BI Norwegian Business School - campus Oslo

GRA 19703

Master Thesis

Thesis Master of Science

ESG and corporate financial performance: A study on differences across countries and industries in Europe

Navn: Sindre Aarak, Hans Wilhelm Werner

Start: 15.01.2021 09.00

Finish: 01.07.2021 12.00

ESG and corporate financial performance: A study on differences across countries and industries in Europe

Master Thesis

by
Sindre Aarak and Hans Wilhelm Werner
MSc in Finance

Supervisor:
Samuli Knüpfer

Oslo, June 30, 2021

ABSTRACT

We study the relationship between ESG performance and financial performance in Europe. We bring new insight to this field of research by dividing the sample into subsamples based on industry and geographic location to investigate differences in the relationship. We use panel data and a fixed effects model to answer our research question. Our results indicate that ESG increases financial performance measured by Tobin's Q. This increase is driven by the social ESG dimension. Further, ESG destroys financial performance measured by ROA. This decrease is driven by the governance ESG dimension. The relationship appears to be strongest among Nordic firms when financial performance is measured by ROA, Central European firms when measured by Tobin's Q and among firms that operate within manufacturing.

Acknowledgements

We would like to thank our supervisor, Professor Samuli Knüpfer for providing us with guidance and feedback throughout the semester. Second, we would Like to thank BI Norwegian Business School for keeping the campus open despite the ongoing pandemic. Lastly, we want to thank our friends and family for encouraging us and keeping us motivated.

Contents

List of Figures

List of Tables

1 Introduction and Motivation

Researchers have studied the possible existence of a relationship between Corporate Financial Performance (CFP) and Corporate Social Performance (CSP) since the 1970's. As our literature review will show, earlier research has provided mixed results on whether CFP and CSP are correlated at all. Furthermore, those that provide evidence of an existing relationship do not always agree on the direction of causality. In other words, there are still many unanswered questions as to how a company's CFP and CSP affect each other.

We believe that the transition from traditional Corporate Social Responsibility (CSR) to Environmental, Social and Governance (ESG) factors has made this relationship increasingly important over the last decade. Firstly, it would appear that investors demand sustainability. Morningstar's release of funds' ESG rating in 2016 resulted in significant outflows for funds with poor ESG rating and inflows for funds with good ESG rating. Secondly, global attention towards climate change increases the general population's concern for a sustainable future. Thirdly, the eruption of social media means that information travels faster than ever before. As a result, firms are vulnerable to bad publicity regarding their ESG performance which we believe could damage their financial performance.

While earlier research has found the relationship between CFP and CSP to be positive, negative, mutual and non-existent, we aim to provide additional value to this field of research by examining the following research question:

"Does the relationship between financial performance and ESG performance differ across industries and countries in Europe?"

This research question has the potential to bring new insight to the research community. We have yet to find a study that successfully identifies a relationship between ESG performance (ESGP) and financial performance (FINP) while simultaneously comparing the results of different industries and countries.

Previous literature has equipped us with insight into this field of research. Furthermore, it has provided us with guidance on how one should proceed when building on existing research. Our thesis is mainly built on the work of Velte (2017) and Ortas et al. (2015). These studies examine the relationship between ESG performance and financial performance measured by both Return on Assets (ROA) and Tobin's Q. We have used the same financial performance measurements and similar control variables to control for firm size, risk and expenditure towards R&D and advertising.

Our data set has been collected through Refinitiv and Bloomberg and contains 490 European firms. We have used a fixed effects model to conduct regression analysis on our data sample and a sample where we have lagged the ESG variable by one year. These regressions use Tobin's Q and ROA as dependent variables and ESG scores and control variables as explanatory variables. We also run separate regressions where we use both the total ESG score and individual ESG pillar scores. Firstly, this allows us to investigate whether a relationship between ESGP and FINP actually exists. Secondly, it allows us to see which of the ESG pillars that are most important in explaining FINP. Lastly, we divide our sample into subsamples categorized by industry and region. We run the same regressions to see if we can identify differences between industries and regions in Europe.

Our findings suggests that the relationship between ESGP and Tobin's Q is positive. This result reflects the notion that investors value sustainability which increases market-based returns. Conversely, we have found the relationship between ESGP and ROA to be negative. This result reflects the expenses associated with investments in ESG activities which lowers accounting-based returns. Moreover, we find that the social dimension of ESG is the main driver of the positive impact of ESG on market-based returns. For accounting-based returns, we find that the governance dimension is the main driver for the negative impact.

The results of our investigation into potential differences across European regions and industries vary in terms of statistical significance. However, we find evidence which suggests that the relationship is stronger in the Nordics when FINP is measured by ROA. In addition, we find evidence which suggests that the relationship is stronger in Central Europe when FINP is measured by Tobin's Q. Lastly, our results indicate that the relationship between ESGP and FINP is strongest for firms that operate within manufacturing. This result is consistent for both measures of FINP. These findings lay the groundwork for which future researchers can build on.

2 Literature Review

This chapter will present previous research that our thesis is meant to build on. Our selected studies focus on the traditional relationship between CSR and CFP, the direction of the causality in the relationship and the transition from traditional CSR to the more holistic ESG approach.

2.1 The Relationship Between CSP and CFP

Since the early 1970's, there has been made numerous attempts to establish a connection between CSR and CFP. Moskowitz (1972) suggested 14 companies as sound financial investments based on their CSP. He subsequently investigated the CFP of these companies, discovering that a portfolio of the 14 stocks would have outperformed the NYSE, Dow Jones and S&P in the following six months. Thus, ascertaining a positive correlation between CSP and CFP. Bragdon and Marlin (1972) were the first to perform an empirical test on the relationship between CSP and CFP. Their study examined the relationship between growth in earnings and measures taken to reduce pollution for 17 firms in the paper and pulp industry. They concluded that some degree of pollution control is likely to increase profits.

Vance (1975) critiqued the aforementioned paper by Moskowitz by investigating 45 firms whose corporate CSP had been ranked by both businessmen and graduate students. He concluded that a portfolio consisting of the 22 stocks with the highest ranked CSP would severely underperform relative to a portfolio made up of the 22 stocks with the lowest ranked CSP. Thus, there are conflicting opinions on whether a firm can bring value to its shareholders through CSP.

2.2 A Discussion of Causality

From the discussion about a relationship between CSP and CFP, a subsequent discussion about the direction of causality emerged. Waddock and Graves (1997) investigated whether CSP causes CFP or vice versa. By using different measures of CFP and ratings on CSP, they found a positive relationship between CSP factors and CFP measures both as the dependent and independent variable. Thus, they suggested that the two factors influence each other simultaneously and that the causality runs in both directions.

Scholtens (2008) made a more recent attempt to determine the relationship between CSP and CFP. He voiced concern over a lack of attention to the causality issue. Furthermore, he used OLS with distributed lags and Granger causation to determine whether social or financial performance precedes the other. His results suggested a positive relationship between CSP and CFP. However, as opposed to the findings of Waddock & Graves, Scholtens found that financial performance precedes social performance in most cases.

Early research on the relationship between social and financial performance has frustratingly yielded mixed results. Orlitzky et al. (2003) conducted a meta-analysis of 52 studies on the relationship between CSP and CFP. Firstly, their aim was to investigate the hypothesis that the relationship between CSP and CFP is positive in general. Secondly, they hypothesized that the relationship is bidirectional. Lastly, they investigated the reason for the inconsistencies in results. They found that CSP is positively related to CFP across studies and that the relationship tends to be bidirectional and simultaneous. Moreover, they concluded that between 15 and 100 percent of correlation variations can be explained by sampling and measurement errors.

Margolis et al. (2009) examined studies from the time period where Orlitzky et al. (2003) ended. Their aim was to determine if the CSP-CFP relationship had strengthened over time in a similar meta-analysis. They found a small but significant positive relationship between social and financial performance. In addition, they introduced a set of criteria that future research should meet, including the use of verifiable third party CSP data and control variables. Control variables should control for size, risk, industry, R&D and advertising as a minimum. Furthermore, they emphasized that research need to consider theoretically meaningful time periods.

Hong and Kacperczyk (2009) studied the effects of social norms on market outcomes. They hypothesized that institutional investors which are subject to norms and scrutiny from their investors, are paying a financial cost of abstaining from certain industries considered as sinful. The authors used the CAPM and a four-factor model with sin stocks' excess return as the dependent variable. Their results showed that the portfolio of sin stocks significantly outperformed a comparable portfolio of stocks. Furthermore, they showed that sin stocks have a higher cost of capital as a result of the risk of legal action and social norms which prevents institutional investors from investing in sin stocks.

Lobe and Walkshaeusl (2016) generated portfolios of global, regional and domestic stocks deemed to be sinful in a similar effort to examine the performance of sin stocks. As opposed to Hong and Kacperczyk (2009), they found no evidence that portfolios consisting of sin stocks outperform or underperform the market. In addition, they concluded that a strategy of going long in the sin stocks portfolio and shorting a socially responsible portfolio did not outperform the market.

2.3 The Shift from CSR to ESG

During the 1980s we saw the start of ESG ratings as a way for investors to evaluate corporations on other dimensions than financial performance, such as social and environmental efforts (Berg et al., 2020). Several third-party entities have emerged with the purpose of providing investors with reliable data on companies' non-financial performance. Even though these third-party entities contribute to making sustainable investing more available for the common investor, ESG ratings are still in their early stages. Berg et al. (2020) identified significant discrepancies between ratings given by the different rating agencies. They point towards measurement divergence, scope divergence and weighting divergence as the underlying reason.

Ortas et al. (2015) investigated the effects of adopting the principles put forward by UNGC. They assessed how the adoption of these principles affected the companies' ESGP and subsequently their FINP. Their research was performed in Japan, Spain and France as these countries held the highest number of businesses which adopted the UNGC principles. FINP was measured by the market-based measure Tobin's Q and the accounting-based measure ROA. Their model included control variables to control for Size, Risk, Industry and R&D. Additionally, the authors investigated the individual effects and relationship of environmental, social and governance performance. They found that the three dimensions were strongly, positively correlated with each other while only the environmental dimension had a positive relationship with FINP (Ortas et al., 2015).

Velte (2017) performed a similar analysis of the relationship between German companies' ESGP and FINP. His study used the same types of data and similar methodology as Ortas et al. (2015). In addition, he controlled for the same variables in his model. To capture a potential time delay in the effect ESGP has on FINP, he lagged the ESG variables in the regression model. The results on the other hand, differed significantly from the research done in Japan, France and Spain. He found that ESGP was positively related to ROA, while there was no significant relationship with Tobin's Q. However, he found that the governance dimension had the strongest effect on FINP. This contrasts the results of Ortas et al. (2015).

In 2016, Morningstar published their first ratings on mutual funds' sustainability performance. The funds were given a rating on a scale of 1 to 5 stars based on their holdings. Hartzmark and Sussman (2019) used these ratings to investigate whether investors value sustainability. These ratings presented no new information. However, it made it more available for investors. The authors used the Morningstar ratings to investigate whether the funds experienced net positive inflow after receiving their rating. Their findings show that low-ranked funds experienced an outflow of $12 billion dollars whereas being ranked high on sustainability led to net inflows of more than $24 billion. This reflects the demand for sustainability among investors. However, they found no evidence of the funds ranked high on sustainability outperforming lower ranked funds.

Based on earlier research, we believe that sustainability is a factor which is highly valued by investors in today's investment community. Therefore, we expect to find a positive relationship between ESGP and FINP.

3 Theory

There are two theories that are useful in explaining the relationship between ESGP and FINP. The shareholder theory and stakeholder theory argues whether the relationship is negative or positive, respectively. Once a relationship between ESGP and FINP is established, the direction of causality can be explained by the slack resources theory, the good management theory and the virtous cycle theory.

3.1 The Shareholder Theory

Friedman (1970) argued that the only objective a firm has is to maximize profits for its shareholders while conforming to the law and following ethical guidelines. Furthermore, he stated that only individuals are capable of having responsibilities. Therefore, a manager can feel a responsibility to make social contributions, but it should be with their own money, time and effort and not the shareholders' money. According to Friedman, a manager who makes expenditure towards mitigating pollution beyond what is required by the law is in fact using the shareholders' money to do so. Consequently, the manager is effectively imposing taxes on the shareholders and also deciding how these taxes should be spent. Friedman argues that this is the role of the government and not something that an agent who acts on the behalf of a principle should concern themself with. If the shareholders wish to spend their money in a way that benefits society, they should make that decision themselves (Friedman, 1970).

Friedman's theory has been subject to critique because shareholders cannot be certain that the manager is working in their best interest. Managers may very well use the money saved from not making investments for the benefit of society to pursue private benefits. While this is true, the argument completely ignores the last part of Friedman's statement: Firms should maximize profits "Without deception or fraud". Smith (2003) defends Freidman's view by using earlier studies to argue that CSR can be beneficial because it can increase firm value which will benefit shareholders. He argued that the shareholder theory would support CSR in these cases. Though he supported Friedman, he admitted that the shareholder theory is based on an unrealistic model. It attempts to separate business from society which is not feasible as these two are so intertwined (Smith, 2003).

3.2 The Stakeholder Theory

Freeman and Mcvea (2001) argued that the current theories were inconsistent with both the quantity and the type of change that the business environment was experiencing. Therefore, the stakeholder theory was introduced in an attempt to counter these challenges. Stakeholders are defined as any group or individual that are affected by or can affect the achievement of an organization's objectives (Freeman and Mcvea, 2001). The stakeholder theory says that managers need to understand and consider the concerns of all parties affected by the company's operations when determining the firm's objective. Thus, the firm will gain the support of its stakeholders which is imperative to gain long-term success. Therefore, managers should seek to explore their relationship with the firm's stakeholders when developing business strategies (Freeman and Mcvea, 2001).

The stakeholder theory has been subject to criticism just like Friedman's shareholder theory. Ambler and Wilson (2006) listed a variety of downsides associated with the stakeholder theory. First, they question whether it is possible to align the interests of all stakeholders. Second, they suggested that different stakeholders will have different opinions on how to measure a firm's success.

They argued that success must be related to the purpose of the firm and that different stakeholders will have different views on what the purpose of the firm truly is. As a result, judging the management's performance is problematic because it is not clear how the success should be measured (Ambler and Wilson, 2006).

3.3 The Good Management Theory

This theory draws parallels to the stakeholder theory by arguing that good CSP will enhance CFP because it strengthens the firm's relationship with its stakeholders. For instance, the theory suggests that good employee relations will boost morale and efficiency which in turn will provide better CFP. In addition, the general population's concern about the environmental impact of industrialization is constantly growing. Therefore, being perceived as a green company that cares about the environment is likely to attract competent employees and more customers as people generally want to be associated with doing good (Prahalad and Hamel, 1994).

3.4 The Slack Resources Theory

The slack resource theory argues that high CFP will lead to high CSP because firms that do well financially will have more money to spend on social investments. This argument is logical as firms with poor financials are more likely to prioritize investments that would strictly benefit the firm. Even though research into the relationship between CSP and CFP has been ongoing for 50 years, most of the research is based on the good management theory (Melo, 2012). One possible reason for why the slack resource theory has been somewhat neglected is that those who believe in a positive correlation between CFP and CSP are usually supporters of the stakeholder theory. As the good management theory is more linked towards the stakeholder theory, it could be that researchers are more eager to investigate how CSP affect CFP. However, this is not to say that there is no evidence that supports the slack resource theory. As mentioned in the literature review, Waddock and Graves (1997) presented evidence which suggested that past CFP is more closely related to CSP than subsequent CFP.

3.5 The Virtuous Cycle Theory

Waddock and Graves (1997) argued that past CFP affects CSP at the same time as CSP has an effect on subsequent CFP. All though it is unclear where the circle starts, the authors offered an interesting theory. They argued that the positive relationship could represent an initial ulterior motive in the management's behavior. Their efforts to improve CSP could be a way to boost employee morale, obtain good publicity or improve their relationship with the local community because they realize that this would allow them to reap financial benefits. Even though the managers' actions are based on the "wrong" reasons, Waddock and Graves argue that firms will eventually adapt their business culture to incorporate CSP expenditure because it serves them financially. Thus, the cycle begins.

This thesis is meant to build on the research of Velte (2017) and Ortas et al. (2015). Both papers found support for the good management theory. Thus, we will investigate whether or ESGP has any effect on FINP. However, being aware of the existence of other theories will be useful in the discussion of our results.

4 Testable Hypothesis

To answer our research question *"Does the relationship between financial performance and ESG performance differ across industries and countries in Europe?"*, we will begin by testing whether ESGP can be useful in explaining FINP at all. We will do this by examining the following relationship between ESGP and FINP

$$ESG_t \rightarrow ROA_t$$

$$ESG_t \rightarrow Tobins'Q_t$$

$$H_0 : \ There\ is\ no\ relationship$$

$$H_A : \ There\ is\ a\ relationship$$

A positive relationship would support the stakeholder theory while a negative relationship would support the shareholder theory. As described in previous literature, it will most likely take some time before the effects of ESG investments are reflected in FINP (Velte, 2017). Therefore, we will proceed by testing the same relationship, but with a lagged version of the ESG variable. This will allow us to test whether past ESGP can be useful in explaining subsequent FINP.

$$ESG_{t-1} \rightarrow ROA_t$$

$$ESG_{t-1} \rightarrow Tobins'Q_t$$

$$H_0 : \ There\ is\ no\ relationship$$

$$H_A : \ There\ is\ a\ relationship$$

A positive relationship would indicate that our analysis supports the good management theory. By comparing the lagged and unlagged model, we will decide which model to use as we proceed. Further, we want to see if the effect on FINP from ESGP originates from good performance within the environmental, social or governance dimension. By decomposing the companies' ESG score into individual pillars for each dimension, we can examine the individual effect they have on FINP.

$$H_0: \; No\ Difference\ in\ effect\ of\ pillar\ scores$$

$$H_A: \; There\ is\ a\ difference\ in\ effect\ of\ pillar\ scores$$

We will conclude on this test by examining the impact of each pillar measured by the coefficient and the level of significance measured by each pillar's p-value. At this point, we should have established whether the relationship exists and if the effects of ESG investments can be identified immediately or not. Furthermore, we should have established which ESG dimension drives the relationship.

Our contribution to this field of research consists of identifying differences in the relationship between ESGP and FINP among different industries and regions in Europe. We believe that the mixed results yielded by previous literature could be explained by sample selection. Most of the previous literature investigated the ESGP-FINP relationship in different countries and time periods. By sorting our main sample into subsamples based on geographic location and industry, we intend to provide new insight on the ESGP-FINP relationship. We will sort the companies in our sample into three industries and three different regions. Once this is done, we will repeat the steps in the analysis above to see how the result from the subsamples compares to those of the total sample.

5 Methodology

In the following section we will discuss the structure of our data sample and the model selection process to define how to best answer our research question. The selected model will be used in all regressions. Furthermore, we will comment on the validity of our selected model.

5.1 Panel Data

Our data sample consists of observations spanning over 10 years for 490 companies. The data is structured as an unbalanced panel, which is a result of missing data throughout our sample period. There are multiple advantages to the panel data structure compared to pure cross-sectional or time-series data. First and foremost, panel data considers the possibility of individual heterogeneity. Moreover, we can control for time and state invariant variables that may affect our dependent variable (Baltagi, 2008). Thus, by structuring our model correctly, we can control for unobservable variables that could affect the FINP and ESGP of the companies in our dataset. Second, to assess the dynamic relationship between variables over time, pure time-series data would require a data set with many time-series observations. ESG data is still in its infancy and by using data for multiple firms we can increase the number of observations in our sample to increase the power of the tests we perform (Brooks, 2014).

5.2 Model Choice

When faced with panel data, there are multiple models that can be applied. To take full advantage of the information contained in panel data, one could perform a seemingly unrelated regression (SUR) framework. However, this model demands that the time series observations, T, per cross-sectional unit, i is at least as large as the number of such units. Moreover, the SUR framework requires estimation of many parameters in addition to the variance-covariance matrix of the errors (Brooks, 2014).

More flexible alternatives can be found in the pooled OLS, fixed effects and random effects model. Velte (2017) used a fixed effects model to investigate the FINP-ESGP relationship in German firms. Ortas et al. (2015) employed several time random effects models to examine the relationship in companies committed to the UNGC principles. To determine which model is better for our data we will run several model specifications tests. Firstly, we will run an individual effects test which will determine if a fixed effects model is preferable to the pooled model. Similarly, the Breusch-Pagan test will determine the presence of random effects. Finally, we will conduct a Hausmann-test to specify whether the fixed or random effects model is preferable.

5.2.1 Pooled Model

The easiest way to deal with panel data is to perform a pooled regression on the data. Pooling the data involves estimating a single equation for the entire data set. In practice, this means stacking all the cross-sectional and time-series data into a single column for the dependent variable. Similarly, the regressors would be stacked in a single column for each independent variable. Then we would estimate the equation using OLS (Brooks, 2014). In this thesis, the equation of the unlagged pooled model is represented by:

$$FINP_{i,t} = \alpha_{i,t} + \beta_1 \, \mathrm{ESG}_{i,t} + \beta_2 \, \mathrm{Size}_{i,t} + \beta_3 \, \mathrm{Risk}_{i,t} + \beta_4 \, \mathrm{RD}_{i,t} + \beta_5 \, \mathrm{AD}_{i,t} + u_{i,t}$$

where $i = 1, ..., 490$ and t = 2010, ..., 2019

Even though this method is desirable in its simplicity and requires estimation of few variables, it has some severe limitations. Most notably, the Pooled OLS assumes that the average values of the variables and the relationship between them is constant over time and across the cross-sectional units in the sample (Brooks, 2014). This assumption implies that we should find the same average values for all variables for all the firms in our sample. Moreover, we should find that these variables affect each other in the same way across entities. For a broad sample of firms across countries and industries, this assumption is unlikely to hold.

5.2.2 Fixed Effects Model

The fixed effects model takes into consideration the individual heterogeneity across entities in the panel data. This is done by breaking the error term, u_{it} into two parts. One captures the time-invariant effects and the other captures the remainder of the unexplained variation in our dependent variable (Brooks, 2014).

By including a dummy variable for all firms, we can capture the individual effect of each firm that does not vary over time. This model is termed the Least Squares Dummy variable (LSDV) model. A concern with LSDV is the large number of variables that needs to be estimated. If the number of entities, N is large the model needs (N-1) dummy variables to capture the time-invariant effects in the data. Baltagi (2008) suggests transforming the data by subtracting the time-mean of each entity from the variables. One would proceed by running a regression on the time-demeaned values. This transformation of the data is known as the within transformation. The equations to be estimated using this model is:

$$\ddot{FINP}_{i,t} = \beta_1\,\ddot{\mathrm{ESG}}_{i,t} + \beta_2\,\ddot{\mathrm{Size}}_{i,t} + \beta_3\,\ddot{\mathrm{Risk}}_{i,t} + \beta_4\,\ddot{\mathrm{RD}}_{i,t} + \beta_5\,\ddot{\mathrm{AD}}_{i,t} + \ddot{u}_{i,t}$$

where $i = 1, ..., 490$ and $t = 2010, ..., 2019$. The double dots above the variables denotes the demeaned values.

The drawback of the within transformation is that we will not be able to retrieve the values of the individual effects, u_i. This value captures time-invariant variables often ascribed to managerial or entrepreneurial skills of the firms' executives. However, these individual effects are not related to the relationship between ESGP and FINP. Therefore, individual effects will not be necessary to investigate further and the within transformation is a viable option.

5.2.3 Random Effects Model

The random effects model proposes different intercept terms for each entity, like the fixed effects model. It suggests an intercept a that is common across cross sectional entities. The individual effects originate from a random variable, ϵ_i which varies cross sectionally but not over time. ϵ_i represents the random deviation from the global intercept found in each entity's intercept Brooks (2014). The random effects equation is constructed as follows:

$$FINP_{i,t} = \beta_1\,\mathrm{ESG}_{i,t} + \beta_2\,\mathrm{Size}_{i,t} + \beta_3\,\mathrm{Risk}_{i,t} + \beta_4\,\mathrm{RD}_{i,t} + \beta_5\,\mathrm{AD}_{i,t} + W_{i,t}$$

where $W_{i,t} = \epsilon_i + v_{i,t}$.

In general, one would prefer the random effects model if the data were randomly selected from a given population (Wooldridge, 2010). However, our sample is based on a set of exclusion criteria such as data availability and geographical location of the firms. As such, we cannot uphold that our data sample has been selected randomly.

5.3 Model Specification Tests

To determine which model is most suitable for our data, we have performed a set of model specification tests. First, we have performed a test for individual effects to see if the cross-sectional entities in our data contain individual effects that need be considered. Second, we performed a Breusch-pagan Lagrange multiplier test to investigate the variance of the individual effects. This is necessary to determine whether the individual effects are random. If we find individual effects in our data, a Hausmann test can be utilized to determine which of the fixed- or random effects model is more suited to our data.

5.3.1 Test for Individual Effects

Our test for individual effects is consistent with Baltagi (2008). In practice, this involves performing a Chow F test on the individual effects, $\mathrm{u_i}$. The unrestricted model follows the fixed effects estimation while the restricted model follows an OLS pooling estimation. This test has the following null and alternative hypothesis:

$$H_0: \ u_i = 0$$

$$H_0: \ u_i \neq =0$$

If the null hypothesis is rejected, the individual effects that are present in the data are statistically different from zero. Thus, we should employ a fixed effects model over a pooled OLS estimation if H0 is rejected.

5.3.2 Breusch-Pagan Test

The Breusch-Pagan LaGrange multiplier test is used to test whether the variance of the individual effects in the data, $\sigma^2_{u_i}$ is statistically different from zero. This test has the following null and alternative hypothesis:

$$H_0 : \ \sigma^2_{u_i} = 0$$

$$H_0 : \ \sigma^2_{u_i} \neq =0$$

If the null hypothesis is rejected, the variance of the individual effects is statistically different from zero. We should employ a random effects model over a pooled OLS model in this case (Breusch and Pagan, 1980).

5.3.3 Hausman Test

In order to decide whether fixed- or random effects are most notable in our data, we perform a Hausman test (Hausman, 1978). This test examines the differences of the estimators, β_{fe} and β_{re} in the fixed and random effects model, respectively. The null and alternative hypothesis is as follows:

$$H_0 : \ \beta_{fe} - \beta_{re} = 0$$

$$H_0 : \ \beta_{fe} - \beta_{re} \neq =0$$

Both the random and fixed effects models are consistent under the null hypothesis. However, only the fixed effects model is consistent if the null hypothesis is rejected (Hausman, 1978).

5.4 Validity

The results from our model specification tests show that a fixed effects model is the best fit for our data. In the following section, we will discuss the underlying assumptions of the fixed effects model and the actions we have taken to secure the validity of our results.

5.4.1 Selection Bias

Selection bias refers to a situation where we have a restricted sample that is not representative for the entire population. In general, when the sample is gathered based on simple random sampling, selection bias will not be a prevalent issue (Wooldridge, 2010). The fixed effects model assumes that the variables are independently and identically distributed across entities. Thus, the fixed effects model is in line with the required standard in relation to selection bias by assumption. Our sample is not collected using random sampling. It is restricted based on several criteria that limits the selection of companies.

Our thesis examines the relationship between FINP and ESGP across Europe. While our sample does not include companies for all European countries, data has been collected from countries included in the MSCI Europe index. This index should be sufficient in representing Europe as a whole. A larger concern is data availability for our variables. ESG scores are largely based on self-reporting by the companies. Consequently, firms can inflate their ratings by reporting statistics where their performance is strong. As a result, our sample could be skewed in the direction of companies with high ESG ratings, whereas companies that perform poorly on ESG criteria are not included.

5.4.2 Omitted Variable Bias

Omitted variable bias becomes an issue when a variable that is useful in explaining the dependent variable has been left out of the regression. This could lead to biased and inconsistent coefficients of the model's explanatory variables. As a result, wrong inferences could be drawn from the regression (Brooks, 2014). One of the assumptions underlying the FE-model states that the expectation of the error term, conditional on the regressors and the time-invariant effects is zero Wooldridge (2010). If the error term is not zero and correlated with the independent and dependent variables, we would get biased and inconsistent estimators. We have chosen our independent variables based on recommendations from previous literature. Therefore, we do not suspect that our model suffers from omitted variable bias.

5.4.3 Multicollinearity

A third assumptions related to the fixed effects model is that there should be no perfect multicollinearity. In a multivariate regression model, there is expected to be some degree of correlation between the regressors. However, a problem occurs when there is an exact relationship between one or more of the variables. In this case, we would not be able to estimate all the coefficients. This specific situation is called perfect multicollinearity. A more common problem is what is called near perfect multicollinearity. This refers to a situation where one or more variables have non-perfect but non-negligible relationships. If near perfect multicollinearity is present but ignored, we would expect to see a high R^2 for the model, while the significance for the individual coefficients would be low. Moreover, standard errors for the coefficients would be high causing significance testing to yield wrong results (Brooks, 2014).

Formally, multicollinearity can be difficult to measure. Through examination of the correlation matrix, we can identify potential strong correlations between our variables. If high correlations are present, we can investigate the issue further by calculating their Variance Inflation Factor (VIF). This factor measures the effect that the independent variables have on each other's variation. VIF is calculated by running a regression with one of the independent variables as the dependent variable against the remainder of the independent variables. This procedure is repeated for all independent variables. Common cut off points for the presence of multicollinearity is a VIF factor of 5 or 10 (Chase, 2013)

5.4.4 Stationarity of Idiosyncratic Errors

One additional assumption that must be satisfied for the fixed effects model to be efficient is that the idiosyncratic errors are stationary. This means that its expected value and variance remain constant over time (Wooldridge, 2010). Stationarity of the idiosyncratic errors is visually examined by plotting the residuals against an independent variable. This assumption translates roughly into an expectation of few large outliers that would distort the distribution of the errors. Wooldridge (2010) provides the same definition of large outliers.

5.4.5 Serial Correlation of Idiosyncratic Errors

Serial correlation or autocorrelation refers to a situation where a variable is dependent on a lagged version of itself. The result of ignoring serial correlation, if present is that the coefficient estimates are no longer BLUE. The estimators will still be unbiased. However, they will no longer be efficient and standard errors may be wrong. As a consequence, we could make wrongful inferences from the regression results (Brooks, 2014). Using the within estimator, Wooldridge (2010) discussed situations where negative serial correlation of the idiosyncratic errors is expected. Regardless, the presence and type of serial correlation needs to be tested. We have performed the Wooldridge test for serial correlation to determine if serial correlation is present.

5.4.6 Measurement Error

Measurement errors can occur in a variety of ways. Most macroeconomic data is based on estimation and is therefore prone to errors compared to real values. Another source could be wrongful data input when handling data which is meant to be used in a model. Measurement error in the explained variable is captured in the disturbance term and should not cause any concerns. More troubling is measurement errors in the explanatory variables which could cause biased estimates (Brooks, 2014). To our knowledge, there has been no mistreatment of our collected data in the process of sample building. The data sample has been collected from the Thomson Reuters database through Eikon Refinitiv and the Bloomberg database. Furthermore, our model does not include any macroeconomic data. Therefore, we are unlikely to find any errors in these variables.

ESG scores are calculated based on data reported from the companies themselves and may be disposed to measurement error. Kotsantonis and Serafeim (2019) discussed the reliability of ESG data and suggested that variations in measurement, computation and model input methods could distort our understanding of ESG scores. Fischer and Sawczyn (2013) pointed to this problem in their investigation of the relationship between CSP and CFP. Refinitiv's process of collecting data on companies' ESG performance is incentivizing correct reporting of ESG data. Lack of reported data or transparency would lead to a lower ESG rating. The process of developing the final ESG scores is described further in the data chapter. Nevertheless, measurement error in the ESG scores could potentially be a threat to the overall validity of our results. As Refinitiv's ESG scores are only an indicator of companies' performance, our results could differ from research that applies other measures of ESG performance.

5.4.7 Simultaneous Causality

Simultaneous causality is a relevant issue when examining the relationship between ESGP and FINP. The problem occurs when the explained variable has an effect on one or more of the explanatory variables. If simultaneous causality is present and ignored, it would lead to biased and inconsistent results (Brooks, 2014). A previously mentioned theory could explain a simultaneous causality issue in our model. The virtuous cycle theory suggests a continuous cycle of causation between the two performance measures (Waddock and Graves, 1997). Examining the effect of ESGP on FINP, this simultaneous causality could pose a threat to the validity of our results. To account for this potential threat, we follow Velte (2017)and Ortas et al. (2015) by considering a model with a time lagged independent variable. This is to account for the possibility that ESG performance will first influence FINP in the subsequent period, and to counter the causality problem.

6 Data

In this chapter, we will begin by explaining the process of firm selection in our thesis. Secondly, we will explain the variables that have been used in our model to answer our research question.

6.1 Data Sources

We collected company data from 2010-2019 for the 15 countries that are part of the MSCI Europe Index. Our initial sample was collected through the Refinitiv database. We gathered data on ESG scores, financial performance and control variables for all publicly traded firms in the 15 countries. Companies in the financial sector were excluded in line with previous research. Velte (2017) suggested that firms in the financial sector should be excluded due to their specific regulations in comparison to other sectors and companies. After adjusting for dual-listed firms and removing companies in the financial sector, we were left with a total of 3,982 firms. However, Refinitiv did not provide us with the necessary data for all those companies. Firms without ESG data for the entire 10-year period were excluded immediately as they would not be useful in explaining the relationship between ESGP and FINP.

Once our sample exclusively contained firms with available ESG data, we discovered that we lacked data on financial performance and control variables for a lot of the remaining firms. We were able to obtain this missing data through Bloomberg. As a result, the number of firms in our sample were decided based on the companies for which Refinitiv had ESG data. We then had to match this data with the financial data for the same companies collected through Bloomberg. As a result, the sample size decreased further as some of the firms that had ESG data lacked financial data. Once the data from Refinitiv and Bloomberg had been successfully matched, we had our final sample which consists of 490 firms.

6.2 ESG Data

We have used the ESG scores provided by Refinitiv as they have one of the most comprehensive ESG databases in the world. Their database contains ESG data for more than 70% of the global market cap. More than 500 ESG data points are collected from annual reports, company and non-governmental organization websites, stock exchange filings, CSR reports and news sources. This data is then analyzed by a team of more than 150 research analysts before a company is rated and the rating is added to the ESG database. Furthermore, their ESG ratings are updated weekly to ensure that the scores are as accurate as possible. The universe for which ESG data is available in Refinitiv's database consists of approximately 9,000 companies, 2,100 of which are located in Europe (Refinitiv, 2021).

In addition to having a highly comprehensive database, Refinitiv's methodology for estimation of their ESG scores is available to the public. Prior research into the effects of ESG investments and performance identifies inconsistencies in the way different companies report ESG data as an issue (Kotsantonis and Serafeim, 2019). The reason is that there is no standardized way of reporting. Therefore, firms can choose to report their ESG performance only on aspects where they are satisfied with their own performance. However, Refinitiv's ESG scores penalizes companies for not reporting on certain ESG data points. The data points are weighted differently such that the impact of the penalty for not reporting on data points that are highly weighted is larger than for a data point that is lower weighted (Refinitiv, 2021). All these data points can be assigned to ten categories which are used to determine a company's score on each ESG pillar. The categories within each pillar are summarized in the table below:

Table 1: Pillar composition

Environmental	Social	Governance
Resource Use	Workforce	Management
Emissions	Human Rights	Shareholders
Innovation	Community	CSR Strategy
	Product Responsibility	

Table 1 illustrates the composition of Refinitiv's ESG pillar scores. The ten categories are used to determine a score for each ESG pillar and ultimately, the total ESG score.

The pillar scores are determined by a given firm's score on each of the ten categories. Scores for each category are calculated in the following way:

$$Score = \frac{\text{Number of companies with lower value} + \frac{\text{Number of Companies with same value}}{2}}{\text{Number of companies with a value}}$$

The weights of each category vary between industries with some exceptions. For instance, the community category is weighted equally across all industries as it is equally important to all industries (Refinitiv, 2021). The magnitude score of each category is summed up and the weights used to determine the pillar score are estimated as follows:

$$Score = \frac{\text{Magnitude weight of a category}}{\text{Sum of magnitudes of all categories}}$$

To obtain the pillar scores, each category score for a firm within a certain industry is multiplied by the corresponding category weight for that same industry. The results are then added together such that the score of a certain pillar equals the sum of category scores multiplied by category weights for categories that fall within that pillar. The total ESG score is then calculated by multiplying the pillar scores with the sum of the category weights within each pillar (Refinitiv, 2021).

6.3 Financial Performance

For our thesis to yield reliable results, we must ensure that our financial performance variables can serve as a true indicator of how well firms are doing financially. The three subdivisions of CFP are market-based measures, accounting-based measures and perceptual measures (Orlitzky et al., 2003).

Market-based measures such as price per share or Tobin's Q reflect the notion that shareholders are a primary stakeholder group. Ultimately, their satisfaction determines the fate of the company. What shareholders decide to do with their shares relies upon expected movements in the share price and their actions decide the market value of the firm (Orlitzky et al., 2003)).

Accounting-based measures such as Return on Assets (ROA) or Return on Equity (ROE) captures the internal efficiency of a firm to some extent. Accounting returns are influenced by managers' allocations of funds to different projects. Thus, they reflect internal decision-making capabilities and managerial performance (Orlitzky et al., 2003).

Lastly, perceptual measures of corporate financial performance ask survey respondents to provide subjective estimates of for instance, the strength of a firm's financial position or how they are positioned compared to competitors (Orlitzky et al., 2003). We will refrain from using these measures in our thesis due to their subjective nature.

We decided to use both market-based and accounting-based measures of financial performance in line with previous research (Choi and Wang, 2009; Velte, 2017). ROA is one of the most common accounting-based measures of financial performance. It represents the profitability of the company in relation to its total assets. Market-based measures are necessary to include in empirical studies as accounting-based variables are often influenced by earnings management decisions (Velte, 2017). For this reason, we will include Tobin's Q as a measure of financial performance as well. Tobin's Q is the ratio between an asset's market value and its replacement value. It has become common practice in finance literature to measure the ratio by comparing the market value of a firm's equity and liabilities with its corresponding book values as the replacement value of a company's assets are hard to evaluate (Velte, 2017).

6.4 Control Variables

As our sample contain firms with different characteristics regarding size, risk, level of innovation and capital structure, we have incorporated control variables in our model to account for these differences. The inclusion of control variables have become common in this area of research (Choi and Wang, 2009; Velte, 2017). In line with suggestions from previous research, we include control variables to control for firm size, risk, R&D and advertising spending (Margolis et al., 2009).

Total assets have been collected from each firm's balance sheet to control for firm size. In our model, firm size is measured by the natural logarithm of total assets in million EUR.

We have used total debt in relation to total assets to control for firm risk. Firms with a high level of ESGP are perceived as less risky in relation to insurance effects and will be associated with lower costs of debt (Velte, 2017). Risk is measured in our model as the total debt of a firm in percentage of that firm's total assets.

We have used the R&D expense from each company's income statement to control for the technological knowledge within the companies in our sample. Being a leading R&D spender within a certain industry could result in improved production processes or products with higher quality than those of the competitors. Thus, it can improve financial performance. Technological knowledge is measured in our model as the natural logarithm of R&D expense in million EUR. Unfortunately, many of the companies in our sample lack data on R&D spending. Excluding the R&D variable from our model would increase the number of observations in our sample. However, we decided to include it as we are concerned that excluding the variable could result in omitted variable bias.

Our last control variable is advertising spending which is measured in our model by the natural logarithm of the selling expense in million EUR. This item can be found in each company's income statement. Bloomberg's definition of selling expense assures us that this item includes the expenditure made towards advertisement. We justify the use of this control variable as firms that spend a lot on advertising will be more visible to the public. This visibility may lead to increased sales and higher profitability. As with R&D data, many of the firms in our sample lack data on advertising spending. We decided to include this variable for the same reason we keep R&D.

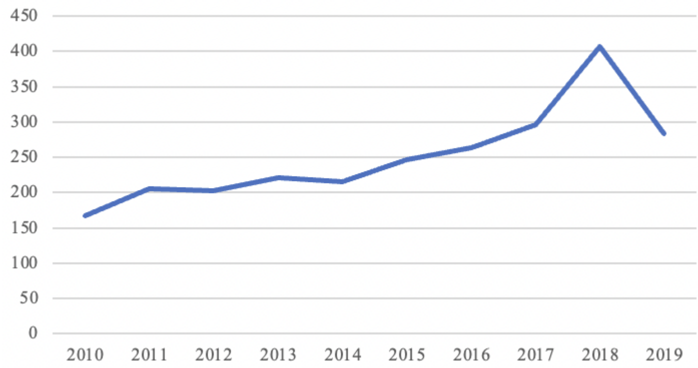
7 Descriptive Statistics

Descriptive statistics are meant to summarize the characteristics of a data set. These statistics can be separated into measures of central tendency and measures of variability. The former includes measures such as mean and median values while the latter includes measures such as standard deviation, variance, maximum and minimum values, skewness and kurtosis (Brooks, 2014). These measures will be discussed throughout this chapter.

7.1 Development of Time Series Data

In this section, we will begin by addressing the development in the total number of observations in our dataset. Next, we will illustrate the development in the mean and median values for our dependent variables, ROA and Tobin's Q and our main variable of interest, ESG scores.

Figure 1: Year by year development in total number of observations



As can be seen from figure 1, the general trend up until 2018 is that the number of observations increased. This trend could be explained by the increasing attention ESG has been getting over the last decade. If there are more companies that have ESG scores in the later part of our sample period compared to the beginning of the period, our number of observations will increase as we get to the later stages of our sample period. However, this does not explain the drastic decrease in observations from 2018 to 2019. We obtain an additional observation only if we are able to collect data on ESG scores, financial performance and for each of the control variables. Refinitiv provided us with fewer ESG scores for 2019 compared to the two years before. As a result, we have fewer total observations in 2019 compared to the two preceding years. The development of yearly observations for each variable can be found in the appendix (figure 6). The number of observations in our dataset ranges from 167 in 2010 to 407 in 2018. In total, we have 2,506 observations in our unbalanced panel where all of the observations contain data on ESG scores, financial performance, size, risk, R&D spending and advertising spending.

Figure 2: Year by year development in ROA for sample firms measured in %

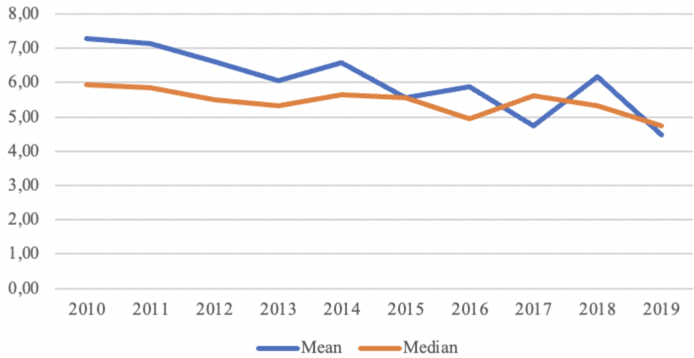
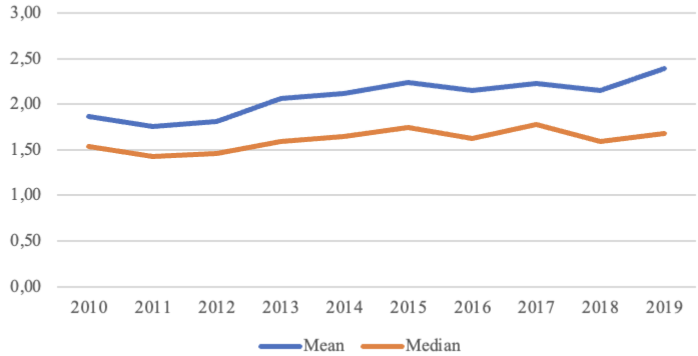


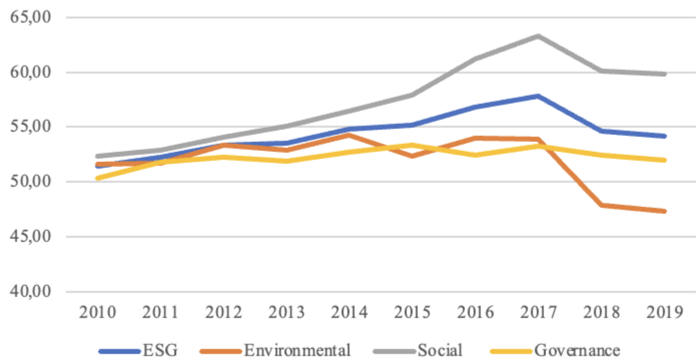
Figure 2 indicates a decreasing trend in ROA during our sample period. Both the average and the median ROA value is lower in 2019 compared to 2010. One interesting thing to note is that the mean value is higher than the median value in every year except from 2017 and 2019. In 2010, the average ROA was 7.30% while the median ROA was 5.93%. This feature indicates that certain firms in our sample had very high ROA which drove the average value up to a level which was higher than the median value. However, when we reach the end of our sample period, the average ROA is 4.49% and the median ROA is 4.75%. Thus, at the end of the sample period, there are firms with sufficiently low ROA to push the average value below the median value.

Figure 3: Year by year development in Tobin's Q for sample firms



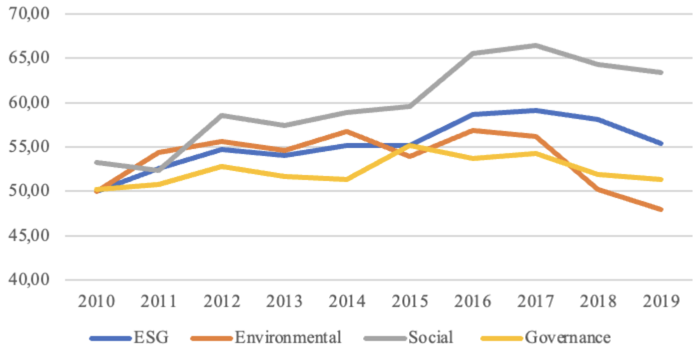
Both mean and median Tobin's Q increases as we reach the end of our sample period which is illustrated by figure 3. As opposed to ROA, the mean ratio for Tobin's Q is higher than the median ratio for the entire sample period. This gap indicates that some firms in our sample have sufficient Tobin's Q to drive the mean above the level of the median. The difference between the mean and median ratio increases as we approach the end of our sample period. While the mean ratio grows from 1.86 in 2010 to 2.39 in 2019, the median ratio only grows from 1.54 to 1.67 in the same time period.

Figure 4: Year by year development in mean ESG scores for sample firms



When looking at figure 4, we note the differences between average total ESG score and individual pillar scores in the more recent years of our sample period compared to the early years. One interpretation could be that as ESG has become more important over the years, more effort has been put into distinguishing between the different aspects of ESG. The average score is higher in 2019 compared to 2010 for all scores except for the Environmental pillar. This development is surprising as the concern for climate change is greater than ever. However, it could be that firms are judged more strictly than before on environmental issues because of the increased attention towards climate change.

Figure 5: Year by year development in median ESG scores for sample firms



By comparing figure 4 and 5, we see that the overall movements of the different ESG scores from 2010 to 2019 are similar when measured by mean and median. The difference between average and median scores are smaller compared to the differences in our financial performance variables. Smaller variations are expected as the ESG scores will always be in the range of 0-100 as opposed to measures of financial performance which are not constrained to a specific interval. Therefore, we do not experience the same deviation of the mean where it is drawn away from the median value due to large outliers.

Both figures also show a drop in ESG scores between 2017 and 2018. An interesting thing to note is that the governance score is less affected by whatever caused this drop. It is difficult to pinpoint exactly why this drop occurred, but 2018 was not a good year for investors in general. In fact, the MSCI Europe index dropped by 10.6% (Morningstar, 2019). As our sample is an extension of this index, most firms in our sample are likely to have had a poor 2018 as well. This may have caused them to shy away from ESG investments which resulted in lower ESG scores. Previous research has found governance to be the driving force of the positive relationship between ESGP and FINP (Velte, 2017). Therefore, firms may decide to prioritize investments into the governance dimension.

We will not elaborate much on the mean and median values of the control variables. However, we would like to point out that they are fairly constant over time. Moreover, the similarities between the mean and median values suggests that there are no large outliers. Yearly development in mean and median values for our control variables can be found in the appendix (figure 7 & 8).

7.2 Distribution of Variables

In this section, we will present some statistics that are useful in explaining the spread among observations in our sample. Furthermore, we will discuss the exclusion of certain observations that in our opinion, are detrimental to the explanatory power of our model. Lastly, we present a correlation matrix that will be useful in explaining the stand-alone effect the variables have on each other.

7.2.1 Full Sample

Table 2: Summary statistics

	Mean	Median	Variance	Std.Dev	Min	Max	Kurtosis	Skewness
Tobins Q	2.107	1.623	2.415	1.554	0.406	19.966	23.804	3.625
ROA	5.940	5.388	141.780	11.907	-297.264	97.603	183.889	-7.206
ESG	54.640	55.871	398.402	19.960	0.629	93.949	2.276	-0.259
Environmental	51.601	53.351	696.347	26.388	0	98.036	2.043	-0.246
Social	58.006	61.223	562.660	23.720	0.437	97.620	2.105	-0.349
Governance	52.332	52.142	487.090	22.070	0.818	97.545	2.142	-0.082
Size	8.036	7.939	2.503	1.582	2.876	12.480	37.255	5.133
Risk	23.119	22.545	255.053	15.970	0	138.472	6.750	1.090
R&D	2.532	2.644	5.655	2.378	0	9.036	64.433	7.342
Advertising	5.099	5.208	3.725	1.930	0	9.914	32.566	4.879
Observations	2506							

This table presents the summary statistics for the variables used in this thesis. Tobins'Q is the ratio between market and book value of a firms equity and liablities. ROA is the firms return on asset in %. ESG scores are provided by Refinitiv. Size is the natural logarithm of our firms book-value. Risk is measured by debt-to-equity ratio in %. R&D and Advertising is the natural logarithm of our firms R&D and advertising expenditure.

All numbers in table 2 are calculated based on all 2,506 observations in our sample. The difference measured in percent between mean and median values are larger for Tobin's Q and ROA than any of the other variables. This feature indicates that our sample contains large outliers that causes the mean to deviate away from the median value. The presence of outliers is also reflected in the kurtosis and skewness. Kurtosis measures the fatness of the tails of the distribution and how peaked at the mean the series is. Skewness defines the shape of the distribution, and measures the extent to which it is not symmetric about its mean (Brooks, 2014).

Tobin's Q has a kurtosis of 23.804 which means that many observations are somewhat similar to the mean. As a result, the distribution peaks at a much higher level compared to the normal distribution. In addition, the maximum Tobin's Q value is much higher than the second largest observation which is approximately 15. Therefore, the tail of the distribution will be fatter than the tail of a normally distributed random variable. A distribution with a higher peak and fatter tail than a normally distributed random variable with the same mean and variance is called a leptokurtic distribution (Brooks, 2014). Moreover, the skewness is positive. A positively skewed distribution indicates that the right hand tail is long and most of the data can be found in the left hand tail (Brooks, 2014). This distribution is the result of large positive outliers.

The kurtosis for ROA is extremely high and the distribution is negatively skewed. This distribution is a result of large negative outliers. The minimum value of -297.26% is almost 200 percentage points lower than the second lowest ROA observation. This spread creates a distribution where the left-hand tail is long and most of the data can be found in the right-hand tail.

For all other variables, the mean and median values are more similar. The kurtosis is also notably lower compared to those of Tobin's Q and ROA. Moreover, the distribution of the explanatory variables is less skewed than those of the dependent variables. The Min and Max columns illustrate the differences in firm characteristics among the firms in our sample. Some firms have been given a score of zero on the environmental pillar while others are close to a perfect score of 100. Furthermore, some firms have no leverage at all while others have negative equity on their balance sheet causing total debt to be larger than total assets. These differences result in many of the variables having a high variance. This is especially true for the variables that have high mean values such as risk and ESG scores.

7.2.2 Dealing with Outliers

We have identified two observations that we will describe as extreme outliers. These are the maximum value observation for Tobin's Q which is the 2015 observation for Fingerprint Card AB and the minimum value observation for ROA which is the 2017 observation for Pharol SGPS S.A.

Fingerprint Cards AB is a Swedish biometrics company. Their incredible growth in 2015 can be explained by the way smartphone manufacturers changed the way people log into their phones. Being able to log into your phone by using your fingerprint resulted in immense growth for biometrics companies. Fingerprint Cards' share price was 20 times higher at the end of 2015 compared to the beginning of the year. Since then, the share price has plummeted to pre-2015 levels (Bloomberg, 2018). Moreover, total assets in 2015 was approximately five times larger than the year before according to the company's annual report. We believe that the drastic increase in their share price reflects an overreaction in the market. Therefore, this observation is excluded.

Pharol SGPS S.A is a Portuguese telecommunication provider. They owned 25.7% of a Brazilian telecommunication firm called Oi which filed for bankruptcy in 2016 (Reuters, 2017a). What followed was the largest ever Latin American restructuring where creditors could swap their debt for up to 75% of Oi's equity which severely diluted the shares owned by Pharol (Reuters, 2017b). Due to losses on their investment in Oi, Pharol's net income fell from -62 mEUR in 2016 to -783 mEUR in 2017 according to their 2017 annual report. These losses are what caused the extremely negative ROA observation which we will exclude from our sample.

Table 3: Summary statistics after removal of outliers

	Mean	Median	Variance	Std.Dev	Min	Max	Kurtosis	Skewness
Tobins'Q	2.101	1.623	2.289	1.513	0.406	15.875	18.801	3.288
ROA	6.036	5.388	103.598	10.178	-108.111	97.603	28.806	-1.097
ESG	54.669	55.875	397.672	19.942	0.629	93.949	2.279	-0.259
Environmental	51.636	53.381	695.294	26.368	0	98.036	2.045	-0.247
Social	58.025	61.228	562.643	23.720	0.437	97.620	2.107	-0.350
Governance	52.367	52.151	486.017	22.046	0.818	97.545	2.142	-0.081
Size	8.038	7.940	2.500	1.581	2.876	12.480	2.661	0.267
Risk	23.137	22.552	254.830	15.963	0	138.472	6.759	1.091
R&D	2.533	2.644	5.657	2.379	0	9.036	2.027	0.391
Advertising	5.102	5.209	3.715	1.928	0	9.914	2.865	-0.226
Observations	2504							

This table presents the summary statistics for the variables used in this thesis after removal of outliers. Tobins'Q is the ratio between market and book value of a firms equity and liablities. ROA is the firms return on asset in %. ESG scores are provided by Refinitiv. Size is the natural logarithm of our firms book-value. Risk is measured by a firms debt-to-equity ratio in %. R&D and Advertising is the natural logarithm of our firms R&D and advertising expenditure.

After excluding the aforementioned observations, we have 2,504 observations left in our sample. In table 3, we can see that the difference between mean and median for Tobin's Q is slightly lower than before as the original maximum value observation has been removed. The difference between mean and median for ROA has increased slightly due to the minimum value being higher than before. We did not expect that removing two observations would impact the mean values by much given the size of our sample.

The kurtosis has decreased for both variables. However, the effect of excluding these observations is larger for ROA compared to Tobin's Q. This result is not surprising as the difference in min and max values has decreased far more for ROA relative to Tobin's Q. We see similar results when looking at the change in skewness. In our opinion, the quality of our sample has been improved by excluding extreme outliers due to the reduced skewness and kurtosis in the distribution of our explanatory variables.

7.2.3 Correlation

Table 4: Correlation matrix

	Tobins'Q	ROA	ESG	Environmental	Social	Governance	Size	Risk	R&D	Advertising
Tobins'Q	1									
ROA	0.444	1								
ESG	-0.132	0.006	1							
Environmental	-0.140	0.027	0.861	1						
Social	-0.073	0.020	0.908	0.749	1					
Governance	-0.146	-0.034	0.661	0.359	0.393	1				
Size	-0.341	-0.075	0.656	0.645	0.586	0.380	1			
Risk	-0.169	-0.224	0.115	0.088	0.103	0.076	0.197	1		
R&D	0.013	-0.014	0.311	0.286	0.306	0.158	0.340	-0.109	1	
Advertising	-0.073	0.098	0.499	0.478	0.443	0.318	0.658	-0.037	0.340	1

This table presents the Person-correlation matrix for our data.

The correlation matrix illustrates the effect that each variable has on each other on a stand-alone basis. In our sample, every variable except from ROA and R&D spending appears to be negatively correlated with Tobin's Q. We would expect the correlation between Tobin's Q and ROA to be positive as they are both measures of financial performance. Increased R&D spending among firms in our sample should increase Tobin's Q according to the correlation matrix. As Tobin's Q is a market-based measure, increased R&D spending could send a signal to the market of expected increased efficiency and competitiveness. Size, risk and advertising spending are negatively correlated with Tobin's Q. Thus, large firms, highly levered firms and firms with high advertising spending should expect lower Tobin's Q according to the correlation matrix.

Judging by the negative correlation between Tobin's Q and the different ESG scores, high performing ESG firms should expect lower market-based returns. The governance pillar score has the largest negative effect on Tobin's Q, followed by the environmental pillar score. The social pillar score appears to be less negatively correlated with Tobin's Q.

ROA and Tobin's Q share a negative correlation with size and risk, but as opposed to Tobin's Q, ROA is negatively correlated with R&D spending. Intuitively, increased R&D spending would increase costs on the income statement which in turn would lower the ROA. Moreover, increased advertising spending should increase ROA according to table 4. This result is surprising as we would expect this relationship to be similar to that of ROA and R&D spending. The environmental and social pillar scores are positively correlated with ROA while the governance score and ROA are more strongly and negatively correlated. As a result, we have a very low, but positive correlation between total ESG score and ROA.

The total ESG score is positively correlated with every variable except Tobin's Q. In addition, it is highly positively correlated with each of the pillar scores which is expected as the pillar scores are used to determine the total ESG score. The social score appears to have the strongest correlation with total ESG score followed by the environmental score and then the governance score. Furthermore, all control variables are positively correlated with each of the ESG scores where size appears to have the strongest correlation.

One thing we should note is the high positive correlation between the pillar scores themselves. Even though a firm might be more concerned with one of the ESG pillars, it is unlikely that they will ignore the issues that lay the groundwork for the other pillar scores. Therefore, we expected a positive correlation, but the strength of the correlation does introduce a concern about multicollinearity. The same applies for the strong positive correlation between size and advertising spending. Consequently, multicollinearity is something we need to address in our analysis.

8 Results

In this chapter, we will begin by presenting the results of the model specification tests. Second, we will attempt to ensure the validity of the model by testing for multicollinearity and serial correlation to address the concerns described in the methodology chapter. Lastly, we will present the results from our fixed effects model and provide a discussion of the different regression results.

8.1 Model Building

Table 5: Model spesification tests

Dependent Variable	Independent Variable(s)	Individual Effects	Breusch-Pagan	Hausmann	Model Choice	Wooldridges	Robust Std.Errors
Tobin's Q_t	ESG_t	Reject H0	Reject H0	Reject H0	Fixed Effects	Reject H0	Yes
ROA_t	ESG_t	Reject H0	Reject H0	Reject H0	Fixed Effects	Reject H0	Yes
Tobin's Q_t	ESG_{t-1}	Reject H0	Reject H0	Reject H0	Fixed Effects	Reject H0	Yes
ROA_t	ESG_{t-1}	Reject H0	Reject H0	Reject H0	Fixed Effects	Failed to reject H0	No
Tobin's Q_t	ESG-pillars$_{t-1}$	Reject H0	Reject H0	Reject H0	Fixed Effects	Reject H0	Yes
ROA_t	ESG-pillars$_{t-1}$	Reject H0	Reject H0	Reject H0	Fixed Effects	Failed to reject H0	No

This table presents the results from the model specification tests.

As mentioned in the methodology chapter, there are three different models that can be applied when working with panel data. Table 5 summarizes the results of the tests that are used to determine which model is most suited. We started with a Chow F test of individual effects where the null hypothesis was rejected in every case. Therefore, a fixed effects model is preferable to a pooled OLS estimation. The Breusch-Pagan test compares the fit of the random effects model to the OLS model. As the null hypothesis is rejected in every case, a random effects model is preferable to a pooled OLS estimation. We also conducted a Hausmann test to see if our data sample is more exposed to fixed or random effects. The null hypothesis is rejected in every case here as well. Therefore, we will use a lagged and unlagged fixed effects model with ROA and Tobin's Q as dependent variables. Detailed results from the model specification tests can be found in the appendix (tables 12, 13 & 14).

8.2 Validity of the Model

We discussed potential threats to the validity of our results in the methodology chapter. This section will present the actions taken to ensure the validity of our model.

Large outliers could cause distortions to the variance of our idiosyncratic error term. The presence of such outliers was investigated, and two observations were removed as explained in the descriptive statistics chapter. After these outliers were removed, we plotted the idiosyncratic error term against the ESG variable to visually inspect its variance (Brooks, 2014). This plot is found in the appendix (figure 9 & 10) . Based on this examination, we conclude that the mean and variance of the idiosyncratic error term remain constant.

Through univariate analysis using the correlation matrix we found some cause for concern in relation to multicollinearity. To further investigate if the regressors are dependent on each other, we calculated the VIF factor for each of the regressors. The formulation of the VIF-factor, acceptance thresholds and individual VIF-factors can be found in the appendix (table 11). The results allow us to continue with our selected model under the assumption of no perfect multicollinearity.

As can be seen from table 5, most of our regression models reject the null hypothesis of the Wooldridge's test for serial correlation. These results indicate that the level of serial correlation in our idiosyncratic errors could cause our estimators to be inefficient. Following Wooldridge (2010), we employ robust standard errors to our models to ensure validity in our results. We have used robust standard errors for the models that did not reject the null hypothesis as well to ensure comparability and consistency of the results across models. This decision will also make our models more conservative and add to the robustness of our statistical inferences.

8.3 Regression Results

In this section, we start by introducing both the unlagged and lagged regression results from our fixed effects model with the total ESG score as the explanatory variable. We will compare the results and decide which model to proceed with. Next, we present the results from the selected model that uses the individual ESG pillar scores instead of the total ESG score. In the two last sections, we divide the firms in our sample into subsamples categorized by industry and region and present the regression results for each subsample.

8.3.1 Section 1: Total ESG Scores

This section examines the relationship between ESGP measured by Refinitiv ESG scores and FINP represented by the accounting-based and market-based measures, ROA and Tobin's Q. The purpose of this section is to establish whether the relationship exists. If we find evidence of an existing relationship between ESGP and FINP, we will compare the results from the lagged and unlagged model to decide which model is best suited to explain the relationship.

Table 6: Full sample total ESG

Panel A: FINP Measured by Tobins'Q

	Coefficient	T-Stat	P-Value
ESG_t	0.0056	1.7491	0.0809
$Size_t$	-0.2923	-2.3238	0.0205
$Risk_t$	-0.0051	-1.4143	0.1579
RD_t	0.0399	1.0995	0.2721
AD_t	0.1184	2.1393	0.0329
Adj.R	n	T	N
0.0185	490	10	2504

Panel B: FINP Measured by ROA

	Coefficient	T-Stat	P-Value
ESG_t	-0.0366	-1.2478	0.2127
$Size_t$	0.5781	0.3070	0.7590
$Risk_t$	-0.1743	-2.7324	0.0065
RD_t	0.1055	0.1813	0.8562
AD_t	0.2940	0.3714	0.7105
Adj.R	n	T	N
0.0365	490	10	2504

This table present the results from the fixed effects regression employing the within transformation. Panel A and B show the results for Tobin's Q and ROA as the dependent variable, respectively. The independent variable is the Total ESG score provided by Refinitiv. Control variables are Size measured by the natural logarithm of book-value in mEUR, Risk measured by the Debt-to-Equity ratio in %, RD measured by the natural logarithm of R&D-expenditure in mEUR, AD measured by the natural logarithm of advertising expenditure in mEUR.

Table 6 show the results from our regression analysis using the unlagged values of ESG as the independent variable. In Panel A we have Tobin's Q as the dependent variable, while in Panel B, ROA is the dependent variable.

The adjusted R^2 is 0.0185 and 0.0365 for panel A and B, respectively. These results imply that the variation in our independent variables performs poorly in explaining the variation in our dependent variables.

Our results show a positive relationship between variations in the total ESG score and variations in Tobin's Q. The relationship is statistically significant at the 10% level. Tobin's Q is a market-based measure of financial performance which reflects the interests of a company's shareholders. Thus, a positive relationship between variations in ESG scores and Tobin's suggests that investors value companies that prioritizes ESG dimensions in their operations. These results support the stakeholder theory, implying that investors value companies that takes a broader responsibility than pure financial gains. Our results are in line with Velte (2017), suggesting a positive relationship between ESG scores and Tobin's Q. In contrast, we find that the relationship is statistically significant. Velte (2017) limits the sample from 2010-2014, while we include data up until 2019. The significance of the ESG-Tobin's Q relationship could be explained by an increased investor focus on ESG dimensions over the last years.

For variations in ROA, we find a negative relationship with variations in ESG score. This result contrasts the findings of Velte (2017) which suggested a positive relationship between the variables. Our results imply that improvements in a company's ESG score is associated with a reduction in ROA. However, the relationship is not statistically significant, and we cannot reject the null hypothesis that no relationship exists between the two variables. Economically, the negative relationship between variations in ESG score and ROA supports the shareholder theory. Investments in ESG dimensions is a direct cost for a company. Thus, for a company to improve its ESG scores, it will have to reduce their net income which directly affects their ROA. It is possible that financial gains from ESG investments will appear in the years following the actual investment (Velte, 2017). This will be examined in the lagged model.

Size and advertising spending are both significant at the 5% level in panel A. However, large firms should have lower Tobin's Q while firms that spend a lot on advertising should have higher Tobin's Q. Neither risk nor R&D spending are significant at any level for Tobin's Q, but risk is the only control variable that is statistically significant in explaining ROA. Increased leverage appears to influence financial performance negatively measured by both Tobin's Q and ROA. Size has a positive, but statistically insignificant impact on ROA.

Table 7: Lagged sample total ESG

Panel A: FINP Measured by Tobins'Q

	Coefficient	T-Stat	P-Value
ESG_{t-1}	0.0074	2.3311	0.0202
$Size_t$	-0.3139	-2.5216	0.0121
$Risk_t$	-0.0021	-0.6564	0.5119
RD_t	0.0514	1.4679	0.1429
AD_t	0.1267	2.1873	0.0293
Adj.R	n	T	N
0.0232	401	10	1943

Panel B: FINP Measured by ROA

	Coefficient	T-Stat	P-Value
ESG_{t-1}	-0.0789	-1.8066	0.0716
$Size_t$	2.5386	1.2578	0.2092
$Risk_t$	-0.2103	-3.9066	0.0001
RD_t	0.2843	2.4683	0.6197
AD_t	-1.0152	-1.2026	0.1669
Adj.R	n	T	N
0.0606	401	10	1943

This table present the results from the fixed effects regression employing the within transformation and lagged ESG scores. Panel A and B show the results for Tobin's Q and ROA as the dependent variable, respectively. The independent variable is the one year lagged Total ESG score provided by Refinitiv. Control variables are Size measured by the natural logarithm of book-value in mEUR, Risk measured by the Debt-to-Equity ratio in %, RD measured by the natural logarithm of R&D-expenditure in mEUR, AD measured by the natural logarithm of advertising expenditure in mEUR.

Table 7 show the results from our regression analysis with the lagged ESG score as the independent variable. Tobin's Q is the dependent variable in panel A. ROA is the dependent variable in panel B.

The explanatory power of our model is improved for both regressions by using the lagged value of ESG score as the independent variable. The adjusted R^2 for panel A is now 0.0232, while we see a larger increase in panel B where the adjusted R^2 is 0.0606. We attribute the increase in explanatory power to the time lag of ESG scores as nothing else has changed.

We find that the lagged ESG-variable has a stronger impact on the dependent variable Tobin's Q. Further, the statistical significance of the relationship is improved. These results strengthen our findings from the unlagged model and provide extended support for the stakeholder theory. By lagging the ESG variable, we also find support for the good management theory which is in line with the previous results of Scholtens (2008).

The relationship between ROA and the lagged ESG score is also stronger than in the unlagged model. In addition, the lagged ESG variable is statistically significant in explaining subsequent ROA at the 10% level. These results provide strengthened support for the shareholder theory, and further contrasts the findings of Velte (2017). Neither the unlagged nor lagged ESG-variable is found to positively affect ROA. We suggested that a positive impact from ESG investments could be found in a later time period than the actual investment. Instead, the negative relationship between ESG scores and ROA in the subsequent year is increased. It could still be possible that the effect requires further lags of the ESG variable to be present. However, this will not be further investigated as we would lose valuable observations.

There are no changes in the Tobin's Q regression regarding which control variables are statistically significant. The sign of the coefficients is also similar to those in Table 6. However, the impact of the control variables on Tobin's Q is larger for all variables except from risk. The impact of size on ROA is larger in the lagged model compared to the unlagged model. Moreover, the size variable has become more significant even though it is still not statistically significant at any level. The negative impact of risk on ROA has become stronger. This variable is the most statistically significant variable in both samples. Neither R&D nor advertising spending are statistically significant after lagging the total ESG score, but we do note that the sign of the advertising spending coefficient is now negative.

Overall, we find a positive relationship between variations in ESG scores and Tobin's Q. Conversely, the relationship is negative with ROA. This result is consistent for both the unlagged and lagged ESG variable. However, the explanatory power of the lagged model is stronger. Thus, when investigating the influence of each individual pillar score, we will apply the lagged ESG variables. This decision is in line with the recommendation of previous research (Velte, 2017).

8.3.2 Section 2: Pillar Scores

In this section, we will investigate the effect of individual ESG pillar scores on FINP in our lagged model. By doing this, we intend to discover possible differences in the relationship between the individual pillars and FINP that is unobserved in the aggregated ESG score.

Table 8: Lagged sample pillar scores

Panel A: FINP Measured by Tobins'Q

	Coefficient	T-Stat	P-Value
$\text{Environmental}_{t\text{-}1}$	0.0035	1.4260	0.1546
$\text{Social}_{t\text{-}1}$	0.0079	2.7174	0.0069
$\text{Governance}_{t\text{-}1}$	-0.0033	-1.6386	0.1021
Size_t	-0.3471	-2.7676	0.0059
Risk_t	-0.0023	-0.7051	0.4812
RD_t	0.0489	1.3747	0.1700
AD_t	0.1325	2.2359	0.0259
Adj.R	n	T	N
0.0430	401	10	1943

Panel B: FINP Measured by ROA

	Coefficient	T-Stat	P-Value
$\text{Environmental}_{t\text{-}1}$	0.0014	0.0489	0.9610
$\text{Social}_{t\text{-}1}$	-0.0222	-0.6948	0.4876
$\text{Governance}_{t\text{-}1}$	-0.0514	-2.8414	0.0047
Size_t	2.3948	1.1685	0.2433
Risk_t	-0.2095	-3.8878	0.0001
RD_t	0.2809	0.4930	0.6223
AD_t	-0.9928	-1.3333	0.1832
Adj.R	n	T	N
0.0624	401	10	1943

This table present the results from the fixed effects regression employing the within transformation. Panel A and B show the results for Tobin's Q and ROA as the dependent variable, respectively. The independent variable is the ESG-pillar scores provided by Refinitiv. Control variables are Size measured by the natural logarithm of book-value in mEUR, Risk measured by the Debt-to-Equity ratio in %, RD measured by the natural logarithm of R&D-expenditure in mEUR, AD measured by the natural logarithm of advertising expenditure in mEUR.

By decomposing the total ESG score into individual pillar scores, the explanatory power of our model is improved further. The adjusted R^2 increases from 0.0232 to 0.043 for panel A. For panel B, it only increases from 0.0606 to 0.0624.

For Tobin's Q, we see from table 8 that the statistical significance of the total ESG score is largely driven by the social pillar score. Neither the environmental nor the governance score are statistically significant in explaining Tobin's Q in the subsequent period. The social score on the other hand, is statistically significant at the 1% level. Breaking down the total ESG score into individual pillar scores is also useful in explaining the size and sign of the ESG coefficient in section 1. The environmental and social score both have positive coefficients. However, their positive impact is countered by the negative impact of the governance score. Consequently, the coefficient of the total ESG score in our lagged model is small, but positive.

The environmental pillar score is based on resource use, emissions and innovation within a firm. Thus, the economic interpretation of the environmental coefficient is that investors will react positively to improvements in these areas. Consequently, Tobin's Q will increase in the subsequent time period. This result is in line with our expectation as climate change has become a global concern.

The social pillar score is based on workforce, human rights, community and product responsibility. Improvements within these categories, will increase Tobin's Q in the subsequent period according to our model. We expected this result as high performance within these categories will mitigate firms' exposure to reputational damage.

The governance pillar score is based on management, shareholders and CSR strategy. According to our model, improvements of performance within these categories will cause Tobin's Q to decrease in the subsequent period. This result is surprising as we would expect investors to react positively to e.g.

improvements in management. Moreover, it contradicts the findings of Velte (2017) who found governance to have a positive, but statistically insignificant impact on subsequent Tobin's Q. However, investors might not be supportive of all investments in the governance dimension. For instance, the development and implementation of a new CSR strategy will require a large amount of time and resources.

Our findings suggest that the social score has the largest impact on Tobin's Q in addition to being the most statistically significant. These findings differ from earlier research. Velte (2017) concluded that none of the individual ESG dimensions were statistically significant in explaining Tobin's Q. Ortas et al. (2015) found the environmental dimension to have the highest impact on FINP measured by Tobin's Q. In addition, they found it to be the most statistically significant variable.

We would expect to see a positive relationship between the pillar scores and ROA if the stakeholder theory holds. The results of this model indicate that investments in ESG activities does not provide better FINP measured by ROA. Albeit small, we do find a positive relationship between the environmental pillar and ROA. However, the relationship is strongly insignificant and any inference from this result would be statistically weak. The impact of both social and governance scores are larger, but negative which causes the negative relationship between the total ESG scores and ROA described in section 1. However, the governance score is the only pillar score that is statistically significant in explaining subsequent ROA.

Velte (2017) and Ortas et al. (2015) both find that the governance dimension is statistically significant in explaining ROA. However, the relationship is deemed to be positive in both studies. Moreover, Ortas et al. (2015) found the environmental dimension to be statistically significant as well. Velte (2017) on the other hand, found all ESG dimensions to be statistically significant in explaining ROA. Thus, our results suggest that the relationship between ESGP and FINP measured by ROA are different from what previous research suggests. The two aforementioned studies are based on older data in addition to focusing on different geographical locations. These differences in sample selection could offer some explanation to the contradicting results. However, it could also be an indication of existing differences in the relationship between ESGP and FINP across countries. We will look further into this possibility in the next section.

For both measures of FINP, we observe that the results for the control variables remain consistent with the total ESG regression. This consistent relationship is expected as these variables are not decomposed in the same way as total ESG score.

8.3.3 Section 3: Region Subsamples

In this section, we will investigate whether there are any identifiable differences in the relationship between ESGP and FINP across the regions in our sample. We have divided the sample into three regions. The Nordics consists of firms in Norway, Denmark, Sweden and Finland. The UK consists of firms in England and Ireland. The last region is Central Europe which contains the remaining firms in our sample. This part of our thesis will solely focus on the existence of differences in the relationship. We will not attempt to explain the underlying reason for potential differences across regions.

Table 9: Region subsample regression

Panel A: FINP measured by Tobins'Q

	Nordic				Central				UK		
	Coefficient	T-Stat	P-Value		Coefficient	T-Stat	P-Value		Coefficient	T-Stat	P-Value
$Environmental_{t-1}$	-0.0099	-1.5786	0.1186	$Environmental_{t-1}$	0.0045	1.4322	0.1538	$Environmental_{t-1}$	0.0038	0.9857	0.3261
$Social_{t-1}$	0.0141	1.5796	0.1184	$Social_{t-1}$	0.0079	1.9672	0.0506	$Social_{t-1}$	0.0050	1.4774	0.1419
$Governance_{t-1}$	-0.0037	-0.5300	0.5976	$Governance_{t-1}$	-0.0030	-1.0229	0.3077	$Governance_{t-1}$	-0.0027	-1.3856	0.1682
$Size_t$	-0.5717	-1.9666	0.0529	$Size_t$	-0.1490	-1.3152	0.1900	$Size_t$	-0.6581	-3.4954	0.0006
$Risk_t$	-0.0131	-1.1773	0.2428	$Risk_t$	0.0018	0.4103	0.6820	$Risk_t$	0.0000	-1.2112	0.2280
RD_t	0.5542	2.1824	0.0322	RD_t	0.0724	1.6028	0.1107	RD_t	-0.0020	-0.0404	0.9678
AD_t	0.6451	1.4191	0.1600	AD_t	0.0894	1.4475	0.1494	AD_t	0.0822	0.6951	0.4882
Adj.R	n	T	N	Adj.R	n	T	N	Adj.R	n	T	N
0.0745	77	10	372	0.0817	190	10	832	0.0964	134	10	733

Panel B: FINP Measured by ROA

	Nordic				Central				UK		
	Coefficient	T-Stat	P-Value		Coefficient	T-Stat	P-Value		Coefficient	T-Stat	P-Value
$Environmental_{t-1}$	-0.0650	-0.6572	0.5130	$Environmental_{t-1}$	0.0372	0.8414	0.4012	$Environmental_{t-1}$	-0.0255	-0.7432	0.4587
$Social_{t-1}$	0.0122	0.2279	0.8204	$Social_{t-1}$	-0.0667	-1.4793	0.1407	$Social_{t-1}$	0.0122	0.3891	0.6978
$Governance_{t-1}$	-0.1660	-2.4661	0.0160	$Governance_{t-1}$	-0.0298	-1.2054	0.2295	$Governance_{t-1}$	-0.0203	-1.1296	0.2607
$Size_t$	5.2372	0.8984	0.3718	$Size_t$	4.2456	1.9584	0.0517	$Size_t$	-1.8302	-0.8856	0.3774
$Risk_t$	-0.4287	-3.5040	0.0008	$Risk_t$	-0.1038	-1.6032	0.1106	$Risk_t$	-0.0025	-3.1156	0.0022
RD_t	3.5251	1.0459	0.2989	RD_t	-0.0611	-0.1661	0.8683	RD_t	0.5570	0.5934	0.5539
AD_t	-8.4779	-2.8494	0.0056	AD_t	-0.2615	-0.4577	0.6477	AD_t	-2.0978	-1.3450	0.1809
Adj.R	n	T	N	Adj.R	n	T	N	Adj.R	n	T	N
0.1382	77	10	372	0.0512	190	10	838	0.1515	134	10	733

This table present the results from the fixed effects regression employing the within transformation for our three sub-regions Nordic, Central Europe and UK. Panel A and B show the results for Tobin's Q and ROA as the dependent variable, respectively. The independent variable is the ESG-pillar scores provided by Refinitiv. Control variables are Size measured by the natural logarithm of book-value in mEUR, Risk measured by the Debt-to-Equity ratio in %, RD measured by the natural logarithm of R&D-expenditure in mEUR, AD measured by the natural logarithm of advertising expenditure in mEUR.

From table 9, we note that the environmental pillar score is statistically insignificant in explaining subsequent Tobin's Q for all regions. Nonetheless, the highest level of significance for this variable is found in the Nordics. Interestingly, we find that the environmental pillar score has a negative impact on Tobin's Q in the Nordics. Conversely, the impact is smaller, but positive for firms in Central Europe and the UK. Thus, the results for Central Europe and the UK are consistent with the findings in section 2.

The social pillar score is statistically significant at the 10% level for the Central Europe subsample. However, it is not statistically significant for neither the Nordic nor the UK subsample. The impact of the social pillar score on subsequent Tobin's Q is positive for all subsamples. This result is consistent with our findings from section 2. Similar to our findings for the environmental pillar, the impact of the social pillar score on subsequent Tobin's Q appears to be largest for the Nordic subsample.

The governance pillar score is also statistically insignificant in explaining subsequent Tobin's Q for all subsamples. In addition, the impact of the governance score is negative for all regions. A negative, but insignificant relationship between governance score and subsequent Tobin's Q is consistent with our findings from section 2. We note that the negative impact is largest for the Nordic subsample in this case as well.

Overall, the results for the control variables are fairly consistent with our findings from section 2. However, there are some deviations in our results. Advertisement spending for instance, is significant at the 5% level in section 2, but it is not statistically significant for any of the region subsamples. Moreover, the adjusted R^2 of each region's regression is higher compared to the regression in section 2 which did not distinguish between geographic location of firms. Therefore, it appears that the firms in each subsample are more comparable which increases the explanatory power of each model.

For panel B, we see that the environmental pillar score is statistically insignificant in explaining subsequent ROA for all subsamples. Moreover, the impact of the environmental score is negative in the Nordics and the UK while it is positive for the firms in Central Europe. In section 2, we found the relationship to be slightly positive and highly insignificant. Thus, these findings are somewhat consistent with the results from section 2. The impact of the environmental score appears to be largest for the firms in the Nordics.

The social pillar score is also statistically insignificant across all subsamples. In addition, the impact of this pillar score is positive for the Nordic and UK subsample, but negative for the firms in Central Europe. The size of the coefficient is larger for the Central Europe subsample. This is also the subsample that contains the most firms. Therefore, these results are consistent with our findings from section 2 as the results from the Central Europe subsample will dominate the other subsamples given its larger coefficient and sample size.

The governance pillar score is statistically significant at the 5% level for the Nordic subsample while statistically insignificant for the Central Europe and UK subsamples. Furthermore, the impact of the governance score on subsequent ROA is negative for all subsamples. In section 2, we found the governance score to be significant, but negative. Therefore, the results are somewhat consistent with the findings from section 2. The impact of the governance score on subsequent ROA is largest for the Nordic subsample in this case as well.

The results of the control variables in Panel B are consistent with our findings in section 2. Furthermore, the adjusted R^2 has increased for the Nordic and UK subsamples while it has decreased for the Central Europe subsample.

In summary, we should be careful about making any conclusions based on these results given the lack of statistical significance in the subsample regressions. However, we do find evidence of the social pillar score having a positive impact on subsequent Tobin's Q for Central European firms. In addition, we find evidence which suggest that the governance pillar score has a negative impact on subsequent ROA for Nordic firms.

8.3.4 Section 4: Industry Subsamples

In this section, we will continue to look for differences in the relationship between ESGP and FINP across subsamples. However, we will separate the firms by industry instead of geographic location. All firms in our sample have been categorized into three industry groups. These groups are Raw Materials, Manufacturing and Services. The companies in our sample have been sorted into these groups based on their Global Industry Classification Standard(GICS) classification. Similar to the previous section, we will not attempt to explain the underlying reason for potential differences.

Table 10: Industry subsample regression

Panel A: FINP measured by Tobins'Q

	Raw Materials				Manufacturing				Services		
	Coefficient	T-Stat	P-Value		Coefficient	T-Stat	P-Value		Coefficient	T-Stat	P-Value
Environmental	0.0013	0.2797	0.7808	Environmental	0.0019	0.5478	0.5844	Environmental	0.0062	1.4601	0.1463
Social	0.0061	1.8379	0.0715	Social	0.0108	2.3652	0.0190	Social	0.0036	0.8143	0.4168
Governance	0.0006	0.0964	0.9236	Governance	-0.0071	-2.2793	0.0237	Governance	-0.0005	-0.2306	0.8179
$Size_t$	-0.2941	-1.4983	0.1398	$Size_t$	-0.3487	-1.8430	0.0669	$Size_t$	-0.3859	-1.7884	0.0757
$Risk_t$	0.0040	0.6906	0.4927	$Risk_t$	-0.0070	-1.1804	0.2393	$Risk_t$	-0.0008	-0.1580	0.8747
RD_t	0.2259	2.0270	0.0475	RD_t	0.1099	2.0974	0.0373	RD_t	-0.0167	-0.3531	0.7245
AD_t	0.0558	0.3987	0.6917	AD_t	0.1444	1.3330	0.1841	AD_t	0.1164	1.5736	0.1177
Adj.R	n	T	N	Adj.R	n	T	N	Adj.R	n	T	N
0.0588	56	10	270	0.0627	194	10	915	0.0335	151	10	758

Panel B: FINP Measured by ROA

	Raw Materials				Manufacturing				Services		
	Coefficient	T-Stat	P-Value		Coefficient	T-Stat	P-Value		Coefficient	T-Stat	P-Value
Environmental	0.1060	1.3054	0.1972	Environmental	0.0162	0.4103	0.6820	Environmental	-0.0327	-0.6440	0.5206
Social	0.0103	0.1892	0.8506	Social	-0.0058	-0.1208	0.9040	Social	-0.0663	-1.0622	0.2898
Governance	-0.0263	-0.6992	0.4874	Governance	-0.0866	-2.9260	0.0038	Governance	-0.0170	-0.5995	0.5497
$Size_t$	4.1134	0.9428	0.3499	$Size_t$	0.4029	0.1891	0.8503	$Size_t$	3.5016	1.0719	0.2855
$Risk_t$	-0.4678	-4.1394	0.0001	$Risk_t$	-0.1627	-3.2694	0.0013	$Risk_t$	-0.1494	-2.2251	0.0276
RD_t	0.6692	0.3695	0.7132	RD_t	-0.1221	-0.2064	0.8367	RD_t	0.8051	1.1069	0.2701
AD_t	-3.7309	-1.8975	0.0630	AD_t	-1.0692	-1.9229	0.0560	AD_t	-0.8730	-0.9739	0.3317
Adj.R	n	T	N	Adj.R	n	T	N	Adj.R	n	T	N
0.3184	56	10	270	0.0502	194	10	915	0.0394	151	10	758

This table present the results from the fixed effects regression for our industry sub-samples. Panel A and B show the results for Tobin's Q and ROA as the dependent variable, respectively. The independent variable is the ESG-pillar scores provided by Refinitiv. Control variables are Size measured by the natural logarithm of book-value in mEUR, Risk measured by the Debt-to-Equity ratio in %, RD measured by the natural logarithm of R&D-expenditure in mEUR, AD measured by the natural logarithm of advertising expenditure in mEUR.

From table 10, we see that the environmental pillar score is statistically insignificant in explaining subsequent Tobin's Q for all subsamples. Moreover, the coefficient is positive for each industry subsample. In section

2, we found the relationship to be positive and insignificant. Thus, these results are consistent. We also note that the impact of the environmental pillar score on subsequent Tobin's Q is largest for the Services subsample.

The social pillar score is statistically significant at the 10% and 5% level for the Raw Materials and Manufacturing subsamples, respectively. However, it is not statistically significant at any level for the Services subsample. Furthermore, the impact of the social score on subsequent Tobin's Q is positive for all subsamples. A positive, significant relationship is consistent with our findings from section 2. Lastly, we note that the impact is strongest for the Manufacturing subsample.

The governance pillar score is statistically significant at the 5% level for the Manufacturing subsample and insignificant for the other industry subsamples. Moreover, the impact of the governance score on subsequent Tobin's Q is negative for the Manufacturing and Services subsamples. However, it is positive for the Raw Materials subsample. Our results from section 2 suggests that the relationship is negative and statistically insignificant. Thus, our findings are somewhat consistent with the results of section 2. The impact on subsequent Tobin's Q originating from the governance pillar score is largest for the Manufacturing subsample.

The results of our control variables are mostly consistent with our findings from section 2. However, we find that advertisement spending is no longer statistically significant for any of the industry subsamples. Moreover, the adjusted R^2 has improved for the Raw Materials and Manufacturing subsamples. Conversely, it has decreased for the Services subsample. Thus, creating subsamples based on industry classification does not appear to have the same positive effect on the explanatory power of the model as the region-based subsamples.

In Panel B we have the regression analysis for our industry groups with ROA as the dependent variable. The environmental pillar is statistically insignificant for all industry groups. This result is in line with our findings from section 2. Further, we find that the relationship is positive within the Raw Materials and Manufacturing subsamples. The impact of the environmental pillar on subsequent ROA is strongest in the Raw Materials group. In contrast to the results from section 2, we observe a negative relationship between the environmental pillar and ROA in the Services subsample.

The social pillar score is statistically insignificant for all industry subsamples. The relationship between the social pillar and subsequent ROA is negative for the Manufacturing and Services sample. In contrast, the relationship within the Raw Materials sample is positive. Most observations are found in the Manufacturing and Services subsamples, which explains the negative relationship found in section 2. The impact on ROA is strongest in the Services subsample for the social pillar.

For the governance pillar we find a statistically significant relationship with subsequent ROA in the Manufacturing subsample. The relationship is significant at the 1% level. In the Raw Materials and Services subsamples, the relationship is statistically insignificant. The governance pillar coefficient is negative across all industry subsamples. This result is consistent with our findings from section 2. We find that the governance pillar has the strongest effect on subsequent ROA in the Manufacturing subsample.

The control variable coefficients are similar to our previous findings. Our models for the Manufacturing and Services subsamples yield a lower adjusted R^2 than for the corresponding model in section 2. However, we have a notably higher adjusted R^2 in the model for the Raw Materials subsample.

Similar to section 3, we should proceed with caution when making inferences based on these results due to the lack of statistical significance. However, we do find evidence that the social pillar score has a positive, statistically significant impact on subsequent Tobin's Q for the Manufacturing subsample. Furthermore, we find evidence of a negative, statistically significant relationship between the governance pillar score and subsequent Tobin's Q for the same subsample. This result is interesting as the governance pillar score is not statistically significant according to our findings from section 2. In addition, we find a negative, significant relationship between the governance score and subsequent ROA for the manufacturing subsample as well. We find no statistically significant relationship between ESGP and FINP for the Services subsample. For the Raw Materials subsample, we only find a statistically significant relationship between the social pillar and subsequent Tobin's Q.

9 Conclusion

The goal with this thesis is to investigate the relationship between European companies' ESGP and FINP. Our thesis addresses the research question:

"Does the relationship between financial performance and ESG performance differ across industries and countries in Europe?"

Using panel data and a fixed effects model with 2,504 observations, we have presented evidence which suggests a positive relationship between ESGP and Tobin's Q. In contrast, the relationship between ESGP and ROA is found to be negative. Furthermore, we find that the relationship is strengthened and more statistically significant when the ESG variable is lagged by one year. This result indicates that it takes time before improvements in ESGP are reflected in FINP.

Decomposing the ESG score into individual pillar scores, we find that the social pillar is the only pillar score that is statistically significant in explaining subsequent Tobin's Q. The relationship is positive which indicates that firms that allocates resources to their workforce, human rights, their community and high product responsibility will attract investors which in turn will increase market-based returns. These firms are likely to be less exposed to reputational damage which could make them an attractive investment. For the accounting-based measure ROA, we find that only the governance pillar is statistically significant in explaining subsequent ROA. This relationship is negative, indicating that firms that allocates resources towards their shareholders, management and CSR strategy are sacrificing accounting-based returns. This relationship could be explained by how costly and time consuming it would be to develop and implement a new CSR strategy.

Our investigation into potential differences in the relationship between ESGP and FINP is done by dividing our data sample into subsamples based on geographic location and industry. The subsample regressions suffer from a lack of statistical significance to a large extent. Consequently, the validity of our results could be subject to critique. We only find evidence of an existing relationship between ESGP and Tobin's Q in the Central Europe subsample. The social pillar is the only statistically significant ESG pillar and the impact on Tobin's Q is positive. Thus, we find support for our overall results in the Central Europe subsample. The only evidence of a relationship between ESGP and ROA is found in the Nordics subsample. Similar to our overall results, the governance pillar is the only statistically significant ESG pillar and the impact on ROA is negative. We do not have evidence to suggest that there exists a relationship between ESGP and FINP in the UK subsample.

The results of our industry subsample regressions are slightly better. We find that the social pillar is statistically significant in explaining Tobin's Q for the Raw Materials subsample. Moreover, the relationship is positive which supports our overall results. However, the positive impact of the social pillar is larger for the Manufacturing subsample. The relationship is more statistically significant in this subsample as well compared to Raw Materials. In addition, the governance pillar has a negative impact on Tobin's Q which is statistically significant in the Manufacturing subsample. The only relationship between ESGP and subsequent ROA we have found is in the Manufacturing subsample as well. The governance pillar is statistically significant in explaining ROA and the impact is negative. This relationship is consistent with our overall results. We have not found evidence to suggest that a relationship between ESGP and FINP exists in the Services subsample.

We conclude that the relationship between ESGP and Tobin's Q is positive and mainly driven by the social pillar. Conversely, the relationship between ESGP and ROA is negative and mainly driven by the governance pillar. The positive relationship between ESGP and Tobin's Q appears to be mostly present in Central European firms while the negative relationship between ESGP and ROA appears to strongest in Nordic countries. Firms that operate within manufacturing appears to have the strongest relationship between ESGP and FINP.

Our contribution to this field of research consists of identifying differences in the relationship between ESGP and FINP among different industries and regions in Europe. Previous literature has yielded mixed results on the ESGP-FINP relationship. With this thesis, we provide results that could explain the inconsistencies of existing research. Our results show that there are differences across regions and industries. Therefore, it is our opinion that general conclusions about the ESGP-FINP relationship cannot be drawn from examining the relationship for one country or industry.

We acknowledge that our thesis has certain limitations. Previous literature has discussed the variation in measurement methods and calculation of ESG score as a problem for ESG research. By basing our inferences solely on the ratings from Refinitiv, our data could be subject to measurement error and our results could vary from other measures of ESGP. Thus, our thesis can only speak to the relationship between Refinitiv's ESG scores and measures of FINP. Furthermore, our data sample was drastically limited due to availability of ESG data. Data selection based on availability could distort the sample and lead to sample selection bias. Moreover, limited ESG data caused our subsamples to be smaller than preferred. Therefore, one could question the validity of our results from the subsample regressions.

For future research, we would recommend the usage of multiple measures of ESGP to ensure that the analysis captures the performance related to each ESG dimension. Additionally, we believe investigating the underlying factors that drive the ESGP-FINP relationship could be an interesting avenue for future research. We have shown that the ESGP-FINP relationship is not homogenous across such subsamples. Therefore, we implore future researchers to gather wide sample data and separating between relevant industries and geographic regions.

APPENDIX

Figure 6: Year by year development in number of observations for all variables

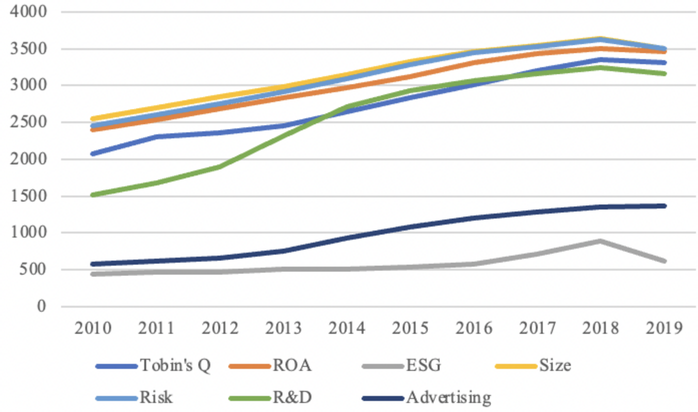


Figure 7: Year by year development in mean control variables for sample firms

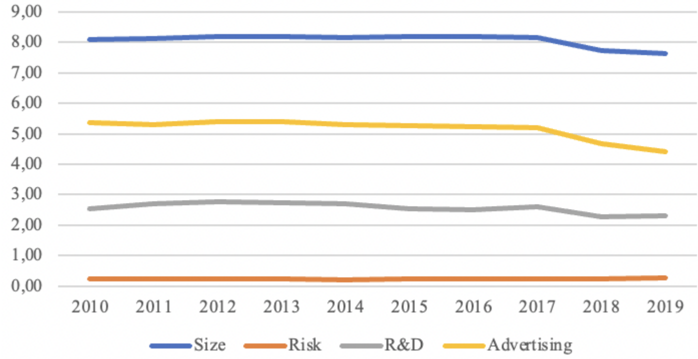


Figure 8: Year by year development in median control variables for sample firms

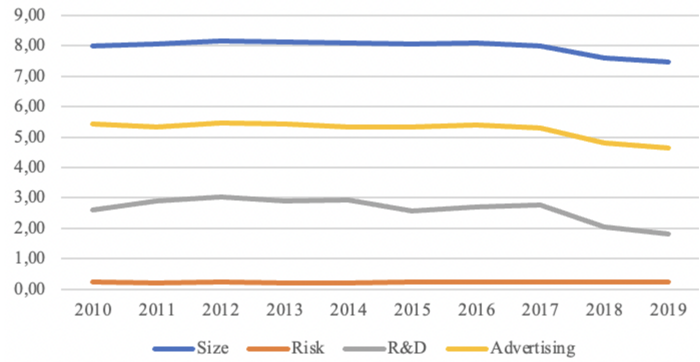


Figure 9: Examination of residuals

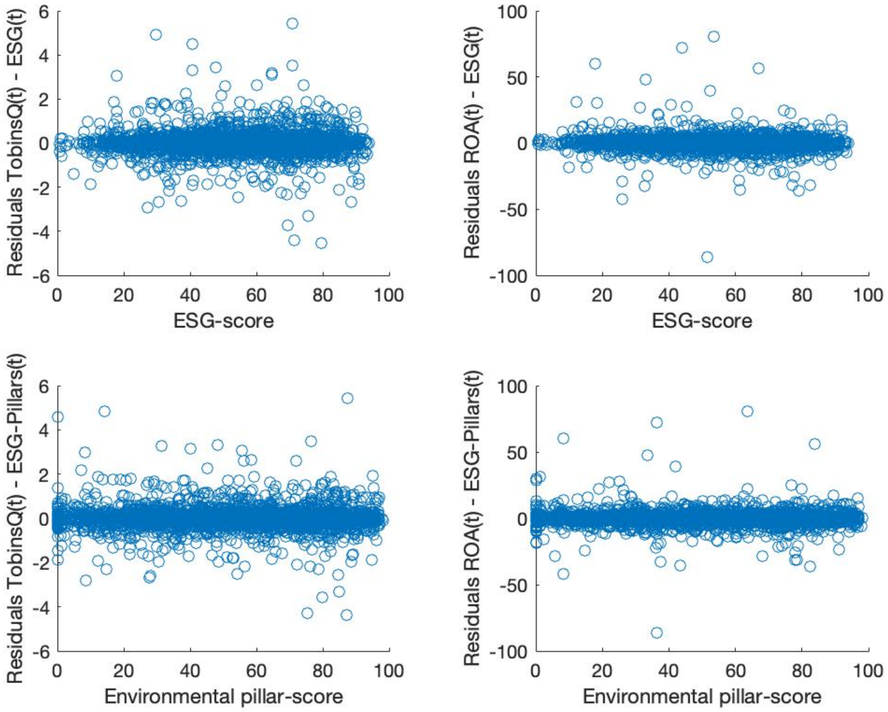


Figure 10: Examination of residuals - lagged ESG

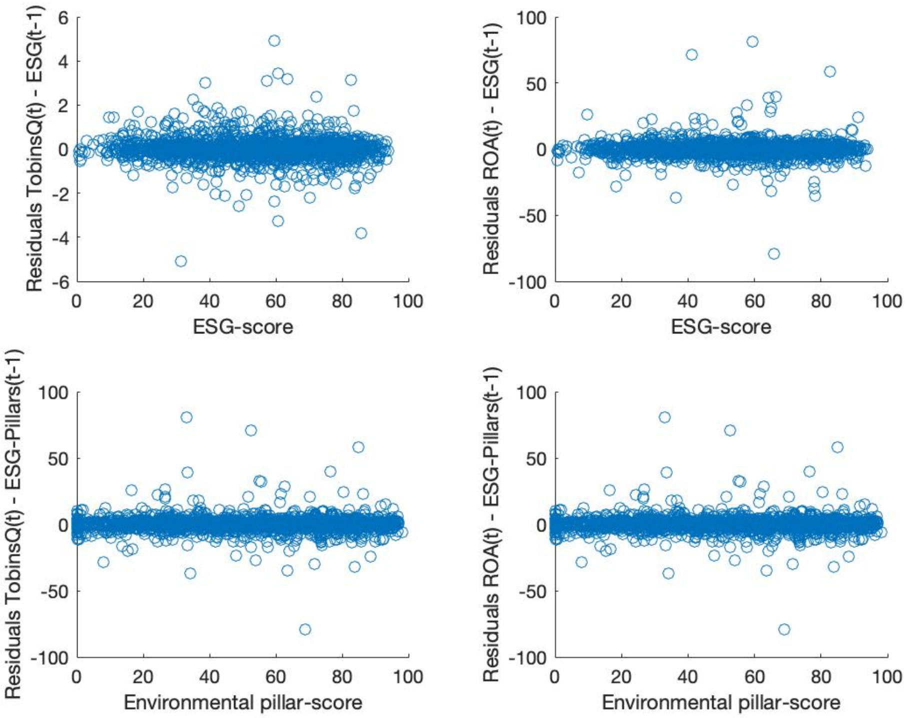


Table 11: Variance Inflation Factor

	ESG	Environmental	Social	Governance	Size	Risk	R&D	Advertsing
VIF	1.146	1.462	1.517	1.099	1.700	1.052	1.103	1.510

This table presents the VIF for our explanatory variables. Common cut off points for the presence of multicollinearity is a VIF factor of 5 or 10 (Chase, 2013). We note that all explanatory variables has VIFs within the lowest threshold.

Table 12: Model specification: Test for individual effects

Sample	Model	F-stat	Prob. >F(P-value)	Reject H0
Full	Tobins_{t}-ESG_{t}	20.156	0.000	Yes
Full	ROA_{t}-ESG_{t-1}	7.086	0.000	Yes
Lagged	Tobins_{t}-ESG_{t-1}	21.169	0.000	Yes
Lagged	ROA_{t}-ESG_{t-1}	5.983	0.000	Yes
Lagged	Tobins_{t}-Pillars_{t-1}	20.972	0.000	Yes
Lagged	ROA_{t}-Pillars_{t-1}	5.856	0.000	Yes
Nordic	Tobins_{t}-Pillars_{t-1}	18.382	0.000	Yes
Nordic	ROA_{t}-Pillars_{t-1}	4.164	0.000	Yes
Central	Tobins_{t}-Pillars_{t-1}	18.354	0.000	Yes
Central	ROA_{t}-Pillars_{t-1}	4.795	0.000	Yes
UK	Tobins_{t}-Pillars_{t-1}	24.839	0.000	Yes
UK	ROA_{t}-Pillars_{t-1}	9.886	0.000	Yes
Raw Materials	Tobins_{t}-Pillars_{t-1}	12.303	0.000	Yes
Raw Materials	ROA_{t}-Pillars_{t-1}	5.759	0.000	Yes
Manufacturing	Tobins_{t}-Pillars_{t-1}	15.331	0.000	Yes
Manufacturing	ROA_{t}-Pillars_{t-1}	6.610	0.000	Yes
Services	Tobins_{t}-Pillars_{t-1}	30.264	0.000	Yes
Services	ROA_{t}-Pillars_{t-1}	5.157	0.000	Yes

This table presents the test-results from the individual effects F-test. We perform a Chow F test on the individual effects, u_i. The unrestricted model follows the fixed effects estimation while the restricted model follows an OLS pooling estimation (Baltagi, 2008) Rejection of H0 suggests the presence of individual effects in the data and gives support for the fixed effects model.

Table 13: Model specification: Breusch-Pagan test

Sample	Model	X2-Stat	Prob. >X2(P-value)	Reject H0
Full	Tobins_{t}-ESG_{t}	2710.936	0.000	Yes
Full	ROA_{t}-ESG_{t}	925.901	0.000	Yes
Lagged	Tobins_{t}-$\text{ESG}_{t\text{-}1}$	2627.440	0.000	Yes
Lagged	ROA_{t}-$\text{ESG}_{t\text{-}1}$	651.899	0.000	Yes
Lagged	Tobins_{t}-$\text{Pillars}_{t\text{-}1}$	2524.580	0.000	Yes
Lagged	ROA_{t}-$\text{Pillars}_{t\text{-}1}$	616.355	0.00	Yes
Nordic	Tobins_{t}-$\text{Pillars}_{t\text{-}1}$	439.589	0.000	Yes
Nordic	ROA_{t}-$\text{Pillars}_{t\text{-}1}$	38.989	0.000	Yes
Central	Tobins_{t}-$\text{Pillars}_{t\text{-}1}$	773.330	0.000	Yes
Central	ROA_{t}-$\text{Pillars}_{t\text{-}1}$	142.413	0.000	Yes
UK	Tobins_{t}-$\text{Pillars}_{t\text{-}1}$	827.065	0.000	Yes
UK	ROA_{t}-$\text{Pillars}_{t\text{-}1}$	315.360	0.000	Yes
Raw Materials	Tobins_{t}-$\text{Pillars}_{t\text{-}1}$	328.090	0.000	Yes
Raw Materials	ROA_{t}-$\text{Pillars}_{t\text{-}1}$	33.842	0.000	Yes
Manufacturing	Tobins_{t}-$\text{Pillars}_{t\text{-}1}$	1229.327	0.000	Yes
Manufacturing	ROA_{t}-$\text{Pillars}_{t\text{-}1}$	294.638	0.000	Yes
Services	Tobins_{t}-$\text{Pillars}_{t\text{-}1}$	857.826	0.000	Yes
Services	ROA_{t}-$\text{Pillars}_{t\text{-}1}$	242.142	0.000	Yes

This table presents the results for the Breusch-Pagan test. We test whether the variance of the individual effects in the data, $\sigma^2_{u_i}$ is statistically different from zero. Rejection of H0 suggests the variance is not equal to zero and supports the use of a random effects model (Breusch and Pagan, 1980).

Table 14: Model specification: Hausman test

Sample	Model	X2-Stat	Prob. >X2(P-value)	Reject H0
Full	$Tobins_t$ -ESG_t	33.271	0.000	Yes
Full	ROA_t -ESG_t	11.621	0.040	Yes
Lagged	$Tobins_t$ -ESG_{t-1}	14.436	0.013	Yes
Lagged	ROA_t -ESG_{t-1}	39.239	0.000	Yes
Lagged	$Tobins_t$ -$Pillars_{t-1}$	16.482	0.000	Yes
Lagged	ROA_t -$Pillars_{t-1}$	38.308	0.000	Yes
Nordic	$Tobins_t$ -$Pillars_{t-1}$	15.368	0.0316	Yes
Nordic	ROA_t -$Pillars_{t-1}$	33.748	0.000	Yes
Central	$Tobins_t$ -$Pillars_{t-1}$	51.72	0.000	Yes
Central	ROA_t -$Pillars_{t-1}$	22.579	0.002	Yes
UK	$Tobins_t$ -$Pillars_{t-1}$	6.730	0.458	No
UK	ROA_t -$Pillars_{t-1}$	38.300	0.000	Yes
Raw Materials	$Tobins_t$ -$Pillars_{t-1}$	11.763	0.109	No
Raw Materials	ROA_t -$Pillars_{t-1}$	44.908	0.000	Yes
Manufacturing	$Tobins_t$ -$Pillars_{t-1}$	10.214	0.1767	No
Manufacturing	ROA_t -$Pillars_{t-1}$	11.680	0.112	No
Services	$Tobins_t$ -$Pillars_{t-1}$	6.520	0.481	No
Services	ROA_t -$Pillars_t$	25.279	0.001	Yes

This table presents the results for the Hausman test. This test examines the differences of the estimators, β_{fe} and β_{re} in the fixed and random effects model. Both models are consistent under H0, while only the fixed effects model is consistent if H0 is rejected (Hausman, 1978)

Table 15: Wooldridge test for serial correlation

Sample	Model	X2-Stat	Prob. >X2(P-value)	Reject H0
Full	$Tobins_t$-ESG_t	38.549	0.000	Yes
Full	ROA_t-ESG_t	35.921	0.000	Yes
Lagged	$Tobins_t$-ESG_{t-1}	47.124	0.000	Yes
Lagged	ROA_t-ESG_{t-1}	16.041	0.000	Yes
Lagged	$Tobins_t$-$Pillars_{t-1}$	43.741	0.000	Yes
Lagged	ROA_t-$Pillars_{t-1}$	14.731	0.000	Yes
Nordic	$Tobins_t$-$Pillars_{t-1}$	11.010	0.002	Yes
Nordic	ROA_t-$Pillars_{t-1}$	0.192	0.663	No
Central	$Tobins_t$-$Pillars_{t-1}$	16.621	0.001	Yes
Central	ROA_t-$Pillars_{t-1}$	24.108	0.000	Yes
UK	$Tobins_t$-$Pillars_{t-1}$	45.708	0.000	Yes
UK	ROA_t-$Pillars_{t-1}$	6.095	0.015	Yes
Raw Materials	$Tobins_t$-$Pillars_{t-1}$	11.474	0.001	Yes
Raw Materials	ROA_t-$Pillars_{t-1}$	2.235	0.142	Yes
Manufacturing	$Tobins_t$-$Pillars_{t-1}$	28.158	0.000	Yes
Manufacturing	ROA_t-$Pillars_{t-1}$	6.073	0.015	Yes
Services	$Tobins_t$-$Pillars_{t-1}$	24.248	0.000	Yes
Services	ROA_t-$Pillars_{t-1}$	9.237	0.003	Yes

This table presents the results from the Wooldridge test for serial correlation. In the presence of serial correlation our estimators will be inefficient and our standard errors could be wrong(Brooks, 2014). If H0 is rejected we employ robust standard errors to add robustness to inferences drawn from the results.

Table 16: Composition of subsamples

Panel A: Composition of Region subsamples

Nordic	Central	UK
Denmark	Austria	Ireland
Finland	Belgium	United Kingdom
Norway	France	
Sweden	Italy	
	Netherlands	
	Portugal	
	Spain	
	Switzerland	

Panel B: Composition of Industries

Raw Materials	Manufacturing	Services
Energy	Health Care	Communication
Materials	Industrials	Consumer Discretionary
	Technology	Consumer Staples
	Utilities	

This table presents the composition of our region and industry subsamples. In panel A we have divided the countries included in our sample into three broader geographic regions. In panel B we have divided industries into three groups based on GICS classification.

Table 17: Companies included in the sample

[HTML]000000888 Holdings PLC	Premier Foods PLC	Swiss Steel Holding AG	Orange SA	Bang & Olufsen A/S
Anglo American PLC	Phoenix Global Resources PLC	Swisscom AG	Compagnie Plastic Omnium SE	Chr Hansen Holding A/S
Associated British Foods PLC	Polypipe Group PLC	Tecan Group AG	Quadient SA	Coloplast A/S
Aggreko PLC	PPHE Hotel Group Ltd	Temenos AG	Remy Cointreau SA	Demant A/S
AO World PLC	Pearson PLC	TX Group AG	Pernod Ricard SA	FLSmidth & Co A/S
Aston Martin Lagonda Global Holdings PLC	Playtech PLC	Valora Holding AG	Vilmorin & Cie SA	GN Store Nord A/S
ASOS PLC	Purplebricks Group PLC	VAT Group AG	Hermes International SCA	H Lundbeck A/S
Auto Trader Group PLC	PZ Cussons PLC	Vifor Pharma AG	SEB SA	Netcompany Group A/S
Avon Rubber PLC	Reach PLC	Ypsomed Holding AG	Soitec SA	Nilfisk Holding A/S
Avast PLC	Relx PLC	Zehnder Group AG	Sopra Steria Group SA	NNIT A/S
AVEVA Group PLC	RHI Magnesita NV	SKF AB	STMicroelectronics NV	Novo Nordisk A/S
AstraZeneca PLC	Royal Mail PLC	Alfa Laval AB	Sodexo SA	Novozymes A/S
Babcock International Group PLC	Rotork PLC	Alimak Group AB	Technicolor SA	Pandora A/S
A.G.Barr PLC	Renishaw PLC	Ambea AB	Tarkett SA	Royal Unibrew A/S
Bakkavor Group Plc	Speedy Hire PLC	Assa Abloy AB	Ubisoft Entertainment SA	Schouw & Co A/S
Boohoo Group PLC	Severfield PLC	Atlas Copco AB	Veolia Environnement SA	Simcorp A/S
Bodycote PLC	SIG PLC	Axfood AB	Vivendi SE	Vestas Wind Systems A/S
BP PLC	Signature Aviation PLC	Betsson AB	Vallourec SA	Basware Oyj
Burberry Group PLC	DS Smith PLC	BioArctic AB	Valneva SE	Cargotec Corp
Breedon Group PLC	Smiths Group PLC	Biotage AB	Acea SpA	Finnair Plc
Britvic PLC	WH Smith PLC	Boliden AB	Aeroporto Guglielmo Marconi di Bologna SpA	F-Secure Oyj
N Brown Group PLC	Senior PLC	Camurus AB	Autogrill SpA	Huhtamaki Oyj
Central Asia Metals PLC	Spirent Communications plc	Catena Media PLC	Biesse SpA	Kesko Oyj
Coca Cola HBC AG	SSE PLC	Clas Ohlson AB	Davide Campari Milano NV	Lehto Group Oyj
Carnival PLC	Stock Spirits Group PLC	Coor Service Management Holding AB	Carel Industries SpA	Metso Outotec Corp
C&C Group PLC	Studio Retail Group PLC	Dometic Group AB	Datalogic SpA	Neles Oyj
Circassia Group PLC	Spectris PLC	Elanders AB	Danieli & C Officine Meccaniche SpA	Nokia Oyj
Coats Group PLC	Synthomer PLC	Electrolux AB	DiaSorin SpA	Oriola Oyj
Croda International PLC	Ted Baker PLC	Elekta AB	Digital360 SpA	Orion Oyj
CRH PLC	Telecom Plus PLC	Eltel AB	De' Longhi SpA	Outokumpu Oyj
Craneware PLC	TI Fluid Systems PLC	Eniro AB	Aquafil SpA	Ponsse Oyj
ConvaTec Group PLC	Travis Perkins PLC	Epiroc AB	El En SpA	Sanoma Oyj
Cranswick PLC	Topps Tiles PLC	Essity AB	ERG SpA	TietoEVRY Corp
DCC PLC	UDG Healthcare plc	Fingerprint Cards AB	FILA Fabbrica Italiana Lapis ed Affini SpA	Tikkurila Oyj
DFS Furniture PLC	Ultra Electronics Holdings PLC	Getinge AB	Italgas SpA	Nokian Tyres plc
Diageo PLC	Unilever PLC	H & M Hennes & Mauritz AB	Immobiliare Grande Distribuzione SIIQ SpA	Uponor Oyj
Daily Mail and General Trust P L C	Victoria PLC	Haldex AB	Interpump Group SpA	Valmet Oyj
Dunelm Group PLC	Victrex PLC	Hexagon AB	Juventus FC SpA	Agrana Beteiligungs AG
Domino's Pizza Group PLC	Vectura Group PLC	Hexpol AB	Moncler SpA	Andritz AG
Dechra Pharmaceuticals PLC	Vodafone Group PLC	Humana AB	Maire Tecnimont SpA	AT & S Austria Technologie & Systemtechnik AG
Diploma PLC	Weir Group PLC	Husqvarna AB	OVS SpA	DO & CO AG
Devro PLC	Naked Wines PLC	ICA Gruppen AB	Pirelli & C SpA	EVN AG
Electrocomponents PLC	Wizz Air Holdings PLC	Indutrade AB	Recordati Industria Chimica e Farmaceutica SpA	FACC AG
Elementis PLC	Whitbread PLC	Kambi Group PLC	Salcef Group SpA	Flughafen Wien AG
Energean PLC	Aryzta AG	LeoVegas AB	Sesa SpA	Kapsch Trafficcom AG
Entain PLC	Autoneum Holding AG	Lindab International AB	Salvatore Ferragamo SpA	Mayr Melnhof Karton AG
Euromoney Institutional Investor PLC	Bachem Holding AG	Mips AB	Saes Getters SpA	OMV AG
EVRAZ plc	Barry Callebaut AG	Modern Times Group MTG AB	Saipem SpA	Palfinger AG
Experian PLC	Belimo Holding AG	Munters Group AB	Tiscali SpA	Porr AG
Easyjet PLC	Bell Food Group AG	Nederman Holding AB	Carlsberg A/S	Oesterreichische Post AG
Frontier Developments PLC	Bossard Holding AG	Nibe Industrier AB	Telecom Italia SpA	Semperit AG Holding
First Derivatives PLC	Bucher Industries AG	Nobia AB	Tinexta SpA	Telekom Austria AG
Flutter Entertainment PLC	Burckhardt Compression Holding AG	Nolato AB	Atresmedia Corporacion de Medios de Comunicacion SA	Verbund AG
Forterra PLC	Burkhalter Holding AG	Paradox Interactive AB	Aena SME SA	voestalpine AG
Future PLC	Coltene Holding AG	RaySearch Laboratories AB	Distribuidora Internacional de Alimentacion SA	Wienerberger AG
Ferrexpo PLC	Comet Holding AG	Saab AB	Ebro Foods SA	Zumtobel Group AG
Games Workshop Group PLC	Daetwyler Holding AG	Sandvik AB	eDreams Odigeo SA	Anheuser Busch Inbev SA
Greencore Group PLC	DKSH Holding AG	SAS AB	Ence Energia y Celulosa SA	AGFA Gevaert NV
Genus PLC	Dormakaba Holding AG	SSAB AB	Enagas SA	argenx SE
Greggs PLC	Dufry AG	Swedish Match AB	Faes Farma SA	Barco NV
GlaxoSmithKline PLC	Emmi AG	Tele2 AB	Fluidra SA	NV Bekaert SA
Headlam Group PLC	Feintool International Holding AG	Telia Company AB	Grifols SA	Ion Beam Applications SA
Halfords Group PLC	Flughafen Zuerich AG	Thule Group AB	Grupo Empresarial San Jose SA	Kinepolis Group NV
Hilton Food Group PLC	Forbo Holding AG	Tobii AB	Industria de Diseno Textil SA	Melexis NV
Hikma Pharmaceuticals PLC	Galenica AG	Trelleborg AB	Lar Espana Real Estate SOCIMI SA	Mithra Pharmaceuticals SA
Hill & Smith Holdings PLC	Geberit AG	Troax Group AB	Naturgy Energy Group SA	Montea Comm VA
Halma PLC	Georg Fischer AG	VBG Group AB	Pharma Mar SA	Orange Belgium SA
Hochschild Mining PLC	Givaudan SA	Volvo AB	Promotora de Informaciones SA	Ontex Group NV
Hunting PLC	Gurit Holding AG	AB Science SA	Repsol SA	Oxurion NV
Howden Joinery Group PLC	Huber+Suhner AG	Aeroports de Paris SA	Laboratorios Farmaceuticos ROVI SA	Proximus NV
International Consolidated Airlines Group SA	Interroll Holding AG	Air France KLM SA	Siemens Gamesa Renewable Energy SA	Recticel NV
Ibstock PLC	Jungfraubahn Holding AG	Airbus SE	Solarpack Corporacion Tecnologica SA	Tessenderlo Group NV
IG Design Group PLC	Kardex Holding AG	Alstom SA	Telefonica SA	Titan Cement International SA
Imperial Brands PLC	Kudelski SA	Biomerieux SA	Tecnicas Reunidas SA	Ucb SA
IMI PLC	LafargeHolcim Ltd	Boiron SA	Zardoya Otis SA	Corticeira Amorim SGPS SA
Inchcape PLC	Landis+Gyr Group AG	Bouygues SA	Aalberts NV	EDP Energias de Portugal SA
Indivior PLC	Lem Holding SA	Capgemini SE	Koninklijke Ahold Delhaize NV	EDP Renovaveis SA
iomart group PLC	Logitech International SA	Carrefour SA	Akzo Nobel NV	Galp Energia SGPS SA
Johnson Matthey PLC	Lonza Group AG	Casino Guichard Perrachon SA	Basic Fit NV	Jeronimo Martins SGPS SA
Kaz Minerals PLC	Medacta Group SA	CGG SA	B&S Group SA	Navigator Company SA
Strix Group PLC	Medartis Holding AG	Societe BIC SA	Corbion NV	NOS SGPS SA
Kingfisher PLC	Metall Zug AG	Danone SA	Koninklijke DSM NV	Pharol SGPS SA
Kenmare Resources PLC	Meyer Burger Technology AG	Christian Dior SE	Fugro NV	Ren Redes Energeticas Nacionais SGPS SA
Kainos Group PLC	Mobilezone Holding AG	CGG SA	Galapagos NV	Sonae SGPS SA
Lookers PLC	Nestle SA	Chargeurs SA	Grandvision NV	Adevinta ASA
McBride PLC	Novartis AG	DBV Technologies SA	Heineken NV	Norske Skog ASA
Micro Focus International PLC	OC Oerlikon Corporation AG Pfaeffikon	Sartorius Stedim Biotech SA	Heineken Holding NV	Norwegian Air Shuttle ASA
Mediclinic International PLC	Orior AG	Dassault Systemes SE	Intertrust NV	Orkla ASA
Marks and Spencer Group PLC	PIERER Mobility AG	EssilorLuxottica SA	Kendrion NV	REC Silicon ASA
Moneysupermarket.Com Group PLC	Schweiter Technologies AG	Europcar Mobility Group SA	Koninklijke KPN NV	Schibsted ASA
Melrose Industries PLC	Sensirion Holding AG	Valeo SA	Signify NV	Telenor ASA
Nostrum Oil & Gas PLC	SFS Group AG	Thales SA	Koninklijke Philips NV	Wallenius Wilhelmsen ASA
Next PLC	Siegfried Holding AG	Ipsen SA	Randstad NV	XXL ASA
Ocado Group PLC	Sika AG	Interparfums SA	SBM Offshore NV	Glanbia PLC
On The Beach Group PLC	Sonova Holding AG	LVMH Moet Hennessy Louis Vuitton SE	Just Eat Takeaway.com NV	Kingspan Group PLC
Oxford Instruments PLC	Straumann Holding AG	Compagnie Generale des Etablissements Michelin SCA	Wolters Kluwer NV	Origin Enterprises PLC
Pendragon PLC	Sulzer AG	Manitou BF SA	ALK-Abello A/S	Ryanair Holdings PLC
Pets at Home Group PLC	Swatch Group AG	L'Oreal SA	Ambu A/S	Smurfit Kappa Group PLC

This table presents all the companies that have been included in our sample.

References

Ambler, T., and A. Wilson. 2006. Problems of Stakeholder Theory. *Business Ethics: A European Review* 4:30–35. doi:10.1111/j.1467-8608.1995.tb00107.x.

Baltagi, B. 2008. *Econometric Analysis of Panel Data*. John Wiley & Sons. Google-Books-ID: oQdx_70Xmy0C.

Berg, F., J. F. Kölbel, and R. Rigobon. 2020. Aggregate Confusion: The Divergence of ESG Ratings. SSRN Scholarly Paper ID 3438533, Social Science Research Network, Rochester, NY. doi:10.2139/ssrn.3438533.

Bloomberg. 2018. Fingerprint Cards Comes Full Circle as All 2015 Gains Now Erased. *Bloomberg.com* .

Bragdon, J. H., and J. A. T. Marlin. 1972. ls Pollution Profitable? Environmental Virtue and Reward: Must Stiffer Pollution Controls Hurt Profits? - Global LAMP Index®.

Breusch, T., and A. Pagan. 1980. The Lagrange Multiplier Test and its Applications to Model Specifications in Econometrics. *Review of Economic Studies* 47:239. doi:10.2307/2297111. Publisher: Oxford University Press / USA.

Brooks, C. 2014. *Introductory econometrics for finance*. 3rd ed. Cambridge: Cambridge Univeristy Press.

Chase, C. W. 2013. *Demand-Driven Forecasting: A Structured Approach to Forecasting*. Somerset, UNITED STATES: John Wiley & Sons, Incorporated.

Choi, J., and H. Wang. 2009. Stakeholder relations and the persistence of corporate financial performance. *Strategic Management*

Journal 30:895–907. doi:https://doi.org/10.1002/smj.759. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/smj.759.

Fischer, T. M., and A. A. Sawczyn. 2013. The relationship between corporate social performance and corporate financial performance and the role of innovation: evidence from German listed firms. *Journal of Management Control* 24:27–52. doi:http://dx.doi.org.ezproxy.library.bi.no/10.1007/s00187-013-0171-5. Num Pages: 27-52 Place: Heidelberg, Netherlands Publisher: Springer Nature B.V.

Freeman, R., and J. Mcvea. 2001. A Stakeholder Approach to Strategic Management. *SSRN Electronic Journal* doi:10.2139/ssrn.263511.

Friedman, M. 1970. A Friedman doctrine The Social Responsibility Of Business Is to Increase Its Profits - The New York Times. *The New York Times* p. 11.

Hartzmark, S. M., and A. B. Sussman. 2019. Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows. *The Journal of Finance* 74:2789–2837. doi:https://doi.org/10.1111/jofi.12841. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.12841.

Hausman, J. A. 1978. Specification Tests in Econometrics. *Econometrica* 46:1251–1271. doi:10.2307/1913827. Publisher: [Wiley, Econometric Society].

Hong, H., and M. Kacperczyk. 2009. The price of sin: The effects of social norms on markets. *Journal of Financial Economics* 93:15–36. doi:10.1016/j.jfineco.2008.09.001.

Kotsantonis, S., and G. Serafeim. 2019. Four Things No One Will Tell You About ESG Data. *Journal of Applied Corporate Finance* 31:50–58. doi:https://doi.org/10.1111/jacf.12346. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jacf.12346.

Lobe, S., and C. Walkshaeusl. 2016. Vice versus virtue investing around the world. *Review of Managerial Science* 10:303–344. doi:10.1007/s11846-014-0147-3. Place: Heidelberg Publisher: Springer Heidelberg WOS:000370873900005.

Margolis, J. D., H. A. Elfenbein, and J. P. Walsh. 2009. Does it Pay to Be Good...And Does it Matter? A Meta-Analysis of the Relationship between Corporate Social and Financial Performance. SSRN Scholarly Paper ID 1866371, Social Science Research Network, Rochester, NY. doi:10.2139/ssrn.1866371.

Melo, T. 2012. Slack-resources hypothesis: A critical analysis under a multidimensional approach to corporate social performance. *Social Responsibility Journal* 8. doi:10.1108/17471111211234879.

Morningstar. 2019. How Did European ESG Funds Stack up in 2018?

Moskowitz, M. R. 1972. Choosing Socially Responsible Stocks. *Business & Society Review (08934398)* p. 71. Publisher: Wiley-Blackwell.

Orlitzky, M., F. L. Schmidt, and S. L. Rynes. 2003. Corporate Social and Financial Performance: A Meta-Analysis. *Organization Studies* 24:403–441. doi:10.1177/0170840603024003910. Publisher: SAGE Publications Ltd.

Ortas, E., I. Álvarez, and A. Garayar. 2015. The Environmental, Social, Governance, and Financial Performance Effects on Companies that Adopt the United Nations Global Compact. *Sustainability* 7:1932–1956. doi:10.3390/su7021932.

Prahalad, C. K., and G. Hamel. 1994. Strategy as a Field of Study: Why Search for a New Paradigm? *Strategic Management Journal* 15:5–16. Publisher: Wiley.

Refinitiv. 2021. Environmental, Social and Governance (ESG) Scores from Refinitiv p. 25.

Reuters. 2017a. BRIEF-Pharol seeks to keep high stake in Oi in restructuring -CEO. *Reuters* .

Reuters. 2017b. Oi creditors approve largest-ever Latin American restructuring. *Reuters* .

Scholtens, B. 2008. A note on the interaction between corporate social responsibility and financial performance. *Ecological Economics* 68:46–55. doi:10.1016/j.ecolecon.2008.01.024.

Smith, N. C. 2003. Corporate Social Responsibility: Not Whether, But How. *California Management Review* 45:52–76. doi:10.2307/41166188.

Vance, S. G. 1975. Are Socially Responsible Corporations Good Investment Risks? *Management Review* 64:18. Publisher: American Management Association.

Velte, P. 2017. Does ESG performance have an impact on financial performance? Evidence from Germany. *Journal of Global Responsibility* 8:169–178. doi:10.1108/JGR-11-2016-0029.

Waddock, S. A., and S. B. Graves. 1997. The Corporate Social Performance–Financial Performance Link. *Strategic Management Journal* 18:303–319. doi:10.1002/(SICI)1097-0266(199704)18:4<303::AID-SMJ869>3.0.CO;2-G. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291097-0266%28199704%2918%3A4%3C303%3A%3AAID-SMJ869%3E3.0.CO%3B2-G.

Wooldridge, J. M. 2010. *Econometric analysis of cross section and panel data.* 2nd ed. Massachusetts Institute of Technology.