis. In: *Strategic Management Journal*, *30*(4), 425-445.

Goergen, M. (2018). Corporate governance: a global perspective. 1st edition, Cengage Learning.

Gold, S., Heikkurinen, P. (2018). Transparency fallacy: Unintended consequences of stakeholder claims on responsibility in supply chains. In: *Accounting, Auditing & Accountability Journal*, *31*(1), 318-337.

Gompers, P., Ishii, J., Metrick, A. (2003). Corporate governance and equity prices. In: *The quarterly Journal of Economics, 118*(1), 107-156.

Gonzales-Perez, M. A., Leonard, L. J. (2017). The UN Global Compact. In: A. de Jonge, R. Tomasic, Research handbook on transnational companies, Cheltenham: Edward Elgar Publisher, 117-138.

Gonzalez, M., Taddonio, K. N., Sherman, N. J. (2015). The Montreal Protocol: how today's successes offer a pathway to the future. In: *Journal of Environmental Studies and Sciences*, *5*(2), 122-129.

Gormley, T. A., Matsa, D. A. (2014). Common errors: How to (and not to) control for unobserved heterogeneity. In: *The Review of Financial Studies*, *27*(2), 617-661.

Grewal, J., Serafeim, G. (2020). Research on corporate sustainability: Review and directions for future research. In: *Foundations and Trends in Accounting*, *14*(2), 73-127.

GRI (2022). Consolidated set of the GRI standards. Global Reporting Initiative, Amsterdam.

GSIA (2021). Global sustainable investment review 2020. Global Sustainable Investment Alliance, Brussels.

Gyönyörová, L., Stachoň, M., Stašek, D. (2021). ESG ratings: relevant information or misleading clue? Evidence from the S&P Global 1200. In: *Journal of Sustainable Finance & Investment*, 1-35.

Handl, G. (2012). Declaration of the United Nations conference on the human environment (Stockholm Declaration), 1972 and the Rio Declaration on Environment and Development, 1992. In: *United Nations Audiovisual Library of International Law, 11*(6).

Harjoto, M. A., Jo, H. (2011). Corporate governance and CSR nexus. In: *Journal of Business Ethics, 100*(1), 45-67.

Hart, O., Zingales, L. (2017). Companies Should Maximize Shareholder Welfare Not Market Value. In: *Journal of Law, Finance, and Accounting, 2*(2), 247-275.

Hartzmark, S. M., Sussman, A. B. (2019). Do investors value sustainability? A natural experiment examining ranking and fund flows. In: *The Journal of Finance, 74*(6), 2789-2837.

Hayat, U., Orsagh, M. (2015). Environmental, social, and governance issues in investing: A guide for investment professionals. Charlottesville, VA: CFA Institute.

Healy, P. M., Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. In: *Journal of Accounting and Economics, 31*(1-3), 405-440.

Heckman, J. J. (1979). Sample selection bias as a specification error. In: *Econometrica*, 47(1), 153-161.

Heeb, F., Kölbel, J. F., Paetzold, F., Zeisberger, S. (2022). Do investors care about impact?. In: *The Review of Financial Studies (forthcoming)*.

Hosmer Jr, D. W., Lemeshow, S. (2000). Applied logistic regression (second edition). Hoboken, NY: John Wiley & Sons.

Howard-Grenville, J. (2021). ESG impact is hard to measure - but it's not impossible. In: *Harvard Business Review*.

Hughey, C. J., Sulkowski, A. J. (2012). More disclosure= better CSR reputation? An examination of CSR reputation leaders and laggards in the global oil & gas industry. In: *Journal of Academy of Business and Economics, 12*(2), 24-34.

Hulme, D. (2009). The Millennium Development Goals (MDGs): a short history of the world's biggest promise. (September 30, 2009). In: *BWPI Working Paper No. 100.*

Ioannou, I., Serafeim, G. (2015). The impact of corporate social responsibility on investment recommendations: Analysts' perceptions and shifting institutional logics. In: *Strategic Management Journal, 36*(7), 1053-1081.

Ioannou, I., Serafeim, G. (2019). Corporate sustainability: a strategy?. In: *Harvard Business School Accounting & Management Unit Working Paper*, (19-065).

IR (2015). Materiality in <IR>. Guidance for the preparation of integrated reports. International Federation of Accountants and the International Integrated Reporting Council, London.

Jegadeesh, N., Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. In: *The Journal of Finance*, *48*(1), 65-91.

Jensen, M. C. (1968). The performance of mutual funds in the period 1945-1964. In: *The Journal of Finance*, *23*(2), 389-416.

Jensen, M. C., Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. In: *Journal of Financial Economics*, *3*(4), 305-360.

Jiraporn, P., Withisuphakorn, P. (2016). The effect of firm maturity on corporate social responsibility (CSR): do older firms invest more in CSR?. In: *Applied Economics Letters*, *23*(4), 298-301.

Johnston, A., Morrow, P. (2016). Fiduciary Duties of European Institutional Investors: Legal Analysis and Policy Recommendations. In: *University of Oslo Faculty of Law Research Paper*, (2016-04).

Kaiser, L. (2020). ESG integration: Value, growth and momentum. In: *Journal of Asset Management*, *21*(1), 32-51.

Kalaitzoglou, I., Pan, H., Niklewski, J. (2021). Corporate social responsibility: How much is enough? A higher dimension perspective of the relationship between financial and social performance. In: *Annals of Operations Research*, *306*(1), 209-245.

Kempf, A., Osthoff, P. (2007). The effect of socially responsible investing on portfolio performance. In: *European Financial Management*, *13*(5), 908-922.

Khan, M., Serafeim, G., Yoon, A. (2016). Corporate sustainability: First evidence on materiality. In: *The Accounting Review*, *91*(6), 1697-1724.

Khandker, S. R., Koolwal, G. B., Samad, H. A. (2009). Handbook on impact evaluation: quantitative methods and practices. Washington, DC: World Bank Publications.

Kinderman, D. (2018). Global and EU-Level Corporate Social Responsibility: Dynamism, Growth, and Conflict. In: Gesellschaftliche Verantwortung von Unternehmen in Deutschland (pp. 101-113). Wiesbaden: Springer.

Kleinbaum, D. G., Dietz, K., Gail, M., Klein, M., Klein, M. (2010). Logistic regression. New York: Springer.

Kolodinsky, J., Stewart, C., Bullard, A. (2006). Measuring economic and social impacts of membership in a community development financial institution. In: *Journal of Family and Economic Issues*, *27*(1), 27-47.

Korteweg, A. (2010). The net benefits to leverage. In: *The Journal of Finance*, *65*(6), 2137-2170.

Kotsantonis, S., Serafeim, G. (2019). Four things no one will tell you about ESG data. In: *Journal of Applied Corporate Finance*, *31*(2), 50-58.

Koundouri, P., Devves, S., Plataniotis, A. (2021). Alignment of the European Green Deal, the Sustainable Development Goals and the European Semester Process: Method and Application. In: *Theoretical Economics Letters*, *11*(4), 743-770.

KPMG (2020). The time has come. The KPMG survey of sustainability reporting 2020. KPMG, Amstelveen.

Krueger, P. (2015). Corporate goodness and shareholder wealth. In: *Journal of financial economics*, *115*(2), 304-329.

Lakonishok, J., Shleifer, A., Vishny, R. W. (1994). Contrarian investment, extrapolation, and risk. In: *The Journal of Finance*, *49*(5), 1541-1578.

Lantos, G. P. (2001). The boundaries of strategic corporate social responsibility. In: *Journal of Consumer Marketing*, *18*(7), 595-632.

Larcker, D., Tayan, B. (2021). Corporate governance matters: A closer look at organizational choices and their consequences. London: Pearson education.

Lashof, D. A., Ahuja, D. R. (1990). Relative contributions of greenhouse gas emissions to global warming. In: *Nature*, *344*, 529-531.

Lee, D. D., Faff, R. W. (2009). Corporate sustainability performance and idiosyncratic risk: A global perspective. In: *Financial Review*, *44*(2), 213-237.

Lee, C. M., Myers, J., Swaminathan, B. (1999). What is the Intrinsic Value of the Dow?. In: *The Journal of Finance*, *54*(5), 1693-1741.

Li, T. T., Wang, K., Sueyoshi, T., Wang, D. D. (2021). ESG: Research progress and future prospects. In: *Sustainability*, *13*(21), 11663.

Liang, H., Renneboog, L. (2017). On the foundations of corporate social responsibility. In: *The Journal of Finance*, *72*(2), 853-910.

Liang, H., Renneboog, L. (2020). Corporate social responsibility and sustainable finance: A review of the literature. In: *European Corporate Governance Institute – Finance Working Paper*, *701*(2020).

Linnér, B. O., Selin, H. (2013). The United Nations Conference on Sustainable Development: forty years in the making. In: *Environment and Planning C: Government and Policy, 31*(6), 971-987.

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

Louche, C., Arenas, D., Van Cranenburgh, K. C. (2012). From preaching to investing: Attitudes of religious organisations toward responsible investment. In: *Journal of Business Ethics, 110*(3), 301-320.

Lucarelli, C., Mazzoli, C., Rancan, M., Severini, S. (2020). Classification of sustainable activities: EU taxonomy and scientific literature. In: *Sustainability, 12*(16), 6460.

Luo, X., Bhattacharya, C. B. (2009). The debate over doing good: Corporate social performance, strategic marketing levers, and firm-idiosyncratic risk. In: *Journal of Marketing, 73*(6), 198-213.

Maas, K., Liket, K. (2011). Social impact measurement: Classification of methods. In Environmental management accounting and supply chain management (pp. 171-202). Dordrecht: Springer.

Macintosh, J. C. (1999). The issues, effects and consequences of the Berle – Dodd debate, 1931-1932. In: *Accounting, Organizations and Society, 24*(2), 139-153.

Malkiel, B. G., Xu, Y. (1997). Risk and return revisited. In: *Journal of Portfolio Management, 23*(3), 9-14.

Margolis, J. D., Elfenbein, H. A., Walsh, J. P. (2009). Does it pay to be good? A meta-analysis and redirection of research on the performance between corporate social and financial performance. Working Paper, available on SSRN at http://ssrn.com/absract=1866371.

Markowitz, H. (1952). Portfolio Selection. In: *The Journal of Finance, 7*(1), 77–99.

Martinez, D. E., Cooper, D. J. (2020). Seeing Through the Logical Framework. In: *International Journal of Voluntary and Nonprofit Organizations, 31*(6), 1239-1253.

Matos, P. (2020). ESG and responsible institutional investing around the world: A critical review. Charlottesville, VA: CFA Institute Research Foundation.

Mattingly, J. E., Berman, S. L. (2006). Measurement of corporate social action: Discovering taxonomy in the Kinder Lydenburg Domini ratings data. In: *Business & Society, 45*(1), 20-46.

McBride (2021). Annual report 2021. https://www.mcbride.co.uk/investors/results-centre/, accessed: 05/15/2022.

McLaughlin, J. A., Jordan, G. B. (2004). Using logic models. In: McLaughlin, J. A., Jordan, G. B., Wholey, J. S., Hatry, H. P., Newcomer, K. E. (Eds.). Handbook of practical program evaluation. San Francisco, CA: Wiley.

McWilliams, A., Siegel, D. (2000). Corporate social responsibility and financial performance: correlation or misspecification?. In: *Strategic Management Journal*, *21*(5), 603-609.

McWilliams, A., Siegel, D. S., & Wright, P. M. (2006). Corporate social responsibility: Strategic implications. In: *Journal of Management Studies*, *43*(1), 1-18.

Meadows, D. H., Meadows, D. L., Randers, J., Behrens III, W. W. (1972). The limits to growth. Club of Rome report. New York: Universe Books.

Meuer, J., Kölbel, J., Hoffmann, V. H. (2020). On the nature of corporate sustainability. In: *Organization & Environment*, *33*(3), 319-341.

MSCI (2022). MSCI ESG ratings methodology. Executive summary. June 2022. MSCI ESG Research LLC, New York City.

Muñoz-Torres, M. J., Fernández-Izquierdo, M. Á., Rivera-Lirio, J. M., Ferrero-Ferrero, I., Escrig-Olmedo, E., Gisbert-Navarro, J. V., Marullo, M. C. (2018). An assessment tool to integrate sustainability principles into the global supply chain. In: *Sustainability*, *10*(2), 535.

Nestlé (2021). Annual report 2020. https://www.nestle.com/investors/annual-report, accessed: 02/16/2022.

Newcomer, K. E., Hatry, H. P., Wholey, J. S. (2004). Meeting the need for practical evaluation approaches: An introduction. In: McLaughlin, J. A., Jordan, G. B., Wholey, J. S., Hatry, H. P., Newcomer, K. E. (Eds.). Handbook of practical program evaluation. San Francisco, CA: Wiley.

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. In: *Quality & quantity*, *41*(5), 673-690.

O'Connell M., Ward A.M. (2020). Shareholder Theory/Shareholder Value. In: Idowu S., Schmidpeter R., Capaldi N., Zu L., Del Baldo M., Abreu R. (Eds). Encyclopedia of Sustainable Management. Cham: Springer.

Oikonomou, I., Brooks, C., Pavelin, S. (2012). The impact of corporate social performance on financial risk and utility: A longitudinal analysis. In: *Financial Management*, *41*(2), 483-515.

Orlitzky, M., Schmidt, F. L., Rynes, S. L. (2003). Corporate social and financial performance: A meta-analysis. In: *Organization studies*, *24*(3), 403-441.

Persson, L., Carney Almroth, B. M., Collins, C. D., Cornell, S., de Wit, C. A., Diamond, M. L., Fantke, P., Hassellöv, M., MacLeod, M., Ryberg, M. W., Jørgensen, P. S., Villarrubia-Gómez, P., Wang, Z., Hauschild, M. Z. (2022). Outside the safe operating space of the planetary boundary for novel entities. In: *Environmental Science & Technology*, *56*(3), 1510-1521.

Petersen, M. A. (2009). Estimating standard errors in finance panel datasets: Comparing approaches. In: *The Review of Financial Studies*, *22*(1), 435-480.

Phillips, R. A., Freeman, R. E. (2010). Stakeholders. Cheltenham: Edward Elgar Publishing.

Porter, M. E., Kramer, M. R. (2006). The link between competitive advantage and corporate social responsibility. In: *Harvard Business Review*, *84*(12), 78-92.

Preston, L. E., O'bannon, D. P. (1997). The corporate social-financial performance relationship: A typology and analysis. In: *Business & Society*, *36*(4), 419-429.

Ramanathan, V., Feng, Y. (2009). Air pollution, greenhouse gases and climate change: Global and regional perspectives. In: *Atmospheric Environment*, *43*(1), 37-50.

Reber, B., Gold, A., Gold, S. (2021). ESG disclosure and idiosyncratic risk in initial public offerings. In: *Journal of Business Ethics*, 1-20.

Refinitiv (2022a). Environmental, social and governance scores from Refinitiv. May 2022. Refinitiv, New York City.

Refinitiv (2022b). The Refinitiv business classification. Methodology. January 2022. Refinitiv, New York City.

Renneboog, L., Ter Horst, J., Zhang, C. (2008). Socially responsible investments: Institutional aspects, performance, and investor behavior. In: *Journal of Banking & Finance*, *32*(9), 1723-1742.

Riedl, A., Smeets, P. (2017). Why do investors hold socially responsible mutual funds?. In: *The Journal of Finance*, *72*(6), 2505-2550.

Roberts, M. R., Whited, T. M. (2013). Endogeneity in empirical corporate finance1. In: *Handbook of the Economics of Finance*, *2*, 493-572.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin III, F. S., Lambin, E., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H. J., Nykvist, B., de Wit, C. A., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P. K., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R. W., Fabry, V. J., Handen, J., Walker, B.,

Liverman, D., Richardson, K., Crutzen, P., Foley, J. (2009). Planetary boundaries: exploring the safe operating space for humanity. In: *Ecology and Society*, *14*(2).

Rogers, J., Serafeim, G. (2019). Pathways to Materiality: How Sustainability Issues Become Financially Material to Corporations and Their Investors. In: *Harvard Business School Accounting & Management Unit Working Paper*, (20-056).

Rogers, P. J. (2008). Using programme theory to evaluate complicated and complex aspects of interventions. In: *Evaluation*, *14*(1), 29-48.

Rosenberg, B., Reid, K., Lanstein, R. (1985). Efficient capital markets: II. In: *Persuasive Evidence of Market Inefficiency*, *11*(3), 9-16.

Saadaoui, K., Soobaroyen, T. (2018). An analysis of the methodologies adopted by CSR rating agencies. In: *Sustainability Accounting, Management and Policy Journal*, *9*(1), 43-62.

Sahin, Ö., Bax, K., Czado, C., Paterlini, S. (2022a). Environmental, Social, Governance scores and the Missing pillar — Why does missing information matter?. In: *Corporate Social Responsibility and Environmental Management*, 1-22.

Sahin, Ö., Bax, K., Paterlini, S., Czado, C. (2022b). The pitfalls of (non-definitive) Environmental, Social, and Governance scoring methodology. In: *Global Finance Journal (forthcoming)*.

SASB (2017). SASB Conceptual framework. Sustainable Accounting Standards Board. February 2017. SASB, San Francisco.

Sassen, R., Hinze, A. K., Hardeck, I. (2016). Impact of ESG factors on firm risk in Europe. In: *Journal of Business Economics*, *86*(8), 867-904.

Schneider, A. (2012). Reifegradmodell CSR – eine Begriffsklärung und -abgrenzung. In: Schneider, A., Schmidpeter, R. (Eds.) Corporate Social Responsibility. Springer, Berlin, Heidelberg, 17-38.

Schreck, P., Raithel, S. (2018). Corporate social performance, firm size, and organizational visibility: Distinct and joint effects on voluntary sustainability reporting. In: *Business & Society*, *57*(4), 742-778.

Schütze, F., Stede, J., Blauert, M., Erdmann, K. (2020). EU taxonomy increasing transparency of sustainable investments. In: *DIW Weekly Report*, *10*(51), 485-492.

Schwalbach, J., Klink, D. (2012). Der Ehrbare Kaufmann als individuelle verantwortungskategorie der CSR-Forschung. In: Corporate social responsibility (pp. 219-240). Berlin, Heidelberg: Springer.

Semenova, N., Hassel, L. G. (2015). On the validity of environmental performance metrics. In: *Journal of Business Ethics*, *132*(2), 249-258.

Serafeim, G. (2021). ESG: Hyperboles and Reality. In: *Harvard Business School Research Paper Series Working Paper*, (22-031).

Serafeim, G., Yoon, A. (2022). Which Corporate ESG News does the Market React to?. In: *Financial Analysts Journal*, *78*(1), 59-78.

Serafeim, G., Zochowski, T. R., Downing, J. (2019). Impact-weighted financial accounts: The missing piece for an impact economy. In: *White Paper, Harvard Business School*, Boston.

SFDR (2019). Regulation (EU) 2019/2088 of the European parliament and of the council of 27 November 2019 on sustainability-related disclosures in the financial services sector. In: *Official Journal of the European Union*, L317.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. In: *The Journal of Finance*, *19*(3), 425-442.

Shleifer, A., Vishny, R. W. (1997). A survey of corporate governance. In: *The Journal of Finance*, *52*(2), 737-783.

Sison, A. J. G. (2011). Aristotelian corporate governance. In: Corporate governance and business ethics (pp. 179-201). Dordrecht: Springer.

Smith, A. (1937). The Wealth of Nations (1776). New York: Modern Library.

Steffen, W., Richardson, K., Rockström, J., Cornell, S. E., Fetzer, I., Bennett, E. M., Bigg, R., Carpenter, S. R., de Vries, W., de Wit, C. A., Folke, C., Gerten, D., Heinke, J., Mace, G. M., Persson, L. M., Ramanathan, V., Reyers, B., Sörlin, S. (2015). Planetary boundaries: Guiding human development on a changing planet. In: *Science*, *347*(6223), 1259855.

Sustainalytics (2020). The ESG risk rating: frequently asked questions – for companies. August 2020. Sustainalytics, Amsterdam.

S&P (2022). S&P Global ESG Scores. Methodology. S&P Global, New York City.

Taubken, N., Feld, T. Y. (2018). Impact measurement and the concept of materiality—new requirements and approaches for materiality assessments. In: *Sustainability Management Forum*, *26*(1), 87-100.

Thomson Reuters (2017). Thomson Reuters ESG scores. March 2017. Thomson Reuters, New York City.

Thomson Reuters (2018). Thomson Reuters ESG scores. May 2018. Thomson Reuters, New York City.

Tirole, J. (2001). Corporate governance. In: *Econometrica*, *69*(1), 1-35.

Trapp, N. L. (2012). Corporation as climate ambassador: Transcending business sector boundaries in a Swedish CSR campaign. In: *Public Relations Review*, *38*(3), 458-465.

UN (2011). Guiding principles on business and human rights: implementing the United Nations "protect, respect and remedy" framework. United Nations, Geneva.

UN (2015). Paris agreement. United Nations, Geneva.

UN Global Compact (2004). Who cares wins: Connecting financial markets to a changing world. Global Compact, New York.

UN PRI (2017). Fiduciary duties in the 21st century. Germany roadmap. United Nations Principles of Responsible Investing, London.

Van Deth, J. (2003). Using Published Survey Data. In: Harkness, J. A., van de Vijver, F. J., Mohler, P. P., Wiley, J. (Eds.). (2003). Cross-cultural survey methods (Vol. 325). Hoboken, NJ: Wiley-Interscience.

Van Duuren, E., Plantinga, A., Scholtens, B. (2016). ESG integration and the investment management process: Fundamental investing reinvented. In: *Journal of Business Ethics*, *138*(3), 525-533.

Verrecchia, R. E. (2001). Essays on disclosure. In: *Journal of accounting and economics*, *32*(1-3), 97-180.

Waddock, S. A., Graves, S. B. (1997). The corporate social performance–financial performance link. In: *Strategic Management Journal*, *18*(4), 303-319.

Walter, I. (2020). Sense and Nonsense in ESG Ratings. In: *Journal of Law, Finance, and Accounting*, *5*(2), 307-336.

Walton, C. C. (1967). Corporate social responsibilities. Belmont, CA: Wadsworth.

Warde, I. (2010). Islamic finance in the global economy. Edinburgh University Press.

Waßmann, J. (2014). Abgrenzung des CSR-Begriffs. In: Grundlagen der CSR aus der Perspektive des Marketings (pp. 3-11). Wiesbaden: Springer Gabler.

Weber, M. (2008). The business case for corporate social responsibility: A company-level measurement approach for CSR. In: *European Management Journal*, *26*(4), 247-261.

WEF (2003). Responding to the leadership challenge: Findings of a CEO survey on global corporate citizenship. World Economic Forum, Geneva.

Weilert, K. (2010). Taming the untamable? Transnational corporations in United Nations law and practice. In: *Max Planck Yearbook of United Nations Law*, *14*, 445-506.

Whelan, T., Atz, U., Van Holt, T., Clark, C. (2021). ESG and financial performance: Uncovering the relationship by aggregating evidence from 1,000 plus studies published between 2015-2020. New York: NYU STERN Center for sustainable business.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. In: *Econometrica: Journal of the Econometric Society*, *48*(4), 817-838.

Whiteman, G., Walker, B., Perego, P. (2013). Planetary boundaries: Ecological foundations for corporate sustainability. In: *Journal of Management Studies*, *50*(2), 307-336.

Windolph, S. E. (2011). Assessing corporate sustainability through ratings: challenges and their causes. In: *Journal of Environmental Sustainability*, *1*(1), 37-57.

Wong, C., Petroy, E. (2020). Rate the raters 2020: Investor survey and interview results. SustainAbility, London.

Wong, C., Brackley, A., Petroy, E. (2019). Rate the raters 2019: Expert views on ESG ratings. SustainAbility, London.

Wood, D. J. (1991). Corporate social performance revisited. In: *Academy of Management Review*, *16*(4), 691-718.

Wooldridge, J. M. (2010). Econometrics of cross section and panel data. Cambridge, MA: MIT Press.

Wooldridge, J. M. (2020). Introductory econometrics: a modern approach (Seventh edition). South-Western, Cengage Learning.

Wu, K., Lai, S. (2020). Intangible intensity and stock price crash risk. In: J*ournal of Corporate Finance*, *64*, 101682.

APPENDIX

Appendix 1: UN PRI - Signatories' commitment[727]

"As institutional investors, we have a duty to act in the best long-term interests of our beneficiaries. In this fiduciary role, we believe that environmental, social, and corporate governance (ESG) issues can affect the performance of investment portfolios (to varying degrees across companies, sectors, regions, asset classes and through time).

We also recognise that applying these Principles may better align investors with broader objectives of society. Therefore, where consistent with our fiduciary responsibilities, we commit to the following:

Principle 1: We will incorporate ESG issues into investment analysis and decision-making processes.

Principle 2: We will be active owners and incorporate ESG issues into our ownership policies and practices.

Principle 3: We will seek appropriate disclosure on ESG issues by the entities in which we invest.

Principle 4: We will promote acceptance and implementation of the Principles within the investment industry.

Principle 5: We will work together to enhance our effectiveness in implementing the Principles.

Principle 6: We will each report on our activities and progress toward implementing the Principles.

The Principles for Responsible Investment were developed by an international group of institutional investors reflecting the increasing relevance of environmental, social and corporate governance issues to investment practices. The process was convened by the United Nations Secretary-General.

In signing the Principles, we as investors publicly commit to adopt and implement them, where consistent with our fiduciary responsibilities. We also commit to evaluate the effectiveness and improve the content of the Principles over time. We believe this will improve our ability to meet commitments to beneficiaries as well as better align our investment activities with the broader interests of society.

We encourage other investors to adopt the Principles."

Appendix 1 UN PRI - Signatories' commitment

[727] https://www.unpri.org/about-us/what-are-the-principles-for-responsible-investment, accessed: 07/06/2022.

Appendix 2: UN Global Compact Principles[728]

"Human Rights

Principle 1: Businesses should support and respect the protection of internationally proclaimed human rights within their sphere of influence; and

Principle 2: make sure that they are not complicit in human rights abuses.

Labour

Principle 3: Businesses should uphold the freedom of association and the effective recognition of the right to collective bargaining;

Principle 4: the elimination of all forms of forced and compulsory labour;

Principle 5: the effective abolition of child labour; and

Principle 6: eliminate discrimination in respect of employment and occupation.

Environment

Principle 7: Businesses should support a precautionary approach to environmental challenges;

Principle 8: undertake initiatives to promote greater environmental responsibility; and

Principle 9: encourage the development and diffusion of environmentally friendly technologies.

Anti-Corruption

Principle 10: Businesses should work against corruption in all its forms, including extortion and bribery."

Appendix 2 UN Global Compact Principles

[728] https://www.unglobalcompact.org/what-is-gc/mission/principles, accessed: 07/07/2021.

Appendix 3: Definitions of main variables

The following table describes the main variables and their data sources applied throughout this study.

Variable	Definition	Source	Formula
CapEx intensity	Capital expenditures / Total assets	Refinitiv Worldscope	$CapEx\ intensity_{i,t} = \frac{capital\ expenditures_{i,t}}{total\ assets_{i,t}}$
Firm age	Years since the firm's IPO	Refinitiv Eikon	
Firm size	Natural logarithm of total assets	Refinitiv Worldscope	
IA intensity	Intangible assets / Total assets	Refinitiv Worldscope	$IA\ intensity_{i,t} = \frac{intangible\ assets\ _{i,t}}{total\ assets_{i,t}}$
Idiosyncratic volatility	Standard deviation of daily residuals from a Carhart four-factor model based on daily stock returns over one year	Refinitiv Worldscope	$IV_{i,y} = std(\varepsilon_{i,t})$
ROA	Return on assets	Refinitiv Worldscope	$ROA_{i,t} = \frac{net\ income_{i,t} + [interest\ expenses_{i,t} * (1 - tax\ rate_t)]}{\frac{1}{2}\sum_{t=-1}^{0} total\ assets_{i,t}}$
ROA volatility	Standard deviation of quarterly ROA over three years prior to the observed year	Refinitiv Worldscope	$ROA\ volatility_{i,t} = std(ROA_{i,q_{t_{[-3;-1]}}})$
R&D intensity	Research and development expenses / Total assets	Refinitiv Worldscope	$R\&D\ intensity_{i,t} = \frac{research\ \&\ development\ expenditures_{i,t}}{total\ assets_{i,t}}$
Leverage	Long-term and short-term debt / Common equity	Refinitiv Worldscope	$Leverage_{i,t} = \frac{(long-term\ debt_{i,t} + short-term\ debt_{i,t})}{common\ equity_{i,t}}$
ESG	Refinitiv ESG performance scores based on the ASSET4 universe as of April 2022	Refinitiv ASSET4	
zESG	Refinitiv ESG performance scores based on the ASSET4 universe as of April 2022 (standardized)	Refinitiv ASSET4	$zESG_t = \frac{ESG_t - \overline{ESG_t}}{std(ESG)_t}$
zadjESG	Refinitiv ESG performance scores based on the ASSET4 universe as of April 2022 (standardized based on yearly firm size deciles)	Refinitiv ASSET4	$zadjESG_{i,t} = \frac{ESG_{i,t} - \overline{ESG}_{I,t}}{std(ESG)_{I,t}}$
ESGmat	The suffix "mat" identifies materiality-adjusted ESG scores based on the approach developed in chapter 5.5.1.	Refinitiv ASSET4	
ESG_adj	The suffix "_adj" indicates whether only fully covered firm-year observations are considered.	Refinitiv ASSET4	
RmRf	Market premium following Fama/French (1993)	Kenneth French online data library	
SMB	Size premium following Fama/French (1993)	Kenneth French online data library	
HML	Value premium following Fama/French (1993)	Kenneth French online data library	
MOM	Momentum premium following Carhart (1997)	Kenneth French online data library	

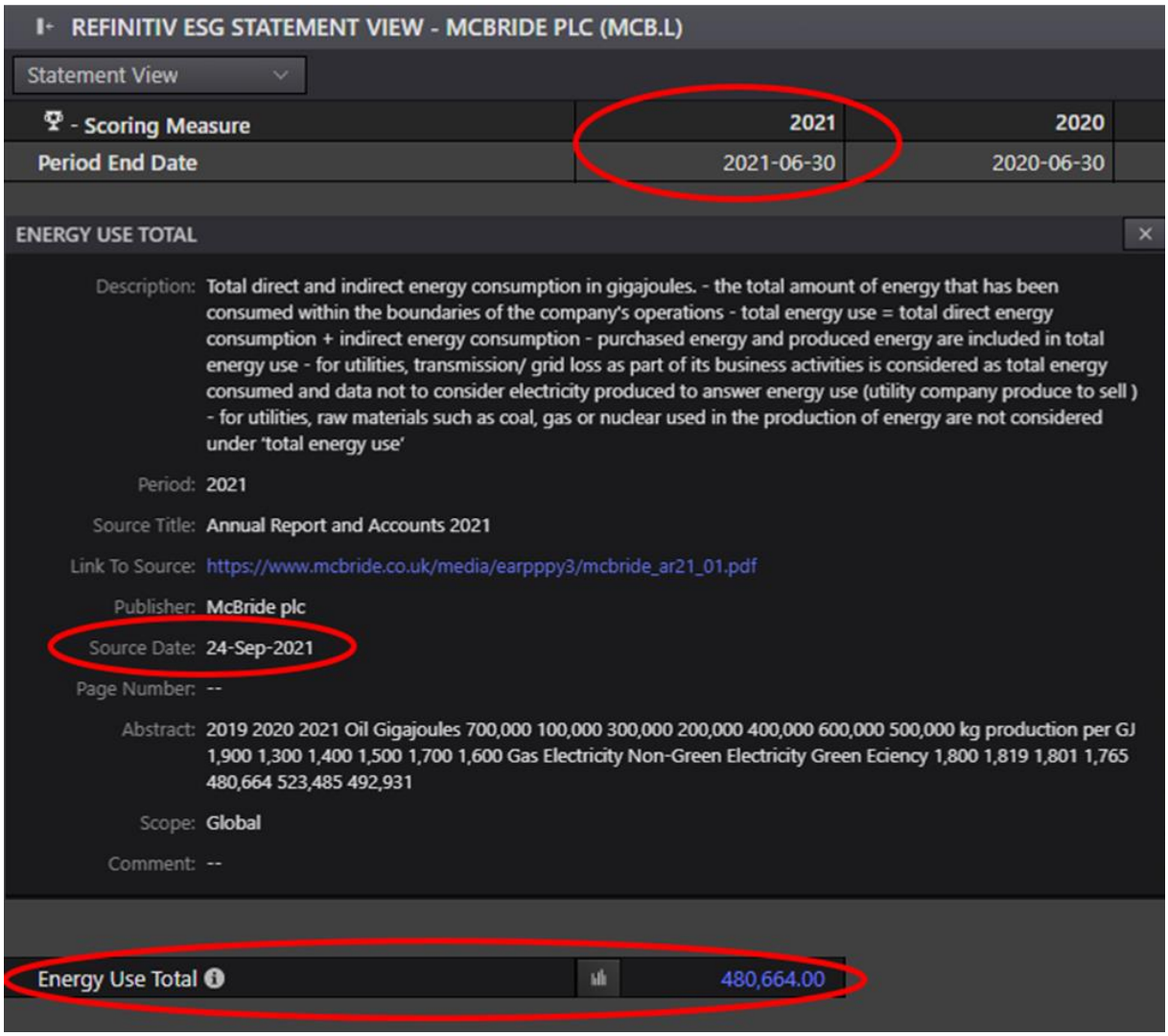
Appendix 3 Definitions of main variables

Appendix 4: ESG scores fiscal year vs. calendar year allocation (example 1)

The following screenshots are extracted from the Refinitiv Eikon application. The screenshots are stripped to the relevant content.

Example 1: McBride plc (GB0005746358)

The screenshot shows the numeric data point "Energy Use Total" and its source for McBride plc. The data point accounts for "direct and indirect energy consumption in gigajoules", which is reflected in the "Resource Use" pillar of the Refinitiv ESG rating universe. The respective value was reported in the annual report of McBride 2021.[729] The data point is allocated to the column 2021, although the value refers to the fiscal year period of June 30th, 2020, to June 30th, 2021. If there are no shifts in the reporting pattern and fiscal years, it can be assumed that yearly values remain comparable among different firms. However, considering that energy consumption might be affected by external factors (e.g., climate and seasonal weather conditions), it might also add some sort of confusion. For example, the fiscal year of McBride captures two different summer seasons. A fiscal year in line with the calendar year captures two different winter seasons. Hence, the best-in-class scoring approach would have different referential periods that might quantitatively differ.



Appendix 4 ESG scores fiscal year vs. calendar year allocation (Example 1: McBride plc)

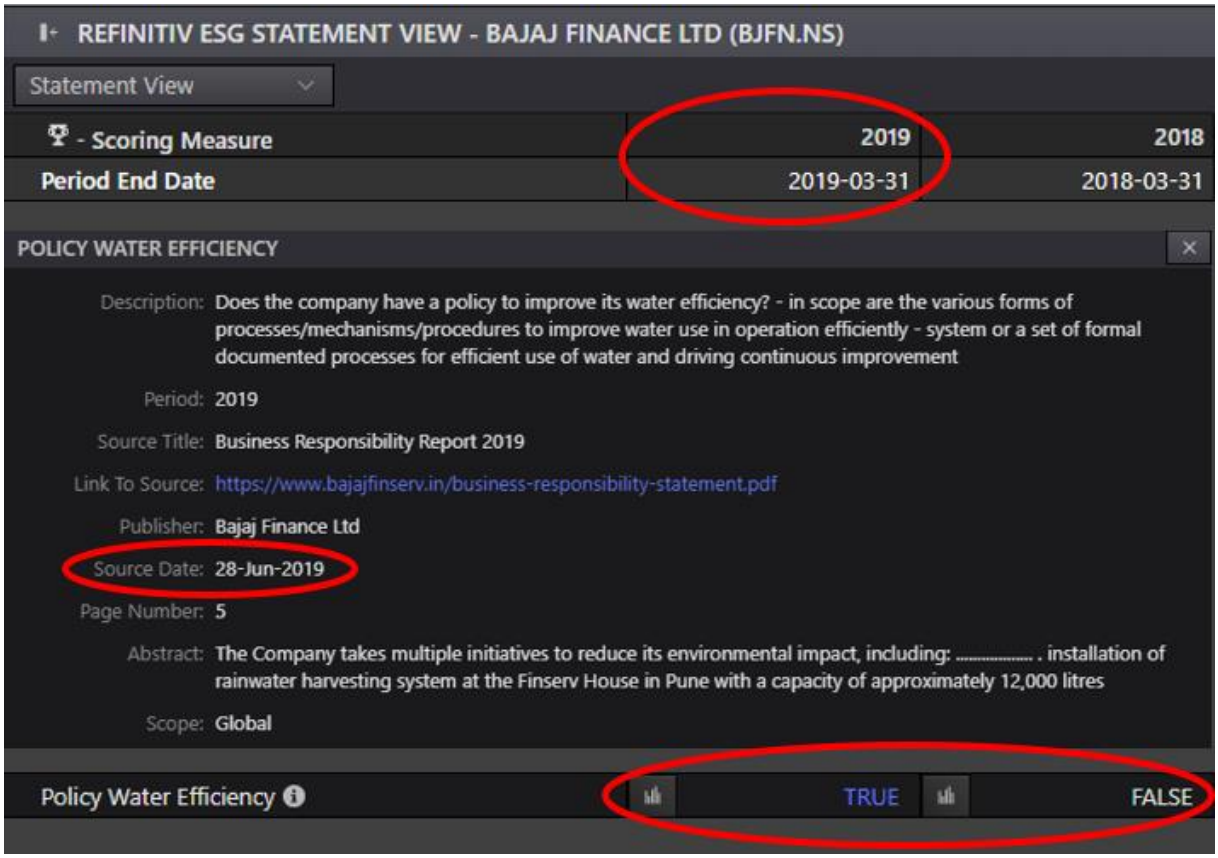
[729] McBride (2021, p. 43)

Appendix 5: ESG scores fiscal year vs. calendar year allocation (example 2)

The following screenshots are extracted from the Refinitiv Eikon application. The screenshots are stripped to the relevant content.

Example 2: Bajaj Finance ltd (INE296A01024)

The screenshot shows the Boolean data point "Policy Water Efficiency", which identifies whether a company implemented policies to enhance the efficient use of water. It refers to the "Resource Use" pillar in the Refinitiv ESG rating universe. According to the data, Bajaj Finance ltd began to report on a respective policy in 2019. The indicator changes from "FALSE" (0) to "TRUE" (1), which due to its polarity, enhances ESG performance. The value is allocated to the column 2019, although it refers to a fiscal year that ended on March 31st, 2019. The data does not provide information on when the policy became effective; the scoring approach, however, accounts for the change not earlier than 2019. Considering the change may become effective just after the fiscal year began in 2018, the best-in-class approach does not account for this in the comparison of different firms in 2018, which might give other firms an advantage.



Appendix 5 ESG scores fiscal year vs. calendar year allocation (Example 2: Bajaj Finance ltd)

Appendix 6: Country distribution

Number and fraction of firms per grouped by country of domicile and following M49 regional groupings

Country	M49 regional group	Number of firms	Fraction of firms
United States	AMERICAS	4,091	34.7%
China	ASIA	953	8.1%
United Kingdom	EUROPE	831	7.0%
Canada	AMERICAS	603	5.1%
Australia	OCEANIA	593	5.0%
Japan	ASIA	514	4.4%
Sweden	EUROPE	345	2.9%
Germany	EUROPE	298	2.5%
Switzerland	EUROPE	231	2.0%
France	EUROPE	199	1.7%
India	ASIA	180	1.5%
South Korea	ASIA	176	1.5%
Taiwan	ASIA	172	1.5%
Hong Kong	ASIA	171	1.5%
South Africa	AFRICA	158	1.3%
Brazil	AMERICAS	155	1.3%
Italy	EUROPE	134	1.1%
Singapore	ASIA	115	1.0%
Thailand	ASIA	115	1.0%
Netherlands	EUROPE	98	0.8%
Spain	EUROPE	92	0.8%
Finland	EUROPE	91	0.8%
Norway	EUROPE	91	0.8%
Turkey	ASIA	84	0.7%
Malaysia	ASIA	73	0.6%
Denmark	EUROPE	69	0.6%
New Zealand	OCEANIA	66	0.6%
Bermuda	AMERICAS	63	0.5%
Mexico	AMERICAS	62	0.5%
Argentina	AMERICAS	59	0.5%
Belgium	EUROPE	57	0.5%
Ireland	EUROPE	57	0.5%
Russian Federation	EUROPE	52	0.4%
Indonesia	ASIA	51	0.4%
Chile	AMERICAS	46	0.4%
Luxembourg	EUROPE	46	0.4%
Qatar	ASIA	46	0.4%
Poland	EUROPE	44	0.4%
Israel	ASIA	42	0.4%
Austria	EUROPE	40	0.3%
Saudi Arabia	ASIA	40	0.3%
Greece	EUROPE	35	0.3%
Peru	AMERICAS	33	0.3%
Guernsey	EUROPE	32	0.3%
Philippines	ASIA	30	0.3%
United Arab Emirates	ASIA	29	0.2%
Colombia	AMERICAS	24	0.2%
Portugal	EUROPE	19	0.2%
Jersey	EUROPE	16	0.1%
Cyprus	ASIA	15	0.1%
Kuwait	ASIA	15	0.1%
Egypt	AFRICA	12	0.1%
Oman	ASIA	10	0.1%
Bahrain	ASIA	9	0.0%
Cayman Islands	AMERICAS	9	0.0%
Iceland	EUROPE	9	0.0%
Malta	EUROPE	7	0.0%
Hungary	EUROPE	6	0.0%
Isle of Man	EUROPE	6	0.0%
Jordan	ASIA	6	0.0%
Czechia	EUROPE	5	0.0%
Monaco	EUROPE	5	0.0%
Pakistan	ASIA	5	0.0%
Puerto Rico	AMERICAS	5	0.0%
British Virgin Islands	AMERICAS	4	0.0%
Kazakhstan	ASIA	3	0.0%
Macau	ASIA	3	0.0%

Appendix 6 Number and fraction of firms per grouped by country of domicile and following M49 regional groupings

Appendix 6: Country distribution (cont.)

Number and fraction of firms per grouped by country of domicile and following M49 regional groupings

Country	M49 regional group	Number of firms	Fraction of firms
Morocco	AFRICA	3	0.0%
Panama	AMERICAS	3	0.0%
Romania	EUROPE	3	0.0%
Viet Nam	ASIA	3	0.0%
Bahamas	AMERICAS	2	0.0%
Faroe Islands	EUROPE	2	0.0%
Gibraltar	EUROPE	2	0.0%
Liechtenstein	EUROPE	2	0.0%
Nigeria	AFRICA	2	0.0%
Papua New Guinea	OCEANIA	2	0.0%
Uganda	AFRICA	2	0.0%
Azerbaijan	ASIA	1	0.0%
Cambodia	ASIA	1	0.0%
Costa Rica	AMERICAS	1	0.0%
Kenya	AFRICA	1	0.0%
Mongolia	ASIA	1	0.0%
Slovenia	EUROPE	1	0.0%
Sri Lanka	ASIA	1	0.0%
Ukraine	EUROPE	1	0.0%
United States Virgin Islands	AMERICAS	1	0.0%
Uruguay	AMERICAS	1	0.0%
Zimbabwe	AFRICA	1	0.0%

Appendix 6 Number and fraction of firms grouped by country of domicile and following M49 regional groupings (cont.)

Appendix 7: Industry distribution of observations per year

Number of observations per year in each TRBC industry group (grouped by TRBC economic sector and TRBC business sector)

Economic Sector | Business Sector | Industry Group	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Sum
Financials																						
Banking & Investment Services																						
Banking Services	4	67	83	127	140	145	164	221	269	302	310	314	316	331	425	565	692	761	812	813	65	6.926
Investment Banking & Investment Services	6	27	36	63	67	71	80	90	108	112	118	112	124	134	154	178	211	244	288	298	20	2.541
Collective Investments																						
Collective Investments		1	3	3	3	3	3	3	4	4	5	5	4	19	44	51	60	67	83	78	14	457
Holding Companies																						
Holding Companies		1	1	5	5	5	5	8	9	9	9	9	10	10	11	11	13	19	23	28	2	193
Insurance																						
Insurance		36	42	73	83	84	93	116	125	143	143	142	151	155	171	200	213	220	234	231	3	2.658
Industrials																						
Industrial & Commercial Services																						
Construction & Engineering		9	22	44	48	48	54	68	81	94	102	102	103	102	110	119	134	161	171	182	15	1.769
Diversified Industrial Goods Wholesalers	3	3	7	8	9	9	9	9	10	12	13	14	14	15	15	16	17	16	19	18		236
Professional & Commercial Services	4	27	39	65	71	72	79	94	111	112	117	119	126	131	171	209	215	253	281	300	16	2.612
Industrial Conglomerates																						
Industrial Conglomerates	1	1	2	3	3	3	3	5	5	5	5	5	5	5	5	5	5	5	5	2		78
Industrial Goods																						
Aerospace & Defense		20	21	25	29	29	29	31	35	37	41	41	41	40	57	62	69	71	76	84	5	843
Machinery, Equipment & Components	3	35	63	100	108	110	116	142	158	189	192	193	190	194	245	279	317	380	428	483	29	3.954
Transportation																						
Freight & Logistics Services	1	13	21	30	32	32	36	44	47	58	61	62	62	67	81	94	107	122	137	142	6	1.255
Passenger Transportation Services	3	11	20	34	34	34	35	44	45	54	54	56	57	59	68	74	80	85	89	88	4	1.028
Transport Infrastructure		5	10	15	15	15	21	28	27	36	40	42	44	48	53	57	65	66	78	76	1	742
Technology																						
Financial Technology (Fintech) & Infrastructure																						
Financial Technology (Fintech) & Infrastructure						2	2	2	3	3	3	3	4	4	4	7	13	16	20	27	2	115
Software & IT Services																						
Software & IT Services	12	30	39	68	75	80	91	109	132	140	141	145	156	174	257	311	378	429	519	635	28	3.949
Technology Equipment																						
Communications & Networking	1	10	11	15	17	17	18	19	21	24	25	24	26	30	41	49	66	70	92	99	9	684
Computers, Phones & Household Electronics	6	10	13	20	23	23	22	27	34	47	47	45	44	41	54	62	67	77	82	88	7	839
Electronic Equipment & Parts	1	3	4	13	14	14	14	18	19	29	31	32	32	32	40	50	65	70	78	90	2	651
Office Equipment	2	5	7	10	11	11	11	11	11	11	11	11	11	11	11	11	12	15	15	16		214
Semiconductors & Semiconductor Equipment	6	22	27	41	45	45	45	58	63	91	90	91	90	93	101	121	137	145	175	181	6	1.673
Telecommunications Services																						
Telecommunications Services	6	25	29	42	50	53	68	95	109	120	122	125	129	131	139	154	164	167	174	166	7	2.075
Consumer Cyclicals																						
Automobiles & Auto Parts																						
Automobiles & Auto Parts	4	18	30	47	52	52	55	68	78	95	99	99	103	105	114	126	148	170	195	207	4	1.869
Cyclical Consumer Products																						
Homebuilding & Construction Supplies	2	17	26	36	37	38	39	44	45	48	51	51	50	52	73	94	98	112	124	132	11	1.180
Household Goods		4	4	10	11	11	12	14	15	18	19	19	19	21	28	32	38	41	48	61	5	430
Leisure Products		4	6	9	9	11	11	11	13	15	15	15	16	20	24	31	39	46	51	57	6	409
Textiles & Apparel	2	11	14	22	24	25	27	28	30	42	42	43	45	47	53	60	71	84	97	97	3	867
Cyclical Consumer Services																						
Hotels & Entertainment Services	2	21	26	42	50	54	59	76	84	92	103	105	108	114	150	174	185	213	241	241	18	2.158
Media & Publishing	3	30	36	59	66	68	71	83	93	99	103	103	106	109	121	132	149	157	168	180	14	1.950
Retailers																						
Diversified Retail	8	12	16	23	23	24	26	33	39	47	49	52	54	58	63	64	69	75	77	80	6	898
Specialty Retailers	12	25	38	61	65	68	75	83	95	103	113	117	115	120	137	159	175	192	201	205	12	2.171

Appendix 7 Industry distribution of observations per year (final sample)

Appendix 7: Industry distribution of observations per year (cont.)

Number of observations per year in each TRBC industry group (grouped by TRBC economic sector and TRBC business sector)

Economic Sector | Business Sector | Industry Group	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Sum
Healthcare																						
Healthcare Services & Equipment																						
Healthcare Equipment & Supplies	4	24	28	38	43	43	49	59	65	67	64	64	67	66	98	142	178	203	235	261	12	1.810
Healthcare Providers & Services		11	12	18	20	20	24	35	38	41	40	42	44	54	69	86	92	104	114	132	7	1.003
Pharmaceuticals & Medical Research																						
Biotechnology & Medical Research		5	5	11	13	13	18	21	24	27	25	25	25	32	76	152	286	340	389	479	9	1.975
Pharmaceuticals	6	30	35	51	53	55	60	66	76	77	82	84	84	95	116	130	181	219	262	275	10	2.047
Basic Materials																						
Applied Resources																						
Containers & Packaging		9	12	16	18	18	21	26	29	29	32	35	39	39	41	42	44	47	55	51	2	605
Paper & Forest Products		4	6	8	10	11	13	17	20	23	20	21	22	23	27	33	38	43	55	59	5	458
Chemicals																						
Chemicals	4	32	46	72	79	80	87	94	107	131	136	138	141	145	163	176	189	216	238	233	10	2.517
Mineral Resources																						
Construction Materials		6	8	13	18	18	19	25	28	39	39	39	41	41	46	53	56	63	68	71	3	694
Metals & Mining	2	15	27	60	65	71	97	179	221	271	287	288	280	274	282	303	329	342	390	399	10	4.192
Real Estate																						
Real Estate																						
Real Estate Operations	3	10	25	52	57	59	65	71	80	110	118	125	131	145	160	171	207	240	297	305	8	2.439
Residential & Commercial REITs	2	19	28	63	71	72	82	106	116	122	131	137	143	150	245	293	310	335	364	360	17	3.166
Consumer Non-Cyclicals																						
Consumer Goods Conglomerates																						
Consumer Goods Conglomerates		9	14	16	18	18	20	23	26	33	35	34	35	35	37	38	39	46	48	46	4	574
Food & Beverages																						
Beverages	2	12	13	26	26	27	31	37	40	47	51	51	50	50	53	57	71	73	86	88	6	897
Food & Tobacco	4	22	27	47	53	53	67	79	98	110	124	129	131	141	173	187	210	249	285	282	21	2.492
Food & Drug Retailing																						
Food & Drug Retailing	8	19	20	33	37	41	45	54	59	64	68	70	71	74	81	85	98	112	122	121	10	1.292
Personal & Household Products & Services																						
Personal &Household Products & Services	4	11	14	21	24	24	24	30	39	40	39	42	43	45	56	62	65	65	74	72	2	796
Energy																						
Energy - Fossil Fuels																						
Coal				3	3	4	6	14	23	32	33	35	33	32	28	29	34	34	35	37	3	418
Oil & Gas		28	31	72	80	84	114	161	180	197	196	199	202	205	217	235	253	249	257	257	7	3.224
Oil & Gas Related Equipment and Services		19	20	44	47	47	58	79	87	90	94	94	96	99	109	118	142	151	154	155	9	1.712
Renewable Energy																						
Renewable Energy		2	2	5	6	7	10	11	14	18	17	15	15	15	15	21	29	31	43	57	5	338
Uranium																						
Uranium			1	2	2	2	5	7	9	8	7	7	6	6	6	7	8	11	12	13		119
Utilities																						
Utilities																						
Electrical Utilities & IPPs	3	25	30	48	56	57	67	95	107	124	121	120	121	131	154	162	170	186	197	194	1	2.169
Multiline Utilities		13	15	20	21	21	24	26	28	28	28	28	30	31	33	34	37	38	37	36		528
Natural Gas Utilities		1	3	5	6	7	9	17	22	24	24	24	24	23	31	35	36	41	44	45	2	423
Water & Related Utilities	2	3	4	6	6	6	6	7	9	11	11	12	13	13	15	19	26	28	31	30	1	259

Appendix 8 Industry distribution of observations per year (final sample – cont.)

Appendix 8: Year-to-year change in the number of firm-year observations per year and industry

Year-to-year change in the number of observations in each TRBC economic sector by TRBC business sector. TRBC economic sectors consisting of only one business sector are not reported (i.e., Utilities, Real Estate)

Economic Sector | Business Sector	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Financials																					
Banking & Investment Services		-29%	1%	-2%	-1%	1%	1%	0%	2%	-1%	1%	0%	0%	-1%	0%	2%	2%	1%	0%	0%	5%
Collective Investments		1%	1%	-1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	3%	0%	0%	0%	1%	0%	8%
Holding Companies		1%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Insurance		27%	-2%	1%	1%	-1%	0%	0%	-2%	1%	-1%	0%	1%	-1%	-3%	-1%	-2%	-1%	-1%	0%	-13%
Industrials																					
Industrial & Commercial Services		-15%	2%	3%	1%	0%	1%	0%	2%	-2%	1%	0%	1%	0%	-1%	1%	-1%	1%	0%	0%	4%
Industrial Conglomerates		-6%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Industrial Goods		24%	-3%	-2%	1%	0%	-2%	-1%	0%	1%	-1%	0%	-1%	-1%	2%	0%	1%	1%	0%	2%	4%
Transportation		-3%	1%	0%	-1%	0%	1%	1%	-2%	2%	0%	0%	0%	1%	-1%	-1%	0%	-1%	0%	-1%	-8%
Technology																					
Financial Technology (Fintech) & Infrastructure		0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	1%
Software & IT Services		-7%	1%	3%	-1%	1%	1%	-1%	2%	-4%	0%	0%	1%	2%	6%	1%	1%	1%	2%	4%	-3%
Technology Equipment		1%	0%	0%	-1%	-2%	-4%	-1%	-1%	6%	0%	-1%	-1%	-1%	-2%	0%	0%	0%	0%	-2%	3%
Telecommunications Services		6%	-2%	-2%	1%	0%	3%	3%	0%	-2%	0%	0%	0%	-1%	-4%	-1%	-2%	-1%	-2%	-2%	-1%
Consumer Cyclicals																					
Automobiles & Auto Parts		1%	3%	0%	0%	-1%	0%	1%	0%	1%	0%	0%	0%	0%	-1%	0%	1%	0%	1%	0%	-11%
Cyclical Consumer Products		13%	0%	-1%	-1%	0%	0%	-2%	-1%	1%	-1%	0%	0%	1%	2%	2%	0%	1%	1%	1%	4%
Cyclical Consumer Services		21%	-4%	1%	2%	0%	0%	1%	0%	-2%	1%	0%	0%	0%	1%	0%	-1%	0%	0%	-1%	7%
Retailers		-35%	1%	0%	-1%	0%	1%	-1%	1%	0%	0%	1%	-1%	0%	-1%	-1%	0%	-1%	-1%	-1%	0%
Healthcare																					
Healthcare Services & Equipment		10%	0%	-3%	1%	-1%	0%	4%	-1%	0%	-2%	0%	1%	-2%	-2%	-2%	-8%	-1%	-1%	-1%	16%
Pharmaceuticals & Medical Research		-10%	0%	3%	-1%	1%	0%	-4%	1%	0%	2%	0%	-1%	2%	2%	2%	8%	1%	1%	1%	-16%
Basic Materials																					
Applied Resources		20%	-2%	-4%	1%	0%	0%	-2%	-1%	-2%	0%	1%	1%	0%	0%	0%	0%	0%	1%	0%	10%
Chemicals		-18%	-2%	-4%	-1%	-1%	-4%	-9%	-1%	0%	0%	0%	0%	1%	1%	0%	0%	2%	-1%	-1%	5%
Mineral Resources		-2%	4%	8%	0%	1%	4%	11%	2%	1%	1%	-1%	-1%	-1%	-2%	0%	0%	-2%	0%	1%	-14%
Consumer Non-Cyclicals																					
Consumer Goods Conglomerates		12%	4%	-5%	0%	0%	0%	0%	0%	1%	0%	-1%	0%	0%	-1%	0%	-1%	0%	-1%	0%	2%
Food & Beverages		13%	-1%	6%	-1%	-1%	3%	0%	1%	1%	2%	0%	0%	1%	1%	0%	1%	1%	1%	0%	2%
Food & Drug Retailing		-18%	-3%	0%	0%	2%	-1%	0%	-2%	-1%	0%	0%	0%	0%	-1%	0%	0%	0%	-1%	0%	3%
Personal & Household Products & Services		-7%	1%	-1%	1%	0%	-2%	1%	1%	-1%	-1%	1%	0%	0%	1%	0%	-1%	-2%	0%	0%	-7%
Energy																					
Energy - Fossil Fuels			-1%	0%	0%	0%	-2%	1%	-1%	0%	1%	1%	0%	0%	0%	-1%	-1%	-1%	-2%	-3%	-7%
Renewable Energy			0%	0%	0%	1%	0%	-1%	0%	1%	0%	-1%	0%	0%	0%	1%	1%	0%	2%	2%	10%
Uranium			2%	0%	0%	0%	1%	0%	0%	-1%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	-3%

Appendix 8 Year-to-year change in the number of firm-year observations per year and industry (final sample)

Appendix 9: Descriptive statistics ESG metrics per TRBC economic sector (final sample)

Number of firm-year observations (N), mean, median, and standard deviation of ESG metrics

	Financials				Industrials			
	N	**mean**	**median**	**sd**	**N**	**mean**	**median**	**sd**
ESG score	12,775	41.58	39.30	19.34	12,517	41.39	39.93	19.92
Controversy score	12,775	40.28	38.63	18.09	12,517	40.29	38.80	19.17
Combined ESG/controversy score	12,765	92.92	100.00	20.35	12,504	92.64	100.00	20.60
Environmental pillar score	12,765	33.21	24.40	28.36	12,504	34.98	31.83	27.68
Social pillar score	12,765	42.11	39.60	21.89	12,504	41.21	38.83	23.24
Governance pillar score	12,775	48.30	49.08	23.06	12,517	47.80	48.15	22.10
Emissions	12,765	29.45	15.00	33.04	12,504	39.05	35.93	32.93
Innovation	12,765	19.23	0.00	29.82	12,504	25.26	3.44	30.96
Resource use	12,765	27.07	6.58	32.62	12,504	38.41	34.86	32.95
Community	12,765	49.71	50.00	28.83	12,504	49.63	50.00	28.94
Human rights	12,765	15.79	0.00	28.28	12,504	25.75	4.39	32.27
Product responsibility	12,765	38.59	34.52	29.55	12,504	41.38	37.40	31.63
Workforce	12,765	50.01	49.41	28.71	12,504	50.16	50.20	28.82
CSR strategy	12,775	26.08	6.25	32.36	12,517	32.99	26.73	31.84
Management	12,775	52.32	53.13	29.45	12,517	49.94	50.34	28.29
Shareholders	12,775	49.70	50.00	28.41	12,517	50.56	50.75	28.64

	Consumer Cyclicals				Technology			
	N	**mean**	**median**	**sd**	**N**	**mean**	**median**	**sd**
ESG score	11,932	40.68	38.58	20.00	10,200	41.47	39.06	20.30
Controversy score	11,932	39.39	37.65	18.92	10,200	40.02	38.05	19.10
Combined ESG/controversy score	11,914	91.76	100.00	21.67	10,197	90.91	100.00	22.69
Environmental pillar score	11,914	31.65	26.29	28.68	10,193	28.63	21.53	28.05
Social pillar score	11,912	41.17	38.16	23.28	10,193	42.73	40.84	23.33
Governance pillar score	11,932	45.70	45.81	21.67	10,196	47.85	48.33	22.95
Emissions	11,914	34.77	27.84	33.40	10,193	32.53	22.92	33.52
Innovation	11,914	20.02	0.00	29.83	10,193	22.23	0.00	29.28
Resource use	11,914	35.44	29.41	33.34	10,193	33.35	24.77	33.44
Community	11,912	49.92	50.00	28.86	10,197	50.18	50.00	28.85
Human rights	11,912	25.57	3.33	32.31	10,193	25.30	2.00	32.21
Product responsibility	11,912	38.75	32.98	32.05	10,193	42.81	35.51	30.33
Workforce	11,912	50.21	50.00	28.81	10,197	50.21	50.00	28.85
CSR strategy	11,932	27.27	14.16	30.60	10,200	27.77	9.69	33.06
Management	11,932	48.92	48.82	27.76	10,200	50.89	51.47	28.44
Shareholders	11,932	47.25	46.62	28.85	10,200	51.06	50.89	29.38

	Basic Materials				Healthcare			
	N	**mean**	**median**	**sd**	**N**	**mean**	**median**	**sd**
ESG score	8,466	42.03	41.57	21.92	6,835	36.96	32.55	19.63
Controversy score	8,466	40.61	40.38	20.71	6,835	35.61	32.20	17.99
Combined ESG/controversy score	8,458	92.29	100.00	21.00	6,827	91.81	100.00	21.68
Environmental pillar score	8,458	39.26	38.84	27.80	6,827	19.12	0.00	26.14
Social pillar score	8,458	39.64	35.73	24.51	6,827	41.50	37.86	22.94
Governance pillar score	8,466	50.25	50.42	23.06	6,835	43.21	41.10	22.59
Emissions	8,458	42.83	41.57	32.10	6,827	21.85	0.00	31.26
Innovation	8,458	22.34	0.00	29.41	6,827	5.46	0.00	16.34
Resource use	8,458	41.36	39.58	32.28	6,827	22.87	0.00	31.31
Community	8,458	49.83	49.72	28.83	6,827	50.14	50.95	28.05
Human rights	8,458	26.90	6.93	32.61	6,827	16.57	0.00	28.75
Product responsibility	8,458	37.85	29.33	32.45	6,827	41.94	35.25	30.90
Workforce	8,458	50.37	50.52	28.70	6,827	50.16	50.17	28.89
CSR strategy	8,466	43.16	42.56	34.18	6,835	18.32	0.00	28.68
Management	8,466	50.70	50.54	28.85	6,835	46.69	44.38	28.74
Shareholders	8,466	53.48	55.13	28.05	6,835	48.20	46.67	28.29

Appendix 9 Descriptive statistics ESG metrics per TRBC economic sector (final sample)

Appendix 9: Descriptive statistics ESG metrics per TRBC economic sector (final sample – cont.)

Number of firm-year observations (N), mean, median, and standard deviation of ESG metrics

	Consumer Non-Cyclicals				Energy			
	N	**mean**	**median**	**sd**	**N**	**mean**	**median**	**sd**
ESG score	6,051	43.08	41.77	22.59	5,811	38.06	34.61	21.15
Controversy score	6,051	41.16	40.24	21.02	5,811	36.60	34.00	19.62
Combined ESG/controversy score	6,043	89.58	100.00	23.97	5,770	91.47	100.00	22.06
Environmental pillar score	6,042	37.52	35.62	29.63	5,770	31.65	26.88	27.27
Social pillar score	6,042	42.51	40.38	25.10	5,770	37.80	32.20	23.68
Governance pillar score	6,050	49.98	50.49	23.29	5,811	48.18	47.33	23.59
Emissions	6,042	39.74	36.99	32.72	5,770	39.38	36.36	32.77
Innovation	6,042	24.66	0.00	31.36	5,662	12.21	0.00	24.26
Resource use	6,042	39.87	37.44	32.73	5,770	35.29	29.32	33.30
Community	6,043	50.07	50.00	28.76	5,770	50.05	50.00	28.72
Human rights	6,042	28.35	11.11	33.05	5,770	20.18	0.00	30.59
Product responsibility	6,042	43.63	42.95	31.59	5,770	35.35	27.52	31.98
Workforce	6,043	50.68	50.71	28.89	5,770	50.06	50.00	28.75
CSR strategy	6,051	39.05	36.84	34.35	5,811	35.25	27.72	34.09
Management	6,051	51.87	52.53	28.55	5,811	49.61	49.32	30.01
Shareholders	6,051	51.08	51.43	28.92	5,811	52.05	52.01	28.39

	Real estate				Utilities			
	N	**mean**	**median**	**sd**	**N**	**mean**	**median**	**sd**
ESG score	5,598	39.26	36.39	19.94	3,379	44.74	44.07	20.10
Controversy score	5,598	39.14	36.26	19.87	3,379	43.46	42.82	19.32
Combined ESG/controversy score	5,605	98.67	100.00	9.08	3,375	91.96	100.00	21.25
Environmental pillar score	5,584	30.57	23.09	29.52	3,375	41.67	40.66	26.52
Social pillar score	5,584	43.77	42.61	21.04	3,375	42.88	40.24	23.07
Governance pillar score	5,584	44.47	44.46	22.24	3,379	52.53	53.83	21.97
Emissions	5,584	31.86	21.17	33.28	3,375	46.72	46.50	30.56
Innovation	5,594	21.77	0.00	30.27	3,375	33.31	27.85	33.10
Resource use	5,584	32.57	22.98	33.35	3,375	43.77	42.89	31.73
Community	5,605	49.59	47.94	28.63	3,375	49.92	50.00	28.92
Human rights	5,584	13.70	0.00	26.90	3,375	23.91	0.00	31.80
Product responsibility	5,584	33.31	36.67	30.28	3,375	40.19	37.80	32.36
Workforce	5,605	50.30	50.30	28.78	3,375	50.35	50.40	28.83
CSR strategy	5,605	25.81	7.65	31.48	3,379	48.98	53.51	32.11
Management	5,605	46.70	46.21	29.10	3,379	54.01	55.38	28.22
Shareholders	5,605	49.16	48.75	28.63	3,379	49.98	50.51	29.38

Appendix 9 Descriptive statistics ESG metrics per TRBC economic sector (final sample – cont.)

Appendix 10: Anti-Competition Controversies of "Alphabet Inc." (2020)

The following table shows the seven controversies of "Alphabet Inc." (US02079K3059) in 2020 referring to the planned acquisition of "Fitbit" within the data point "Anti-Competition" (source: Refinitiv Eikon).

Date	Title
05/13/2020	EU consumer group warns against 'game-changer' Google-**Fitbit deal**
06/18/2020	Australian regulator says Google's $2.1 bln **Fitbit deal** could harm competition
07/09/2020	Google's **Fitbit deal** will face EU antitrust probe if no concessions
07/24/2020	Seven Democratic senators urge caution on Google's purchase of **Fitbit**
07/30/2020	Google's $2.1 bln **Fitbit deal** faces EU antitrust probe
08/04/2020	Google's $2.1 bln **Fitbit deal** hits roadblock as EU opens probe
10/19/2020	Japan can open probe into any merger involving **Fitbit** - new antitrust watchdog head

Appendix 10 Anti-Competition controversies of "Alphabet Inc." (2020)

Appendix 11: Business Ethics Controversies of "Volkswagen AG" (2017)

The following table shows the controversies of "Volkswagen AG" (DE0007664039) in 2017 according to the respective data point in the Refinitiv ESG data universe (source: Refinitiv Eikon).

Date	Title
01/19/2017	FBI arrests Volkswagen exec on fraud charges
01/09/2017	VW managers authorised concealment of **diesel cheating**
01/11/2017	Volkswagen confirms $4.3 billion U.S. settlement over **diesel emissions**
01/26/2017	VW set to plead guilty in U.S. **diesel emissions case**
01/27/2017	**German prosecutors open fraud inquiry into former VW CEO**
02/01/2017	VW, Robert Bosch agree to pay $1.55 billion to settle U.S. **diesel** claims
02/20/2017	VW says it did not mislead UK customers on **diesel emissions**
02/22/2017	**German lawmakers call for inquiry into VW compliance chief exit**
03/07/2017	VW expects to sanction more employees in **emissions scandal**
03/10/2017	Volkswagen pleads guilty in U.S. court in diesel **emissions scandal**
03/15/2017	Volkswagen says German prosecutors search its headquarters
03/30/2017	Volkswagen, State AGs Reach More **Emissions-Cheating** Settlements
06/30/2017	VW's French unit reported false delivery figures for years
07/07/2017	Munich prosecutors confirm arrest of Audi employee in **emissions probe**
07/09/2017	VW bosses told costs of **emissions** saga a month before disclosure
07/10/2017	**Stuttgart prosecutor investigates Porsche for fraud, false advertising**
07/20/2017	EU examines **emissions** collusion by German car makers
07/26/2017	Volkswagen Executive to Plead Guilty in **Diesel Emissions** Case
08/01/2017	**EU anti-fraud office send VW probe findings to German prosecutors**
08/07/2017	**German regulator probes VW, Daimler for disclosure violations**
09/28/2017	**EU regulators fine Scania 880 mln euros for truckmakers cartel**
11/14/2017	Volkswagen agrees to pay $69 mln to settle New Jersey **emissions suits**
12/29/2017	Germany's top court rejects VW's bid to suspend **emissions audit**

Appendix 11 Business ethics controversies of "Volkswagen AG" (2017)

Appendix 12: ESG scores and IV: Coefficient β_{ESG} in rolling window panel regressions

The results are based on specification 3 (**Table 19**) and applied to different sub-samples of two- to six-year periods selected from the full sample of firm-year observations.

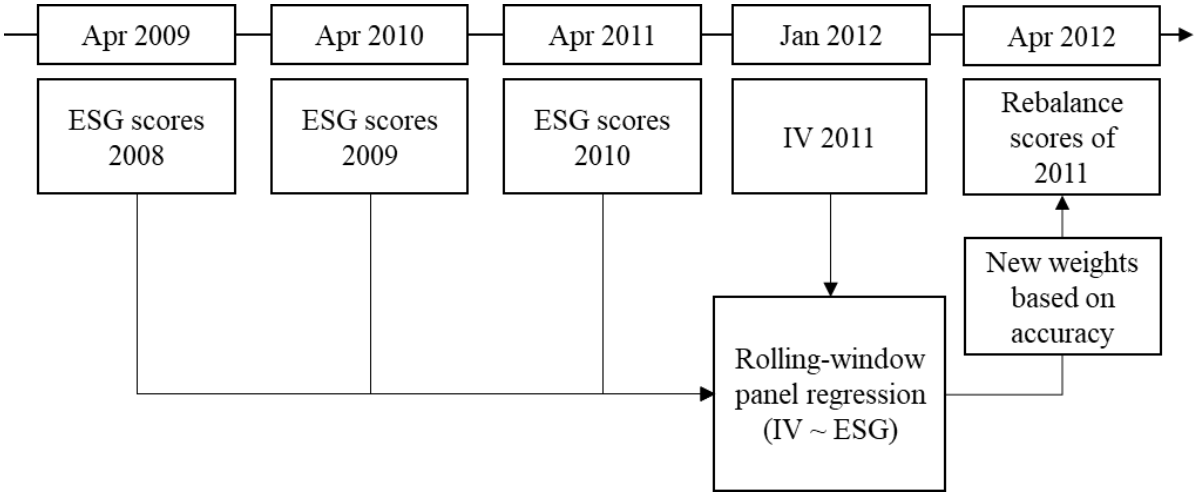
	β_{ESG} coefficient				
Base year	**2-years**	**3-years**	**4-years**	**5-years**	**6-years**
2020	-0.0156	-0.0133	-0.0128	-0.012	-0.0116
2019	-0.0096	-0.0109	-0.01	-0.0098	-0.01
2018	-0.0098	-0.0095	-0.0094	-0.0098	-0.0094
2017	-0.0077	-0.0088	-0.0096	-0.0092	-0.0086
2016	-0.009	-0.0105	-0.0095	-0.0085	-0.0079
2015	-0.0098	-0.0088	-0.0075	-0.0068	-0.0064
2014	-0.0066	-0.006	-0.0057	-0.0054	-0.0054
2013	-0.0041	-0.0049	-0.0047	-0.0049	-0.0058
2012	-0.0032	-0.0046	-0.005	-0.0062	-0.007
2011	-0.0035	-0.0054	-0.007	-0.008	-0.0078
2010	-0.0057	-0.0083	-0.0093	-0.0086	-0.0079
2009	-0.0102	-0.0112	-0.0098	-0.0086	-0.008
2008	-0.0112	-0.0093	-0.0076	-0.0069	
2007	-0.0054	-0.0047	-0.0044		
2006	-0.0038	-0.0036			
2005	-0.0029				

Notes: Results are based on rolling window sub-samples of various lengths with ESG scores as provided by Refinitiv (standardized based on yearly firm size deciles). The dependent variable, which is Idiosyncratic volatility is the yearly standard deviation of the residual risk derived from a Carhart four-factor model. Control variables are: CapEx intensity is capital expenditures divided by total assets. Firm age is the number of years since the firm's IPO. Firm size (in $US million) is the natural logarithm of total assets reported on the balance sheet. IA intensity is the ratio of intangible assets and total assets. Return on assets is net income divided by the book value of total assets. ROA volatility is the standard deviation of ROA over the last three fiscal years. R&D intensity is R&D expenditures divided by total assets. Leverage is the ratio of long- and short-term debt and shareholders' equity. Business sectors, years, and countries are fixed. Values reported in local currencies are converted into USD. Variables are winsorized at the yearly 1%- and 99%-percentiles. All explanatory variables are lagged by one year. Robust test statistics calculated with group-clustered standard errors are reported below coefficients. Significance levels are indicated as: * = 0.1, ** = 0.05, *** = 0.01.

Appendix 12 ESG scores and IV: Coefficient β_{ESG} in rolling window panel regressions

Appendix 13: Timeline of materiality adjustment based on historical accuracy

Timeline and data implemented in rolling window panel regressions to calculate materiality-adjusted ESG scores



Appendix 13 Timeline of materiality adjustment based on historical accuracy