BI Norwegian Business School - campus Oslo

GRA 19703

Master Thesis

Thesis Master of Science

Does Materiality Matter?

Links Between ESG Issues and Firm Performance

Navn:	Filip Mihajlovski, Nikita Gukovas

Start:	15.01.2020 09.00
Finish:	01.09.2020 12.00

Does Materiality Matter?
Links Between ESG Issues and Firm Performance

Master thesis

Authors:
Filip Mihajlovski and Nikita Gukovas

Supervisor:
Leon Bogdan Stacescu

Programme:
Master of Science in Business, Major Finance and Master of Science in Finance

BI Norwegian Business School
Oslo,
1st of September, 2020

Abstract

Using material classifications of sustainability categories, we map industry-specific sustainability indicators into materiality-adjusted firm-level ESG scores in a unique geographical sample of the Nordics and the United Kingdom for the 2009-2018 period. Common to the return predictability literature, top and bottom portfolios are sorted based on the residuals from an orthogonalization of changes in material ESG scores with respect to changes in common firm characteristics. Mixed evidence is found on the link between material ESG performance and future stock returns, with only one portfolio exhibiting a significant alpha. Furthermore, panel data fixed effects models are used to test the implications of material ESG performance on year-ahead accounting performance. We distinguish between top performers and improvers on material ESG by using changes and levels in the scores. Results show that the top improvers on material ESG scores exhibit positive future ROA, whereas there is no significant link between being a top performer and future profitability.

Acknowledgments

We would like to thank our supervisor Leon Bogdan Stacescu for his support, guidance and insights during the whole process. Additionally, we would like to thank our families and friends for their unconditional support.

Table of Contents

List of abbreviations

MESG – material ESG scores

MTB – market-to-book ratio

ROE – return on equity

ROA – return on assets

ESG – environmental, social and governance

CSR – corporate social responsibility

SRI – socially responsible investment

FEE – fixed effects estimator

LSDV – least-squares dummy variables estimator

REE – random effects estimator

OLS – ordinary least squares estimator

CLRM – classical linear regression model

CSP – corporate social performance

CFP – corporate financial performance

SASB – Sustainability Accounting Standards Board

KLD – Kinder, Lydenberg and Domini

SMB – small minus big factor

HML – high minus low factor

PR1YR – momentum factor

GRI – Global reporting initiative

List of tables

1. Introduction

For several decades, researchers have been examining the implications of company sustainable behaviour on its future financial performance. Corporate social responsibility (CSR), defined as "the responsibility of enterprises for their impacts on society", has been growing in importance as a part of investor and company decision-making procedures. However, there are differing schools of taught concerning CSR. In line with the concept of shareholder value maximization, Milton Friedman (1970) stated that the "social responsibility of a business is to increase its profits." The other school is led by the seminal work of Freeman (1984) and the notion of stakeholder theory. He argues that a company can achieve long-term success only if it understands and manages the interests of a broad group of stakeholders. The most important notions of this theory, such as the company engaging with multiple stakeholder groups and aiming for maximization of value for all stakeholders, can be seen as largely connected to the concept of CSR[1].

Nowadays, $90tn of assets under management globally come from signatories of the UN Principles for Responsible Investments, one of the most well-known initiatives that promote sustainable investing. The CEO of BlackRock (BlackRock, 2020), stated that there ought to be a "fundamental reshaping of finance" and that companies cannot achieve long-term profits without having a purpose and entertaining the interests of a broad range of stakeholders. In a monumental joint statement by the CEOs of 181 of the largest global companies, the "purpose of the corporation" was redefined to incorporate the interests of all stakeholders (Business Roundtable, 2019). Thus, not only governments, NGOs, customers, and employees require information about the impact that a company has on the environment and people, but also investors, asset managers, and other stakeholder groups.

In this thesis, we examine the implications that performance on sustainability issues has on the future financial performance of a company. Using environmental, social,

[1] In this thesis, we will refer to the concepts of ESG and CSR interchangeably.

and governance (ESG[2]) scores as a measure of sustainable behavior, we focus on testing the relationship between these scores and future stock returns, as well as future accounting performance. The idea is that ESG scores contain information that is financially material to the operations of a given company.

The school of taught proposing a positive link between ESG and financial performance states that high ESG scores are a sign of managerial concern with long-term sustainability, and therefore should signal higher long-term value and returns for shareholders (Gerard, 2019). This is the "doing good by doing well" argument. Krüger (2015) defines the argument more broadly, stating that companies engage in ESG activities in line with the interests of key stakeholders for value-enhancing purposes (Krüger, 2015).

However, a lot of previous research has been plagued by the definition and construction of the ESG ratings (Gerard, 2019). Different data providers gather data and construct ratings in very different manners. Therefore, using publicly available ESG ratings creates issues as, in most cases, a company might have very different scores across data providers. Moreover, there is also the notion of which part of ESG information is financially material to a given company. Issues that are financially material to a company within a given sector or industry might not be relevant to a company in another sector or industry. As materiality classifications of ESG data were not publicly available until recently, most of the previous research disregards this argument.

The Sustainability Accounting Standards Board (SASB) is one of the first sustainability reporting organizations to develop an industry-specific set of standards that provide financially material classifications of ESG metrics. The standards relay valuable information to investors and management on what part of sustainable behaviour is financially material to companies' operations. One of the first papers that are aimed to discern the material information from total ESG scores is the work of Khan et al. (2016). They apply the material classifications provided by SASB to ESG data metrics by KLD. The results show that materiality-adjusted

[2] Environmental, social and governance (ESG) scores are used as a measure of company performance on sustainability topics.

and readily available ESG scores relay different information. Whereas materiality-adjusted ratings are shown to be good return predictors, the information provided by total KLD ESG scores is mostly not informative of future financial performance (Khan et al., 2016).

Based on the SASB classifications, this thesis examines the implications that ESG has on future financial performance, both from the perspective of the investor and the company, and extends the work of Khan et al. (2016). Similarly, it focuses on discerning what part of the information used to create ESG scores is financially material for a given company. We hand-map the 423 ESG metrics that are gathered by Thomson Reuters to general issue categories from the standards provided by SASB. At the industry level, we find that the number of material ESG metrics ranges between 10 and 30, compared to 70 and 170 items used by Thomson Reuters in their ESG score creation. Our sample is based on publicly listed companies in Norway, Sweden, Denmark, Finland, and the United Kingdom in the 2009-2018 period.

Using a time-series methodology, we find mixed evidence on the link between material ESG score performance and future stock returns. Most of the alphas of material ESG sorted portfolios are insignificant, besides a bottom quintile portfolio, which exhibits an annualized underperformance of 6%. We complement this analysis using a panel data approach with fixed effects models. Using either levels or changes in material ESG scores as the independent variable of interest, we do not find a conclusive link between good material ESG performance and future stock returns. The finding is in line with the results of the time-series methodology.

Using a time-series methodology, we do not find any conclusive evidence between TR ESG performance and future stock returns. The estimates of abnormal performance (alphas) are statistically insignificant for all of the top and bottom portfolios. The analysis is as well complemented by a panel data approach with fixed effects models. Similarly, using levels or changes in the Thomson Reuters ESG scores as the independent variable of interest, we do not find any conclusive link between good TR ESG performance and future stock returns.

Further, we examine the relationship between material ESG performance and year-ahead ROA. Using fixed effects models with changes in material ESG scores, we find that company improvement on material ESG scores is followed by an increase in future ROA. On the contrary, there is no link between top material ESG performance and future profitability when levels of material ESG scores are used instead.

Lastly, we examine the relationship between performance on Thomson Reuters scores and year-ahead ROA. Using fixed effects models with changes in TR ESG scores, we do not find any conclusive link between top improvers on TR ESG scores and their future ROA. However, using levels of TR ESG scores, we find that top performers on TR ESG scores have a decrease in their future ROA.

The thesis is organized as follows: in section 2., we review the theory and previous related literature; section 3 presents the research design and testable hypotheses; section 4 is comprised of the empirical approach used; section 5 presents the data collection, sample construction, materiality mapping, and material ESG score creation; section 6 shows the main results and robustness checks; in section 7 we present the conclusion.

2. Theory & literature review

We start the following section by discussing the theory and the basic sustainability concepts in 2.1. The literature review is presented in 2.2. There we focus on a review of studies regarding the link of ESG-financial performance, meta-studies aggregating previous research, as well as the importance of the concept of materiality in terms of ESG score construction.

2.1 Theory & the basics of sustainability

To understand the notion of sustainability, one must define the essential ideas underlying the concept, representing both the investor and company perspective. Those ideas are corporate social responsibility (CSR), environmental, social, governance (ESG), as well as socially responsible investment (SRI). Corporate social responsibility (CSR) is defined as the responsibility of the company for its

impacts on society (European Commission, 2011). It represents the dimension through which companies affect sustainable development. Many authors use the names CSR and ESG interchangeably. We adopt the same notation, and whereas we use the term CSR to present more of the theoretical concepts on sustainable company behavior, we use the term ESG to show the numerical measurement (ESG scores) of sustainable company behavior. Socially responsible investment (SRI) can be defined as an investment strategy that considers both financial returns and social good (Robecco, n.d). Many tools such as the UN Global Compact, the Global Reporting Initiative (GRI), and the Sustainable Development Goals (SDG) have been presented as guidelines of how a company can incorporate responsible behavior in its operations.

Increased engagement by companies in CSR activities is a sign that they try to align not only the shareholders but rather a broader set of stakeholders incentives. Moving from the premise that the role of a company is to maximize value for its shareholders, towards inclusion and value creation for a broader set of stakeholders is in line with Freeman's (1984) theory. It is thereby paramount for the company to successfully acknowledge and manage the interconnected interests of the stakeholders if it is to obtain long-term profits. This argument is the connection point between CSR and stakeholder theory. Moreover, as stated by Kotler et al. (2012, pg.1):

> "Corporate social responsibility can only be successful if it is understood and practiced as an exchange and cooperation between a company and its stakeholders."

The key proposals of the stakeholder theory are opposite to the shareholder theory, etched by Friedman (1974). He famously stated that: "the social responsibility of business is to increase its profits.". This theory gives priority to shareholders and sees CSR engagement as a cost rather than a way of creating long-term value.

The growing interest in socially responsible investment by asset owners has resulted in an increased demand for CSR reporting. Moreover, there is a complementary need for a comparative sustainability reporting framework to make informed investment decisions that include sustainability considerations. Stock indexes consisting of companies with high CSR standards, such as the Domini 400 Social

Index (nowadays MSCI KLD 400 Social Index), already appeared in the 1990s. However, at that time, there were no standard criteria for measuring CSR performance, defining what a relevant CSR category for a given company is, what should be the form of CSR reporting or if there should be unique CSR reporting standards.

The further development of socially responsible investing has been contingent on finding a way to measure CSR performance consistently. The most widely known and used form of measuring sustainable behavior is through the creation of Environmental, Social, and Governance (ESG) scores. The first idea to measure ESG scores of companies to determine their sustainability impact was introduced in the report "Who Cares Wins" (UN Global Compact, 2004). Today, ESG scores are widely known as the most crucial metric of sustainable behavior that is used in investment decisions globally, similarly to what credit ratings represent for assessing the creditworthiness of a company.

2.2 Literature review

The link between ESG and firm value

Eccles et al. (2014) examine the impact of voluntary integration of social and environmental issues on organizational and financial performance. They identify *High* and *Low Sustainability* groups, based on the adoption of social and environmental corporate policies in the 1990s. The authors look into the stock return performance of the two groups in the 1993-2010 period. The findings show that *High Sustainability* firms outperform their counterparts, using both value and equal-weighted portfolios. The authors report that investing $1 in 1993 in the *High Sustainability* value (equal) weighted portfolio would grow to $22.6 ($14.3), compared to $15.4 ($11.7) of the *Low Sustainability* portfolio.

Clark et al. (2015) review different studies that investigate the effects of sustainability practices on various metrics, such as the cost of capital, operational performance, and stock prices. The authors conclude that strong sustainability (ESG) scores lead to better operational performance and less risk. Moreover,

strategies integrating ESG issues outperform comparable non-ESG ones, and that active ownership creates value for both companies and investors.

Flammer (2013) investigates the impact of the environmental dimension on stock prices. Using an event study in the 1980-2009 period, the author looks at the reactions of stock prices of companies related to their environmental performance. The main finding shows that stock prices of companies who experience eco-friendly events increase, on average, 0.84% over two days after the event. On the opposite, if an adverse event happens, the average decline is 0.65% in the same period (Flammer, 2013).

Attig et al. (2013) report that CSR performance in itself reflects a substantial amount of non-financial information that rating agencies might use in assessing the creditworthiness of a company. Moreover, the authors find that CSR investments that are above and beyond what is needed for compliance can lead to lower financing costs, mainly through their effect on a potential increase in credit ratings.

Hartzmark & Sussman (2019) investigate whether investors collectively see sustainability as being a positive, negative, or neutral attribute to the company. Their variable of interest is US mutual fund flows. They use a particular event that represents a shock to the visibility and understanding of sustainability ratings by investors. The event is the Morningstar issuance of sustainability rankings of over 20,000 mutual funds based on a percentile ranking and a 1-5 globe system. The main finding of the paper is that the universe of US mutual fund investors, with over $8tn of assets under management, collectively put a positive value on sustainability. The authors report that the effect is the largest on the extreme values of 5 and 1 globe. Hartzmark & Sussman (2019) find that before the publication, all of the funds exhibit a similar level of inflows. After the publication, the top-rated funds exhibited high inflows of approximately 4% of the fund size, while the bottom rated showed an outflow of around 6% of fund size.

Aggregate evidence and criticism

Although a part of the previous research states that there is a positive link between ESG and financial performance, there is no consensus on this topic among all

researchers. The study of Friede et al. (2015) aggregates evidence from more than 2200 papers dating back to the 1970s, thereby making it one of the most extensive overviews of academic research on the ESG-financial performance link. About 90% of the reviewed papers show a nonnegative relation between ESG and corporate financial performance (CFP), whereby the majority of the studies find a positive relationship that is stable over time.

Gerard (2019) does an extensive review of the effect of CSR[3] and ESG on stock performance, debt value, credit risk, "green bond" labeling, etc. The findings presented on the effects of ESG performance on firm value are the most relevant for our research.

When it comes to governance, Gerard (2019) argues that good corporate governance reduces the agency problem and enhances long-term value. The author points out to the work of Compers et al. (2003), who find that the outperformance of the companies with the strongest corporate governance compared to companies with the worst, was around 8.5% annually in the 1990-1999 period. Bebchuk et al. (2009) show that investing in firms that exhibit low management entrenchment provides a positive risk-adjusted return.

Regarding economic arguments behind a possible positive link between CSR and firm value, Gerard (2019) points out the "good company" and "good management" hypotheses. In the first case, building good relationships with stakeholders by addressing their needs through effective CSR management creates reputational capital that affects corporate valuation through increased profitability. In the second case, implementing appropriate CSR policies represents a sign of managerial quality needed to manage the intersection of multiple stakeholders' claims. Higher managerial quality might translate to higher profitability and lower agency problems (Gerard, 2019).

One of the most important takeaways for our topic lies in Gerard's (2019) criticism regarding the divergence of ESG measures. The main problem is not having unique

[3] Gerard (2019) refers to CSR as the E and S part of ESG, so the overall definition that he proposes is that ESG is CSR and governance.

standards that could be used to transform company reported sustainability data into comparable ESG ratings. Therefore, investors might be misled by ratings that contain ESG information that is irrelevant to a given company. Moreover, they can obtain different ratings for the same company from various ESG data providers. It is also an issue for companies in the sense that the ESG score assigned to them by a data provider might not reflect the ESG issues that are relevant to its operations. Furthermore, if there are no unified standards for sustainability reporting, companies might not be aware of what ESG issues they should focus on.

Divergent rankings are a problem for researchers when testing the link between CSR and firm value, as they face the joint hypothesis problem. Many studies are testing at the same time whether the given score measures CSR performance correctly and that the same score is related/unrelated to financial performance (Gerard, 2019). Moreover, as material classifications were not available previously, a lot of the previous literature uses aggregated scores that might not reflect the true nature of what is materially relevant in terms of ESG scores for a given company.

The importance of materiality

Berg, Koelbel & Rigobon (2020) investigate the divergence of ESG ratings among the five most prominent rating agencies. The authors compare this divergence to that of credit ratings, mentioning that the correlation between credit ratings of S&P and Moody's is about 0.99. In contrast, the average correlation of ESG ratings between the five agencies is around 0.61, ranging from 0.42 to 0.73.

The authors define three sources of divergence: *scope divergence* – different sets of metrics are used to constitute a rating; *weight divergence* – different rating agencies have different views on the relative importance and therefore weighting of the metrics in the creation of a rating; *measuring divergence* – where the same metric could be calculated using different indicators, leading to different results (Berg et al., 2020). They find that 53% of the overall divergence is due to measurement, 44% from the scope, and 3% from weight divergence. Having this divergence in ratings supports the notion that investors might find it challenging to discern relevant ESG data to make informed decisions, or what data provider and score to use in their investment procedures.

The motivation for our research topic lies in the work of Khan et al. (2016). They create materiality-adjusted scores by mapping KLD ESG scores based on industry-specific standards of ESG issues that are deemed material by SASB. As of recently, non-financial reporting regarding ESG issues has been on the rise, and the amount of data that is available to investors is large. The crucial question lies in how much and which part of this data is financially material for a given company. As Eccles and Serafeim (2013) point out, different metrics underlying the environmental, social, and governance pillar are more or less material for a given company depending on its industry or sector.

The work of Khan et al. (2016) tries to provide a way for discerning the material from the immaterial dimension of ESG data. For that reason, the authors set out to use the SASB materiality map, a unified framework of universal sustainability reporting standards. The materiality map is based on 77 industry-specific standards that represent material classifications of ESG issues at the industry level. Materiality-adjusted ESG scores are created, and their implication on future stock and accounting performance is tested. The empirical work of Khan et al. (2016) is distinct compared to previous ESG-financial performance research, on several dimensions:

1) After creating the material ESG scores, the authors orthogonalize the yearly changes in the score with respect to annual changes in the most common firm characteristics, such as size, book-to-market, profitability, leverage, and sector membership. Using this approach, they obtain the residuals from the cross-sectional regressions to use as signals for portfolio construction. The motivation is to obtain a portfolio formation signal that is unencumbered by the effects of the firm characteristics. In other words, by orthogonalizing, the residuals obtained represent changes in material sustainability investments that are unexplained by the changes in firm characteristics. If orthogonalization is not performed, it is likely that the changes in material ESG scores will be partially affected by some of the firm characteristics. Therefore, the orthogonalization process is performed to obtain a portfolio construction signal that can be attributed more safely to changes in material ESG

scores rather than potentially also incorporate the effects of correlated firm characteristics.

2) There is no consensus in terms of using ESG levels compared to changes in previous literature. The motivation behind the work of Khan et al. (2016) is to use changes to work with companies that had the most significant upgrade/downgrade in ESG performance during a year. On the contrary, using levels, the focus of previous research is on working with companies that have achieved a top ESG score at a given point in time, which as a more static approach.

The work of Khan et al. (2016) has some significant findings. Most of the top/bottom portfolios sorted on the material ESG characteristic have significant positive/negative alphas. The alphas of the top portfolios range from 2.88% to 5.16%, and the differential between the top/bottom alphas is substantial and ranges from 2.69% to 7.47% (Khan et al., 2016). Using the immaterial or the total KLD index as a signal for portfolio formation, the results are ambiguous as most of the alphas are statistically insignificant. The authors additionally estimate the same relationship using panel data analysis. They report that the results are very similar, as the top quintile of companies sorted on material ESG outperform by 6.47% annually (Khan et al., 2016). Using an indicator for the total ESG scores, they find a positive estimate of its coefficient, however only marginally significant. When the indicator variable is based on immaterial ESG scores, the estimate is not significant. The authors also estimate the panel data regressions using the future return on sales as a dependent variable. For the top quintile of material ESG, they find that there is positive future growth in ROS. When the indicator variable is based on total or material ESG scores, its estimates are insignificant.

There are several economic interpretations from the presented results:

1) The results show that materiality guidance helps improve the informativeness of ESG scores. The significant alphas are interpreted as being due to investors not being able to apply materiality classifications to ESG scores in the past. The authors argue that the alpha, therefore, was realized through the materiality investments translating into positive future profitability (Khan et al., 2016).

2) The authors show that neither total nor immaterial scores are good predictors of future financial performance. It is only the material component of the scores that affects future performance.

3. Research design and testable hypotheses

The following section serves the purpose of explaining the suitable research design, showing the motivation of why it was used and how it helps us understand the implications of our research questions. In 3.1, we present the research design and connect to the underlying theory. In 3.2, we define our research questions based on the theory and present them in terms of testable hypotheses.

3.1 Research design and underlying theory

Previous research shows that there are differing theories on explaining the potential relationship between ESG and future financial performance. Some authors argue that ESG performance represents a manifestation of the agency problem within a company (Bénabou & Tirole, 2010; Cheng et al., 2013). The argument is that if a company is a good corporate citizen, it is the managers who benefit the most at the expense of shareholders, in terms of the reputation gained being at the helm of such a company.

The line of thought relating a positive link between ESG and financial performance is the "doing well by doing good" argument (Gerard, 2019; Krüger, 2015). The argument states that engagement in ESG activities in line with key stakeholder interests helps create long-term value for the shareholders. The "good management" and "good company" hypotheses are underlying this argument.

The "good management" hypothesis states that having effective CSR practices in place is a signal of good managerial quality (Gerard, 2019). This hypothesis connects to the notion of Freeman's (1984) stakeholder theory. Having good CSR policies in place entails balancing and managing several key stakeholder groups. It is argued that engaging in CSR thus improves relationships with these stakeholder groups which ultimately leads to better financial performance (Waddock & Greaves, 1997).

The “good company” hypothesis states that engagement with key stakeholders and addressing their interests regarding sustainability issues creates reputational capital for the firm, which in turn leads to an enhanced corporate valuation through better profitability and lower impact from adverse events (Gerard, 2019). Lins et al. (2017) support this argument by showing that during the 2008-2009 global financial crisis, high-CSR firms outperformed their peers in terms of stock returns.

However, as Gerard (2019) points out, there are weaknesses in many of the previous studies in the research regarding the link between ESG performance and firm value. Namely, there are a large number of different measures for ESG that differ in specificity and informativeness, as well as the data gathering process and methodology to obtain the final scores. The main issue is that there has not been a specified set of standards that would show how the ESG scores should be computed on a standardized basis or what metrics should constitute the score (Gerard, 2019). Moreover, there has not been a set of standards to include and show what dimension of the ESG score is material to a given company. If there are no set of standards and different ESG scores are created with various issues deemed material, then researchers examining the link between ESG scores and firm value might obtain biased conclusions (Gerard, 2019).

It is the work of Khan et al. (2016) and Grewal et al. (2020) that focuses on providing a solution to the issue mentioned above. The authors use the standards for sustainable reporting as presented by SASB to disseminate between material and immaterial issues and see their effect on future stock and accounting performance.

Our thesis takes its basis on the long-standing previous research on the implications of ESG performance on future financial performance and focuses explicitly on issues of material classifications. Therefore, it builds upon the work of Khan et al. (2016) and adds value by extending their research on two specific dimensions:

1) Khan et al. (2016) focus on a sample from 1991-2014, whereas our research focuses on a more recent 2009-2018 period. The choice of the sample period is important since the 2000s is the period where ESG performance

has started to become an increasingly important issue for investors and other stakeholders.

2) Secondly, whereas Khan et al. (2016) focus on US-listed companies, we extend their research to the unique geographical sample of the United Kingdom, Norway, Sweden, Finland, and Denmark. In that regard, we can test the robustness of SASB's materiality classifications by applying it to a broader and different set of markets.

3.2 Research questions and testable hypotheses

Our thesis intends to answer several research questions, defined as follows:

RQ_1: *Does performance on material sustainability issues affect the future stock performance of publicly listed firms in the United Kingdom and the Nordics?*

RQ_2: *Does performance on all sustainability issues (both material and immaterial) affect the future stock performance of publicly listed firms in the United Kingdom and the Nordics?*

RQ_3: *Does performance on material sustainability issues affect year-ahead accounting performance (ROA) of publicly listed firms in the United Kingdom and the Nordics?*

RQ_4: *Does performance on all sustainability issues (both material and immaterial) affect year-ahead accounting performance (ROA) of publicly listed firms in the United Kingdom and the Nordics?*

As argued by Gerard (2019):

> *"The economic arguments in favor of CSR are similar but less straightforward: high CSR scores suggest managerial concern with long—term sustainability and hence should signal higher long— term shareholder value and returns. This is "the doing well by doing good" argument. However, achieving high CSR scores might require large expenditures, significantly affecting short—term and perhaps long-term firm profitability. Although this might enhance the reputation of the firm*

> *and its managers, the costs of improved CSR immediately affect the bottom line, while the benefits might be quite uncertain and far in the future" (Gerard, 2019, pg.3).*

Our research questions are partially designed to attempt to examine the arguments posed in the quote above. RQ_1 and RQ_2 help us understand whether ESG scores are good future return predictors. By examining the effect of materiality-adjusted ESG scores and total ESG scores separately, we can observe if there is a difference in informativeness between the two ratings. Therefore, RQ_1 and RQ_2 allow us to examine the first part of the argument above made by Gerard (2019), namely whether a company can do well by doing good.

RQ_3 and RQ_4 are designed to help us understand whether ESG scores are good predictors of future accounting performance. They are in line with the second part of the argument presented by Gerard (2019). Engaging in ESG activities likely entails costs to the bottom line of a company, however as Khan et al. (2016) point out, there might be a different effect on the bottom line based on whether a company focuses on total or material components of ESG scores.

It is here that we also make the distinction between the usage of level compared to changes in ESG scores. The distinction in our thesis is the following:

1) Using the time-series methodology, we form portfolios based on annual changes in ESG scores, in line with Khan et al. (2016).
2) Using the panel data methodology, we do an extra step compared to Khan et al. (2016) and present results using both changes and levels in ESG scores.

The motivation behind the adoption of two sets of results is that they will help us examine one additional dimension of the link between ESG performance and future financial performance. *Namely, the difference between using levels or changes helps us distinguish between top/bottom ESG scores achieved in a given point of time and firms with the biggest upgrades/downgrades on ESG scores during a given year.*

Using changes, for the panels where we examine ESG impact on stock returns, we will be able to understand whether the market prices ESG considerations based on a static approach or more dynamic – ESG momentum approach (Giese & Nagy, 2018). The distinction is also important when examining the effect of ESG on ROA. In essence, we will be able to understand whether improving on material ESG scores affects future profitability. When using levels, we can just examine whether a company's high material ESG score achieved at a point affects its future profitability or stock performance. There are nuances in the interpretation between both cases. The main focus of our thesis is to understand the implications of the improvements in ESG scores on future financial performance. Levels are used as a complement to provide additional different interpretations.

In the paragraphs below, we will present our research questions in terms of testable hypotheses, whereas the exact definitions of the statistical tests will be presented in *Section 4.3.*

ESG implications on future stock performance

RQ_1 and RQ_2 help us understand whether material and total ESG scores are good stock return predictors. If material ESG scores are good return predictors relative to total scores, it would mean that material ESG sorted portfolios exhibit significant abnormal performance. That performance cannot be attributed to the common variation between the returns of the portfolios and common risk factors from the contemporary asset pricing models used. It is worth mentioning that irrespective of materiality classifications, portfolios formed on top ESG performers outperformed in the 1990s, while the effect has slowed down during the 2010s and has mostly disappeared since 2010 (Gerard, 2019). RQ_1 and RQ_2 will help us examine whether the result is still present and simultaneously add value for investors who deem using sustainability issues in their investment decisions worldwide.

In that light, concerning RQ_1, we present our first testable hypothesis:

Hypothesis 1

H_0: For companies in the Nordics and the UK, top performers on material ESG issues do not exhibit abnormal future stock performance.

H_1: For companies in the Nordics and the UK, top performers on material ESG issues exhibit abnormal future stock performance.

To see if the total Thomson Reuters scores are good return predictors, concerning RQ_2, we present our second testable hypothesis:

Hypothesis 2

H_0: For companies in the Nordics and the UK, top performers on all ESG (aggregated material and immaterial) issues do not exhibit abnormal future stock performance.

H_1: For companies in the Nordics and the UK, top performers on all ESG (aggregated material and immaterial) issues exhibit abnormal future stock performance.

ESG implications on future accounting performance

RQ_3 and RQ_4 help us understand the implications of good ESG performance on future profitability. Obtaining high ESG scores might require expenditures in the short-term, thereby affecting the bottom line of the company, whereas benefits may manifest far in the future (Gerard, 2019). However, there might be a distinction of costs entailed to have good material or total ESG scores. RQ_3 and RQ_4 will then add value at the company level, by discerning the future costs or benefits that a company incurs on its bottom line by either focusing on material or aggregated ESG issues.

In line with RQ_3, we thereby present our third testable hypothesis:

Hypothesis 3

H_0: For companies in the Nordics and the UK, a statistically significant relationship between top performers on material ESG issues and year-ahead return on assets (accounting performance) is not present.

H_1: For companies in the Nordics and the UK, a statistically significant relationship between top performers on material ESG issues and year-ahead return on assets (accounting performance) is present.

Focusing on total ESG scores entails convergence towards a large number of ESG indicators[4], some of which have no material importance to a company's operations. Materially mapping the TR indicators, we find that our material ESG scores contain only 10-20% of the indicators used by TR to form their scores, on average. This is in line with the findings of Khan et al. (2016). A company which focuses on total ESG issues is likely to entail higher costs compared to a company focusing on material ESG issues. This is an additional source of motivation for specifying our research questions in terms of both TR ESG and material ESG scores.

Concerning RQ_4, we present our last testable hypothesis:

Hypothesis 4

H_0: For companies in the Nordics and the UK, a statistically significant relationship between top performers on all ESG (aggregated material and immaterial) and year-ahead return on assets (accounting performance) is not present.

H_1: For companies in the Nordics and the UK, a statistically significant relationship between top performers on all ESG (aggregated material and immaterial) and year-ahead return on assets (accounting performance) is present.

From a company perspective, the research questions could help understand whether management should focus exclusively on sustainability issues that are deemed material to the operations of the company. Examining implications on both stock and accounting performance, we can potentially provide answers to the trade-off between the costs of implementing ESG practices and the benefits obtained. Hilman & Keim (2001) state that building better relationships with key stakeholders such as employees, suppliers, customers, and the overall community helps firms build intangible value that can increase shareholder wealth. In line with this study, we hypothesize that the reputational gain from good ESG practices could potentially translate into better future stock performance.

From an investor's perspective, our research will show whether investors can make well-informed investment decisions by focusing on material rather than total ESG

[4] Thomson Reuters reports using 70-170 indicators depending on industry.

scores. Knowing what dimension of sustainability investments is financially material to a company's operations would serve a purpose for a wide variety of investors.

4. Empirical methodology

In 4.1 we define and discuss the dependent, control, and independent variables of interest used. In 4.2, we present the time-series methodology and orthogonalization. We discuss the panel data methodology in 4.3. In 4.4, we focus on the economic meaning and interpretation of potential regression coefficient estimates.

4.1 Dependent, independent and control variables

Variables used in orthogonalization and time-series methodology

Orthogonalization and time-series are exclusively used in the analysis of RQ_1 and RQ_2. In terms of the orthogonalization, annual changes in material/Thomson Reuters ESG[5] scores are used as the dependent variable. They are orthogonalized with respect to annual changes in company size, market to book ratio, leverage, and profitability. Top and bottom portfolios are created based on the residuals from this process and time-series regressions are further used. There, the dependent variables are the value/equal-weighted returns of the material/Thomson Reuters sorted portfolios. The independent variables are the returns of common risk factors, such as the proxy for market portfolio, size, value, and momentum factors.

Variables used in panel data methods

Panel data methods are used to examine RQ_3 and RQ_4 explicitly, as well as to serve as an additional methodology for RQ_1 and RQ_2. When we examine the impact of ESG on future profitability[6], the dependent variable in panel regressions is year-ahead ROA. Month-ahead stock returns are used for estimating the impact of ESG on future stock performance. The control variables in all four research questions are the same and include past 52-week stock returns, company market capitalization, market to book ratio, leverage, share turnover, return on equity, and

[5] Throughout this thesis, Thomson Reuters ESG scores are proxy for all (aggregated) ESG issues.
[6] In RQ_3 and RQ_4.

capital expenditures/sales. All of the control variables represent firm characteristics that are most likely to affect the relationship between ESG and stock or accounting performance. The same control variables are used in all of the panels presented.

The difference comes from the independent variable of interest used in each panel. In general, the independent variable of interest is an indicator representing the top quintile of firms on TR/material ESG score levels/changes each year. When using changes, the indicator takes a value of 1 for the top quintile of companies that had the most significant improvement in material/TR ESG scores from the end of year t-2 to the end of year t-1, and 0 otherwise. When using levels, the indicator takes a value of 1 for the top quintile of companies that achieved the highest material/TR ESG score at the end of year t-1, and 0 otherwise. Through time, the top quintiles change as firms exit and enter, allowing the indicators to be time-variant, and therefore available to be used in panel data estimation.

4.2 Time-series regression methodology

Orthogonalization and time-series methodology

In order to test the implications of company performance on material and total ESG on future stock returns, we follow the approach used by Khan et al. (2016). Material ESG scores are constructed with the idea of capturing the part of the information contained in the total ESG is financially material to a company. Furthermore, we attempt to isolate the effect that changes in common firm characteristics might have on the changes in the material ESG score. For that purpose, we orthogonalize changes in the material ESG scores with respect to changes in size, market-to-book ratio, return on assets (ROA), leverage, and sector membership. This procedure is operationalized by estimating the following model cross-sectionally each year:

$$\Delta MaterialESG_{it} = \beta_1 + \beta_2 \Delta lnSize_{it} + \beta_3 \Delta MTB_{it} + \beta_4 \Delta ROAit + \beta_5 \Delta Leverage_{it} + \beta_6 DSector^{7} + e_{i,t} \quad (1)$$

[7] We have also run the same models including an industry dummy. The model exhibited a very similar explanatory power, so regressions with sector dummies were used as the portfolio signal.

Our goal is to focus on the residuals as they would reflect as closely as possible the underlying changes in material sustainability investments, rather than the changes in our scores also being affected by changes in size, MTB, ROA or leverage. Using this approach, we work with companies who had the biggest upgrades/downgrades in material/Thomson Reuters ESG scores during a given year.

ESG data by Thomson Reuters is published following the fiscal year-end of a company. In our sample, the latest fiscal year-ends are in December, although there is a sizable amount of companies with fiscal year-ends in earlier quarters of the calendar year. To avoid look-ahead bias and make sure that the strategy would have been feasible to implement for investors, we create portfolios using a 3-month gap. This gap helps us mitigate concerns of when both the financial and ESG data will have been made public. We form portfolios at the end of March each year ***t*** and hold them from April in year ***t*** to March in year ***t+1***. For example, we use residuals obtained from changes in material ESG scores from FY2009 to FY2010 to form portfolios at the end of March 2011 that are held until the end of March 2012, before being rebalanced.

Equal and value-weighed returns for top/bottom quintile and decile portfolios are calculated during the sample period. Our approach is common in the return-predictability literature, where portfolios are sorted on a given characteristic and their future return performance is tested using established asset pricing models (Khan et al. 2016). If the intercept (alpha) is significant, then portfolios sorted on the material/Thomson Reuters ESG score characteristic exhibit abnormal performance. This performance is the part of the portfolio returns that cannot be attributed to common variation with the returns of the common risk factors hypothesized by the asset pricing modes.

The time-series regression approach is a well-known methodology used by Black, Jensen, and Scholes (1972) and Fama & French (Fama & French, 1993). The asset pricing models used to examine the implications of material/TR ESG performance on future stock returns are the following:

- Capital Asset Pricing Model – CAPM (Sharpe 1964; Lintner 1965)

$E(R_i) = R_f + [E(R_m) - R_f]\beta_{i,mkt}$,

where $E(R_i)$ is the expected return of security *i,* R_f is the risk-free rate and $[E(R_m)-R_f]$ is the expected excess return on a proxy for the market portfolio.

- Fama-French 3-factor model (Fama & French, 1993)

  $E(R_i) = R_f + \beta_{i,mkt}[E(R_m) - R_f] + \beta_{i,smb}\, SMB + \beta_{i,hml}HML$

  where $[E(R_m)-R_f]$ is the expected excess return on a proxy for the market portfolio, SMB is a mimicking long/short portfolio based on size and HML is a mimicking long/short portfolio based on book-to-market value of equity.

- Fama-French-Carhart 4-factor model (Carhart, 1997)

  $E(R_i) = R_f + \beta_{i,mkt}[E(R_m) - R_f] + \beta_{i,smb}\, SMB + \beta_{i,hml}HML + \beta_{i,pr1yr}PR1YR$

  where PR1YR represents the momentum factor, a mimicking long/short portfolio ranked on prior 12 months momentum (winner vs loser stocks).

The goal is to evaluate if the intercept (alpha) of the time-series regression is statistically significant. Linear factor models usually use excess returns of the proxy for a market portfolio and long/short zero-investment mimicking portfolios, which implies that the time series intercept should be statistically insignificant (Cochrane, 2009). For the research questions where we use this methodology, namely RQ_1 and RQ_2, the testable assets are excess returns of the materiality or total ESG sorted portfolio returns over the risk-free rate. The time-series models that we estimate are:

RQ_1

$$R_{materialptf}[8] - R_f = \alpha + \beta_{materialptf,\,mkt}[R_{mkt}(t) - R_f(t)] + h_{materialptf,\,hml}HML(t) + s_{materialptf,\,smb}SMB(t) + e_t \quad (2)$$

RQ_2

$$R_{totalESGptf} - R_f = \alpha + \beta_{totalESGptf,\,mkt}[R_{mkt}(t) - R_f(t)] + h_{totalESGptf,\,hml}HML(t) + s_{totalESGptf,\,smb}SMB(t) + e(t) \quad (3)$$

[8] The asset pricing models used in the thesis are the CAPM, the Fama-French 3-factor model and the Fama-French-Carhart 4-factor model. However, we only present here the equation with respect to the FF3.

4.3 Panel-data regression methodology

The second methodology that we employ is panel data regressions. There are multiple arguments for introducing panel data regressions: 1) the structure of our data, where we have observations over time for a cross-section of firms, can be accounted for with panel data estimators; 2) it serves as a complementary methodology towards time-series regressions in terms of future stock performance, and it is of common use in previous literature that examines relationships between accounting variables and ESG performance (Marti et al., 2015; Lo & Sheu, 2007; Lee et al., 2015); 3) It allows us to include additional measurable firm characteristics as control variables; 4) it enables us to control for unobservable heterogeneities between firms.

One of the important advantages of panel data structures is the acknowledgment that entities are distinct from each other. In other words, companies may have their unobservable uniqueness (heterogeneity) that affects the dependent variable alongside the regressors used in a model. In econometric terms, panel data would have an equation as follows (Gujarati, 2009):

$$Y_{it} = \beta_1 + \beta_2 X_{it} + \beta_3 X_{it} + \alpha_i + u_{it},$$

where α_i represents the individual effects (heterogeneity) of entities.

In the case of our research topic, some possible sources of this heterogeneity are: quality and diversity of the management and workforce; knowledge, commitment, and engagement towards promoting sustainability; internal firm values and codes of conduct; type of leadership and prior experience with implementing policies of sustainability; management philosophy towards improving ESG performance, etc. It can be assumed that these, and other unobservable variables are distinct to firms but do not change over time.

Pooled OLS estimator

The most well-known estimators in the panel data context are the pooled OLS, fixed effects, and random effects estimators. As the name suggests, with the pooled

OLS, we would pool all our cross-sectional time-series observations and estimate a regular OLS regression. However, there are potential disadvantages to this approach. When using this estimator, we disregard the panel structure of the data and ignore the notion that each entity might behave differently (Gujarati, 2009).

If we consider our sample, by using a pooled OLS estimator, we would not account for company-specific individual effects, and they would be subsumed by the error term. If that is true, and these individual effects are correlated with the regressors that we use in the models, then one of the main assumptions of the classical linear regression model will be violated. The assumption underlying the CLRM is that the correlation between regressors and the disturbance term should be $Cov(X_k,u) = 0$. In other words, we would incorporate endogeneity (through the omitted variables), obtaining estimates that are potentially biased and inconsistent (Gujarati, 2009).

The endogeneity problem in previous literature

Endogeneity is an issue that arises due to violation of one of the key assumptions of the CLRM, that being the violation of the conditional mean independence, or $E(u_i|X_i) = 0$ (Verbič, 2018). There are different potential sources of endogeneity: omitted explanatory variables that are correlated with the regressors included in the model; simultaneity between the dependent and independent variables; measurement errors regarding the dependent and independent variables (Verbič, 2018). For our research topic, if there are omitted variables, such as management quality or board reputation, that are at the same time determinants of financial performance and correlated with ESG scores, we would have introduced endogeneity. If a regressor such as ESG scores is correlated with the error term, then ESG scores will increase when the error term increases and vice versa (Gujarati, 2009). We then cannot estimate the actual effect that an ESG score has on financial performance. The outcome is that the presence of endogeneity leads to potentially biased and inconsistent estimates when using regular estimators such as pooled OLS.

Endogeneity has been documented as an issue that was not accounted in much of the previous ESG-financial performance literature. As Gerard (2019) reports,

based on a meta-analysis of 214 research papers examining the relationship between CSR and financial performance (Margolis et al. 2009), most of the previous studies do not deal appropriately with the endogeneity problem. In this context, they do not take into account that decisions to engage in CSR activities are likely to be correlated with unobservable firm characteristics that also affect the firm's financial performance (Gerard, 2019). The same discussion is presented by Garcia-Castro et al. (2010), where they also report that at the time of their study, only a few other CSR-financial performance studies have considered endogeneity problems.

We argue that due to our data being firm-level, unobservable heterogeneity between firms is present and likely correlated with the ESG score as a regressor. If this is true, we would be facing an obvious problem if we were to use a pooled OLS or the random effects model[9], both of which have an assumption that regressors used are not correlated with the error term. Moreover, we present the arguments that Khan et al. (2016) report regarding alleviating endogeneity concerns using both the time-series methodology as well as fixed effects models.

Khan et al. (2016) state the following:

> "*Collectively the tests mitigate concerns about endogeneity by using empirical approaches from the forefront of the return predictability literature: (i) The returns tests are predictive rather than contemporaneous regressions; (ii) The return prediction signal is the change in the materiality score orthogonalized with respect to changes in a number of firm characteristics; (iii) The portfolio tests control for conventional risk factors, allowing attribution of the alpha to material investments. This inferential approach is standard in the asset pricing literature; (iv) The portfolio tests are supplemented by firm-level return prediction regressions saturated with controls for known return predictors, a host of firm characteristics, and time and firm fixed effects*"
> (Khan et al., 2016, pg.4).

[9] The difference comes in that random effects models account for individual effects, while pooled OLS do not. However, contrary to fixed effects models, the random effects models assume that the individual effects are a part of the error term and are not correlated with any of the regressors used.

These arguments are also valid for our study, as the methodology used in this thesis closely follows Khan et al. (2016). It is important to distinguish the arguments above in terms of *unobservable* or *observable* omitted variables as a source for the omitted variable bias. The first part of argument *(iv)* relates to alleviating concerns associated with the omission of *observable* variables. In that sense, the mitigation comes from specifying as control variables the most used return predictors and firm characteristics. The usage of time and firm fixed effects relates to the omitted *unobservable* variables. If the unobservable variables are time-invariant, then the usage of fixed effects models mitigates potential endogeneity issues from this source. To see whether our arguments are supported empirically, we turn to present the fixed and random effects estimators, and statistical tests that allow us to determine the most suitable estimator for our research.

Fixed effects estimator and random effects estimator

Seemingly, the fixed effects estimator has properties that are desirable and would potentially fit the structure of our data. Namely, the estimator deals with company-specific effects by introducing them as time-invariant (fixed) effects - α_i. In the FEE, no general constant term is present, as the individual effects α_i represent entity (company)-specific intercepts. These intercepts differ across entities (companies) but are constant over time, so a generalized equation of the FEE can be written as follows:

$$Y_{it} = \beta_{1i} + \beta_2 X_{2it} + \beta_3 X_{3it} + \beta_{4it} + u_{it},$$

where the intercept term contains only a subscript ***i*** referring to the individual effects assumed to be time-invariant (Gujarati, 2009).

The individual effects are called fixed effects and contain all of the effects that are specific to an entity (company) but are constant over time. Moreover, the FEE assumes that the time-invariant effects are correlated with included regressors (Verbič, 2018). The FEE can be operationalized in two mathematically equivalent ways (Gujarati, 2009). First, we could use the *LSDV* technique in which we would need to implement company dummies. The other approach is the *within estimator*. This estimator demeans the values of each of the dependent and independent variables for each entity. In other words, for each company, the values of the

dependent and independent variables are expressed as deviations from their respective mean values (Gujarati, 2009). Compared with the pooled OLS, the within estimator controls for the company-specific effects by removing them, using the within transformation. As this procedure entails time-demeaning of values of variables, all of the time-invariant variables will drop out of estimation, as for each time period their value is going to be the same.

However, this also means that we cannot specify variables that are time-invariant to be regressors, as the FEE estimation will also wipe them out. For example, a company is likely to stay in the same industry throughout the sample period. Therefore, using a within estimator, we cannot explicitly include an industry membership dummy as an explanatory variable. It is a trade-off, as we would have to accept the fact that we cannot use time-invariant regressors to be able to control for the unobservable heterogeneities (Gujarati, 2009).

The random effects estimator, on the other hand, introduces individual effects as a part of the disturbance term, and the individual effects are therefore assumed to be random and uncorrelated with the regressors (Verbič, 2018). The general equation of the model can be written as follows:

$$Y_{it} = \beta_1 + \beta_2 X_{2it} + \beta_3 X_{3it} + \beta_4 X_{4it} + w_{it},$$

where $\boldsymbol{w_{it}}$ is a compound disturbance term containing a random variable $\boldsymbol{v_i}$ representing the individual effects $\boldsymbol{\alpha_i}$, and the independently and identically distributed error term $\boldsymbol{u_i}$ (Verbič, 2018). The assumption underlying the model is that the compound disturbance term is not correlated with any of the regressors. However, if the individual effects that are part of the error term are correlated with the regressors, coefficient estimates would be inconsistent (Gujarati, 2019). It is thereby paramount to perform a multitude of statistical tests to decide what model fits best our data empirically.

Restricted F-test for individual effects

Firstly, we use the restricted F-test (Verbič, 2018) to test whether all the company-specific intercepts are jointly insignificant. The null of the F-test is that all the differential (company) intercepts are jointly equal to 0, meaning that companies do

not differ among each other in the cross-section (Gujarati, 2009). If the null is rejected, we conclude from the alternative that the companies are indeed heterogeneous, and using pooled OLS would be inappropriate.

Hausman test for model specification

Further, a Hausman test is performed to assess the suitability of the FEE and the REE. The null hypothesis under the test states that the FEE and REE do not substantially differ (Gujarati, 2009). If we fail to reject the null, the REE is the preferred model. If we reject the null, FEE is the model to be used. The notion is that if the unobservable individual effects are correlated with the regressors, the FEE is the most appropriate estimator (Baltagi et al., 2012).

F-test for joint significance of time fixed effects

In line with previous research (Khan et al., 2016), we also argue that we should control for factors that have the same impact on the cross-section of firms but change over time. These are called time fixed effects, and are proxies for macroeconomic variables that are assumed to affect the cross-section of firms in the same manner but are time-varying. For example, change in technological progress, changes in government regulation, tax changes, etc. (Gujarati, 2009). To test whether time fixed effects should be included in our models, we use an F-test. Under the null, the test examines whether all of the time dummies are jointly equal to zero (Verbič, 2018). If the null is rejected, we conclude that time fixed effects should be added to a model.

Modified Wald's test for groupwise heteroskedasticity

In a panel data context, we test for the presence of groupwise heteroskedasticity by using the modified Wald test. Under the null, the test examines whether the variances of the error terms across entities (companies) are equal. One source of heteroskedasticity can be group membership (industry or sector membership). If observations in the sample are pre-defined as members of groups, the variance of the error terms across groups will likely be heteroskedastic (Baum, 2006). As this is precisely the case with our sample, where firms are pre-defined as members of a given industry, we suspect the presence of groupwise heteroskedasticity.

Wooldridge test for AR (1) serial correlation

The other source of potential issues is the presence of autocorrelation of the error term. Usually, in panel data, the first-order autocorrelation is tested (Verbič, 2018). We use Wooldridge's AR (1) serial correlation test. Under the null, the test would indicate that first-order autocorrelation is not present.

Clustered standard errors as potential remedies

Remedial measures are needed when there is a presence of heteroskedasticity and serial correlation, to have correct statistical inference. One possibility is to estimate clustered standard errors that are heteroskedasticity and autocorrelation robust. The assumption is that there is a correlation of the error term *within a cluster*, but independence of the error terms *across clusters* (Cameron & Miller, 2015). If remedial measures are not undertaken, the OLS standard errors would potentially be biased, affecting the t-statistics and p-values and confounding statistical inference (Petersen, 2009). Therefore, we follow previous research (Khan et al., 2016) and cluster at the firm level.

Pesaran test for cross-sectional dependence

Another form of correlation between residuals is cross-sectional or spatial correlation (Gujarati, 2009). Cross-sectional dependence could be an issue for our sample as we have companies nested within industries, meaning that there is a likelihood of correlation between the error terms of companies. Therefore, we use a Pesaran CD test. Under the test, the null states that the residuals between groups (companies) are not correlated (Pesaran, 2015).

An interesting argument for the presence of cross-sectional dependence is brought by De Hoyos & Serafidis (2006). They state that it is the ever-growing financial and economic cross-country integration that creates interdependencies between companies in the cross-section. If we assume that the cross-sectional dependence is due to unobservable common factors that are uncorrelated with regressors, fixed effects and random effects models can be used, however estimated standard errors will be biased (De Hoyos & Serafidis, 2006). In that light, the remedial measure would be to obtain the estimates from a FEE/REE and compute Driscoll-Kraay standard errors (Driscoll & Kraay, 1998). The Driscoll-Kraay standard errors are

autocorrelation, heteroskedasticity, and spatial correlation robust. That being said, for each of our models, we perform all of the aforementioned statistical tests. As it will be shown later, we present our models using firm-level clustered as well as Driscoll-Kraay standard errors.

Gujarati (2009) argues that in panel data regressions, one needs to be very careful about the assumptions underlying the error term. Because panel entails data through two dimensions, time ***t*** and cross-section ***i,*** the classical linear model assumptions regarding the error term have to be modified. He points out to 3 permutations of the error term, namely: assuming that the error variance differs across entities (it is heteroskedastic); assuming that for each entity there is a presence of AR (1) autocorrelation and; assuming that for a given time period, error terms of entities are correlated between themselves (Gujarati, 2009). We perform the aforementioned diagnostic tests to examine each of these three issues for each of our specified models.

Defining specific forms of econometric models

The following section outlines the final specifications of the models used after each of the tests mentioned above is implemented. In that sense, the final models are fixed effects models with both firm and time fixed effects being employed. The specification of the models is similar to Khan et al. (2016):

Hypothesis 1

$$1\text{-month return}_{i,t+1} = \beta_0 + D_1 HighMaterialESG_{i,t} + \beta_2 Past52weekreturns_{i,t} + \beta_3 lnSize_{i,t} + \beta_4 MTB_{i,t} + \beta_5 Leverage_{i,t} + \beta_6 Shareturnover_{i,t} + \beta_7 ROE_{i,t} + \beta_8 CAPEX_{i,t} + FirmFixedEffects_i + TimeFixedEffects_t + u_{it} \quad (4)$$

where ***i*** (company identifier) = 1, 2, 3…435 and ***t*** (time-month identifier) = 1, 2, 3…108.

For one version of this panel, *HighMateriaESG* is an indicator for top material ESG performers each year, *based on the levels of the ESG score*. For the other

version of the same panel, *HighMaterialESG* is an indicator of top material ESG improvers, *based on yearly changes in material ESG scores*[10].

Hypothesis 2

$$\textit{1-month return}_{i,t+1} = \beta_0 + D_1 HighTRESG_{i,t} + \beta_2 Past52weekreturns_{i,t} + \beta_3 lnSize_{i,t} + \beta_4 MTB_{i,t} + \beta_5 Leverage_{i,t} + \beta_6 Shareturnover_{i,t} + \beta_7 ROE_{i,t} + \beta_8 CAPEX_{i,t} + FirmFixedEffects_i + TimeFixedEffects_t + u_{it} \quad (5)$$

For one version of this panel, *HighTRESG* is an indicator for top Thomson Reuters ESG performers each year, *based on the levels of the TR ESG score.* For the other version of the same panel, HighTRESG is an indicator for top TR ESG improvers, based on *yearly changes in the TR ESG scores.*

Hypothesis 3

$$ROA_{i,t+12} = \beta_0 + D_1 HighMaterialESG_{i,t} + \beta_2 Past52weekreturns_{i,t} + \beta_3 lnSize_{i,t} + \beta_4 MTB_{i,t} + \beta_5 Leverage_{i,t} + \beta_6 Shareturnover_{i,t} + \beta_7 ROE_{i,t} + \beta_8 CAPEX_{i,t} + FirmFixedEffects_i + TimeFixedEffects_t + u_{it} \quad (6)$$

For one version of this panel, *HighMaterialESG* is an indicator for top material ESG performers each year, *based on the levels of the ESG score.* For the other version of the same panel, *HighMaterialESG* is an indicator for top material ESG improvers, *based on yearly changes in material ESG scores.*

Hypothesis 4

$$ROA_{i,t+12} = \beta_0 + D_1 HighTRESG_{i,t} + \beta_2 Past52weekreturns_{i,t} + \beta_3 lnSize_{i,t} + \beta_4 MTB_{i,t} + \beta_5 Leverage_{i,t} + \beta_6 Shareturnover_{i,t} + \beta_7 ROE_{i,t} + \beta_8 CAPEX_{i,t} + FirmFixedEffects_i + TimeFixedEffects_t + u_{it} \quad (7)$$

For one version of this panel, *HighTRESG* is an indicator for top Thomson Reuters ESG performers each year, *based on the levels of the TR ESG score.* For the other version of the same panel, *HighTRESG* is an indicator for top TR ESG improvers, *based on yearly changes in the TR ESG scores.*

[10] When yearly changes are used, the indicator variable is based on the sort of companies obtained from the orthogonalization process.

To test our research questions, we would use the following statistical tests. Namely, for time-series regressions, it would entail having a t-test of the following form:

$$H_0: \alpha_i = 0$$

$$H_1: \alpha_i \neq 0$$

where we would test the statistical significance of the intercept (alpha) from the time-series regressions of the returns on our portfolios on the specified asset pricing models. We would therefore report tests including a CAPM, Fama-French 3 factor, and Fama-French 4-factor alphas.

Similarly, using the panel data regression approach, we would be able to test our hypotheses by using a t-test, but this time on the estimated regression coefficient for our variable of interest. The variables of interest are the indicator variables of the top performers on ESG score levels/changes.
Therefore, testing the hypotheses would entail:

$$H_0: MESG\ Dummy_1 = 0$$

$$H_1: MESG\ Dummy_1 \neq 0$$

and

$$H_0: TRESG\ Dummy_1 = 0$$

$$H_1: TRESG\ Dummy_1 \neq 0$$

4.4 Meaning of statistical significance and signs of estimates

If we are successful in creating material ESG scores that relay the information deemed material by SASB, then we would expect to see potentially significant positive/negative alphas for the top/bottom portfolios sorted on material ESG scores. If true, the explanation could be that as investors did not have materiality classifications available, then when ESG scores were published, they could not react immediately. Therefore significant alphas were present in their sample period (Khan et al. 2016). A counter-argument arises because, for at least some part of our sample period, investors had SASB classifications publicly available.

In their meta-analysis of 215 of CSR-financial performance studies in the 1973-2007 period, Margolis et al. (2009) find a small positive effect of CSR on financial

performance, which is declining over time. Moreover, as Gerard (2019) notes, it seems like that firms who performed well on ESG in the 1990s outperformed their peers. In contrast, that effect halved in the 2000s and has been almost completely gone since the 2007 financial crisis, meaning that markets seem to largely price in ESG performance today.

When it comes to examining the implications of performance on total or material ESG on future ROA, we follow the arguments by Gerard (2019) and Garcia-Castro et al. (2010). Incorporating ESG issues likely entails costs for the company on the short-term that are to be incurred on its bottom line. If a company performs well on total ESG scores, it means that it needs to incorporate a large number of ESG indicators in its operations, and we would expect to see a negative sign on the relationship with its one-year ahead ROA. If it performs well on materiality-adjusted ESG scores and thereby focuses only on a small portion of indicators that are material to its operations, we would not expect to see a negative year-ahead ROA. If the relationships tested by the relevant hypotheses of RQ_3 and RQ_4 are significant and with signs that we expect, that would confirm the argument that firms should focus on material ESG items to in order not to decrease their future profitability.

5. Data & materiality mapping

5.1 Materiality data

Our data source for financially-material sustainability information is the Sustainability Accounting Standards Board (SASB). It is an independent non-profit organization whose mission is to develop unified sustainability accounting standards. SASB standards focus on financially material issues in sustainability topics that matter most to investors (SASB - a, n.d). By the end of 2018, SASB had developed and published a set of 77 industry standards, which are the ones used for our study. SASB's standard-setting process is based on a 6 stage project model (from identification and assessment to development, proposal, update, and monitoring). Although the SASB research staff is responsible for this process, the final decision is made by the Standards Board, which consists of industry professionals with an assignment of a minimum of 3 board members to each sector.

Moreover, individuals from corporations, financial institutions, and other stakeholders with industry expertise constitute the SASB Standards Advisory Group, whose role is to provide feedback on standards development and implementation, as well as to raise emerging issues that should be considered by SASB.

In the presence of several sustainability reporting frameworks, we argue that SASB is the one that helps us examine our research questions in the manner that they were set. Besides SASB, the other most relevant framework is the Global Reporting Initiative (GRI). This organization similarly develops sustainability reporting standards with a focus on materiality[11]. However, there is a specific difference between GRI and SASB that can be seen in the following quote from a mutual article by the boards of both organizations (Mohin & Rogers, 2017):

> "*GRI and SASB are intended to meet the unique needs of different audiences. The GRI standards are designed to provide information to a wide variety of stakeholders and consequently, include a very broad array of topics. SASB's are designed to provide information to investors and consequently, focus on the subset of sustainability issues that are financially material.*"

Moreover, as SASB states (SASB – b, n.d):

> *"What distinguishes SASB standards is a focus on financially material information covering a range of industry-specific sustainability areas, including environmental and social topics and the governance of those topics. SASB focuses on financially material issues because our mission is to help companies around the world report on the sustainability topics that matter most to investors."*

Industry professionals, as well as researchers, cite SASB as their choice of a suitable sustainability reporting framework. For example, BlackRock[12] CEO Larry Fink stated: *"While no framework is perfect, BlackRock believes that the Sustainability Accounting Standards Board (SASB) provides a clear set of*

[11] 75% of the largest 250 companies in the world use GRI sustainability reporting standards according to KPMG Survey of Corporate Responsibility Reporting 2017.
[12] Michael Bloomberg has also been the Chair Emeritus of SASB in the 2014-2018 period.

standards for reporting sustainability information across a wide range of issues, from labour practices to data privacy to business ethics." (BlackRock, 2020)

As our research questions are focused on discerning the material component of ESG scores and seeing its effect on firm financial performance, we find SASB's standards an obvious choice.

5.2 Sustainability performance data

The ESG scores that we use the Thomson Reuters ESG scores available in Eikon. Thomson Reuters calculates 423 company level ESG metrics, of which 186 are used[13] in their scoring process. These metrics are then grouped into ten categories, such as resource use, emissions, workforce, human rights, management, shareholders, etc. Those ten categories then comprise the environmental, social, and governance pillars, after which the final company-level ESG scores are created. Furthermore, Thomson Reuters also reports an ESGC score, which takes the regular ESG score and overlays it with 23 controversies categories that might be reported in the media for a specific company (Refinitiv, n.d). The basis for the creation of our material ESG scores are the 423 ESG metrics that are also the basis for the Thomson Reuters score creation. The sample period for our research is ten years, starting with 2009 and ending in 2018. The geographical focus of this study is on the United Kingdom, Norway, Sweden, Finland, and Denmark.

5.3 Financial data

The source for the financial data is the Thomson Reuters Datastream database. As a part of the orthogonalization process and the subsequent use of time-series methodology, yearly data points for company market capitalization, market-to-book, leverage, and return on assets (ROA), as well as monthly stock returns are used. For the time-series methodology, tests are done using three different asset pricing models. For the CAPM, the relevant data used is the excess returns of the market portfolio over the 1-month risk-free rate. For the Fama-French 3-factor model, additional data on the long/short small minus big and high minus low

[13] Thomson Reuters reports that depending on the industry, 70 to 170 are relevant.

factors was used. For the Fama-French-Carhart 4-factor model, the three factors above are complemented by the momentum factor. Data for all asset pricing models were downloaded from the online library of Professor Kenneth French (Data library, n.d). Table 1 provides detailed specifications of the variables used in the time-series and orthogonalization process.

Table 1: Specification on dependent, independent and control variables used for orthogonalization and time-series regressions

Name	Type of variable	Definition
Size	Control	Natural logarithm of calendar-year end market capitalization in USD
Market-to-book	Control	Calendar year-end market value to book value of common equity
Leverage	Control	(Short Term Debt & Current Portion of Long-Term Debt + Long Term Debt) / Total Assets * 100
Return-on assets	Control	(Net Income + ((Interest Expense on Debt-Interest Capitalized) *(1-Tax Rate))) / Average of Last Year's and Current Year's Total Assets * 100
Sector membership	Control	Dummy variable indicating SICS Sector membership
Monthly returns	Dependent	1-month total return incorporating the price changes and any dividends for the previous month
Yearly changes in material ESG scores	Dependent	Hand-mapped material ESG scores based on SASB-Thomson Reuters ESG indicator level matching
Yearly changes in Thomson Reuters ESG scores	Dependent	Thomson Reuters ESG score provided by Refinitiv

For panel data regressions as our second methodology, all of the data used is based on a monthly frequency. Stock returns, company market capitalization, market-to-book ratio, leverage, and ROA have the same definitions as above. Additionally, we use additional control variables, such as past 52-week returns, ROE, share turnover, and CAPEX. Table 2 provides detailed specifications of the variables used in the panel data regressions.

Table 2: Specification on dependent, independent and control variables used for panel regressions

Name	Type of variable	Definition
Month-ahead stock returns	Dependent	One-month ahead stock returns. For year t, monthly returns are taken from the end of April and matched with all the other control variables from the end of March of the same year.
Year-ahead ROA	Dependent	One-year ahead return on assets. For year t, ROA at the end of April is matched with data of one year prior at the end of March of year t-1.
HighMaterialESG (levels)	Independent	Indicator variable with a value of 1 for the top quintile of performers on material ESG scores (levels) each year, and 0 otherwise. [14]
HighMaterialESG (changes)	Independent	Indicator variable with a value of 1 for the top quintile of improvers on material ESG scores, and 0 otherwise.
TRESG (levels)	Independent	Indicator variable with a value of 1 for the top quintile of performers on the Thomson Reuters ESG scores (levels) each year, and 0 otherwise.
TRESG (changes)	Independent	Indicator variable with a value of 1 for the top quintile of improvers on TR ESG scores, and 0 otherwise.
Size	Control	Natural logarithm of end of month market capitalization in USD
MTB	Control	End of month market value to book value of common equity

[14] All of the dummy variables indicating material/total ESG performance are time-variant, meaning that they can be used in fixed effects estimation. They are time-variant because each year, the quintile of top performers has firms entering/exiting.

Leverage[15]	Control	(Short Term Debt & Current Portion of Long-Term Debt + Long Term Debt) / Total Assets * 100
Past 52 week returns	Control	Past 52-week return incorporating the price changes and any dividends for the last 52 weeks
ROE	Control	(Net Income – Preferred Dividend Requirement) / Average of Last Year's and Current Year's Common Equity * 100
Share turnover	Control	Shares traded on a particular day (at the end of a month) / total number of shares outstanding (at the end of a month)
CAPEX/Sales	Control	Capital expenditure / Net Sales or Revenues * 100

5.4 Data collection and sample construction

Table 3 shows the sample construction starting from the Thomson Reuters ESG database to obtaining the final sample. We begin by screening all companies that have been listed on a stock exchange in the UK, Norway, Finland, Sweden, and Denmark during the 2009-2018 period, with 3325 total number of available companies obtained. Our first filter is based on a company having an ESG score for at least one year in our sample period, after which 562 companies remain. This shows that there are more than 2700 companies that have not reported on ESG issues in our sample period. Our second filter relates to the materiality matching of SASB-Thomson Reuters. As the matching is done on the industry level, several industries totalling 74 companies are excluded because of particular issues that are required by SASB but not available in the Thomson Reuters ESG dataset. Finally, as the last screen, we exclude a further 20 companies due to missing firm fundamentals, leaving our final sample size to have 468 unique companies over ten years.

[15] In the panel data context, we use monthly data. Therefore, variables such as Leverage, ROE, CAPEX/Sales and ROA have same values for each month of a company fiscal year.

Table 3: Sample construction

Sample construction	**Screens**	**No. of firms**
Screen 1	*Country of exchange: UK, Norway, Denmark, Sweden, Finland*	3325
Screen 2	*ESG score in at least one year in the 2009-2017/2018 period*[16]	562
Excluded	*Specific industry categories required by SASB, not available in Thomson Reuters*	*74*
	of which: Meat & Poultry	5
	of which: Casinos & Gaming	6
	of which: Mortgage Finance	2
	of which: Security & Commodity Exchanges	1
	of which: Real Estate Services	8
	of which: Investment Trusts	52
Excluded	*Missing firm fundamentals*	*20*
Final sample	***Unique number of firms***	***468***[17]

The frequency of companies available in the sample is changing by year, ranging from 316 in the 2009 fiscal year to 465 companies for the 2017 fiscal year, as shown in Table 4. All of the companies are allocated to the 11 sectors provided by the Sustainable Industry Classification System (SICS) that is made available by SASB (SASB - c, n.d). This sector allocation is needed to facilitate the SASB-

[16] The last year a company can be included in the sample is if it has ESG data for FY 2017. That is because we use changes in ESG scores from FY 2017 to FY 2018, in order to form the final portfolios in the beginning of April in 2019 and hold them until March 2020.

[17] Our sample grows to 465 unique firms available at the end of FY 2017. It is however 468 *unique* firms available overall in the sample period. The difference comes because for some years, there are companies who are removed from the sample for that specific year, but were in the sample prior. Reasons for exclusions range from delisting to bankruptcy.

Thomson Reuters matching and creation of the material ESG scores that are assigned to each company.

Table 4: Number of companies by year

Year	Number of firms
FY 2009	316
FY 2010	324
FY 2011	341
FY 2012	343
FY 2013	348
FY 2014	361
FY 2015	420
FY 2016	439
FY 2017/2018	465

Table 5 shows the number of unique firms and associated unique years that constitute the sample, by sector. There are 44 companies from the consumer goods, 54 companies from the extractives & minerals processing, 67 from the financials, 38 from the food & beverage, 26 from the health care, 64 from the infrastructure, 8 from the renewable resources & alternative energy, 66 from the resource transformation, 35 from the services, 38 from the technology & communications and 28 from the transportation sector. Overall, 468 unique companies are comprising 3820 unique years.

Table 5: Unique firms and years by sector

SICS sector	Number of unique firms	Number of firm years
Consumer goods	44	348
Extractives & Minerals Processing	54	472
Financials	67	530
Food & Beverage	38	312
Health Care	26	192
Infrastructure	64	528
Renewable Resources & Alternative Energy	8	66
Resource Transformation	66	551
Services	35	315

Technology & Communications	38	288
Transportation	28	216
Total	**468**	**3818**

5.5 Summary statistics and correlation matrix

Table 6 presents summary statistics of the sample using the level values[18] of variables. If we compare the Q1, median and Q3 value for the two types of scores, we can conclude that TR ESG scores seem to have more extreme scores, both on the top and bottom sides. Additionally, the standard deviation of the materiality-adjusted scores is lower, a finding that is also supported by Khan et al. (2016) using KLD scores.

Table 6: Summary statistics of the sample

Variables	(1) Obs	(2) Mean	(3) SD	(4) Q1	(5) Median	(6) Q3
Material ESG[19]	3,818	52.16	14.31	43.33	52.63	60.71
TRESG	3,818	51.41	18.60	38.20	51.75	65.42
lnSize	3,818	21.86	1.437	20.93	21.77	22.75
MTB	3,818	3.582	31.25	1.160	2.070	3.780
ROA	3,818	7.680	14.68	2.850	6.190	10.98
Leverage	3,818	23.03	18.12	8.730	21.70	32.72

In Table 7, we show pairwise correlations using variable levels. Similar to Khan et al. (2016), we find positive correlations between ESG scores and firm size. Moreover, the correlation between size and Thomson Reuters scores is 0.586, whereas size and materiality-adjusted scores have a correlation of 0.269. This is as well supported by the findings of Khan et al. (2016). We provide several interpretations:

1) a larger company has more resources, staff, and organizational opportunities to pursue sustainability performance than a smaller company;
2) a bigger company is usually more covered in the media than a smaller company. This implies higher reputational risk coming from possible

[18] Summary statistics using changes in variables are shown in the Appendix.
[19] Please note that in the orthogonalization, changes in material ESG scores were done using unscaled scores. Summary statistics and pairwise correlations using changes in variables are shown in the Appendix. When reporting levels here, we scale material ESG scores on a 0-100 scale for easier interpretation and comparability with Thomson Reuters scores.

controversies being made public and creates more pressure to engage in sustainability to limit the potential reputation loss;

The correlation between material and TR ESG scores is 0.301. The finding shows that our scores do relay different information from TR scores. Khan et al. (2016) find a correlation closer to 0.6 between KLD material and total scores. One of the reasons for this discrepancy might be the specificity of the required indicators by SASB and their relative availability in the Thomson Reuters database.

Table 7: Correlation matrix with level variables

Variables	(1)	(2)	(3)	(4)	(5)	(6)
(1) lnSize	1.000					
(2) MTB	0.009	1.000				
	(0.574)					
(3) ROA	0.130*	0.478*	1.000			
	(0.000)	(0.000)				
(4) Leverage	0.006	-0.058*	-0.115*	1.000		
	(0.701)	(0.000)	(0.000)			
(5) Material ESG	0.269*	0.038*	0.060*	0.002	1.000	
	(0.000)	(0.020)	(0.000)	(0.881)		
(6) Total TR ESG	0.586*	-0.030	-0.030	0.101*	0.301*	1.000
	(0.000)	(0.067)	(0.061)	(0.000)	(0.000)	

**** p<0.01, ** p<0.05, * p<0.1*

In Table 8, we also show correlations between material, Thomson Reuters, and Bloomberg scores for our sample. The correlations between our scores and Bloomberg scores are even lower than those compared to TR scores, which is understandable given that our scores are based on indicators from the TR database. However, what is most interesting and in line with previous research, is that the correlation between scores of different ESG providers is in the 0.6-0.7 range.

Table 8: Correlation matrix of TR, Bloomberg and materiality-adjusted scores[20]

Variables	(1)	(2)	(3)
(1) Material ESG	1.000		

[20] Please note that Bloomberg scores were not available fully for our sample. Therefore, this correlation matrix takes our sample and excludes all firm-year observations for which Bloomberg does not have a score reported.

(2) Thomson Reuters ESG	0.267*	1.000	
	(0.000)		
(3) Bloomberg ESG	0.163*	0.702*	1.000
	(0.000)	(0.000)	

**** p<0.01, ** p<0.05, * p<0.1*

5.6 Materiality mapping - methodology

One of the most critical tasks for our research is to map as closely as possible Thomson Reuters ESG indicators to categories deemed material. SASB summarizes its sustainability standards in a unique framework known as the "Materiality map" (SASB – d, n.d). This is the primary tool that we used in our research to map materially relevant issues to our sample of companies. The map offers five dimensions: environment, social capital, human capital, business model & innovation, and leadership & governance. The dimensions entail 26 general issue categories such as GHG emissions, human rights & community relations, labour practices, etc.

For each industry in our sample, we use its SASB standard[21] containing the general issue categories that include the material industry-specific indicators. Although the 26 general issue categories are the same across each industry, the indicators within an issue category can vary across industries. For example, if we take GHG emissions as an issue category, it can be that two different indicators are relevant for the two industries: scope 1 emissions or NOx emissions. It is therefore essential that matching is done on the industry-by-industry level to capture the specific industry metrics and differences that arise between industries across different sectors, as well as industries within the same sector.

We find that only 155 unique indicators out of 423 are mapped to SASB's categories. Based on our matching, more than 60% of the Thomson Reuters ESG indicators represent issues that are not financially material to investors. Moreover, we find that the range of material indicators per industry is 3 to 38, compared to Thomson Reuters scores, which uses 70-170 indicators. The sector that has the

[21] One of the students takes lead in one half of the sectors whereas the other student takes lead in the second half of the sectors. Each of us then maps all of the industries within the sectors. We then cross-check our work and make decisions on discrepancies.

most materially relevant indicators is the Extractives & Minerals Processing. On average, industries within the sector had 27 indicators, and an overall 61 unique indicators were used. Table 9 shows an extract of indicators deemed material for three companies in different industries.

Table 9: Comparation of environmental material indicators across 3 industries (extract)

SICS Sector[22]	**Extractives & Minerals Processing**	**Consumer Goods**	**Consumer Goods**
SICS Industry	*Oil & Gas – Exploration & Production*	*Household & Personal Products*	*Apparel, Accessories & Footwear*
Company name	***British Petroleum***	***Unilever***	***Hennes & Mauritz***
Policy Water Efficiency	Immaterial	TRUE	Immaterial
Policy Sustainable Packaging	Immaterial	TRUE	Immaterial
Policy Environmental Supply Chain	Immaterial	TRUE	TRUE
Environment Management Team	TRUE	Immaterial	Immaterial
Environmental Materials Sourcing	Immaterial	Immaterial	TRUE
Toxic Chemicals Reduction	Immaterial	TRUE	TRUE

5.7 Materiality-adjusted ESG score creation

After all of the matching is done, we score the indicators to create material ESG scores needed for our research. Previous research on this topic uses mainly KLD ratings (Statman & Glushkov, 2009; Servaes & Tamayo, 2013), which are designed as a binary system of strengths and concerns. KLD uses 1 to indicate the presence of each criterion and 0 to indicate its absence (Khan et al., 2016). On the other hand, Thomson Reuters indicators can be qualitative, where they are marked with "TRUE" or "FALSE" to indicate presence/absence, or quantitative where numerical data is present.

Therefore we adopted an approach to re-state the combination of qualitative and quantitative indicators in a 1/0 binary system to be in line with previous research. The transformation is done in two steps. First, we transform all qualitative and quantitative data into the 1/0 system. Second, we apply polarity to discern the concerns from the strengths to create the final scores. Overall, the formation of

[22] The following table represents an extract of ESG indicators for three industries. For a full overview of the materially-mapped indicators by industry, please refer to the Appendix.

materiality scores is similar to that of Khan et al. (2016). To get to a materiality score for firm ***i*** in year ***t***, they subtracted the sum of concerns from the sum of strengths to arrive at a single net score:

$MaterialESG_{it} = \Sigma$ *Thomson Reuters Strength, SASB* $- \Sigma$ *Thomson Reuters* $Concern_{it}$*, SASB*

We start by transforming qualitative data by assigning 1 for "TRUE " and 0 for "FALSE ". Thomson Reuters report quantitative data either as a ratio or unscaled numerical value. For data that is unscaled, we follow the approach used by Quantitative Management Associates (2018). We manually scale all the unscaled indicators for each company in each year by the company's market capitalization of the same year-end. After that, in each specific year, for each ESG indicator, we use sector-specific median values as cut-off points. The total sample that was considered for calculation of medians was that of all listed companies on the stock exchanges in the UK, Norway, Sweden, Denmark, and Finland[23]. A quantitative ESG indicator receives 1 if it is above or equal to the median value, and 0 if it is below[24]. When all of the qualitative and quantitative indicators are transformed into the binary 1/0 system, we apply polarity to disseminate the concerns from strengths.

Finally, for a given company, there can be indicators that are deemed material by SASB, but data is not be reported in the Thomson Reuters database. We could not find previous discussions of how this was dealt with in other studies. Therefore we applied several assumptions: 1) even though a company does not report data for a given indicator, if it is deemed material by SASB, it should be included in the scoring in some way; 2) if an indicator is positively polarized and there is missing data, it gets assigned a 0; if an indicator is negatively polarized (all indicators besides controversies) and has missing data it gets a -1; if an indicator represents a controversy and there is missing data, it gets assigned a 0.

[23] The sample used for median calculation closely resembles our own sample. This is because the only companies that have data available to be included in median calculations each year, are the companies who actually have ESG score data for the same year.

[24] In general, this is true when the polarity of the indicator is positive. If polarity is negative, we assign -1 for "true" and 0 for "false" for qualitative data. Similarly, for negatively polarized quantitative data, we assign -1 for above or equal to median and 0 for below median.

As Thomson Reuters reports, all the data collection is done by their analysts from multiple sources such as corporate and sustainability reports or the media. It is specifically defined that all controversy indicators are based on the company being mentioned in the media. Therefore, we assumed if there is no data for a controversy indicator, then it is likely that the company did not have a controversy published in the media and should get a 0. Negatively polarized indicators that are not controversies, on the other hand, are mostly based on information coming from the company's reports. Therefore, for other negatively polarized indicators, we assign a -1, assuming that the company either chose not to disclose information or it does not focus on that specific issue yet. Because negatively polarized indicators represent concerns, we assume that companies with missing data should be penalized and therefore reduce the overall score.

The same logic was applied to positively polarized items. Namely, if there is no data, it is as if a company has chosen not to disclose or does not focus on the given issue. Here, we assign a 0 for missing data. This is because positively polarized items present a potential strength, so an absence of data should not reduce the score of the company, but rather not inflate it unnecessarily. In the end, we sum up the values of all the material indicators for each company in a given year to obtain its material ESG score.

6. Results and analysis

The following section presents results from the performed econometric models. In 6.1, we discuss results from orthogonalization and subsequent portfolio formation. We then present results from time-series OLS regressions using common risk factors from various asset pricing models as independent variables. These regressions help us answer Hypothesis 1 and Hypothesis 2. In 6.2, we focus on the panel data approach. Discussion is made with regards to model specification, model diagnostics, and potential remedies. In 6.2, we also present results for the four panel regression models shown in equations (4)-(7), which are done based on changes in ESG scores. Besides, we perform the analysis using the same four panels from equations (4)-(7), where we use level ESG scores.

6.1 Time series regression and orthogonalization

Materiality-adjusted ESG sorts

Table 10 shows the results from time-series regressions of portfolios sorted on the material ESG characteristic using the Fama-French 3-factor model. We estimate equation (1) cross-sectionally each year to obtain the residuals for every year. Companies are then sorted based on the value of their residuals obtained from the orthogonalization, and top/bottom quintile and decile cut-offs are used for the portfolio construction. The residuals represent yearly changes in material ESG scores that are not due to changes in firm size, its book-to-market ratio, leverage, ROA, or sector membership. Additionally, equation (1) was estimated using industry dummies, but the results remained unchanged[25].

Most of the variation in the returns of the material ESG sorted portfolios over time is explained by the market factor. All of the loadings on the market factor are significant at the 1% level. Most of the portfolios also load on the HML factor, but none of them has a significant SMB coefficient. Our research design examines whether the portfolios generate abnormal performance that cannot be attributed to common risk factors presented in the asset pricing models. We find that almost all of the FF3 alphas are insignificant beside the bottom quintile portfolio, which has an annualized underperformance of around 6%, significant at the 5% level.

Table 10: Fama-French 3-factor model using material ESG sorts

	Equal-weighted				**Value-weighted**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variable	Top Quin.	Bottom Quin.	Top Dec.	Bottom Dec.	Top Quin.	Bottom Quin.	Top Dec.	Bottom Dec.
MktRF	0.819***	0.831***	0.803***	0.829***	0.740***	0.808***	0.676***	0.830***
	(0.0773)	(0.0728)	(0.0834)	(0.0832)	(0.0736)	(0.0637)	(0.0818)	(0.0704)
HML	0.287***	0.148	0.208*	0.234**	0.0277	0.0593	-0.148	0.190**
	(0.104)	(0.0979)	(0.112)	(0.112)	(0.0991)	(0.0858)	(0.110)	(0.0948)
SMB	0.0491	0.0956	0.0979	0.161	0.0171	-0.160	0.128	-0.119

[25] Using industry dummies, we find that on average top/bottom quintiles/deciles are comprised of the same companies as when sector dummies are used.

	(0.129)	(0.122)	(0.140)	(0.139)	(0.123)	(0.107)	(0.137)	(0.118)
FF3 alpha	-0.224	-0.229	-0.330	-0.445	-0.176	-0.504**	-0.160	-0.350
	(0.284)	(0.267)	(0.306)	(0.305)	(0.270)	(0.234)	(0.300)	(0.258)
N	108	108	108	108	108	108	108	108
R^2	0.617	0.637	0.565	0.594	0.550	0.643	0.455	0.631

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Our results do not support the notion that the biggest improvers/decliners on material ESG achieve future abnormal performance. This is in contrast with Khan et al. (2016), who find that some top and bottom portfolios exhibit significant alphas. One possible way to interpret our results comes from the notion good ESG performance might be oriented more towards avoiding the downside risk rather than obtaining abnormal returns. One dimension of this argument is examined by Hoepner et al. (2018). The study examines the links between investor engagement in the company's ESG undertakings and the company downside risk. They specifically examine the engagement of large institutional investors and find that their engagement leads to reduced downside risk for the company. If an increasing number of institutional investors are aware and actively engage in ESG issues, then this is also a confirmation towards the argument that significant alphas based on material/total ESG sorts should not be present today. However, based on our results, we cannot fully reject the notion that being a bad performer on material ESG leads to future underperformance.

We also acknowledge that the FF3 model has relatively more difficulty in explaining the average returns of top quintile/decile portfolios compared to their bottom counterparts. This is evident given the R^2 is always lower for a top compared to a bottom portfolio across every decile/quintile.

Although not backed by statistical significance, it is important to discuss the economic intuition behind the other alpha estimates. All of the top quintile/decile portfolios have alphas that are less negative than their bottom quintile/decile counterparts. Although there is inconclusive evidence in the data, it seems that the notion of being a bad performer on material ESG issues might negatively affect the future stock returns of a company. This is in line with Khan et al. (2016), who find

negative alphas that are statistically significant for a larger number of portfolios than we do.

Given our results so far, we present the following possible interpretations:

1) Whereas Khan et al. (2016) use a sample spanning 1991-2014, we focus on a more recent period. As Gerard (2019) and other studies (Margolis et al., 2011) argue, it seems that good ESG performance translated into return outperformance in the 1990s and 2000s. Still, this effect has largely disappeared in more recent times. Khan et al. (2016) present the notion that the unavailability of material classifications in the past left investors unable to react properly when ESG scores were made public. They argue that the alphas realized through the effect of the company investments in material ESG issues had on its future profitability. However, for more than 50% of our sample period, investors had access to some form of SASB classifications. Even though the full set of standards was provided in 2018, the first provisional standards were published already in early 2013 (SASB – e, n.d). If investors have materiality classifications available and use them accordingly, it can be argued that the alphas should disappear. In this sense, investors would be able to immediately understand companies' performance on material issues when ESG scores are made public. Therefore, they can price in this information much faster than in the period when materiality classifications were not available. The arguments presented above lead more towards the notion that alphas shown in the work of Khan et al. (2016) are some sort of market inefficiency.
2) On the contrary, there is also the question of whether ESG scores can be a proxy for a common risk source that is not accounted for in contemporary asset pricing models. As known in the asset pricing literature, the key notion of an anomaly is its persistence. Most of the discovered anomalies, however, are largely non-existent after their discovery is made public. Given our results, we are inclined to adopt the argument that material ESG score alphas are some sort of temporary market inefficiency.

Thomson Reuters ESG sorts

Further, we present results from a Fama-French 3-factor model regression of portfolios sorted on the Thomson Reuters ESG scores characteristic. We replicate the orthogonalization process, but here we orthogonalize changes in the total TR ESG scores with respect to changes in the same firm characteristics. Our motivation is twofold: 1) we want to see whether our material ESG scores are a good predictor of returns relative to the TR ESG scores; 2) we want to observe whether the TR ESG scores relay useful information for future stock performance, as they are used by investors regularly. The second point is very important, as a multitude of investment decisions every day globally depend on the usage of TR ESG scores.

As Table 11 shows, we find insignificant alphas across all of the TR-sorted portfolios. Therefore, we fail to reject the null under each t-test, meaning that each of the portfolio alphas is indistinguishable from zero. We conclude that TR ESG scores are not a good predictor of future returns.

Table 11: Fama-French 3-factor model using Thomson Reuters ESG sorts

	Equal-weighted				**Value-weighted**			
Variables	(1) Top Quin.	(2) Bottom Quin.	(3) Top Dec.	(4) Bottom Dec.	(5) Top Quin,	(6) Bottom Quin.	(7) Top Dec.	(8) Bottom Dec.
MktRF	0.870***	0.801***	0.919***	0.831***	0.795***	0.781***	0.821***	0.815***
	(0.0776)	(0.0779)	(0.0841)	(0.0827)	(0.0710)	(0.0690)	(0.0768)	(0.0720)
HML	0.201*	0.286***	0.243**	0.334***	0.0691	0.0252	0.0804	0.0381
	(0.104)	(0.105)	(0.113)	(0.111)	(0.0956)	(0.0929)	(0.103)	(0.0969)
SMB	0.0892	0.123	0.117	0.171	-0.0483	-0.146	-0.0667	-0.0851
	(0.130)	(0.130)	(0.141)	(0.138)	(0.119)	(0.115)	(0.128)	(0.120)
FF3 alpha	-0.225	-0.0624	-0.353	-0.109	-0.236	-0.303	-0.367	0.0155
	(0.285)	(0.286)	(0.309)	(0.303)	(0.260)	(0.253)	(0.282)	(0.264)
N	108	108	108	108	108	108	108	108
R^2	0.633	0.614	0.627	0.615	0.598	0.586	0.575	0.595

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

If we compare the magnitude of the alpha estimates here with the material ESG alphas, we observe an important pattern. Although inconclusive based on significance, the magnitude of the alphas on the bottom portfolios of the material-sort are bigger than the bottom portfolio alphas of the TR-sort. Khan et al. (2016) find a similar phenomenon, however their results are supported with the statistical significance. Moreover, a similar economic occurrence can be seen on the top portfolios. For the material sorts, all of the top portfolios have a more positive alpha estimate than the equivalent top TR sorted portfolio.

From an economic perspective, we argue that it might be probable there is a difference between performing poorly on material issues compared to performing poorly on total ESG scores. The intuition is the following, performing poorly on material issues means not focusing on critical aspects of sustainability issues from a business operations perspective. However, performing poorly on TR ESG scores is mostly equivalent to performing poorly on immaterial issues. As we show, more than 75-80% of the indicators in Thomson Reuters ESG scores are deemed immaterial by SASB.

6.2 Panel data methods

Before we present the results, in this section, we focus on making clear the distinction between each of the panels. The results section will present four main panels, Panel A1, A2, B1, and B2. The models that are used to estimate each of these panels are equations (4)-(7). *A very important notion is that all of these four panels use as an independent variable of interest an indicator of top improvers based* ***on changes*** *in material or Thomson Reuters ESG scores.*

In the same context, we additionally present the results of Panels A1.1, A2.1, B1.1, and B2.1. *These four panels are done in addition to the main ones and use as an independent variable of interest an indicator of top performers based* ***on levels*** *in material or Thomson Reuters ESG scores.*

Until the work of Khan et al. (2016), previous research on our topic was focusing on the usage of levels of ESG scores. Therefore, our thesis tries to add value by combining the two approaches. Our primary motivation is in line with Khan et al.

(2016), where we want to understand how an improvement/decline in sustainability investments over time affects future financial performance. However, it is also of importance to understand how obtaining a high ESG score at one point in time affects future financial performance. We provide the following arguments for using both changes and levels in ESG scores:

1. The question that arises is whether investors care more about how a company improves/declines on its ESG performance or do they predominantly choose companies based on their static score at a given point in time. More sophisticated investors would likely be interested in the ongoing process of the sustainable behaviour of a company over time, where changes in scores would be appropriate. However, there might be other investors who predominantly incorporate ESG issues in their portfolio decisions based on a more static approach, where levels in ESG scores would be the choice.
2. The distinction is also important from a firm perspective. A similar group of companies will likely dominate the top quintiles for an extended period of time if ESG levels are used. Companies that constitute a quintile for a given year are likely to have built the infrastructure, obtained the policies, and put teams in place in terms of ESG activities. Therefore, it would be relatively more straightforward for those companies to maintain their high ESG score once that they achieve it. Moreover, given the costs incurred as well as the infrastructure already being in place, it can be assumed that it would be difficult to see significant decreases in ESG scores from top performers since the assets are already in use. Furthermore, it is difficult for a company that has achieved a high ESG score in a given year to decline significantly, as that movement can lead to loss of reputational capital that was previously created. To conclude, we can examine whether being a top improver in terms of changes or top performer based on levels is beneficial to the company.

Model diagnostics

In this section, we present model specifications and diagnostics. All of the models presented in this section use changes in ESG scores as the indicator variable. We

present the results for the same tests using the versions of the models with levels of ESG score as an indicator in the Appendix.

In order to build towards the final models used, we start by specifying a fixed effects model of the following form, which is in line with **Hypothesis 1**:

$$1\text{-month return}_{i,t+1} = \beta_0 + D_1 HighMaterialESG_{i,t} + \beta_2 Past52weekreturns_{i,t} + \beta_3 lnSize_{i,t} + \beta_4 MTB_{i,t} + \beta_5 Leverage_{i,t} + \beta_6 Shareturnover_{i,t} + \beta_7 ROE_{i,t} + \beta_8 CAPEX_{i,t} + FirmFixedEffects_i + u_{it} \quad (8)$$

For each of the four hypotheses, we use a slightly modified model than equation (8) as a starting point. The modifications come from the usage of different dependent and independent variables of interest.

For **Hypothesis 2**, the starting model is the same as in equation (8), besides the *HighMaterialESG* dummy being replaced with a *TRESG dummy*. **For Hypothesis 3**, the starting model is the same as equation (8), besides the *1-month return* being replaced with *1-year ahead ROA*. **For Hypothesis 4**, the starting model is the same as equation (8), besides the *HighMaterialESG dummy* being replaced with a *TRESG dummy* and the *1-month return* being replaced *with 1-year ROA.* These are the four basic models on which we separately apply all of the specification and model diagnostic tests shown below, to obtain the final forms for each model. The final forms of the models used in estimation are equations (4)-(7).

Firm level heterogeneities

The first test that is performed is a restricted F-test that examines the presence of individual effects - α_i in the data (Verbic, 2018). As Table 12 shows, for each of the four specified models, we reject the null hypothesis and conclude that there is a presence of individual effects, making the use of pooled OLS inappropriate.

Table 12: Restricted F-test (Pooled OLS vs. FE)

Model version tested	F- statistic	Prob. > F (p-value)	Reject H_0
Month-ahead returns and material ESG (changes) scores	2.68	0.0000	Yes

Month-ahead returns and TR ESG (changes) scores	2.69	0.0000	Yes
Year-ahead ROA and material ESG (changes) scores	133.94	0.0000	Yes
Year-ahead ROA and TR ESG (changes) scores	133.97	0.0000	Yes

For hypotheses 1 and 2, the dependent variables are month-ahead stock returns. For hypotheses 3 and 4 the dependent variables are year-ahead ROA. For hypotheses 1 and 3 the independent variable of interest is an indicator variable for top quintile improvers on material ESG scores (changes). For hypotheses 2 and 4 the independent variable of interest is an indicator variable for top quintile improvers on Thomson Reuters ESG scores (changes).

Fixed or random effects estimator

Further, a Hausman test is performed to see whether the REE or FEE would be preferred as our estimator. As shown in Table 13, for each of the four model specifications, we reject the null and conclude that the FEE is the preferred estimator.

Table 13: Hausman test (FE vs RE)

Model tested	**$\chi 2$ statistic**	**Prob. > $\chi 2$ (p-value)**	**Reject H_0**
Month-ahead returns and material ESG (changes) scores	1089.41	0.0000	Yes
Month-ahead returns and TR ESG (changes) scores	1090.49	0.0000	Yes
Year-ahead ROA and material ESG (changes) scores	6979.10	0.0000	Yes
Year-ahead ROA and TR ESG (changes) scores	6896.51	0.0000	Yes

Testing for joint significance of time dummies

Subsequently, we test the joint significance of adding time (month) dummies to our models. Table 14 shows that we reject the null and conclude that time-dummies should be a part of the regressions or each of the four models. Therefore, based on the results of the presented statistical tests until now, we specify the models shown in equations (4), (5), (6), and (7).

Table 14: Joint F-test for time dummies

Model tested	F-statistic	Prob. > F (p-value)	Reject H_0
Month-ahead returns and material ESG (changes) scores	60.33	0.0000	Yes
Month-ahead returns and TR ESG (changes) scores	60.33	0.0000	Yes
Year-ahead ROA and material ESG (changes) scores	3.04	0.0000	Yes
Year-ahead ROA and TR ESG (changes) scores	3.00	0.0000	Yes

Groupwise heteroskedasticity

After all of the models are estimated, we perform contemporary tests of model diagnostics in a panel data context. We start by using the modified Wald's test for groupwise heteroskedasticity. For each of the four models, the null hypothesis that the variance of the error terms is equal across entities is rejected. As shown in Table 15, we conclude that groupwise heteroskedasticity is present.

Table 15: Modified Wald's test of groupwise heteroskedasticity

Model tested	$\chi 2$ statistic	Prob. > $\chi 2$ (p-value)	Reject H_0
Month-ahead returns and material ESG scores (changes)	34515.36	0.0000	Yes
Month-ahead returns and TR ESG scores (changes)	34440.15	0.0000	Yes
Year-ahead ROA and material ESG scores (changes)	3.3+09	0.0000	Yes
Year-ahead ROA and TR ESG scores (changes)	3.4+09	0.0000	Yes

Serial correlation

Afterward, we turn towards testing for serial correlation in Table 16. For all of the four models, we reject the null and conclude that there is a presence of order one autocorrelation of the residuals.

Table 16: Wooldridge test of AR(1) autocorrelation

Model tested	**F-statistic**	**Prob. > F (p-value)**	**Reject H_0**
Month-ahead returns and material ESG scores (changes)	187.224	0.0000	Yes
Month-ahead returns and TR ESG scores (changes)	187.212	0.0000	Yes
Year-ahead ROA and material ESG scores (changes)	1.914e+06	0.0000	Yes
Year-ahead ROA and TR ESG scores (changes)	1.917+06	0.0000	Yes

Pesaran test for cross-sectional dependence (spatial correlation)

Eventually, we test for spatial correlation by running a post-estimation Pesaran CD test. For each of the models, we reject the null and find the presence of a cross-sectional correlation as shown in Table 17.

Table 17: Pesaran CD test of cross-sectional dependence

Model tested	**CD statistic**	**p-value**	**Reject H_0**
Month-ahead returns and material ESG scores	443.523	0.000	Yes
Month-ahead returns and TR ESG scores	443.544	0.000	Yes
Year-ahead ROA and material ESG scores	48.942	0.000	Yes
Year-ahead ROA and TR ESG scores	48.073	0.000	Yes

Model remedies

Given the results of the tests above, model remedies are needed to proceed with correct statistical inference. The approach that we undertake is estimating fixed effects models and correcting the standard errors post-estimation. For that reason, our regression results include estimates with standard errors clustered[26] at the firm level as well as Driscoll-Kraay standard errors.

Clustered standard errors

Clustered standard errors correct for the presence of heteroskedasticity and autocorrelation, but not spatial correlation. Usage of unadjusted (OLS) SEs in the presence of heteroskedasticity and autocorrelation leads to misleadingly high t-statistics and low p-values (Cameron & Miller, 2015). For example, in un-tabulated results, we find that the OLS SEs for our models are several times lower than the clustered SEs.

The motivation for using clustered standard errors in our study is twofold: 1) our model diagnostic tests showed the presence of autocorrelation and heteroskedasticity, which calls for using clustered SEs; 2) the approach of Khan et al. (2016), which uses a similar type of data also acknowledges the use of clustered standard errors.

However, there is no consensus among researchers regarding the choice of the level of clustering. Therefore, we follow Khan et al. (2016) and cluster standard errors at the firm level. Clustered standard errors that are reported are now autocorrelation and heteroskedasticity robust.

Driscoll-Kraay standard errors

In reporting our results, we go one step further than Khan et al. (2016) and calculate Driscoll-Kraay (1998) standard errors that correct for the presence of spatial correlation. The assumption underlying the clustered SEs is that the error terms are independent across firms at a given point in time (Vogelsgang, 2012).

[26] Also known as Rogers (1993) standard errors.

Assuming that there is no spatial correlation might be problematic given our dataset. As Norway, Sweden, Denmark, and Finland are part of the Nordics, we hypothesize that their financial markets are integrated and cross-dependent. The underlying reasons for this argument are numerous: geographical proximity, shared values, similar code of business ethics, etc. Moreover, there might be a cross-sectional correlation through membership in the same industry, as it is likely that firms will have some sort of mutual dependence of their error terms. This notion is supported empirically, as the Pesaran CD test shows the presence of cross-sectional dependence. Therefore, we also compute Driscoll-Kraay standard errors, which assume an error structure that is heteroskedastic, autocorrelated within groups as well as possibly correlated between groups (Hoechle, 2007). These standard errors are then heteroskedasticity-autocorrelation-spatial correlation robust (HACSC).

Panel regressions results

This section presents the results from panel data regressions. The panels will be presented as follows:

- for **Hypothesis 1**, Panel A1 estimating equation (4) using *changes* in ESG scores and Panel A1.1 estimating (4) using *levels* in ESG scores;
- for **Hypothesis 2**, Panel A2 estimating equation (5) using *changes* in ESG scores and Panel A2.1 estimating (5) using *levels* in ESG scores;
- for **Hypothesis 3**, Panel B1 estimating equation (6) using *changes* in ESG scores, and Panel B1.1 estimating (6) using *level*s in ESG scores;
- for **Hypothesis 4**, Panel B2 estimating equation (7) using *changes* in ESG scores, and Panel B2.2 estimating (7) using *levels* in ESG scores.

Hypothesis 1 – Panel A1 and A1.1

Panel A1 below is based on estimating equation (4) and uses as dependent variable month-ahead stock returns. The results shown in Panel A1 are in line with the findings using the time-series methodology. The within estimator is used, meaning that all of the estimates are based on the within-group variation.

Results from Panel A1 show that the estimate on the indicator variable for top performance on material ESG score changes is negative and insignificant. It is in

line with our previous findings using time-series regressions. Additionally, we find that most of the control variables are significant at the 1 or 5% level. Size is negatively related to future returns. This finding is consistent with the notion of the size premium, as initially shown by Banz (1981) and later by Fama & French (1992). Past 52 week returns have a significant positive coefficient, indicating some presence of momentum. Higher leverage affects returns negatively, while higher ROE affects it positively. Capital expenditures seem to affect future returns negatively. The only control that is insignificant is the market-to-book ratio.

Panel A1[27]: month-ahead stock returns and material ESG scores (changes)

Variables	(1) Clustered SE	(2) Driscoll-Kraay SE
Top Material ESG (changes)	-0.111	-0.111
	(0.150)	(0.137)
Past 52 week returns	0.00952***	0.00952***
	(0.00152)	(0.00358)
Size	-3.095***	-3.095***
	(0.297)	(0.375)
MTB	-0.00193	-0.00193
	(0.00435)	(0.00202)
Leverage	-0.0693***	-0.0693***
	(0.0124)	(0.0117)
Share turnover	0.946*	0.946**
	(0.488)	(0.407)
ROE	0.00466**	0.00466***
	(0.00205)	(0.000998)
CAPEX	-0.00301*	-0.00301**
	(0.00177)	(0.00133)
Constant	70.86***	70.86***
	(6.550)	(8.030)
Observations	37,194	37,194
R-squared	0.153	0.153
Number of firms	432	432
Firm FE	Yes	Yes
Month FE	Yes	Yes

The dependent variable is monthly returns starting at the end of April of year t. The Independent variable of interest is Top Material ESG. This is an indicator variable that takes a value of 1 for the top quintile of improvers based on the annual changes in material ESG scores, and 0 otherwise. Monthly returns in year t are matched with ESG data that represents the change in material ESG score from the end of t-2 to the end of t-1. All of the other variables are controls, where their values at the end of March in year t are matched with monthly returns in April of year t. Robust standard errors are shown in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

[27] All of the panels using as indicator ESG score changes were estimated using 96 month dummies as time fixed effects. In the interest of space, those estimates were dropped when results were exported.

As an addition, we also report results using material ESG score levels in Panel A1.1. The results are very similar to Panel A1, as the estimate of the indicator of top performance on material ESG score levels is also statistically insignificant. Results show similar significance for the controls as in Panel A1.

Panel A1.1: Month-ahead stock returns and material ESG scores (levels)

Variables	(1) Firm-level clustered SE	(2) Driscoll-Kraay SE
Top Material ESG (levels)	0.0928	0.0928
	(0.202)	(0.217)
Past 52 week returns	0.00937***	0.00937***
	(0.00136)	(0.00311)
Size	-2.647***	-2.647***
	(0.242)	(0.379)
MTB	-0.00202	-0.00202
	(0.00378)	(0.00192)
Leverage	-0.0617***	-0.0617***
	(0.0104)	(0.0113)
Share turnover	0.875	0.875**
	(0.553)	(0.410)
ROE	0.00363***	0.00363***
	(0.00130)	(0.000967)
CAPEX	-0.00260***	-0.00260***
	(0.000928)	(0.000859)
Constant	59.64***	59.64***
	(5.245)	(8.038)
Observations	41,370	41,370
R-squared	0.161	0.161
Number of Firms	433	433
Firm FE	Yes	Yes
Month FE	Yes	Yes

The dependent variable is monthly returns starting at the end of April of year t. The independent variable of interest is Top Material ESG. It is an indicator variable where a value of 1 reflects the top quintile of material ESG performers based on the level scores at the end of December in year t-1 and 0 for all other companies in the same period. All other variables are control variables, whereas their values at the end of March in year t are matched with monthly returns at the end of April in year t. Robust standard errors are shown in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Overall, the results of Panel A1 and A1.1 are in line with the results of the time-series methodology. We do not find conclusive evidence of a link between material ESG performance and future stock returns. This is opposite to the findings of Khan et al. (2016), who find positive estimates for the material ESG indicator variables in their panel regressions.

We adopt the following argument for this discrepancy:
In line with the findings of Margolis et al. (2011), where they report that the relationship between ESG and stock performance was present in the past, but has since decreased and is almost non-existent, we argue that the markets are pricing material ESG performance more efficiently in our sample period. Investors are now more aware of material classifications and have had SASB standards available for some years, allowing them to incorporate the standards faster into their ESG considerations of investment decisions.

Hypothesis 2 – Panel A2 and A2.1
We present the results of Panels A2 and A2.1, estimated using equation (5) and Thomson Reuters ESG score changes/levels. The results from Panel A2 are as well in line with the results from the time-series methodology.

We do not find a conclusive link between top performers/improvers on Thomson Reuters total ESG scores and future stock returns. From an economic interpretation, we do see that the estimates of the ESG performance indicator are more negative when using TR ESG scores compared to material ESG scores.

Overall, using both time-series and panel regressions, we find mixed evidence regarding the link between material or total ESG performance and future stock performance. The only significant finding is that some bottom portfolios sorted on changes in material ESG scores exhibit negative alphas.

Panel A2: month-ahead stock returns and TR ESG (changes)

Variables	(1) Clustered SE	(2) Driscoll-Kraay SE
Top TR ESG (changes)	-0.162	-0.162
	(0.134)	(0.130)
Past 52 week returns	0.00947***	0.00947***
	(0.00152)	(0.00358)
Size	-3.090***	-3.090***
	(0.297)	(0.374)
MTB	-0.00195	-0.00195
	(0.00437)	(0.00203)
Leverage	-0.0693***	-0.0693***
	(0.0124)	(0.0117)
Share turnover	0.946*	0.946**
	(0.487)	(0.408)
ROE	0.00464**	0.00464***
	(0.00206)	(0.000998)
CAPEX	-0.00302*	-0.00302**
	(0.00177)	(0.00133)
Constant	70.77***	70.77***
	(6.547)	(8.010)
Observations	37,194	37,194
R-squared	0.153	0.153
Number of Firms	432	432
Firm FE	Yes	Yes
Month FE	Yes	Yes

The dependent variable is monthly returns starting at the end of April of year t. The independent variable of interest is Top TR ESG. This is an indicator variable that takes a value of 1 for the top quintile of performers based on the annual changes in TR ESG scores, and 0 otherwise. Monthly returns in year t are matched with ESG data that represents the change in TR ESG score from the end of t-2 to the end of t-1. All of the other variables are controls, where their values at the end of March in year t are matched with monthly returns in April of year t. Robust standard errors are shown in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Panel A2.1: Month-ahead stock returns TR ESG scores (levels)

Variables	(1) Clustered SE	(2) Driscoll-Kraay SE
Top TR ESG (levels)	-0.127	-0.127
	(0.156)	(0.163)
Past 52 week returns	0.00936***	0.00936***
	(0.00137)	(0.00310)
Size	-2.646***	-2.646***

	(0.242)	(0.378)
MTB	-0.00203	-0.00203
	(0.00377)	(0.00192)
Leverage	-0.0617***	-0.0617***
	(0.0104)	(0.0113)
Share turnover	0.877	0.877**
	(0.553)	(0.411)
ROE	0.00362***	0.00362***
	(0.00130)	(0.000963)
CAPEX	-0.00261***	-0.00261***
	(0.000926)	(0.000858)
Constant	59.69***	59.69***
	(5.251)	(8.045)
Observations	41,370	41,370
R-squared	0.161	0.161
Number of Firms	433	433
Firm FE	Yes	Yes
Month FE	Yes	Yes

The dependent variable is monthly returns starting at the end of April of year t. The independent variable of interest is Top TR ESG. It is an indicator variable where a value of 1 reflects the top quintile of material ESG performers based on the level scores at the end of December in year t-1 and 0 for all other companies in the same period. All other variables are control variables, whereas their values at the end of March in year t are matched with monthly returns at the end of April in year t. Robust standard errors are shown in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

We outline one important notion that is connected with ESG data limitations, namely the way that Thomson Reuters and KLD scores are constructed differs. Previous research predominantly uses KLD scores (Waddock & Greaves, 1997; Hong & Kostovetsky, 2012; Cheng et al., 2013; Khan et al., 2016). KLD scores are favoured to Thomson Reuters, because the ESG indicators used to construct the scores in KLD are showed in a comparable standardized format. In contrast, indicators used in Thomson Reuters[28] or Bloomberg scores are predominantly "true" or "false" statements showing only a presence or absence of company disclosure on a given indicator (Khan et al., 2016). The difference in score

[28] TR standardizes the information that is gathered from annual reports, websites, sustainability reports and etc. so that the indicators that contain the information can be comparable across the universe of companies. However, TR still presents most of the indicators in a "true" or "false" format that just indicates presence or absence of company disclosure on that indicator. In that light, TR does not distinguish how good each company is compared to another on a given indicator.

construction and meaning of underlying indicators may have an impact on the significance of our estimates.

Hypothesis 3 – Panel B1 and B1.1

Further, we present the results of Panel B1 and B1.1 using equation (6) and changes/levels in material ESG scores. Gerard (2019) states that a company engaging in ESG entails expenditures that affect the bottom line in the short-run, whereas benefits obtained are uncertain and far in the future. Gerard (2019) also mentions that it is probable that:

> "*Improved CSR is attractive only as long as the costs are lower than the expected benefits in terms of future profitability and returns.*"

We use 1-year ahead ROA to examine the implication of this argument. In that way, we make a distinction from the work of Khan et al. (2016), who use a 2-year ahead ROA. Therefore, we are focusing on the short-run impact that increased ESG engagement has on a company's bottom line.

There will likely be some costs entailed by the company to have good performance. Still, there is probably a distinction between the amount of costs incurred for total or material ESG engagement. Our findings support the notion that being good on total ESG scores entails making investments on four to five times more indicators than being good on material ESG scores does. Therefore, firms that do not distinguish what component of sustainable issues is material to their industry are incurring expenses for engagement in ESG issues that are not relevant to their operations, which can affect their future profitability negatively.

The positive effects on future profitability, on the other hand, can come from increased reputation, better management of stakeholder interests, actively supporting sustainable development, reduced exposure to climate risks or social capital controversies, etc. Irrespective of which ESG issues a company focuses, there is likely to be some benefit for taking into account sustainability in their business model.

As shown in Panel B1, using changes in material ESG scores, we find that the estimate of the indicator for the top quintile of improver companies is positive and

statistically significant. Firms that increase their material ESG score and thus move towards the top quintiles attain an increase in their future ROA. Khan et al. (2016) find the same results when using 2-year to 5-year ahead ROA.

Panel B1.1 examines the same issue using levels in material ESG scores. Results show that the estimate of the indicator for the top quintile of performers is insignificant. Therefore, it does not seem that achieving a high material ESG score in terms of levels has any effect on future profitability.

We provide several arguments for the difference:

1) In line with Khan et al. (2016), it seems that the improvements on material ESG scores made by a company have a positive effect on its future ROA. As the material ESG score increases, relative to the average material ESG score of the company, its future ROA increases, relative to the firm average ROA.
2) However, we do not find similar results using material ESG score levels. One explanation comes from the usage of the FEE, which uses the within-firm variation of the variables. If there are many firms where the ESG ratings stay the same through time, the within-firm variation of the ESG score is low, and this might work against finding significance in estimates when using levels. In our sample, we find that the quintiles formed on levels have less variety of companies compared to quintiles formed on changes.
3) Moreover, we provide an economic interpretation. Once a company achieves a high score, it intends to maintain it at an approximately same level for a prolonged period. The motivation from a firm-perspective is that it would be very costly to entail a significant decline in ESG scores, mainly from a reputational perspective. Therefore, companies that achieve top scores through expensive investments in the first place need to incur ongoing costs to maintain that level. It can be argued that the benefits obtained from an increase of reputational, social, or operational standpoint are declining over time and are offset by the costs of maintaining the high scores, especially as sustainable behaviour becomes the norm rather than the exception worldwide. Therefore, even though a company can be a top

performer based on the ESG level score, it does not mean that this should affect future ROA.

4) It can be assumed that as long as the companies are improving in terms of sustainability investments and moving towards the top quintiles of material ESG scores, the costs they entail from the upgrades are less than their future potential benefits. Thus the material sustainability investments can have a positive effect on their future profitability. In general, this finding is line with the idea of SASB that improvements on the material component of sustainability should affect profitability. Khan et al. (2016) also argue that the profitability channel is where the company sees the effect of its material investments pay off.

Panel B1: Year-ahead ROA and material ESG scores (changes)

Variables	(1) Clustered SE	(2) Driscoll-Kraay SE
Top Material ESG (changes)	0.362	0.362*
	(0.374)	(0.195)
Past 52 week returns	0.0171***	0.0171***
	(0.00539)	(0.00454)
Size	-0.240	-0.240
	(0.607)	(0.267)
MTB	0.00827	0.00827
	(0.00760)	(0.00515)
Leverage	-0.0414	-0.0414***
	(0.0385)	(0.0151)
Share turnover	-0.431	-0.431
	(0.583)	(0.672)
ROE	0.0128**	0.0128***
	(0.00577)	(0.00382)
CAPEX	-0.00738	-0.00738**
	(0.00624)	(0.00281)
Constant	13.18	13.18**
	(13.15)	(5.909)
Observations	35,464	35,464
R-squared	0.030	0.030
Number of Firms	430	430
Firm FE	Yes	Yes
Month FE	Yes	Yes

The dependent variable is ROA starting in April of year t+1. The independent variable of interest is Top Material ESG. This is an indicator variable of the top quintile of performers based on changes in material ESG scores from year t-2 to t-1. All other variables are control variables, whereas their values at the end of March in year t is matched with April of year t+1 ROA data. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Panel B1.1: Year-ahead ROA and material ESG score (levels)

Variables	(1) Clustered SE	(2) Driscoll-Kraay SE
Top Material ESG (levels)	-0.279	-0.279
	(0.722)	(0.284)
Past 52 week returns	0.0202***	0.0202***
	(0.00511)	(0.00376)
Size	0.196	0.196
	(0.569)	(0.275)
MTB	0.0251	0.0251**
	(0.0185)	(0.0123)
Leverage	-0.0782**	-0.0782***
	(0.0340)	(0.0209)
Share turnover	-0.318	-0.318
	(0.683)	(0.705)
ROE	0.0144*	0.0144***
	(0.00832)	(0.00449)
CAPEX	-0.00273	-0.00273
	(0.00448)	(0.00240)
Constant	3.893	3.893
	(12.35)	(5.854)
Observations	39,975	39,975
R-squared	0.051	0.051
Number of Firms	431	431
Firm FE	Yes	Yes
Month FE	Yes	Yes

The dependent variable is ROA starting in April of year t+1. The independent variable of interest is Top Material ESG, an indicator variable where a value of 1 reflects the top quintile of material ESG performers at the end of December in year t-1, and 0 for all other companies in the same period. All other variables are control variables, whereas their values at the end of March in year t is matched with April of year t+1 ROA data. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Hypothesis 4 – Panel B2 and B2.1

In the end, we present the results of Panels B2 and B2.1 estimating equation (7) and using changes/levels in TR ESG scores. The results in Panel B2 show that the estimate of the indicator of top quintile performers on changes in TR ESG scores is not significant. However, the estimate of the indicator using levels of TR ESG scores is negative and significant at the 1% level.

From the results in this section, we present the following conclusions:

1) When it comes to Thomson Reuters scores, there is no link between a company improving on its score and its future profitability. This is in line with our thinking, as an improvement on TR scores entails improvement mostly on issues that are financially immaterial to a company's operations. This means that an improvement in TR compared to material ESG issues is likely to entail higher costs, as TR scores are based on 4 to 5 times more indicators for sustainability issues than material ESG scores are. Therefore, it seems to be a difference between improving on material compared to total ESG issues and the implications of both actions on future profitability.
2) Similarly to levels of material ESG scores, we find that the top quintiles on levels of TR scores have less variety than the TR quintiles based on changes. This supports our argument that companies who once achieve a high ESG score intend to continue with the investments to maintain it. However, the nuance here compared to material scores is that the investments in TR scores also incorporate improvements on immaterial issues. Therefore, the costs might be even higher, but the benefits are less, compared to the previous discussion of material ESG scores. Overall, if a company continues to invest each year to maintain its good performance mostly on issues deemed immaterial, the actual costs of doing so will likely prevail over the expected future benefits, potentially having a negative effect on future profitability.

Panel B2: Year-ahead ROA and TR ESG score (changes)

VARIABLES	(1) Clustered SE	(2) Driscoll-Kraay SE
Top TR ESG (changes)	-0.0438	-0.0438
	(0.353)	(0.158)
Past 52 week returns	0.0171***	0.0171***
	(0.00544)	(0.00455)
Size	-0.236	-0.236
	(0.607)	(0.267)
MTB	0.00832	0.00832
	(0.00765)	(0.00517)
Leverage	-0.0410	-0.0410***
	(0.0385)	(0.0152)
Share turnover	-0.426	-0.426
	(0.584)	(0.670)
ROE	0.0128**	0.0128***
	(0.00578)	(0.00382)
CAPEX	-0.00740	-0.00740**
	(0.00625)	(0.00282)
Constant	13.15	13.15**
	(13.15)	(5.909)
Observations	35,464	35,464
R-squared	0.030	0.030
Number of Firms	430	430
Firm FE	Yes	Yes
Month FE	Yes	Yes

The dependent variable is ROA starting in April of year t+1. The independent variable of interest is Top TR ESG. This is an indicator variable of the top quintile of performers based on changes in TR ESG scores from year t-2 to t-1. All other variables are control variables, whereas their values at the end of March in year t is matched with April of year t+1 ROA data. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Panel B2.1: Year-ahead ROA TR ESG scores (levels)

Variables	(1) Clustered SE	(2) Driscol-Kraay SE
Top TR ESG (levels)	-0.839*	-0.839***
	(0.446)	(0.239)
Past 52 week returns	0.0202***	0.0202***
	(0.00511)	(0.00375)
Size	0.189	0.189

	(0.569)	(0.276)
MTB	0.0251	0.0251**
	(0.0185)	(0.0123)
Leverage	-0.0774**	-0.0774***
	(0.0339)	(0.0209)
Share turnover	-0.302	-0.302
	(0.686)	(0.703)
ROE	0.0144*	0.0144***
	(0.00830)	(0.00448)
CAPEX	-0.00276	-0.00276
	(0.00447)	(0.00241)
Constant	4.164	4.164
	(12.35)	(5.900)
Observations	39,975	39,975
R-squared	0.051	0.051
Number of Firms	431	431
Firm FE	Yes	Yes
Month FE	Yes	Yes

The dependent variable is ROA starting in April of year t+1. The Independent variable of interest is now the Top TR ESG, an indicator variable where a value of 1 reflects the top quintile of Thomson Reuters ESG performers at the end of December in year t-1, and 0 for all other companies in the same period. All other variables are control variables, whereas their values at the end of March in year t is matched with April of year t+1 ROA data. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Examining the implications of ESG engagement on short-run profitability is also a dimension where our research provides value and extends the work of Khan et al. (2016). Overall, we find that firms who are good improvers on material investments see their future ROA improve. On the other hand, when companies have already achieved a high score, being a top performer constantly does not have an effect on future ROA. Using total TR scores, we find that top improvers on total sustainability issues do not have their future ROA affected in any way. However, if companies already have a top TR ESG scores, maintaining it through the years seems to affect their profitability negatively.

6.3 Robustness checks

Alternative asset pricing models

In this section, we present a series of robustness tests. First, we use alternative factor models to investigate the future stock returns of the top/bottom material ESG sorts. As shown in Table 26, we find results that support the regressions using the Fama-French 3-factor model. Results using CAPM are similar to our main regressions, with the same bottom quintile having a significant annualized underperformance of 5.5%. Moreover, the bottom decile equal-weighted portfolio alpha is significant, and annualized underperformance is 7.5%. However, such results using CAPM are expected since it has been shown that model has more difficulties explaining the cross-section of average stock returns compared to the FF3 model. When using the Fama-French-Carhart 4-factor model, we see that the only portfolio remaining with a significant alpha is the bottom quintile portfolio, the same as in our main regressions.

Table 26: CAPM and Fama-French-Carhart 4-factor alphas for material ESG sorts

	Equal-weighted				**Value-weighted**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variable	Top Quintile	Bottom Quintile	Top Decile	Bottom Decile	Top Quintile	Bottom Quintile	Top Decile	Bottom Decile
Jensen alphas	-0.410	-0.354	-0.490	-0.645**	-0.199	-0.474**	-0.126	-0.411
	(0.281)	(0.259)	(0.298)	(0.300)	(0.258)	(0.226)	(0.290)	(0.252)
FFC 4 alphas	-0.172	-0.208	-0.285	-0.410	-0.162	-0.489**	-0.150	-0.320
	(0.280)	(0.268)	(0.304)	(0.305)	(0.272)	(0.235)	(0.303)	(0.258)

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Excluding non-December fiscal year-ends

Next, we exclude companies that have fiscal year-ends different from December. As we form portfolios based on changes in scores, in the original sample, there are more than 100 companies that have fiscal year-ends in months other than December. For example, companies with fiscal-year ends in August have their TR ESG scores available around the end of November, following corporate reporting patterns. Effectively, there is an implied four-month gap between TR ESG

publications and portfolio formation in March of the next year. Thus, for this sub-sample of firms, it means that investors have a longer time-period to understand and acknowledge the scores compared to firms with December fiscal year-ends. This is the motivation for testing our models using only companies with a December fiscal year-end. Moreover, this approach of robustness testing is common in asset pricing literature, as many seminal finance papers include this robustness exercise (Fama & French, 1993).

The estimates that are shown in Table 27 are very similar to the estimates of the original sample. Most of the variation in returns of the material ESG sorted portfolios is explained by the returns of the market factor, the SMB factor, and the HML factor. Similarly, we find that only the bottom quintile portfolio has statistically significant alpha, indicating annualized underperformance of about 6.4% compared to the 6% estimate of the same portfolio in the original sample.

Table 27: Fama-French 3-factor model excluding non-December year-ends

	Equal-weighted				**Value-weighted**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables	Top Quint.	Bottom Quint.	Top Dec.	Bottom Dec.	Top Quint.	Bottom Quint.	Top Dec.	Bottom Dec.
MktRF	0.850***	0.893***	0.866***	0.916***	0.754***	0.859***	0.753***	0.878***
	(0.0860)	(0.0804)	(0.100)	(0.100)	(0.0842)	(0.0692)	(0.0975)	(0.0903)
SMB	0.0287	0.0817	0.0375	0.173	0.0743	-0.202*	0.179	-0.194
	(0.144)	(0.135)	(0.167)	(0.168)	(0.141)	(0.116)	(0.163)	(0.151)
HML	0.250**	0.143	0.147	0.309**	-0.0148	0.113	-0.110	0.434***
	(0.116)	(0.108)	(0.135)	(0.135)	(0.113)	(0.0931)	(0.131)	(0.122)
FF3 alpha	-0.358	-0.360	-0.593	-0.523	-0.230	-0.555**	-0.261	-0.465
	(0.316)	(0.295)	(0.367)	(0.368)	(0.309)	(0.254)	(0.358)	(0.331)
N	108	108	108	108	108	108	108	108
R^2	0.573	0.619	0.492	0.558	0.496	0.636	0.432	0.571

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Nordics-only subsample

In the end, we present another robustness check using only Nordic listed firms in our sample. As Stradman & Freeman (2015) point out, the Nordic countries have long been considered among the leaders in CSR and sustainability globally. They point out that the area has long-standing traditions of stakeholder engagement and that the concept of shared value has had its roots in the Nordics.

Therefore, using only Nordic-listed companies in our sample, we create material ESG sorted portfolios and obtain estimates from the Fama-French 3-factor model regression. It is worth mentioning that the sample size is smaller, starting with 102 companies in 2009 and growing to 146 by the end of 2017.

According to results from Table 28, for all of the top and bottom portfolios, FF3 alphas are insignificant. From an economic interpretation, we see that the intercept estimates are more negative for each bottom portfolio compared to its top counterpart. Moreover, this is the only sample specification where we find positive alpha estimates for the top portfolios. However, as all of the alphas are insignificant, we conclude that material ESG scores are not good predictors of future stock performance for Nordic-listed companies. The finding of insignificant results might be due to the small sample of the Nordics and possible small variation of changes in their ESG scores. If a lot of the companies in the Nordics already have high scores, then it is difficult to identify the effect of material sustainability improvements on their financial performance.

Table 28: Fama-French 3-factor model using Nordics-only material ESG sorts

	Equal-weighted				**Value-weighted**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variable	Top Quint.	Bottom Quint.	Top Dec.	Bottom Dec.	Top Quint.	Bottom Quint.	Top Dec.	Bottom Dec.
MktRF	0.821***	0.911***	0.879***	0.965***	0.734***	0.796***	0.736***	0.776***
	(0.0872)	(0.0910)	(0.103)	(0.105)	(0.0851)	(0.0821)	(0.104)	(0.0978)
SMB	0.0866	0.128	0.132	0.183	0.00534	-0.125	0.0271	-0.140
	(0.146)	(0.152)	(0.172)	(0.176)	(0.142)	(0.137)	(0.174)	(0.164)
HML	0.0296	0.0653	0.140	0.172	-0.211*	-0.0498	-0.310**	-0.0683
	(0.117)	(0.123)	(0.138)	(0.142)	(0.115)	(0.110)	(0.140)	(0.132)

FF3 alphas	-0.0258	-0.217	-0.188	-0.392	0.0256	-0.0430	0.0327	-0.389
	(0.320)	(0.334)	(0.377)	(0.386)	(0.312)	(0.301)	(0.382)	(0.359)
N	108	108	108	108	108	108	108	108
R^2	0.527	0.565	0.498	0.541	0.452	0.505	0.359	0.401

7. Conclusion

Using the materiality classifications provided by SASB, we hand-map relevant metrics of the Thomson Reuters database into material ESG scores. We find that the newly created scores seem to relay different information than the original TR scores. The correlation between the material and Thomson Reuters ESG scores is around 0.31. Even if there are discrepancies between specific definitions by SASB and the availability of such particular metrics in the TR database, we still find that most of the metrics used to compute TR ESG scores are deemed immaterial. For reference, we find that 10 to 30 of the TR ESG indicators are considered material depending on the industry. At the same time, Thomson Reuters reports using 70 to 170 indicators in the construction of their ESG scores.

In this research, we find mixed evidence on the link between material ESG performance and future stock returns. Most of the alphas of the material ESG sorted portfolios are insignificant, besides one bottom portfolio that entails a negative annualized alpha of 6%. Our results contrast the findings of Khan et al. (2016), who report that most of the top/bottom portfolios have significant alphas.

Replicating the process with Thomson Reuters ESG scores, we do not find any significant links to future stock performance as all of the portfolio alphas are insignificant. It seems that the Thomson Reuters total ESG scores are not a good predictor of future stock performance. Similarly, Khan et al. (2016) find mostly insignificant alphas when using total KLD scores.

To complement the analysis of the link between changes in ESG scores and future stock returns, we additionally use panel data models. The panel data models include additional firm characteristics as controls as well as firm and time fixed effects.

The estimates of the indicator variables that represent the top quintile performers and improvers on material and TR ESG scores are negative and insignificant.

When focusing on the link between material ESG performance and future ROA, we find that changes in material ESG scores positively affect future profitability. This positive effect is not present when we use levels of material ESG scores rather than changes. Therefore, it seems that the beneficial effect on future profitability comes through the improvement in material investments through the years, rather than obtaining a high score and maintaining it on the same level.

On the contrary, when examining the link between Thomson Reuters ESG scores and future ROA, we do not find that changes in TR ESG scores have any effect on future profitability. The coefficient on the indicator for the top quintile of improvers is not significant. However, when we use TR level scores rather than changes, the indicator for top performers is negative and significant. It seems that there is no link between improving on TR ESG scores and future profitability, whereas investing to obtain high TR ESG scores and remaining a top performer on TR ESG has a detrimental effect on future profitability.

We provide the following interpretation for the different results obtained using future stock returns and ROA:

1) One explanation for the results is that nowadays, markets are pricing in ESG performance and especially the material component, more efficiently than in the past. The argument of Khan et al. (2016) is that the reason for significant alphas in their sample period is that investors did not have publicly available material classifications. It can be argued that for our sample period, SASB classifications have been public, and therefore available to investors to implement more informed decisions on firm ESG performance. Moreover, the alphas discovered by Khan et al. (2016) should persist if material ESG performance is a proxy for some sort of common risk factor for the whole cross-section of companies. On the contrary, if those alphas are not present, then they are likely to be a result of some sort of temporary market inefficiency.

2) When examining results where future ROA is used, we discover that changes on issues that are deemed material to a company's operations have a positive effect on its future profitability. This is in line with yet one more argument of Khan et al. (2016), who state that the alphas that they discover are realized through the profitability channel. For their sample period, they argue that the material sustainability investments translate into future profitability, which then over time results in the alphas that they find. The difference in our findings is that even though material sustainability investments seem to affect future profitability, there is no alpha as a result of these developments. Even if the markets do price the material ESG performance more efficiently nowadays, that does not constrain engagement into material investments to be translated into higher future profitability.

Our research has implications, both for investors and companies. As our results suggest, making investment decisions based solely on Thomson Reuters scores does not have any predictive power in terms of future stock performance. Moreover, investors should consider the possible negative effect on profitability, when investing into the companies with top TR scores each year. Similarly, if using exclusion screening, they might screen out companies who have low scores for performing poorly on issues that are irrelevant to their operations.

From a company perspective, not being able to distinguish the material ESG issues that a company needs to focus on can entail significant usage of resources for purposes that do not bring any future benefits. To obtain high TR ESG scores and maintain them, a company has to incur high costs, with its future profitability being negatively affected. On the contrary, investments that lead to improving material sustainability have a positive effect on future profitability.

This thesis leaves many different dimensions open for future research. One can apply materiality mappings provided by SASB to ESG data from various sources. In that light, researchers would be able to distinguish whether there is an issue of ESG data collection and methodology of forming the scores by the data providers. At the same time, this procedure will shed light on the potential robustness of

SASB's materiality classifications to ESG data coming from different sources. Another avenue for research is to expand the materiality classifications to ESG data for a more comprehensive geographical sample, such as the rest of Europe, Asia, Latin America, and other global markets. In this way, researchers will be able to see whether SASB's classifications are specific to the US markets and if there is a need to expand them in a more international context. Finally, management behaviour regarding investments in material and immaterial sustainability issues can be further researched. It will be useful to know what are the drivers of decision-making and what are the trade-offs between focusing on all or just material ESG issues, from an agency theory perspective.

REFERENCE LIST

Attig, N., El Ghoul, S., Guedhami, O., & Suh, J. (2013). Corporate social responsibility and credit ratings. *Journal of business ethics, 117*(4), 679694.

Baltagi, B. H., Feng, Q., & Kao, C. (2012). A Lagrange Multiplier test for cross-sectional dependence in a fixed effects panel data model. *Journal of Econometrics, 170*(1), 164-177.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of financial economics, 9*(1), 3-18.

Bartels, B. (2008). Beyond" fixed versus random effects": a framework for improving substantive and statistical analysis of panel, time-series cross-sectional, and multilevel data. *The Society for Political Methodology, 9*, 1-43.

Baum, C. F. (2006). Stata tip 38: Testing for groupwise heteroskedasticity. *The Stata Journal, 6*(4), 590-592.

Bebchuk, L., Cohen, A., & Ferrell, A. (2009). What matters in corporate governance?. *The Review of financial studies, 22*(2), 783-827.

Bénabou, R., & Tirole, J. (2010). Individual and corporate social responsibility. *Economica, 77*(305), 1-19.

Berg, F., Koelbel, J. F., & Rigobon, R. (2020). Aggregate confusion: the divergence of ESG ratings. *Available at SSRN 3438533.*

BlackRock (2020). A fundamental reshaping of finance. Retrieved from: https://www.blackrock.com/corporate/investor-relations/larry-fink-ceo-letter

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of human resources, 50*(2), 317-372.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance, 52*(1), 57-82.

Cheng, I. H., Hong, H., & Shue, K. (2013). *Do managers do good with other people's money?* (No. w19432). National Bureau of Economic Research.

Clark, G. L., Feiner, A., & Viehs, M. (2015). From the stockholder to the stakeholder: How sustainability can drive financial outperformance. *Available at SSRN 2508281.*

Cochrane, J. H. (2009). *Asset pricing: Revised edition.* Princeton University press.

Compact, U. G. (2004). Who cares wins: Connecting financial markets to a changing world. *New York.*

Data library (n.d). Fama-French research factors. Retrieved from: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

De Hoyos, R. E., & Sarafidis, V. (2006). Testing for cross-sectional dependence in panel-data models. *The stata journal, 6*(4), 482-496.

Dowell, G., Hart, S., & Yeung, B. (2000). Do corporate global environmental standards create or destroy market value?. *Management science, 46*(8), 1059-1074.

Driscoll, J. C., & Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Review of economics and statistics, 80*(4), 549-560.

Eccles, R. G., & Serafeim, G. (2013). The performance frontier. *Harvard business review, 91*(5), 50-60.

Edmans, A. (2011). Does the stock market fully value intangibles? Employee satisfaction and equity prices. *Journal of Financial economics, 101*(3), 621-640.

European Commission (2011). Corporate Social Responsibility. Retrieved from: https://ec.europa.eu/growth/industry/sustainability/corporate-social-responsibility_en

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, *47*(2), 427-465.

Fama, E. F., & Kenneth, R. (1993). French, 1993, Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1), 3-56.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of financial economics, 116*(1), 1-22.

Flammer, C. (2013). Corporate social responsibility and shareholder reaction: The environmental awareness of investors. *Academy of Management Journal, 56*(3), 758-781.

Freeman, R. E. (2010). *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*(4), 210-233.

Friedman, M. (1970). A Friedman doctrine: The social responsibility of business is to increase its profits. *The New York Times Magazine, 13*(1970), 32-33.

Gerard, Bruno, 2019, ESG and Socially Responsible Investment: A Critical Review. *Beta, Scandinavian Journal of Business Research, 53*(1), pp. 61-83. https://doi.org/10.18261/issn.1504-3134-2019-01-05.

Giese, G. & Nagy, Z. (2018). How markets price ESG. MSCI ESG Research. Retrieved from: https://www.msci.com/documents/10199/f3f252a7-0dbd-035b-e67c-964a87fbcd03

Grewal, J., Hauptmann, C., & Serafeim, G. (2020). Material sustainability information and stock price informativeness. *Journal of Business Ethics*, 1-32.

GSIA. (2018). Global Sustainable Investment Review. Retrieved from: http://www.gsi-alliance.org/

Gujarati, D. N. (2009). *Basic econometrics*. Tata McGraw-Hill Education.

Hartzmark, S. M., & Sussman, A. B. (2019). Do investors value sustainability? A natural experiment examining ranking and fund flows. *The Journal of Finance, 74*(6), 2789-2837.

Hillman, A. J., & Keim, G. D. (2001). Shareholder value, stakeholder management, and social issues: What's the bottom line?. *Strategic management journal, 22*(2), 125-139.

Hoechle, D. (2007). Robust standard errors for panel regressions with cross-sectional dependence. *The stata journal, 7*(3), 281-312.

Hoepner, A. G., Oikonomou, I., Sautner, Z., Starks, L. T., & Zhou, X. (2018). ESG shareholder engagement and downside risk.

Hong, H., & Kostovetsky, L. (2012). Red and blue investing: Values and finance. *Journal of Financial Economics, 103*(1), 1-19.

Hou, K., Xue, C., & Zhang, L. (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies, 28*(3), 650-705.

Jensen, M. C., Black, F., & Scholes, M. S. (1972). The capital asset pricing model: Some empirical tests.

Khan, M., Serafeim, G., & Yoon, A. (2016). Corporate sustainability: First evidence on materiality. *The accounting review, 91*(6), 1697-1724.

Krüger, P. (2015). Corporate goodness and shareholder wealth. *Journal of financial economics, 115*(2), 304-329.

Lee, K. H., Min, B., & Yook, K. H. (2015). The impacts of carbon (CO2) emissions and environmental research and development (R&D) investment on firm performance. *International Journal of Production Economics, 167*, 1-11.

Lins, K. V., Servaes, H., & Tamayo, A. (2017). Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis. *The Journal of Finance, 72*(4), 1785-1824.

Lintner, J. (1975). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. In *Stochastic optimization models in finance* (pp. 131-155). Academic Press.

Lo, S. F., & Sheu, H. J. (2007). Is corporate sustainability a value-increasing strategy for business?. *Corporate Governance: An International Review, 15*(2), 345-358.

Margolis, J. D., Elfenbein, H. A., & Walsh, J. P. (2009). Does it pay to be good... and does it matter? A meta-analysis of the relationship between corporate social and financial performance. And does it matter.

Marti, C.P., Rovira-Val, M.C., Drescher, L.G.J., 2015. Are firms that contribute to sustainable development better financially? *Corporate Social Responsibility and Environmental Management 22*, 305–319.

Mohin, T. and Rogers, J. (2017, March 16). *How to approach corporate sustainability reporting in 2017*. Greenbiz.

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of Financial Studies, 22*(1), 435-480.

QMA (2018). Integration ESG in Portfolio Construction: An innovative data lens gives a wider view on sustainable investing. Retrieved from: https://cdn.pficdn.com/cms/qma/sites/default/files/static_files/pdf/QMA_Integrating_ESG_in_Portfolio_Construction_Brief.ADA.pdf

Refinitiv (n.d). ESG data. Retrieved from: https://www.refinitiv.com/en/financial-data/company-data/esg-data

Robecco (n.d). Glossary – Socially responsible investment. Retrieved from: https://www.robeco.com/en/key-strengths/sustainable-investing/glossary/socially-responsible-investing.html

Rogers, W. (1993). Regression standard errors in clustered samples. *Stata technical bulletin, 13*, 19-23.

SASB – a (n.d). Front page. Retrieved from: https://www.sasb.org/

SASB – b (n.d). Standards overview. Retrieved from: https://www.sasb.org/standards-overview/sasb-and-others/

SASB – c (n.d). Sustainable Industries Classification System. Retrieved from: https://www.sasb.org/find-your-industry/

SASB – d (n.d). Materiality map. Retrieved from: https://materiality.sasb.org/

SASB – e (n.d). Standard setting archive. Retrieved from: https://www.sasb.org/standard-setting-archive/

Servaes, H., & Tamayo, A. (2013). The impact of corporate social responsibility on firm value: The role of customer awareness. *Management science, 59*(5), 1045-1061.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance, 19*(3), 425-442.

Statman, M., & Glushkov, D. (2009). The wages of social responsibility. *Financial Analysts Journal, 65*(4), 33-46.

UN PRI. (2005) Principles for responsible investment. Retrieved from: https://www.unpri.org/pri/about-the-pri

Verbič, M. (2018). *Lecture on panel data analysis*. School of Business and Economics, University of Ljubljana, Slovenia.

Vogelsang, T. J. (2012). Heteroskedasticity, autocorrelation, and spatial correlation robust inference in linear panel models with fixed-effects. *Journal of Econometrics, 166*(2), 303-319.

Waddock, S. A., & Graves, S. B. (1997). The corporate social performance–financial performance link. *Strategic management journal, 18*(4), 303-319.

APPENDIX

Sample summary statistics using changes in variables

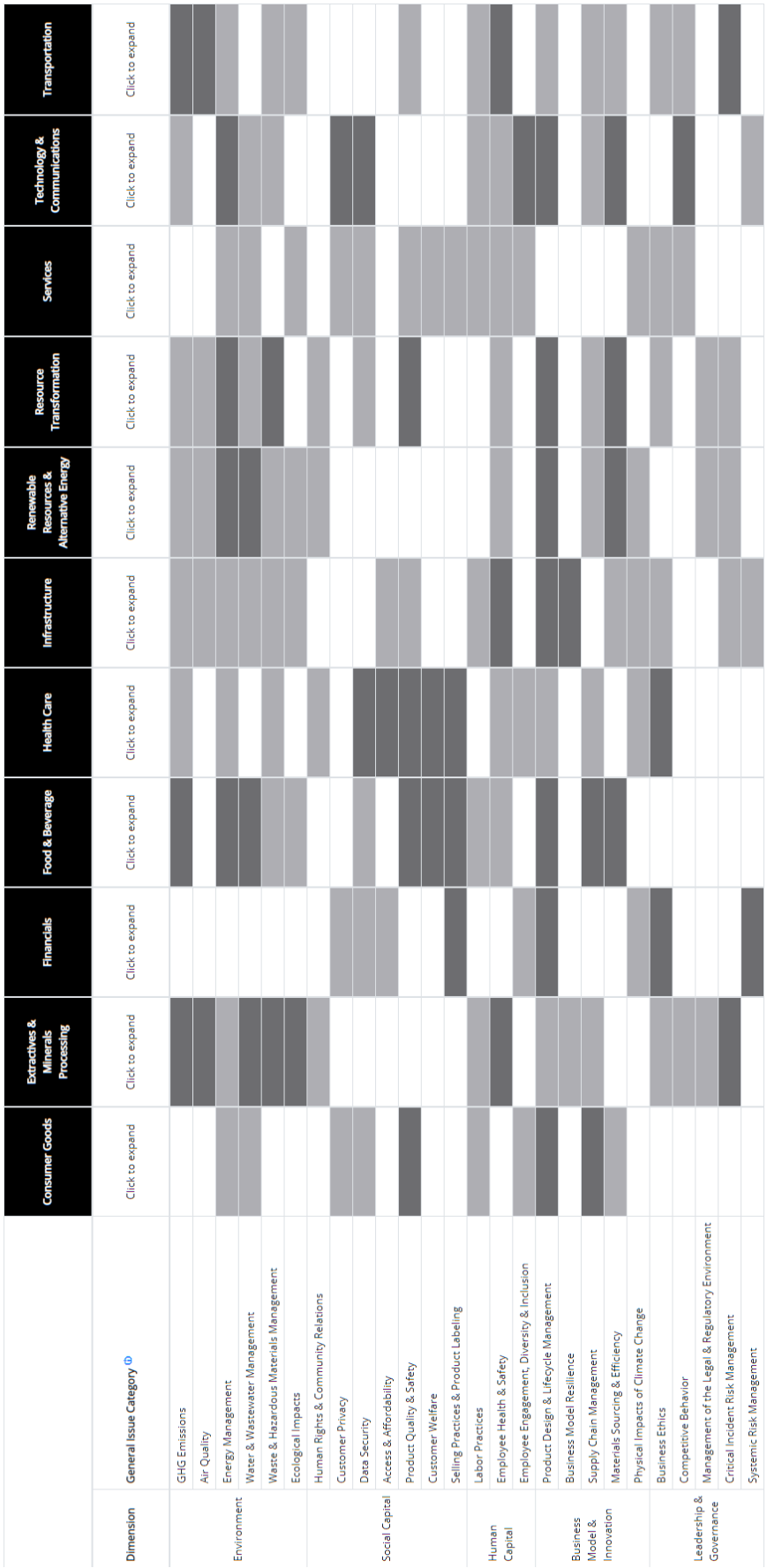
Variables	(1) N	(2) Mean	(3) SD	(4) Q1	(5) Median	(6) Q3
ΔMaterial ESG	3,355	0.135	1.052	0	0	1
ΔTotal TR ESG	3,355	1.582	6.502	-2.220	1.280	5.170
ΔlnSize	3,355	0.0317	0.389	-0.168	0.0484	0.247
ΔMTB	3,355	-0.0709	36.60	-0.430	-0.0200	0.390
ΔROA	3,355	0.0761	9.713	-2.050	0.0700	2.070
ΔLeverage	3,355	-0.274	6.047	-2.630	-0.180	1.550

Correlation matrix using changes in variables

Variable	(1)	(2)	(3)	(4)	(5)	(6)
(1) ΔMaterial ESG_t	1.000					
(2) ΔTotal TR ESG_t	0.174*	1.000				
	(0.000)					
(3) $\Delta lnSize_t$	0.042*	-0.020	1.000			
	(0.014)	(0.249)				
(4) ΔMTB_t	0.037*	0.012	0.005	1.000		
	(0.032)	(0.495)	(0.787)			
(5) ΔROA_t	0.022	0.003	0.193*	0.036*	1.000	
	(0.209)	(0.854)	(0.000)	(0.037)		
(6) $\Delta Leverage_t$	-0.031	-0.019	-0.243*	0.021	-0.227*	1.000
	(0.076)	(0.268)	(0.000)	(0.235)	(0.000)	

**** p<0.01, ** p<0.05, * p<0.1*

Sector level SASB Materiality Map®



TR ESG indicators defined as material according to SASB standards – part I

Consumer Goods	Extractives & Minerals Processing	Financials	Food & Beverage	Health Care
VOC Emissions Reduction	Targets Emissions	Climate Change Commercial Risks Opportunities	Targets Emissions	Waste Total
VOC Emissions	Policy Emissions	Environmental Products	Policy Emissions	Waste Recycling Ratio
Policy Water Efficiency	Biodiversity Impact Reduction	Environmental Assets Under Mgt	CO2 Equivalent Emissions Direct	Hazardous Waste
Policy Energy Efficiency	CO2 Equivalent Emissions Direct	Environmental Project Financing	Climate Change Commercial Risks Opportunities	Policy Sustainable Packaging
Policy Sustainable Packaging	Carbon Offsets/Credits	Revenue from Environmental Products	Ozone-Depleting Substances	Policy Environmental Supply Chain
Policy Environmental Supply Chain	Emissions Trading	Policy Diversity and Opportunity	Waste Total	Toxic Chemicals Reduction
Environmental Materials Sourcing	Climate Change Commercial Risks Opportunities	Women Employees	Waste Recycling Ratio	Renewable Energy Use Ratio
Toxic Chemicals Reduction	Flaring Gases	Women Managers	Policy Water Efficiency	Energy Use Total
Renewable Energy Use Ratio	NOx Emissions	Whistleblower Protection	Policy Sustainable Packaging	Fleet Fuel Consumption
Energy Use Total	SOx Emissions	Community Lending and Investments	Policy Environmental Supply Chain	Take-back and Recycling Initiatives
Water Withdrawal Total	Particulate Matter Emissions Reduction	Bribery, Corruption and Fraud Controversies	Environmental Materials Sourcing	Product Environmental Responsible
Environmental Supply Chain Management	VOC Emissions	Crisis Management Systems	Renewable Energy Use Ratio	Training and Development Policy
Environmental Supply Chain Monitoring	Waste Total	Business Ethics Controversies	Energy Use Total	Policy Career Development
Environmental Products	Waste Recycling Ratio	Policy Data Privacy	Water Withdrawal Total	Flexible Working Hours
Eco-Design Products	Hazardous Waste	Policy Responsible Marketing	Environmental Supply Chain Management	Total Injury Rate Total
Labeled Wood Percentage	Water Discharged	Product Responsibility Monitoring	Environmental Supply Chain Monitoring	Lost Days To Total Days
Product Impact Minimization	Environmental Restoration Initiatives	Customer Satisfaction	Environmental Controversies	Employees Health & Safety Controve
Take-back and Recycling Initiatives	Accidental Spills	Product Access Low Price	Fleet Fuel Consumption	Voluntary Turnover of Employees
Product Environmental Responsible Use	Environmental Expenditures	Consumer Complaints Controversies	Organic Products Initiatives	Involuntary Turnover of Employees
Revenue from Environmental Products	Environmental Provisions	Product Quality Controversies	GMO Products	HSMS Certified Percent
Policy Employee Health & Safety	Policy Water Efficiency	Responsible Marketing Controversies	Policy Supply Chain Health & Safety	Policy Bribery and Corruption
Policy Supply Chain Health & Safety	Policy Environmental Supply Chain	Controversies Privacy	Supply Chain Health & Safety Training	Policy Business Ethics
Policy Diversity and Opportunity	Targets Water Efficiency	Policy Board Diversity	Supply Chain Health & Safety Improvements	Diseases of the Developing World
Supply Chain Health & Safety Training	Environment Management Team	Board Cultural Diversity	Salaries and Wages from CSR reporting	Bribery, Corruption and Fraud Contr
Supply Chain Health & Safety Improvements	Toxic Chemicals Reduction	Executive Members Gender Diversity	Trade Union Representation	Crisis Management Systems
Employees Health & Safety OHSAS 18001	Renewable Energy Use Ratio	Insider Dealings Controversies	Strikes	Public Health Controversies
Salaries and Wages from CSR reporting	Energy Use Total	CSR Sustainability Reporting	Total Injury Rate Contractors	Business Ethics Controversies
Women Employees	Renewable Energy Produced		Total Injury Rate Employees	Policy Customer Health & Safety
Women Managers	Green Buildings		Contractor Accidents	Policy Data Privacy
Supplier ESG training	Fresh Water Withdrawal Total		Employee Accidents	Policy Responsible Marketing
Wages Working Condition Controversies	Water Recycled		Employee Fatalities	Product Responsibility Monitoring
Diversity and Opportunity Controversies	Environmental Supply Chain Management		Contractor Fatalities	ISO 9000
Voluntary Turnover of Employees	Land Environmental Impact Reduction		Supplier ESG training	QMS Certified Percent
Involuntary Turnover of Employees	Environmental Controversies		Wages Working Condition Controversies	Product Access Low Price
Human Rights Contractor	Fleet Fuel Consumption		Diversity and Opportunity Controversies	Consumer Complaints Controversies
Employee Engagement Voluntary Work	Renewable/Clean Energy Products		Employees Health & Safety Controversies	Product Quality Controversies
Policy Customer Health & Safety	Sustainable Building Products		Voluntary Turnover of Employees	Responsible Marketing Controversie
Policy Data Privacy	Real Estate Sustainability Certifications		Involuntary Turnover of Employees	Controversies Customer Health & Sa
Product Responsibility Monitoring	Policy Employee Health & Safety		Human Rights Contractor	Controversies Responsible R&D
Controversies Customer Health & Safety	Health & Safety Training		Policy Customer Health & Safety	Controversies Privacy
Controversies Privacy	Trade Union Representation		Policy Data Privacy	Controversies Responsible Marketin
Product Recall	Strikes		Policy Responsible Marketing	Controversies Product Access
	Employee Health & Safety Training Hours		Product Responsibility Monitoring	FDA Warning Letters
	Total Injury Rate Total		Healthy Food or Products	Product Delays
	Total Injury Rate Contractors		Obesity Risk	Product Recall
	Total Injury Rate Employees		Responsible Marketing Controversies	
	Accidents Total		Controversies Customer Health & Safety	
	Contractor Accidents		Controversies Privacy	
	Employee Accidents		Controversies Responsible Marketing	
	Occupational Diseases		Product Recall	
	Employee Fatalities			
	Contractor Fatalities			
	Employees Health & Safety Controversies			
	Policy Human Rights			
	Policy Bribery and Corruption			
	Policy Community Involvement			
	Extractive Industries Transparency Initiative			
	Political Contributions			
	Crisis Management Systems			
	Anti-competition Controversies			
	Critical Countries Controversies			

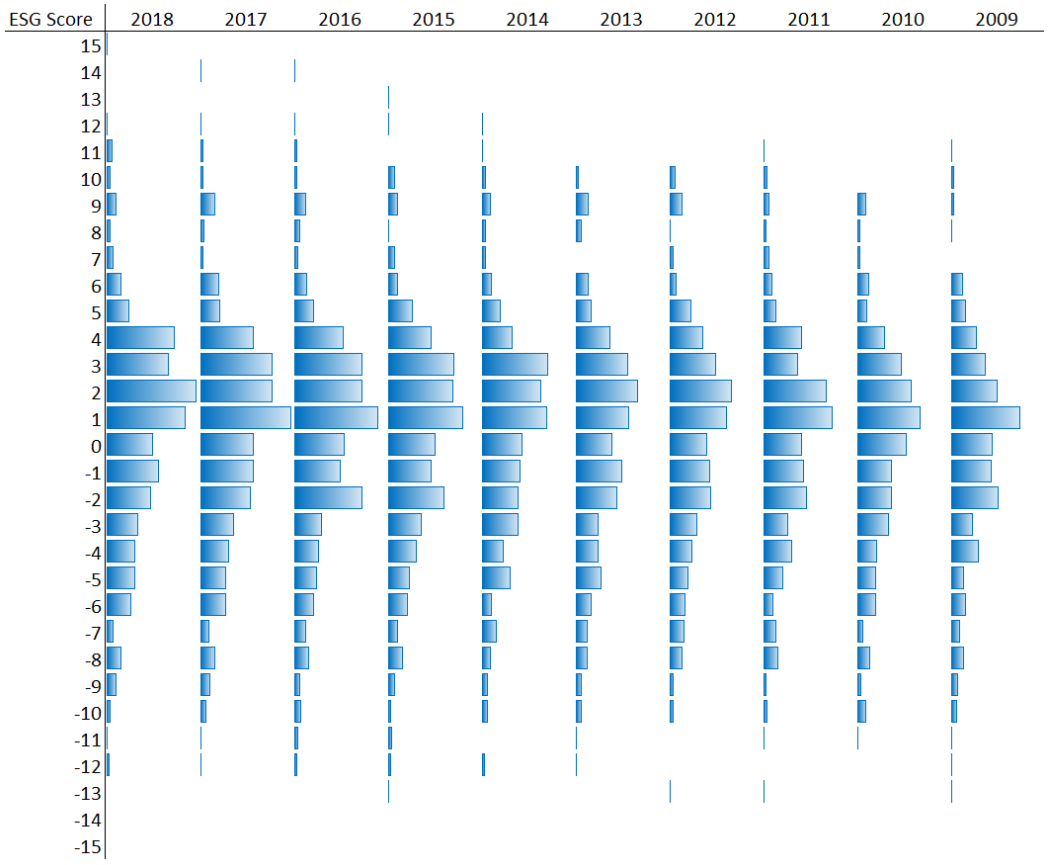
TR ESG indicators defined as material according to SASB standards – part II

Infrastructure	Renewable Resources & Alternative Energy	Resource Transformation	Services	Technology & Communications	Transportation
Targets Emissions	Targets Emissions	Targets Emissions	Biodiversity Impact Reduction	Targets Emissions	Targets Emissions
Policy Emissions	Policy Emissions	Policy Emissions	Environment Management Team	Policy Emissions	Policy Emissions
Biodiversity Impact Reduction	CO2 Equivalent Emissions Direct	CO2 Equivalent Emissions Direct	Renewable Energy Use Ratio	CO2 Equivalent Emissions Direct	Biodiversity Impact Reduction
CO2 Equivalent Emissions Direct	Carbon Offsets/Credits	Carbon Offsets/Credits	Energy Use Total	Waste Recycling Ratio	CO2 Equivalent Emissions Direct
Emissions Trading	Emissions Trading	Emissions Trading	Water Withdrawal Total	Hazardous Waste	Emissions Trading
Climate Change Commercial Risks Opportunities	Ozone-Depleting Substances	Ozone-Depleting Substances	Policy Diversity and Opportunity	e-Waste Reduction	Climate Change Commercial Risks Opportunities
Flaring Gases	NOx Emissions	NOx Emissions	Salaries and Wages from CSR reporting	Policy Water Efficiency	NOx Emissions
NOx Emissions	SOx Emissions	SOx Emissions	Women Employees	Policy Energy Efficiency	SOx Emissions
SOx Emissions	VOC Emissions	VOC Emissions	Women Managers	Environmental Materials Sourcing	Waste Total
VOC Emissions	Waste Recycling Ratio	Waste Recycling Ratio	Total Injury Rate Contractors	Renewable Energy Use Ratio	Waste Recycling Ratio
Waste Total	Hazardous Waste	Hazardous Waste	Total Injury Rate Employees	Energy Use Total	Hazardous Waste
Waste Recycling Ratio	Accidental Spills	Accidental Spills	Wages Working Condition Controversies	Water Withdrawal Total	Accidental Spills
Hazardous Waste	Policy Water Efficiency	Policy Water Efficiency	Diversity and Opportunity Controversies	Environmental Supply Chain Monitoring	Renewable Energy Use Ratio
e-Waste Reduction	Policy Sustainable Packaging	Policy Sustainable Packaging	Employees Health & Safety Controversies	Take-back and Recycling Initiatives	Energy Use Total
Water Pollutant Emissions	Toxic Chemicals Reduction	Toxic Chemicals Reduction	Voluntary Turnover of Employees	Product Environmental Responsible Use	Environmental Supply Chain Monitoring
Environmental Expenditures	Renewable Energy Use Ratio	Renewable Energy Use Ratio	Involuntary Turnover of Employees	Policy Employee Health & Safety	Environmental Controversies
Policy Water Efficiency	Energy Use Total	Energy Use Total	Policy Fair Competition	Policy Diversity and Opportunity	Eco-Design Products
Policy Energy Efficiency	Energy Produced Direct	Energy Produced Direct	Policy Business Ethics	Strikes	Fleet Fuel Consumption
Resource Reduction Targets	Renewable Energy Produced	Renewable Energy Produced	Employee Engagement Voluntary Work	Women Employees	Hybrid Vehicles
Targets Water Efficiency	Water Withdrawal Total	Water Withdrawal Total	Bribery, Corruption and Fraud Controversies	Women Managers	Fleet CO2 Emissions
Targets Energy Efficiency	Environmental Supply Chain Management	Environmental Supply Chain Management	Business Ethics Controversies	Total Injury Rate Contractors	Take-back and Recycling Initiatives
Renewable Energy Use Ratio	Environmental Controversies	Environmental Controversies	Policy Data Privacy	Total Injury Rate Employees	Product Environmental Responsible Use
Renewable Energy Supply	Eco-Design Products	Eco-Design Products	Policy Responsible Marketing	Employees Health & Safety Controversies	Health & Safety Policy
Energy Use Total	Fleet Fuel Consumption	Fleet Fuel Consumption	Product Responsibility Monitoring	Policy Child Labor	Policy Employee Health & Safety
Green Buildings	Product Impact Minimization	Product Impact Minimization	Product Quality Controversies	Policy Forced Labor	Employees Health & Safety OHSAS 18001
Water Withdrawal Total	Take-back and Recycling Initiatives	Take-back and Recycling Initiatives	Responsible Marketing Controversies	Policy Human Rights	Salaries and Wages from CSR reporting
Water Recycled	Product Environmental Responsible Use	Product Environmental Responsible Use	Controversies Customer Health & Safety	Human Rights Controversies	Trade Union Representation
Land Environmental Impact Reduction	GMO Products	GMO Products	Controversies Privacy	Child Labor Controversies	Strikes
Environmental Controversies	Renewable/Clean Energy Products	Renewable/Clean Energy Products		Employee Engagement Voluntary Work	Total Injury Rate Total
Equator Principles	Revenue from Environmental Products	Revenue from Environmental Products		Intellectual Property Controversies	Total Injury Rate Contractors
Environmental Project Financing	Health & Safety Policy	Health & Safety Policy		Crisis Management Systems	Total Injury Rate Employees
Nuclear	Policy Employee Health & Safety	Policy Employee Health & Safety		Anti-competition Controversies	Employee Fatalities
Sustainable Building Products	Policy Supply Chain Health & Safety	Policy Supply Chain Health & Safety		Policy Data Privacy	Contractor Fatalities
Real Estate Sustainability Certifications	Total Injury Rate Total	Total Injury Rate Total		Product Responsibility Monitoring	Lost Days To Total Days
Policy Employee Health & Safety	Total Injury Rate Contractors	Total Injury Rate Contractors		Customer Satisfaction	Wages Working Condition Controversies
Trade Union Representation	Total Injury Rate Employees	Total Injury Rate Employees		Controversies Privacy	Employees Health & Safety Controversies
Strikes	Accidents Total	Accidents Total			Voluntary Turnover of Employees
Total Injury Rate Total	Contractor Accidents	Contractor Accidents			Involuntary Turnover of Employees
Total Injury Rate Contractors	Employee Accidents	Employee Accidents			Bribery, Corruption and Fraud Controversies
Total Injury Rate Employees	Employee Fatalities	Employee Fatalities			Anti-competition Controversies
Employee Fatalities	Contractor Fatalities	Contractor Fatalities			Public Health Controversies
Contractor Fatalities	Employees Health & Safety Controversies	Employees Health & Safety Controversies			Business Ethics Controversies
Employees Health & Safety Controversies	Policy Fair Competition	Policy Fair Competition			Critical Countries Controversies
Policy Fair Competition	Policy Bribery and Corruption	Policy Bribery and Corruption			Consumer Complaints Controversies
Policy Bribery and Corruption	Policy Business Ethics	Policy Business Ethics			Product Quality Controversies
Bribery, Corruption and Fraud Controversies	Policy Community Involvement	Policy Community Involvement			Controversies Customer Health & Safety
Crisis Management Systems	Political Contributions	Political Contributions			Product Recall
Anti-competition Controversies	Bribery, Corruption and Fraud Controversies	Bribery, Corruption and Fraud Controversies			
Public Health Controversies	Anti-competition Controversies	Anti-competition Controversies			
Critical Countries Controversies	Public Health Controversies	Public Health Controversies			
Policy Customer Health & Safety	Critical Countries Controversies	Critical Countries Controversies			
Quality Mgt Systems	Policy Data Privacy	Policy Data Privacy			
Product Access Low Price	Product Quality Controversies	Product Quality Controversies			
Consumer Complaints Controversies	Controversies Customer Health & Safety	Controversies Customer Health & Safety			
Product Quality Controversies	Controversies Privacy	Controversies Privacy			
Controversies Customer Health & Safety	Product Recall	Product Recall			
Controversies Product Access					
Product Delays					

Number of material indicators used per sector

SICS Sector	Unique data items used per sector	Number of industries within the sector	Average number of data items used per industry within the sector
Consumer Goods	42	7	11
Extractives & Minerals Processing	61	8	27
Financials	27	7	10
Food & Beverage	50	8	18
Health Care	45	6	15
Infrastructure	58	8	15
Renewable Resources & Alternative Energy	56	6	16
Resource Transformation	56	5	19
Services	28	7	9
Technology & Communications	36	6	14
Transportation	47	9	15

Distribution of material ESG scores by year



Number of material indicators used per industry

SICS Industry	Number of material indicators used
Advertising & Marketing	8
Aerospace & Defense	17
Agricultural Products	29
Air Freight & Logistics	16
Airlines	12
Alcoholic Beverages	12
Apparel, Accessories & Footwear	12
Appliance Manufacturing	7
Asset Management & Custody Activities	15
Auto Parts	11
Automobiles	9
Biotechnology & Pharmaceuticals	19
Building Products & Furnishings	13
Chemicals	30
Commercial Banks	11
Construction Materials	27
Consumer Finance	5
Containers & Packaging	24
Cruise Lines	19
E-commerce	16
Electric Utilities & Power Generators	24
Electrical & Electronic Equipment	16
Engineering & Construction Services	19
Food Retailers & Distributors	21
Gas Utilities & Distributors	7
Hardware	14
Health Care Delivery	24
Health Care Distributors	12
Home Builders	10
Hotels & Lodging	11
Household & Personal Products	6
Industrial Machinery & Goods	9
Insurance	9
Internet Media & Services	14
Investment Banking & Brokerage	15
Iron & Steel Producers	22
Leisure Facilities	7
Marine Transportation	18
Media & Entertainment	8
Medical Equipment & Supplies	12
Metals & Mining	36
Multiline and Specialty Retailers & Distributors	15
Non-Alcoholic Beverages	20
Oil & Gas – Exploration & Production	38
Oil & Gas – Midstream	16
Oil & Gas – Refining & Marketing	23
Oil & Gas – Services	29
Processed Foods	24
Professional & Commercial Services	11
Pulp & Paper Products	14
Rail Transportation	12
Real Estate	10
Restaurants	18
Road Transportation	17
Software & IT Services	16
Solar Technology & Project Developers	15
Telecommunication Services	10
Tobacco	3
Toys & Sporting Goods	11
Waste Management	22
Water Utilities & Services	13
Wind Technology & Project Developers	11

Diagnostic tests for equations (4) - (7) using levels in ESG scores

Restricted F-Test (Pooled OLS vs FE)

Model version tested	F- statistic	Prob. > F (p-value)	Reject H_0
Month-ahead returns and material ESG scores	2.64	0.0000	Yes
Month-ahead returns and TR ESG scores	2.63	0.0000	Yes
Year-ahead ROA and material ESG scores	139.43	0.0000	Yes
Year-ahead ROA and TR ESG scores	139.22	0.0000	Yes

Hausman test (FE vs RE)

Model tested	$\chi 2$ statistic	Prob. > $\chi 2$ (p-value)	Reject H_0
Month-ahead returns and material ESG scores	1053.23	0.0000	Yes
Month-ahead returns and TR ESG scores	1049.08	0.0000	Yes
Year-ahead ROA and material ESG scores	1309.56	0.0000	Yes
Year-ahead ROA and TR ESG scores	1101.29	0.0000	Yes

Joint F-test for time dummies

Model tested	F-statistic	Prob. > F (p-value)	Reject H_0
Month-ahead returns and material ESG scores	65.11	0.0000	Yes
Month-ahead returns and TR ESG scores	65.11	0.0000	Yes
Year-ahead ROA and material ESG scores	3.04	0.0000	Yes
Year-ahead ROA and TR ESG scores	3.03	0.0000	Yes

Modified Wald's test of groupwise heteroskedasticity

Model tested	$\chi 2$ statistic	Prob. > $\chi 2$ (p-value)	Reject H_0
Month-ahead returns and material ESG scores	30965.37	0.0000	Yes
Month-ahead returns and TR ESG scores	30969.99	0.0000	Yes
Year-ahead ROA and material ESG scores	8.3e+08	0.0000	Yes
Year-ahead ROA and TR ESG scores	8.4e+08	0.0000	Yes

Wooldridge test of AR(1) autocorrelation

Model tested	F-statistic	Prob. > F (p-value)	Reject H_0
Month-ahead returns and material ESG scores	164.737	0.0000	Yes
Month-ahead returns and TR ESG scores	164.751	0.0000	Yes
Year-ahead ROA and material ESG scores	3.699e+0.6	0.0000	Yes
Year-ahead ROA and TR ESG scores	3.711e+0.6	0.0000	Yes

Pesaran test for cross-sectional dependence (spatial correlation)

Model tested	CD statistic	p-value	Reject H_0
Month-ahead returns and material ESG scores	485.349	0.000	Yes
Month-ahead returns and TR ESG scores	485.336	0.000	Yes
Year-ahead ROA and material ESG scores	57.156	0.000	Yes
Year-ahead ROA and TR ESG scores	56.673	0.000	Yes