Market Efficiency, Information Processing, and Trading Strategies: Empirical Evidence from ESG Ratings and Cryptocurrency Markets

Faculty of Social Sciences, Economics and Business Administration
University of Bamberg

A thesis presented for the degree of

– doctor rerum politicarum –

Submitted by
Thomas Johannes Schmidl

Primary reviewer: Prof. Dr. Matthias Muck

Secondary reviewer: Prof. Dr. Frank Schiemann

Tertiary reviewer: Prof. Dr. Thomas Egner

Bamberg, 2026

Diese Arbeit hat der Fakultät Sozial- und Wirtschaftswissenschaften der Otto-Friedrich-Universität Bamberg als Dissertation vorgelegen.

Gutachter: Prof. Dr. Matthias Muck

Gutachter: Prof. Dr. Frank Schiemann

Gutachter: Prof. Dr. Thomas Egner

Tag der mündlichen Prüfung: 16.03.2025

URN: urn:nbn:de:bvb:473-irb-115316x

DOI: https://doi.org/10.20378/irb-115316

List of thesis papers

This thesis consists of an introductory section (Chapter 1) and four full-text studies. Study I (Chapter 2) and study IV (Chapter 5) had already been published. Study III (Chapter 4) was under review at the time of the dissertation's publication (version dated 17 May 2025) and has since been published as well. Study II (Chapter 3) is a working paper. The bibliographic details are listed below.

Study I:

Muck, Matthias und Schmidl, Thomas (2024): Comparing ESG score weighting approaches and stock performance differentiation. *Finance Research Letters*.
doi:10.1016/j.frl.2024.105924

Study II:

Schmidl, Thomas (2025): ESG data characteristics and stock price differentiation. *Working Paper*.
doi:10.2139/ssrn.5076561

Study III:

Muck, Matthias und Schmidl, Thomas (2026): Relative to whom? The impact of peer groups on ESG ratings and financial performance. *Journal of Asset Management*.
doi:10.1057/s41260-026-00444-4

Study IV:

Muck, Matthias, Schmidl, Thomas und Wolf, Julian (2025): Wish or reality? On the exploitability of triangular arbitrage in cryptocurrency markets. *Finance Research Letters*.
doi:10.1016/j.frl.2024.106508

Danksagung (Acknowledgements)

Das Anfertigen einer Dissertation ist immer mit Höhen und Tiefen verbunden. Auch wenn die Fertigstellung einer solchen Arbeit letztendlich eine individuelle Leistung darstellt, haben mich viele Menschen auf diesem Weg begleitet, unterstützt und mit mir gelitten. Ihnen allen möchte ich an dieser Stelle danken, denn sie haben maßgeblich zum Gelingen dieser Arbeit beigetragen.

Ein besonderer Dank gilt meinem Promotionsbetreuer Prof. Dr. Matthias Muck, der mich seit Beginn der Promotion mit großem Engagement begleitet und gefördert hat. Besonders dankbar bin ich ihm dafür, dass er mir die akademische Freiheit ließ, eigene Ideen zu entwickeln und zu verfolgen. Unsere akademischen Diskussionen waren nicht immer einfach, aber die gemeinsamen Publikationen zeigen, dass sie letztlich stets zu erfolgreichen Ergebnissen geführt haben.

Die Zeit am Lehrstuhl für Banking und Finanzcontrolling wird für mich unvergesslich bleiben. Dafür möchte ich mich ganz besonders bei meinen ehemaligen Kolleginnen und Kollegen bedanken, die mich während dieses langen Weges begleitet haben. Insbesondere gilt dieser Dank Pascal Albert, Michael Herold, Tobias Herzing, Tobias Kaufmann, Julia Lang und Gerlinde Weinhäupl. Ihr habt den Alltag am Lehrstuhl, ebenso wie die Zeit außerhalb davon - meistens im Schwarzen Schaf - besonders gemacht.

Ein letzter großer Dank gilt meiner Familie. Meiner zukünftigen Ehefrau Ann-Kathrin Abt, die mich immer bedingungslos unterstützt und motiviert hat. Meinen Eltern Marie Luise und Franz Schmidl, die mir jederzeit mit Zuversicht und Vertrauen zur Seite standen. Und meinem Bruder Andreas Schmidl, der stets ein offenes Ohr für mich hatte, mich kontinuierlich unterstützt hat und vor derselben akademischen Herausforderung steht. Ich hoffe, dass ich in gleicher Weise dazu beitragen kann, dass auch seine Arbeit von ähnlichem Erfolg geprägt sein wird.

März, 2026 Thomas Schmidl

Contents

List of figures **xi**

List of tables **xiii**

1 Introduction **1**

1.1 Introduction . . . 1

1.2 Sustainable finance . . . 3

1.3 Digital finance . . . 6

1.4 Overview of the chapters . . . 8

1.5 Summary of the chapters . . . 9

2 Comparing ESG score weighting approaches and stock performance differentiation **21**

2.1 Introduction . . . 22

2.2 Related literature and research question . . . 24

2.3 Data and methodology . . . 26

2.4 Results . . . 28

2.5 Robustness . . . 31

2.6 Conclusion . . . 36

2.7 Appendix . . . 38

3 ESG data characteristics and stock price differentiation **51**

3.1 Introduction . . . 52

3.2 Related literature . . . 54

3.3 Data and methodology . . . 55

3.4 Results 57
3.5 Robustness 60
3.6 Conclusion 63
3.7 Appendix 65

4 Relative to whom? The impact of peer groups on ESG ratings and financial performance 77
4.1 Introduction 78
4.2 Related literature 80
4.2.1 Sustainability and financial returns 81
4.2.2 The market of ESG ratings 81
4.2.3 Methodological challenges in ESG ratings 82
4.3 The Refinitiv rating 83
4.3.1 Rating categories 84
4.3.2 Calculation 85
4.3.3 Impact of the peer group 87
4.3.4 Differences in ESG ratings 91
4.4 Methodology and data 93
4.4.1 Methodology 93
4.4.2 Data 95
4.5 Results 98
4.5.1 Fit of recalculated best-in-industry ratings 98
4.5.2 ESG ratings and financial returns 99
4.5.3 Alternative model specifications 105
4.6 Robustness 108
4.7 Conclusion 111
4.8 Appendix 113
4.8.1 Numerical scores: firms with identical ratings and the average score . 113
4.8.2 ESG score estimate 113
4.8.3 Additional ESG factor 113

5 Wish or reality? On the exploitability of triangular arbitrage in cryptocurrency markets 121

5.1 Introduction . . . 122

5.2 Related literature . . . 124

5.3 Methodology and data . . . 126

5.4 Results . . . 127

5.5 Conclusion . . . 136

6 Conclusion 143

List of figures

2.1 Histogram ESG scores 40

3.1 Distribution of ESG scores 72

4.1 Scatter plot ESG scores 117

5.1 Yield curve within an arbitrage opportunity 129

5.2 Exchange rate product over time 130

List of tables

1.1 Overview of publications . . . 9

2.1 Descriptive statistics: ESG scores and ESG pillars . . . 29
2.2 Descriptive statistics: ESG scores by industry . . . 30
2.3 Regression results: performance differences . . . 32
2.4 Regression results: Paris Agreement . . . 34
2.5 Regression results: robustness tests . . . 35
2.6 Proxy data points for industry magnitudes . . . 41
2.7 ESG category assignment . . . 42
2.8 SASB industry group weightings . . . 43
2.8 SASB industry group weightings - continued . . . 44

3.1 Descriptive statistics: ESG scores and ESG pillars . . . 58
3.2 Descriptive statistics: ESG scores by industry . . . 59
3.3 Regression results: performance differences . . . 61
3.4 Robustness tests . . . 62
3.5 Numeric calculation example 1 . . . 66
3.6 Numerical calculation example 2 . . . 67
3.7 Numerical calculation example 3 . . . 67
3.8 Numerical calculation example 4 . . . 68
3.9 Boolean calculation example 1 . . . 68
3.10 Boolean calculation example 2 . . . 69
3.11 Boolean calculation example 3 . . . 69
3.12 Boolean calculation example 4 . . . 70

4.1 ESG score range and grade description 86
4.2 Descriptive statistics 96
4.3 Change in ratings 97
4.4 Validation of recalculated ESG category scores 100
4.5 Regression results of different ESG scores 101
4.6 Regression results of rating strategy 103
4.7 Additional tests 106
4.8 Robustness tests 109
4.9 TRBC sector and industry classification 115
4.10 Regression results for the original Refinitiv score 116

5.1 Tradable quantities for an exemplary triangular arbitrage opportunity 128
5.2 Number of profitable arbitrage opportunities by Binance VIP level 131
5.3 Number of arbitrage opportunities by return, duration, and value 133
5.4 Number of profitable arbitrage opportunities by return and value 134
5.5 Quantities by trading pairs 135

Chapter 1

Introduction

1.1 Introduction

The primary function of financial markets is to allocate capital. Based on the central assumption that every action by market participants is intended to generate profit, prices of the underlying assets have an informative value. As a consequence, prices reflect the aggregated information from individuals' and institutions' decisions based on their future expectations (Goldstein, 2023). This initial idea of Hayek (1945), that prices serve as a communication mechanism by acting as a source of information about goods and services, is retrieved in the efficient market hypothesis (Fama, 1970). It states that all financial market participants have the same information and that prices fully reflect all available information.[1] Since they are processed instantaneously, the rational expectations lead to the result that investors are not able to obtain excess returns, because price changes are completely unpredictable and not the result of better informed market participants.

This conclusion of the efficient market hypothesis only applies if no trading or information processing costs are incurred (Grossman and Stiglitz, 1980). In practice, the economy, with its dynamically changing framework conditions, constantly tests the efficiency of financial

[1] Depending on the degree to which different types of information are processed, a distinction is made between three forms. The weak form of the efficient market hypothesis states that prices reflect past prices, while the semi-strong form claims that prices additionally include publicly available information such as company disclosures. The strongest form emphasizes that information that is known to everyone is priced. For example, this also includes insider information which is known to operating managers (Fama, 1970).

markets. Under real world circumstances, market frictions such as transaction costs and irrational participants violate the theorem, which leads to an incomplete pricing mechanism. Empirical finance literature identifies such anomalies and shows that abnormal returns are possible. For example, they find earnings (Basu, 1977), size (Banz, 1981), or momentum (Jegadeesh and Titman, 1993) patterns that cause excess returns. In addition, the observed volatility of stock prices does not adequately reflect the degree of new information about dividends (Shiller, 1981), which illustrates an overreaction of financial market participants (De Bondt and Thaler, 1985). Consequently, the desire for full information processing in asset prices appears to be an unattainable goal.

Nevertheless, it must be noted that financial markets are constantly developing and are close to frictionless markets. The increasing levels of liquidity, more sophisticated market participants, and higher amounts of available information have led to a continuous improvement of price quality over time (Goldstein, 2023). This observed development is also accelerated by the introduction of technologies and regulatory requirements. In this context, this dissertation examines two important drivers to enhance the information processing of financial market participants and, consequently, increase the degree of market efficiency: sustainable and digital finance.

The first mentioned addresses the trade-off between short-termism and long-term perspectives. Sustainability is acknowledged as an important driver of long-term development to enhance the resilience of societies. In particular, sustainable finance focuses on the environmental and social risks by including future costs (e.g., air pollution and climate change) in the investment process. It is also the result of changing investor preferences that penalises low sustainable engagement with higher risk premiums (Heeb et al., 2023). In contrast, corporations with superior sustainable profiles benefit from reduced cost of capital (El Ghoul et al., 2011). In combination with increasing levels of regulatory and disclosure requirements, this contributes to more sophisticated and forward-looking decisions by market participants. Consequently, this could lead to an enhanced pricing mechanism in financial markets.

Another important driver of improved decision-making and market efficiency is the rising degree of digitalisation across all sectors. In the context of financial markets, this means that the activity of collecting, processing, and analysing large sets of financial data becomes more efficient by using digital technologies. This enhances the ability of investors or financial institutions to make more sophisticated decisions. Moreover, digitalisation could reduce entry barriers, which leads to potentially lower transaction costs. For example, the introduction of blockchain technology reduces the need for intermediaries since the decentralized information about property rights and transactions are fully traceable and cannot be manipulated. Besides technological and efficiency improvements, this trend creates new technical opportunities for the functionality of the financial system.

1.2 Sustainable finance

According to Friedman (1970), the responsibilities of a corporation are to efficiently use its resources to increase profits, assets, and firm value. The only constraints for this objective are that corporations engage in free competition and do not conduct fraud or crime. Hence, if managers as representatives of the corporations use resources to pursue ecological or social objectives that are not compatible with the goal of maximising profits, they misallocate the owners' capital.

This normative view of a corporation, known as shareholder theory, states that common wealth is created if the corporations maximise value for their shareholders. Freeman (1984) broadened the shareholder value concept by developing the stakeholder theory to emphasize additional interactions with other interest groups. The managers should align their decisions with the interests of employees, customers, suppliers, the environment, and the government. If implemented perfectly, this enables companies to enhance their reputation, attract a talented workforce or reduce legal risks, while generating a competitive advantage that leads to long-term value.[2] However, this approach stands in contrast to the former, as it drifts away from the pure focus of profit maximization by incorporating other stakeholders which directly or indirectly affect the success of the corporation. This simultaneous fulfilment of several objectives increases the complexity of the decision-making process of managers (Jensen, 2001).

[2] For a little overview of how stakeholders could increase firm value and performance, see for example, Deng et al. (2013), Harrison and Wicks (2013) and Tantalo and Priem (2016).

Some key aspects of the stakeholder value approach also play a central role in the foundation of sustainable finance. In theory, it suggests that financial market participants should incorporate environmental and social risks into their decision-making process to include external costs and reduce potential information asymmetries. However, the term sustainability in the financial context is not clearly defined. In the 18th century, von Carlowitz et al. (1732) are the first who provided a definition. Their definition originates from the forestry business and emphasises the idea that at least as many trees should be replanted as are cut down to preserve the long-term productivity of the forest. In the following years, the Brundtland (1987) report describes sustainability as an overarching goal for the three dimensions ecology, economy, and social to preserve the living conditions for future generations. In the following, the normative term CSR (corporate social responsibility) and the operationalised term ESG (environmental, social, and governance) reflect the sustainability engagement from both corporate and investment perspectives.

In particular, the importance of sustainability gained significant attention following the Paris Agreement of the United Nations (Abdel-Karim and Kollmer, 2022). The signatories identified global warming and the goal of limiting temperature increase as a key challenge for future generations. In addition, the allocation of financial resources should contribute to a more climate-resilient development. In the following years governments introduced several policies such as the E.U. Taxonomy which defines six major environmental objectives, to increase the sustainability engagement and transparency requirements of corporations. Especially financial institutions (banks and insurance companies) play a central role in this transmission. The United Nations report "Who cares wins" emphasised that including ESG issues in corporations' decision-making processes could result in competitive advantages (UNGC, 2004). This means that corporations committed to sustainability are more likely to attract investors (Hartzmark and Sussman, 2019) and, consequently, receive financial funding on better terms (El Ghoul et al., 2011). In turn, not investing in ESG as prevention, could result in potential future risks for corporations. Moody's (2021) commonly states three main fields of ESG-related risks: regulatory and legal risks, reputational risks, and long-term risks. For example, reputational risks in the form of decreasing sales or regulatory and legal risks in the form of higher costs have direct impact on the profitability of a company. In comparison, long-term risks, such as global warming or the rising sea levels, are more difficult to quantify.

Since those sustainable risk exposures are relevant in terms of future cash flows or profit expectations, financial market participants should price this information. Hence, academics and practitioners deal with the crucial question how ESG issues affect the financial performance of a corporation. In this field, meta-analyses find a general positive relationship (Friede et al., 2015; Whelan et al., 2021), which results in a reduced risk assessment of more sustainable firms (Albuquerque et al., 2019). ESG activities reduce the cost of equity (El Ghoul et al., 2011) and corporations with high ESG scores benefit from lower cost of debt (Apergis et al., 2022). Moreover, sustainable stocks are used as financial hedge due to their resiliency during times of high volatility (Lins et al., 2017; Albuquerque et al., 2020; Broadstock et al., 2021). Those findings are confirmed in different settings (e.g., Hübel and Scholz, 2020) and exhibit that corporations with poor ESG profiles receive a sustainability risk premium that compensates for the higher inherent ESG risk, while their sustainable counterparts benefit from reduced return expectations from investors.

Through ongoing efforts of voluntary publication obligations, corporations increasingly disclose sustainable information for the capital market. In principle, this leads to the reduction of information asymmetries. Researchers use individual parts of this data in their examinations for the construction of different proxies for environmental (Bolton and Kacperczyk, 2021) and social performance (e.g., Hong and Kacperczyk, 2009; Edmans, 2011). Moreover, information intermediaries such as rating agencies collect, process, analyse, and interpret large sets of sustainability data to make it available in the form of an easily interpretable measure. Since a growing number of investors includes those ESG ratings or scores in their investment strategies, they have a serious impact on capital flows (Hartzmark and Sussman, 2019).

However, this measurement is part of a problem. Considering the consequences, building a sustainable metric is a challenging task. It includes the selection of indicators and an adequate weighting procedure to result in an aggregated measure (Gan et al., 2017). Hence, assessing the exact value of sustainability for a corporation is difficult. The lack of globally accepted regulation and supervision (ESMA, 2021) leads to a diverse market of ESG ratings (Berg et al., 2022). In addition, much reporting is voluntary and prone to selective disclosure practices. This leads to a situation where companies are just disclosing favourable information and could withhold the rest.

This dissertation strives to examine the function of ESG ratings and their ability to reflect the inherent sustainability risk exposure of a corporation. ESG ratings or scores are the result of collected and processed sustainability data and should act as a metric that guides investors through their investment decision. However, the assessment of rating providers is diverse, as are their methodologies. Consequently, recent literature finds a disagreement among rating agencies (Berg et al., 2022). This means that there is still a need to improve existing metrics. The thesis gives insights into whether the weighting approach, the peer groups, or the characteristics of indicators contribute to the development of more accurate measures. Since the semi-strong form of the efficient market hypothesis implies that all sustainability information is reflected in the pricing of assets, the validation takes place with the help of risk premiums. As a result, the dissertation gives implications for the enhancement of ESG ratings.

1.3 Digital finance

The impact of new information technologies on the financial sector is diverse. Digitalisation in finance aims to process real-time information more quickly and efficiently, (trading) tools are becoming increasingly sophisticated, and the use of machine learning and artificial intelligence could reveal systematic patterns in large data sets. In addition to the development of new business fields (e.g., FinTechs), digitalisation contributes to a steady reduction in information asymmetries in the financial system.

From a technical perspective, the introduction of new technologies in financial markets enhances the speed and efficiency of information processing and trade execution. It becomes possible to analyse larger amounts of data in less time in order to obtain an informational advantage over other market participants. For example, on the major stock exchanges, electronic trading has replaced the traditional human market makers through automated systems such as high-frequency algorithms. According to Brogaard et al. (2014), these high-frequency traders represent a new class of intermediaries, which are mainly characterised by their increasing use of technology. This evolution towards automated trading systems has led to diminishing risk-free arbitrage opportunities across different financial market segments over time (e.g., Ito et al., 2020). Since arbitrage-free trading is a key condition for market efficiency, the increasing use of new technologies plays a beneficial role in the price discovery process.

Besides the rise in efficiency, the new tools or access points of digital finance offer opportunities in the form of lower market barriers and decreasing costs. The technological progress mitigates transaction and information processing costs, which consequently leads to broader market access. In addition, this also appeals to the group of digital natives, who continue to drive the disruptive introduction of innovations. For example, the digitalisation of services allows newcomers without large amounts of financing funds to enter markets and compete with established corporations. The increasing competition aims to foster more efficient pricing behaviour.

The use of blockchain technology for electronic cash transactions represents one of the major innovations in recent years. The most popular example is Bitcoin. Nakamoto (2008) developed a peer-to-peer version for transactions as a counterpart to the conventional payment system. In this case, the hash-based proof-of-work concept ensures the integrity of every individual transaction within the information chain. Each new transaction is attached to the data block and encrypted in such a way that any subsequent manipulation is immediately apparent. Decentralized data storage, transparency and non-manipulability are therefore the key advantages of using blockchain as a means of payment (Bundesamt für Sicherheit in der Informationstechnik, 2025). Decentralized organizations such as the Uniswap exchange use liquidity pools for cryptocurrency trading, which reduces the need for market makers. This may be seen as an innovation with large disruptive potential, since traditional intermediaries as central counterparties for financial transactions are not needed. Taking this idea further, these decentralized autonomous systems can contribute to a democratisation of finance. This means that financial market participants are independent of banks and have easier access to the financial system.

However, the efficiency gains achieved through innovations are not without negative consequences. Increasing complexity could make the financial system more vulnerable, especially due to externalities. Flash crashes triggered by algorithmic trading errors, fraud, or the excessive energy consumption (associated with certain blockchain applications) are just some potential risks. Furthermore, data ownership is becoming a critical issue, since the control over large datasets may enable a small number of market participants to consolidate market dominance. In this context, academic research plays a crucial role in assessing these trade-offs and providing evidence-based guidance for future developments in the financial markets.

This thesis attempts to examine whether the market for cryptocurrencies exhibits a high degree of market efficiency. The introduction of new cryptocurrency exchanges offers an opportunity to empirically test for the existence of arbitrage opportunities in a new, developing market segment. In this setting, the advanced technological infrastructure within such exchanges is contrasted with a still young, largely unregulated and immature market environment. Sophisticated trading tools in combination with technology enthusiasts should lead to a relatively high level of efficiency in the early stage of market development. However, empirical literature regarding efficiency in cryptocurrency markets remains mixed (e.g., Vidal-Tomás and Ibañez, 2018; Burggraf and Rudolf, 2021; Aloosh et al., 2022). By implementing an automated high-frequency arbitrage strategy on a cryptocurrency exchange, the thesis aims to contribute practical insights into market efficiency.

1.4 Overview of the chapters

The remainder of the dissertation is structured as follows: The following is a brief summary of the individual chapters with respect to research questions, contributions, data, and main results. Table 1.1 provides an overview of the four included studies of the dissertation and their current publication status. Chapters 2 to 4 present the research studies on ESG ratings and excess returns, while Chapter 5 contains a research study on triangular arbitrage in cryptocurrency markets. Chapter 6 concludes.[3]

[3]The formatting of the following studies has been adapted to meet the requirements of this dissertation. Furthermore, the appendices contain additional result tables or supplementary material that is not part of the published studies.

Table 1.1: Overview of publications

Chapter	Title	Publication	
		Status	Journal
2	Comparing ESG score weighting approaches and stock performance differentiation	Published	Finance Research Letters
3	ESG data characteristics and stock price differentiation	-	-
4	Relative to whom? The impact of peer groups on ESG ratings and financial performance	Published	Journal of Asset Management
5	Wish or reality? On the exploitability of triangular arbitrage in cryptocurrency markets	Published	Finance Research Letters

1.5 Summary of the chapters

Comparing ESG score weighting approaches and stock performance differentiation

Material sustainability risks vary substantially across industries (Eccles et al., 2012). As Khan et al. (2016) mention, most empirical studies on the relationship between sustainability and financial performance have not accounted for the importance of different ESG risks across industries. Simultaneously, rating providers try to diminish this shortcoming through initial weighting approaches. For example, Refinitiv, as one of the major rating agencies, assesses these industry-specific sustainability risks using a data-driven methodology. However, this approach is completely independent of expected future profits or cash flows.

This study tests whether the intra-rating weighting in the Refinitiv ESG scores leads to better risk and return assessments. To do this, two alternative scores are constructed as benchmarks to compare their effect on the financial market with the original score. The first score treats ESG categories equally. Such a naively unweighted approach should be selected if the financial significance of the sustainability information is completely unknown. The second score extends the Refinitiv data with SASB's materiality assessment for the different industries. In this case, the weighting is derived from expert assessments.

The study analyses 3,098 U.S. firms with an ESG rating during the period from 2010 to 2022. Contributing to previous research, this study validates whether the ESG score has predictive power over future financial performance. As a result, the Refinitiv scores could differentiate risk premiums but lead to a smaller performance difference between sustainable and unsustainable portfolios than an uninformed score with equal weighting. This shows that the financial relevance of the ESG score is not improved by the data-driven approach. Consequently, investment strategies based on a naively unweighted score would achieve on average higher returns. Comparing the results with SASB's materiality approach for the same period, this materiality-weighted score leads to the greatest differences in returns between the sustainable and unsustainable portfolios. Hence, the SASB assessment is suitable for better differentiation. In any case, this provides a further indication that not taking into account the relationship between future earnings and sustainability information significantly limits the suitability of an ESG rating for predicting financial performance.

The results of this study draw attention to the challenging task of determining the potential ESG risk exposures of different industry groups through intra-rating weighting. In doing so, Refinitiv's data-driven approach is tempting since it mitigates potentially subjective interpretations of sustainability information. However, the results show that this method is less precise in predicting than, for example, using a rating that naively weights the information equally. A potential solution to this problem is an index that extends Refinitiv's ESG rating by including SASB's materiality assessment. As a result, this rating could provide the best forecast for financial returns.

ESG data characteristics and stock price differentiation

The task of rating agencies is to collect, analyse and interpret sustainability information (Berg et al., 2022). Besides NGO and company websites, the main source to fulfil this function is firms' disclosures such as annual and CSR reports. However, this task is difficult since most of the sustainability data is still emerging, voluntary, and not standardised. To display relevant ESG issues for financial market participants adequately, ESG rating providers present them only as binary data points. Whether a company is engaged in a specific theme has informational value but does not say much about the way it is implemented.

Most rating providers combine those binary indicators and numerical data for developing their ESG company ratings. Nevertheless, the binary representation obscures information and, consequently, makes it difficult to differentiate between companies as this merely creates two comparison groups: corporations that are engaged in this theme and their counterparts that are not active in this particular sustainability category. Against the background of the weaknesses of binary data, the question arises whether these reflect reliable signals to capital market participants. The study examines a sample of 23,551 firm-year observations by constructing two separate ESG scores from the original Refinitiv database (one numerical and one binary score).

The results of the study indicate that numerical and binary measures could capture inherent sustainability risk as reflected in stock returns. Firms with low ESG scores exhibit higher risk premiums compared to stocks with higher ratings. This differentiation is more pronounced for the binary ESG score. Using alternative factor models, time periods, and portfolio rebalancing methodologies as robustness tests, the results show a larger difference in unexpected returns for the binary score. This result is particularly noticeable if the extraordinary effects of the COVID-19 crisis are excluded. These results confirm earlier empirical evidence that investors are guided by simplified attributes when deciding in favour of sustainable financial instruments.

Relative to whom? The impact of peer groups on ESG ratings and financial performance

Simultaneously having a superior ESG rating from one agency and a neutral rating from another is not uncommon, since there is an ongoing disagreement among rating providers (Chatterji et al., 2016; Berg et al., 2022). Moreover, there are discrepancies within providers' rating methodologies, which lead to confusing inconsistencies in their assessments. For example, a worse ESG score of a solar firm in comparison to a superior score of a coal producer seems counterintuitive. The reason for that is the best-in-industry rating approach. Using this methodology, the rating agency collects firms' sustainability data and compares it with a peer group of the same industry. In other words, it is enough that a firm's environmental and social performance only stands out within its industry group to receive a superior rating. On the one hand, this approach has the advantage that it includes every type of company and thus creates general incentives to improve. On the other hand, environmentally or socially harmful corporations seem to be more sustainable through their ratings than their engaged counterparts. As ESG ratings allocate high amounts of financial funds (Hartzmark and Sussman, 2019), the question arises whether such peer group ESG measures provide valuable information for financial market participants' investment decisions.

This study investigates to what extent the industry-based ratings of a firm reflect the risk-return signal under the assumption that the information content of the ESG ratings is reflected in stock prices (e.g., Fama, 1970; Hübel and Scholz, 2020). By recalculating the ESG scores at the superordinate economic sector level, the study compares the excess returns (measured as Alpha) of portfolios constructed with best-in-industry and best-in-sector ESG scores. Both scores contain financially relevant information since stocks with lower sustainability profiles exhibit higher risk premiums than their sustainable peers. However, the recalculated best-in-sector scores could better discriminate stocks according to their returns. This finding indicates that sustainable measures constructed with broader peer groups offer more financially significant information for market participants.

In addition, the study analyses investment strategies that only allow investing in companies with an ESG rating above a defined threshold. Through the recalculation of the ratings, some companies receive a rating downgrade and, therefore, are no longer investable. The results show that these corporations make a significant contribution to the observed return of the investment strategies. Since fund managers have to meet the return expectations of their investors, they will choose the rating that also allows them to invest in high-yielding firms. Those are mostly in industries that are not commonly considered as sustainable (Hong and Kacperczyk, 2009). Consequently, the best-in-industry approach favours greenwashing suspicions or problems for mutual funds.

Although sustainable investment opportunities have reached a total amount of 30.3 trillion U.S. Dollar in assets under management in 2022 (GSIA, 2022), the different definitions or assessment methodologies still leave too much room for interpretation. This study has practical relevance, as ESG ratings or scores are relevant for financing conditions, managers' salaries, or for inclusion in indices. The findings of the study suggest that there is still room for regulatory improvement to ensure a more straightforward use of ESG ratings by market participants.

Wish or reality? On the exploitability of triangular arbitrage in cryptocurrency markets

Risk-free profits are a primary objective for many financial market participants. But the more efficient and regulated capital markets become, the more difficult it is to achieve this with arbitrage strategies. As a new phenomenon, cryptocurrencies remain largely unregulated, and research states that they tend to be inefficient (e.g., Aloosh et al., 2022) or weakly efficient (e.g., Vidal-Tomás and Ibañez, 2018). As a consequence, traditional trading strategies such as triangular arbitrage could result in risk-free profits on centralized cryptocurrency markets for traders.

Therefore, the study examines a high-frequency data-set with 30.9 million exchange rate pairs from the Binance Exchange. It answers the research question of whether there are arbitrage opportunities in cryptocurrency markets and to what extent retail investors could benefit from them. Implementing a triangular arbitrage trading strategy with Bitcoin, Litecoin, and the U.S. Dollar, the study identifies 4,879 triangular arbitrage opportunities in one week. But only a small proportion of these are profitable for traders, after including real transaction fees. Even if these opportunities are exploited, the resulting profits would be very small due to limited trading volumes in the order book. In particular, the least liquid cross-rate (Bitcoin and Litecoin) limits the quantity and shows similar characteristics as in traditional foreign exchange markets. In addition, since there is a time lag between the observed and realized prices, slippage is a crucial factor when implementing high-frequency trading strategies. By estimating the execution time for a trade, the study provides guidance on how long an arbitrage opportunity should exist to result in a positive net return.

In summary, the study extends research about the efficiency of centralized cryptocurrency markets (e.g., Burggraf and Rudolf, 2021). It includes real-world circumstances and indicates that this relatively new market segment exhibits a high degree of efficiency. Moreover, the sheer number of arbitrage opportunities alone is a weak indicator of market efficiency.

Bibliography

Abdel-Karim, B. M., Kollmer, F. X. (eds.), 2022. Sustainable Finance: Herausforderungen und technologische Lösungen für Banken und Finanzdienstleister. Springer Fachmedien, Wiesbaden.

Albuquerque, R., Koskinen, Y., Yang, S., Zhang, C., 2020. Resiliency of Environmental and Social Stocks: An Analysis of the Exogenous COVID-19 Market Crash. The Review of Corporate Finance Studies 9, 593–621.

Albuquerque, R., Koskinen, Y., Zhang, C., 2019. Corporate Social Responsibility and Firm Risk: Theory and Empirical Evidence. Management Science 65, 4451–4469.

Aloosh, A., Ouzan, S., Shahzad, S. J. H., 2022. Bubbles across Meme Stocks and Cryptocurrencies. Finance Research Letters 49, 103155.

Apergis, N., Poufinas, T., Antonopoulos, A., 2022. ESG scores and cost of debt. Energy Economics 112, 106186.

Banz, R. W., 1981. The relationship between return and market value of common stocks. Journal of Financial Economics 9, 3–18.

Basu, S., 1977. Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis. The Journal of Finance 32, 663–682.

Berg, F., Kölbel, J. F., Rigobon, R., 2022. Aggregate Confusion: The Divergence of ESG Ratings. Review of Finance 26, 1315–1344.

Bolton, P., Kacperczyk, M., 2021. Do investors care about carbon risk? Journal of Financial Economics 142, 517–549.

Broadstock, D. C., Chan, K., Cheng, L. T. W., Wang, X., 2021. The role of ESG performance during times of financial crisis: Evidence from COVID-19 in China. Finance research letters 38, 101716.

Brogaard, J., Hendershott, T., Riordan, R., 2014. High-Frequency Trading and Price Discovery. The Review of Financial Studies 27, 2267–2306.

Brundtland, G. H., 1987. Our Common Future: Report of the World Commission on Environment and Development.

Bundesamt für Sicherheit in der Informationstechnik, 2025. Blockchain macht Daten praktisch unveränderbar. https://www.bsi.bund.de/DE/Themen/Verbraucherinnen-und-Verbraucher/Informationen-und-Empfehlungen/Technologien_sicher_gestalten/Blockchain-Kryptowaehrung/blockchain-kryptowaehrung.html?nn=131544.

Burggraf, T., Rudolf, M., 2021. Cryptocurrencies and the low volatility anomaly. Finance Research Letters 40, 101683.

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

De Bondt, W. F. M., Thaler, R., 1985. Does the Stock Market Overreact? The Journal of Finance 40, 793–805.

Deng, X., Kang, J.-k., Low, B. S., 2013. Corporate social responsibility and stakeholder value maximization: Evidence from mergers. Journal of Financial Economics 110, 87–109.

Eccles, R. G., Krzus, M. P., Rogers, J., Serafeim, G., 2012. The Need for Sector-Specific Materiality and Sustainability Reporting Standards. Journal of Applied Corporate Finance 24, 65–71.

Edmans, A., 2011. Does the stock market fully value intangibles? Employee satisfaction and equity prices. Journal of Financial Economics 101, 621–640.

El Ghoul, S., Guedhami, O., Kwok, C. C. Y., Mishra, D. R., 2011. Does corporate social responsibility affect the cost of capital? Journal of Banking & Finance 35, 2388–2406.

ESMA, 2021. ESMA letter to EC on ESG ratings.

Fama, E. F., 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance 25, 383–417.

Freeman, R. E., 1984. Strategic Management: A Stakeholder Approach. Cambridge University Press.

Friede, G., Busch, T., Bassen, A., 2015. ESG and financial performance: Aggregated evidence from more than 2000 empirical studies. Journal of Sustainable Finance & Investment 5, 210–233.

Friedman, M., 1970. The Social Responsibility of Business is to Increase its Profits. The New York Times Magazine .

Gan, X., Fernandez, I. C., Guo, J., Wilson, M., Zhao, Y., Zhou, B., Wu, J., 2017. When to use what: Methods for weighting and aggregating sustainability indicators. Ecological Indicators 81, 491–502.

Goldstein, I., 2023. Information in Financial Markets and Its Real Effects*. Review of Finance 27, 1–32.

Grossman, S. J., Stiglitz, J. E., 1980. On the Impossibility of Informationally Efficient Markets. The American Economic Review 70, 393–408.

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

Harrison, J. S., Wicks, A. C., 2013. Stakeholder Theory, Value, and Firm Performance. Business Ethics Quarterly 23, 97–124.

Hartzmark, S. M., Sussman, A. B., 2019. Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows. The Journal of Finance 74, 2789–2837.

Hayek, F. A., 1945. The Use of Knowledge in Society. The American Economic Review 35, 519–530.

Heeb, F., Kölbel, J. F., Paetzold, F., Zeisberger, S., 2023. Do Investors Care about Impact? The Review of Financial Studies 36, 1737–1787.

Hong, H., Kacperczyk, M., 2009. The price of sin The effects of social norms on markets. Journal of Financial Economics .

Hübel, B., Scholz, H., 2020. Integrating sustainability risks in asset management: The role of ESG exposures and ESG ratings. Journal of Asset Management 21, 52–69.

Ito, T., Yamada, K., Takayasu, M., Takayasu, H., 2020. Execution Risk and Arbitrage Opportunities in the Foreign Exchange Markets.

Jegadeesh, N., Titman, S., 1993. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. The Journal of Finance 48, 65–91.

Jensen, M., 2001. Value Maximisation, Stakeholder Theory, and the Corporate Objective Function. European Financial Management 7, 297–317.

Khan, M., Serafeim, G., Yoon, A., 2016. Corporate Sustainability: First Evidence on Materiality. The Accounting Review 91, 1697–1724.

Lins, K. V., Servaes, H., Tamayo, A., 2017. Social Capital, Trust, and Firm Performance: The Value of Corporate Social Responsibility during the Financial Crisis. The Journal of Finance 72, 1785–1824.

Moody's, 2021. General Principles for Assessing Environmental, Social and Governance Risks Methodology. https://ma.moodys.com/rs/961-KCJ-308/images/Rating%20Methodologies%20-%20General-Principles-for-Assessing-Environmental-Social-and-Governance-Ri...%20-%2028Jun21.pdf.

Nakamoto, S., 2008. Bitcoin: A Peer-to-Peer Electronic Cash System. Working Paper .

Shiller, R., 1981. Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends? American Economic Review 71, 421–436.

Tantalo, C., Priem, R. L., 2016. Value creation through stakeholder synergy. Strategic Management Journal 37, 314–329.

UNGC, 2004. Who Cares Wins: Connecting Financial Markets to a Changing World. https://documents1.worldbank.org/curated/en/280911488968799581/pdf/113237-WP-WhoCaresWins-2004.pdf.

Vidal-Tomás, D., Ibañez, A., 2018. Semi-strong efficiency of Bitcoin. Finance Research Letters 27, 259–265.

von Carlowitz, H. C., von Rohr, J. B., Braun, J. F., 1732. Sylvicultura Oeconomica. Bey Johann Friedrich Brauns sel. Erben, second ed.

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 .

Chapter 2

Comparing ESG score weighting approaches and stock performance differentiation

Abstract: This study examines how weighting methodologies in ESG ratings for sustainability categories align with their financial relevance. We analyse Refinitiv's data-driven weights against equal weights and SASB's expert-based weights. Our findings show all methods differentiate firms by returns, but equal and SASB weights lead to a stronger effect, particularly after the Paris Agreement. This suggests alternative weighting approaches might better capture the financial significance of ESG categories.

Keywords: ESG ratings, category weightings
JEL Classification: M14, G24

2.1 Introduction

Sustainability information on environmental (E), social (S), and governance (G) factors has become increasingly important for investors in recent years. As such, firms are increasingly mandated to provide ESG reporting, satisfying the growing demand from investors (Ilhan et al., 2023). ESG ratings aggregate this sustainability information into a single number, but the market for these ratings remains unregulated (ESMA, 2021). This lack of regulation can lead to disagreement between rating agencies (e.g., Chatterji et al., 2016; Berg et al., 2022). ESG ratings combine information from various categories, such as CO2 emissions, human rights policies, or corporate social responsibility (CSR) strategies, with an assessment of their relevance to investors. This materiality concept ensures the information used significantly impacts a company's financial return. Material information increases the price informativeness of the rating, while immaterial information has no financial impact (Grewal et al., 2021). The relevance of each category is reflected in its weighting, which signifies its relative importance for the final ESG score. Typically, category weighting is industry-specific. For example, a category highly relevant in the energy sector (e.g., carbon emissions) may be less important in the financial services sector.

In this paper, we explore to what extent ESG weighting systems can be used to distinguish between different types of investments and assess their suitability for investors. Specifically, we focus on Refinitiv's ESG score[4] and weighting methodology, which is referenced in over 1,200 academic articles by the end of 2020 (Berg et al., 2020). According to an OECD report about ESG investing practices, Refinitiv is one of the three major rating providers (Boffo and Patalano, 2020) and is one of the considered ratings in the white paper "Seeking Return on ESG" as an intermediary for sustainable information during the World Economic Forum (WEF, 2019). Standard methods for assessing the materiality of ESG categories involve analysing economic channels within an industry, which can be prone to errors. Refinitiv's approach aims to circumvent this issue through a "data-driven methodology". Refinitiv assigns weights to categories, based on quantifiable data, such as a company's CO2 emissions (Refinitiv, 2023). While this method simplifies the calculation and promotes transparency, it

[4] For consistency within this paper, we will refer to Refinitiv, even though they became LSEG (London Stock Exchange Group) data and analytics in 2023.

does not analyse the direct impact of these categories (e.g., CO2 emissions) on a company's performance as reflected in the profit and loss statements or cash flows. We raise the question whether such a weighting system aligns with the financial relevance of categories.

To assess the relationship between financial performance and ESG weighting, we compare Refinitiv's ESG score to two alternative indices: an informed rating and an uniformed rating. All ESG ratings use the same category scores but different weightings. The informed benchmark rating enriches the Refinitiv rating by incorporating the Sustainability Accounting Standards Board's (SASB) materiality assessment for different industries. This materiality assessment is based on expert group judgement and aims to identify ESG material sustainability categories. The uniformed benchmark rating, in contrast, assigns equal weight to all ESG categories, offering a baseline comparison. Following Khan et al. (2016), we sort the stocks into decile and quintile portfolios based on their ESG rating. We then estimate the Alpha (the abnormal return) using the Fama and French (2015) five-factor model. We measure the performance difference between stocks with the highest and lowest ESG ratings as the differences in abnormal returns (Δ Alpha).

Our findings contribute to previous research in two important ways. First, we demonstrate that the weighting of ESG ratings could enhance the differentiation of firms based on stock returns. Stocks with the highest sustainability ratings exhibit lower returns compared to those with lower ratings. This effect is more pronounced for weights that incorporate industry-specific materiality through SASB information and for the naive equal-weighted approach. While Refinitiv's weights also differentiate between financially under- and over-performing stocks, the performance differences are smaller.

Second, to ensure the robustness of our results, we divided the sample into pre- and post-Paris Agreement (2015) sub-periods. Interestingly, significant return differences are primarily observed in the post-Agreement period. This suggests that the Agreement potentially leads investors to re-evaluate the importance of sustainability. Furthermore, when comparing the ESG indices, we find that the performance differences were larger for the naive and SASB ratings compared to Refinitiv's weighting.

In summary, our findings suggest that both the uninformed, equally weighted score and the SASB-based score may be more effective in differentiating stocks based on their future financial performance compared to the existing Refinitiv score. This enhanced effectiveness is particularly evident after the Paris Agreement, likely driven by an increased recognition of ESG factors by investors.

The remainder of this study is structured as follows: Section 2.2 addresses the related literature and details the research question. Section 2.3 describes the data and the methodology used. Section 2.4 provides the results. Section 2.5 checks for robustness. Section 2.6 concludes.

2.2 Related literature and research question

Sustainability ratings have become increasingly important for asset managers. They factor into best-in-class investment decisions (Drempetic et al., 2020) and influence how mutual funds align their strategies with ESG ratings (Hartzmark and Sussman, 2019). Investors' willingness to pay a premium for sustainable investments further strengthens this trend (Heeb et al., 2023). Additionally, successful investor engagement in sustainability can lead to increased institutional ownership (Dimson et al., 2015). Recent findings highlight that superior sustainable firms have lower systematic risk (Albuquerque et al., 2019), which translates to a lower cost of capital[5] (El Ghoul et al., 2011; Apergis et al., 2022; Wong et al., 2021) due to their perceived resiliency during economic downturns. This makes them function as a financial hedge for investors (Albuquerque et al., 2020; Broadstock et al., 2021). Consequently, a link between sustainability information and financial performance is evident (e.g., Whelan et al., 2021; Friede et al., 2015). Interestingly, studies by Drempetic et al. (2020) and Dobrick et al. (2023) identify a potential size bias in Refinitiv ESG scores, where larger companies tend to have better ratings.

The relationships described above highlight the importance of ESG ratings. These ratings make sustainability measurable by combining various social (e.g., Hong and Kacperczyk, 2009; Edmans, 2011) and environmental proxies (e.g., Bolton and Kacperczyk, 2021). In addition to traditional credit rating agencies, several financial data providers offer their own

[5]However, the probability of default is not normally distributed but depends on the respective industry. See, for example, Palmieri et al. (2023)

sustainability assessments. Refinitiv, MSCI, and Bloomberg are listed as the three major rating agencies, according to the OECD.[6] Alongside these, specialized providers exist, such as the Carbon Disclosure Project (CDP) or the Institutional Shareholder Services (ISS) (Douglas et al., 2017).

ESG ratings combine category scores and weights. Category scores are firm-specific characteristics that measure various aspects of sustainability. Weights measure the importance of the category and are industry-specific.[7] This acknowledges that a category important in one industry may be less relevant in another. Specifically, consider firm i from industry j. Its ESG score is given by

$$ESG_i = \sum_{c=1}^{N} C_{c,i} \cdot w_{c,j}, \tag{2.1}$$

where $C_{c,i}$ is the firm's score for category $c \in \{1, ..., N\}$ and $w_{c,j}$ is the weight for category c in industry j.

ESG ratings differ with regard to categories and weighting methods, despite similar data collection. For example, Refinitiv includes the most indicators, while MSCI only contains 37 issues in their calculation (Billio et al., 2021). Notably, MSCI and Refinitiv are similar concerning their methodology. Both are industry-based measures with weights reflecting the relevance of different themes (Eccles and Stroehle, 2018; Refinitiv, 2023).

The literature on ESG rating construction mainly addresses the choice of categories. For example, Chatterji et al. (2016) and Berg et al. (2022) analyse rating disagreements between providers and the relative importance of different ESG categories for ratings. These differences in ratings are considered as financial risk, and Gibson et al. (2021) find that financial market participants price that divergence with a risk premium. To reduce such a divergence, regulatory frameworks like the EU Taxonomy could be used. For example, it defines a set of environmental properties, and Dumrose et al. (2022) analyse the consistency of rating categories with the EU Taxonomy for several providers. Their findings show that ESG measures from three out of four agencies are significantly related to the EU Taxonomy.

[6] See the report "ESG Investing: Practices, Progress and Challenges" (Boffo and Patalano, 2020).

[7] This aligns with, for example, Eccles et al. (2012) who state that sustainability risks vary significantly across industries.

While significant research focuses on selecting relevant categories for ESG ratings, less attention has been paid to how those categories are weighted. The literature suggests that the materiality of categories drives their financial importance for investors, but this materiality may vary across industries (Eccles et al., 2012). Therefore, the Sustainability Accounting Standards Board (SASB) also defines the materiality of categories with respect to the industry. Using categories from the MSCI sustainability ratings, Khan et al. (2016) find that mainly material sustainability categories matter to discriminate stocks according to their performance. Moreover, (Schiehll and Kolahgar, 2021) and (Consolandi et al., 2022) consider financial disclosure and show that material information is most relevant for stock returns.

In this paper, we investigate whether the weighting of categories in Refinitiv's ESG ratings aligns with their financial importance for investors. Refinitiv is considered one of the most relevant ESG ratings and employs a data-driven approach similar to MSCI ratings. We explore whether the weights assigned to categories in Refinitiv's ESG ratings better reflect their financial importance compared to a simple, equal weighting scheme (detailed in the next section) and further benchmark these weights against those obtained from SASB's materiality matrix, a comprehensive framework that defines industry-specific materiality of ESG categories based on expert assessments.

2.3 Data and methodology

Our sample consists of 3,098 firms having their headquarters in the U.S. The examination period starts in the year 2010 and ends in 2022, resulting in 18,352 firm-year observations. ESG and price data are taken from the Refinitiv database. We sort the stocks into equally weighted decile and quintile portfolios based on their ESG score, as detailed below. Companies with the lowest (highest) scores are assigned to the bottom (top) portfolios. In line with Refinitiv's updating of sustainability measures, we rebalance portfolios annually at the beginning of the fiscal year.

We consider three different sets of weights, including "Refinitiv weights", "SASB weights", and "naive weights" leading to different ESG scores. Refinitiv weights reflect Refinitiv's assessment of a category for an industry. They apply a data-driven approach: Each industry group and ESG category is assigned a magnitude weight. The proxy for the magnitude weight can be quantitative (measured with units, e.g., tons of carbon emissions) or binary (yes/no answer, e.g., availability of a firm's human rights policy). Industry medians are determined for quantitative categories. For binary categories, the share of companies with a positive answer within an industry is determined. Industry results are then ranked and mapped on scales from 1 (low exposure) to 10 (high exposure).[8] Category weights are calculated as

$$w_{c,j} = \frac{MagnitudeWeight_{c,j}}{\sum_{c=1}^{N} MagnitudeWeight_{c,j}}. \tag{2.2}$$

Governance categories use a different approach, which is not elaborated on here.[9]

The SASB introduced investor-focused guidance for sustainability disclosure in February 2014. This guidance identifies specific sustainability risks with different financial impacts in industries. SASB distinguishes between material and immaterial sustainability risks for each industry based on expert group judgement (Khan et al., 2016). We compute SASB weights to address the economic channels of Refinitiv's categories by enriching them with data from the SASB's materiality matrix. For coherence, SASB categories and industries are mapped to Refinitiv's categories and industries (see Table 2.7 in the Online Appendix). Magnitude weights are then assigned as 10 for material categories and 5 for immaterial categories, reflecting the best possible score and Refinitiv's default value. This weighting scheme essentially doubles the influence of material categories compared to immaterial ones. Weights are calculated using the same formula as Refinitiv's weights (Equation (2.1)). By using SASB weights, we make a strong assumption: the identification of material and immaterial categories was independent of stock returns in our sample. This means past stock performance did not influence the classification of sustainability factors. We address this potential issue in greater detail in the robustness section 2.5. Finally, naive weights[10] serve

[8]See Online Appendix for details.

[9]The weighting of the governance categories deviates from the number of underlying indicators. In other words, the more data points are included in the calculation, the higher the weighting. For a more detailed explanation, see Table 2.6 (Online Appendix). For the next steps, we will use the same weighting approaches as for the other categories.

[10]Equal weighting could be used, if the evidence about different themes and their contributions are not clear. Nearly 45 percent of sustainability indices, such as the Human Development Index, use an equal-weight approach (Gan et al., 2017).

as a benchmark, assuming no information about the importance of categories is available. In this case, we assume that all weights are equal, resulting in an equally weighted sum of the magnitude weights.

We explore the financial informativeness of ESG scores by comparing the excess returns of the top and bottom portfolios.[11] We follow the most common approach in the literature and compute excess stock returns Alpha according to the Fama and French (2015) five-factor model, which captures the factors market (MKT), size (SMB), book-to-market (HML), profitability (RMW), and investment (CMA). As a market benchmark, we follow the standard procedure. This benchmark is computed from all stocks listed on the NYSE, AMEX, and NASDAQ and represents approximately the whole U.S. market capitalization.[12] Note that this procedure implicitly assumes the efficient market hypothesis in the semi-strong form, that is, all publicly available ESG information is included in prices (Fama, 1970). The advantage of our standard approach is that results are comparable to the existing literature on ESG ratings (in particular Khan et al. (2016)). An alternative performance measure would be to use forward-looking information like analyst-based implied cost of capital (ICC) instead of Alpha. This has been done, for example, by Pástor et al. (2022). However, this procedure implicitly assumes that analysts' financial forecasts are accurate and unbiased.[13]

2.4 Results

Table 2.1 presents descriptive statistics of ESG scores. The naive-unweighted and SASB scores have similar means and standard deviations. While the Refinitiv score also matches the other scores in terms of standard deviations, it has substantially higher means. These observations also apply to the environmental, social, and governance pillars considered. However, the discrimination of firms is indicated by the dispersion of scores. Across all weighting methods considered, the environmental categories have the highest dispersion in terms of

[11]See for similar methodical approaches, for example, Hübel and Scholz (2020), Avramov et al. (2022), and Pedersen et al. (2021).

[12]The market benchmark and the other factors are obtained through Kenneth R. French's web-page: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[13]For a discussion, see, for example, Gebhardt et al. (2001), Easton and Monahan (2005), and Guay et al. (2011).

Table 2.1: Descriptive statistics: ESG scores and ESG pillars

	Obs.	Mean	Std. Dev.	Minimum	P25	Median	P75	Maximum
Original score								
ESG	18,352	40.90	19.34	0.63	25.76	37.76	54.71	95.16
ENV	18,352	24.08	27.12	0.00	0.00	13.96	43.02	98.55
SOC	18,352	41.56	20.73	0.44	25.76	38.03	55.55	98.01
GOV	18,352	48.10	22.35	0.16	30.15	48.33	66.00	99.46
Naive unweighted score								
ESG^{Equal}	18,352	35.25	19.59	0.34	20.36	28.85	47.78	95.51
ENV^{Equal}	18,352	22.34	26.23	0.00	0.00	10.61	40.51	98.48
SOC^{Equal}	18,352	39.99	20.13	0.45	25.00	35.91	52.18	98.07
GOV^{Equal}	18,352	41.84	20.40	0.08	26.55	39.88	55.85	99.48
SASB risk-weighted score								
ESG^{SASB}	18,352	36.20	20.02	0.50	21.00	29.93	49.04	96.39
ENV^{SASB}	18,352	23.20	27.25	0.00	0.00	10.65	42.96	98.69
SOC^{SASB}	18,352	40.99	20.27	0.41	25.83	37.26	53.76	98.31
GOV^{SASB}	18,352	41.59	20.50	0.08	26.11	39.49	55.52	99.48

Notes: This table shows summary statistics of the original ESG score (ESG), its environmental pillar (ENV), social pillar (SOC), governance pillar (GOV). The naive-unweighted scores are ESG^{Equal}, ENV^{Equal}, SOC^{Equal}, GOV^{Equal}. The SASB risk-weighted scores are ESG^{SASB}, ENV^{SASB}, SOC^{SASB}, GOV^{SASB}. The mean, standard deviation, minimum, 25% quantile, median, 75% quantile, and maximum are reported. The sample includes data points between February 2010 and December 2022.

standard deviations of scores. Specifically, the standard deviations of the environmental scores are around 27 compared to standard deviations around 20 for the social and the governance scores.

Table 2.2 details the distributions of ESG scores across industries. Consistent with the overall data, naive-unweighted and SASB scores again show similar means across industries. Refinitiv scores, however, exhibit higher means compared to the others. This difference is most pronounced in the financial, real estate, and academic & educational services industries. Conversely, utilities and consumer non-cyclicals are less affected. Importantly, the standard deviations of all scores remain similar across industries.

Table 2.3 shows the regression results of a Fama and French (2015) five-factor model applied to top and bottom portfolios formed by ESG scores calculated with Refinitiv weights (Panel A), SASB weights (Panel B), and naive weights (Panel C). Concerning Fama and French (2015) factors, the estimated models show no major differences across the ESG metrics. Moreover, for all scores, the top portfolio has significantly negative Alphas. This suggests

Table 2.2: Descriptive statistics: ESG scores by industry

	Obs.	ESG	SD	ESG^{Equal}	SD^{Equal}	Diff.	ESG^{SASB}	SD^{SASB}	Diff.
Academic & Educational Services	77	34.52	15.25	27.29	13.23	**7.22**	27.63	12.84	**6.89**
Basic Materials	975	45.39	20.48	43.29	19.10	**2.10**	43.08	19.95	**2.31**
Consumer Cyclicals	2,615	41.37	18.91	35.89	18.95	**5.48**	36.05	19.82	**5.32**
Consumer Non-Cyclicals	901	48.07	24.48	46.25	24.00	1.82	45.32	25.25	**2.75**
Energy	1,050	39.04	20.74	36.51	19.70	**2.53**	38.69	20.04	0.35
Financials	2,983	38.59	15.90	29.50	16.11	**9.08**	31.90	16.37	**6.68**
Healthcare	2,966	35.81	18.29	28.66	17.53	**7.15**	30.69	18.04	**5.12**
Industrials	2,265	41.48	18.81	37.35	18.83	**4.12**	37.24	19.77	**4.24**
Real Estate	1,345	42.73	20.11	35.09	18.43	**7.64**	34.57	19.10	**8.16**
Technology	2,589	42.05	19.51	36.85	20.83	**5.20**	38.30	21.25	**3.74**
Utilities	586	50.39	18.76	48.65	18.95	1.74	48.32	20.45	**2.07**
Total	18,352	35.25	19.34	40.90	19.59	**5.65**	36.20	20.02	**4.69**

Notes: This table presents the differences between the original Refinitiv, the naive-unweighted, and the SASB risk-weighted ESG score across industries. The number of observations, the mean, the standard deviation of all scores, and the average differences are reported. Bold coefficients are significant at the 10% level. The sample includes data points between February 2010 and December 2022.

that sustainability is negatively associated with stock returns, which aligns with existing literature.

The performance difference (Δ Alpha) is the key metric of our analysis. It is printed in bold and measures the difference in unexplained returns (Alpha) between top and bottom portfolios. These differences are statistically significant. Decile portfolios, compared to quintile portfolios, exhibit higher levels of significance. This suggests that all ESG scores can differentiate firms based on their financial performance, with decile portfolios providing a clearer distinction. The largest difference is observed for SASB weights, followed by naive weights. However, these two metrics show very similar abilities to distinguish. In contrast, Refinitiv weights lead to substantially smaller performance differences (Δ Alpha) with lower levels of statistical significance, especially for quintile portfolios. In summary, our results imply that while all metrics use the same information, SASB weights seem to marginally improve the identification of financially relevant categories compared to naive weights. Refinitiv weights, however, appear less effective in this regard. These findings are relevant for investment approaches that select their stocks based on Refinitiv ESG ratings. Their data-driven approach is not as effective in measuring financial risk as an uninformed equal-weighted rating. As a consequence, investors do not receive the expected risk-return relation, which is reflected by the rating. However, they could address this shortcoming by integrating, for example, a SASB weighting to enhance the informativeness (Khan et al., 2016) of Refinitiv ESG ratings.

2.5 Robustness

As a robustness test, we follow Bolton and Kacperczyk (2021) and consider the Paris Agreement (December 2015) as an exogenous shock. According to Duong et al. (2022) “the Agreement is considered as the most significant event in climate finance history”. The Agreement’s focus on climate change makes it particularly relevant in our context for two reasons. First and as indicated above, environmental categories have the highest dispersion of scores (in terms of standard deviations), implying the most pronounced differentiation between firms. Second, it is reasonable to assume the Agreement served as an important turning point, prompting investors to recognise the broader relevance of sustainability issues (including

Table 2.3: Regression results: performance differences

	Deciles				Quintiles			
	Bottom		Top		Bottom		Top	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Panel A: Refinitiv ESG score ($N = 155$)								
MKT	0.9701***	22.8602	1.0101***	53.8631	1.0212***	27.0196	1.0228***	52.0335
SMB	0.7699***	9.3446	0.0959***	2.6344	0.7177***	9.7803	0.1817***	4.7623
HML	0.0091	0.1238	0.2053***	6.3134	0.0430	0.6558	0.2049***	6.0099
RMW	-0.4079***	-4.1548	0.1258***	2.8999	-0.3521***	-4.0266	0.1402***	3.0838
CMA	-0.0308	-0.2717	0.0812	1.6203	-0.1276	-1.2640	0.0630	1.2008
$Intercept$	0.1546	0.8552	-0.2278***	-2.8514	0.0958	0.5950	-0.1616*	-1.9295
R^2	0.8757		0.9626		0.9029		0.9615	
Annualized Alpha	1.87		-2.70***		1.16		-1.92*	
Δ Alpha				**4.57****				**3.08***
Panel B: SASB risk-weighted ESG score ($N = 155$)								
MKT	0.9581***	22.1052	1.0112***	55.1359	0.9928***	27.6461	1.0383***	52.2352
SMB	0.7252***	8.6189	0.0510	1.4335	0.7204***	10.3317	0.1903***	4.9316
HML	-0.0965	-1.2844	0.2180***	6.8543	-0.0448	-0.7194	0.2243***	6.5057
RMW	-0.3786***	-3.7764	0.1236***	2.9128	-0.3188***	-3.8369	0.1336***	2.9052
CMA	-0.0663	-0.5724	0.0753	1.5372	-0.0467	-0.4866	0.0425	0.7998
$Intercept$	0.3743**	2.0272	-0.1696**	-2.1707	0.1608	1.0510	-0.1595*	-1.8838
R^2	0.8638		0.9637		0.9058		0.9621	
Annualized Alpha	4.59**		-2.02**		1.95		-1.90*	
Δ Alpha				**6.61*****				**3.85****
Panel C: Naive-unweighted ESG score ($N = 155$)								
MKT	0.9649***	22.6174	1.0125***	55.0927	0.9917***	27.5742	1.0300***	51.4703
SMB	0.7039***	8.4984	0.0361	1.0120	0.7254***	10.3884	0.1766***	4.5466
HML	-0.0501	-0.6772	0.2388***	7.4904	-0.0293	-0.4698	0.2237***	6.4441
RMW	-0.4283***	-4.3395	0.1132***	2.6619	-0.4239***	-5.0948	0.1567***	3.3850
CMA	-0.1144	-1.0039	0.0515	1.0484	-0.0856	-0.8905	0.0315	0.5888
$Intercept$	0.3442*	1.8938	-0.1634**	-2.0877	0.1832	1.1956	-0.1527*	-1.7918
R^2	0.8697		0.9637		0.9081		0.9608	
Annualized Alpha	4.21*		-1.94**		2.22		-1.82*	
Δ Alpha				**6.15*****				**4.04****

Notes: This table shows coefficient estimates and t-statistics of the Refinitiv, SASB risk-weighted, and naive-unweighted ESG regressions. All values are presented in percent. The portfolios are sorted by ESG scores. Companies with the worst scores are in the bottom portfolio, and companies characterized by the best scores are assigned to the top portfolio, respectively. Those portfolios are rebalanced on a yearly basis. The dependent variable is the equal-weighted monthly return of the ESG portfolios minus the risk-free rate. Independent variables are Fama and French (2015): MKT, SMB, HML, RMW, CMA. The results are based on the monthly average returns between February 2010 and December 2022. The difference in annualized Alphas reflects the returns of a long-short portfolio, with long positions in stocks with low ESG ratings and short positions in high ESG stocks. ***, ** and * reflect significance at the 1%, 5%, and 10% level.

climate risks) to asset pricing. Therefore, we divide our sample into two subsamples: the first from February 2010 to November 2015 (Before Paris Agreement) and the second from January 2016 to February 2020 (After Paris Agreement). We exclude data from March 2020 onwards to eliminate potential influence from the COVID-19 crisis on our findings.

Results in Table 2.4 reveal that after the Paris Agreement, all ESG metrics displayed a statistically significant ability to differentiate investments, but with varying degrees of effectiveness as measured by the performance difference (Δ Alpha). Naive-unweighted and SASB risk-weighted ESG scores exhibit the strongest distinction, with significantly higher returns (positive Alpha) for the bottom quintile portfolios. This effect is also observed for decile portfolios. Refinitiv ESG scores, on the other hand, show the weakest differentiation, with significant Alphas only being observed for the top quintile (and decile) portfolios. Hence, naive-unweighted and SASB risk-weighted ESG metrics may be more sensitive to identifying stocks with differing financial performance in the post-Agreement era.

In the pre-Agreement period, none of the performance differences (Δ Alpha) are statistically significant for quintile portfolios. For decile portfolios, only the top decile with SASB weights exhibited a significant performance difference. However, since the SASB guidance was not established before February 2014, we cannot rule out that this finding might be due to endogeneity. Performance differences for other ESG weightings were not significant during this period. Interestingly, we also find that top portfolios across all ESG metrics were associated with potentially lower future returns before the Paris Agreement.

In summary, our results suggest a shift in investor behaviour after the Paris Agreement. The Paris Agreement might have influenced how investors perceive ESG factors when making investment decisions, leading to performance differences captured by ESG scores to varying degrees.

Table 2.5 displays a series of additional robustness tests. To generalize our results, we first re-run the analysis with firms headquartered in the U.S., Europe, Asia, Canada, and Australia. The Alpha differences increase for all weightings, but the SASB and naive approaches better reflect the differentiation. In the next series of robustness tests, we exclude observations from the aftermath of the Financial Crisis and the COVID-19 period. All alternative sub-samples confirm our main results. To capture the dynamic nature of ESG scores, we rebalance the

Table 2.4: Regression results: Paris Agreement

	Deciles				Quintiles			
	Bottom		Top		Bottom		Top	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Panel A: Before Paris Agreement ($N = 70$)								
Refinitiv								
Intercept	-0.0320	-0.1590	-0.2684***	-2.7136	-0.0137	-0.0903	-0.2255***	-2.9717
Annualized Alpha	-0.38		-3.17***		-0.16		-2.67***	
Δ Alpha				**2.79**				**2.51**
SASB risk-weighted								
Intercept	0.1597	1.0608	-0.2959***	-3.6265	-0.0168	-0.1336	-0.2065***	-2.8106
Annualized Alpha	1.93		-3.49***		-0.20		-2.45***	
Δ Alpha				**5.42*****				**2.25**
Naive-unweighted								
Intercept	0.0134	0.0797	-0.2633***	-3.1756	-0.0515	-0.4176	-0.1516**	-2.0564
Annualized Alpha	0.16		-3.11***		-0.62		-1.80**	
Δ Alpha				**3.27**				**1.18**
Panel B: After Paris Agreement ($N = 50$)								
Refinitiv								
Intercept	0.1998	0.7695	-0.2997**	-2.6326	0.2754	1.3196	-0.2693**	-2.1724
Annualized Alpha	2.42		-3.54**		3.36		-3.18**	
Δ Alpha				**5.96*****				**6.54*****
SASB risk-weighted								
Intercept	0.4638**	2.0617	-0.1329	-1.0100	0.3508*	1.8848	-0.3076**	-2.2145
Annualized Alpha	5.71**		-1.58		4.29*		-3.63**	
Δ Alpha				**7.29*****				**7.92*****
Naive-unweighted								
Intercept	0.5198**	2.2734	-0.1376	-1.1142	0.4364**	2.1810	-0.3128**	-2.2910
Annualized Alpha	6.42**		-1.64		5.36**		-3.69**	
Δ Alpha				**8.06*****				**9.05*****

Notes: This table shows coefficient estimates of the intercepts and t-statistics of the Refinitiv, naive-unweighted, and SASB risk-weighted ESG regressions before and after the Paris Agreement in December 2015. All values are presented in percent. The portfolios are sorted by ESG scores. Companies with the worst scores are in the bottom portfolio, and companies characterized by the best scores are assigned to the top portfolio, respectively. Those portfolios are rebalanced on a yearly basis. The dependent variable is the equal-weighted monthly return of the ESG portfolios minus the risk-free rate. Independent variables of the Fama and French (2015) model are hidden. The period before the Paris Agreement starts in February 2010 and ends in November 2015, and the period after the Paris Agreement is between January 2016 and February 2020. The difference in annualized Alphas reflects the returns of a long-short portfolio, with long positions in stocks with low ESG ratings and short positions in high ESG stocks. ***, ** and * reflect significance in the difference of alphas at the 1%, 5%, and 10% level.

Table 2.5: Regression results: robustness tests

	Decile Δ Alpha			Quintile Δ Alpha		
	Refinitiv	SASB	Unweighted	Refinitiv	SASB	Unweighted
Alternative Regions						
Developed Countries	6.80***	7.63***	7.52***	4.84***	6.29***	6.34***
Alternative Sample Periods						
Excluding COVID	3.80**	6.08***	5.23***	3.96***	4.49***	4.28***
Excluding Fin. Crisis	4.48*	6.53***	6.65***	2.35	3.48*	3.98*
Excluding Both Events	3.49*	5.93***	5.57***	3.60**	4.42***	4.51**
Alternative Rebalancing						
Monthly Rebalancing	4.79**	6.03***	5.68***	3.04*	3.82**	4.05**
Alternative Portfolio Sorting						
Residual ESG	5.02**	5.35**	5.67**	2.47	3.73**	3.09*

Notes: This table presents the annualized Alpha differences of the robustness tests. All values are presented in percent. Companies with the worst scores are in the bottom portfolio, and companies characterized by the best scores are assigned to the top portfolio, respectively. Those portfolios are rebalanced on a yearly basis. The dependent variable is the equal-weighted monthly return of the ESG portfolios minus the risk-free rate. Independent variables are Fama and French (2015): MKT, SMB, HML, RMW, CMA. The results are based on the monthly average returns between February 2010 and December 2022. The difference in annualized alphas reflects the returns of a long-short portfolio, with long positions in stocks with low ESG ratings and short positions in high ESG stocks. The portfolios are sorted by ESG scores. Under "Alternative Regions", we examine a different sample and extend it to 6,900 firms with their headquarters in the developed countries. Under "Alternative Sample Periods", we report differences in Alpha for sub-periods. The "Excluding COVID" period starts in February 2010 and ends in February 2020. The "Excluding Fin. Crisis" period does not include the aftermath of the financial crisis and starts in January 2013 and ends in December 2022. The "Excluding Both Events" period includes observations between January 2013 and February 2020. Under "Alternative Rebalancing", we rebalance the portfolios on a monthly basis. Under "Alternative Portfolio Sorting", we estimate an ESG score with respect to firm size, market-to-book ratio, return-on-assets, and leverage and conduct the sorting based on those estimates. ***, ** and * reflect significance at the 1%, 5%, and 10% level.

portfolios on a monthly basis. Notably, the quintile portfolio results diverge only minimally from the yearly rebalancing. This is because most disclosed ESG data is updated once a year (Refinitiv, 2023). In the final robustness test, we address possible endogeneity concerns. We estimate an ESG score based on firm characteristics (size, market-to-book ratio, leverage and return on assets) and sort the decile and quintile portfolios based on the new estimates. The Alpha differences are more similar, but the general trend remains evident. Across all additional tests, we confirm our main findings: the largest differences are displayed for SASB weights and naive weights.

2.6 Conclusion

Sustainability and financial performance are closely linked. Consequently, ESG ratings are valuable tools for managers and academics, providing information about the sustainability of investments and enhancing transparency in financial markets. From a policymaker's point of view, understanding how weighting methodologies impact the differentiation of firms based on ESG factors can inform policymakers seeking to establish robust ESG rating frameworks. This study investigates the weighting of categories in Refinitiv's ESG rating, recognized as one of the most important metrics in the field. Refinitiv employs a data-driven approach similar to the methodology of MSCI ratings. We focus on whether the weighting of categories in Refinitiv's ESG ratings aligns with their financial importance for investors. We compare Refinitiv's weights with a simple, equal weighting scheme and with SASB weights calculated based on expert assessments.

We find that all weighting methods can differentiate firms based on stock returns. However, equal weights and SASB weights lead to a more pronounced differentiation of firm performance compared to Refinitiv's data-driven weighted score, especially following the Paris Agreement. This suggests that alternative weighting approaches might better capture the financial relevance of ESG categories.

Specifically, one future research direction is to identify procedures that lead to weights enabling a clearer differentiation of investment opportunities based on ESG criteria. From a regulatory perspective, it is important to identify weights that better align with the financial relevance of categories. This would further enhance market transparency and inform investment decisions. Our work could be extended by analysing the weighting methodology of other providers in detail to see if our results hold true for these other major rating agencies. Finally, while the Fama and French (2015) five-factor model was used to test return differences, another interesting avenue for future research could involve incorporating more forward-looking information, such as analyst-based implied cost of capital.

2.7 Appendix

Calculation of the Refinitiv Materiality Matrix (Refinitiv, 2023):

For the environmental and social pillars the weights are based on specific data points representing the whole category. Two methods are applied for the numeric (industry median) and the Boolean data (transparency weights) points:

1. Industry median

For every TRBC industry group a median for the particular data point is calculated. After that, all industry medians are ranked and sorted into deciles. The decile rank determines an industry magnitude weight which is between 1 and 10.

2. Transparency weights

For the transparency weights a disclosure percentage for every TRBC industry group is calculated. After that, all industry disclosure percentages are ranked and sorted into deciles. The decile rank determines an industry magnitude weight between 1 and 10.

The industry magnitude weights of the categories within each industry are summed up. Each category's magnitude weight is divided by the sum of magnitude weights of the respective industry group to derive the category weight.

$$w_{c,j} = \frac{MagnitudeWeight_{c,j}}{\sum_1^{10} MagnitudeWeight_{c,j}}$$

Additional notes for the calculation:

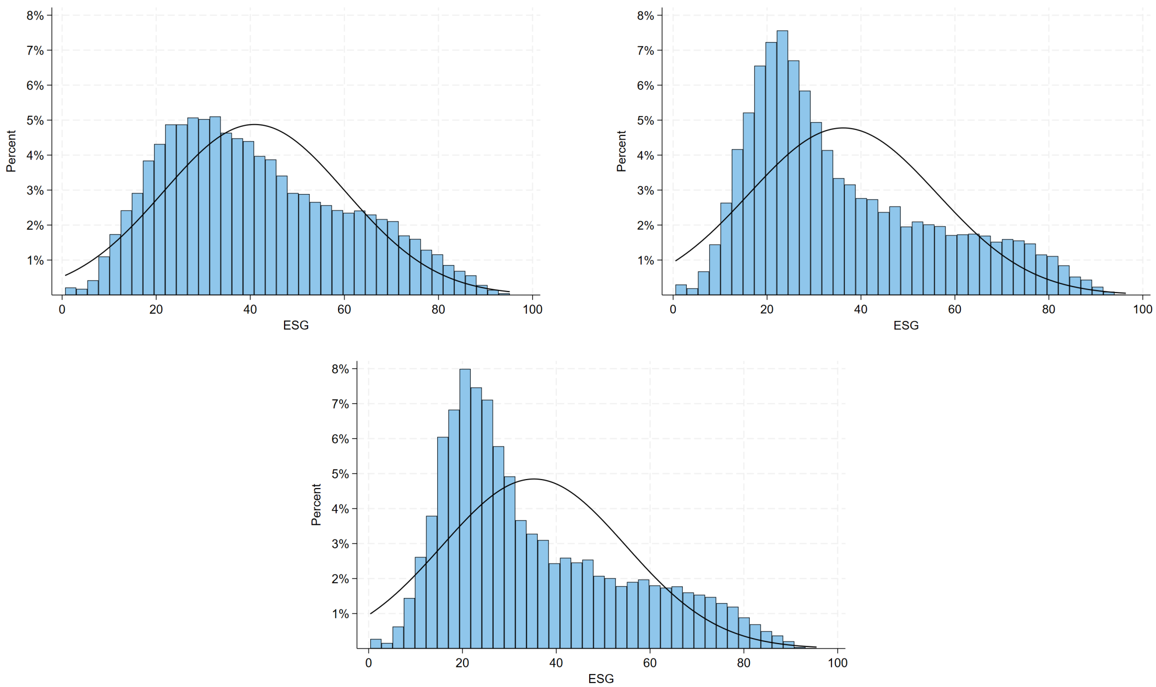
- If a category has more than one data point as reference for the importance (e.g., emissions category with emissions and waste), the average of the decile weights is considered to derive the magnitude weight for that category.

- Corporate governance has a different calculation method: data points in each governance category/total data points in the governance pillar, multiplied by the default category weights of 15.

- Due to data availability reasons, half of the industries have a default value of 1 in the environmental R&D expenditure (innovation category) theme.

- Because responsible product marketing (product responsibility category) is an industry specific data-point and most industries do not disclose this theme, around 80 percent of the industry groups have a default value of 1.

Figure 2.1: Histogram ESG scores



Notes: This figure shows the distribution of the weighted Refinitiv ESG score (left figure) the SASB risk-weighted ESG score (middle figure), and the naive-unweighted ESG score (right figure). The sample consists of 3,098 U.S. firms between February 2010 and December 2022 resulting in 18,352 firm-year observations

Table 2.6: Proxy data points for industry magnitudes

Categories	**Themes**	**Data points**	**Weight method**
Emissions	Emissions	TR.AnalyticCO2	Quant industry median
	Waste	TR.AnalyticTotalWaste	Quant industry median
	Biodiversity		no data point available
	Environmental management systems		no data point available
Innovation	Product innovation	TR.EnvProducts	Transparency weights
	Green revenues, research and development (R&D) and capital expenditures (CapEx)	TR.AnalyticEnvRD	Quant industry median
Resource use	Water	TR.AnalyticWaterUse	Quant industry median
	Energy	TR.AnalyticEnergyUse	Quant industry median
	Sustainable packing		no data point available
	Environmental supply chain		no data point available
Community	Equally important to all industry groups		median weight of 5 is assigned to all
Human rights	Human rights	TR.PolicyHumanRights	Transparency weights
Product responsibility	Responsible marketing	TR.PolicyResponsibleMarketing	Transparency weights
	Product Quality	TR.ProductQualityMonitoring	Transparency weights
	Data privacy	TR.PolicyDataPrivacy	Transparency weights
Workforce	Diversity and inclusion	TR.WomenEmployees	Quant industry median
	Career development and training	TR.AvgTrainingHours	Transparency weights
	Working conditions	TR.TradeUnionRep	Quant industry median
	Health and safety	TR.AnalyticLostDays	Transparency weights
CSR strategy	CSR strategy		(Count of data points in each governance category / all data points in governance pillar) multiplied by 15
	ESG reporting and transparency		
Management	Structure (independence, diversity, committees)		
	Compensation		
Shareholders	Shareholder rights		
	Takeover defences		

Source: (Refinitiv, 2023).

Table 2.7: ESG category assignment

Refinitiv Pillar	Refinitiv Categories	SASB Category Issues
Environment	Resource Use	Energy Management Water and Wastewater Management Product Design and Lifecycle Management
	Emissions	GHG Emissions Air Quality Waste and Hazardous Materials Management Ecological Impacts
	Innovation	-
Social	Workforce	Employee Health and Safety Employee Engagement, Diversity and Inclusion Supply Chain Management
	Human Rights	Human Rights and Community Relations Labour Practices Supply Chain Management
	Community	Human Rights and Community Relations Business Ethics Competitive Behaviour
	Product Responsibility	Customer Privacy Data Security Access and Affordability Customer Welfare Selling Practices and Product Labelling Product Quality and Safety Materials Sourcing and Efficiency
Governance	Management	Management of the Legal and Regulatory Environment
	Shareholders	Management of the Legal and Regulatory Environment
	CSR Strategy	Business Model Resilience

Table 2.8: SASB industry group weightings

	RESU	EMIS	ENVI	WORK	HUMR	COMM	PROD	MGMT	SHAR	CSRS
Investment Holding Companies	10	5	5	10	5	10	10	5	5	5
Banking Services	5	5	5	5	5	10	10	5	5	5
Investment Banking & Investment Services	10	5	5	10	5	10	5	5	5	5
Collective Investments	10	5	5	10	5	10	10	5	5	5
Insurance	10	5	5	5	5	5	10	5	5	5
Biotechnology & Medical Research	5	5	5	10	10	10	10	5	5	5
Pharmaceuticals	5	5	5	10	10	10	10	5	5	5
Healthcare Equipment & Supplies	10	10	5	10	10	10	10	5	5	5
Healthcare Providers & Services	5	5	5	5	5	5	10	5	5	5
Software & IT Services	10	5	5	10	5	10	10	5	5	5
Semiconductors & Semiconductor Equipment	10	10	5	10	5	10	10	5	5	5
Communications & Networking	10	5	5	10	5	10	10	5	5	5
Electronic Equipment & Parts	10	10	5	10	10	5	10	5	5	5
Computers, Phones & Household	10	5	5	10	10	5	10	5	5	5
Integrated Hardware & Software	10	5	5	10	10	5	10	5	5	5
Office Equipment	10	5	5	10	10	5	10	5	5	5
Telecommunications Services	10	5	5	10	5	10	10	5	5	5
Financial Technology (Fintech) & Infrastructure	10	5	5	10	5	10	10	5	5	5
Metals & Mining	10	10	5	10	10	10	5	5	5	5
Construction Materials	10	10	5	10	10	10	5	5	5	10
Chemicals	10	10	5	10	10	10	5	10	10	5
Containers & Packaging	10	10	5	10	10	5	10	5	5	5
Paper & Forest Products	10	10	5	10	10	10	5	5	5	5
Real Estate Operations	10	5	5	5	5	10	5	5	5	5
Residential & Commercial REITs	10	5	5	5	5	5	5	5	5	5
Construction & Engineering	10	5	5	5	5	10	5	5	5	5
Professional & Commercial Services	10	10	5	10	10	10	10	5	5	10
Diversified Industrial Goods Wholesale	10	5	5	10	10	5	10	5	5	5
Machinery, Tools, Heavy Vehicles, Tr	10	10	5	10	5	10	10	5	5	5
Aerospace & Defence	10	10	5	5	5	10	10	5	5	5
Passenger Transportation Services	10	10	5	5	10	10	10	5	5	5
Freight & Logistics Services	5	10	5	10	10	10	5	5	5	5
Transport Infrastructure	5	5	5	10	5	10	10	5	5	5
Oil & Gas	10	10	5	10	10	10	5	10	10	5
Oil & Gas Related Equipment and Services	10	10	5	10	5	10	5	10	10	5

(continued on the next page)

Table 2.8: SASB industry group weightings - continued

	RESU	EMIS	ENVI	WORK	HUMR	COMM	PROD	MGMT	SHAR	CSRS
Coal	10	10	5	10	10	10	5	5	5	10
Renewable Energy	10	10	5	10	5	5	10	5	5	5
Electric Utilities & IPPs	10	10	5	10	5	5	10	5	5	10
Multiline Utilities	10	10	5	10	5	5	10	5	5	10
Natural Gas Utilities	5	5	5	5	5	5	10	5	5	10
Water & Related Utilities	10	5	5	5	5	5	10	5	5	10
Automobiles & Auto Parts	10	10	5	5	10	10	10	5	5	5
Textiles & Apparel	5	5	5	10	10	5	10	5	5	5
Homebuilding & Construction Supplies	10	10	5	10	5	5	5	5	5	10
Household Goods	10	5	5	10	10	5	10	5	5	5
Leisure Products	5	5	5	5	10	5	10	5	5	5
Hotels & Entertainment Services	10	10	5	10	10	10	10	5	5	5
Media & Publishing	5	5	5	10	5	10	10	5	5	5
Speciality Retailers	10	5	5	10	10	5	10	5	5	5
Diversified Retail	10	5	5	10	10	5	10	5	5	5
Food & Tobacco	10	10	5	10	10	5	10	5	5	5
Beverages	10	10	5	10	10	5	10	5	5	5
Food & Drug Retailing	10	10	5	10	10	5	10	5	5	5
Consumer Goods Conglomerates	10	10	5	5	5	10	10	5	5	5
Personal & Household Products & Services	10	5	5	10	10	5	10	5	5	5
Miscellaneous Educational Service P	5	5	5	5	5	5	10	5	5	5
Professional & Business Education	5	5	5	5	5	5	10	5	5	5
Schools, Colleges & Universities	5	5	5	5	5	5	10	5	5	5

Bibliography

Albuquerque, R., Koskinen, Y., Yang, S., Zhang, C., 2020. Resiliency of Environmental and Social Stocks: An Analysis of the Exogenous COVID-19 Market Crash. The Review of Corporate Finance Studies 9, 593–621.

Albuquerque, R., Koskinen, Y., Zhang, C., 2019. Corporate Social Responsibility and Firm Risk: Theory and Empirical Evidence. Management Science 65, 4451–4469.

Apergis, N., Poufinas, T., Antonopoulos, A., 2022. ESG scores and cost of debt. Energy Economics 112, 106186.

Avramov, D., Cheng, S., Lioui, A., Tarelli, A., 2022. Sustainable investing with ESG rating uncertainty. Journal of Financial Economics 145, 642–664.

Berg, F., Fabisik, K., Sautner, Z., 2020. Rewriting History II: The (Un)Predictable Past of ESG Ratings. SSRN Electronic Journal .

Berg, F., Kölbel, J. F., Rigobon, R., 2022. Aggregate Confusion: The Divergence of ESG Ratings. Review of Finance 26, 1315–1344.

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

Boffo, R., Patalano, R., 2020. ESG Investing: Practices, Progress and Challenges. https://www.oecd.org/finance/ESG-Investing-Practices-Progress-Challenges.pdf.

Bolton, P., Kacperczyk, M., 2021. Do investors care about carbon risk? Journal of Financial Economics 142, 517–549.

Broadstock, D. C., Chan, K., Cheng, L. T. W., Wang, X., 2021. The role of ESG performance during times of financial crisis: Evidence from COVID-19 in China. Finance research letters 38, 101716.

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

Consolandi, C., Eccles, R. G., Gabbi, G., 2022. How material is a material issue? Stock returns and the financial relevance and financial intensity of ESG materiality. Journal of Sustainable Finance & Investment 12, 1045–1068.

Dimson, E., Karakaş, O., Li, X., 2015. Active Ownership. The Review of Financial Studies 28, 3225–3268.

Dobrick, J., Klein, C., Zwergel, B., 2023. Size bias in refinitiv ESG data. Finance Research Letters 55, 104014.

Douglas, E., van Holt, T., Whelan, T., 2017. Responsible Investing: Guide to ESG Data Providers and Relevant Trends. Journal of Environmental Investing 8, 92–114.

Drempetic, S., Klein, C., Zwergel, B., 2020. The Influence of Firm Size on the ESG Score: Corporate Sustainability Ratings Under Review. Journal of Business Ethics 167, 333–360.

Dumrose, M., Rink, S., Eckert, J., 2022. Disaggregating confusion? The EU Taxonomy and its relation to ESG rating. Finance Research Letters 48, 102928.

Duong, H. N., Kalev, P. S., Kalimipalli, M., Trivedi, S., 2022. Do Firms Benefit from Carbon Risk Management? Evidence from the Credit Default Swaps Market. Working Paper .

Easton, P. D., Monahan, S. J., 2005. An Evaluation of Accounting-Based Measures of Expected Returns. The Accounting Review 80, 501–538.

Eccles, R. G., Krzus, M. P., Rogers, J., Serafeim, G., 2012. The Need for Sector-Specific Materiality and Sustainability Reporting Standards. Journal of Applied Corporate Finance 24, 65–71.

Eccles, R. G., Stroehle, J. C., 2018. Exploring Social Origins in the Construction of ESG Measures.

Edmans, A., 2011. Does the stock market fully value intangibles? Employee satisfaction and equity prices. Journal of Financial Economics 101, 621–640.

El Ghoul, S., Guedhami, O., Kwok, C. C. Y., Mishra, D. R., 2011. Does corporate social responsibility affect the cost of capital? Journal of Banking & Finance 35, 2388–2406.

ESMA, 2021. ESMA letter to EC on ESG ratings.

Fama, E. F., 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance 25, 383–417.

Fama, E. F., French, K. R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116, 1–22.

Friede, G., Busch, T., Bassen, A., 2015. ESG and financial performance: Aggregated evidence from more than 2000 empirical studies. Journal of Sustainable Finance & Investment 5, 210–233.

Gan, X., Fernandez, I. C., Guo, J., Wilson, M., Zhao, Y., Zhou, B., Wu, J., 2017. When to use what: Methods for weighting and aggregating sustainability indicators. Ecological Indicators 81, 491–502.

Gebhardt, W. R., Lee, C. M. C., Swaminathan, B., 2001. Toward an Implied Cost of Capital. Journal of Accounting Research 39, 135–176.

Gibson, B. R., Krueger, P., Schmidt, P. S., 2021. ESG Rating Disagreement and Stock Returns. Financial Analysts Journal 77, 104–127.

Grewal, J., Hauptmann, C., Serafeim, G., 2021. Material Sustainability Information and Stock Price Informativeness. Journal of Business Ethics 171, 513–544.

Guay, W., Kothari, SP., Shu, S., 2011. Properties of implied cost of capital using analysts' forecasts. Australian Journal of Management 36, 125–149.

Hartzmark, S. M., Sussman, A. B., 2019. Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows. The Journal of Finance 74, 2789–2837.

Heeb, F., Kölbel, J. F., Paetzold, F., Zeisberger, S., 2023. Do Investors Care about Impact? The Review of Financial Studies 36, 1737–1787.

Hong, H., Kacperczyk, M., 2009. The price of sin The effects of social norms on markets. Journal of Financial Economics .

Hübel, B., Scholz, H., 2020. Integrating sustainability risks in asset management: The role of ESG exposures and ESG ratings. Journal of Asset Management 21, 52–69.

Ilhan, E., Krueger, P., Sautner, Z., Starks, L. T., 2023. Climate Risk Disclosure and Institutional Investors. The Review of Financial Studies 36, 2617–2650.

Khan, M., Serafeim, G., Yoon, A., 2016. Corporate Sustainability: First Evidence on Materiality. The Accounting Review 91, 1697–1724.

Palmieri, E., Ferilli, G. B., Stefanelli, V., Geretto, E. F., Polato, M., 2023. Assessing the influence of ESG score, industry, and stock index on firm default risk: A sustainable bank lending perspective. Finance Research Letters 57, 104274.

Pástor, Ľ., Stambaugh, R. F., Taylor, L. A., 2022. Dissecting green returns. Journal of Financial Economics 146, 403–424.

Pedersen, L. H., Fitzgibbons, S., Pomorski, L., 2021. Responsible investing: The ESG-efficient frontier. Journal of Financial Economics 142, 572–597.

Refinitiv, 2023. Environmental, social and governance scores from LSEG. LSEG .

Schiehll, E., Kolahgar, S., 2021. Financial materiality in the informativeness of sustainability reporting. Business Strategy and the Environment 30, 840–855.

WEF, 2019. Seeking return on ESG: Advancing the reporting ecosystem to unlock impact for business and society (produced in collaboration with Allianz SE and Boston Consulting Group). https://www3.weforum.org/docs/WEF_ESG_Report_digital_pages.pdf.

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 .

Wong, W. C., Batten, J. A., Ahmad, A. H., Mohamed-Arshad, S. B., Nordin, S., Adzis, A. A., 2021. Does ESG certification add firm value? Finance Research Letters 39, 101593.

Chapter 3

ESG data characteristics and stock price differentiation

Abstract: This study analyses how sustainability data characteristics affect stock price differentiation. Rating providers collect binary and numerical indicators to merge them into an ESG rating. Based on Refinitiv data, we construct a boolean and a numerical ESG score to test which indicators lead to a better risk assessment. The results show that both constructed ESG scores differentiate stocks by returns, but this effect is more pronounced for the boolean score. Using alternative factor models, sub-periods, and portfolio rebalancing methods, the findings remain robust. This suggests that simplified sustainability indicators could support investors in their investment decisions.

Keywords: ESG ratings, numerical ESG data, boolean ESG data
JEL Classification: M14, G24

3.1 Introduction

In 2023, a total of 5,391 PRI (Principles for Responsible Investment) signatories with over 120 trillion U.S. Dollar in assets under management committed themselves to integrating sustainable metrics in their decision-making processes (PRI, 2023). This demonstrates that investments according to responsible guidelines are not a temporary trend but rather a complementary strategy for financial market participants. Besides their own assessments of ESG (Environmental, Social, and Governance) risks, investors orient themselves on the sustainability measures of external rating agencies. The service provided by such agencies consists of collecting, processing, and bundling sustainable information (Berg et al., 2022). In order to construct those ratings, the providers require a large amount of ESG data. This offering is still in the process of development, which is why some sustainability dimensions cannot yet be mapped adequately. An efficient way to overcome this problem is to express important ESG issues only as binary data points. For example, the question of whether a company has a policy on emissions' reduction could be answered easily. However, this characteristic says nothing about the actual degree of implementation on the company level.

Refinitiv as one of the three major rating providers (Billio et al., 2021), uses those binary indicators and supplements them with numerical data for assessing the sustainable engagement of a company. For example, the binary data point "TR.PolicyEmissions" gives an impression of whether the company is endeavouring to reduce its emissions. But, it does not provide any precise evidence of the company's involvement. In contrast, the measure "TR.AnalyticCO2" gives further insights about this behaviour by measuring the sum of CO2 emissions of a company. This value could be used to compare the company's CO2 emissions with the previous years or its industry peers. Thus, this kind of data provides better insights about a company's actual sustainability efforts or developments and could be utilized to determine its inherent ESG risk exposure. However, empirical research demonstrates, that investors ignore highly detailed sustainability information and instead orient themselves on simplified indicators (Hartzmark and Sussman, 2019). This raises the question, which kind of indicators could better reflect the sustainability risks of a company?

In this study, I test which type of ESG indicators leads to better risk and return assessments by measuring the relationship between financial performance and sustainability metrics. Therefore, I construct two alternative ESG scores from the Refinitiv data: a boolean and a numerical ESG score. The boolean ESG score consists exclusively of binary data points, while the numerical ESG score is based solely on numerical data points. For both measures, I calculate the ESG scores according to the original Refinitiv methodology.

For testing the ability of both self-constructed scores, the analysis follows the efficient market hypothesis. According to its semi-strong form, all publicly available information is reflected in stock prices (Fama, 1970). This therefore applies to all sustainability information in the form of ESG scores, which should be price-relevant for financial market participants. Since ESG scores should signal a systematic risk for an inherent company, stock prices reflect that risk. In other words, the increased sustainability risk of companies with low ESG scores is reflected in a price premium, and vice versa (e.g., El Ghoul et al., 2011). To measure the informative value of the scores, I sort the stocks into decile (quintile) portfolios depending on their ESG scores and compare the unexplained returns (Alpha) of the bottom (very unsustainable) and top (very sustainable) portfolios. A larger return difference in Alphas should signal a better risk assessment in terms of sustainability (Khan et al., 2016).

The study contributes to previous research in several ways. First, the results show that both constructed sustainability measures could differentiate companies based on stock returns. Stocks with the lowest ESG scores exhibit significantly higher annual risk premiums compared to stocks with higher ratings. Hence, they are appropriate metrics to measure the risk premiums of very unsustainable companies.

Second, I find that this differentiation is more pronounced for the boolean ESG score. Using alternative factor models, time periods, and portfolio rebalancing methodologies as robustness tests, the results show a larger difference in Alphas for the binary score. This result is particularly noticeable if the extraordinary effects of the COVID-19 crisis are excluded. In summary, this finding indicates that simplified sustainability indicators could support financial market participants in their investment decisions.

The remainder of this study proceeds as follows: Section 3.2 reviews the related literature. Section 3.3 describes the sample and introduces the methodology. Section 3.4 presents the main results, while Section 3.5 provides additional robustness checks. Section 3.6 concludes.

3.2 Related literature

The initial idea of ESG ratings is that they measure the degree of how financially vulnerable a company is to so-called sustainability risks. This leads to the understanding that sustainable investments are a financial hedge (Albuquerque et al., 2020). For example, Broadstock et al. (2021) show that stocks with high sustainable performance are more resilient during times of crisis and ESG engagement by shareholders could substantially reduce the downside risk of companies (Hoepner et al., 2024). Moreover, research pronounces the reduced systematic risk of superior sustainable companies (Albuquerque et al., 2019), which leads to lower cost of equity (El Ghoul et al., 2011) or cost of debt (Apergis et al., 2022). In contrast, investing in so-called sin industries increases the inherent risk and, consequently, the corresponding return of a company (Hong and Kacperczyk, 2009).

Institutional and retail investors care about sustainability and guide their fund flows according to this attribute (Hartzmark and Sussman, 2019). This trend is reinforced by regulation as, for example, the E.U. taxonomy. Becker et al. (2022) find that E.U.-based funds, that are affected by the Sustainable Finance Disclosure Regulation (SFDR) increase their ESG metrics. As a consequence, the enhanced ESG ratings lead to larger fund net inflows. The SFDR marks ESG-related funds, but does not directly show if they are impact-related or not (Scheitza and Busch, 2024). In addition, behavioural experiments show that social characteristics of investors determine the preference for socially responsible investments (Riedl and Smeets, 2017). In order to identify such sustainable investments, investors mainly focus on certification. Consequently, Gutsche and Ziegler (2019) find that the willingness to pay more is higher for ESG-certificated assets than for their non-certificated counterparts. Investors are therefore generally willing to pay for sustainability, but they do not increase their payments to create more actual impact (Heeb et al., 2023; Barber et al., 2021).

Since ratings, like certificates, are intended to simplify the selection process for sustainable products, it is important that they also correctly reflect the desired risk-return relationship. As a consequence, it is an essential task for ratings to be able to distinguish between sustainable and non-sustainable companies to guide capital flows accordingly (Chatterji et al., 2016). However, missing regulation (ESMA, 2021) leads to different methodologies to measure sustainability. In this context, Billio et al. (2021) show that there is a lack of fully accepted metrics in terms of ESG characteristics, attributes, and standards. In addition, the risk-adjusted industry weighting approaches used for the different sustainability categories to calculate the aggregated ESG ratings are also not standardized and result in risks being better or worse reflected in the ratings (Khan et al., 2016; Muck and Schmidl, 2024). This leads to rating disagreement across agencies (Berg et al., 2022). But even within the universe of the individual rating provider, there are some contradictions. For example, Refinitiv ESG ratings exhibit a size bias. To be more precise, Drempetic et al. (2020) and Dobrick et al. (2023) demonstrate that larger companies with more financial resources and reporting activities systematically have higher ESG scores than their smaller peers.

The study supplements recent sustainable finance literature that contributes to the aim of constructing better sustainability metrics (e.g., Dumrose et al., 2022). Besides research about the construction of a particular ESG score, this study analyses, how the different emphases (binary or numerical) of data points affect the differentiation of stocks in terms of returns. This study tests, whether the limited informational content of binary data points impairs the signal of ESG scores and whether the numerical components enhance this effect.

3.3 Data and methodology

To test whether the boolean or the numerical ESG score leads to an enhanced risk differentiation in terms of stock returns, I use a sample consisting of U.S.-based companies. The examination period begins in January 2011 and ends in June 2023. As a result, the sample consists of 23,551 company-year observations. Price and ESG data are obtained through the Refinitiv database, while the Fama and French (2015) factors are taken from Kenneth R. French's database.

I construct the two sustainability measures pursuant to the original Refinitiv (2023) percentile rank scoring methodology. According to the best-in-industry approach, the following ranking procedure is dependent on the TRBC (The Refinitiv Business Classification) industry group. The values of all companies within an industry for a specific data point are ranked. This means that the best company is ranked first and the worst company last. With the percentile rank formula, those rankings are translated into individual data point scores. They are calculated as follows:

$$Score_{d,i} = \frac{no.\ worse + \frac{no.\ same}{2}}{no.\ value}$$

where $Score_{d,i}$ is company i's percentile score of data point d. Additionally, $no.worse$ denotes the number of companies with a worse value, $no.same$ denotes the number of companies with the same value, and $no.value$ denotes all companies with a value for a certain data point. If there is no company with the same value, $no.same$ is set to 1. The intuitive understanding of the score is that the better the company's value of a data point is compared to its industry peer group, the higher the score is.

As a consequence, every sustainability data point has an indicative score. I divide the data point scores into two groups: the numerical and boolean measures. Depending on their characteristics, the individual data points are assigned to two different categories. For example, the boolean data point "TR.EmissionsTrading" is utilized to calculate the boolean emission category score, while the numerical measure "TR.AnalyticCO2" is part of the numerical emission category score. To calculate the final category scores $C^n_{c,i}$ and $C^b_{c,i}$, I sum the percentile scores of a company i and rank them. The first rank is assigned to the highest value and the last rank is assigned to the lowest value. After that, I again translate that ranking with the percentile formula into a score between 0 and 100. To further calculate the final numerical and boolean ESG scores, I follow:

$$NumericalESG_i = \sum_{c=1}^{N} C^n_{c,i} \cdot w_{c,j} \qquad BooleanESG_i = \sum_{c=1}^{N} C^b_{c,i} \cdot w_{c,j}$$

where $C^n_{c,i}$ is company i's numerical category score and $C^b_{c,i}$ is company i's boolean category score for category $c \in \{1, ..., N\}$, and $w_{c,j}$ is the industry weighting for category c of the industry group j.[14]

To test whether the boolean or the numerical ESG score leads to an enhanced risk differentiation, I follow Khan et al. (2016) by allocating the companies into equally-weighted decile and

[14]See 3.7 for a comprehensive calculation example.

quintile portfolios based on their ESG metrics. Companies with the worst scores are located in the bottom portfolios, and companies characterized by the best scores are assigned to the top portfolios, respectively. Since Refinitiv updates its sustainability measures on a yearly basis, I rebalance the portfolios at the beginning of the fiscal year. To estimate excess stock returns (i.e., Alpha), I utilize the Fama and French (2015) five-factor model. It captures company characteristics with respect to the market (MKT), size (SMB), book-to-market (HML), profitability (RMW), and investment (CMA).

3.4 Results

Table 3.1 shows the summary statistics. The standard deviations of the original Refinitiv and the boolean score are almost identical, while the mean deviates from the original measure. The workforce category with its high degree of numerical data points exhibits the largest difference of the mean. Consequently, this category has the highest average values in the numerical score. The numerical score has substantially lower means in comparison to the other two measures, because of a high degree of missing values. Since the differentiation of companies is mainly driven by the distribution of the scores, the standard deviations could give first insights. Despite the fact that the characteristic of the numerical metric would indicate that there should be a greater differentiation, I find lower standard deviations in this case.

Table 3.2 displays the mean and the standard deviation of the ESG scores by industry. In line with former findings, means of the boolean ESG score are slightly lower or similar, while the means of the numerical ESG score are much smaller than the original Refinitiv measure. For example, consumer cyclicals and consumer non-cyclicals exhibit the highest differences in means. In contrast, the boolean ESG score for basic materials and industrials is less affected. I find similar results for the standard deviations across industry groups.

Table 3.3 displays coefficient estimates and t-statistics of the Fama and French (2015) five-factor model for the decile and quintile portfolios formed by the boolean and numerical ESG scores. Panel A shows the results of the boolean ESG score. I find a significant annualized Alpha of 9.85 percent (p-value < 0.01) for the bottom decile portfolio and 6.21 percent

Table 3.1: Descriptive statistics: ESG scores and ESG pillars

	Obs.	Mean	Std. Dev.	Minimum	P25	Median	P75	Maximum
Original score								
$RESU$	23,551	26.68	32.44	0.00	0.00	8.33	51.61	99.91
$EMIS$	23,551	25.92	31.20	0.00	0.00	10.17	48.71	99.88
$ENVI$	23,551	16.11	26.88	0.00	0.00	0.00	27.83	99.50
$WORK$	23,551	41.55	26.40	0.12	19.38	37.65	61.40	99.90
$HUMR$	23,551	19.85	30.14	0.00	0.00	0.00	38.89	99.43
$COMM$	23,551	60.32	24.13	0.00	43.40	60.64	80.95	99.94
$PROD$	23,551	38.93	26.88	0.00	19.83	33.95	57.44	99.84
ESG	23,551	39.17	19.31	0.44	24.16	35.60	52.18	95.16
Boolean score								
$RESU$	23,551	24.78	31.62	0.00	0.00	4.96	47.83	99.82
$EMIS$	23,551	23.99	31.06	0.00	0.00	5.65	45.45	99.79
$ENVI$	23,551	15.07	26.22	0.00	0.00	0.00	24.61	99.40
$WORK$	23,551	32.16	29.31	0.00	5.74	25.31	53.51	99.83
$HUMR$	23,551	19.03	29.48	0.00	0.00	0.00	36.27	99.25
$COMM$	23,551	55.29	26.97	0.00	39.70	58.11	76.90	99.78
$PROD$	23,551	37.73	27.11	0.00	18.13	32.59	54.63	99.57
ESG	23,551	36.66	19.36	0.05	21.61	32.93	49.46	93.49
Numeric score								
$RESU$	23,551	13.68	26.89	0.00	0.00	0.00	9.78	99.76
$EMIS$	23,551	14.28	25.84	0.00	0.00	0.00	18.39	99.83
$ENVI$	23,551	1.10	7.85	0.00	0.00	0.00	0.00	99.54
$WORK$	23,551	36.61	26.00	0.00	15.22	33.44	55.08	99.91
$HUMR$	23,551	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$COMM$	23,551	13.25	27.74	0.00	0.00	0.00	0.00	99.76
$PROD$	23,551	2.27	11.64	0.00	0.00	0.00	0.00	98.08
ESG	23,551	23.62	13.49	0.06	13.43	21.46	31.45	76.83

Notes: This table shows summary statistics of the original, boolean, and numerical ESG scores (ESG). $RESU$ is the resource use category score, $EMIS$ the emissions category score, $ENVI$ the environmental innovation category score, $WORK$ the workforce category score, $HUMR$ the human rights category score, $COMM$ the community category score, $PROD$ the product responsibility category score. The mean, standard deviation, minimum, 25% quantile, median, 75% quantile, and maximum are reported.

Table 3.2: Descriptive statistics: ESG scores by industry

	Obs.	ESG	SD	ESG^{Bol}	SD^{Bol}	Diff.	ESG^{Num}	SD^{Num}	Diff.
Academic & Educational Services	103	33.19	14.51	31.12	14.20	2.07	22.48	11.69	**10.70**
Basic Materials	1,278	42.25	20.73	41.12	21.51	1.13	26.18	14.99	**16.07**
Consumer Cyclicals	3,489	39.52	18.93	35.18	18.61	**7.21**	19.62	11.77	**19.90**
Consumer Non-Cyclicals	1,200	46.73	24.14	43.36	24.57	**3.37**	25.16	14.99	**21.57**
Energy	1,518	37.01	20.58	34.64	20.27	**2.37**	23.18	13.44	**13.83**
Financials	3,871	37.62	15.82	35.49	16.55	**2.14**	26.40	12.21	**11.23**
Healthcare	3,494	34.61	18.54	32.32	18.34	**2.29**	20.75	12.28	**13.86**
Industrials	2,916	39.28	18.81	37.88	18.80	**1.41**	23.25	13.10	**16.04**
Real Estate	1,662	39.33	20.01	37.11	19.94	**2.22**	25.59	14.95	**13.74**
Technology	3,214	40.33	19.43	37.55	19.08	**2.77**	23.22	13.50	**17.11**
Utilities	806	48.31	19.08	46.21	19.99	2.09	33.42	15.22	**14.89**
Total	23,551	39.17	19.31	36.66	19.36	**2.51**	23.62	13.49	**15.56**

Notes: This table presents the differences between the original Refinitiv, boolean, and the numerical ESG scores across industries. The number of observations, the mean, the standard deviation of all scores, and the average differences of the means are reported. ESG^{Bol} is the difference between the original and the Boolean score and $ESG^{N}um$ is the difference between the original and numeric score, respectively. Bold coefficients are significant at the 1% level.

(p-value < 0.01) for the quintile portfolio. The Alpha estimates for both top portfolios are statistically insignificant. The difference in Alphas ($\Delta Alpha$) between the bottom and top portfolios is 11.59 percent per year for the decile and 7.18 percent for the quintile portfolio. Panel B uses the numerical ESG score. In line with the results of the boolean ESG score, both bottom portfolios are significant while the top portfolios remain insignificant. The bottom decile portfolio yields an annualized excess return of 6.44 percent (p-value < 0.05) and the quintile portfolio 5.32 percent (p-value < 0.05). The reported Alpha differences of the numerical portfolios are smaller than the boolean ones.

Despite the fact that binary indicators are criticized for their reduced information content, the results show that, in aggregated form, they are suitable. They could differentiate companies on stock returns, which means that they reflect the sustainability risk profile of the corresponding company. One possibility for the adequate assessment of the simplified indicators is that sustainable investment strategies follow negative or positive screening approaches in their decision-making. This means, for example, that they include or exclude stocks of their portfolios based on the companies' will to reduce their emissions, but not necessarily define the degree of the implementation.

3.5 Robustness

Table 3.4 presents a battery of robustness tests. Panel A displays the results of the boolean ESG score, and Panel B of the numerical ESG score. To verify the robustness of the findings, I follow (Khan et al., 2016) and use alternative factor models. I estimate the annualized Alphas through the Fama and French (1993) three-factor model which excludes the profitability and the investment factor and a six-factor model including an additional momentum factor (Carhart, 1997). The results remain constant. In Panel A, the Alpha differences in the decile portfolios are 9.82 percent (three-factor model) and 10.98 (six-factor model), respectively. In comparison, the numerical ESG decile portfolios exhibit annualized Alpha differences of 5.36 percent (three-factor model) and 7.21 percent. Using alternative factor models, the same tendency can be seen for the quintile portfolios. As a consequence, the boolean ESG score reflects an enhanced differentiation according to financial returns. The study period spans

Table 3.3: Regression results: performance differences

	Deciles				Quintiles			
	Bottom		Top		Bottom		Top	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Panel A: Boolean ESG score ($N = 150$)								
MKT	0.9627***	(19.0057)	1.0018***	(31.3039)	1.0098***	(23.8708)	1.0222***	(40.4151)
SMB	0.9763***	(10.1232)	0.1936***	(3.1766)	0.8894***	(11.0432)	0.3060***	(6.3550)
HML	0.0442	(0.5161)	0.2324***	(4.2995)	0.0027	(0.0383)	0.2266***	(5.3030)
RMW	-0.3646***	(-3.1903)	0.0878	(1.2165)	-0.3260***	(-3.4155)	0.1563***	(2.7389)
CMA	-0.0567	(-0.4416)	-0.0339	(-0.4177)	-0.0298	(-0.2780)	-0.0397	(-0.6190)
$Intercept$	0.7860***	(3.7807)	-0.1462	(-1.1128)	0.5029***	(2.8966)	-0.0809	(-0.7791)
R^2	0.8556		0.9055		0.8935		0.9438	
Annualized Alpha	9.85***		-1.74		6.21***		-0.97	
Δ Alpha				**11.59******				**7.18******
Panel B: Numeric ESG score ($N = 150$)								
MKT	0.9949***	(20.1266)	0.9545***	(28.7305)	1.0677***	(24.5816)	0.9929***	(40.3531)
SMB	0.8786***	(9.3355)	0.2175***	(3.4388)	0.9101***	(11.0059)	0.3340***	(7.1307)
HML	-0.0092	(-0.1100)	0.2705***	(4.8196)	-0.0399	(-0.5445)	0.2839***	(6.8298)
RMW	-0.3301***	(-2.9589)	0.0655	(0.8732)	-0.2978***	(-3.0380)	0.1032*	(1.8596)
CMA	-0.0387	(-0.3088)	-0.1329	(-1.5779)	0.0324	(0.2945)	-0.0943	(-1.5124)
$Intercept$	0.5216**	(2.5708)	-0.0087	(-0.0635)	0.4329**	(2.4286)	-0.0311	(-0.3079)
R^2	0.8568		0.8941		0.8951		0.9462	
Annualized Alpha	6.44**		-0.10		5.32**		-0.37	
Δ Alpha				**6.54****				**5.69****

Notes: This table shows coefficient estimates and t-statistics of the Refinitiv boolean and numerical ESG regressions. All values are presented in percent. The portfolios are sorted by ESG scores. Companies with the worst scores are in the bottom portfolio, and companies characterized by the best scores are assigned to the top portfolio, respectively. Those portfolios are rebalanced on a yearly basis. The dependent variable is the equal-weighted monthly return of the ESG portfolios minus the risk-free rate. Independent variables are the Fama and French (2015) factors: MKT, SMB, HML, RMW, CMA. The results are based on the monthly average returns between January 2011 and June 2023. The difference in annualized Alphas reflects the returns of a long-short portfolio, with long positions in stocks with low ESG ratings and short positions in high ESG stocks. ***, ** and * reflect significance at the 1%, 5%, and 10% level.

Table 3.4: Robustness tests

	Deciles			Quintiles		
	Annualized Alpha			Annualized Alpha		
	Bottom	Top	Difference	Bottom	Top	Difference
Panel A: Boolean ESG score						
Alternative factor models						
Three-factor Alpha	8.22***	-1.60	9.82***	5.16**	-0.62	5.78 **
Six-factor Alpha	9.92***	-1.06	10.98***	6.57***	-0.33	6.90***
Alternative sub-periods						
Excluding COVID-19	7.68***	-3.40***	11.08***	5.70***	-2.59***	8.29***
Excluding Fin. Crisis	10.11***	-1.95	12.06***	6.18**	-0.83	6.41***
Excluding Both	8.43***	-3.69***	12.12***	5.96***	-2.38**	8.34***
Alternative rebalancing						
Monthly rebalancing	9.56***	-1.76	11.31 ***	6.21***	-0.97	7.18***
Panel B: Numeric ESG score						
Alternative factor models						
Three-factor Alpha	5.36**	-0.24	5.60**	4.52**	-0.30	4.82*
Six-factor Alpha	6.65***	0.56	7.21**	5.22**	0.17	5.39**
Alternative sub-periods						
Excluding COVID-19	4.26**	-1.63	5.89***	4.56***	-1.72*	6.28***
Excluding Fin. Crisis	6.72**	-0.09	6.81**	5.11**	-0.10	5.21**
Excluding Both	4.49**	-1.52	6.01***	4.44**	-1.31	5.75***
Alternative rebalancing						
Monthly rebalancing	6.44**	-0.10	6.54**	5.32**	-0.37	5.69**

Notes: This table shows the annualized Alpha differences of the bottom and top portfolios. All values are presented in percent. The portfolios are sorted by ESG scores. Companies with the worst scores are in the bottom portfolio, and companies characterized by the best scores are assigned to the top portfolio, respectively. Those portfolios are rebalanced on a yearly basis. The dependent variable is the equal-weighted monthly return of the ESG portfolios minus the risk-free rate. Independent variables are the Fama and French (2015) factors: MKT, SMB, HML, RMW, CMA. The results are based on the monthly average returns between January 2011 and June 2023. The difference in annualized alphas reflects the returns of a long-short portfolio, with long positions in stocks with low ESG ratings and short positions in high ESG stocks. Under alternative factor models, we report the Alpha differences of the Fama and French (1993) three-factor model and a six factor model consisting of five Fama and French (2015) factors and an additional Carhart (1997) momentum factor. Under alternative sub-periods, we report differences in Alpha for sub-periods. The excluding COVID-19 period starts in January 2011 and ends in February 2020. The excluding financial crisis period does not include the aftermath of the financial crisis and starts in January 2013 and ends in June 2023. The excluding both events period includes observations between January 2013 and February 2020. Under alternative rebalancing, we rebalance the portfolios on a monthly basis. ***, ** and * reflect significance at the 1%, 5%, and 10% level.

significant economic events, including the aftermath of the financial crisis and the COVID-19 pandemic. These events may have influenced ESG scores and stock returns in ways that are not accounted for in the analysis, leading to potentially skewed results. Hence, I follow Muck and Schmidl (2024) and exclude observations before January 2013 (aftermath of the financial crisis) and months since February 2020 (COVID-19), to separately analyse those sub-periods. In general, they confirm the main results. In a third examination, I exclude both events, resulting in highly significant Alphas for the top and bottom decile and quintile portfolios of the boolean ESG score. Consequently, by not taking into account the influence of COVID-19 pandemic, the results become even stronger and confirm the main findings. An important concern may be that, the rebalancing of portfolios on a yearly basis might not capture the dynamic nature of ESG scores and their impact on stock returns. ESG scores and their financial implications can change more frequently than annually, suggesting that a more granular rebalancing approach might be necessary. Therefore, the next test assesses the robustness of changing the portfolio rebalancing methodology on a monthly basis. The Alpha estimates and differences between the top and bottom portfolios are almost equal. The reason for this is that Refinitiv calculates ESG scores using data from annual disclosures. This means, that there is only a change when the financial statements are published. The minimal difference in the estimates can be explained by the fact that not every company publishes at fiscal year's end (Refinitiv, 2023).

3.6 Conclusion

Rating agencies are regarded as intermediaries, and their ESG ratings are intended to reflect certain signals to market participants. They are a useful tool to visualize a large amount of sustainable information in a single number. However, the market for sustainability metrics is an unregulated field with much disagreement and controversies about the measurement of sustainability.

The study contributes to the current discussion of constructing better measures. I analyse whether the characteristics of ESG data are important to adequately reflect the sustainability risk exposure of a company. Therefore, I construct a numerical and a boolean ESG score.

The results indicate that both are able to differentiate companies based on stock returns. However, this effect is more pronounced for the boolean ESG score. In addition, they are very suitable to measure the risk premiums of unsustainable companies. The results are robust after using alternative factor models, sub-periods, and rebalancing methods.

In summary, it is important to understand how individual characteristics of ESG indicators could impact the assessment of sustainability risks. This would enhance the transparency and facilitate the decision-making process for investors. Hence, the results could provide further insights for policymakers to establish an uniform ESG framework.

3.7 Appendix

Calculation example

The following calculation steps derive from the original Refinitiv (2023) ESG score methodology. In our approach, we deviate from this by considering the numerical and boolean data points separately. This means, that we construct two scores for the environmental (resource use, emissions, environmental innovation) and social (workforce, human rights, community, product responsibility) categories, resulting in two final ESG measures.

According to the original methodology, we use The Refinitiv Business Classifications (TRBC) industry group in which the company is located as a peer group. This benchmarking is adopted because sustainable topics are more material and comparable to similar business models. Hence, if a data point is not relevant for the industry, it is excluded from the calculation. In the following example, we use companies in the automobiles and auto parts industry. This industry group includes 128 companies for the year 2022.

Central ranking formula

The percentile score calculation formula is utilized for every data point and for the final construction of all category scores and the ESG controversies score. The score is calculated by adding the number of companies with a worse value to half of the number of companies with the same value and dividing this sum by the total number of companies. The key advantage of using this methodology is that it is not sensitive to outliers.

$$Score = \frac{no.\, worse + \frac{no.\, same}{2}}{no.\, value}$$

where:

$$\begin{aligned} no.\, worse &= \text{Number of companies with a worse value} \\ no.\, same &= \text{Number of companies with the same value} \\ & \quad\ \text{(including the observed one)} \\ no.\, value &= \text{Number of companies with a value} \end{aligned}$$

Calculation of numerical data points:

Beginning with the numerical indicators, we first rank the numerical values due to their polarity. Positive polarity means the higher, the better, and negative polarity means the lower, the better, respectively[15]. The exemplary data point "TR.AnalyticCO2" is characterized by a negative polarity, and consequently, the ranking begins with companies that disclose the lowest values in this indicator. Table 3.5 illustrates company "THULE.ST" with the smallest value of the indicator "TR.AnalyticCO2". 40 companies of the 128 disclose no values, hence, they are excluded from the following calculation. 87 companies have a worse value than company "THULE.ST" and if there is no additional company with the same value as the company, the number of companies with the same value is set to 1. Following this methodology, company "INCH.L" on the second rank has 86 companies with a worse value and, again, no company with the same value, so it is set to 1.

Table 3.5: Numeric calculation example 1

Description	**THULE.ST**	**INCH.L**	**TSLA.O**
Number of companies with a worse value	87	86	85
Number of companies with the same value	1	1	1
Number of companies with a value	88	88	88

This information is the basis for calculating the individual percentile score. Table 3.6 displays the following steps: First, every company is ranked based on their CO2 emissions. Beginning with company "THULE.ST" characterized by the lowest value of 3.6609016, descending to the highest value (1262.6737) of company "LEVE3.SA". Depending on the rank, the numbers of companies with a worse value and the same value are (like in Table 3.5) inserted into the percentile score formula. The resulting score illustrates companies sustainable endeavours for that indicator in relation to their industrial peer group. The interpretation of this score is intuitive in the way that a higher score means better performance in this explicit indicator. The same calculation steps are applied to all other data points in the category. Table 3.7 displays all indicators of company "DOMETIC.ST" that are assigned to the emissions' category. There is a corresponding percentile score for each data point with a value. Additionally, a

[15]The individual measure indicates the polarity, whether a higher value is positive or negative. A higher degree of recycled water is positive, but more emissions are negative.

Table 3.6: Numerical calculation example 2

Company name	Eikon code	Value	Rank	Formula	Score
THULE.ST	TR.AnalyticCo2	3.6609016	1	(87+(1/2))/88	.9943182
INCH.L	TR.AnalyticCo2	3.9870382	2	(86+(1/2))/88	.9829546
TSLA.O	TR.AnalyticCo2	7.488154	3	(85+(1/2))/88	.9715909
DOMETIC.ST	TR.AnalyticCo2	8.4551959	4	(84+(1/2))/88	.9602273
7276.T	TR.AnalyticCo2	9.0678941	5	(83+(1/2))/88	.9488636
...	...	...	...	...	...
APLO.NS	TR.AnalyticCo2	268.3257	84	(4+(1/2))/88	.0511364
SMPV.VI	TR.AnalyticCo2	393.20124	85	(3+(1/2))/88	.0397727
600660.SS	TR.AnalyticCo2	462.6623	86	(2+(1/2))/88	.0284091
601966.SS	TR.AnalyticCo2	489.53204	87	(1+(1/2))/88	.0170455
LEVE3.SA	TR.AnalyticCo2	1262.6737	88	(0+(1/2))/88	.0056818

default value of 0 is assumed for data points with no value. Finally, all percentile scores are summed up at the company level.

Table 3.7: Numerical calculation example 3

Company name	Eikon code	(Default) Value	Score
DOMETIC.ST	TR.AnalyticCO2	8.4551959	.9602273
DOMETIC.ST	TR.CO2IndirectScope3	8.514641	.9042553
DOMETIC.ST	TR.OZONE	0	0
DOMETIC.ST	TR.NOX	0	0
DOMETIC.ST	TR.SOX	0	0
DOMETIC.ST	TR.VOC	0	0
DOMETIC.ST	TR.AnalyticTotalWaste	4.4688752	.8092105
DOMETIC.ST	TR.AnalyticHazardousWaste	.06993545	.9590164
DOMETIC.ST	TR.SELFFINE	0	0
DOMETIC.ST	TR.WATERPOLEMISSIONS	0	0
DOMETIC.ST	TR.WASTE	.76	.2916667
DOMETIC.ST	TR.EMS	48	.1101695
Sum of all (percentile) scores			4.034545

Table 3.8 shows the last calculation step and the final numerical ESG score. All companies in the industry group are ranked (highest to lowest) according to their sum of percentile scores. To derive the final numerical emission category score, we again apply the percentile scores formula. Moreover, companies without data points, and consequently without a sum of percentile scores, receive the value 0 for their emission category score in our study. For purposes of presentation, we multiply the numerical ESG score by 100.

Table 3.8: Numerical calculation example 4

Company name	Sum of scores	Formula	Emission category score
MBGn.DE	7.517672	(107+(1/2))/108	.9953704
0175.HK	6.679072	(106+(1/2))/108	.9861111
XPEV.K	6.439848	(105+(1/2))/108	.9768519
000338.SZ	5.889572	(104+(1/2))/108	.9675926
...	...	...	...
DOMETIC.ST	4.034545	(88+(1/2))/108	.8194444
...	...	...	...
MEC	.059322	(1+(1/2))/108	.0138889
LEVE3.SA	.0163201	(0+(1/2))/108	.0046296
...	...	...	...
NKLA.O	-	-	0
SW10n.H	-	-	0

Calculation of boolean data points

The second form of indicators are boolean data points. This means that an indicator could have just two different emphases: 1 and 0. Table 3.10 illustrates a calculation example with the data point “TR.PolicyEmissions” which is characterized by a positive polarity. It shows whether the company has an emissions reduction efficiency policy or not. In the event that this is true (“Yes”), the indicator is 1, and otherwise (“No”) 0. Moreover, if the company does not disclose any information about this (“Null”), the default value is 0 (see Table 3.9). The majority of those measures in the Refinitiv universe have a positive polarity, meaning that 1 is associated with positivity and 0 is negative.

Table 3.9: Boolean calculation example 1

value		
Positive	“Yes” = 1	"No" / “Null” = 0
Negative	“Yes” / “Null” = 0	“No” = 1

Table 3.10 shows the transformation of a binary data point to a percentile score. We rank the companies according to the characteristic of the indicator “TR.PolicyEmissions”. All companies with a default value of 1 rank first, and companies with a default value are at the bottom of the ranking. In our example, 112 companies of 128 have a policy to reduce their emissions and 16 do not have one, respectively. After applying the percentile score

formula, this leads to a score for the companies with such a policy and a default value of 0 for the others. In principle, the more companies have a score of 1, the lower the score of these companies, and vice versa.

Table 3.10: Boolean calculation example 2

Company name	Eikon code	Value	Default value	Formula	Score
DOMETIC.ST	TR.PolicyEmissions	Yes	1	(16+(112/2))/128	.5625
600418.SS	TR.PolicyEmissions	Yes	1	(16+(112/2))/128	.5625
000581.SZ	TR.PolicyEmissions	Yes	1	(16+(112/2))/128	.5625
000800.SZ	TR.PolicyEmissions	Yes	1	(16+(112/2))/128	.5625
600660.SS	TR.PolicyEmissions	Yes	1	(16+(112/2))/128	.5625
0425.HK	TR.PolicyEmissions	Yes	1	(16+(112/2))/128	.5625
FTON.S	TR.PolicyEmissions	Yes	1	(16+(112/2))/128	.5625
...	...	...	...	...	...
002151.SZ	TR.PolicyEmissions	No	0	0	0
002151.SZ	TR.PolicyEmissions	No	0	0	0
FUV.O	TR.PolicyEmissions	No	0	0	0

The same calculation steps are applied to all other data points in the category. Table 3.11 shows all indicators for the company 600660.SS. There is a corresponding percentile score for each data point with the characteristic "Yes". For the rest of indicators, a 0 is displayed. Simultaneously to the calculation methodology of the numerical ESG score, we sum all percentile scores.

Table 3.11: Boolean calculation example 3

Company name	Eikon code	(Default) Value	Score
600660.SS	TR.PolicyEmissions	1	.5625
600660.SS	TR.TargetsEmissions	1	.6953125
600660.SS	TR.BiodiversityImpactReduction	1	.828125
600660.SS	TR.ClimateChangeRisksOpp	1	.6992188
600660.SS	TR.NOxSOxEmissionsReduction	1	.875
600660.SS	TR.TR.AnalyticVOCorPMReduction	1	.75
600660.SS	TR.eWasteReduction	0	0
600660.SS	TR.StaffTransportationReduction	0	0
600660.SS	TR.EnvironmentalExpendituresInves	1	.765625
600660.SS	TR.EnvPartnerships	0	0
600660.SS	TR.EnvRestorationInitiatives	0	0
600660.SS	TR.EmissionsTrading	0	0
Sum of all (percentile) scores			5.175781

Table 3.12 shows the last calculation step and the final boolean ESG score. All companies in the industry group are ranked (highest to lowest) according to their sum of percentile

scores. To derive the final boolean emission category score, we again apply the percentile scores formula. Moreover, companies without data points, and consequently without a sum of percentile scores, receive the value 0 for their emission category score in our study. For purposes of presentation, we multiply the numerical ESG score by 100.

Table 3.12: Boolean calculation example 4

Company name	Sum of scores	Formula	Emission category score
BMWG.DE	8.679688	(112+(1/2))/113	.9957265
RENA.PA	8.601563	(111+(1/2))/113	.9871795
7269.T	8.582031	(110+(1/2))/113	.9786325
MBGn.DE	7.917969	(109+(1/2))/113	.9700854
6902.T	7.867188	(108+(1/2))/113	.9615384
601633.SS	7.777344	(107+(1/2))/113	.9529914
...	...	...	...
600660.SS	5.175781	(78+(1/2))/113	.7051282
...	...	...	...
PSIX.PK	.5625	(0+(5/2))/113	.0213675
WKHS.O	.5625	(0+(5/2))/113	.0213675

Calculation of the final ESG metrics

The calculation of the final numerical and boolean ESG score is similar. After all the individual category scores have been calculated, they are multiplied by an industry weighting in the final step. The weighting should reflect the importance of the individual sustainability category for the industry. The industry weighting for the appropriate industry is taken from Refinitiv.

Numerical ESG score of "MBGn.DE"

$$NumericalESG_i = \sum_{c=1}^{N} C_{c,i}^{m} \cdot w_{c,j}$$

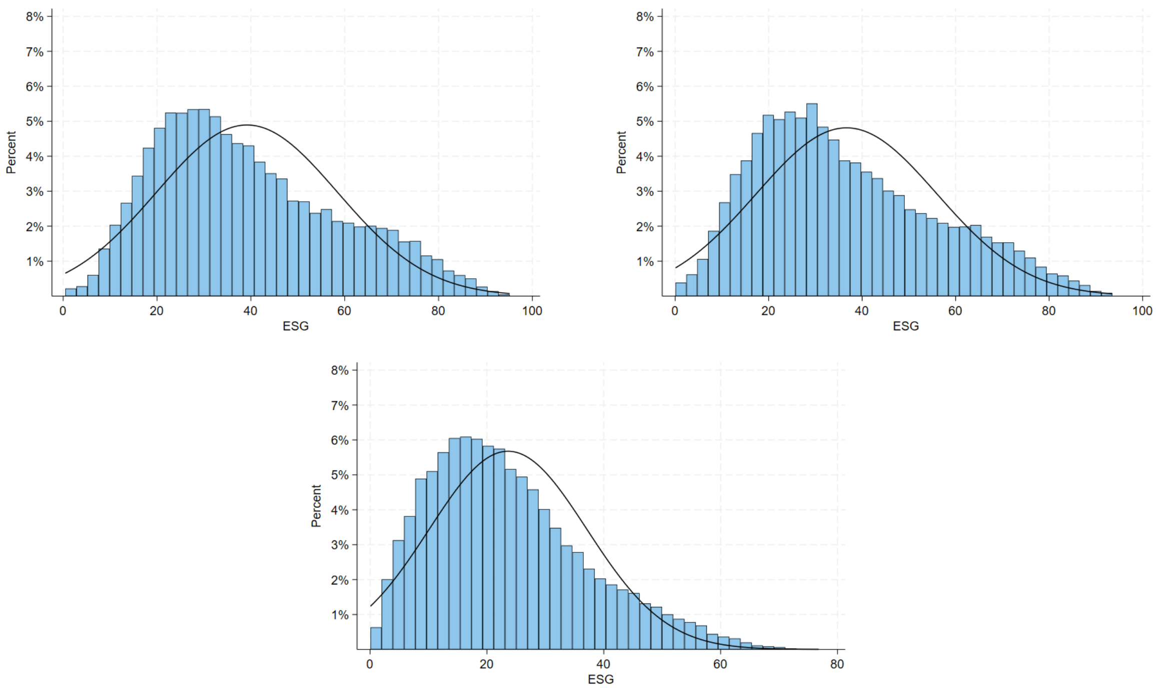
$$
\begin{aligned}
NumericalESG_{MBGn.DE} &= 0.081 \cdot 93.88889 + 0.097 \cdot 99.53704 + 0.162 \cdot 14.70588 \\
&+ 0.101 \cdot 97.26563 + 0.146 \cdot 0 + 0.081 \cdot 62.30769 + 0.087 \cdot 0 \\
&+ 0.162 \cdot 92.985612 + 0.049 \cdot 95.863309 + 0.032 \cdot 94.779116 \\
&= 57.3071
\end{aligned}
$$

Boolean ESG score of "MBGn.DE"

$$
BooleanESG_i = \sum_{c=1}^{N} C_{c,i}^{b} \cdot w_{c,j}
$$

$$
\begin{aligned}
BooleanESG_{MBGn.DE} &= 0.081 \cdot 94.95798 + 0.097 \cdot 97.00854 + 0.162 \cdot 77.02703 \\
&+ 0.101 \cdot 98.8189 + 0.146 \cdot 90.56604 + 0.081 \cdot 97.2 + 0.087 \cdot 64.34426 \\
&+ 0.162 \cdot 92.985612 + 0.049 \cdot 95.863309 + 0.032 \cdot 94.779116 \\
&= 89.04821
\end{aligned}
$$

Figure 3.1: Distribution of ESG scores



Notes: This figure shows the distribution of the weighted Refinitiv ESG score (left figure) the SASB risk-weighted ESG score (middle figure), and the naive-unweighted ESG score (right figure). The sample consists of 3,098 U.S. firms between February 2010 and December 2022 resulting in 18,352 company-year observations

Bibliography

Albuquerque, R., Koskinen, Y., Yang, S., Zhang, C., 2020. Resiliency of Environmental and Social Stocks: An Analysis of the Exogenous COVID-19 Market Crash. The Review of Corporate Finance Studies 9, 593–621.

Albuquerque, R., Koskinen, Y., Zhang, C., 2019. Corporate Social Responsibility and Firm Risk: Theory and Empirical Evidence. Management Science 65, 4451–4469.

Apergis, N., Poufinas, T., Antonopoulos, A., 2022. ESG scores and cost of debt. Energy Economics 112, 106186.

Barber, B. M., Morse, A., Yasuda, A., 2021. Impact investing. Journal of Financial Economics 139, 162–185.

Becker, M. G., Martin, F., Walter, A., 2022. The power of ESG transparency: The effect of the new SFDR sustainability labels on mutual funds and individual investors. Finance Research Letters 47, 102708.

Berg, F., Kölbel, J. F., Rigobon, R., 2022. Aggregate Confusion: The Divergence of ESG Ratings. Review of Finance 26, 1315–1344.

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

Broadstock, D. C., Chan, K., Cheng, L. T. W., Wang, X., 2021. The role of ESG performance during times of financial crisis: Evidence from COVID-19 in China. Finance research letters 38, 101716.

Carhart, M. M., 1997. On Persistence in Mutual Fund Performance. The Journal of Finance 52, 57–82.

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

Dobrick, J., Klein, C., Zwergel, B., 2023. Size bias in refinitiv ESG data. Finance Research Letters 55, 104014.

Drempetic, S., Klein, C., Zwergel, B., 2020. The Influence of Firm Size on the ESG Score: Corporate Sustainability Ratings Under Review. Journal of Business Ethics 167, 333–360.

Dumrose, M., Rink, S., Eckert, J., 2022. Disaggregating confusion? The EU Taxonomy and its relation to ESG rating. Finance Research Letters 48, 102928.

El Ghoul, S., Guedhami, O., Kwok, C. C. Y., Mishra, D. R., 2011. Does corporate social responsibility affect the cost of capital? Journal of Banking & Finance 35, 2388–2406.

ESMA, 2021. ESMA letter to EC on ESG ratings.

Fama, E. F., 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance 25, 383–417.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fama, E. F., French, K. R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116, 1–22.

Gutsche, G., Ziegler, A., 2019. Which private investors are willing to pay for sustainable investments? Empirical evidence from stated choice experiments. Journal of Banking & Finance 102, 193–214.

Hartzmark, S. M., Sussman, A. B., 2019. Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows. The Journal of Finance 74, 2789–2837.

Heeb, F., Kölbel, J. F., Paetzold, F., Zeisberger, S., 2023. Do Investors Care about Impact? The Review of Financial Studies 36, 1737–1787.

Hoepner, A. G. F., Oikonomou, I., Sautner, Z., Starks, L. T., Zhou, X. Y., 2024. ESG shareholder engagement and downside risk. Review of Finance 28, 483–510.

Hong, H., Kacperczyk, M., 2009. The price of sin The effects of social norms on markets. Journal of Financial Economics .

Khan, M., Serafeim, G., Yoon, A., 2016. Corporate Sustainability: First Evidence on Materiality. The Accounting Review 91, 1697–1724.

Muck, M., Schmidl, T., 2024. Comparing ESG score weighting approaches and stock performance differentiation. Finance Research Letters 67, 105924.

PRI, 2023. PRI Annual Report .

Refinitiv, 2023. Environmental, social and governance scores from LSEG. LSEG .

Riedl, A., Smeets, P., 2017. Why Do Investors Hold Socially Responsible Mutual Funds? The Journal of Finance 72, 2505–2550.

Scheitza, L., Busch, T., 2024. SFDR Article 9: Is it all about impact? Finance Research Letters 62, 105179.

Chapter 4

Relative to whom? The impact of peer groups on ESG ratings and financial performance

Abstract: This paper investigates how peer groups affect the financial informativeness of ESG ratings. Many ESG rating agencies evaluate firms relative to their industry or sector peers, but the consequences of this methodological choice remain largely anecdotal and have received little systematic empirical attention. Using the Refinitiv ESG framework (which applies a best-in-industry approach) we construct an alternative best-in-sector rating and assess their respective abilities to reflect sustainability-related financial risks. Our findings show that firms with low ESG ratings consistently earn higher abnormal returns than those with high ratings, consistent with the existence of a sustainability risk premium. While this pattern holds across both methodologies, the return spread is significantly larger under the best-in-sector approach. This suggests that broader peer groups are more effective in distinguishing firms with financially material sustainability risks.

Keywords: ESG ratings, ESG category scores, recalculation

JEL Classification: M14, G24

4.1 Introduction

According to the Global Sustainable Investment Alliance review, the demand for sustainable investment opportunities has increased by 20 percent in the last two years and reached a total amount of 30.3 trillion assets under management in 2022 (GSIA, 2022). This displays the fact that fund managers align their flows and investment strategies with environment, social, and governance (ESG) objectives (Hartzmark and Sussman, 2019). This growing demand has driven the importance of ESG ratings, which are widely used to measure firms' sustainability performance and financial risks.

A key feature of many ESG rating systems is their reliance on relative benchmarking. Firms are commonly evaluated against their industry or sector peers. While this approach controls for structural differences across industries, it may also lead to counter-intuitive outcomes. For example, superior ESG ratings for firms in environmentally harmful industries and low ratings for firms in sustainability-focused industries. Such divergences could affect investment decisions, capital allocation, and ultimately the cost of capital.

Although such inconsistencies are frequently cited in public discourse, they remain largely anecdotal and have received little systematic empirical attention. It remains unclear to what extent the peer group influences ESG ratings and whether this affects their informativeness for investors. This paper aims to close this gap by empirically investigating how peer groups shape the financial informativeness of ESG ratings. Specifically, we study Refinitiv's ESG rating methodology, which employs a best-in-industry approach. This means, that firms are evaluated relative to others in the same industry group. We then construct an alternative best-in-sector rating, in which firms are benchmarked against a broader set of peers within the same economic sector.

This peer group change allows us to examine whether ESG ratings based on broader peer groups offer more relevant signals to financial markets. In doing so, we compare the performance of U.S. stocks sorted by both rating types and analyse the return spread (Δ-Alpha) between firms with low and high ESG scores. If ESG ratings are informative, we expect stocks with poor sustainability engagement to earn a return premium that compensates for higher sustainability-related risk. The magnitude of this spread provides a measure of how well the rating distinguishes between financially relevant sustainability information.

Our results contribute to the existing literature in several ways. First, we examine the mechanisms through which peer groups affect ESG ratings under the Refinitiv methodology, in which final ratings are calculated as weighted averages of category-specific scores. Both the category scores and the associated weights are influenced by the composition of the peer group. On average, firms tend to receive more favourable scores when benchmarked against less sustainable peer groups. However, the effect on category scores is not necessarily uniform across all firms within an industry. Moreover, the impact of changing weights is theoretically ambiguous. As a result, the overall effect of the peer group on ESG ratings is ex ante unclear.

Second, we construct best-in-industry and best-in-sector ESG ratings following the Refinitiv (2023) methodology. Our recalculated best-in-industry scores closely match the original Refinitiv ratings, validating our approach. Nevertheless, small deviations arise due to data limitations and differences in peer group composition. In particular, we do not observe the full global universe of firms used by Refinitiv, nor do we have access to historical industry assignments. Additional discrepancies may stem from firm turnover (e.g., delistings, mergers) and potential refinements in the Refinitiv (2023) methodology over time. Despite these limitations, our analysis of U.S. firms shows that while some ratings remain unchanged across both methodologies, a substantial number of firms receive different ratings if the peer group is changed.

Third and most importantly, we examine the financial informativeness of ESG ratings by sorting firms into portfolios based on their best-in-industry and best-in-sector scores. We find that firms with low ESG ratings consistently earn higher abnormal returns than those with high ratings. This finding holds across both peer group definitions, indicating that both best-in-class measures contain financially relevant information. However, the return spread between the lowest and highest rated portfolios is significantly larger under the best-in-sector approach. This implies that broader peer groups are more effective in identifying financially material sustainability differences between firms. The results are robust across various model specifications, portfolio rebalancing strategies, and sample periods.

Our study is related to two strands of sustainable finance literature. The first strand focuses on the construction and evaluation of ESG measures. Previous research compares ESG rating providers (e.g., Billio et al., 2021) and documents significant disagreement among them (Chatterji et al., 2016), which is often based on the relative importance of different ESG attributes (Berg et al., 2022). Moreover, ESG ratings are subject to construction biases, as shown by Drempetic et al. (2020) and Dobrick et al. (2023) in the context of Refinitiv. While a growing number of studies aim to improve ESG measurement frameworks (e.g., Dumrose et al., 2022), little attention has been paid to the role of peer groups in score construction. To the best of our knowledge, we are the first to explicitly analyse the best-in-industry approach as implemented in the Refinitiv (2023) methodology.

The second related strand of literature addresses the relevance of sustainability in asset pricing. Prior research suggests that sustainable firms exhibit lower downside risk and greater resilience in times of market stress (e.g., Albuquerque et al., 2020), implying that ESG risks may be priced by financial market participants. Empirical evidence shows that sustainability-related information, including ESG scores, is reflected in stock returns (e.g., Hübel and Scholz, 2020). Our study contributes to this by examining how the construction of ESG measures, and specifically the peer groups, affects the financial informativeness and usefulness for investment decision-making.

The remainder of this study is structured as follows: Section 4.2 gives an overview of related literature. Section 4.3 addresses the construction of the Refinitiv rating and the difference between best-in-sector and best-in-industry scores. Section 4.4 shows the used data and methodology. Section 4.5 provides our empirical results and a discussion of implications, while Section 4.6 presents the robustness checks. The last section concludes.

4.2 Related literature

ESG considerations have become a central topic in academic and practical discussions on corporate finance and investment strategies due to a growing emphasis on sustainability from regulatory initiatives and investor changing preferences. In this context, ESG ratings play a crucial role in assessing corporate sustainability performance and informing investment decisions. However, the methodologies underlying these ratings, as well as their financial implications, remain widely debated. This section reviews the existing literature on sustainability and financial returns (Section 4.2.1), the ESG rating market (Section 4.2.2), and the methodological challenges associated with ESG ratings (Section 4.2.3).

4.2.1 Sustainability and financial returns

The significance of sustainability information has grown considerably in recent years. From a societal point of view, sustainability is widely acknowledged as a critical driver of long-term development, enhancing corporate resilience while mitigating adverse environmental and social impacts. In the context of investment decisions, the importance of sustainability-related data has also gained substantial attention. There is a growing body of literature, that examines the connection between firms' increasing sustainability engagement and their financial performance (Friede et al., 2015; Whelan et al., 2021). Specifically, sustainable firms are often perceived as less risky (e.g., Albuquerque et al., 2019; Hübel and Scholz, 2020). Moreover, being rated as sustainable also has a positive effect on fund inflows (Hartzmark and Sussman, 2019). But, motives for investors are diverse. For example, retail investors may deviate their decisions from positive emotions connected with the term sustainability (Heeb et al., 2023), independently of a financial risk-return consideration. Others view stocks with superior ESG performance as financial hedges, due to their enhanced resiliency during times of the financial crisis (Lins et al., 2017) or the COVID-19 pandemic (Albuquerque et al., 2020; Broadstock et al., 2021). As a result, sustainable firms may benefit from lower financing costs (El Ghoul et al., 2011).

4.2.2 The market of ESG ratings

In response to the growing demand for sustainability information, regulatory frameworks have introduced reporting obligations designed to enhance transparency and ensuring that firms provide comprehensive insights into their ESG practices. These regulations have led to an increasing emphasis on sustainability reporting across various countries, with regulatory bodies recognizing the necessity of clear standards to ensure the availability of reliable and comparable ESG data (ESMA, 2021). The primary objective of these reporting requirements is to foster transparency and provide investors with essential information to facilitate informed decision-making. ESG ratings, as combined indicators of environmental (e.g., Bolton and Kacperczyk, 2021) and social (e.g., Hong and Kacperczyk, 2009; Edmans, 2011) performance, play a crucial role in this context by consolidating complex and detailed sustainability information into a single, easily interpretable metric (Berg et al., 2022). This simplification allows financial market participants to efficiently assess a firm's sustainability performance.

The ESG rating market is highly diverse (Billio et al., 2021). Specialized agencies, such as the Carbon Disclosure Project (CDP) or RepRisk, focus on particular aspects of ESG spectrum. In contrast, larger providers like MSCI, Bloomberg, and Refinitiv combine their ESG data with investment tools to extend their services (Douglas et al., 2017). These providers generally rely on public sources to collect sustainability data, which they supplement with proprietary information, such as internal company questionnaires (e.g., RobecoSAM) or disclosures from non-profit organizations. Despite the fact that their data basis is very similar to a large extent, their approaches of processing and interpretation of this is in turn very different. Hence, rating agencies substantially disagree in their sustainable assessment (Chatterji et al., 2016; Berg et al., 2022). Notably, this divergence tends to increases with the volume of disclosed ESG information (Christensen et al., 2022), highlighting the inherent difficulties in objectively assessing ESG engagement of firms.

4.2.3 Methodological challenges in ESG ratings

ESG ratings have been critically examined in the literature. For example, Drempetic et al. (2020) and Dobrick et al. (2023) identify a size bias in Refinitiv's ESG ratings, where larger firms tend to receive higher ratings. Furthermore, Muck and Schmidl (2024) analyse the aggregation mechanism of the Refinitiv ESG rating methodology and find that the weighting of different ESG categories plays a crucial role in determining the financial informativeness of the final rating.

However, a fundamental concern in ESG rating methodologies is the widespread adoption of the best-in-industry approach. Under this framework, a firm's ESG performance is assessed by comparing it only to firms within the same industry, rather than using an objective benchmark to define sustainability. This method reflects the assumption that each industry may face unique sustainability risks, which could vary across industries (Eccles et al., 2012). As a result, firms in different industries disclose different types of ESG information, and failure to disclose certain data may negatively impact their ratings. Moreover, the best-in-industry methodology is believed to incentivize firms in lower ESG-performing industries to enhance sustainability efforts, as targeted investments can substantially improve their ratings.

Despite its theoretical justifications, the best-in-industry approach presents several limitations. A key drawback is that companies operating in industries with lower sustainability standards may receive higher ESG ratings than firms in inherently more sustainable industries. This undermines the comparability of ESG ratings across sectors, making direct comparisons of sustainability performance challenging. Additionally, this approach can lead to situations where firms obtain high ESG ratings, even though their industries as a whole might be excluded from sustainable investment portfolios due to misalignment with broader sustainability principles.

While there are anecdotal concerns regarding the best-in-industry approach, empirical evidence on its financial implications remains scarce. In particular, little is known about how this methodology affects the financial informativeness of ESG ratings. An ESG rating is considered financially informative if it enables investors to identify and assess the risks associated with investment opportunities. Despite theoretical concerns regarding the best-in-industry approach, its actual influence on investment decision-making has not yet been thoroughly analysed. This study empirically examines whether and to what extent this methodology introduces biases in financial decision-making, with a particular focus on the ESG rating model employed by Refinitiv.

4.3 The Refinitiv rating

This section outlines the methodological foundations of the Refinitiv (2023) ESG rating system, which follows a best-in-industry approach. Accordingly, our focus lies on understanding the role of the peer group in shaping ESG scores and ratings. In particular, we examine how the exclusion of peers influences individual scores and the overall rating outcome.

The section is structured as follows: Section 4.3.1 provides an overview of the ESG categories and associated letter grades. Section 4.3.2 describes the calculation methodology. Section 4.3.3 analyses the impact of peer groups on individual scores, while Section 4.3.4 investigates how these peer effects translate into differences in overall ESG ratings.

4.3.1 Rating categories

The Refinitiv (2023) ESG Rating is a numerical value ranging from 0 to 100, with higher scores indicating stronger ESG performance. These scores are subsequently translated into letter grades as shown in Table 4.1. The ESG rating is derived from ten underlying categories, grouped into three pillars: environmental, social, and governance.

The environmental pillar assesses a firm's resource use, emissions, and environmental innovation. The resource use ($RESU$) category evaluates factors such as energy efficiency, water consumption, and reliance on renewable energy sources. The emissions ($EMIS$) category captures sustainability metrics, including total CO2 emissions relative to revenue, NOx emissions, and corporate emission reduction policies. Environmental innovation ($ENVI$) reflects efforts to develop environmentally friendly products and investments in sustainable research and development.

The social pillar consists of four categories. Workforce ($WORK$) reflects diversity metrics, trade union representation, employee turnover rates, training programs, and workplace safety. Human rights ($HUMR$) considers corporate policies on fundamental labour rights, including freedom of association and the prohibition of forced or child labour. Community ($COMM$) focuses on a firm's business ethics, fair competition policies, anti-corruption measures, and community engagement initiatives. Product responsibility ($PROD$) assesses whether a company upholds high standards in customer health and safety, data privacy, and product accountability.

The governance pillar encompasses three categories. Management evaluates board composition, leadership skills, audit committee independence, and executive compensation relative to revenue. Shareholders examines shareholder rights, equal voting power, and corporate governance policies. Corporate social responsibility (CSR) strategy assesses whether a firm discloses its sustainability initiatives in alignment with established frameworks such as the

Global Reporting Initiative.

Each category score is derived from multiple underlying data points, which can be either numerical (e.g., emissions, water consumption, waste production) or Boolean (e.g., presence of an emission reduction policy, diversity policies). The ESG scores are calculated based on a firm's relative performance within its peer group, as outlined in the next section.

4.3.2 Calculation

The Refinitiv (2023) ESG score is a weighted average of all ten category scores:

$$ESG_i = \sum_{c=1}^{10} w_{c,j} C_{c,i} \tag{4.1}$$

where ESG_i is the ESG score of firm i, and $C_{c,i}$ and $w_{c,j}$ denotes the category score and category weight of category c for firm i within peer group j.

Category scores are determined relative to the peer group. In the case of Refinitiv, peer groups are defined as all firms within the same Thomson Reuters Business Classification (TRBC) industry group (best-in-industry approach) for the categories within the environmental and social pillars. For the governance pillar, the peer group comprises all firms headquartered in the same country. For each data point of a category, all firms within a peer group are ranked according to sustainability. For a numerical data point, such as total emissions, firms are ranked in ascending order, while for a Boolean data point, firms that meet a given sustainability criterion (score = 1) are ranked above those that do not meet the criterion or do not report the metric (score = 0). The score $Score_{d,i}$ for a data point d of firm i is computed as:

$$Score_{d,i} = \frac{no.\ worse + \frac{no.\ same}{2}}{no.\ value} \cdot 1_{eligible} \tag{4.2}$$

where $Score_{d,i}$ is firm i's percentile score for data point d. The variable $no.worse$ represents the number of firms with a worse sustainability performance for the data point, $no.same$ denotes the number of firms with identical performance, and $no.value$ is the total number of firms within the peer group reporting the given data point. Some firms are assigned a score of

Table 4.1: ESG score range and grade description

Score Range	Grade	Description
$0.0000 \leq \text{score} \leq 8.3333$	D-	'D' score indicates poor relative ESG performance and insufficient degree of transparency in reporting material ESG data publicly.
$8.3333 < \text{score} \leq 16.6666$	D	
$16.6666 < \text{score} \leq 25.0000$	D+	
$25.0000 < \text{score} \leq 33.3333$	C-	'C-' score indicates satisfactory relative ESG performance and moderate degree of transparency in reporting material ESG data publicly.
$33.3333 < \text{score} \leq 41.6666$	C	
$41.6666 < \text{score} \leq 50.0000$	C+	
$50.0000 < \text{score} \leq 58.3333$	B-	'B-' score indicates good relative ESG performance and above-average degree of transparency in reporting material ESG data publicly.
$58.3333 < \text{score} \leq 66.6666$	B	
$66.6666 < \text{score} \leq 75.0000$	B+	
$75.0000 < \text{score} \leq 83.3333$	A-	'A-' score indicates excellent relative ESG performance and high degree of transparency in reporting material ESG data publicly.
$83.3333 < \text{score} \leq 91.6666$	A	
$91.6666 < \text{score} \leq 100.0000$	A+	

Notes: This table presents the translation from the Refinitiv ESG score (displayed as numbers between 0 and 100) to the Refinitiv ESG Rating (displayed as grades between A+ and D-). Source: (Refinitiv, 2023).

zero by construction. This applies, for instance, in Boolean scoring frameworks when a firm fails to meet a predefined sustainability criterion or when no information is available. These firms are treated as non-eligible ("non-eligible firm") and are represented by an indicator variable $1_{eligible}$ equal to zero. Firms meeting the criterion receive a value of one ("eligible firms"). The individual data point scores within a category are then aggregated into a single category score using a similar scoring mechanism.

Category weights capture the relative importance of each ESG category for a given industry and are specific to the respective peer group. Refinitiv employs a data-driven methodology that assigns higher weights to categories considered more relevant based on an industry's exposure to negative sustainability screening criteria. For example, industries with high emission levels are assigned greater weight in the emissions category. This approach is operationalized by comparing the median values of sustainability indicators across industries. Industries are then sorted into deciles based on these medians. Categories linked to industries in less sustainable deciles receive higher weights, reflecting their presumed materiality. As a result, category weights vary across industries and aim to align ESG assessments with the specific sustainability challenges most relevant to each sector. A modified procedure is applied to the categories within the governance pillar, as detailed below in Section 4.3.4.

4.3.3 Impact of the peer group

In this section, we examine the impact of peer groups on scores. Consider a sector that is divided into two industries, A and B, consisting of M and N firms, respectively. The effect of the peer group is isolated by comparing ESG scores based on best-in-industry versus best-in-sector benchmarking. The analysis generalizes to sectors with more than two industries, as each industry can be subdivided into smaller groups without loss of generality.

Definitions: Let $s_A = \frac{M}{M+N}$ and $s_B = 1 - s_A = \frac{N}{M+N}$ denote the relative sizes of industries A and B. In industry A (B), all x_0 (y_0) firms are non-eligible, where higher scores reflect better sustainability performance. Let $z_S = \frac{x_0+y_0}{M+N}$ be the share of non-eligible firms in the sector, and $z_A = \frac{x_0}{M}$ and $z_B = \frac{y_0}{N}$ the corresponding shares within industries A and B.

For a given firm i, define $v_{A,i} = \frac{x_{H,i}}{M}$ and $w_{A,i} = \frac{x_{L,i}}{M}$, where $x_{H,i}$ and $x_{L,i}$ represent the number of firms in industry A with strictly higher and lower scores, respectively. Similarly, define $v_{B,i} = \frac{y_{H,i}}{N}$ and $w_{B,i} = \frac{y_{L,i}}{N}$ for industry B. Note that $x_{L,i} + x_{H,i} \leq 1$ and $y_{L,i} + y_{H,i} \leq 1$, as some firms may have identical scores to firm i.

Average Scores: The score of a firm is defined in Equation (4.2). For calculating average scores, we assume without loss of generality that all eligible firms have distinct, non-zero scores.[16] The average best-in-sector score μ_S is given by:

$$\mu_S = \frac{\sum_i Score_i}{M+N} = \frac{\sum_{i=x_0+y_0}^{M+N-1} \frac{x+\frac{1}{2}}{M+N}}{M+N} = \frac{1}{2}\left(1 - z_S^2\right)$$

Similarly, the average best-in-industry scores $\mu_{A,B}^{ind}$ are:

$$\mu_{A,B}^{ind} = \frac{1}{2}\left(1 - z_{A,B}^2\right)$$

Average scores equal $\frac{1}{2}$ when all firms are eligible, implying that peer group definitions do not influence average outcomes in such cases.

If some firms are non-eligible, the average scores may deviate from the benchmark value of $\frac{1}{2}$ and can be lower, but still positive. The magnitude of this deviation increases with the share of non-eligible firms in the respective peer group. Define $\bar{\Delta}$ as the difference between average best-in-industry and best-in-sector scores:

$$\begin{aligned}\bar{\Delta} &= \left(s_A \mu_A^{ind} + (1 - s_A)\mu_B^{ind}\right) - \mu_S \\ &= -\frac{1}{2}(1 - s_A)s_A(z_A - z_B)^2\end{aligned}$$

Thus, the average best-in-industry score across all firms can never exceed the average best-in-sector score. The difference is non-positive and increases in absolute value with the squared difference $(z_A - z_B)^2$.

Remark: In the case of Boolean scores, where only firms meeting a sustainability criterion are eligible and receive a score of 1, it holds that $w_{A,i} = \bar{w}_A = z_A$ and $w_{B,i} = \bar{w}_B = z_B$. Hence, $(z_A - z_B)^2 = (\bar{w}_A - \bar{w}_B)^2$. This implies that the greater the heterogeneity in sustainability performance between industries – that is, the larger the difference between $\bar{w}_A$ and $\bar{w}_B$ – the more pronounced the divergence between best-in-industry and best-in-sector scores becomes.

[16] See Appendix 4.8.1 for details.

Scores Using Equation (4.2), the scores of eligible firms satisfy:

$$\begin{aligned} S_{S,i} &= \frac{1}{2} + \frac{1}{2}\left[s_A(w_{A,i} - v_{A,i}) + (1-s_A)(w_{B,i} - v_{B,i})\right] \\ S^{ind}_{D,i} &= \frac{1}{2} + \frac{1}{2}(w_{D,i} - v_{D,i}), \quad D \in \{A, B\} \end{aligned} \tag{4.3}$$

where $S_{S,i}$ denotes the best-in-sector score. Depending on the industry, $S^{ind}_{A,i}$ or $S^{ind}_{B,i}$ is the corresponding best-in-industry score. The scores of non-eligible firms are always zero.

Intuitively, $r_{A,i} = w_{A,i} - v_{A,i}$ and $r_{B,i} = w_{B,i} - v_{B,i}$ represent firm i's peer-relative scores within industries A and B, respectively. Recall that w denotes the share of less sustainable peers, while v denotes the share of more sustainable ones. A firm receives a best-in-sector score above $\frac{1}{2}$ if and only if its peer-relative score is positive. To obtain a positive score, the firm must therefore outperform the median firm in its industry. The overall best-in-sector score is then computed as a weighted average of the firm's peer-relative scores across industries, with weights corresponding to the relative sizes of the respective industries.

Remark For Boolean scores, if a firm has a score of 1, then $v_{A,i} = v_{B,i} = 0$. The corresponding scores are:

$$\begin{aligned} S_{S,i} &= \frac{1}{2} + \frac{1}{2}\left[s_A\bar{w}_A + (1-s_A)\bar{w}_B\right] \\ S^{ind}_{D,i} &= \frac{1}{2} + \frac{1}{2}w_{D,i} \quad D \in \{A, B\} \end{aligned}$$

The best-in-industry score is positively related to the number of non-eligible peers within the same industry. Hence, firms that meet a sustainability criterion benefit more from industry-specific benchmarking when a large share of their industry peers is non-eligible.

Changes in Scores of a Firm: Let $\Delta_{i,A}$ and $\Delta_{i,B}$ denote the difference in scores between best-in-industry and best-in-sector benchmarks. Then, for an eligible firm, it holds that:

$$\begin{aligned} \Delta_{A,i} &= S^{ind}_{A,i} - S_{S,i} = -\frac{1}{2}(1-s_A)\left[(w_{B,i} - v_{B,i}) - (w_{A,i} - v_{A,i})\right] \\ \Delta_{B,i} &= S^{ind}_{B,i} - S_{S,i} = -\frac{1}{2}(1-s_B)\left[(w_{A,i} - v_{A,i}) - (w_{B,i} - v_{B,i})\right]. \end{aligned} \tag{4.4}$$

This is the central result of the section. It shows that firms with identical sustainability characteristics may receive different ESG scores under the best-in-industry approach, depending on their peer-relative scores within the industries that make up the sector. A firm benefits from best-in-industry benchmarking – specifically, when it receives a higher score than under sector-based evaluation – if its peer-relative score within its own industry exceeds that in the other industry. This mechanism tends to favour firms in weaker-performing industries and penalize those in stronger ones, although the effect may not be uniform (as illustrated in the example below). Moreover, the effect is amplified when the firm belongs to a relatively small industry. Firms that are non-eligible receive a score of zero under both the best-in-industry and best-in-sector approaches.

Example (Non-Uniform Effect within an Industry): Consider a highly skewed distribution within industry A, where a single firm I achieves the highest score in the entire sector, while all other firms in A score below all firms in B. All firms are eligible, and the sustainability criterion is numerical (e.g., tons of waste). In this case, we have $0 < w_{A,I} < 1$, $w_{B,I} = 1$, and $v_{A,I} = v_{B,I} = 0$. Thus, firm I's peer-relative score is less than one within industry A, but exactly one within industry B.

For all other firms $i \neq I$ in industry A, it holds that $0 \leq w_{A,i} < 1$, $w_{B,i} = 0$, $0 \leq v_{A,i} < 1$, and $v_{B,i} = 1$. Their peer-relative score in industry A is greater or equal than in industry B, where it equals -1. From Equation (4.4) it follows that:

$$\begin{aligned} \Delta_{A,I} &< 0 \\ \Delta_{A,i} &\geq 0 \quad \forall i \neq I \end{aligned}$$

This example illustrates that the direction of score changes within a single industry may not be uniform. Although such cases may be rare, even a high-performing firm can be penalized under industry-specific benchmarking if it belongs to an otherwise weak industry.

Remark In the Boolean case, Equation (4.4) simplifies to:

$$\begin{aligned} \Delta_{A,i} &= -\frac{1}{2}(1 - s_A)(\bar{w}_B - \bar{w}_A) \\ \Delta_{B,i} &= -\frac{1}{2}(1 - s_B)(\bar{w}_A - \bar{w}_B) \end{aligned}$$

In this case, it is sufficient to consider the share of firms that do not meet the sustainability criterion and are therefore non-eligible. Firms operating in industries with a higher proportion of non-eligible peers receive higher best-in-industry scores, and vice versa.

4.3.4 Differences in ESG ratings

Differences in ESG ratings between best-in-industry and best-in-sector benchmarking, denoted by ΔESG_i, arise from differences in both category scores and the associated category weights. According to the Refinitiv methodology, category-level scores are first computed as described in Section 4.3.3 and then aggregated using the scoring mechanism defined in Equation (4.2), applied to the sum of individual data point scores (SIDPS) within each category. Consequently, differences in category scores are expected to be positively related to differences in individual scores. These category scores are subsequently weighted to reflect their relative importance in the overall ESG assessment.

The weights are based on negative sustainability screening criteria (e.g., tons of emissions or the absence of specific social policies) considered to be key for each category (key indicator). For each industry, the medians of these key indicators are computed and ranked into deciles.[17] Industries with the worst median value for a given sustainability criterion are assigned a "magnitude weight" of 10, whereas those with the best median receive a weight of 1. An exception applies to the Governance pillar, which uses fixed magnitude weights.[18] The final category weight is calculated as the ratio of a category's magnitude weight to the sum of all magnitude weights. Thus, magnitude weights and category weights are positively related.

Definitions: The analytical setup follows Section 4.3.3. The sector consists of two industries, A and B, comprising M and N firms, respectively. For a given ESG category, let the difference between best-in-industry and best-in-sector scores be denoted by $\Delta_{D,i} = S^{ind}_{D,i} - S_{S,i}$, where $D \in \{A, B\}$, and $\Delta_{D,i}$ is defined in Equation (4.4).

Let $\bar{S}^{ind}_{D,i}$ and $\bar{S}^{S}_{D,i}$ denote the weighted average of category scores across all other categories (excluding the one under consideration), under best-in-industry and best-in-sector bench-

[17] If a category contains multiple indicators, the value of the screening criterion is the average of them.

[18] In the Governance pillar, the magnitude weights for the categories Management, Shareholders, and CSR Strategy are fixed at 10, 3, and 2, respectively.

marking, respectively. To isolate the effect of the category in question, we assume $\bar{S}^{ind}_{D,i} = \bar{S}^{S}_{D,i}$. The firm's weight for the respective category is denoted by $\omega^{ind}_{D,i}$ (best-in-industry) and $\omega^{S}_{D,i}$ (best-in-sector), with both weights bounded between 0 and 1.

ESG Ratings Difference: The overall difference in ESG scores is given by:

$$\begin{aligned} \Delta ESG_i &= \omega^{ind}_{D,i} S^{ind}_{D,i} + \left(1 - \omega^{ind}_{D,i}\right) \bar{S}^{ind}_{D,i} - \left[\omega^{S}_{D,i} S^{S}_{D,i} + \left(1 - \omega^{S}_{D,i}\right) \bar{S}^{S}_{D,i}\right] \\ &= \omega^{ind}_{D,i} \Delta_{D,i} + \left(\omega^{ind}_{D,i} - \omega^{S}_{D,i}\right) \left(S_{S,i} - \bar{S}_{S,i}\right). \end{aligned} \tag{4.5}$$

The first term shows that ESG rating differences are directly related to category score differences. The second term reflects the impact of weight differences between the two benchmarking approaches. This effect is positive when the category score exceeds the average of the other categories, and negative otherwise. It follows that the overall impact on ESG ratings depends on both the magnitude and the direction of category score and weight differences.

Weights: As a key indicator, consider a numerical variable such as emissions or waste, where higher values indicate lower sustainability performance. Let M_A and M_B denote the medians of this indicator in industries A and B, respectively, and assume that $M_A > M_B$. Then the sector-wide median M_S lies within the interval $[M_B, M_A]$.

Because medians are positively related to firm rankings – and these rankings, in turn, determine category weights – firms from industries with higher medians (e.g., A) tend to receive higher weights under sector-based benchmarking. Conversely, firms from industries with lower medians (e.g., B) are more likely to receive lower weights.

This effect is further amplified by differences in industry size. The number of firms in each industry influences the position of the overall sector median. When the share of firms from industry A (s_A) is small, the sector median M_S will lie closer to M_B. In this case, removing industry B through best-in-industry benchmarking induces a more pronounced shift in the effective median, thereby intensifying the resulting change in category weights. A symmetric argument applies to industry B

Total Effect: The firm's category score difference $\Delta_{D,i}$ influences the direction of the ESG rating change positively through the first term in Equation (4.5). Each category score is derived from a positively defined sustainability metric (SIDPS). Let m_D denote the median of the SDPS within industry D. According to Equation (4.4), the category score *tends* to be positively related to m_A and negatively related to m_B. As previously noted, this effect is particularly pronounced in smaller industries.

However, the overall impact on an individual firm's ESG rating remains ambiguous. As motivated by the example in Section 4.3.4, a lower median $m_{A,i}$ compared to $m_{B,i}$ does not necessarily imply that $\Delta_{A,i} > 0$ (and vice versa), since the influence of medians on score differences may not be uniform across firms.

Even when abstracting from this non-uniformity and assuming a plausible empirical scenario – namely, that the key indicators and SDPS underlying category weights and scores, and thus M_D and m_D, are perfectly aligned in how they reflect sustainability performance across industries – the total effect on the ESG rating remains unclear. In such a setting, changes in category scores and weights would tend to move in the same direction. However, the final impact on the ESG rating also depends on how a firm's best-in-sector score in a given category compares to its average score across all categories. Since this deviation can be either positive or negative, the two components in Equation (4.5) – score and weight effects – may either reinforce or offset one another.

4.4 Methodology and data

This section outlines the empirical methodology to evaluate the impact of peer group on the informativeness of ESG ratings. We begin by introducing the methodological framework, including the construction of best-in-industry and best-in-sector ESG scores and the portfolio sorting approach used to assess financial performance, in Section 4.4.1. We then describe the underlying dataset, which combines ESG information from Refinitiv with financial market data to estimate return differentials across rating-based portfolios, in Section 4.4.2.

4.4.1 Methodology

Our main objective is to assess the impact of the peer group on the financial informativeness of ESG ratings. Specifically, we examine how the choice of the peer group affects the ability of ESG scores to reflect the inherent sustainability risk exposure of a firm. Therefore, we compare two types of ESG ratings: best-in-industry and best-in-sector ratings. Both are based on the Refinitiv methodology.

In practice, only best-in-industry ratings are available at Refinitiv database. However, (historical) ESG scores could include small discrepancies in terms of the underlying methodology and benchmark, which may introduce sample selection biases. To address this issue, we reconstruct best-in-industry ratings ourselves to ensure that the benchmark is consistent with our available sample. This approach also guarantees that average category ratings do not exceed 50, as discussed in Section 4.3.3.

For industry group and economic sector classification, we rely on Refinitiv's TRBC scheme. TRBC is a five-level system consisting of 13 economic sectors, 33 business sectors, 62 industry groups, 154 industries, and 898 activities. The Refinitiv (2023) ESG methodology utilizes the industry group as the relevant benchmark. Accordingly, we use this level to recalculate the best-in-industry scores. In contrast, best-in-sector scores are based on the broader economic sector. This is the highest level at which peer-specific weights are still applied under the Refinitiv approach. This distinction affects the peer groups used for calculating scores within the environmental and social pillars. In contrast, scores for categories in the governance pillar remain unaffected, as firms are benchmarked against firms in the same country.

We sort firms into decile and quintile portfolios based on their best-in-industry or best-in-sector ESG score. Firms with the lowest (highest) scores are assigned to the bottom (top) portfolios. The Refinitiv scores are updated at fiscal year's end. Hence, we rebalance our portfolios at the end of December in year t, and holding them from the beginning of January until the end of December in the following year $t+1$.

To analyse the informativeness of the ESG scores, we use the five-factor model by Fama and French (2015):

$$R_t = \alpha + \beta_1 MKT_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 RMW_t + \beta_5 CMA_t + \epsilon_t, \tag{4.6}$$

where R_t is the return reduced by the risk-free rate at time t, and α is the unexplained abnormal return (i.e., the sustainability premium) by the model. The Fama and French (2015) five-factor model captures the return of the market (MKT), the return of a size factor (SMB), the return of a value factor (HML), the return of a profitability factor (RMW), and the return of an investment factor (CMA).

To evaluate the financial informativeness of the ESG scores, we rely on the semi-strong form of the efficient market hypothesis, that all publicly available information is priced by financial market participants (Fama, 1970). Specifically, we assume that all sustainability information in the form of ESG scores is reflected in the prices. If ESG scores contain priced information, we expect α to be lower for high-ESG firms and, consequently, higher for low-ESG firms. Following Khan et al. (2016), we focus on the Δ-Alpha, defined as the difference in abnormal returns between the top and bottom portfolios. Consequently, a higher Δ-Alpha signals greater informativeness of ESG scores. From an investment perspective, this corresponds to the return of an investment strategy that has long positions in stocks with low ESG scores and short positions in stocks with high ESG scores, based on the respective peer-group-specific ESG score.

4.4.2 Data

The focus of the study is the U.S. stock market. However, the recalculation of the best-in-industry and best-in-sector ratings requires all worldwide rated firms by Refinitiv as reference. Hence, we use an initial universe of 10,510 companies that are included in the Refinitiv ESG rating procedure in the year 2021. For each of these firms, we obtained the required data to recalculate both best-in-industry and best-in-sector scores. Since the examination period starts in the year 2010 and ends in 2022, our final sample consists of 59,978 firm-year observations.

In the following examination, we use a subsample consisting of 3,098 firms having their headquarters in the U.S., resulting in 18,078 firm-year observations. With 29.24 percent of the firms, U.S.-based firms represent the largest share in our dataset. For each of these U.S. firms, price, accounting, and ESG data are taken from Refinitiv database. Factor data for the Fama and French (2015) five-factor model is obtained from Kenneth R. French's data library.[19]

Table 4.2 presents the best-in-industry (Panel A) and best-in-sector (Panel B) environmen-

[19] Web-page: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Table 4.2: Descriptive statistics

	U.S. Sample			Whole Sample		
	Count	Mean	S.D	Count	Mean	S.D
Panel A: Industry-based Score						
RESU	18,078	25.25	31.65	59,978	36.80	33.10
EMIS	18,078	24.18	30.22	59,978	37.48	33.03
ENVI	18,078	16.06	26.71	59,978	23.27	30.48
WORK	18,078	37.62	25.82	59,978	49.39	29.16
HUMR	18,078	20.02	29.86	59,978	26.71	32.56
COMM	18,078	56.43	25.14	59,978	47.24	30.13
PROD	18,078	39.12	25.77	59,978	43.12	30.43
ESG	18,078	38.38	18.93	59,978	42.70	20.69
Panel B: Sector-based Score						
RESU	18,078	25.31	31.46	59,978	37.15	33.20
EMIS	18,078	24.09	29.90	59,978	37.88	33.11
ENVI	18,078	16.17	27.36	59,978	23.42	31.37
WORK	18,078	37.14	25.05	59,978	49.68	29.22
HUMR	18,078	20.28	30.09	59,978	26.83	32.71
COMM	18,078	56.51	24.27	59,978	47.31	30.26
PROD	18,078	38.21	26.29	59,978	43.53	30.89
ESG	18,078	32.67	21.20	59,978	39.91	22.63

Notes: This table presents the number, the mean, and the standard deviation of the recalculated ESG scores. The U.S. sample is the examination sample and includes firms having their headquarters in the U.S., while the whole sample contains every recalculated firm-year-observation.

tal and social category scores. In the whole sample, the means of the best-in-sector category scores are slightly higher than those of the best-in-industry category scores, while the standard deviations (reflecting the extent to which ESG ratings could differentiate between firms based on sustainability performance) stay on a similar level across both peer group approaches.

Turning to the U.S. sample, we observe that companies located in the U.S. exhibit substantially lower average ESG category scores. The largest differences of around 12 points are observed in the resource use ($RESU$), emissions ($EMIS$), and workforce ($WORK$) categories. These differences are the equivalent of one and a half letter grade. The only exception

Table 4.3: Change in ratings

Best-in-industry	Best-in-sector												
	D-	D	D+	C-	C	C+	B-	B	B+	A-	A	A+	Total
D-	**193**	49	3	0	0	0	0	0	0	0	0	0	245
D	455	**902**	251	20	0	0	0	0	0	0	0	0	1,628
D+	267	1,716	**826**	311	19	1	0	0	0	0	0	0	3,140
C-	12	1,126	1,304	**737**	265	53	4	1	0	1	0	0	3,503
C	0	246	1,002	702	**533**	248	56	6	0	2	1	0	2,796
C+	0	23	350	438	499	**461**	256	56	7	2	0	0	2,092
B-	0	0	46	119	206	386	**412**	256	69	2	1	0	1,497
B	0	0	0	18	48	161	321	**429**	265	47	1	0	1,290
B+	0	0	0	0	2	27	108	278	**377**	234	20	0	1,046
A-	0	0	0	0	0	1	7	40	177	**298**	77	1	601
A	0	0	0	0	0	0	0	3	12	80	**117**	11	223
A+	0	0	0	0	0	0	0	0	0	1	14	**2**	17
Total	927	4,062	3,782	2,345	1,572	1,338	1,164	1,069	907	667	231	14	18,078

Notes: This table displays a matrix of the recalculated industry-based ESG ratings (vertical axis) and the sector-based ESG ratings (horizontal axis) of U.S. firms. Firms on the diagonal line (in bold) have no rating change, companies above the diagonal line receive a rating upgrade through the peer group change, and companies below the diagonal line are downgraded.

is the Community ($COMM$) category, where the average score increases from circa 47 to 56. Overall, the U.S. results are consistent to the global patterns: the means and standard deviations of scores remain broadly stable across benchmarking approaches.

Table 4.3 illustrates the rating transition matrix between best-in-industry and best-in-sector ESG ratings for U.S. firms. After the change of the benchmark, around 29 percent (5,287 of 18,078 ratings) of the companies remain at the identical rating level, while 56 percent (10,195 of 18,078 ratings) of the companies receive a downgrade. For example, 240 of the 841 firm-year observations rated between A+ and A-, according to the industry-based approach, receive a downgrade that results in a B+ rating or worse. The largest decreases are observed in one firm-year observation from A- to C+ and in three firm-year-observations from A to B.

4.5 Results

This section presents the empirical results of our analysis. We begin by validating the plausibility of our recalculated best-in-industry ESG scores by comparing them to the original Refinitiv scores (Section 4.5.1). We then examine the relationship between ESG scores and stock returns using the Fama and French (2015) five-factor model, comparing the informativeness of best-in-industry and best-in-sector scores (Section 4.5.2). Finally, we explore several alternative model specifications to disentangle the roles of category weights, firm characteristics, and sustainability-related risk factors (Section 4.5.3).

4.5.1 Fit of recalculated best-in-industry ratings

As discussed in Section 4.4.1, we use the original Refinitiv (2023) methodology to recalculate the industry-based category scores. To validate the appropriateness of this approach, we compare them with the original Refinitiv scores. Table 4.4 shows the results by category and economic sector. We find that the correlations between the original and recalculated category scores are consistently high and close to 1. On average, the correlation coefficients range between 0.95 to 1.00, with one matching the original score ($ENVI$ in *Academic & Educational Services*). To further examine the fit of the self-constructed scores, we regress the recalculated scores (dependent variable) with the original Refinitiv scores (independent variable) and year-fixed effects. The estimated coefficients are close to 1 and highly significant, indicating a nearly perfect linear relationship, and consequently, show that the recalculated scores closely replicate the original ratings. In addition, we find a minor systematic deviation in means between the category scores. These are within the range of 0.5 to 3 points in a 100-point scale.

There are three main reasons why the fit between original and recalculated scores may not be perfect. First, ESG ratings depend on a specific peer group used for the benchmarking. Differences in peer group composition will lead to differences in scores. In our recalculation procedure, we take 10,510 companies for 2021 as the basis and assume that these are also rated in the previous years. As a result, our list does not include companies that are part of the rating calculation process in the past but no longer appear in the base year. This may be the case if a respective company was involved, for example, in past insolvency issues, mergers, or acquisitions. We must therefore assume that our list is partially incomplete, and we do not match the exact peer groups for the recalculation.

Second, it is possible for firms to be assigned to a different industry over time. According to the TRBC classification of Refinitiv, the assignment to an industry is based on sales generated in an individual business segment. Since the rated companies usually generate sales in several industry segments and some of them are thematically very close to each other (e.g., *Aerospace & Defense* and *Government Activity*), volatility in sales leads to a different industry allocation over time. This leads to changes in the peer group and, consequently, in the recalculated scores. However, due to the lack of historical industry classification of the individual firms, these changes cannot be taken into account.

Third, the scores are fixed after 5 years (Refinitiv, 2023). This means, if 2021 is the current fiscal year in which Refinitiv discloses its category scores, all scores prior to fiscal year 2017 are no longer updated, even if the underlying data points are corrected or the calculation method has changed. As a result, scores in our sample dating further back in time may reflect outdated information and are particularly affected by this inaccuracy.

Despite these limitations, we find a strong overall fit between the recalculated and original Refinitiv scores – both statistically and economically.

4.5.2 ESG ratings and financial returns

Financial informativeness

Table 4.5 presents the regression coefficients of the Fama and French (2015) five-factor model for the top and bottom decile and quintile portfolios. Panel A shows the results of the industry-based ESG scores and Panel B of the sector-based ESG scores. We find significant abnormal returns (Alphas) across all portfolios. For the industry-based and the sector-based scores, the top portfolios exhibit negative Alphas, while the bottom portfolios display positive Alphas. In line with prior literature, these findings point to the negative relationship between stock returns and sustainability engagement (e.g., Bolton and Kacperczyk, 2021; Hübel and Scholz, 2020). In addition, this suggests the existence of a sustainability premium, and that

Table 4.4: Validation of recalculated ESG category scores

	RESU	EMIS	ENVI	WORK	HUMR	COMM	PROD
Academic & Educational Serv.							
Correlation	0.9572	0.9608	1.0000	0.8989	0.9529	0.8562	0.9550
Regression	0.9358	0.9428	1.000	0.8799	0.9282	0.8378	0.9467
Mean	-1.6487	0.0638	0.0000	-1.0980	0.0658	1.2755	0.0233
Basic Materials							
Correlation	0.9820	0.9802	0.9681	0.9699	0.9944	0.9807	0.9890
Regression	0.9826	0.9780	0.9852	0.9797	0.9983	1.0081	0.9839
Mean	-1.1605	-1.5082	1.0249	-2.0510	-0.1521	-1.8768	-1.0591
Consumer Cyclicals							
Correlation	0.9488	0.9607	0.9743	0.8964	0.9944	0.9320	0.9701
Regression	0.9409	0.9488	0.9612	0.9253	0.9813	0.9833	0.9628
Mean	-1.8804	-1.9238	-0.9066	-2.9315	-0.6165	-3.6699	-1.3697
Consumer Non-Cyclicals							
Correlation	0.9517	0.9542	0.9648	0.9064	0.9890	0.9362	0.9623
Regression	0.9449	0.9508	0.9550	0.9330	0.9757	0.9831	0.9564
Mean	-2.0927	-2.0023	-0.4856	-3.7461	-0.6779	-3.2996	-2.5723
Energy							
Correlation	0.9760	0.9776	0.9881	0.9704	0.9852	0.9596	0.9809
Regression	0.9540	0.9588	0.9921	0.9721	0.9597	0.9742	0.9580
Mean	-2.7765	-2.8702	0.0353	-3.6522	-0.9463	-3.4714	-1.4846
Financials							
Correlation	0.9936	0.9779	0.9936	0.9804	0.9949	0.9609	0.9897
Regression	0.9809	0.9606	0.9840	0.9812	0.9919	1.0060	0.9832
Mean	-0.5576	-0.9613	-0.4607	-2.1838	-0.1304	-4.1361	-1.1422
Healthcare							
Correlation	0.9812	0.9806	0.9687	0.9608	0.9928	0.9795	0.9758
Regression	0.9605	0.9701	0.9711	0.9530	0.9707	1.0037	0.9682
Mean	-0.7510	-0.4230	-0.0862	-2.4773	-0.5461	-3.1609	-1.9311
Industrials							
Correlation	0.9774	0.9775	0.9755	0.9450	0.9932	0.9700	0.9652
Regression	0.9664	0.9628	0.9636	0.9539	0.9838	0.9932	0.9699
Mean	-0.7656	-1.1395	-0.6131	-2.3416	-0.3928	-2.7793	-0.8122
Real Estate							
Correlation	0.9871	0.9856	0.9805	0.9730	0.9932	0.9727	0.9850
Regression	0.9761	0.9628	0.9830	0.9610	0.9845	1.0012	0.9698
Mean	-1.3641	-1.3661	-0.3310	-2.0765	-0.1137	-2.7694	-1.2508
Technology							
Correlation	0.9687	0.9693	0.9671	0.9337	0.9939	0.9633	0.9706
Regression	0.9517	0.9511	0.9541	0.9446	0.9740	0.9943	0.9645
Mean	-1.7319	-1.7350	-0.8158	-3.8414	-0.6489	-3.4790	-2.4715
Utilities							
Correlation	0.9768	0.9857	0.9122	0.9682	0.9958	0.9822	0.9921
Regression	0.9686	0.9796	0.8835	0.9795	0.9965	0.9770	0.9854
Mean	-1.4386	-1.9181	-2.8188	-2.4839	0.0377	-1.2876	-1.4711
Average							
Correlation	0.9749	0.9745	0.9743	0.9470	0.9931	0.9612	0.9767
Regression	0.9620	0.9595	0.9648	0.9557	0.9816	0.9935	0.9693
Mean	-1.3322	-1.4737	-0.4683	-2.7297	-0.4163	-3.1553	-1.4977

Notes: This table presents the correlation coefficients, regression coefficients, and differences in means between the original Refinitiv category scores and the recalculated industry-based category scores. We regress the recalculated category scores as a dependent variable with the Refinitiv original scores as an independent variable and year-fixed effects. All correlation and regression coefficients are highly significant (p-value < 0.01). The mean is the difference in means between the recalculated and the original Refinitiv category scores.

Table 4.5: Regression results of different ESG scores

	Deciles				Quintiles			
	Bottom		Top		Bottom		Top	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Panel A: Industry-based Score ($N = 155$)								
MKT	0.9694***	(20.6642)	1.0153***	(43.1018)	0.9975***	(25.7156)	1.0027***	(50.2973)
SMB	1.0009***	(10.9887)	0.1573***	(3.4388)	0.8646***	(11.4800)	0.2485***	(6.4214)
HML	0.0790	(0.9711)	0.2178***	(5.3325)	0.0853	(1.2678)	0.2181***	(6.3088)
RMW	-0.2934***	(-2.7036)	0.1570***	(2.8805)	-0.2628***	(-2.9280)	0.2087***	(4.5249)
CMA	-0.2044	(-1.6311)	-0.0563	(-0.8951)	-0.0864	(-0.8336)	-0.0395	(-0.7421)
$Intercept$	0.5001**	(2.5024)	-0.2938***	(-2.9282)	0.4553***	(2.7553)	-0.1795**	(-2.1138)
R^2	0.8700		0.9446		0.9021		0.9605	
Annualized Alpha	6.17**		-3.47***		5.60***		-1.82**	
Δ Alpha				9.64***				7.42***
Panel B: Sector-based Score ($N = 155$)								
MKT	0.9024***	(17.2796)	0.9985***	(51.9741)	0.9448***	(21.8361)	1.0312***	(51.4980)
SMB	0.8729***	(8.6096)	0.1067***	(2.8599)	0.8506***	(10.1267)	0.2286***	(5.8794)
HML	-0.2283**	(-2.5210)	0.2553***	(7.6611)	-0.1582**	(-2.1088)	0.2559***	(7.3688)
RMW	-0.8128***	(-6.7279)	0.1300***	(2.9240)	-0.6790***	(-6.7834)	0.1628***	(3.5151)
CMA	-0.2866**	(-2.0548)	0.0496	(0.9659)	-0.2003*	(-1.7334)	0.0289	(0.5402)
$Intercept$	0.8499***	(3.8206)	-0.2475***	(-3.0244)	0.6248***	(3.3902)	-0.1661*	(-1.9473)
R^2	0.8380		0.9609		0.8812		0.9621	
Annualized Alpha	10.69***		-2.93***		7.76***		-1.98*	
Δ Alpha				13.62***				9.74***

Notes: This table presents coefficient estimates and t-statistics of the recalculated industry-based and sector-based ESG score decile and quintile regressions. All values are presented in percent. The portfolios are sorted by ESG scores. Companies with the worst scores are in the bottom portfolio, and companies characterized by the best scores are assigned to the top portfolio, respectively. Those portfolios are rebalanced on a yearly basis. The dependent variable is the equal-weighted monthly return of the ESG portfolios minus the risk-free rate. Independent variables are the following Fama and French (2015) factors: MKT, SMB, HML, RMW, CMA. The results are based on the monthly average returns between the years 2010 and 2022. The difference in annualized Alphas reflects the returns of a long-short portfolio, with long positions in stocks with low ESG ratings and short positions in high ESG stocks. ***, ** and * reflect significance at the 1%, 5%, and 10% level.

both peer-group-specific ESG scores are capable of identifying it.

However, differences emerge when comparing best-in-industry and best-in-sector results. For industry-based scores, we find an annualized Alpha difference (Δ-Alpha) of 9.64 percent for the decile portfolios and 7.42 percent for the quintile portfolios. In comparison, when sorting according to the sector-based scores, the Δ-Alpha increases to 13.62 percent for decile portfolios and 9.74 percent for quintile portfolios. The bottom portfolios primarily drive those differences. This means, firms with the weakest ESG scores. Differences in unexplained returns (Alphas) for the top portfolios are much smaller.

In summary, the results suggest that the peer group has a meaningful impact on the financial informativeness of ESG ratings. In particular, the best-in-sector approach appears to more effectively detect firms with weak sustainability performance. This improves the identification of a return premium required by investors for firms with greater sustainability risk exposures.

Firms with downgraded ratings

The previous section highlights that the choice of the peer group primarily affects firms in the bottom ESG portfolios. But what does this mean in practice for portfolio managers of sustainable funds that invest in firms with high sustainability commitments? Does it matter if they use best-in-industry or a best-in-sector ratings to assess a firm's ESG risk? To answer this question, we examine three sustainable investment strategies based on minimum ESG thresholds: "Strategy A-" targets firms rated at least $A-$, "Strategy B" includes all firms rated $B+$ or higher, and "Strategy B" applies a $B+$ cutoff. These thresholds are based on best-in-industry ratings, reflecting how such strategies might be implemented in practice using the standard Refinitiv ESG data.

However, under the best-in-sector approach, some firms meeting these thresholds while others would fall short. We focus on these "downgraded" firms by forming a portfolio and computing their Alphas. Table 4.6 presents the Alpha estimates of the three possible investment strategies. Panel A displays portfolios consisting of $A+$ to $A-$ firms according to the recalculated industry-based ESG ratings. Following this, the one-level downgrade portfolio contains all superior rated firms (according to the best-in-industry approach) with

Table 4.6: Regression results of rating strategy

	Best-in-Class		1 Level Downgrade		2 Level Downgrade		3 Level Downgrade	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Panel A: Strategy A-								
Intercept	-0.0280	(-0.1989)	0.4795	(1.2345)	1.2325**	(2.1998)	0.0285	(0.0252)
Annualized Alpha	-0.34		5.91		15.84**		0.34	
Panel B: Strategy B+								
Intercept	0.2387	(1.1313)	1.1467**	(2.4399)	2.7781*	(1.7583)	0.9487*	(1.7031)
Annualized Alpha	2.90		14.66**		38.93*		12.00*	
Panel C: Strategy B								
Intercept	0.2430	(0.9725)	0.8555*	(1.6706)	2.2776	(1.4601)	0.3469	(0.8234)
Annualized Alpha	2.96		10.76*		31.03		4.24	

Notes: This table shows coefficient estimates and t-statistics of four investment strategies. Best-in-Class represents a strategy that invests in all companies that have a better industry-based rating than a defined lower bound. For Panel A, this lower bound is an A- ESG rating, for Panel B, a B+ ESG rating, and for Panel C, a B ESG rating. 1 Level Downgrade represents an investment portfolio that includes the part of companies of the former strategy that have at least a one-level downgrade of their ESG rating through the peer group change. 2 Level Downgrade and 3 Level Downgrade represent investment portfolios of companies with at least two level or three level downgrades in the ESG rating. All values are presented in percent. Those investment portfolios are rebalanced on a yearly basis. The dependent variable is the equal-weighted monthly return of the ESG portfolios minus the risk-free rate. Independent variables are omitted, and according to Fama and French (2015): MKT, SMB, HML, RMW, CMA. ***, ** and * reflect significance at the 1%, 5%, and 10% level.

at least one rating decrease through the sector-based recalculation, which are now below the threshold. The two- and three-level downgrade portfolios include firms with at least two or three rating downgrades, respectively. If sustainable funds are just allowed to invest in at least $A-$ rated firms (Panel A), we would expect, that the downgraded firms should exhibit higher positive excess returns since market participants price the risk according to the actual sustainability risk and not to the third-party rating.

According to the hypothesis, in Panel A, the two-level downgrade portfolio exhibits significantly positive returns. Moreover, the annualized Alpha estimates are significant for the one- and two-level downgrade portfolios for the B+ (Panel B) and B (Panel C) strategy. The results show that especially the downgraded firms (according to the best-in-sector approach) are responsible for the positive financial returns within a sustainable portfolio. As a consequence, the best-in-industry ratings allow portfolio managers to invest in superior rated firms without having to forgo positive returns. We assume, that those high yields originate to a large extent from firms in environmentally harmful industries. In summary, our results show that the best-in-industry approach favours green-washing suspicions for sustainable investment strategies of mutual funds.

4.5.3 Alternative model specifications

This section explores alternative model specifications to better understand the observed differences in the financial informativeness of ESG scores across peer groups. Specifically, we aim to unbundle the influence of weighting schemes, address potential endogeneity concerns, and examine the explanatory power of sustainability-related risk factors. Therefore, we introduce three complementary approaches: a constant-weighting method to isolate the impact of the peer group (Section 4.5.3), a two-step procedure that accounts for firm-level fundamentals (Section 4.5.3), and an extended factor model that includes an ESG-specific risk factor (Section 4.5.3).

Weights

As shown in Section 4.3.2, the difference between best-in-sector and best-in-industry ESG scores depends not only on differences in category scores but also on deviations in category weights. The latter effect is ambiguous and may relativize the impact of score differences. As noted by Muck and Schmidl (2024), the weighting scheme has a significant empirical influence on the financial informativeness of ESG ratings. To isolate the effect of weighting, we fix category weights while allowing scores to vary. Specifically, for each firm i, we set the best-in-sector weight equal to the firm i's best-in-industry weight, i.e., $\omega_{S,i} = \omega_{D,i}^{ind}$, where D denotes the industry and S denotes the sector to which firm i belongs.

Table 4.7 presents the regression estimates of the decile and quintile portfolios. Panel A displays the results of the industry-based ESG scores. To facilitate the comparison, we replicate these results from Table 4.5. Panel B reports the coefficients for the sector-based ESG scores with constant weighting. For both decile and quintile portfolios, the estimated Δ-Alpha is higher under the best-in-sector approach than under the best-in-industry benchmark, even though category weights are held constant. This suggests that differences in category scores alone contribute meaningfully to the financial informativeness of ESG scores. However, the Δ-Alpha estimates remain lower than those reported in Table 4.5.

In particular, we find that the differences in abnormal returns between best-in-sector and best-in-industry scores in the bottom portfolios are substantially reduced when best-in-

Table 4.7: Additional tests

	Deciles			Quintiles		
	Annualized Alpha			Annualized Alpha		
	Bottom	Top	Difference	Bottom	Top	Difference
Panel A: Industry-based Score						
Constant Weighting						
Intercept	0.5001**	-0.2938***		0.4553***	-0.1795**	
Annualized Alpha	6.17**	-3.47***	9.64***	5.60***	-1.82**	7.42***
Two Step Procedure						
Intercept	0.7811***	-0.1145		0.5813***	-0.1117	
Annualized Alpha	9.79***	-1.37	11.16***	7.20***	-1.33	8.53***
ESG Adjusted Model						
Intercept	-0.0532	-0.1702		-0.0577	-0.0845	
Annualized Alpha	-0.64	-2.02	1.38	-0.69	-1.01	0.32
Panel B: Sector-based Score						
Constant Weighting						
Intercept	0.6440***	-0.1812**		0.5784***	-0.1332	
Annualized Alpha	8.01***	-2.15**	10.16***	7.17***	-1.59	8.76***
Two Step Procedure						
Intercept	0.7936***	-0.1932**		0.6751***	-0.1786*	
Annualized Alpha	9.95***	-2.29**	12.24***	8.41***	2.12*	10.53***
ESG Adjusted Model						
Intercept	0.2800	-0.1037		0.0540	-0.0135	
Annualized Alpha	3.41	-1.24	4.65**	0.65	-0.16	0.81

Notes: This table shows annualized Alpha estimates and differences of the additional tests. All values are presented in percent. The portfolios are sorted by ESG scores. Companies with the worst scores are in the bottom portfolio, and companies characterized by the best scores are assigned to the top portfolio, respectively. Those portfolios are rebalanced on a yearly basis. The dependent variable is the equal-weighted monthly return of the ESG portfolios minus the risk-free rate. Independent variables are the following Fama and French (2015) factors: MKT, SMB, HML, RMW, CMA. The results are based on the monthly average returns between February 2010 and December 2022. The difference in annualized Alphas reflects the returns of a long-short portfolio, with long positions in stocks with low ESG ratings and short positions in high ESG stocks. Under Constant Weighting, we report the Alpha estimates and differences of industry-weighted ESG scores. Under Two Step Procedure, we report the Alpha estimates and differences for estimated ESG scores with respect to firm size, market-to-book ratio, profitability, year fixed effects, and sector fixed effects. Under ESG Adjusted Model, we report Alpha estimates and differences for a Fama and French (2015) five-factor model with an additional ESG-factor. ***, ** and * reflect significance at the 1%, 5%, and 10% level.

industry weights are applied instead of best-in-sector weights. For example, using best-in-sector weights, the abnormal return in the bottom decile portfolio is 4.52 percentage higher than under the best-in-industry benchmark. Isolating the benchmark effect, this difference decreases to 1.84 percentage. These observations suggest that differences in scores and weights tend to operate in the same direction. More importantly, they provide empirical evidence that variation in category weights is a relevant driver of the observed differences in financial informativeness between best-in-sector and best-in-industry ESG scores.

Two step procedure

As discussed in Section 4.2, ESG ratings and financial performance may be affected by each other. To address potential endogeneity concerns, we follow a two-step procedure. First, we estimate an ESG score with respect to firm size, market-to-book ratio, profitability, and leverage, while controlling for year and sector fixed effects.[20] In the second step, we sort firms into decile and quintile portfolios according to their estimated best-in-industry and best-in-sector ESG scores and repeat the regression analysis. This procedure mitigates concerns that ESG scores might merely reflect firm characteristics rather than sustainability-related information.

Table 4.5 presents the results of this two-step procedure. Panel A shows the results of the industry-based ESG scores. We find statistically significant abnormal returns for the bottom portfolios, while the estimates for the top decile and quintile portfolios are not statistically significant. Nevertheless, the difference in abnormal returns (Δ-Alpha) between the bottom and top portfolios is significant, suggesting that best-in-industry ESG scores reflect financial informativeness after controlling for specific firm characteristics.

Panel B reports the results of the sector-based scores. In contrast to the industry-based case, all estimated abnormal returns in this setting are statistically significant across decile and quintile portfolios. The estimated differences in abnormal returns (Δ-Alpha) are also larger than those obtained under the best-in-industry approach. In line with previous findings, these results confirm the conclusion that broader peer groups enhance the financial informativeness of ESG ratings.

[20]For further explanation, see Appendix 4.8.2.

ESG adjusted model

To further investigate the source of the return difference between bottom and top portfolios and to ensure that our results are driven by the portfolio sorting procedure, we include an additional ESG factor into the standard Fama and French (2015) five-factor model. Therefore, we construct the factor in order to isolate a sustainability premium and repeat the empirical analysis.[21]

Table 4.7 presents the results of the ESG adjusted model. Across all regression models, the abnormal returns (Alphas) for the industry-based and sector-based ESG score portfolios are no longer statistically significant. Hence, the former excess returns are largely absorbed by the additional ESG factor. This suggests that the informativeness in form of abnormal returns can be traced back to the related ESG risk and our sorting methodology.

4.6 Robustness

We conduct further robustness checks to validate our previous findings. Table 4.8 provides a battery of additional tests for the recalculated industry-based (Panel A) and the sector-based ESG score (Panel B). First, we need to take into account that extraordinary events such as the aftermath of the Financial Crisis or the outbreak of COVID-19 have a significant impact on financial markets and could bias our results. Consequently, we analyse different sub-samples of the original sample. For the period January 2013 to December 2022 (excluding the aftermath of the Financial Crisis), we find similar results. The sector-based score points higher Alpha differences than the industry-based score. This is in line with the results between February 2010 and February 2020 (exclusion of the COVID-19 observations). Moreover, we exclude both events, which results in an analysis period of January 2013 to February 2020. Decile and quintile bottom portfolios exhibit significant risk premiums, and the top portfolios show significant negative returns, while the sector-based ESG score points higher Alpha differences. These results indicate that if we neglect the effects of extraordinary events, our constructed sector-based ESG score leads to better differentiation and could therefore enhance the measurement of sustainability risk.

[21] For further explanation, see Appendix 4.8.3.

Table 4.8: Robustness tests

	Deciles			Quintiles		
	Annualized Alpha			Annualized Alpha		
	Bottom	Top	Difference	Bottom	Top	Difference
Panel A: Industry-based Score						
Alternative Sub-periods						
Excluding Fin. Crisis	6.71**	-3.78**	10.49***	5.92**	-2.40**	8.32***
Excluding COVID-19	5.30**	-3.29***	8.59***	5.04***	-2.34***	7.38***
Excluding Both	6.15**	-3.33***	9.48***	5.43**	-2.42**	7.85***
Alternative Factor Models						
Three-factor Alpha	4.54*	-3.21***	7.75***	4.49**	-1.66	6.15***
Six-factor Alpha	6.76***	-2.86**	9.62***	6.38***	-1.55	7.93***
Alternative Rebalancing						
Monthly Rebalancing	6.25**	-3.46***	9.71***	5.60***	-2.13**	7.73***
Pillar Models						
Environmental Pillar	4.45**	-2.64**	7.09***	4.99***	-2.50**	7.49***
Social Pillar	5.99***	-2.25**	8.24***	4.21**	-1.22	5.43***
Panel B: Sector-based Score						
Alternative Sub-periods						
Excluding Fin. Crisis	12.84***	-2.89***	15.73***	8.80***	-2.01*	10.81***
Excluding COVID-19	9.43***	-3.76***	13.19***	7.09***	-2.78***	9.87***
Excluding Both	11.39***	-3.67***	15.06***	8.19***	-2.77***	10.96***
Alternative Factor Models						
Three-factor Alpha	7.06**	-2.39**	9.45***	4.94*	-1.41	6.35**
Six-factor Alpha	11.44***	-2.13**	13.57***	8.38***	-1.21	9.59***
Alternative Rebalancing						
Monthly Rebalancing	10.69***	-2.93***	13.62***	7.76***	-1.98*	9.74**
Pillar Models						
Environmental Pillar	5.68***	-2.36**	8.04***	5.68***	-2.25**	7.93***
Social Pillar	7.38***	-2.20**	9.58***	6.23**	-1.31	7.54***

Notes: This table shows annualized Alpha estimates and differences of the robustness tests. All values are presented in percent. The portfolios are sorted by ESG scores. The analyses for the pillar (environmental and social) models are conducted in the same way as for the ESG scores. Companies with the worst scores are in the bottom portfolio, and companies characterized by the best scores are assigned to the top portfolio, respectively. Those portfolios are rebalanced on a yearly basis. The dependent variable is the equal-weighted monthly return of the ESG portfolios minus the risk-free rate. Independent variables are the following Fama and French (2015) factors: MKT, SMB, HML, RMW, CMA. The results are based on the monthly average returns between February 2010 and December 2022. The difference in annualized Alphas reflects the returns of a long-short portfolio, with long positions in stocks with low ESG ratings and short positions in high ESG stocks. Under Alternative Sub-periods, we divide the sample into three sub-periods. Under Alternative Factor Models, we report the Alpha estimates and differences of the (Fama and French, 1993) three and six-factor regression models. Under Alternative Rebalancing, we sort the portfolios on a monthly basis. Under Pillar Models, the portfolios are sorted by the environmental and the social pillar score. ***, ** and * reflect significance at the 1%, 5%, and 10% level.

The next series of robustness tests includes a subset of different factor models. We re-estimate the Alphas with two alternative factor models to assess the robustness of our findings. Using the Fama and French (1993) three-factor and the Fama and French (2015) five-factor model with a momentum factor (Carhart, 1997), we re-run our analysis. While the Alpha difference of the Fama and French (1993) three-factor model is 1.7 percent per year (7.75 percent and 9.45 percent) for decile portfolios, we find an Alpha difference of around 4 percent per year (9.62 percent and 13.57 percent) for the six-factor model. For both ESG scores, the returns of the top quintile portfolios remain negative, but are no longer statistically significant. In general, the results of the alternative factor models confirm the previous out-performance of the sector-based recalculated ESG scores.

The third robustness check assesses the reliability of the findings if we use an alternative rebalancing methodology for the portfolios. To capture the dynamic nature of sustainability and its impact on stock returns, we use a more granular approach. Hence, we rebalance the ESG score portfolios on a monthly basis and re-run the analysis. We find significantly higher Alpha differences for the recalculated sector-based ESG scores (13.62 percent) as for the recalculated industry-based ESG score (9.71 percent) in the decile portfolios. In addition, the Alpha difference slightly decreases for the quintile portfolios (7.73 percent and 9.74 percent). Regarding our main findings, the Alpha estimates of this robustness check are close to the previous regression results. This is because Refinitiv calculates its ESG scores on the basis of companies' annual reports. As the majority of companies in our study disclose sustainability information at fiscal year's end, the underlying data of the ESG scores is also updated on a yearly basis.

The fourth series of robustness tests the recalculated environmental and social pillars. To rule out the possibility that our results are mainly driven by a single ESG pillar, we analyse the effects at a more granular level. Therefore, we sort the portfolios according to the environmental and social pillar scores. Regarding the Alpha differences in the decile and quintile portfolios for the environmental pillar, the coefficients are significant and similar for both. The change in the calculation method has also an effect in this case. The differentiation, especially in the environmental portfolios, is as pronounced as in the previous results. In contrast, the recalculated social score estimates exhibit higher Alpha differences for the decile (9.58 percent and 8.24 percent) and quintile (7.54 percent vs. 5.43 percent) portfolios. This implies that the recalculation procedure strongly enhances the explanatory power of the social component of the ESG rating, while slightly affecting the environmental pillar.

In summary, we find across all specifications similar results. In line with our main findings, the bottom portfolios, including companies with worse ESG scores, exhibit a risk premium, while financial market participants forgo returns for companies in the top portfolios. The recalculated sector-based ESG scores point to larger differences in abnormal returns, suggesting that those scores better reflect the inherent sustainability risks of a firm.

4.7 Conclusion

In this paper, we investigate the role of peer groups for the financial informativeness of ESG ratings. Due to the methodology, differences in ESG scores arise from both category scores and category weights, which reflect the sustainability performance of the industry and the firm, respectively. Therefore, we construct best-in-industry and best-in-sector ESG ratings and analyse the financial performance of U.S. stocks under each approach. We find that both ratings are financially informative, but best-in-sector scores yield a larger differentiation in returns between firms with strong and weak sustainability profiles. In particular, the return spread between bottom and top ESG portfolios is significantly larger under the best-in-sector methodology. This suggests that broader peer groups could more effectively capture sustainability-related financial risks.

These findings have important implications for retail and institutional investors who rely on ESG ratings for portfolio construction and risk assessments. Best-in-industry ratings may be less suitable than best-in-sector ratings for identifying sustainability risks and estimating the associated risk premium. Moreover, this methodological approach could potentially favour green-washing suspicions or problems for sustainable funds that align their investment strategies with those ratings. As a consequence, those findings highlight the limitations of relying solely on relative, best-in-class ESG scores. Instead, financial market participants should complement such ratings with their own absolute assessments of firms' environmental and social exposures. For asset managers, this adds another layer of complexity to the integration of ESG criteria into investment decisions.

Our analysis focuses on the Refinitiv ESG framework. A promising avenue for future research is to examine whether similar effects arise under alternative ESG methodologies, which may apply different peer-group definitions or weighting schemes. Further research could also explore whether peer group effects are similarly relevant in global or emerging markets, or in sectors with particularly salient sustainability risks, such as those related to climate change.

4.8 Appendix

4.8.1 Numerical scores: firms with identical ratings and the average score

Assume that there are k firms with the same rating and that there are $\xi > 0$ firms with a worse score than that. The total number of firms in a peer group is N. It holds that:

$$\sum_{j=1}^{k} \frac{\xi + \frac{k}{2}}{N} = \frac{k\left(\xi + \frac{k}{2}\right)}{N} = \frac{\sum_{j=0}^{k-1} \xi + j + \frac{1}{2}}{N}.$$

The right-hand side shows the corresponding sum for k firms with different numerical scores, where the first is better than ξ firms, the second better than $\xi + 1$ firms, the third better than $\xi + 2$ firms, and so on. Hence, the average is not affected by the number of firms with identical numerical scores.

4.8.2 ESG score estimate

For the estimation of the ESG scores based on firm characteristics, we employ the following cross-sectional regression model:

$$\widehat{ESG}_{i,t} = \alpha_{i,t} + \beta_1 SIZE_{i,t} + \beta_2 MTB_{i,t} + \beta_3 ROA_{i,t} + \beta_4 LEV_{i,t} + \gamma_t + \delta_j + \epsilon_{i,t}$$

where $SIZE_{i,t}$ is the natural logarithm of firm i's book value of total assets, $MTB_{i,t}$ is the ratio of firm i's market capitalization and book value of equity, $ROA_{i,t}$ is the ratio of firm i's EBIT (earnings before interest and taxes) and the book value of total assets, $LEV_{i,t}$ is the ratio of firm i's total debt and the book value of total assets, γ_t are year fixed effects, and δ_j are sector fixed effects.

4.8.3 Additional ESG factor

Following the Fama and French (1993) methodology, we construct an ESG factor. We sort firms into six portfolios based on their market capitalization and ESG scores. As breakpoints for market capitalization, we utilize the median. For the ESG score, we use the bottom 30 percent, the middle 40 percent, and the top 30 percent. We only apply four portfolios and omit the two middle portfolios: $SmallLowESG$ stands for a portfolio with small market capitalization and low ESG scores, $BigLowESG$ for big market capitalization and low ESG scores, $SmallHighESG$ for small market capitalization and high ESG scores, and $BigHighSOC$ for big market capitalization and high ESG scores. As all Refinitiv ESG scores and pillars are updated on a yearly basis, we adjust the sorting of portfolios on a yearly basis. Finally, we receive the monthly return of our ESG factor (ESG_t):

$$ESG_t = 0.5(SmallLowESG_t + BigLowESG_t) - 0.5(SmallHighESG_t + BigHighESG_t)$$

According to Hübel and Scholz (2020), we interpret this sustainability factor as the return of a long-short investment strategy, that has long positions in low ESG firms and short positions in firms with high ESG scores. To measure the excess stock returns in terms of Alpha, we use the Fama and French (2015) five-factor model plus the constructed ESG_t factor, which captures the factors market (MKT_t), size (SMB_t), book-to-market (HML_t), profitability (RMW_t), investment (CMA_t), and sustainability (ESG_t):

$$R_t = \alpha + \beta_1 MKT_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 RMW_t + \beta_5 CMA_t + \beta_6 ESG_t + \epsilon_t$$

where R_t is the return on the diversified decile or quintile portfolio in month t reduced by the risk-free return. Alpha (α) is the excess return, that is not explained by the other firm characteristic factors, and ϵ_t is the error term.

Table 4.9: TRBC sector and industry classification

Academic.	**Basic Materials**
Miscellaneous Educational Service Providers	Chemicals
School, College & University	Metals & Mining
Professional & Business Education	Construction Materials
	Paper & Forest Products
	Containers & Packaging
Consumer Cyclicals	**Consumer Non-Cyclicals**
Automobiles & Auto Parts	Beverages
Textiles & Apparel	Food & Tobacco
Homebuilding & Construction Supplies	Personal & Household Products & Services
Household Goods	Food & Drug Retailing
Leisure Products	Consumer Goods Conglomerates
Hotels & Entertainment Services	
Media & Publishing	
Diversified Retail	
Specialty Retailers	
Energy	**Financials**
Coal	Banking Services
Oil & Gas	Investment Banking & Investment Services
Oil & Gas Rel.	Insurance
Renewable Energy	Collective Investments
Uranium	Investment Holding Companies
Industrials	**Healthcare**
Aerospace & Defense	Healthcare Equipment & Supplies
Machinery, Tools, Heavy	Healthcare Providers & Services
Vehicles, Trains & Ships	Pharmaceuticals
Construction & Engineering	Biotechnology & Medical Research
Diversified Industrial	
Goods Wholesale	
Professional & Commercial Serv.	
Freight & Logistics Serv.	
Passenger Transportation Serv.	
Transport Infrastructure	
Real Estate	**Technology**
Real Estate Operations	Semiconductors & Semiconductor Equipment
Residential & Commercial REITs	Communications & Networking
	Electronic Equipment & Parts
	Office Equipment
	Computers, Phones & Household Electronics
	Integrated Hardware & Software
	Software & IT Services
	Financial Technology (Fintech) & Infrastructure
	Telecommunications Services
Utilities	
Electrical Utilities & IPPs	
Natural Gas Utilities	
Water & Related Utilities	
Multiline Utilities	

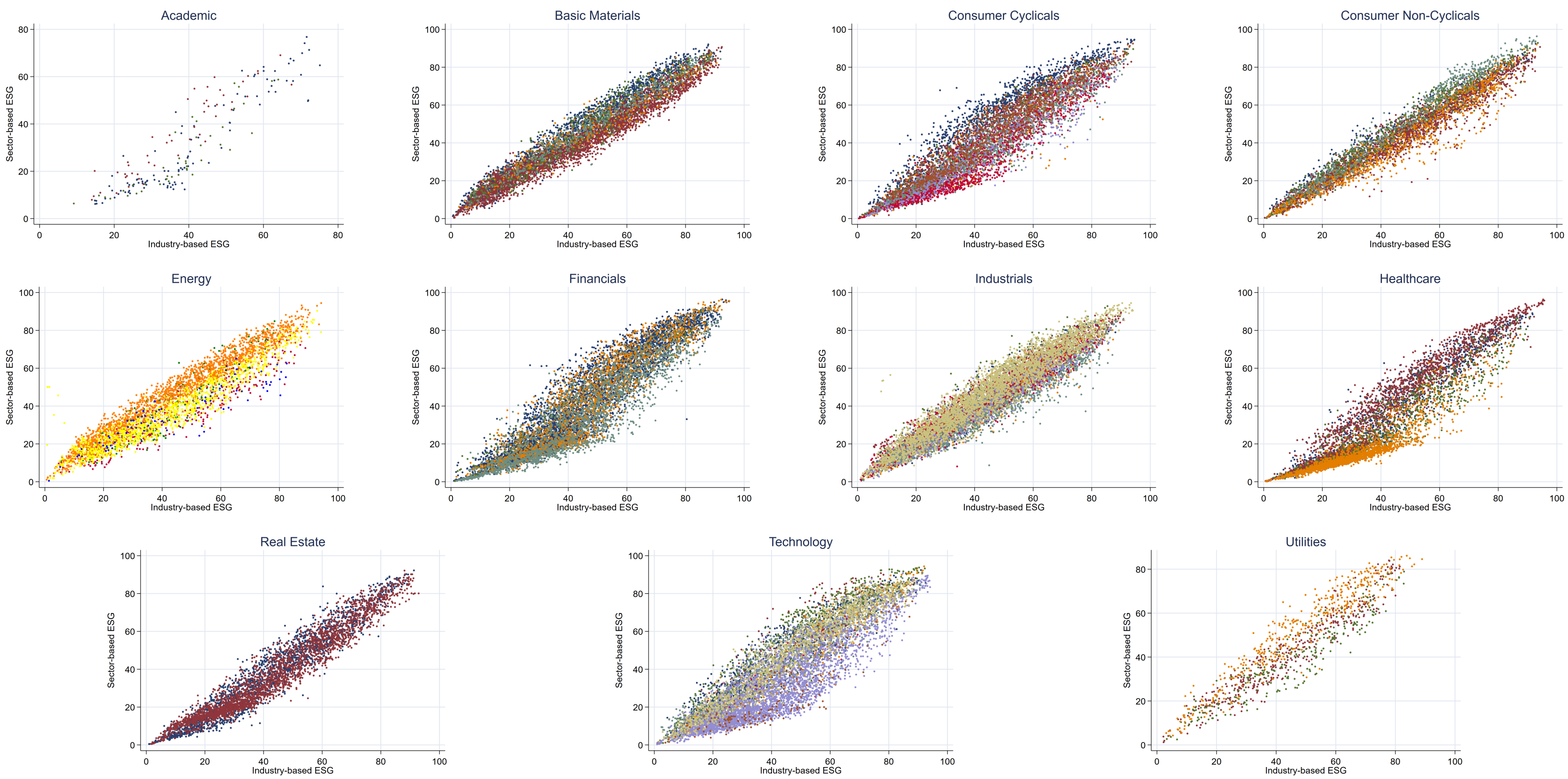
Notes: This table shows sector and industry classifications according to Thomson Reuters Business Classification (TRBC) system. The sectors Government Activity and Institutions are not displayed, since they are equal to their subordinate industry groups.

Table 4.10: Regression results for the original Refinitiv score

	Deciles			Quintiles		
	Annualized Alpha			Annualized Alpha		
	Bottom	Top	Difference	Bottom	Top	Difference
Panel A: Main Results						
Main Results	8.01***	-2.81***	10.82***	5.61***	-1.82*	7.43***
Panel B: Additional Tests						
Alternative Weighting	6.40***	-2.44***	8.84***	6.03***	1.93**	7.96***
Endogeneity Concern	10.32***	-1.34	11.46***	7.20***	-1.33	8.53***
Noise Concern	0.78	-1.78*	2.56	-0.65	-0.85	0.20
Panel C: Robustness Tests						
Alternative Sub-periods						
Excluding Fin. Crisis	9.01***	-2.76**	11.77***	5.78**	-1.74	7.52***
Excluding COVID-19	6.44***	-3.44***	9.88***	5.11***	-2.77***	7.88***
Excluding Both	7.49***	-3.33***	10.82***	5.47**	-2.67***	8.14***
Alternative Factor Models						
Three-factor Alpha	6.34**	-2.11**	8.45***	4.06**	-1.16	5.22**
Six-factor Alpha	8.77***	-2.34**	11.11***	6.59***	-1.25	7.84***
Alternative Rebalancing						
Monthly Rebalancing	7.79***	-2.77***	10.56***	5.55***	-1.82*	7.37***
Pillar Models						
Environmental Pillar	6.55***	-2.23**	8.78***	6.00***	-1.93**	7.93***
Social Pillar	4.94**	-1.69*	6.63***	4.07**	-0.60	4.67***

Notes: This table shows annualized Alpha estimates and differences of the original Refinitiv ESG scores. All values are presented in percent. The portfolios are sorted by ESG scores. Companies with the worst scores are in the bottom portfolio, and companies characterized by the best scores are assigned to the top portfolio, respectively. Those portfolios are rebalanced on a yearly basis. The dependent variable is the equal-weighted monthly return of the ESG portfolios minus the risk-free rate. Independent variables are the following Fama and French (2015) factors: MKT, SMB, HML, RMW, CMA. The results are based on the monthly average returns between February 2010 and December 2022. The difference in annualized Alphas reflects the returns of a long-short portfolio, with long positions in stocks with low ESG ratings and short positions in high ESG stocks. ***, ** and * reflect significance at the 1%, 5%, and 10% level.

Figure 4.1: Scatter plot ESG scores



Notes: This figure shows the distribution of the recalculated and original industry-based score for the sectors academic, basic materials, consumer cyclicals, consumer non-cyclicals, financials, industrials, healthcare, utilities. The sample consists of data points between the years 2010 and 2022.

Bibliography

Albuquerque, R., Koskinen, Y., Yang, S., Zhang, C., 2020. Resiliency of Environmental and Social Stocks: An Analysis of the Exogenous COVID-19 Market Crash. The Review of Corporate Finance Studies 9, 593–621.

Albuquerque, R., Koskinen, Y., Zhang, C., 2019. Corporate Social Responsibility and Firm Risk: Theory and Empirical Evidence. Management Science 65, 4451–4469.

Berg, F., Kölbel, J. F., Rigobon, R., 2022. Aggregate Confusion: The Divergence of ESG Ratings. Review of Finance 26, 1315–1344.

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

Bolton, P., Kacperczyk, M., 2021. Do investors care about carbon risk? Journal of Financial Economics 142, 517–549.

Broadstock, D. C., Chan, K., Cheng, L. T. W., Wang, X., 2021. The role of ESG performance during times of financial crisis: Evidence from COVID-19 in China. Finance research letters 38, 101716.

Carhart, M. M., 1997. On Persistence in Mutual Fund Performance. The Journal of Finance 52, 57–82.

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

Christensen, D. M., Serafeim, G., Sikochi, A., 2022. Why is Corporate Virtue in the Eye of The Beholder? The Case of ESG Ratings. The Accounting Review 97, 147–175.

Dobrick, J., Klein, C., Zwergel, B., 2023. Size bias in refinitiv ESG data. Finance Research Letters 55, 104014.

Douglas, E., van Holt, T., Whelan, T., 2017. Responsible Investing: Guide to ESG Data Providers and Relevant Trends. Journal of Environmental Investing 8, 92–114.

Drempetic, S., Klein, C., Zwergel, B., 2020. The Influence of Firm Size on the ESG Score: Corporate Sustainability Ratings Under Review. Journal of Business Ethics 167, 333–360.

Dumrose, M., Rink, S., Eckert, J., 2022. Disaggregating confusion? The EU Taxonomy and its relation to ESG rating. Finance Research Letters 48, 102928.

Eccles, R. G., Krzus, M. P., Rogers, J., Serafeim, G., 2012. The Need for Sector-Specific Materiality and Sustainability Reporting Standards. Journal of Applied Corporate Finance 24, 65–71.

Edmans, A., 2011. Does the stock market fully value intangibles? Employee satisfaction and equity prices. Journal of Financial Economics 101, 621–640.

El Ghoul, S., Guedhami, O., Kwok, C. C. Y., Mishra, D. R., 2011. Does corporate social responsibility affect the cost of capital? Journal of Banking & Finance 35, 2388–2406.

ESMA, 2021. ESMA letter to EC on ESG ratings.

Fama, E. F., 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance 25, 383–417.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fama, E. F., French, K. R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116, 1–22.

Friede, G., Busch, T., Bassen, A., 2015. ESG and financial performance: Aggregated evidence from more than 2000 empirical studies. Journal of Sustainable Finance & Investment 5, 210–233.

GSIA, G. S. I. A., 2022. Global Sustainable Investment Review 2022. https://www.gsi-alliance.org/wp-content/uploads/2023/12/GSIA-Report-2022.pdf.

Hartzmark, S. M., Sussman, A. B., 2019. Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows. The Journal of Finance 74, 2789–2837.

Heeb, F., Kölbel, J. F., Paetzold, F., Zeisberger, S., 2023. Do Investors Care about Impact? The Review of Financial Studies 36, 1737–1787.

Hong, H., Kacperczyk, M., 2009. The price of sin The effects of social norms on markets. Journal of Financial Economics .

Hübel, B., Scholz, H., 2020. Integrating sustainability risks in asset management: The role of ESG exposures and ESG ratings. Journal of Asset Management 21, 52–69.

Khan, M., Serafeim, G., Yoon, A., 2016. Corporate Sustainability: First Evidence on Materiality. The Accounting Review 91, 1697–1724.

Lins, K. V., Servaes, H., Tamayo, A., 2017. Social Capital, Trust, and Firm Performance: The Value of Corporate Social Responsibility during the Financial Crisis. The Journal of Finance 72, 1785–1824.

Muck, M., Schmidl, T., 2024. Comparing ESG score weighting approaches and stock performance differentiation. Finance Research Letters 67, 105924.

Refinitiv, 2023. Environmental, social and governance scores from LSEG. LSEG .

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 .

Chapter 5

Wish or reality? On the exploitability of triangular arbitrage in cryptocurrency markets

Abstract: This study investigates the efficiency of cryptocurrency markets by examining the presence and exploitability of arbitrage opportunities. Using high-frequency data from the Binance Exchange, we implement a triangular arbitrage strategy, considering Bitcoin, Litecoin, and the U.S. Dollar. We find 4,879 possible arbitrage opportunities. Although these findings suggest potential inefficiencies, transaction costs and limited trading volumes in the order book eliminate their profitability. Consequently, centralized cryptocurrency markets exhibit a high degree of efficiency. Moreover, our results suggest that the mere number of triangular arbitrage opportunities is not a reliable indicator of market inefficiency.

Keywords: Triangular arbitrage, cryptocurrencies, market efficiency

JEL Classification: G14

5.1 Introduction

In recent years, the market for cryptocurrencies has experienced a tremendous growth. Binance, the largest cryptocurrency exchange (Disli et al., 2022), handled over 300 billion transactions in 2022, with an average daily trading volume exceeding 65 billion U.S. Dollar (Binance, 2024). Due to the increasing scale of these markets, ensuring consistent and contradiction-free pricing across assets is a relevant issue. In financial markets, arbitrage mechanisms enforce pricing coherence by eliminating discrepancies between the prices of similar assets. In the relatively new and rapidly growing environment of cryptocurrency markets, the question arises whether such self-correcting mechanisms effectively operate.

In our study, we investigate whether arbitrage opportunities exist in cryptocurrency markets and assess to what extent market participants can exploit them under real-world conditions. To explore this, we implement a triangular arbitrage trading strategy on the Binance Exchange. This strategy has long been employed in traditional currency markets to secure riskless trading returns (Fenn et al., 2009). However, due to the rise of automated trading systems and increasing market efficiency, such opportunities have significantly diminished in foreign exchange markets over time (Ito et al., 2020). We focus on the U.S. Dollar, Bitcoin, and Litecoin, identifying potential triangular arbitrage opportunities. Thus, we analyse how the exchange rates of two cryptocurrencies and the U.S. Dollar are related to each other. By using Bitcoin as the most recognized cryptocurrency and Litecoin as one of its first substitutes (Miglietti et al., 2019), we are in line with existing literature that especially examines those two cryptocurrencies (e.g., Tu and Xue, 2019; Tiwari et al., 2019; Miglietti et al., 2019; Sapkota and Grobys, 2021).

Unlike most previous studies, our paper goes beyond the mere identification of arbitrage possibilities. We conduct an in-depth analysis of the duration of these opportunities and their practical exploitability. Specifically, we examine how factors like transaction costs, latency issues, and order book depth limit the extent to which identified arbitrage opportunities can be leveraged. While textbook cases suggest that triangular arbitrage offers theoretically unlimited profit potential, in reality, these constraints have a restricting character. We raise the question of whether the identified arbitrage opportunities can truly be exploited and, therefore, serve as an indication of potential inefficiencies in this particular market segment.

We make several contributions to the literature on cryptocurrency market efficiency. First, we create a highly granular, high-frequency dataset consisting of bid and ask quotes for the three exchange rates. These quotes are directly obtained from the Binance Exchange through the Binance Websocket. In high-frequency trading, latency plays a crucial role in determining the viability of arbitrage strategies. Although the exact location of Binance's servers is not publicly known, it is assumed that they are based in Japan. Our latency measurements support this assumption. In an initial run with a server located in Europe, we measure an 80-millisecond latency, which we reduce to 4 milliseconds by using a server in Japan. Consequently, we opt for a Tokyo-based server for the analysis. Over the course of a randomly selected week, we collect 30.9 million quotes in total.

Second, during the examined week, we identify a total of 4,879 potential triangular arbitrage opportunities. However, the majority of these offered only marginal returns, with most ranging between 0 percent and 0.025 percent. Taking real transaction fees on the Binance Exchange into account, the profitability of most of these opportunities is effectively eliminated. For regular traders, a small subset of 18 opportunities remains profitable after accounting for transaction costs. For example, traders could realize a total of 2 percent if these opportunities are taken immediately. In a best-case scenario, this sum of relative returns increases to 3 percent. The depth of the order book further restricts the quantities that could be traded. Since the average tradable amount is 4,070.86 U.S. Dollar for the 18 opportunities, this results in marginal arbitrage profits between 12.43 U.S. Dollar and 17.73 U.S. Dollar for the entire week. Therefore, we conclude that cryptocurrency markets exhibit a high degree of efficiency. Our results also suggest that the mere number of triangular arbitrage opportunities is not a reliable indicator of market inefficiency, as many of these opportunities are unprofitable to be exploited in practice.

Third, there is a time lag between the observed and realized prices. Hence, slippage[22] is a critical issue for the implementation of triangular arbitrage strategies. Under the condition that the expected value of the arbitrage trades should be profitable, we estimate the execution

[22]Slippage means the price change between the time a trader identifies or places a market order and the settled price. This is specifically relevant for volatile assets and illiquid market segments.

time for a trade in our high-frequency data. Results suggest that a trader needs to execute the triangular arbitrage strategy in not more than 146 milliseconds to be profitable. Otherwise, the slippage risk would exceed the profitability of arbitrage trades.

The remainder of the study proceeds as follows: Section 5.2 reviews the related literature and details our research question. Section 5.3 describes the data and methodology used. Section 5.4 presents and discusses our results, and Section 5.5 concludes.

5.2 Related literature

Cryptocurrencies such as Bitcoin and others have seen huge price explosions in the past (Bouri et al., 2019). Such new, promising market segments are intended to generate rapid, risk-free returns. For example, Makarov and Schoar (2020) find profitable arbitrage on different cryptocurrency exchanges across countries and, therefore, suggest that a lack of regulation supports those price deviations. Those findings are confirmed by Crépellière et al. (2023). However, examining a longer period, they point out a significant decline in risk-free arbitrage opportunities for market participants over time. They suggest, that increased institutional involvement and retail trader attention lead to higher informed activity in this market segment. In addition, using different specifications of statistical arbitrage, Fischer et al. (2019) and Leung and Nguyen (2019) observe significant returns for their strategies. Besides these strategies, triangular arbitrage is another method to exploit inefficiencies in currency markets. Triangular arbitrage capitalizes on discrepancies between three different trading pairs or currencies (Aiba et al., 2002). This strategy has been extensively studied in foreign exchange markets by, for example, Aiba and Hatano (2006), Fenn et al. (2009), and Akram et al. (2008), who investigate how these pricing inefficiencies arise and diminish in traditional markets. More recently, research has turned to examining arbitrage in decentralized cryptocurrency markets. Wang et al. (2022) implement a cyclical arbitrage strategy on Uniswap, where trades occur via smart contracts rather than through a traditional order book, and where converting cryptocurrencies into fiat money is not feasible. In addition, Bewaji et al. (2021) explore various arbitrage strategies and find that the exchange platform itself plays a critical role in identifying price asymmetries. However, both studies leave open the question of why those are not being exploited and omit key real-world factors, such as transaction costs for traders and the risk of slippage, which can significantly impact the feasibility and profitability of arbitrage strategies in practice. By incorporating these considerations, our study aims to provide a more comprehensive understanding of arbitrage opportunities in cryptocurrency markets.

Arbitrage-free trading is one condition for efficient markets. Moreover, it is a fundamental requirement for a consistent pricing[23] system. Vice versa, the efficiency of a market segment also implies whether arbitrage opportunities could be observed and potentially exploited by market participants. Given that the markets for cryptocurrencies are unregulated, this question is of particular relevance. The semi-strong form of the efficient market hypothesis (EMH) states that the prices of securities fully and immediately reflect all publicly available information (Fama, 1970). Consequently, there is no systematic mispricing in an efficient market, and risk-free exploitation through arbitrage is not possible. In contrast, literature exhibits inefficiencies in cryptocurrency markets. Joo et al. (2020) show that information is not immediately priced, and Aloosh and Ouzan (2020) find evidence of a small-price bias, whereby investors react more strongly to news when the price of a cryptocurrency is low. However, there are several studies suggesting that cryptocurrency markets are, to a certain degree, efficient (e.g., Wang and Chong, 2021; Vidal-Tomas and Ibanez, 2018; Alvarez-Ramirez and Rodriguez, 2021). This development is strengthened by the introduction of Bitcoin futures (Köchling et al., 2019). Moreover, Burggraf and Rudolf (2021) conclude in their analysis of low-volatility anomalies for cryptocurrencies that the markets are more efficient than previously assumed. An additional perspective views cryptocurrency markets through the lens of an adaptive market hypothesis, where market efficiency is not static but instead varies. For example, Chu et al. (2019), Al-Yahyaee et al. (2020), and Khuntia and Pattanayak (2021) find that cryptocurrencies exhibit different levels of efficiency over time.

[23]Recent literature examines the pricing of cryptocurrencies. For example, trading volume and volatility in cryptocurrency markets are subject to seasonal patterns (Kaiser, 2019). Therefore, there is a growing literature explaining these returns. For example, in line with Fama and French (1992), Liu et al. (2022) construct a cryptocurrency market, size and momentum factor to explain the returns through a standard asset-pricing model. Alternative approaches exist, like Liu and Tsyvinski (2021) suggesting that network effects, such as user adoption, are additional value drivers. Moreover, the interaction of investors and traders on social platforms contributes to the price increases of meme stocks and cryptocurrencies (Aloosh et al., 2022).

5.3 Methodology and data

According to Fenn et al. (2009), a triangular arbitrage opportunity can be quantified through an exchange rate product and is given as:

$$Y(t) = \prod_{i=1}^{3} r_{\mathrm{i}}(t), \tag{5.1}$$

where $r_i(t)$ represents the respective exchange rate and $Y(t)$ the exchange rate product at time t. A profitable triangular arbitrage opportunity exists if $Y(t) > 1$. We consider the three currencies for the triangular trading strategy: U.S. Dollar (USD), Bitcoin (BTC), and Litecoin (LTC). The triangular arbitrage sequence is $USD \rightarrow BTC \rightarrow LTC \rightarrow USD$. Starting with an initial amount of X_S in U.S. Dollar, this is converted into Bitcoin, then into Litecoin, and finally back into an end amount X_E in U.S. Dollar. The trading strategy is profitable if $X_E > X_S$. As the market tends to be more efficient, the difference $X_E - X_S$ diverges towards zero. For the considered sequence of the three trading pairs, the rate product is given as:

$$Y(t) = \frac{1}{BTC/USDT_{\mathrm{Ask}}(t)} \cdot \frac{1}{LTC/BTC_{\mathrm{Ask}}(t)} \cdot LTC/USDT_{\mathrm{Bid}}(t), \tag{5.2}$$

where $Y(t)$ is the exchange rate product at time t. The trading strategy exploits the arbitrage opportunity if $Y(t) > 1$.

We collect high-frequency data from the order book of the Binance Exchange between the 06/03/2024 and the 06/09/2024.[24] Using a websocket connection containing real-time order book data, we gather the best-ask and best-bid prices for the three trading pairs, as well as the quantities offered and demanded at these prices. Additionally, time and latency data are collected. We calculate the exchange rate product and the corresponding return percentage for each data point separately. To measure the duration, the period between the moment when the exchange rate product exceeds 1 and the moment it falls back below 1 is recorded. The latency is measured as the temporal difference between the actual time of the change and the time when the corresponding message was received. The final data set

[24] Beginning at 0:00 AM GMT and ending at 0:00 PM GMT.

contains 30.9 million exchange rate products. It is important to note, that each change in one of the variables, for example, a change in ask-quantities, leads to the calculation of a new exchange rate product. Under this assumption, a triangular arbitrage opportunity can therefore comprise multiple rate products. Table 5.1 illustrates such a case. While the return percentage stays the same, the quantities are changing rapidly due to high market volatility. To address this issue, we clustered the data associated with each arbitrage opportunity.

Since there could be a deviation between the observed and the actually settled price, we assume the absence of slippage. We further assume, that a trader only trades those quantities for which there are supply/demand quantities in the order book at the underlying ask and bid prices. Under this assumption, the lowest quantity offered/demanded is relevant in each case, as this is the limiting factor. The maximum tradable value $MaxUSD(t)$ is defined as:

$$\mathrm{MaxUSD(t)} = \min \begin{pmatrix} \mathrm{BTC/USDT_{Ask\_Qty}(t)} \times \mathrm{BTC/USDT_{Ask}(t)}, \\ \mathrm{LTC/BTC_{Ask\_Qty}(t)} \times \mathrm{LTC/BTC_{Ask}(t)} \times \mathrm{BTC/USDT_{Ask}(t)}, \\ \mathrm{LTC/USDT_{Bid\_Qty}(t)} \times \mathrm{LTC/USDT_{Bid}(t)} \end{pmatrix} \tag{5.3}$$

5.4 Results

Before testing whether there are inefficiencies in the cryptocurrency market, we have to uniquely define the emergence of an arbitrage opportunity. The most obvious approach is that a triangular arbitrage opportunity occurs when the exchange rate product rises above the threshold of 1 and falls back below 1. As a result, the achievable return during this period is continuously positive. Since high-frequency exchange rates are subject to volatility, different levels of returns and tradable quantities can occur within this time period. Therefore, a second approach is to consider every change in the rate product and define it as an individual opportunity. This results in significantly more executable trades. Figure 5.1 shows the temporal development of returns within a triangular arbitrage opportunity with a total duration of approximately 1.9 seconds. In the absence of transaction costs, an infinitely fast trader could execute every change in the exchange rate product, resulting in over 140 possible trades. However, since the exchange rates change over a fraction of a millisecond

Table 5.1: Tradable quantities for an exemplary triangular arbitrage opportunity

Return	BtcUsdtAskQty	LtcBtcAskQty	LtcUsdtBidQty	Time
0.015490622	1.15597	22.841	21.144	23:07:49.896938
0.015490622	1.15597	22.841	36.998	23:07:49.897050
0.015490622	1.15597	22.841	41.399	23:07:49.897150
0.015490622	1.15597	22.841	59.519	23:07:49.897276
0.015490622	1.15597	22.841	59.519	23:07:49.897374
0.015490622	1.15597	22.841	43.665	23:07:49.897474
0.015490622	1.15597	54.008	43.665	23:07:49.897624
0.015490622	1.15597	22.841	43.665	23:07:49.898155

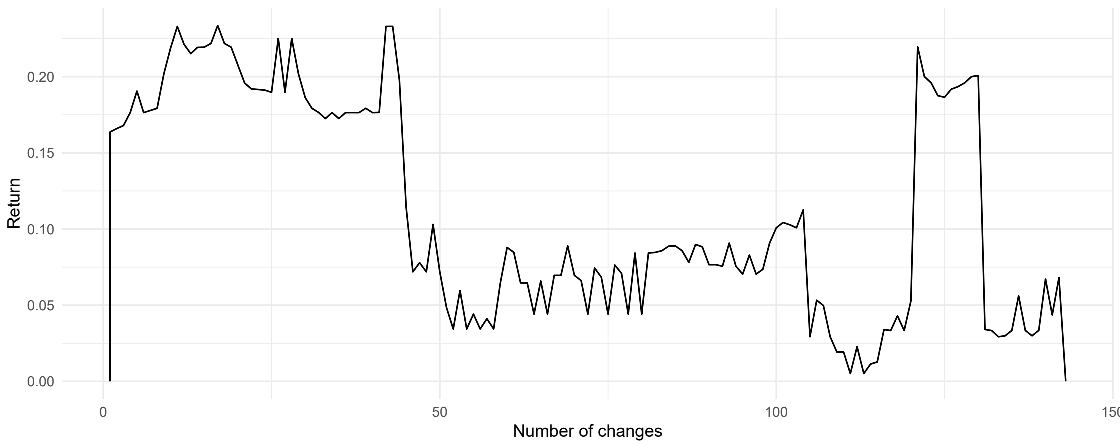
Notes: This table shows tradable quantities of the triangular arbitrage strategy ($USD \rightarrow BTC \rightarrow LTC \rightarrow USD$) and the corresponding return. The return, without including transaction costs, is presented in percent. BtcUsdtAskQty means the ask quantity of the exchange pair of Bitcoin and U.S. Dollar; LtcBtcAskQty means the ask quantity of the exchange pair of Litecoin and Bitcoin; and LtcUsdtBidQty means the bid quantity of the exchange pair of Litecoin and U.S. Dollar.

and the duration cannot be determined in advance, realistically, no trader is fast enough to take advantage of these opportunities. The majority of traders will therefore exploit the opportunity only once when the expected return exceeds the transaction costs by a certain margin. For the following analysis, we follow the first approach and count the initial passing of the threshold as one arbitrage opportunity.

Figure 5.2 displays the development of the rate product over the observed period. Assuming an efficient market, the value of the exchange rate product should be 1 or just below this threshold (Fenn et al., 2009). Our findings show that, for the majority of the time, the exchange rate product remains below 1. During these periods, it is not possible to profit from a triangular arbitrage trading strategy. However, there are profitable instances when the exchange rate product exceeds the threshold of 1. This suggests that the cryptocurrency markets contain inefficiencies that can theoretically be exploited by arbitrage traders.

Table 5.3 presents the number of arbitrage opportunities by return (Panel A), duration (Panel B), and tradable quantity (Panel C). We identify 4,879 potential triangular arbitrage opportunities. The majority (4,337 observations) have a return between 0 percent and 0.025 percent. But we find three observations with a return larger than 0.5 percent. Simultaneously, approximately 65 percent of these arbitrage opportunities are very short and last one second or less, while 767 observations (15.72 percent) exhibit a relatively long duration of

Figure 5.1: Yield curve within an arbitrage opportunity

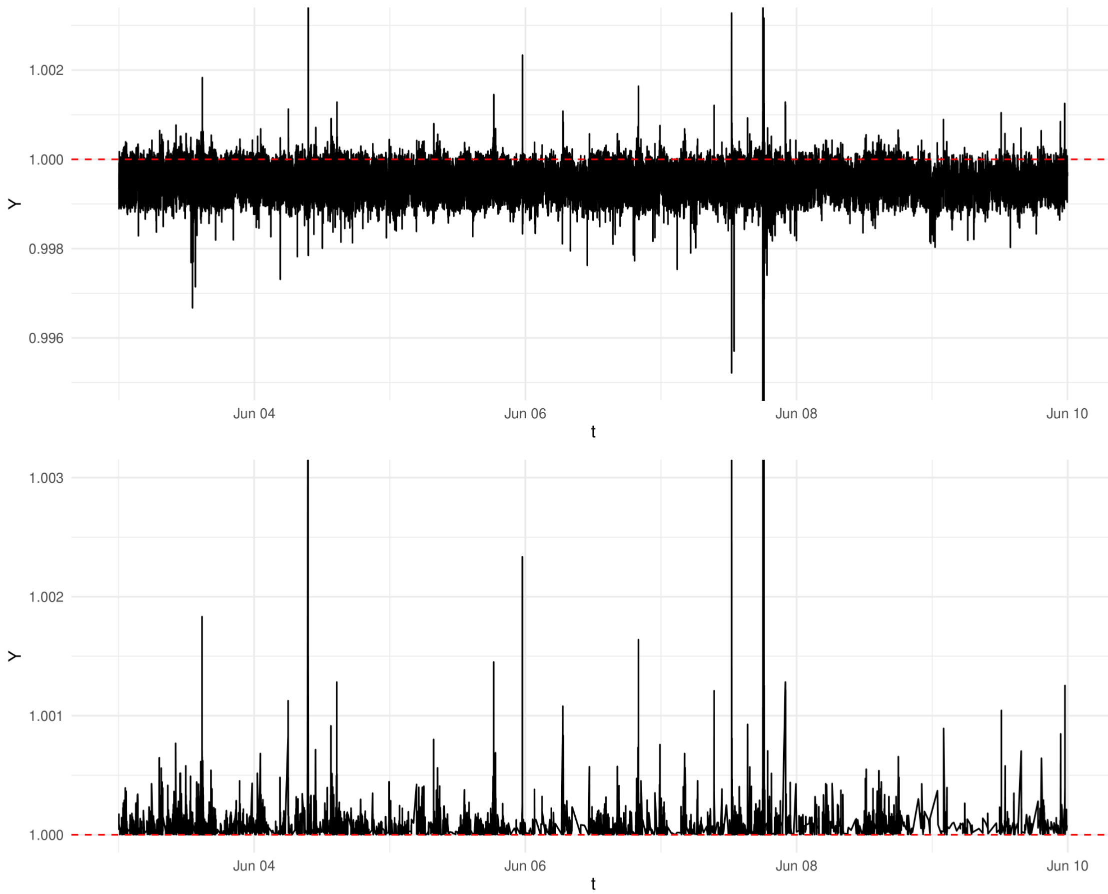


Notes: This figure shows the number of profitable possible trades during a triangular arbitrage opportunity. The vertical axis displays the return without including transaction costs for one potential trade. The return is presented in percent. The horizontal axis displays the number of changes during the profitable arbitrage opportunity.

more than 5 seconds. This observation aligns with Fenn et al. (2009), who find similar values for the forex market. Therefore, the time to react is an inherent factor for the success of the trading strategy.

Binance offers different trading fee levels depending on how much volume a market participant trades per month. They are called VIP levels, and VIP 9 indicates the lowest transaction costs, while a regular trader has to pay the highest fees. Table 5.2 illustrates the number of profitable triangular arbitrage opportunities after considering transaction costs. Not all the 4,879 observed triangular arbitrage opportunities are profitable, as the transaction costs reduce the potential return. 96.93 percent of the triangular arbitrage opportunities are not even profitable for VIP 9 traders. This finding is in line with Wang et al. (2022), who find that a large proportion of cryptocurrency arbitrage opportunities are not exploited. Consequently, this is due to the fact that these are not profitable for market participants paying regular trading fees. However, when considering transaction costs for a VIP 9 trader, there are still 150 opportunities with a return greater than the transaction costs. The lower the VIP level, the fewer arbitrage opportunities are profitable. As a consequence, 18 trading opportunities remain for a regular investor. Transaction costs are accounted for based on the Binance spot trading fees. The transaction fees on the Binance exchange are determined by the 30-day trading volume, the balance of Binance Coins (BNB), as well as the selected

Figure 5.2: Exchange rate product over time



Notes: Those figures show the development of the exchange rate product ($USD \rightarrow BTC \rightarrow LTC \rightarrow USD$) over the examination period. The vertical axis displays the exchange rate product, and the horizontal axis displays the time. The upper figure presents all exchange rate products. All products above the red line reflect a profitable triangular arbitrage opportunity without including transaction costs. The lower figure displays only the profitable arbitrage opportunities without including transaction costs. The period starts on the 06/03/2024 and ends on the 06/09/2024.

order type.[25] Consequently, a triangular arbitrage opportunity is profitable only when the achievable return exceeds the transaction costs. To analyse the profitability of triangular arbitrage opportunities for traders, we have to define the exact moment of execution. We consider four different cases, including the first, the best profit, the highest tradable value, and the average execution opportunity leading to different levels of return. The first case assumes that the trader exploits the arbitrage opportunity if the exchange rate product passes over the threshold of 1 and the return rises above the transaction costs. The second case assumes that the trader could realize the maximum profit in each triangular arbitrage trade. This value can be achieved in reality by specifying a certain return level that must be

[25] See appendix for further explanation.

Table 5.2: Number of profitable arbitrage opportunities by Binance VIP level

VIP Level	30-Day Trade Volume	and/or	BNB Balance	Transaction Costs	Rel. Frequency	Total
/	/	/	/	/	96.93	4,729
VIP 9	$\geq$ 4,000	and	$\geq$ 5,500	0.0540	3.07	150
VIP 8	$\geq$ 2,000	and	$\geq$ 4,500	0.0675	2.07	101
VIP 7	$\geq$ 800	and	$\geq$ 3,000	0.0810	1.62	79
VIP 6	$\geq$ 400	and	$\geq$ 1,750	0.0945	1.27	62
VIP 5	$\geq$ 150	and	$\geq$ 1,000	0.1080	1.09	53
VIP 4	$\geq$ 100	and	$\geq$ 500	0.1215	0.94	46
VIP 3	$\geq$ 20	and	$\geq$ 250	0.1350	0.78	38
VIP 2	$\geq$ 5	and	$\geq$ 100	0.2250	0.37	18
VIP 1	$\geq$ 1	and	$\geq$ 25	0.2250	0.37	18
Regular	$<$ 1	or	$\geq$ 0	0.2250	0.37	18

Notes: This table shows the number of profitable triangular arbitrage opportunities, including transaction costs. The "30-Day Trade Volume" is presented in million U.S. Dollars. The "BNB Balance" (Binance Coin) is presented in Binance Coins. The transaction costs are presented in percent. The relative frequency is displayed in percent, and Total is the absolute number of arbitrage opportunities.

reached at least for the trade to be executed. The next case assumes that the trader realizes the highest tradable quantity for a given return. For the average arbitrage opportunity, we consider the mean return (reduced by the transaction costs) and the mean quantity during the arbitrage period. Table 5.4 illustrates the distribution of returns after considering transaction costs (Panel A) for a regular trader (highest transaction fees) and a VIP 9 trader (lowest transaction fees). If traders exploit the arbitrage opportunity at the first moment when the return exceeds the transaction fees, they achieve a sum of returns of 2.01 percent during the week. Investors exploiting the best return opportunity could increase their return to 2.97 percent. As mentioned earlier, in practice, traders will only trade those quantities for which there are supply/demand quantities in the order book at the underlying ask and bid prices. This limits the possible net profits to 12.43 U.S. Dollar or 17.73 U.S. Dollar, respectively. Panel B displays the tradable value of the four cases. The majority of arbitrage opportunities range between a tradable value of 0 U.S. Dollar and 500 U.S. Dollar. The sum of initial quantities (10,650.26 U.S. Dollar) under the first strategy is substantially lower than the amount achieved when traders follow the highest tradable value strategy to maximize trading volume (17,947.96 U.S. Dollar), while the average traded amount is 4,070.86 U.S. Dollar. In addition, we find more profitable triangular arbitrage opportunities for a VIP 9 trader with higher corresponding quantities. Consequently, they could achieve higher net returns, up to a maximum of 170.76 U.S. Dollar. However, if we relate this finding to the condition that those market participants need at least a trading turnover of 4 billion U.S. dollars per month, the possible profit is negligible. In summary, our results indicate that transaction costs and order book depth restrict possible profits in such a way that market participants could not achieve a risk-free return on cryptocurrency markets.

Since the trading quantity of the three exchange rates is a crucial factor for the profitability of the triangular arbitrage, we further identify the "bottleneck" of our strategy. Table 5.5 presents the relative number of exchange rates as the limiting factor. We find that for a VIP 9 trader, the cross-rate of Bitcoin and Litecoin ($BTC \rightarrow LTC$) reduces the possible profits of the trading strategy in around half of the cases. If the arbitrage opportunity is exploited immediately, the most liquid trading pair, U.S. Dollar and Bitcoin ($USD \rightarrow BTC$), limits the quantity of our strategy in 10 percent of the opportunities. In contrast, the results for traders with the highest fees, on the other hand, are not unambiguous. The two Litecoin exchange rates ($BTC \rightarrow LTC$ and $LTC \rightarrow USD$) limit the arbitrage opportunities. However, this finding is strongly influenced by outliers due to the small number of observations. In summary, this finding reflects the characteristics of foreign exchange markets for traditional currencies.[26]

An important concern is that the time to identify and execute the arbitrage opportunity exceeds the time period when it is profitable. When collecting the data, it takes an average of 4 milliseconds to identify a triangular arbitrage opportunity. However, executing the trades involves an additional latency. As mentioned earlier, the maximum return cannot be achieved over the entire duration of the arbitrage opportunity. Instead, the achievable return fluctuates during phases with high volatility. Therefore, for a practical assessment of these opportunities, the entire duration of the arbitrage opportunity cannot be considered. Rather, the relevant duration is the period during which the return exceeds the transaction costs. Within the 18 profitable opportunities for a trader, we identify a total of 34 phases with a return greater than the transaction costs. We assume that a trader uses the first phase, in which the return exceeds the transaction fees. This results in 18 phases, each with

[26] Triangular arbitrage involves exchange rates (with respect to the USD) and cross-rates (exchange rate of one currency with respect to another one but not the USD). For example, trading EUR and JPY with respect to the USD is more relevant than the EUR with the JPY directly.

Table 5.3: Number of arbitrage opportunities by return, duration, and value

	Frequency	Rel. Frequency	Total	Rel. Total
Panel A: Return				
0.000 - 0.025	4,337	88.89	4,337	88.89
0.025 - 0.050	371	7.60	4,708	96.50
0.050 - 0.075	85	1.74	4,793	98.24
0.075 - 0.100	26	0.53	4,819	98.77
0.100 - 0.200	41	0.84	4,860	99.61
0.200 - 0.300	9	0.18	4,869	99.80
0.300 - 0.400	5	0.10	4,874	99.90
0.400 - 0.500	2	0.04	4,876	99.94
> 0.500	3	0.06	4,879	100.00
Panel B: Duration				
0.000 - 0.500	2,825	57.90	2,825	57.90
0.500 - 1.000	322	6.60	3,147	64.50
1.000 - 2.000	440	9.02	3,587	73.52
2.000 - 3.000	223	4.57	3,810	78.09
3.000 - 4.000	167	3.42	3,977	81.51
4.000 - 5.000	135	2.77	4,112	84.28
> 5.000	767	15.72	4,879	100.00
Panel C: Tradable Value				
0 - 1,000	1,636	33.53	1,636	33.53
1,000 - 2,000	1,044	21.40	2,680	54.93
2,000 - 3,000	936	19.18	3,616	74.11
3,000 - 4,000	435	8.92	4,051	83.03
4,000 - 5,000	396	8.12	4,447	91.15
5,000 - 10,000	392	8.03	4,839	99.18
> 10,000	40	0.82	4,879	100.00

Notes: This table shows summary statistics of the return, the duration, and the tradable value of the triangular arbitrage opportunities. Panel A displays the number of arbitrage opportunities by return without transaction costs. The return is presented in percent. Panel B displays the number of arbitrage opportunities by duration. The duration is presented in seconds. Panel C displays the number of arbitrage opportunities by tradable value. The tradable value is presented in U.S. Dollar.

Table 5.4: Number of profitable arbitrage opportunities by return and value

	Regular				VIP 9			
	First	Best Profit	Highest Value	Average	First	Best Profit	Highest Value	Average
Panel A: Return								
0.000 - 0.025	6	5	3	3	86	68	58	58
0.025 - 0.050	1	2	1	1	26	24	18	22
0.050 - 0.075	2	1	2	2	18	18	13	7
0.075 - 0.100	2	1	1	0	7	10	4	9
0.100 - 0.200	5	6	4	3	7	17	7	8
0.200 - 0.300	0	0	0	1	4	6	4	3
0.300 - 0.400	1	1	1	0	2	4	4	4
0.400 - 0.500	0	1	1	0	0	0	0	0
>0.500	1	1	0	0	0	3	2	0
$\sum$ Total Return	2.01	2.97	1.66	0.90	6.21	10.83	6.98	5.89
Panel B: Value								
0 - 500	12	12	7	7	101	100	62	67
500 - 1,000	2	3	0	2	15	14	5	17
1,000 - 1,500	1	0	0	1	4	7	4	5
1,500 - 2,000	1	3	3	0	4	7	8	7
2,000 - 3,000	2	0	1	0	13	10	12	8
3,000 - 4,000	0	0	1	0	4	4	3	1
4,000 - 5,000	0	0	0	0	1	1	2	2
5,000 - 10,000	0	0	1	0	6	5	8	2
>10,000	0	0	0	0	2	2	6	2
$\sum$ Total Tradable Value	10,650.26	9,257.36	17,947.96	4,070.86	188,505.13	179,800.77	254,966.18	123,588.79
Panel C: Total Net Profit								
	12.43	17.73	30.42	4.77	82.72	100.48	170.76	50.53

Notes: This table presents profitable arbitrage opportunities after subtracting transaction costs. Panel A displays the number of profitable opportunities by return. The return is presented in percent. Panel B displays the number of profitable arbitrage opportunities by tradable value. The value is presented in U.S. Dollars. The column "First" assumes that the trader exploits the triangular arbitrage opportunity immediately if the exchange rate product passes the threshold of 1 and the return is higher than the transaction costs; the column "Best Profit" assumes that the trade is realized at the maximum return by the corresponding tradable value; the column "Highest Value" assumes that the trade is realized at the maximum tradable value by the corresponding return; and the column "Average" assumes the means of return and tradable value during an arbitrage opportunity. Panel C displays the total net profit in U.S. Dollar during the examined week.

Table 5.5: Quantities by trading pairs

	Regular				VIP 9			
	Trading Value	$USD \rightarrow BTC$	$BTC \rightarrow LTC$	$LTC \rightarrow USD$	Trading Value	$USD \rightarrow BTC$	$BTC \rightarrow LTC$	$LTC \rightarrow USD$
Panel A: First								
Total	10,650.27	451,624.21	365,133.72	25,768.32	188,505.13	35,863,109.19	1,041,792.30	595,830.92
Average per Day	1,521.47	64,517.74	52,161.96	3,681.19	26,929.30	5,123,301.31	148,827.47	85,118.70
Rel. Amount		22.22	44.44	33.33		10.67	50.67	38.67
Panel B: Best Profit								
Total	9,257.36	393,741.86	345,506.23	17,085.52	179,800.77	34,876,736.82	1,017,469.32	576,552.09
Average per Day	1,322.48	56,248.84	49,358.03	2,440.79	25,685.82	4,982,390.97	145,352.76	82,364.58
Rel. Amount		16.67	33.33	50.00		14.00	45.33	40.67
Panel C: Highest Value								
Total	17,947.97	290,463.55	332,273.82	24,029.92	254,966.18	34,549,980.87	848,097.84	503,110.71
Average per Day	2,564.00	41,494.79	47,467.69	3,432.85	36,423.74	4,935,711.55	121,156.83	71,872.96
Rel. Amount		0.00	41.67	66.67		4.55	53.64	41.82
Panel D: Average								
Total	4,070.86	159,681.01	162,992.96	11,565.93	123,588.79	26,517,295.01	555,177.72	394,512.36
Average per Day	581.55	22,811.57	23,284.71	1,652.28	17,655.54	3,788,185.00	79,311.10	56,358.91
Rel. Amount		0.00	50.00	50.00		5.41	53.15	41.44

Notes: This table presents the trading pair quantities of the profitable arbitrage opportunities after subtracting transaction costs. "Total" means the U.S. Dollar amount of the sum of the quantities for every trading. "Rel. Amount" means the percentage of cases when the appropriate quantity of the trading pair limits the profit of the arbitrage opportunity. Panel A "First" assumes that the trader exploits the triangular arbitrage opportunity immediately if the exchange rate product passes the threshold of 1 and the return is higher than the transaction costs; Panel B "Best Profit" assumes that the trade is realized at the maximum return by the corresponding tradable value; Panel C "Highest Value" assumes that the trade is realized at the maximum tradable value by the corresponding return; and Panel D "Average" assumes the means of return and tradable value during an arbitrage opportunity.

a specific duration and a certain achievable return. If a trader manages to trade quickly enough, they achieve a profit equal to the return minus the transaction costs. If they do not manage to trade quickly enough, we assume that they trade at the next available price. In this case, they realize a loss as the transaction costs exceed the return. The expected value for each possibility is then calculated as:

$$EV_i(T) = \begin{cases} G_i & \text{if } T \leq D_i \\ -L_i & \text{if } T > D_i, \end{cases} \tag{5.4}$$

where G_i is the possible gain, L_i is the possible loss, D_i is the duration of the arbitrage possibility i, and T the trading duration. For the total expected value $GEV(T)$ as a function of the trading duration, the following applies:

$$GEV(T) = \sum_{i=1}^{18} EV_i(T). \tag{5.5}$$

For a trading strategy to be profitable, the sum of gains must exceed the sum of losses. This means that the total expected value must be positive. Through an iteration process, the duration of T can be determined up to the point where the expected value remains positive. Consequently, a regular trader has to execute trades on average within a maximum of 146 milliseconds, based on the data. In order to reduce the risk of slippage, the strategy must be implemented that, if the arbitrage opportunity is recognized, trading should take place within these 146 milliseconds in order to avoid making a loss on average. This result emphasizes the location advantage in high-frequency trading. We interpret this finding as a potential filter criteria that allows us to distinguish between real arbitrage opportunities and those, which exist just in theory, as their slippage risk exceeds the potential profits.

5.5 Conclusion

By implementing a triangular arbitrage strategy under real circumstances at the Binance Exchange, we analyse 4,879 arbitrage opportunities in a highly granular, high-frequency data set covering a randomly chosen trading week. Considering practical frictions such as transaction costs, a limited order book, and latencies, we find that these arbitrage opportunities only offer marginal net profits. In relation to the necessary trading volumes, the possible trading profits seem negligible. Thus, our findings indicate that the market segment for Bitcoin, Litecoin, and the U.S. Dollar is free of arbitrage and suggest that the mere size of triangular arbitrage opportunities is a weak indicator of market efficiency, since most of these opportunities are not exploitable in a realistic setting. In addition, we propose a detailed filter criteria for the duration of arbitrage opportunities to be profitable in terms of slippage. This has practical implications for future research and traders to distinguish between real exploitable arbitrage opportunities and those which exist just in theory.

Possible limitations of our study are that we could only analyse the best bid- and ask-quotes. We did not have further information on the depth of the order book, implying potentially further trading opportunities. However, these trading opportunities can only be executed at less favourable prices. We leave this question for further research to determine whether these more unfavourable quotes lead to exploitable triangular arbitrages under realistic conditions. Moreover, we examine a combination of a "major" cryptocurrency (Bitcoin) and a "smaller" cryptocurrency (Litecoin). By including Litecoin, we also align with the existing literature. However, it remains to be evaluated to what extent our results could be generalized to other adjustments, such as using two "major" (e.g., Bitcoin and Ethereum) or two "smaller" cryptocurrencies. Differences may arise due to the lower trading depth in Litecoin, potentially leading to fewer arbitrage opportunities. Conversely, larger order books and potentially smaller bid-ask spreads in "major" cryptocurrencies like Bitcoin and Ethereum could facilitate greater arbitrage volumes.

Furthermore, we randomly selected an examination period, that is free of extraordinary market events. An open question for future research is the existence and characteristics of triangular arbitrage in stress scenarios. It is likely that such opportunities would become more prevalent in highly volatile markets, potentially accompanied by turbulence, forced sales, or restricted capital access for arbitrageurs. Since the examination of these additional circumstances could contribute to further generalizability, we leave those questions for future research.

Bibliography

Aiba, Y., Hatano, N., 2006. A microscopic model of triangular arbitrage. Physica A: Statistical Mechanics and its Applications 371, 572–584.

Aiba, Y., Hatano, N., Takayasu, H., Marumo, K., Shimizu, T., 2002. Triangular arbitrage as an interaction among foreign exchange rates. Physica A: Statistical Mechanics and its Applications 310, 467–479.

Akram, Q. F., Rime, D., Sarno, L., 2008. Arbitrage in the foreign exchange market: Turning on the microscope. Journal of International Economics 76, 237–253.

Al-Yahyaee, K. H., Mensi, W., Ko, H.-U., Yoon, S.-M., Kang, S. H., 2020. Why cryptocurrency markets are inefficient: The impact of liquidity and volatility. The North American Journal of Economics and Finance 52, 101168.

Aloosh, A., Ouzan, S., 2020. The psychology of cryptocurrency prices. Finance Research Letters 33, 101192.

Aloosh, A., Ouzan, S., Shahzad, S. J. H., 2022. Bubbles across Meme Stocks and Cryptocurrencies. Finance Research Letters 49, 103155.

Alvarez-Ramirez, J., Rodriguez, E., 2021. A singular value decomposition approach for testing the efficiency of Bitcoin and Ethereum markets. Economics Letters 206, 109997.

Bewaji, O., Hamid, S., Aerts, T., Byck, S., Heijmans, R., van der Woerd, E., 2021. Are crypto currencies cryptic or a source for arbitrage?, Payments Canada and De Nederlandsche Bank.

Binance, 2024. https://www.binance.com/en/about.

Bouri, E., Shahzad, S. J. H., Roubaud, D., 2019. Co-explosivity in the cryptocurrency market. Finance Research Letters 29, 178–183.

Burggraf, T., Rudolf, M., 2021. Cryptocurrencies and the low volatility anomaly. Finance Research Letters 40, 101683.

Chu, J., Zhang, Y., Chan, S., 2019. The adaptive market hypothesis in the high frequency cryptocurrency market. International Review of Financial Analysis 64, 221–231.

Crépellière, T., Pelster, M., Zeisberger, S., 2023. Arbitrage in the market for cryptocurrencies. Journal of Financial Markets 64, 100817.

Disli, M., Abd Rabbo, F., Leneeuw, T., Nagayev, R., 2022. Cryptocurrency comovements and crypto exchange movement: The relocation of Binance. Finance Research Letters 48, 102989.

Fama, E. F., 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance 25, 383–417.

Fama, E. F., French, K. R., 1992. The Cross-Section of Expected Stock Returns. Journal of Finance 47, 427–465.

Fenn, D. J., Howison, S. D., McDonald, M., Williams, S., Johnson, N. F., 2009. The Mirage of Triangular Arbitrage in the Spot Foreign Exchange Market. International Journal of Theoretical and Applied Finance 12, 1105–1123.

Fischer, T. G., Krauss, C., Deinert, A., 2019. Statistical Arbitrage in Cryptocurrency Markets. Journal of Risk and Financial Management 12, 31.

Ito, T., Yamada, K., Takayasu, M., Takayasu, H., 2020. Execution risk and arbitrage opportunities in the foreign exchange markets. Working paper / National Bureau of Economic Research, Inc., Cambridge, MA : National Bureau of Economic Research.

Joo, M. H., Nishikawa, Y., Dandapani, K., 2020. Announcement effects in the cryptocurrency market. Applied Economics 52, 4794–4808.

Kaiser, L., 2019. Seasonality in cryptocurrencies. Finance Research Letters 31.

Khuntia, S., Pattanayak, J., 2021. Adaptive calendar effects and volume of extra returns in the cryptocurrency market. International Journal of Emerging Markets 17, 2137–2165.

Köchling, G., Müller, J., Posch, P. N., 2019. Does the introduction of futures improve the efficiency of Bitcoin? Finance Research Letters 30, 367–370.

Leung, T., Nguyen, H., 2019. Constructing cointegrated cryptocurrency portfolios for statistical arbitrage. Studies in Economics and Finance 36, 581–599.

Liu, Y., Tsyvinski, A., 2021. Risks and Returns of Cryptocurrency. The Review of Financial Studies 34, 2689–2727.

Liu, Y., Tsyvinski, A., Wu, X., 2022. Common Risk Factors in Cryptocurrency. The Journal of Finance 77, 1133–1177.

Makarov, I., Schoar, A., 2020. Trading and arbitrage in cryptocurrency markets. Journal of Financial Economics 135, 293–319.

Miglietti, C., Kubosova, Z., Skulanova, N., 2019. Bitcoin, Litecoin, and the Euro: An annualized volatility analysis. Studies in Economics and Finance 37, 229–242.

Sapkota, N., Grobys, K., 2021. Asset market equilibria in cryptocurrency markets: Evidence from a study of privacy and non-privacy coins. Journal of International Financial Markets, Institutions and Money 74, 101402.

Tiwari, A. K., Kumar, S., Pathak, R., 2019. Modelling the dynamics of Bitcoin and Litecoin: GARCH versus stochastic volatility models. Applied Economics 51, 4073–4082.

Tu, Z., Xue, C., 2019. Effect of bifurcation on the interaction between Bitcoin and Litecoin. Finance Research Letters 31.

Vidal-Tomas, D., Ibanez, A., 2018. Semi-strong efficiency of Bitcoin. Finance Research Letters 27, 259–265.

Wang, Q., Chong, T. T.-L., 2021. Factor pricing of cryptocurrencies. The North American Journal of Economics and Finance 57, 101348.

Wang, Y., Chen, Y., Wu, H., Zhou, L., Deng, S., Wattenhofer, R., 2022. Cyclic Arbitrage in Decentralized Exchanges. In: *Companion Proceedings of the Web Conference 2022*, Association for Computing Machinery, New York, NY, USA 12–19.

Chapter 6

Conclusion

Sustainable and digital finance emerged as important drivers for capital markets by reshaping investor behaviour and regulatory frameworks. Both topics have recently gained increasing attention in academic discussion and practical implementation. This thesis contributes to this field by analysing selected aspects of those topics. The key objective of all four studies is to examine if the new drivers could foster enhanced market efficiency.

Sustainable finance outlines the major trend in corporations to reduce the short-term pursuit of profits, while transforming their strategies to sustainable long-term wealth. This is accompanied by substantial transparency requirements for corporations. The ESG data could lead to a reduction of information asymmetries. However, the major task is to collect, process, and interpret the additional sustainability information. In this case, rating agencies should serve as intermediaries for the capital market. This measurement is still in its infancy, and, consequently, needs valuations and back-testing. This thesis examines critical approaches of the rating providers. It focuses on the adoption of the best-in-industry benchmarking, the weighting of industry groups for the final rating and the use of binary figures and contributes to the academic discussion of constructing better measures. According to the findings of the thesis, the construction of ESG ratings could benefit from comparing corporations on larger benchmarks and weighing approaches that measure the sustainability risk of a corporation's business model. Moreover, the use of binary variables to identify a sustainability dimension is appropriate to reflect the risk-return relationship of a corporation.

The future direction for sustainable finance is not clear. The missing general definition of sustainability, different objectives of national governments and divergence of ESG measures make it difficult to take uniform steps to a beneficial long-term development. Nevertheless, the results of the thesis show that financial market participants incorporate ESG information and price it accordingly. For example, corporations with low sustainable profiles have to pay a risk premium for their inherent long-term risk exposure. Consequently, the capital market takes a steering function with the allocation of cash flows. However, it should be noted that the direction taken by financial market participants does not necessarily have to coincide with socially favourable preferences. At this point it is necessary for regulators to intervene in order to achieve the desired objectives. Academics with their research in this field can provide empirical evidence to support and guide the regulatory initiatives. In summary and despite initial positive developments, sustainable finance's journey is far from over.

Digital finance, with its increasing use of technologies, is disruptive across all sectors of finance. Through an increasing number of automated processes and sophisticated tools, market participants obtain the opportunity to incorporate more information into their decision-making processes. This leads to a higher degree of market efficiency, and consequently, to a decrease in information asymmetries. Moreover, cryptocurrencies emerged as a potential tender or a new type of investment instrument. This thesis combines a few of these topics in a practical test of whether new market segments are efficient. It implements a high-frequency trading strategy on the largest cryptocurrency exchange under real-world conditions. The absence of profitable arbitrage opportunities points to a high degree of market efficiency. This finding of the thesis suggests that the use of technology prevents inefficiencies in this new and unregulated market segment at an early stage.

Similar to other areas of life, digitalisation will constantly reshape the capital market. The speed and efficiency of market participants' actions will increase and cause more informative pricing. In addition, the introduction of cryptocurrencies, and especially the blockchain in the background has the potential to substantially change the financial system. It allows users to transfer currencies through a traceable and decentralised encryption technology. In this process, intermediaries such as banks and financial institutions are no longer needed, which would result in a kind of democratisation of the financial sector. This paradigm shift seems to be beneficial, but also contains certain disadvantages. The high energy consumption, illegal activities, and problems of scaling are still disproportionate to the potential of the blockchain technology. However, research can make a significant contribution in the development of this innovative technology.

In summary, the findings of this thesis demonstrate how sustainable finance and digital finance could contribute to more efficient capital markets. The construction of ESG ratings plays an important role in accurately reflecting the sustainability-related risk exposure of a corporation. From a regulatory perspective, this points to a future direction for improving the uniform measurement of a firm's sustainability profile. In addition, the capital market benefits from the increasing digitalisation through automated processes and new digital assets. In particular, high-frequency traders contribute to a more efficient price discovery process.