What Matters in ESG Investing?

Exploring the underlying drivers of ESG and their impact on abnormal returns

Master's Thesis
Copenhagen Business School

Jakob Ravn Hansen - 91875
M.Sc. Applied Economics & Finance

Tobias Qwist Fenger - 93701
M.Sc. Finance & Strategic Management

Supervisor: Finn Lauritzen

Date of submission: 15th of May 2019

Number of characters: 244,121
Number of standard pages: 107

Jakob Ravn Hansen
Tobias Qwist Fenger

May 15, 2019

Abstract

This thesis examines the relationship between environmental, social and governance (ESG) factors and abnormal returns and, specifically, which underlying drivers of ESG that are most important. Regression analyses are conducted with ESG data from Thomson Reuter's Asset4 database, using a sample of US and European firms in three time periods within 2003-2017. We find that 11 ESG variables are significantly associated with abnormal returns, of which eight ESG variables are positively associated. There is a regional divergence in the significance of ESG variables, as no variables are significant both in the US and Europe in the same time period. Additionally, there is no temporal persistence in ESG variables, as no variables are significant in consecutive time periods. It is also tested whether investors could have utilized ESG information to form portfolios with abnormal returns. Two of the 22 tested portfolios obtained significant abnormal returns, which calls for more sophisticated methods in leveraging ESG data to construct portfolios. The findings of this thesis underscore that both academic researchers and investors can benefit from decomposing ESG scores in examining the association with abnormal returns. The decomposition of ESG scores is useful in understanding the underlying drivers in the relationship between ESG and abnormal returns.

Table of Contents

1. Introduction

The world is facing monumental challenges, ranging from climate change, economic inequality, social injustice and more. These issues span the globe and businesses are affected to varying degrees by them. Increasingly, businesses are being pressured to become part of the solution instead of being part of the problem. However, trust in business among the general public is low. In a survey of 33,000 global respondents, only 52% trusted businesses to "do what is right" (Harrington, 2017). Firms are trying to regain trust by launching environmental, social and governance (ESG) initiatives, but do these initiatives create value for firms' investors in addition to its stakeholders? Specifically, is high ESG performance associated with abnormal returns? Many investors seem to think so. A survey by KPMG International in 2018 among 900 executives and board members uncovered that more than one-third had increased their focus on ESG factors as a result of pressure from investors (KPMG, 2018). Why are investors driving firms to change? From an economic perspective, ESG investing incorporates non-financial dimensions to improve investment performance (Duuren, Plantinga, & Scholtens, 2016). ESG factors are thus expected to either increase the investment return or reduce the investment risk. Higher investment returns could accrue from ESG factors correlating with firm-specific factors known to induce superior returns. For example, Duuren, Platinga and Scholtens (2016) speculate that a high governance score is associated with high-quality management and that high social and environmental scores correlate with strong strategic planning capabilities. Regarding lower investment risk, firms with a high ESG score might be less susceptible to corporate scandals. Sassen et al. (2016) find that firms with high ESG scores receive positive moral capital among its stakeholders, generating "insurance-like" protection for the company. From an ethical perspective, ESG investing nudges firms towards becoming more socially responsible members of society (Kell, 2018). This can also improve firm performance by strengthening relationships with key stakeholders (Renneboog, Horst, & Zhang, 2008a). There are thus numerous ways ESG factors can improve firm performance and, by extension, investor returns.

On the other hand, there are also counter-arguments against both the economic and ethical perspectives, suggesting that investors' focus on ESG is detrimental to firm performance. From the ethical perspective, traditional economic theory emphasizes that the primary duty of organizations' management is to maximize shareholder value by increasing the value of its stock (Shank, Manullang, & Hill, 2005). This notion raises the question of how social concerns are balanced against the need

to create value for shareholders. Friedman (1970) argues that social concerns should be of no interest to management. Their sole responsibility is to increase profits, as long as they conform to the rules and regulations of society. From an economic perspective, focusing on non-profit goals can cause management to destroy value for shareholders. Non-profit performance is arguably harder to evaluate than profit-oriented performance. Bundling non-profit and profit-oriented performance thus makes it harder to accurately evaluate management's performance, which reduces accountability and increases agency costs (Renneboog, Horst, & Zhang, 2008a).

This discussion highlights that there are theoretical arguments both for and against a positive effect of ESG on investor returns. Empirically, there is a wealth of literature on the relationship between ESG and investor returns. A meta-study concluded that half of the studies found a positive relationship, 40% found no or a mixed relationship and 10% found a negative relationship (Friede, Busch, & Bassen, 2015). Given this empirical ambiguity, how does this thesis add value to the existing literature? We suggest two ways to create more clarity in the relationship between ESG and investor returns. First, breaking down an overall ESG score into its components could reveal important underlying drivers of ESG and abnormal returns. ESG is often aggregated to a combined ESG score in the literature, which might conceal important information. A more nuanced picture of how ESG affects investor returns might emerge by decomposing firms' ESG scores. Second, valuable insights could be uncovered by comparing the effect of ESG in different regions. The literature often uses one region rather than several to study the financial impact of ESG. Duuren et al. (2016) found that US investors were much more skeptical than European investors about the positive effects of ESG on investment performance. This suggests that the association of ESG variables with investor returns might diverge between the US and Europe.

The over-arching purpose of this thesis to clarify which specific ESG factors that are important in relation to investor returns and whether investors could exploit ESG factors to earn abnormal returns. Investigating different aspects of how ESG influences investor returns will hopefully contribute with much-needed clarity to both academic researchers, investors and firms. Academic researchers could uncover new insights about which components of ESG to focus on. Investors would get an indication of which ESG factors to focus on in the investment process. Lastly, firms benefit from knowing which ESG factors that are rewarded by capital markets when prioritizing ESG investments. The following section will translate the purpose of this thesis into a research question that will guide our approach.

2. Research Question

The research question will be the core of this thesis and thus serve as a guide for our research. The aim of this thesis is to examine the relationship between ESG performance and abnormal returns across several markets and time periods. A research question should further be specific, measurable and relevant (Mourougan & Sethuraman, 2017), which leads to the following question:

What was the relationship between ESG factors and abnormal returns in the US and Europe during 2003-2017? Moreover, could investors have exploited this potential relationship to construct portfolios with abnormal returns?

The above research question will act as a pointer for the remainder of the thesis and will be supported by several sub-questions. The aim of the sub-questions is to create structure and coherent flow throughout the paper, resulting in a comprehensive answer to the research question. The first sub-question will help examine the possible effects of ESG performance. The question will guide an analysis of existing theory, which will provide the arguments about the relationship between ESG and investor returns:

1) *Which theoretical arguments can explain the relationship between ESG factors and abnormal returns?*

Enhanced knowledge about the potential implications of ESG performance will enable us to form hypotheses about the relationship between ESG and investor returns. Hypotheses development will provide a clear direction of the thesis and make the research question directly testable (Mourougan & Sethuraman, 2017). Having testable hypotheses allows us to analyze the gathered data and report the results of our regression analysis. This analysis will be guided by the second sub-question:

2) *What was the relationship between specific ESG variables and abnormal returns from 2003-2017?*

The above question will guide the first part of our analysis, which aims at determining a potential relationship between ESG factors and abnormal returns, defined as alpha. This allows a second analysis that aims at determining whether the results of the regression analyses could have been exploited by investors. The third sub-question will thus help illustrate the practical implications of our analysis:

3) *Could investors have exploited information about significant ESG variables to construct portfolios with abnormal returns?*

After performing the second part of our analysis, it is important to relate the findings to our theoretical expectations to help explain the results. The last sub-question will thus support a discussion of the results based on previously defined theories. The question will further aim at determining the potential implications of our findings.

4) *Which theories can explain the results of our analyses, and what are the implications of our findings?*

2.1 Delimitation

The research question has provided a clear direction for the thesis, but it is important to specify what is inside and outside of scope. This thesis will only evaluate the long-term effect of ESG performance, and any short-term impact will thus not be clearly observed. ESG initiatives affect companies over different time-horizons, which will be assessed in further detail in the following 'Theoretical Overview' section. Certain environmental initiatives may have an immediate impact on investor return due to a direct effect such as cost reductions or a signaling effect. Other initiatives may only have an effect on long-term investor returns, creating value by establishing legitimacy among stakeholders, which may take several years. Additionally, this thesis aims at examining the relationship between ESG and investor returns, which implies that an investor perspective has been chosen. The literature illustrates that ESG performance can have a significant influence on firm performance by affecting the profitability of companies. However, this paper will not evaluate firm performance by examining accounting performance measures directly. Instead, the paper will establish the importance of ESG for investors, which of course will be influenced by the effect on firm performance. Firm performance and investor return are correlated, but several factors influence investor return and not firm performance, which means that the methodology of the two analyses differs. The abnormal return of investors, measured by alpha, is thus the focus of this thesis.

It is important to define what ESG investing is prior to examining the relationship between ESG and alpha. Additionally, understanding the historical context of how ESG investing emerged from socially responsible investing (SRI), and the drivers behind its emergence, is useful to understand the nature of ESG investing today. The following section will address these aspects by outlining the historical development of SRI and how ESG investing emerged.

3. Socially Responsible Investing

SRI can be categorized as investment processes that integrate social, environmental and ethical considerations into the investment decision-making process (Renneboog, Horst, & Zhang, 2008a). SRI is not a new phenomenon, as individuals have included ethical considerations in their investment strategies for centuries. Traditionally, ethical investing was based on religious traditions. In the 1920s, the church in the United Kingdom avoided investing in 'sinful' industries such as alcohol, gambling, tobacco and firearms (ibid.). Ethical investing is similarly bound to Islamic traditions, as Islamic investors have conventionally avoided companies involved in pork production, pornography and gambling. Modern SRI is increasingly based on personal values and convictions of individuals. In the 1940s, SRI became visible, as government agencies started to avoid investment opportunities in organizations that did not live up to certain standards regarding labor practices (Shank, Manullang, & Hill, 2005). SRI received further attention during the 1970s, as a direct consequence of the Vietnam War and the Apartheid regime in South Africa (Kell, 2018). In 1971, the Pax World Fund was established, which refrained from investing in companies that profited from the Vietnam War. Politics and human rights thus became influential factors in investment strategies during this period, and investors were suddenly confronted with the fact that investment choices could have significant social consequences. In 1986, the Chernobyl nuclear power plant in the former Soviet Union exploded, which resulted in environmental damages due to the effects of radioactive materials. Three years later, the oil tanker Exxon Valdez ran aground near Alaska and spilled more than 11 million gallons of crude oil into the ocean. These events made investors aware of the potential negative environmental consequences of industrial development, which started to affect investment decisions (Renneboog, Horst, & Zhang, 2008a). Furthermore, a series of corporate scandals at the end of the 1990s and early 2000s turned corporate governance and corporate transparency into a focal point for SRI investors (ibid.). This overview illustrates that the focus of SRI has evolved over time to include various factors in investment decisions.

Recently, investors have started to integrate ESG factors into their investment decisions, and ESG investing has become the new norm for socially responsible investing (Kell, 2018). ESG investing began in 2004 when the former UN Secretary, Kofi Annan, invited CEOs of major financial institutions to participate in a joint initiative under the UN Global Compact. A year later the initiative produced a report entitled "Who Cares Wins", which reported that ESG factors can increase firm

profitability, result in more sustainable markets and have superior outcomes for societies. ESG cover a wide range of issues that traditionally have not been part of financial analysis yet may have financial relevance. This may include how corporations respond to climate change, how they deal with resource management, which governance policies they have in place and how they educate and train their employees (ibid.). ESG investing is thus based on the assumption that ESG factors have financial relevance, which has not always been the primary focus of previous SRI. ESG investing has thus arguably changed SRI from a charitable phenomenon to a full-fledged investment alternative for both the institutional and the individual investor.

In 2006, the Principles for Responsible Investment (PRI) was launched at the New York Stock exchange to encourage the incorporation of ESG issues into investment analysis (United Nations, 2016). In particular, the PRI works with an international network of asset managers to enforce six principles of responsible investing, which are listed in Appendix 1. The principles illustrate that signatories commit to incorporate and promote ESG factors into the investment industry wherever it is consistent with fiduciary responsibilities. The mission of PRI is thus to support an economically efficient and sustainable global financial system that results in long-term value creation. PRI has since its establishment in 2006 evolved and had more than 1600 global members at the end of 2018 representing over $70 trillion assets under management (AUM) (Kell, 2018). This does not imply that every signatory incorporates ESG factors into every investment decision, but it certainly supports that ESG AUM has accelerated significantly in recent years. The US Social Investment Forum keeps a record of all the ESG AUM in the US. Total US-domiciled AUM applying ESG strategies grew from $8.7 trillion at the start of 2016 to $12 trillion at the start of 2018, which is equivalent to a 38% increase. The same number was mere $0.6 trillion in 1995, which further highlights the rapid growth of ESG investments (SIF, 2018). A similar trend is apparent in Europe, where ESG AUM has increased at an annual rate of 27% in the last two years to reach €23.5 trillion at the start of 2018 (Eurosif, 2018). The US Social Investment Forum further states that about fifth of all funds under professional management screens on ESG criteria in 2018, which was up from a ninth in 2012 (SIF, 2018). This illustrates that ESG investing constitutes a large and rapidly increasing part of all investing.

This raises the questions of whether investors actually benefit from incorporating ESG variables into their investment processes. Specifically, can ESG investing result in abnormal returns for investors?

In contrast to previous SRI strategies, ESG investing is based on the assumption that ESG factors have financial relevance and thus impact investor returns. The following section will provide a theoretical overview of the central theories to understand how ESG factors can influence investor returns.

4. Theoretical Overview

This section will provide a theoretical overview of the relevant theories that can help explain the results of the analyses in this thesis. Some of the theories are appropriate to explain the results of the primary analysis that examines the relationship between ESG factors and firm alpha. Others are relevant for the secondary analysis, which examines whether knowledge of ESG factor can be utilized by investors to form portfolios with a significant alpha. The following theories will thus constitute the theoretical foundation of the 'Discussion' section.

4.1 Institutional Theory

This thesis will examine the relationship between ESG factors in both Europe and the United States. The reason for analyzing the effect of ESG factors in both of these regions is that there are historical, cultural and regulative differences, which might create differences in results. Institutional theory argues that organizations and their strategies are significantly influenced by the institutional setting in which they operate (Doh & Guay, 2006). Accordingly, organizations are affected by the history, culture, norms and regulation of their specific country or region. Scott (1995) defines three institutional pillars, namely the regulatory pillar, the normative pillar and the cognitive-cultural pillar. The paper argues that these pillars induce actors in similar environmental settings to become and remain similar. The paper further contends that the pillars induce actors in dissimilar environments to act differently, which is of great importance for this thesis. The regulatory pillar consists of rules and regulations that promote a certain kind of behavior while it restricts other kinds of behavior. This could be a specific ESG regulation that requires organizations to disclose certain environmental, social or governance initiatives. The regulatory pillar is enforced by coercive means (ibid.). The normative pillar consists of the norms and habits that are shared by individuals in the environment. Individual and collective actions are highly influenced by normative processes even though these processes are not made explicit in the same way as laws and regulations. Instead, the normative pillar is governed through a tacit agreement of what is right and wrong. Lastly, the cognitive-cultural pillar is a shared perception of what is typical or taken for granted (ibid.). This includes cultural differences, which are enforced through mimicking the actions of other actors. The three institutional pillars create the structure that provides meaning and stability to social behavior, and actors are induced to conform due to isomorphic pressures (Dimaggio & Powell, 1983). Institutional theory is useful when

analyzing similarities within and differences between environments, which will be done in this thesis. Accordingly, the differences in the results between Europe and the United States can be partly explained by discussing institutional differences.

4.2 Shareholder Theory

Shareholder theory can be used to explain the potential impact of ESG and how specific ESG initiatives can both drive and diminish firm value. Shareholder theory was originally proposed by Friedman (1970). The theory states that the sole responsibility of managers is to increase profits, as long as it conforms to the basic rules and regulations of society. The theory is based on the premise that management is hired as an agent of the shareholders to run the firm on their behalf. Consequently, management is both legally and morally obliged to serve the interests of the shareholders. Friedman argues that when executives are investing in ESG initiatives, they are basically spending other peoples' money for general social interest and are thus not serving as agents of the shareholders. As a result, the executive becomes a public employee instead of a company employee (ibid.). According to shareholder theory, if ESG initiatives lower firms' profits due to compromises with other stakeholders, then firms should not implement these initiatives. It is more efficient to charge lower prices and allow consumers to make their own charitable contributions and let governments deal with the provision of public goods and the existence of externalities (Renneboog, Horst, & Zhang, 2008a). Agency theory suggests that a principal-agent problem naturally exists, as the incentives of shareholders and managers often diverge (Fama & Jensen, 1983). Managers may thus derive a higher level of utility by pursuing strategies that do not maximize shareholder profits, such as ESG increasing initiatives. Nyberg et al. (2010) state that one possible solution to the principal-agent problem is to align the incentives of shareholders and management by an appropriate compensation structure. The principal-agent problem diminishes if the compensation structures are designed in a way that incentivizes management to maximize shareholder value.

Shareholder theory has received significant criticisms in literature, as it fails to answer how social concerns are balanced against the need to create shareholder value (Shank, Manullang, & Hill, 2005). The social welfare theorems supported by Adam Smith's theory of the "invisible hand" states that there should be no conflict between social welfare and value maximization of firms (Renneboog, Horst, & Zhang, 2008a). The basic idea is that any competitive equilibrium results in a Pareto efficient allocation of resources and thus a social optimum. However, in reality, it can be difficult to reach a

social optimum, as the maximization of shareholder value often conflicts with the interests of other stakeholders such as employees, customers, governments and so forth. Jensen (2002) states that the existence of externalities can be one factor that creates a gap between shareholders and other stakeholders. Externalities arise when the benefits and costs of an agent's actions are affected by the actions of other agents. A common example is when a company maximizes profits by increasing pollution, which can have severe consequences for the surrounding environment (Renneboog, Horst, & Zhang, 2008a). Friedman (1970) argues that company management should only be concerned with the interest of the shareholders and thus not focus on the relationship with other stakeholders. However, a company's relationship with its other stakeholders may have a significant impact on the firm profitability according to stakeholder theory, which will be described in the section below.

4.3 Stakeholder theory

Stakeholder theory can be applied to explain the impact of ESG similarly to shareholder theory. However, the two theories diverge in several ways. The basic idea in stakeholder theory is that management must formulate and implement processes that please all groups that have a stake in the company (Freeman & McVea, 2001). The main task in this process is to manage the concerns and interests of both shareholders, employees, customers, communities and other groups in a way that ensures the long-term success and survival of the company. Stakeholder theory thus supports shareholder theory's aim of maximizing shareholder wealth, but it opposes the notion that this should be the only objective for management. Freeman and Mcvea (2001) state that a stakeholder approach is intended to provide a framework that is flexible enough to deal with environmental changes without requiring managers to adopt new strategic paradigms on a regular basis. A stakeholder approach is thus arguably proactive, rather than reactive, since management must understand the stakeholders that impact the firm in order to effectively change the strategic course. Management thus has a never-ending task of balancing multiple relationships and objectives (ibid.). Rennebog et al. (2008a) argue that high ESG performance is a consequence of a stakeholder focus, as ESG is inevitably linked to the relationship with certain stakeholder groups. However, which relationship should management prioritize to increase firm performance the most? Moreover, what ESG initiatives should be implemented to improve the relationship with the prioritized stakeholder group? Stakeholder theory is useful to explain the impact of ESG, but the theory is less useful to help determine which ESG initiatives to introduce. The theory of stakeholder identification and salience will be helpful for this purpose.

4.5 Stakeholder Identification and Salience

The above sections illustrate that both shareholder theory and stakeholder theory hold great explanatory power in the relationship between ESG and firm performance. Stakeholder theory is useful to explain how ESG initiatives can drive value, but it does not specify which initiatives will be most beneficial for firms to implement. Prioritizing the appropriate ESG initiatives will improve the relationship with key stakeholders and arguably drive value. Contrarily, spending significant resources on relatively irrelevant stakeholders will decrease firm value, which highlights the need for identifying salient stakeholders. Mitchell, Agle and Wood (1997) propose that stakeholders should be identified and assessed by their possession of three attributes: legitimacy, power and urgency. Each of these attributes will be examined, as they can be highly helpful in explaining the results of the analysis in later sections.

Suchman (1995) defines legitimacy as the perception that an entity's actions are desirable in a socially accepted system of values, norms and beliefs. A stakeholder is thus considered legitimate if it has legitimate standing in society or a legitimate claim on the firm. However, legitimacy works both ways, as firms also need to prove their legitimacy towards their stakeholders. The survival of a firm is highly dependent on whether the firm is able to operate within the norms of a given society and its stakeholders. The rules of legitimacy change, as societies are dynamic, and norms may change over time. This requires firms to continuously evaluate whether its operations are legitimate by examining the requirements of its legitimate stakeholder groups.

The paper by Mitchell, Agle and Wood (1997) argues that legitimacy needs to be paired with power in order to create authority and make stakeholders important. A stakeholder may have a legitimate standing in society, but unless it has the power to enforce its will, it should not achieve attention from company managers (ibid.). Mitchell, Agle and Wood (1997) define power as "the ability of those who possess the power to bring about the outcomes they desire". Power can come in different forms, which affect how it is utilized. Firstly, power can be coercive, which is based on physical resources of force, restraint and violence. Coercive power is highly driven by threats and fear is needed for its existence. Secondly, power can be utilitarian and based on material and financial resources. Stakeholders with essential resources will naturally require more attention from company management. Lastly, stakeholders can utilize normative power, which is based on symbolic resources

(ibid.). The access to resources that provides a stakeholder with power can vary, and power can thus be categorized as transitory similarly to legitimacy.

Legitimacy and power are important factors to consider among stakeholders, but they do not capture the full dynamics of stakeholder-manager relationships. The urgency factor can thus help move the analysis from a static state to a more dynamic state. Urgency is defined as "calling for immediate attention" (ibid.). This can either be because a matter is time-sensitive or because it is of great importance to an agent. Accordingly, the claim of specific stakeholders become more important to consider if the claim is considered urgent by the stakeholder. In situations of corporate scandals, the urgency of the mistreated stakeholders' claim increases the importance of an immediate reaction from management. The theoretical model by Mitchell, Agle and Wood (1997) is a tool that can help to determine the effect of ESG initiatives in specific situations, which is valuable in the discussion sections later in the thesis.

To this point, the theoretical overview has illustrated how different theories can explain the relationship between ESG and firm performance. According to shareholder theory, investing in ESG initiatives will likely decrease profitability. Stakeholder theory suggests that ESG initiatives can drive firm value and diminish firm risk by improving relationships with key stakeholders. Moreover, evaluating the appropriate ESG initiatives requires identifying and balancing various relationships with stakeholders. However, this thesis is examining the relationship between ESG and abnormal return and not firm performance. This thesis is thus concerned with how ESG affect the share price of companies, and further how ESG information can be utilized by investors. For this purpose, it is essential to address capital market theory, which will be done in the following section.

4.6 Capital Market Theory

Capital market theory is generally concerned with the trade-off between the expected returns of investors and the inherent risks involved, and the theory thus aims at pricing securities appropriately. This thesis studies the relationship between ESG factors and abnormal return. Consequently, it will be examined whether ESG factors can explain an abnormal return that is not accounted for by previous asset-pricing models. To perform such an analysis, it is essential to assess the theory on asset-pricing models, which will be done in the following sections.

CAPM

The Capital Assets Pricing Model (CAPM) is a single-factor model of risk and return. The model assumes that the investor's required return of an investment is linearly related to the asset's covariance with the market portfolio. There are three pivotal assumptions in the CAPM model (Sharpe, 1964):

1. Investors can buy and sell any security at competitive market prices and can borrow and lend at the risk-free rate
2. All investors choose a portfolio of securities that provide the highest possible expected return given the level of volatility they are willing to accept
3. Investors have homogenous expectations about the volatility, correlation, and expected return of securities

Some of the above assumptions may be unrealistic and, consequently, the model will not always hold. However, the model is still useful and widely applied by researchers as a performance measure for stocks and portfolios (Berk & DeMarzo, 2017). The standard version of the model was developed by Sharpe (1964) and can be written as the following equation:

$$E[R_{i,t}] = R_{f,t} + \beta_i(R_{m,t} - R_{f,t})$$

Where:

$E[R_{i,t}]$ is the expected return of asset i in period t
$R_{f,t}$ is the risk-free market rate in period t
β_i is the beta of asset i with respect to the market return
$R_{m,t}$ is the return of the market portfolio

Source: (Sharpe, 1964)

The model thus implies that the expected return of an asset is equal to the risk-free rate plus the beta times the market risk premium. Investors expect to be compensated for the time value of money and for risk. The risk-free rate accounts for the time value of money, while the other part of the formula accounts for risk. The beta of a security measures the security's sensitivity to market risk and is calculated using the following formula:

$$\beta_i = \frac{COV(R_{i,t}; R_{m,t})}{VAR(R_{m,t})}$$

Source: (Sharpe, 1964)

The expected return of a security is thus related to the security's covariance with the market portfolio. A security with a high covariance, that fluctuates more than the market, should thus yield a higher expected return than a security with low covariance. The beta of the market portfolio itself is equal to one according to the formula. The relationship between expected return and beta can be illustrated using the security market line (SML). Under the CAPM assumptions, the SML is the line along which all securities should lie when plotted according to their beta and expected return (Berk & DeMarzo, 2017).

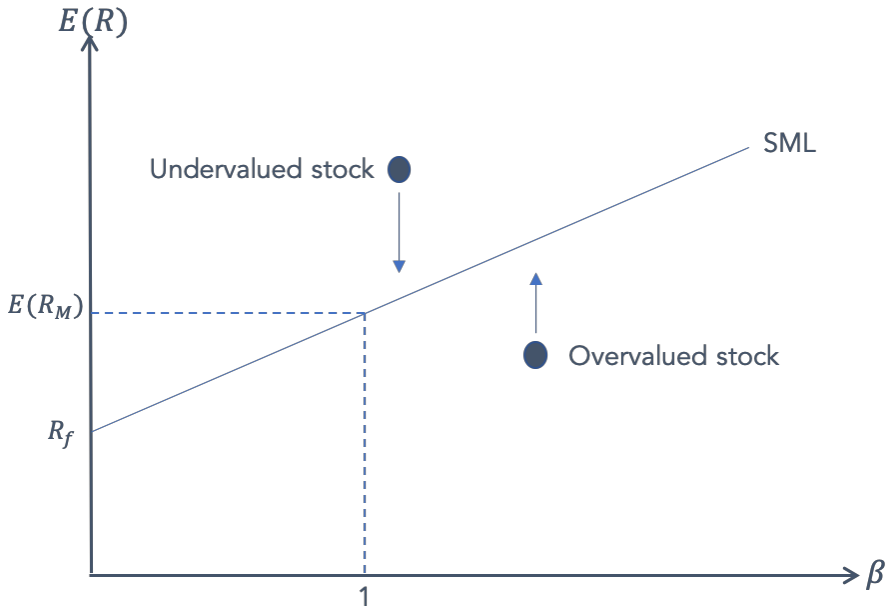


Figure 1; Source: (Sharpe, 1964)

In case a security is plotted above the SML, it implies that the security is undervalued. In an efficient market, investors will then increase their share of this security in their portfolios, which will increase its price and lower its expected return until the stock is back on the SML. Contrarily, the security will be overvalued if it is located below the SML. Investors of the overvalued security will shift their portfolios to contain less of the security, which will reduce its price and thereby increase its expected return to bring it back to the SML. The CAPM model is widely used in academic papers for performance measurement (Mănescu, 2011). Normally, papers perform statistical tests using the excess return form of the CAPM, which is given by the equation below:

$$R_{i,t} = a_i + \beta_i R_{m,t} + \varepsilon_{i,t}$$

Where:

$R_{i,t}$ is the excess return of the asset i in period t
a_i is the intercept, or alpha, of asset i
β_i is the beta of asset i with respect to the market
$R_{m,t}$ is the excess return of the market in period t
$\varepsilon_{i,t}$ is the error term of the asset i in period t

Source: (Mănescu, 2011)

The equation implies that the excess return is a function of an asset's beta with respect to the excess return of the market. In equilibrium, the intercept, denoted by alpha, of all assets should equal zero, and all variation in returns should be explained by asset betas (Berk & DeMarzo, 2017). The equation is similar when calculating the excess return of a portfolio instead of a single asset. The beta of the portfolio will then be a weighted average of all the betas of the assets in the portfolio. Similarly, the alpha of a portfolio is the weighted average alpha of the assets in the portfolio, and investors will thus benefit from buying stocks that yield a positive alpha and selling stocks that yield a negative alpha. According to the CAPM model, investors will not be able to get a constant alpha because markets are assumed to be efficient and undervalued stocks will rapidly return to the SML. However, investors may be able to obtain a consistent positive alpha in case markets are not efficient by taking advantage of idle information. Markets are efficient by order of degree and the following section will outline the different degrees of market efficiency.

Efficient market hypothesis

The efficient market hypothesis (EMH) was developed by Eugene Fama and states that security prices fully reflect all available information (Fama, 1970). Eugene Fama has also formulated the EMH as prices fully reflecting intrinsic values at every point in time (Fama, 1965). However, Eugene Fama recognizes that the EMH is an extremely strong hypothesis that is not expected to hold in all instances and that markets can be efficient by order of degree (Fama, 1970). To specify the degree to which a market is efficient, Eugene Fama broke down the EMH into three versions: weak, semi-strong and strong form. The weak form hypothesis states that stock prices reflect all market trading data such as past prices. If stock prices fully reflect historical stock prices, it should not be possible to create a profitable trading strategy based on past stock prices, which are widely available and nearly costless to obtain. The semi-strong hypothesis asserts that stock prices reflect all publicly available information. In addition to past prices, all firms' publicly available information is expected to be

reflected in stock prices. This includes financial and extra-financial information such as ESG performance. It should thus not be possible to consistently profit from trading on publicly available information such as firms past financial performance or ESG performance, if the semi-strong form hypothesis holds. Lastly, the strong form hypothesis states that stock prices reflect all public as well as private information. In other words, stock prices reflect all information, even if some information is only available to company insiders (Bodie, Kane, & Marcus, 2014). This would effectively imply that no one should be able to profit from insider trading, which is clearly not the case empirically. Eugene Fama also states that the strong from hypothesis should be viewed as a useful benchmark from which deviations from market efficiency can be tested (Fama, 1970).

Testing whether markets are efficient is difficult. In fact, Fama (1991) argues that market efficiency in itself cannot be tested. To know whether an information set is fully reflected in stock prices, a model based on the given information set must be specified and exhaustively explain the variation in stock prices. Once such a model has been specified, stock price deviations from the model's predictions can either result from markets not being efficient, the model being misspecified or both. It is often ambiguous whether the model is flawed, or the examined market is inefficient, which makes empirical research on market efficiency inconclusive. Nonetheless, these models seeking to explain variation in stock prices, called asset pricing models, can yield valuable insights into how stock prices behave. Eugene Fama is, along with Kenneth French, among the pioneers of constructing asset pricing models. The following section will explain how multi-factor asset pricing models have expanded upon the single-factor CAPM to better explain the behavior of stock prices.

Arbitrage pricing theory

Recent papers question the adequacy of the CAPM as a model that explains expected returns. Specifically, researchers argue that the beta alone may not suffice to explain stock returns and point towards arbitrage pricing theory for better accountability (Fama & French, 1996). The arbitrage pricing theory was developed by Ross (1976), as an alternative to the CAPM model to derive a relationship between risk and return. According to arbitrage pricing theory, the returns of an asset can be described by a factor model. A factor model is a model that describes the variance in returns as a function of two specific components; a systemic component and a firm-specific component. Below is the equation for a single-factor model:

$$R_{i,t} = E\left[R_{i,t}\right] + \beta_i F_t + \varepsilon_{i,t}$$

Where:

$R_{i,t}$ is the excess return of the asset i in period t
$E\left[R_{i,t}\right]$ is the expected return of asset i in period t
F_t is the deviation of the common factor from its expected value in period t
β_i is the sensitivity of firm i to that factor
$\varepsilon_{i,t}$ is a firm-specific risk component

Source: (Ross, 1976)

Arbitrage pricing theory assumes that markets sometimes misprice assets, before the market then corrects, and assets return to their fair value. Markets are thus not assumed to be completely efficient, which contradicts an assumption of the CAPM model. Using arbitrage pricing theory, investors hope to take advantage of deviations from fair market values to make a profit (Bodie, Kane, & Marcus, 2014). The above equation is an example of a single-factor model; however, it is more likely that excess return is expressed by more than one factor. For example, GDP and interest rate levels are two factors that could both influence the excess return of an asset and both factors could thus potentially be added to the model. The arbitrage theory model is thus more flexible than the CAPM model, but it is also more complex, as it may be difficult to determine which factors to include. Roll and Ross (1980) argue that there are only a few systemic components of risk existing in nature. Consequently, many portfolios are close substitutes and must have similar expected returns. The following sections will explain two of the most famous and empirically applied factors models, namely, Fama and French's three-factor and five-factor models.

The three-factor model

Fama and French propose a three-factor model, which is meant to explain a security's expected return. The model has become one of the dominant approaches for describing investor returns in academic papers (Renneboog, Horst, & Zhang, 2008a). According to the model, a security's expected return depends on the sensitivity of its return to the market return and the returns of two additional portfolios meant to mimic additional risk factors (Fama & French, 1996). The authors find that adding two systemic risk factors, namely the SMB and the HML factors, to the market factor helps the model explain a higher percentage of variability in returns. The model is expressed by the equation below:

$$R_{i,t} = \alpha_i + \beta_i R_{M,t} + s_i SMB_t + h_i HML_t + \varepsilon_{i,t}$$

Where:

$R_{i,t}$ is the return of the asset i in period t, in excess of the risk-free rate
α_i is the return not explained by any risk factor
β_i is the sensitivity of firm i to the market factor
$R_{M,t}$ is the return on the market factor in period t, in excess of the risk-free rate
s_i and h_i are the sensitivities of firm i to the SMB_t and HML_t factors, respectively
SMB_t and HML_t is the return on these factors in period t.
$\varepsilon_{i,t}$ is a firm-specific risk component

Source: (Fama & French, 1996)

SMB stands for 'small minus big' and is the difference between the returns of a portfolio of small stocks and a portfolio of big stocks. This factor seeks to incorporate a size effect, as empirical research supports that small stocks outperform large stocks (ibid.). HML is an acronym for high minus low, and it is the difference between the returns on a portfolio of high-book-to-market-equity (BE/ME) stocks and a contrasting portfolio of low BE/ME stocks. The HML factor seeks to incorporate that value stocks tend to outperform growth stocks (ibid.). Fama and French (1993) illustrate that these three factors capture much of the spread in the cross-section of average returns on portfolios, which is known to cause problems in the CAPM model. Since the publication of the paper introducing the three-factor model, the theoretical arguments for including the additional risk factors have been heavily debated in academic literature (Renneboog, Horst, & Zhang, 2008a). Some papers argue that the three-factor model has performed poorly in explaining returns when it has been applied to stocks in emerging markets. One problem is that the SMB factor behaves differently in emerging markets compared to developed markets. However, our thesis exclusively examines stocks from developed markets, meaning that this issue is less relevant for this thesis. Other papers provide evidence that much of the variation in returns that is related to profitability and investment is left unexplained by the three-factor model (Fama & French, 2015). Additional factors could thus be added to the model to improve its applicability.

Extending the three-factor model

One extension of the three-factor model was developed by Carhart (1997). This study argues that momentum is another factor that has a significant effect on excess stock returns. The study argues that stocks that have performed well in the past tend to outperform stocks that have performed poorly in the past. Carhart (1997) finds that the applicability of the three-factor model will increase by adding

the additional momentum-factor and making it a four-factor model. Supporters of behavioral finance argue that irrationality among investors is what causes the momentum-factor to persist. Generally, individuals believe that specific events are more likely to happen if they have happened in the near past (Munk, 2017). As a result, investors might underreact to stock related news, and the effect will not be incorporated in stock prices immediately. The four-factor model has been used by several papers that examine the relationship between ESG and abnormal returns (Halbritter & Dorfleitner, 2015). However, the four-factor model does not account for profitability and investment factors, which the original three-factor model was criticized for. An additional model was thus developed by Fama and French in 2015 to account for these factors.

Fama and French (2015) expanded upon their original paper from 1993 by extending the three-factor model with two additional factors. The model is given by the equation below:

$$R_{i,t} = \alpha_i + \beta_i R_{M,t} + s_i SMB_t + h_i HML_t + r_i RMW_t + c_i CMA_t + \varepsilon_{i,t}$$

Source: (Fama & French, 2015)

In the equation, RMW is the difference between returns on diversified portfolios of stocks with strong and weak profitability. CMA is the difference between returns on diversified portfolios of high and low investment firms, which can be classified as conservative and aggressive firms. Similar to the three-factor model, the alpha will equal zero if the coefficients of the risk factors capture all variation in expected returns. This knowledge can be used to test which model most accurately captures the variation in stock returns. Interestingly, Fama and French (2015) provide evidence that the five-factor model directed at capturing both size, value, profitability and investment patterns performs better than their previous three-factor model and the Carhart four-factor model. The authors add that the superior performance of the five-factor model is not sensitive to the way its factors are defined. This supports that the five-factor model could be more appropriate to examine excess returns. Roll and Ross (1980) state that arbitrage pricing theory is more complex than the CAPM since it can be difficult to determine which factors to include in the model. This thesis will thus apply both the CAPM and the five-factor model, which will be elaborated further in the ‘Methodology’ section.

The above sections have highlighted the models that are useful for pricing securities and portfolios. These models will be applied when computing the abnormal returns of firms in the later analysis. The

following section will examine portfolio choice theory, which is useful to evaluate the performance of portfolios. The section will thus be relevant for explaining the results of ESG portfolios in the 'Discussion' section.

Portfolio choice theory

The most well-known framework for choosing portfolios is based on mean-variance analysis by Markowitz (1952). Mean-variance analysis can be used as a tool to determine the optimal portfolio for a given set of rules about investor behavior. The two main rules about investor behavior are as follows. First, investors only consider the expected return and variance of the portfolios over a fixed time-horizon. Second, investors prefer as high expected return and as low variance as possible (Munk, 2017). An investor following these rules is called a mean-variance optimizer, who will only choose portfolios that are mean-variance efficient (ibid.). A portfolio is mean-variance efficient when it is not possible to choose another portfolio with a lower standard deviation and the same expected return. All mean-variance efficient portfolios lie on the upward sloping part of the efficient frontier, denoted 'frontier-risky' in Figure 2.

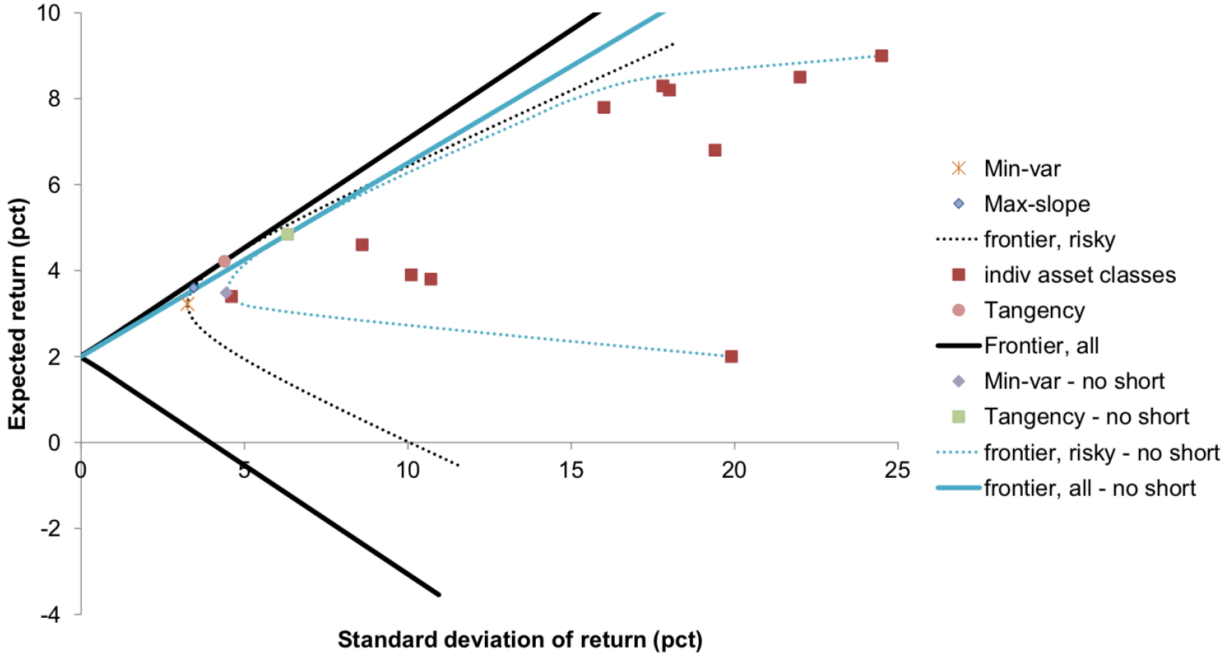


Figure 2; Source: (Munk, 2017)

The two-fund separation theorem states that all mean-variance efficient portfolios can be formed through some combination of the so-called minimum-variance and maximum-slope portfolios (ibid.).

The minimum-variance portfolio, denoted 'Min-var', is the unique combination of risky assets that yields the lowest possible variance (ibid.). In this context, a risky asset is any asset that is not risk-free. The variance of stocks comprises a firm-specific and a systemic risk component. Combining risky assets lowers the variance by reducing the firm-specific risk component due to the diversification benefit of combining imperfectly correlated risky assets (ibid.). The maximum-slope portfolio, denoted 'Max-slope', is the unique portfolio that yields the largest ratio of expected return to the standard deviation (ibid). By iterating a multitude of combinations of the minimum-variance and maximum slope portfolios, the efficient frontier of risky assets is traced out.

For the purposes of this thesis, the key insight from Figure 2 is the behavior of the efficient frontier when imposing constraints. One such constraint could be not allowing investors to short-sell assets, meaning that investors cannot borrow funds to sell stocks they do not own (ibid.). This no-short frontier is denoted by 'frontier, risky – no short'. It is evident from Figure 2 that the efficient frontier shrinks after imposing the no-short constraint. In fact, any constraint imposed on the efficient frontier will shrink it or at best leave it unaffected, which implies that the constrained frontier has an equal or lower return for a given standard deviation compared to the unconstrained efficient frontier. In other words, investors are either unaffected or worse-off with a portfolio on the constrained efficient frontier. The reason is that the minimum-variance and maximum-slope portfolios utilize the covariances between the risky assets to create the two portfolios with the lowest variance and highest return-risk trade-off, respectively. Any constraint that prohibits the optimal utilization of the stocks' covariances will shrink the efficient frontier, leaving investors worse-off. Therefore, all mean-variance optimizing investors would prefer a portfolio on the unconstrained efficient frontier compared to a constrained efficient frontier. The mean-variance analysis is important to consider in this thesis, as utilizing ESG information to screen firms implies constraining the efficient frontier. The theory is thus useful to explain the results of the portfolio analysis in the 'Discussion' section.

The theoretical section has now presented the most relevant theories that allow us to explain the results of the regression analyses in later sections. These theories can thus help explain the relationship between ESG and abnormal returns. The theoretical overview explains the reasons for and against investors being able to utilize ESG information to create portfolios with significant, positive alphas.

4.7 Theoretical Arguments in the Debate

The presented theories can help guide the research methodology and explain the results of the analyses. The main arguments of the theories, and how they can be utilized to answer the research question of the thesis, will be presented below.

The net benefit argument

Stakeholder theory has illustrated that there can be benefits associated with ESG initiatives, which can increase company revenue. However, introducing ESG initiatives can also have significant costs and the investor return will thus be affected by the net effect (Mănescu, 2011). ESG will impact firm performance positively, in case the benefits of ESG outweigh the costs of implementation. However, if the costs of implementing ESG initiatives exceed the incremental company revenue, then firm performance will be impacted negatively. Investor return is highly influenced by firm performance, which illustrates the importance of the net benefit argument. This thesis examines the relationship between ESG and abnormal returns, and a positive net effect on firm performance will not necessarily provide abnormal returns. This leads to the second argument concerning the mispricing of securities, which will be presented below.

The mispricing argument

The net benefit of ESG may be fully reflected in the share prices of companies. In this case, it will not be possible for investors to earn an abnormal return as the risk-adjusted return of companies that perform well and poorly on ESG will be the same. However, the net benefit of ESG may not be fully reflected in share prices if the information is unavailable or not used appropriately, which translates into either higher or lower risk-adjusted returns (Mănescu, 2011). This can be illustrated by the aforementioned Figure 1, where ESG could cause firms to be either above or below the security market line, which indicates the fair value of stocks according to the CAPM. If the benefits of ESG outweigh the costs, but investors consistently underestimate those benefits, then the risk-adjusted returns of high-ESG firms should be higher than low-ESG firms. Shares would then be mispriced as underestimating the benefits of ESG could cause the stock performance to subsequently beat expectations. Mispricing could thus be caused by differing beliefs among investors about the benefits of ESG, which some evidence supports. A survey by the Chartered Financial Analysts Institute shows that investors differ in how they incorporate ESG information into their investment strategies (CFA,

2015). In the United States, 67% of investors take ESG information into account, while 82% of investors from Europe, the Middle-East and Africa (EMEA) take ESG information into account. This suggests that some investors could misprice the net benefit of ESG to a certain degree, resulting in an inefficient market with scope for earning abnormal returns.

The mean-variance efficiency argument

The net benefit and the mispricing arguments are the main arguments in the debate regarding the relationship between ESG and abnormal return on a firm level. However, investors usually invest in a portfolio of companies to diversify non-systematic risk. For this purpose, investors can apply specific investment strategies to improve the performance of their portfolios. However, such strategies can potentially impose constraints on the efficient frontier, as described in the 'Portfolio Choice Theory' section (Munk, 2017). This thesis examines whether investors can utilize significant relationships between ESG variables and abnormal return in an investment strategy. However, this implies that investors screen on certain ESG criteria, which lowers the number of investable assets. Investors may thus benefit by utilizing ESG information in their investment strategies, as their portfolios will include companies that benefit from having high ESG performance. However, the positive effect of having firms with high ESG performance in the portfolio could be offset by the potential constraints on the efficient frontier. Constraining the efficient frontier results in portfolios that are not mean-variance efficient as illustrated by Figure 2.

The main arguments regarding the relationship between ESG and abnormal return have now been presented on both a firm and portfolio level. The theory illustrates that different mechanisms can affect the relationship and the outcome of this thesis is thus uncertain from a theoretical perspective. The following section will examine the empirical findings in the literature to establish whether other papers find a clear relationship between ESG and abnormal return.

5. Empirical Findings

This section will evaluate the current literature on ESG and abnormal returns. The first part will examine the empirical findings on ESG and firm performance. This part will thus highlight how specific ESG initiatives can increase or diminish value for firms based on arguments related to shareholder and stakeholder theory. The second part will examine the empirical findings regarding the relationship between ESG and abnormal returns. Examining the literature enables an identification of research gaps and the development of hypotheses, which will support the research question and guide the thesis going forward.

5.1 ESG and Its Effect on Firm Value

The theoretical overview provided explanations of both shareholder and stakeholder theory and how each theory relates to ESG. However, the theoretical overview did not examine how each of the two theories is supported in the literature. Critics of stakeholder theory argue that the theory has problems in terms of accountability and managerial incentives (Renneboog, Horst, & Zhang, 2008a). In shareholder theory, managers are expected to invest when a project's return exceeds the cost of capital. In contrast, managers are expected to balance the interests of all stakeholders to maximize aggregate welfare according to stakeholder theory. The question is whether this goal is possible to achieve, and further how managerial incentives are affected by focusing on stakeholder maximization. Jensen (2002) argues that the difficulty of making tradeoffs between stakeholders makes stakeholder theory a prescription for destroying firm value. The article further states that a stakeholder perspective makes it very difficult to design performance measures and incentive schemes for managers. Consequently, agency costs will increase, and the internal control systems of managers will worsen. Tirole (2001) contributes by stating that management can almost always justify any action by highlighting its impact on the welfare of a specific stakeholder. For example, a CEO can validate a costly and value-destroying acquisition by emphasizing that it will save jobs. Other researchers add that ESG initiatives often imply lowering profits, which is not feasible in a competitive economy (Baumol & Blackman, 1991). However, ESG initiatives may also result in higher profits and thereby drive value creation for investors. The following section will look further into the value drivers of ESG in accordance with stakeholder theory.

A stakeholder focus can drive value in several ways. Turban and Greening (1997) find that firms with high ESG scores have more positive reputations and are more attractive employers than firms with low scores. This suggests that applicants are aware of a firms' ESG profiles and that this knowledge influences the applicants' decision-making, providing a competitive advantage in attracting talent to firms with high ESG scores. The article argues that certain ESG factors are more likely to be appealing to applicants and may thus be more important for firms to consider. The treatment of women and minorities along with employee relations are identified as factors that are especially important for applicants. A positive reputation can further help firms retain quality employees for longer, which can also improve firm performance (ibid.).

Cheng et al. (2014) argue that firms with higher ESG performance face lower capital constraints. The article argues that this correlation materializes via two distinct mechanisms. Firstly, high ESG performance is associated with superior stakeholder engagement, which can reduce the possibility of opportunistic behavior. It is argued that stakeholder engagement based on shared trust reduces potential agency costs and forces management to have a long-term rather than short-term orientation. Secondly, firms with higher ESG performance is likely to disclose more information, which makes these firms more transparent. A high level of transparency reduces asymmetries between the firm and its investors and thus lowers the perceived level of risk (ibid.). Lower capital constraints can have a high impact on firm performance, as it improves the ability of firms to undertake profitable investments.

Green and Peloza (2011) find a consistent correlation between ESG performance and positive consumer support. The paper argues that ESG performance can provide three types of value to consumers: emotional, social and functional. Each of these factors can enhance or diminish the overall value proposition for consumers and the total value will thus depend on the relative fit of the ESG initiative. A company will thus only be able to extract a higher level of value from its customers with an appropriate ESG strategy. Berman et al. (1999) contribute by stating that ESG may function similar to advertising by increasing the demand for products and services by reducing consumer price sensitivity. ESG performance can thus be used to differentiate firms' products or services, which can enhance value creation.

In addition to driving value creation, stakeholder theory also supports how high ESG performance can reduce value destruction by minimizing firm risk. Sassen et al. (2016) state that a high ESG score can help generate positive moral capital among stakeholders, which can provide "insurance-like" protection for the company. The paper explains that a high level of moral capital leads to a higher brand value, which affects the attitude and loyalty towards a company. This alleviates sanctions against the company in the event of a company crisis and thus leads to more stable future cash flows and reduced risk. Berman et al. (1999) support this argument and add that engaging in ESG activities implies that companies are proactive instead of reactive and thus avoid the costs of negative reactions from stakeholders such as customers, employees and governments. Nofsinger and Varma (2014) find that companies with a high level of socially responsible attributes are less risky in market crisis periods. The study finds that SRI funds outperformed conventional funds during the Great Financial Crisis by having significantly higher alphas. This outcome supports the findings by Sassen et al. (2016), and it can be concluded that ESG can help mitigate both firm-specific and systemic risk.

The above sections illustrate how ESG can both drive and diminish firm value in different ways. The arguments of shareholder and stakeholder theory have thus been specified by empirical findings. However, an empirical examination of the general direction in the relationship between ESG and abnormal returns is yet to be conducted.

5.2 ESG and Abnormal Returns

This section will review the literature to establish whether ESG is predictably associated with either high or low firm performance and abnormal returns. Halbritter and Dorfleitner (2015) investigate the link between ESG and abnormal return for a sample of the US market from 1991 to 2012. The paper applies a portfolio approach based on a single ESG indicator score and performs regressions according to the Carhart (1997) four-factor model. The authors find no significant return difference between high and low ESG rated portfolios. Similarly, Luther et al. (1992) find that Jensen's alphas of ethical companies have a mean of 0.3% per month, which was not significantly different from zero. This study focuses on the UK market in a time period from 1984 to 1990. Although the findings are insignificant in these two papers, other papers do find a significant relationship between ESG performance and abnormal returns. Renneboog et al. (2008b) find that ethical funds in the US and UK strongly underperform the Fama-French-Carhart benchmark in the period from 1991 to 2003. In particular, the adjusted alphas of funds in these countries are -2.2% and -3.4%, respectively. The

authors argue that the firms included in ethical funds, and hence meet high ethical/social standards and governance criteria, may be overpriced in the market. Accordingly, the paper argues that investors pay a premium for ethics. In contrast, Ortas et al. (2015) find that ESG performance has a significant positive impact on firm performance for companies in Spain, France and Japan. The paper examines the implications of committing to the United Nation Global Compact (UNGC), which has an effect on both ESG and financial performance. Contrary to the previously mentioned articles, this implies that firms generally benefit from improving their ESG performance.

The above examples illustrate that the literature on ESG and investor return is generally unclear with very mixed findings. Furthermore, analyzing just a fraction of studies makes it difficult to generalize findings. Friede et al. (2015) provide aggregated evidence from more than 2000 empirical studies on ESG and financial performance. The study extracts all available primary and secondary data of previous meta-studies and finds that the business case for ESG is very well founded. About 40% of the studies report a positive relationship between ESG and financial performance compared with 10% of studies finding a negative relationship. This is a significant finding when taking the large number of papers included in this study into account. It is important to consider that abnormal returns are not directly related to firm performance, which was established in previous sections. Furthermore, the article does not examine specifically which ESG initiatives that have the most significant effect, and it argues that this is a relevant research area for future research. Lastly, the study argues that the relation between ESG and firm performance depends on the examined time period. This suggests that short-term results may differ from long-term results.

Nollet et al. (2016) provide evidence of a non-linear relationship between ESG and firm performance. The study finds a negative correlation between ESG and firm performance in the short-term and a positive correlation in the long-term. The study examines the relationship between ESG and firm performance using the S&P500 firms as the sample in the period from 2007-2011. This finding is supported by Shank et al. (2005) who state that the relationship between ESG and investor return is neutral in three and five-year periods, while there is a positive relationship over a ten-year period. Nollet et al. (2016) further argue that ESG expenditures only pays off after a certain threshold is reached, which requires considerable planning and dedication of resources. These findings are in line with the theoretical argument by McWilliams and Siegel (2001), who argue that an optimal level of

ESG exists. In conclusion, it is important to consider the examined time period and the ESG variables included in the study.

Most papers on this topic apply a single ESG measure, which is expected to cover every aspect of CSR (Friede, Busch, & Bassen, 2015). However, the different inputs used to create an ESG measure often have opposing effects, which can explain the mixed findings to some extent (Mănescu, 2011). Some studies decompose the ESG score by examining the relationship between the subcategories (environmental, social and governance) and abnormal returns. First, Derwall et al. (2005) construct equity portfolios based on environmental performance indicators and measure the performance of the portfolios using the four-factor model presented by Carhart (1997). This study finds that the portfolio of firms with high environmental scores outperforms a portfolio of firms with low environmental scores by 6% per annum. The paper thus suggests that environmental screens can help investors improve the performance of their portfolios. Cohen et al. (1997) also constructed portfolios based on environmental characteristics. In contrast, this study suggests that there is neither a premium nor penalty for investing in environmentally friendly companies, as no significant abnormal return was identified.

Several studies examine the social aspect of ESG by focusing on the correlation between stakeholder management and abnormal return. Hillman and Keim (2001) investigate the impact of improved relationships with primary stakeholders such as employees, customers, suppliers and communities. The paper finds that a focus on stakeholder value can also result in increased shareholder value. In contrast, the study further concludes that participation in social issues, such as abandoning nuclear power or avoiding 'sin' industries, will often diminish shareholder value. Mánescu (2011) examines the effect of several social factors including community relation, diversity, employee relationship and production safety. This study finds that community relations had a positive effect on risk-adjusted stock returns in one time period and a negative effect in another period. The paper argues that the positive result could be due to mispricing, while the negative result could be compensation for ESG inducing lower systemic risk exposure.

Lastly, numerous studies assess the correlation between corporate governance and abnormal returns, and much of the empirical literature suggests a positive relationship. Gompers et al. (2003) study the relationship between a set of governance provisions and a firm's long-term performance in the 1990s.

The authors build a governance index, based on 24 governance provisions, that aims at capturing the relative power of shareholder rights. The paper finds a strong association between corporate governance and stock returns, as buying firms with strong shareholder rights and selling firms with low shareholder rights resulted in abnormal returns of 8.5% per annum. Similarly, Cremers and Nair (2005) divide governance mechanisms into internal and external governance and examine how these governance mechanisms are associated with equity returns. This study finds that an investment strategy based on a firm's internal governance mechanisms generates abnormal returns of 8% per annum. However, other studies are not able to find a strong correlation between high corporate governance performance and equity returns. Bebchuk et al. (2013) find that the correlation Gompers et al. (2003) find in the period from 1991-1999 disappears in the subsequent period from 2000-2008. The paper argues that the disappearance of correlation is due to market participants gradually learning to account for differences between companies scoring well and poorly on governance performance indicators. This paper supports that findings on the relation between corporate governance and equity returns differ, and it further emphasizes how results can be affected by the chosen time period.

5.3 Research Gaps

The review of the literature illustrates that the findings of the relationship between ESG and abnormal returns are mixed and often contradictive. This suggests that additional research is needed on this topic in order to draw more definite conclusions. Some researchers argue that contradictive findings are caused by different research methodologies. Others argue that the results may be highly dependent on the examined time period. The inconclusive results suggest there are research gaps worthy of further investigation. The sections below will highlight the prevailing research gaps that will be examined in this thesis.

First, the examination of how ESG initiatives impact firm performance illustrates that ESG initiatives can affect firm performance in a variety of ways (Mănescu, 2011). An overall ESG measure does not highlight the value-drivers of ESG and may thus lack explanatory power. The literature review highlighted that most papers examine the relationship between ESG and abnormal return by using a single measure of ESG. Consequently, new literature could benefit from focusing on subcategories of ESG and how these underlying ESG factors influence abnormal return.

Second, papers in this research area often focus on a single geographic area. Renneboog et al. (2008b) is an exception as the paper finds that the return of socially responsible investment (SRI) funds and conventional mutual funds differs at a statistically significant level depending on the country. There are relatively few studies performing a direct comparison of how ESG performance differs geographically with respect to how it affects abnormal returns. Consequently, it will be valuable to examine how results vary in different geographical areas, where the regulations and views on ESG generally differ.

Third, the literature review shows that the relationship between ESG and abnormal return has been examined in many different time periods. However, few studies have applied similar methodologies to investigate whether the relationship between ESG variables and abnormal return specifically depends on the examined time period. Bebchuk et al. (2013) replicate a previous study by Gompers et al. (2003) and highlight that the relation between governance variables and abnormal return may be highly dependent on the examined time period. However, a few other papers have studied if and how the relationship between ESG and abnormal return depends on the time period, which will thus be valuable to investigate in further detail. In conclusion, the literature on ESG illustrates that many questions are still unanswered and further research is required. This thesis aims at investigating several of the research gaps identified above. For this purpose, several hypotheses will be developed in the following section to support the research question and guide the analysis going forward.

5.4 Hypotheses Development

A hypothesis is a testable statement about a causal relationship. It is an important tool in academia because until a problem has been reduced to a hypothesis form, it cannot be scientifically tested (Prasad, Rao, & Rehani, 2001). The purpose of developing hypotheses is thus to decompose the research question into specific parts. The hypotheses should be testable, falsifiable and realistic (Mourougan & Sethuraman, 2017). This ensures that the problem is well-defined and within scope to solve with the available resources and time. The falsifiability criterion implies that the direction of the examined relationship should be clearly stated, even though the context is ambiguous. It can be equally valuable to reject or not reject a hypothesis. Therefore, the hypotheses will state the expected direction of the examined relationship even if there is conflicting evidence of the direction. The

research hypotheses are not null-hypotheses but empirically supported hypotheses of the expected results. The following section will highlight the hypotheses of this thesis and the rationale for each.

Hypothesis 1:

There is a positive relationship between one or more ESG variables and abnormal return, alpha, at the 5% significance level.

As highlighted in the literature review, there is conflicting evidence about the direction of the relationship between ESG and alpha. However, the meta-study of more than 2000 empirical studies on ESG and financial performance concludes that 40% of the studies find a positive relationship compared to 10% of studies finding a negative relationship (Friede, Busch, & Bassen, 2015). This makes it more reasonable to expect a positive relationship than a negative one. However, the timing, impact and the way ESG variables affect firm performance differ greatly. Hence, the relationship with alpha is likely to differ for different ESG variables.

Hypothesis 2:

No ESG variable is significant at the 5% level in both the US and Europe in the same time period.

There is significant evidence in the literature that the relationship between ESG and investment performance differ geographically. For example, Renneboog et al. (2008b) find that it differs among countries if the return differentials between socially responsible investment (SRI) funds and conventional mutual funds are statistically significant. This could be caused by historical, regulative and cultural differences. From an investor perspective, Duuren et al. (2016) conduct a survey and find that US investors are much more skeptical about the impact of ESG on financial performance than European investors. On the other hand, the survey finds that US investors believe ESG to be more important for long-term risk reduction than European investors. Due to the differing perspectives of ESG and investment performance, we expect different ESG variables to be significant in the US and Europe.

Hypothesis 3:

No ESG variable is consistently either positive or negative at the 5% significance level across consecutive time periods in the same region.

ESG literature on the relationship with abnormal returns often examines different time periods and research suggests that the examined time period is important for the results. However, few studies apply a similar methodology to subsequent time periods, as highlighted in the 'Research Gaps' section. Bebchuk et al. (2013) replicate the methodology of Gompers et al. (2003) in a later time period and find that the abnormal returns disappear. This suggests that the significant ESG variables will vary by time period. Additionally, abnormal returns associated with an ESG variable is unlikely to persist over long time periods due to the no-arbitrage condition of efficient markets. Eventually, more investors will recognize the abnormal returns associated with the ESG variable and buy stocks from firms scoring high on the given ESG variable. This will drive up the price of investing in these firms, diminishing the returns for new investors until the abnormal returns converge towards zero.

Hypothesis 4:

Positive alpha portfolios can be formed by using significant ESG variables as screening criteria.

There is a wealth of studies on using ESG factors to form portfolios with abnormal returns and, at a general level, the results are mixed. However, there are two main reasons that it could be possible to form positive alpha portfolios from significant ESG variables. First, the aforementioned meta-study of 2000 empirical studies found that there are four time as many studies concluding a positive relationship with between ESG and financial performance than a negative relationship. Second, if our primary regression analysis finds a positive, significant relationship between an ESG variable and alpha, it indicates that firms scoring high on this ESG variable have higher abnormal returns. It is plausible that a portfolio, consisting of firms scoring high on the ESG variable positively related to alpha, could yield a positive alpha. However, there are two important counter-arguments. Firstly, a portfolio of firms scoring high on a desirable ESG variable might also score high on undesirable ESG variables. It becomes a complex optimization problem to construct a portfolio of firms based on whether the firms score high or low on a range of ESG variables. Secondly, forming portfolios based on ESG criteria will likely impose constraints on the efficient frontier. This could result in lower mean-variance optimization for investors, who could be worse off as a result. Despite these counter-arguments, we expect it to be possible to construct positive alpha portfolios. This is based on the findings of the meta-study and the prior indication of which ESG variables are positively related to alpha, highlighted by our regression analysis.

The formulation of hypotheses enables empirical testing of the research question. It follows that a methodology must be developed to create an effective research design. Further, it is important to thoroughly explain the logic of the research design, which is done in the following section. The explanation of the research design will provide clarity of how the results will be achieved and enable others to replicate our methodology in different settings, which could provide additional insights.

6. Methodology

The following sections will highlight the methodological considerations and choices. Two separate regression analyses are needed to provide a comprehensive answer to the research question. The first section will provide an overview of the research design to illustrate how the two regression analyses complement each other.

6.1 Overview of Research Design

The purpose of this thesis is to investigate whether ESG factors have a significant association with abnormal returns and, additionally, exploring whether investors could have profited from investing based on ESG-factors. Recall that the research question driving our thesis is:

"What was the relationship between ESG scores and abnormal returns in the US and Europe during 2003-2017? Moreover, could investors have exploited this potential relationship to construct portfolios with abnormal returns?"

Our research design to answer the research question is illustrated in Figure 3 below:

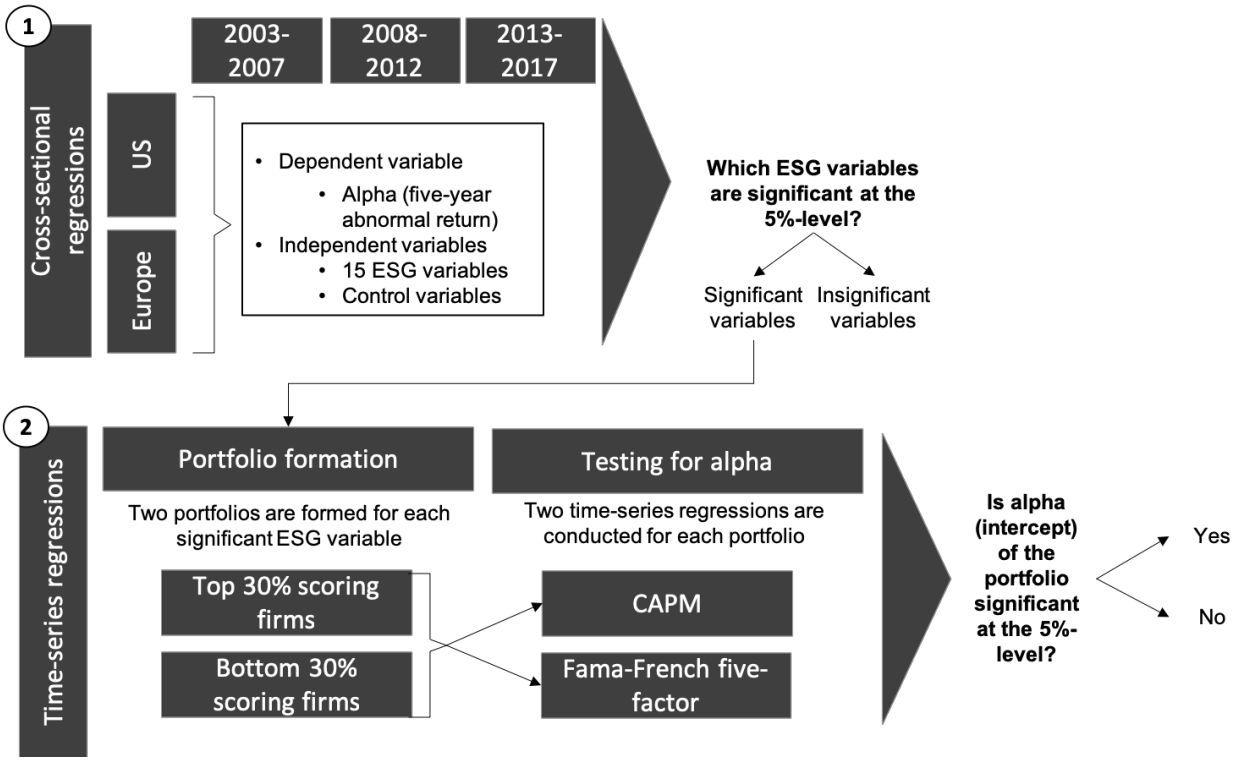


Figure 3

This section will use Figure 3 to provide a brief overview of the methodology, followed by a more thorough explanation of each element. We start by conducting separate regressions for the US and Europe over three consecutive time periods. These are three five-year time periods of 2003-2007, 2008-2012 and 2013-2017. This results in a total of six cross-sectional regressions. The dependent variable is alpha, which is defined as the cumulative abnormal return over the five-year period. Alpha is computed for each firm in each period using the CAPM. The independent variables comprise 15 ESG variables and control variables. The values of the ESG variables are the ESG scores prior to the given five-year period. In this way, we are testing whether strong ESG performance is associated with subsequent abnormal returns. Control variables are included to account for factors that influence alpha in addition to ESG. These six cross-sectional regressions will then be analyzed to see if any ESG variables are significant at the 5% level. The insights from analyzing the relationship between ESG and alpha in the cross-sectional regressions will enable us to answer the first half of the research question. The other half aims at understanding whether the observed relationship between ESG and alpha can be utilized by investors to construct portfolios with significant alphas. To test whether this is the case, we form portfolios based on the ESG variables significantly associated with alpha in the cross-sectional regressions. For each significant ESG variable, the top 30% and bottom 30% scoring firms are grouped in separate portfolios. There are thus two portfolios for each significant ESG variable. It is subsequently tested whether the portfolios obtained a significant alpha over the following five-year period. This is done by conducting two time-series regression for each portfolio, using both the CAPM and the Fama French five-factor model. It will then by analyzed if any of the portfolios' time-series regressions have significant alphas at the 5% level. The analysis of the time-series regressions will help answer the second half of the research question.

The cross-sectional and time-series regressions supplement each other in answering the research question comprehensively. However, there are several prior steps needed to construct these regressions in a meaningful way. The first step is deciding how to collect data, which will be assessed in the section 'Method of data collection'. This section comprises the choice of database, the choice of sample and the collection of ESG and non-ESG data. The next step is explaining how the data was analyzed in the section 'Method of data analysis'. This section starts by detailing the pre-regression considerations of dealing with outliers and multicollinearity among other issues. The test results of the pre-regression tests will be highlighted and the implications for our approach will be discussed.

The next section in 'Method of data analysis' will further explain how the regressions were constructed. Afterward, an indicative test of endogeneity will be described to show whether there is reverse causality between ESG and alpha in our sample. This will highlight whether ESG scores depend on past abnormal returns. Sensitivity tests will also be described as they help check the robustness of the results. The methodology section will conclude with a discussion of the most important limitations in the outlined research design.

6.2 Method of Data Collection

Choice of database

Thomson Reuter's DataStream database was chosen to extract both ESG and non-ESG data. The Asset4 database, a part of DataStream, was used to obtain ESG data. The availability, granularity and informativeness of ESG data were the main criteria for choosing the database. The availability of non-ESG data, such as Total Shareholder Return, is available across most databases and was thus not as important. Asset4 is one of the most widely used ESG databases in academic research of ESG due to the highly detailed ESG data, a large number of firms with ESG data and the availability of ESG data from 2002 forth (Halbritter & Dorfleitner, 2015). One reason for Asset4's widespread use is that Asset4 was the first agency to provide raw ESG data to investors back in 2009 (Huber, Comstock, & Polk, 2017). The key drawback of the other ESG database that Copenhagen Business School has access to, Bloomberg, is how the components of E, S and G are broken down (Appendix 2). The social component exclusively measures firms' internal factors and not how firms interact with their external stakeholders. Collectively, Bloomberg's measures are easily quantifiable, such as the ratio of greenhouse gasses emitted to revenue and the average age of directors, but the majority of Bloomberg's measures are arguably not meaningful and relevant across firms. A firm might have a very low ratio of greenhouse gas emissions to revenues but emit other harmful toxins not captured by Bloomberg. Another example is the average age of directors, as it is unclear whether this variable drives or diminishes firm value. The average age of directors could be high due to the retention of skilled and experienced directors or due to directors being entrenched and thus hard to displace. Therefore, some degree of qualitative assessment of how each firm score on ESG parameters is likely needed to make each parameter meaningful across firms. Asset4 excels in this regard as Thomson Reuters has more than 100 analysts to continuously assess how firms rank on ESG parameters

(Thomson Reuters, 2011). The informativeness of the ESG parameters are pivotal for this thesis so the ESG scoring methodology will be explained in the following section.

Asset4's ESG scoring methodology

Asset4's ESG ratings are computed based on a hierarchical structure illustrated in Figure 4.

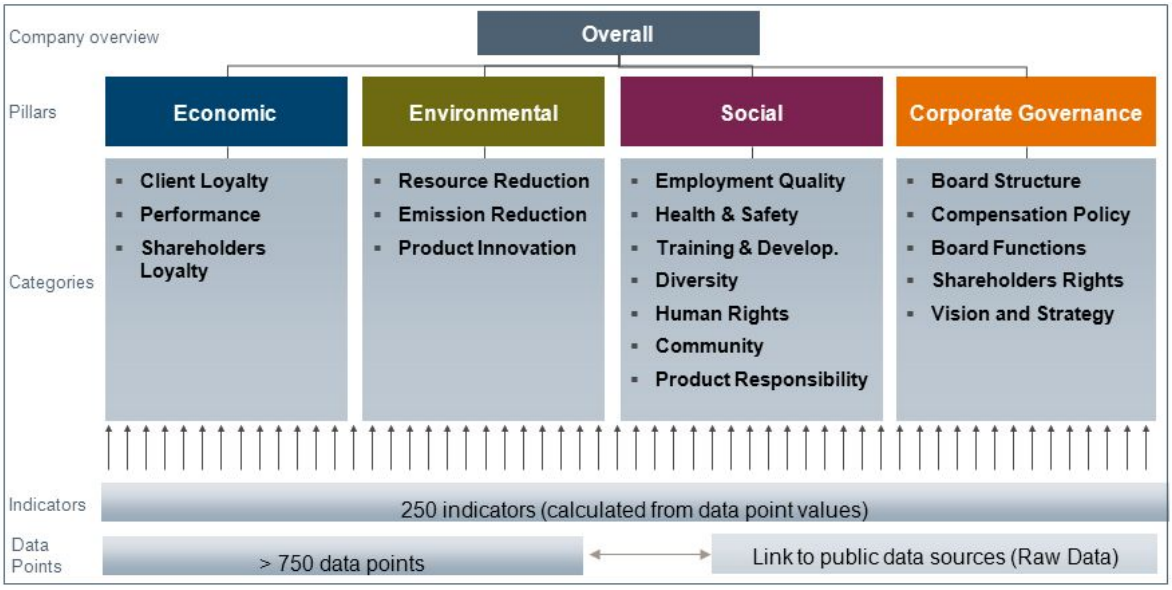


Figure 4; Source: (Thomson Reuters, 2011)

At the top, there is an overall ESG score based on the average scores of the four pillars of environmental, social, governance and economic performance. This thesis has excluded the economic performance pillar, as this pillar is separate from ESG. Each E, S and G score is divided into between 3-7 category scores with 15 category scores in total (Figure 4). These 15 category scores are the chosen ESG variables in the regression. By breaking E, S and G down into category scores, this thesis aims to specify which underlying drivers of ESG that influence abnormal returns the most. Category scores are computed as the average of indicator scores. An indicator score is a Z-score between 0 and 100% indicating how a firm ranks among all other firms in Asset4's database on a given parameter (Thomson Reuters, 2015). Indicator values are used as the input for determining firms' indicator scores and thus the firms' ranking. The indicator values are collectively based on more than 750 data points (Thomson Reuters, 2011), and Asset4's scoring methodology is thus very data-driven. The relative scoring of firms is important because in isolation a value, such as greenhouse gas emissions, is less useful. However, when a value is compared to other firms' values, it becomes insightful. An indicator score of 90% thus means that the firm is at the top 10% percentile of the indicator. The

category scores are simple averages of the related indicator scores, so a category score of 90% means that a firm is, on average, at the top 10% percentile of the relevant indicator scores. The following will explore the benefits and drawbacks of Asset4's ESG scoring methodology.

Besides relative scoring, the main benefits of Asset4's ESG rating methodology are its consistent measurement across firms and objectivity in scoring criteria. The same method of measuring ESG performance is used across all firms in the database. If a different measurement methodology was applied to energy firms compared to technology firms, separate regressions for each group of firms would be required. This would significantly reduce the sample size, which would limit our ability to make statistical inferences and generalize our results. Regarding objectivity, the hierarchical structure of Asset4's ESG ratings limits the impact of subjectivity. Qualitative assessments are only conducted at the indicator value level. Below this level are data points and above this level are relative scoring. Most of the qualitative assessments at the indicator value level are either a yes/no assessment or an observable value. An example of a yes/no assessment related to emission reduction is "does the company monitor its emission reduction performance?" (Thomson Reuters, 2015). Limiting the analysts' subjective assessments to mostly yes/no answers on specific data points makes the ESG ratings less subjective as a whole than if analysts had the autonomy to assign E, S and G scores directly to firms. The objectivity in scoring criteria is important because if the degree of subjectivity is too high, non-ESG factors could influence the ESG ratings. Past firm performance might influence the analysts' perceptions of the firm, which could bias the ESG scores. Such bias could cause endogeneity between past firm performance and ESG scores. Subjective scoring could also lead to omitted variable bias if the non-ESG factors that influenced the ESG ratings are not included in the regression. However, though there is inevitably some bias in Asset4's ESG ratings, the impact of subjectivity is limited by requiring all qualitative assessments to be based on specific data points.

The key drawback of the Asset4 ESG scoring methodology is that firms' ESG scores are compared to all firms and not firms in its peer group. Consistent measurement was mentioned as a benefit but the relative ranking of firms on a given parameter would be more insightful if firms were ranked against their peers. For example, with the current methodology, an oil firm might score low on the environmental pillar because it generates more emissions compared to all other firms due to the nature of its business. In other words, oil firms' environmental scores might be primarily driven by the industry they are in rather than specific environmental actions taken by the oil firms. However, an oil

firm can improve its environmental score significantly by improving transparency and frequency of its reporting on emissions, which are key factors in calculating the environmental pillar score. The variety of inputs used to calculate ESG scores allows firms, regardless of industry, to take steps to improve its ESG score. Additionally, benchmarking against a peer-group of firms carries its own pitfalls. Introducing peer groups also implies a layer of subjectivity in terms of defining the peer group. Additionally, firms might change their peer group over time. This would make ESG scores less comparable over time. In sum, Asset4's ESG scoring methodology is transparent, meaningful and relatively objective but the methodology is limited by comparing firms' ESG performance to all other firms rather than peers.

Sample selection

Choosing the sample comprises both deciding on the composition and size of the sample. The sample consists of public firms with an average market capitalization of $17 million USD over the sample period. These firms were chosen mainly because they are covered by Asset4's ESG ratings. The firms are US and European firms, which are covered by major equity indices (Thomson Reuters, 2011). Most ESG studies focus on one geographic region as highlighted in the 'Empirical Findings' section. However, there could be differences in how ESG factors translate into abnormal returns in different regions. We address this aspect by having two sub-samples, where one includes companies from the US and the other includes companies from Europe. Europe comprises the UK, Germany, France, Italy and Spain. Other European countries were excluded as they only marginally increased the number of firms due to the lack of firms with ESG data. Additionally, the country-specific differences impacting firm performance would proliferate if many countries were included. It is important to note that the firms from the chosen five European countries are not representative of Europe as a whole and are thus a potential source of sampling bias. One way this sampling bias could have been mitigated is through random sampling. However, we chose all firms with ESG data for each region in a given year to have a larger sample. Choosing a subset of random firms with ESG data within each region would reduce the sample size significantly and thus reduce the precision of the coefficient estimates in the regression. Another source of sampling bias is that the firms with ESG data could have different characteristics important to firm performance than firms without ESG data. A factor mitigating this sampling bias is that Asset4 selects which equity indices to cover and not individual firms (Thomson Reuters, 2011). Since Asset4 has no influence on which firms are included in the indices, it limits the

ability of Asset4 to "cherry-pick" certain firms to be scored on ESG. However, there is some scope for sampling bias in choosing which indices to cover as well.

A key benefit in increasing the sample size is reducing the standard error of the regression coefficients (Wooldridge, 2012). This means that the coefficient estimates become more precise and thus move closer to their true values. Since this thesis uses the ESG variables' coefficients to make inferences about abnormal returns, it is crucial that the sample size is large. Table 1 highlights the sample size across regions and time periods.

Number of firms	2002	2008	2012
US	290	480	691
Europe	219	419	478
Total	**509**	**899**	**1169**

Table 1

The number of firms in the sample more than doubles from 2002 to 2012 and, crucially, the minimum number of firms for a region at any time is 219 in Europe in 2002. If the number of European firms was substantially smaller in 2002, it could have been necessary to merge the two regions to get a sufficiently large sample size.

In addition to collecting ESG data across geographic regions, ESG data was also collected across time. Annual ESG data was retrieved from 2002 to 2017. The key reason for collecting ESG data at multiple points in time is that the impact of ESG factors on total shareholder return might have changed over time. As highlighted in the 'Empirical findings' section, some ESG factors could be more important now than at previous points in time. Collecting ESG data over the chosen time period further allows for interesting analyses about the impact of ESG on abnormal returns before, during and after the Great Financial Crisis.

6.3 Method of Data Analysis

The high-level approach to answer our research question is a two-step process as highlighted in Figure 3. First, a cross-sectional regression will be performed across several time periods to see if any ESG

variable has a significant impact on alpha in the subsequent five-year period. The dependent variable, alpha, is accumulated to a single, five-year value used in the regression. Thus, although the dependent variable is measured over time, it is a cross-sectional regression because the dependent variable represents a single time period and not several time periods as in time series or panel regression. Second, in case some ESG variables are significant, portfolios will be formed on the basis of each significant ESG variable and time-series regressions will be performed to examine if the portfolios have significant alphas or abnormal returns. This will indicate whether investors could have traded on knowing which ESG variables were significant and, as a result, obtained a significant alpha over the subsequent five-year time period. These two steps will produce the main results of this thesis. Additionally, sensitivity tests will be conducted to test the robustness of the results.

Cross-sectional pre-regression tests and considerations

This section will highlight the pre-regression considerations and tests to ensure that the regression is properly designed so that meaningful inferences can be extracted from the results.

Outliers

It is important to check the data set for outliers because "unusually" large or small observations can greatly affect the coefficient estimates of the ordinary least squares (OLS) regression (Wooldridge, 2012). In an OLS regression, the regression line is the line that minimizes the squared residuals from the line to the data points. If a data point has an abnormally large value compared to the other observations, its squared residual will carry a relatively large weight in fitting the regression line (Wooldridge, 2012). Cook's distance (Cook's D) can be applied to assess the impact of outliers on a regression. This indicates how much the predicted values are affected by a specific observation (Cook, 1977). As a rule of thumb, a Cook's D value above 0.5 indicates the observation might be influential and a value above 1 indicates that the observation is quite likely to be influential (PennState, 2018). If Cook's D is abnormally high for a few observations, these observations will be evaluated individually and potentially excluded. An alternative to exclusion is mitigating the impact of the outlying values through winsorizing. Winsorizing is a frequently used technique in academia to reduce the influence of outliers on OLS coefficient estimates (Brown & Caylor, 2006). Winsorizing at the 1% and 99% level sets all values below 1% of the distribution equal to the 1% percentile value and all values above 99% equal to the 99% percentile value. This greatly reduces the impact of

extreme values, but it is inevitably a source of bias. The next section will highlight the results of the Cook's D tests and the implications for our approach.

Test result of outliers

Only two firms recorded a Cook's distance above 0.5. They were European firms and both observations were recorded in the 2013 to 2017 time period. The data for these firms did not contain any errors upon closer inspection. The first observation had an alpha below the 1% percentile and the second observation above the 99% percentile. However, neither observation's alpha value was very different from other observations at the tail ends of the distribution. Thus, removing the two aforementioned observations could be a source of bias. This argument, combined with no evident typing errors, resulted in keeping the observations in the regression. A lower threshold of Cook's distance was also applied to identify other potential outlying observations, but neither of these observations contained errors. Therefore, no observations were excluded from the regression as outliers. Winsorizing was not applied as the low level of Cook's distance values indicated that the results were not heavily affected by outliers. Additionally, test regressions were run to estimate the impact of winsorizing at the 1st and 99th percentiles. These regressions showed that the results were only marginally affected by winsorizing at the 1st and 99th percentiles. No significant variables became insignificant when winsorizing or vice versa.

Multicollinearity

Multicollinearity is defined as "a high (but not perfect) correlation between two or more independent variables" (Wooldridge, 2012). The effect on the dependent variable of two or more independent variables is entangled when multicollinearity is present. This is a potential issue as the aim of this study is examining the effect of individual ESG variables on alpha. Equation 1 is useful in understanding how multicollinearity affects the independent variables' coefficient estimates.

$$\mathrm{Var}(\hat{\beta}_j) = \frac{\sigma^2}{\mathrm{SST}_j(1 - \mathrm{R}_j^2)},$$

Equation 1 (Wooldridge, 2012)

The left-hand side of Equation 1 is the variance of the coefficient estimate and R_j^2 in Equation 1 shows how multicollinearity affects this variance. Note that R_j^2 is the goodness-of-fit of one

independent variable against all other independent variables and not the conventional R^2 that measures the goodness-of-fit of the independent variables against the dependent variable (ibid.). When R_j^2 is high, the variation in the independent variable x_j is well explained by the variation in other independent variables. In other words, a high R_j^2 implies a high correlation between x_j and the other independent variables, which is the definition of multicollinearity. As this thesis aims to examine the effect of single ESG variables on investment performance, it is important to examine how each ESG variable is correlated with the other independent variables. The variance inflation factor (VIF) is a useful measure to gauge the degree of multicollinearity. The formula for VIF is denoted by Equation 2.

$$VIF_j = \frac{1}{1 - R_j^2}$$

Equation 2 (Wooldridge, 2012)

VIF highlights how much a coefficient estimate's variance is inflated by correlation with other independent variables. Equation 3 makes this clear by inserting VIF in Equation 1.

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{\text{SST}_j} \cdot \text{VIF}_j,$$

Equation 3 (Wooldridge, 2012)

It is thus clear that multicollinearity entangles the effect of independent variables. Additionally, an inflated variance due to multicollinearity also makes coefficients less significant, which will be demonstrated in the following. Taking the square root of the variance yields the standard deviation of the coefficient in Equation 4.

$$\text{sd}(\hat{\beta}_j) = \sigma/[\text{SST}_j(1 - R_j^2)]^{1/2}.$$

Equation 4 (Wooldridge, 2012)

However, neither the variance nor the standard deviation of the coefficient is directly observable because the error standard deviation, σ, cannot be observed and must be estimated. The estimated error standard deviation, $\hat{\sigma}$, is used to compute the standard error of the coefficient in Equation 5.

$$\text{se}(\hat{\beta}_j) = \hat{\sigma}/[\text{SST}_j(1 - R_j^2)]^{1/2}.$$

Equation 5 (Wooldridge, 2012)

A clear link has now been established between the coefficient variance, standard deviation and sample standard error. The standard error directly impacts the statistical significance of the coefficient estimate by influencing the t-statistic (ibid.).

$$t_{\hat{\beta}_j} \equiv \hat{\beta}_j / \text{se}(\hat{\beta}_j).$$

Equation 6 (Wooldridge, 2012)

It follows from Equation 6 that a higher standard error will lower the t-statistic. A lower t-statistic implies a higher p-value or a less significant coefficient (ibid.). It has now been established by Equation 1-6 that a higher coefficient variance ultimately results in a less significant coefficient by inflating the standard error. Multicollinearity can thus cause significant variables to become insignificant, which greatly impacts our results. Therefore, this thesis will scrutinize the VIF for each ESG variable to assess the severity of multicollinearity. Multicollinearity among control variables is of little interest because the collective ability of the control variables to control for variance in the dependent variable is important, not the isolated effect of each control variable.

A correlation matrix is a useful supplement to VIF because a correlation matrix highlights the correlation between pairs of independent variables. A high pair-wise correlation is problematic because it entangles the effect of two variables on the dependent variable. Therefore, correlation matrices among independent variables will be analyzed.

Potential multicollinearity can be remedied by either dropping a highly correlated variable or by increasing the sample size. In deciding whether to drop a variable due to multicollinearity, it is ill-advised to use a cut-off value, such as 10, for VIF as a strict criterion (ibid.). This is because the effect of multicollinearity can be mitigated by increasing the sample size. Increasing the sample size increases the total sum of squares (SST), which reduces the variance of the coefficient estimate (ibid.). This is also clear from Equation 3 as increasing SST clearly lowers the variance.

Test result of multicollinearity

The degree of multicollinearity was modest across time periods and regions, which can be inferred from Table 2. There were only three variables exceeding a VIF of 3 across all time periods, which

are highlighted in the table. These comprised the governance variable 'Vision and Strategy' and the two environmental variables 'Emission Reduction' and 'Resource Reduction'. Correlation matrices showed that these three variables were also highly correlated in multiple time periods.

VIF values	**US**			**Europe**		
	2002	2007	2012	2002	2007	2012
Board Function (G)	1.87	1.21	1.25	1.78	1.59	1.60
Board Structure (G)	1.52	1.13	1.23	1.92	1.86	2.17
Compensation Policy (G)	1.18	1.28	1.26	1.70	1.90	1.75
Shareholder Rights (G)	1.44	1.11	1.13	1.19	1.10	1.15
Vision and Strategy (G)	**2.37**	**3.38**	**4.72**	**2.17**	**3.77**	**3.42**
Emission Reduction (E)	**4.76**	**4.91**	**5.74**	**3.34**	**3.97**	**4.00**
Product Innovation (E)	2.10	1.95	1.83	1.86	1.93	1.99
Resource Reduction (E)	**4.89**	**4.14**	**4.53**	**2.66**	**3.18**	**3.34**
Community (S)	2.06	1.80	2.27	1.55	1.69	1.92
Diversity and Opportunity (S)	2.09	1.90	1.87	2.09	2.01	2.07
Employment Quality (S)	1.60	1.46	1.38	1.82	1.53	1.61
Human Rights (S)	1.24	1.48	1.68	1.24	2.05	2.09
Health and Safety (S)	2.20	2.06	1.92	1.87	1.77	1.72
Product Responsibility (S)	1.46	1.31	1.46	1.51	1.61	1.47
Training and Development (S)	1.83	1.96	2.15	1.44	1.81	1.94
Leverage Ratio	1.14	1.11	1.14	1.16	1.12	1.18
Market Capitalization (log)	1.64	1.70	1.72	1.45	2.37	1.84
Price-To-Book Ratio (log)	1.16	1.11	1.14	1.34	1.13	1.21
Average	2.03	1.94	2.13	1.78	2.02	2.03

Table 2

For the US regression from 2013-2017, neither of these variables was statistically significant. This is partially due to the higher VIFs of these variables. A higher VIF lowers the significance levels of the coefficients as evident from Equation 1-6. The 'Emission Reduction' variable had the largest VIF of 5.74 in the US regression for the 2013 to 2017 period. Recall that the square root of a coefficient's variance equals its standard deviation, which is approximated for the sample by the standard error. Taking the square root of 5.74 yields a squared VIF value of 2.40. This implies that the standard error of the coefficient for 'Emission Reduction' was 2.40 times higher due to multicollinearity in the US regression for the 2013 to 2017 period. A VIF of 5.74 also implies an R-squared of 82.59% when regressing all other independent variables on 'Emission Reduction'. In other words, the other independent variables capture 82.59% of the variation in 'Emission Reduction'.

Neither of the variables was excluded due to multicollinearity. Omitting 'Vision and Strategy', 'Emission Reduction' and 'Resource Reduction' in the US 2013-2017 regression does not make any insignificant variables significant and vice versa. The two significant variables in the US 2013-2017 regression, which will be elaborated in the 'Results' section, were the social variables 'Diversity and Opportunity' and additionally 'Health and Safety'. The standard errors and coefficients for these two variables were only marginally affected by omitting 'Vision and Strategy', 'Emission Reduction' and 'Resource Reduction'. Therefore, the correlated variables were not omitted because their omission could induce omitted variable bias and, additionally, their omission only marginally affected the significant variables. Having the same variables in all regressions also has the benefit of making the regressions more comparable. The following section will discuss the issue of heteroskedasticity and assess the degree of heteroskedasticity in the regressions.

Heteroskedasticity

Heteroskedasticity is defined as non-constant variance of the error term in the regression (Wooldridge, 2012). In other words, it varies with the value of the independent variables how precise the regression model is in predicting the dependent variable. This is opposed to homoskedasticity, which is defined as constant variance of the error term. Homoskedasticity is one of the assumptions of multiple linear regression highlighted in Appendix 3. To understand why homoskedasticity is important, recall Equation 1.

$$\mathrm{Var}(\hat{\beta}_j) = \frac{\sigma^2}{\mathrm{SST}_j(1 - \mathrm{R}_j^2)},$$

Equation 1 (Wooldridge, 2012)

The variance of the coefficient on the left-hand side is a function of constant variance of the error term, σ^2, on the right-hand side. Therefore, if the variance of the error term is not constant, the variance of the coefficients is not precisely estimated. An imprecise variance of the coefficients implies imprecise standard errors, which leads to flawed inferences about significance levels (ibid.). Thus, the standard errors should be adjusted to account for potential heteroskedasticity. The heteroskedasticity-robust standard error can be used for this purpose, which is the square root of the following predicted variance of the coefficient.

$$\widehat{\text{Var}}(\hat{\beta}_j) = \frac{\sum_{i=1}^{n} \hat{r}_{ij}^2 \hat{u}_i^2}{\text{SSR}_j^2},$$

Equation 7 (Wooldridge, 2012)

Robust standard errors are more precise than regular standard errors in the presence of heteroskedasticity and when the sample size is large. When the sample size is small, the robust t-distribution can be a poor approximation of the t-distribution used to estimate significance levels (ibid.) Robust standard errors are commonly used in academic research as data often exhibits heteroskedasticity, and when the sample size is "large", the difference between robust and regular standard errors is negligible (ibid.). The degree of heteroskedasticity will be assessed by the Breusch-Pagan test for heteroskedasticity (BP test). This is an F-test of whether the squared residuals are a function of any of the independent variables as illustrated in Equation 7.

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \ldots + \delta_k x_k + error$$

Equation 7 (Wooldridge, 2012)

The null-hypothesis of homoskedasticity is defined in Equation 8.

$$H_0: \delta_1 = \delta_2 = \ldots = \delta_k = 0.$$

Equation 8 (Wooldridge, 2012)

Thus, we reject the null hypothesis of homoskedasticity if the squared residuals are a function of one or more of the independent variables, which implies heteroskedasticity or non-constant variance of the residuals. In addition to BP-tests, plots of residuals and fitted values of the regression will be used to graphically analyze whether heteroskedasticity is present.

Test results for heteroskedasticity

There was significant heteroskedasticity in all the cross-sectional regressions. The Breusch-Pagan tests for heteroskedasticity rejected the null hypothesis homoskedasticity in all the cross-sectional regressions at the 1% significance level. The presence of heteroskedasticity can also be observed by analyzing the residual plots in Appendix 4. In a homoscedastic residual plot, the observations will

form a band around the residual line (Williams, 2015). This would indicate that the magnitude of residuals does not depend on the fitted values. By contrast, the residuals vary in magnitude as a function of the fitted values when heteroskedasticity is present. The plots in Appendix 4 clearly demonstrate that heteroskedasticity is present in the cross-sectional regressions. Therefore, robust standard errors were used to mitigate the impact of heteroskedasticity. It is more conservative to use robust standard errors in making statistical inferences because, empirically, robust standard errors are typically larger than normal standard errors (Wooldridge, 2012). Smaller standard errors imply smaller p-values, and thus a higher risk of making false positive inferences. In our sample, the normal standard errors are also consistently smaller than the robust standard errors. Hence, our statistical inferences are more conservative as a result of using robust standard errors. This concludes the discussion of pre-regression tests for the cross-sectional regressions. The next section will outline the cross-sectional research design by highlighting the chosen variables and by explaining the rationale for each.

Firm-level cross-sectional regression design

The variables in the cross-sectional regressions are divided into a dependent variable and independent variables. Independent variables comprise the ESG variables of interest as well as control variables.

Dependent variable

Recall that the purpose of this thesis is the investigate the impact of ESG factors on abnormal returns. This thesis uses Jensen's alpha as the definition of abnormal returns. Jensen's alpha, denoted by α, is illustrated in Figure 5 below:

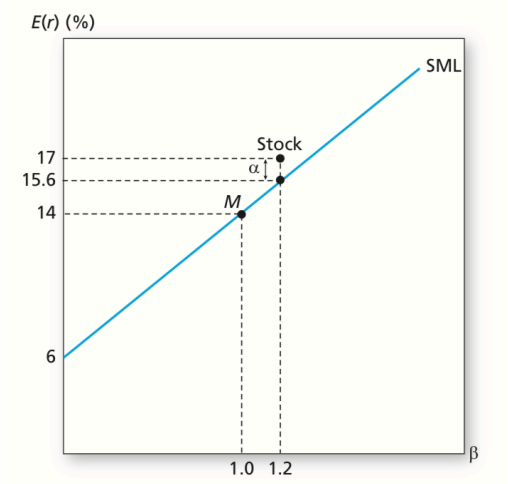


Figure 5; Source: (Bodie, Kane, & Marcus, 2014)

As argued by Jensen in his seminal paper (Jensen, 1967), portfolio managers should be judged for their ability to increase returns relative to the exposure to systemic risk, which cannot be diversified. The exposure to systematic risk is captured by beta. The beta of the market portfolio is one by definition, so portfolio managers holding portfolios with a beta of one should yield a return exceeding the market return to have "beaten the market". The degree to which investors beat the market is given by α in Figure 5 and will be this thesis' measure of abnormal returns.

The 'Empirical Findings' section highlighted several studies concluding that the relationship between ESG and abnormal returns was stronger over the long-term compared to the short-term. Therefore, we hypothesize that the impact of ESG factors on abnormal returns is more long-term than short-term in nature. The dependent variable is Jensen's alpha accumulated over five years to capture long-term investment performance. There is research indicating that choosing a longer time period of ten-years could be fruitful (Shank, Manullang, & Hill, 2005). However, prolonging the time period proliferates the factors besides ESG that influence abnormal returns. Another reason for using five-year rather than ten-year returns is that we use the level of ESG variables in a given year and then test for subsequent abnormal returns. The level of ESG performance is arguably more informative about abnormal returns in the subsequent five-year period compared to a ten-year period. This is because there is more scope for the ESG performance of firms to fluctuate from its initial level over ten years than five years.

Three five-year time periods will be examined: 2003-2007, 2008-2012 and 2013-2017. The specific formula applied to compute the alpha of firm i over a time period of five years ($t = 5$) is highlighted in Equation 9, which is a rearrangement of the CAPM:

$$\alpha_i = TSR_{i,t} - Rf_t - \beta_i(Rm_t - Rf_t)$$

Equation 9 (Jensen, 1967)

The total shareholder return (TSR) is calculated by dividing the end-of-period TSR index value with the start-of-period TSR index value and subtracting one. This yields the percentage growth in the TSR index over the period, which is the total return an investor would receive over the period. The

reason for using TSR instead of stock prices is that TSR includes dividends by assuming the dividends are reinvested in the stock. By including dividends, TSR is a more accurate representation of the total income you would receive as an investor over a given period for holding a given stock.

The risk-free (Rf) rate and the market return (Rm) are retrieved from Kenneth French's website (French, 2019a), which is commonly used in academic research (Sassen, Hinze, & Hardeck, 2016). The Rf rate for US firms is the 1-month T-bill return. The ideal risk-free rate would be one that matches the maturity of the investment to avoid reinvestment risk (Damodaran, 2002). One such risk-free rate could be a five-year US bond, but the Rf rate from Kenneth French's website is applied to US firms despite its limitations. The rationale is that retrieving variables from the same data source ensures consistency in computations and greater comparability with the numerous other studies using the same data source. However, for our sample of European firms, it is more sensible to use a Rf rate in the same currency. Otherwise, there would be exchange rate risk in converting returns to US dollars to buy the 1-month T-bills that are issued in the US. The applied Rf rate is the 12-month London Interbank Offered Rate (LIBOR) rate, which is retrieved from the Federal Reserve Bank of St. Louis (St. Louis Fed, 2019). 12 months is the longest available maturity for LIBOR rates. LIBOR is the interest rate at which major international banks agree to when lending to each other (Moneyland, n.d.). It is not completely risk-free as even major banks can default. The LIBOR rate was deemed preferable to government bonds for each of the five European countries in the sample as these bonds contain widely differing degrees of default risk, ranging from very secure German debt to riskier Italian and Spanish debt. The LIBOR was based on the British pound (GBP) for UK firms and on euro (EUR) for the remaining European firms. The consecutive annual Rf returns are compounded to obtain a five-year cumulative risk-free return.

The market return is retrieved separately for the US and Europe via Kenneth French's website. The reason for using different market portfolios is that US and European firms are exposed to differing location-specific systemic risks that a global market portfolio would not capture. In other words, a downturn in the European equity market is expected to affect European stocks more on average than US stocks. The market portfolio for US firms is a value-weighted portfolio that includes all NYSE, AMEX and NASDAQ firms (French, 2019b). The European market portfolio is a value-weighted portfolio that included all publicly listed firms with available equity data in the 16 European countries listed in Appendix 5. Both market portfolios assume reinvestment of dividends (French, 2019b)

(French, 2019c), which makes the market return comparable to each firm's TSR that also includes reinvestment of dividends. Each firm's beta for a given time period is computed by Equation 10:

$$\beta_i = \frac{\mathrm{Cov}[r_i, r_m]}{\mathrm{Var}[r_m]}$$

Equation 10 (Munk, 2017)

The covariance between the firm's monthly return in excess of the risk-free rate and the relevant market portfolio's monthly return in excess of the risk-free rate is computed. Subtracting the risk-free rate from both the firm and market return is standard practice because, by subtracting the risk-free rate, the intercept, alpha, should be zero in the regression if CAPM holds (Damodaran, 2002). The beta is computed directly without conducting a regression, however, by using Equation 10. A new beta is computed for each firm in each period because a firm's beta, or the sensitivity of a firm's stock price to the market portfolio, can change over time. Beta is generally determined by three factors: the industry in which the firm operates, the firm's degree of operating leverage and financial leverage (Damodaran, 2002). Different industries have different exposures to the equity market. Operating leverage, defined as a high proportion of fixed to total costs, and financial leverage both increase the volatility of a firms operating income, which leads to a higher beta, all else being equal (Damodaran, 2002). Because a firm can change all three factors over time, it is sensible to compute a new beta for each five-year time period.

Now all inputs have been outlined to compute each firm's alpha for a given period. This alpha constitutes the dependent variable in the regression. The following will explain what independent variables are included and why they are included.

Independent variables

The independent variables comprise ESG variables as well and control variables. 15 ESG variables are included in the regression and the description of each is listed in Appendix 6. The intent of this thesis to understand what the underlying drivers are behind E, S and G and abnormal returns. Therefore, the 15 ESG category scores in Appendix 6 are used as independent variables in the regression. The granular categorization of ESG variables allows for interesting analyses of the important variables within the environmental, social and governance categories, respectively. An

important consideration is whether to use the change or level of ESG scores. Specifically, is using the change or the initial level of ESG scores most appropriate to explain the variation of alpha in the five-year time period.

Our hypothesis is that the initial ESG scores is a better predictor of alpha for two key reasons. Firstly, we hypothesize that being in the top 10% on ESG is more important for firm performance than an improvement of, say, five percentage points. This is because an improvement of five percentage points can mean both increasing the ESG score from 20% to 25% or 90% to 95%, which may have different effects on alpha. It was highlighted in the literature review that a high ESG score indicates that stakeholders generally have a positive perception of the firm, which can improve firm performance by strengthening relationships with suppliers, customers and so on. By contrast, a marginal improvement of ESG score is a less clear signal of improved stakeholder relationships. It takes time for the marginal improvement of ESG score to translate into more trust and legitimacy and thus stronger stakeholder relationships. In other words, the initial level is likely to be more informative than the marginal change in ESG score about the strength of stakeholder relationships.

Secondly, most of the ESG variables are expected to have a delayed effect on total shareholder return. A notable exception is the environmental variable 'Resource Reduction', which can increase profitability in the short-term by reducing production costs. In addition, there might be an immediate signaling effect on the stock price of a firm increasing its ESG score, but this signaling effect would be better captured through an event study methodology. Variables related to the other two pillars, 'Social' and 'Governance', can be reasonably expected to have a delayed effect on total shareholder return and thus alpha. A firm might increase its 'Governance' score by changing its board structure, but it can take years for the decisions of the newly structured board to translate into new actions of the firm and for these actions to impact total shareholder return. Likewise, improved relationships with stakeholders, resulting in a firm increasing its 'Social' score, can take years before translating into improved firm performance and thereby total shareholder return. Therefore, we hypothesize that, for example, whether a board if well-functioning at the beginning of the 5-year measurement period is a better predictor of subsequent alpha than changes in the functioning of the board during the 5-year period. For this reason, we use the initial ESG scores of each firm as independent variables in the regression. The ESG scores are measured year-end, so if alpha is computed from 2003 to 2007, the ESG score for 2002 is applied.

Control variables are included in the regression to control for variation in alpha that is unrelated to ESG factors. Control variables in this regression comprise leverage, market capitalization, price-to-book ratio and dummy variables indicating industry and country. The country dummy variable is only included in the regressions for European firms to control for country-specific influences on alpha. The risk of not including important control variables is omitted variable bias (Wooldridge, 2012). In other words, by not including factors influencing alpha, such as leverage, in the regression, the coefficient estimates for the other independent variables become biased. By including leverage as a control variable, the effect of ESG variables on alpha holds constant the differing degree of leverage among firms. More leverage can increase earnings' volatility, which influences alpha (Damodaran, n.d.). Therefore, the debt to total assets ratio is included as a control variable for leverage. Leverage is often used as a control variable in corporate governance research due to its direct influence on firm profitability and performance (Bebchuk, Cohen, & Ferrell, 2009). The other control variables will be discussed in turn in the following.

Market capitalization is included for two main reasons. First, market capitalization reflects investors' expectations of the firm's future profitability because the stock price is a component of market capitalization. Second, market capitalization is a proxy for firm size, which affects future profitability through economies of scale. Market capitalization is log transformed using the natural logarithm to normalize the distribution and reduce the impact of outliers (Wooldridge, 2012). The price-to-book ratio, or the ratio of market capitalization to the book value of shares, is also included as a control variable because it indicates whether a firm is a value firm (low price-to-book ratio) or a growth firm (high price-to-book). In other words, the price-to-book ratio indicates whether a firm is in the growth or mature phase of its lifecycle. If a firm is in the mature phase, it is more likely to be profitable now than in the future and vice versa for growth firms. The price-to-book ratio is thus included because it contains information about the pattern of future profitability. Histograms of price-to-book ratios illustrated a heavily left-skewed distribution with several outliers. To normalize the distribution, the price-to-book ratio is also log transformed.

Dummy variables indicating industry are included to control for variation in alpha due to industry-specific factors. For example, profitability can differ widely across industries. Thomson Reuter's 28 business sectors are used to classify the firms by industry. Choosing industry classification is a trade-

off between broad classifications, that has more firms in each industry, and narrow classifications, that are often more meaningful but contain fewer firms. Appendix 7 highlights the classification hierarchy. At the top, there are 10 economic sectors, then the chosen 28 business sectors, then 54 industry groups and so on. If the industry group classification had been used, with a sample of 219 European firms in 2002, there would be a low average number of firms in each industry and many industries without firms. By contrast, classifying firms into 10 economic sectors would group firms with few similarities. Therefore, using the 28 business sectors was seen as the optimal balance. The exact number of industry dummy variables will depend on the number of industries the given sample of firms encompasses. Additionally, one industry dummy variable is omitted to avoid the dummy variable trap or perfect multicollinearity, which would violate the OLS regression assumption 3 in Appendix 3.

Country dummy variables are incorporated in the regressions for European firms to control for country-related drivers of alpha. Firm profitability might vary across countries, which ultimately translates into alpha by influencing total shareholder return. As mentioned, five European countries are included, so four country dummy variables are used to avoid the dummy variable trap.

Firm-level regression summary

The regression model for the cross-sectional regressions has now been fully described. The purpose of these regressions is to identify ESG variables associated with alpha across firms. The second part of the research design in Figure 3 aims at using the information of significant ESG variables to form positive alpha portfolios. The rationale behind this analysis is exploring whether investors could have used their knowledge of firms ESG scores to form portfolios and subsequently earn abnormal returns. In other words, exploring whether the regression model used to identify ESG variables significantly associated with alpha can also be used practically to form positive alpha portfolios. The following will elaborate the method for forming portfolios and the time-series regressions for testing if the portfolios' alphas are significant, starting with the pre-regression tests.

Time-series pre-regression tests and considerations

The following section will highlight the results of the pre-regression tests for the time-series regressions. The main difference compared with the cross-sectional pre-regression tests is the additional test for autocorrelation. Autocorrelation is a common feature in time-series data,

potentially causing flawed inferences if standard errors are not adjusted to account for it (Wooldridge, 2012). The issues of heteroskedasticity and autocorrelation are grouped because the potential remedy, heteroskedastic and autocorrelation consistent standard errors, or HAC standard errors, addresses both issues.

Heteroskedasticity and autocorrelation

As in the cross-sectional regression, heteroskedasticity can be present in time-series regression (Wooldridge, 2012). In addition, time-series regressions can exhibit autocorrelation, meaning that the error terms are correlated across time (ibid.). This implies that the error terms are not independently distributed, which can invalidate the standard errors used to make statistical inferences. Autocorrelation in our sample would oppose the random walk hypothesis that changes in short-term stock prices are random and unpredictable (Munk, 2017). The random walk theory states that past stock returns are not useful in predicting future stock returns (Fama, 1965). This is equivalent to stating that stock returns do not exhibit significant autocorrelation. If past stock prices could be used to predict the next period's return, an investor could obtain an arbitrage profit simply by trading on this information. However, spurious correlation of the returns is a common feature of time-series regressions (Wooldridge, 2012). This implies that autocorrelation might be observed even if the returns are not meaningfully related.

Similarly to the cross-sectional regression, the degree of heteroskedasticity will be assessed by the Breusch-Pagan test for heteroskedasticity (BP test). Additionally, the Breusch-Godfrey test will be used to test the null hypothesis of no autocorrelation (ibid.). If there is wide-spread heteroskedasticity and autocorrelation in the time-series data, heteroskedastic and autocorrelation consistent standard errors, or HAC standard errors, will be applied.

Test results for heteroskedasticity and autocorrelation

The majority of time-series regressions exhibited both heteroskedasticity and autocorrelation. The wide-spread autocorrelation was somewhat surprising given the expectation of stock returns following a random walk. A plausible reason for the observed autocorrelation is the aforementioned spurious correlation of the returns (ibid.). Our time-series sample of five years contains 60 monthly returns, which is relatively small. This makes it more likely that returns are autocorrelated due to random chance than if more observations had been used. Heteroskedasticity and autocorrelation

(HAC) standard errors were applied across all time-series regressions to mitigate the impact of both heteroskedasticity and autocorrelation. These standard errors, also called Newey-West standard errors, correct for heteroskedasticity and autocorrelation up to a maximum number of lags (Newey & West, 1987). Greene (2012) claims it is empirical practice to use Equation 11 as a rule of thumb to determine the maximum number of lags. This is similar to Newey & West (1994), who proposed $n^{1/3}$ as a rule thumb, where n denotes the number of time-series observations, which is 60 in our case.

$$\#\ of\ lags = n^{1/4} = 60^{1/4} = 2.78$$

Equation 11 (Greene, 2012)

Greene (2012) suggest using the smallest integer equal to or lower than this number. Therefore, the maximum number of lags used to compute the Newey-West standard errors was set to two. This concludes the pre-regression considerations for the time-series regression, shifting the focus towards describing the regression design.

Portfolio-level time-series regression design

The first step for forming the portfolios is establishing the criteria for which firms to include in the portfolios. There are two criteria used to choose the sample of firms. First, the given ESG variable must be significantly associated with alpha at the 5% significance level or below in the given time period and region. Variables negatively associated with alpha are also used because the portfolio of firms scoring high in this variable could be shorted by investors. Second, firms must score above or below certain thresholds, defined as the top 30% or bottom 30%, to be included in the portfolio. These thresholds will be relaxed in the sensitivity analyses to assess the thresholds' impact on the results. Once the portfolios have been formed, separate time-series regressions for each portfolio will be conducted. The following will explain how the time-series regressions are constructed.

Dependent and independent variables

The dependent variable in the time-series regression is the total shareholder return (TSR) in excess of the risk-free rate, Rf_t, for the given portfolio, which is the left-hand side of Equation 12 below:

$$TSR_{i,t} - Rf_t = \alpha_i + \beta_i(Rm_t - Rf_t) + \varepsilon_{i,t}$$

Equation 12 (Jensen, 1967)

This is the formula for computing Jensen's alpha, α_i, which is the intercept in the regression. β_i denotes beta and $\varepsilon_{i,t}$ denotes the error term. Equation 12 highlights the CAPM alpha, where market exposure is assumed to be the only source of systemic risk. We also use Fama French's five-factor model as a robustness-check of whether Jensen's alpha is still significant for a given portfolio when adding more systemic risk factors. The Fama French five-factor model is also applied to see whether the alpha depends on the model specification.

$$TSR_{i,t} - Rf_t = \alpha_i + \beta_i(Rm_t - Rf_t) + s_i SMB_t + h_i HML_t + r_i RMW_t + c_i CMA_t + \varepsilon_{i,t}$$

Equation 13 (Fama & French, 2015)

SMB_t: Small firms' returns minus big firms' returns

HML_t: High book-to-market-value firms' returns minus low book-to-market-value firms' returns

RMW_t: Robustly profitable firms' returns minus weakly profitable firms' returns

CMA_t: Conservatively investing firms' returns minus aggressively investing firms' returns

The purpose of both the CAPM and the five-factor model is to see if alpha is significantly different from zero. If alpha for a portfolio is significantly positive, investors could have earned an abnormal return corresponding to the alpha value over the given time period. If the alpha was significantly negative, investors could have earned abnormal returns by shorting the portfolio. If positive alpha portfolios can be constructed, it indicates that specific ESG scores contain valuable information about future abnormal returns, which is highly useful for investors. However, this does not imply that an ESG variable significantly associated with alpha in one period can be used to form positive alpha portfolios in other or future periods. The wider purpose is testing whether the methodology developed in this thesis, namely identifying significant ESG variables and using these variables to form portfolios, is useful for investors.

Monthly observations are used for both the dependent and independent variables, which yields 60 observations over the five-year time period. An alternative method was using daily observations, which would proliferate the number of observations. Using a daily frequency would yield more variation across the five-year time period, likely leading to more precise coefficient estimates (Wooldridge, 2012). However, daily frequency for estimating beta also introduces more "noise" in the regression because some stocks are more frequently traded than other stocks (Damodaran, n.d.). Low trading frequency implies lower sensitivity, or covariance, to market movements, which directly

impacts the beta estimate as evident in Equation 10. With monthly data, trading frequency has a lower impact on beta because there are fewer stocks being scarcely traded in a month than in a day. Additionally, retrieving daily frequency of total shareholder return data for the 1183 firms in the sample in 2012 over five years would yield around 1.5m observations (1183 firms * 5 years * 253 average trading days per year) versus approximately 71,000 observations using monthly data. Monthly observations thus keep the size of the data set at a manageable level. For these reasons, monthly observations were used in the time-series regression. The monthly risk-free rate for US firms is the 1-month T-bill. For UK firms, the 1-month LIBOR rate based on GBP is applied and LIBOR based on EUR for the remaining European firms. The T-bill rate was retrieved from Kenneth French's website (French, 2019a) and the 1-month LIBOR rate from the Federal Bank of St. Louis (St. Louis Fed, 2019).

No additional control variables are added in the time-series regressions for two main reasons. The Fama French five-factor model is designed to capture the maximum variation in expected returns (Fama & French, 2015). Therefore, adding control variables would likely shift explanatory power across variables rather than improve the collective explanatory power, or R-squared, of the model. The five-factor model is an extension of the CAPM, which implies that the CAPM has omitted variable bias if the five-factor model explains more variation in expected returns. However, using the CAPM and five-factor model without modifications makes this study's findings more comparable to the numerous other studies applying these models.

Both the time-series and cross-sectional regression models have now been fully specified. The two types of regressions supplement each other in answering the research question. To answer the research question affirmatively, additional sensitivity tests will be conducted.

6.4 Sensitivity Tests

The purpose of the sensitivity tests is to assess how robust the results are to changing key aspects of the regression models. In the cross-sectional regression, the time period will be shortened from five years to three years. Another sensitivity test will aggregate the ESG variable scores into higher-level E, S and G scores as well as an overall ESG score to see if grouping variables changes any prior conclusions. In the time-series regression, the breakpoint for firms' ESG performance will change

from top and bottom 30% to top and bottom 10%. The results of the sensitivity tests will be shown following our results in the 'Results' section

3-year time period

As discussed earlier, the timing of ESG variables' impact on alpha might differ. An ESG variable, such as 'Resource Reduction', might reduce production costs and affect abnormal returns in the short-term. By contrast, scoring high on the 'Community' variable could strengthen stakeholder relationships in the long-term. The impact of 'Community' on abnormal returns may thus be more long-term in nature. The timing difference implies that the effect of ESG variables on alpha might be sensitive to the length of the examined period. Therefore, it could be insightful to examine if the significance levels of ESG variables change markedly by shortening the time period.

Aggregation of ESG variables

The literature review highlighted that there might be synergetic effects among ESG variables (Mănescu, 2011). In other words, there is perhaps interaction among ESG variables that yields a stronger relationship with alpha when grouping ESG variables. To test for the presence of synergies, the environmental, social and governance variables will be averaged into E, S and G scores, respectively. Thomson Reuters also computes the higher-level E, S and G scores by taking the simple average of the lower-level category scores (Thomson Reuters, 2015). The E, S, and G scores will also be averaged into an ESG score. The regressions will then be re-run with either E, S and G scores or the overall ESG score. Next, it will be analyzed which variables are significant at the 5% level. More importantly, the magnitude of the coefficient for significant variables will be compared with the lower-level variables' coefficient size to gauge whether there are synergetic effects from grouping variables.

Top and bottom 10% portfolios

The percentile cut-off value when forming portfolios greatly influences both the return, composition and the number of firms in the portfolios. Additionally, the effect of ESG on alpha might be diluted by using a broader 30% cut-off value than the stricter top 10% cut-off value. The drawback of using top and bottom 10% cut-off values is that the portfolios will contain fewer firms and thus the portfolio return be more susceptible to the extreme performance of single firms.

In addition to using sensitivity tests to check the robustness of the results, it is important to address the potential issue of endogeneity thoroughly. The statistical inferences from the cross-sectional model depend on the assumption that ESG affects alpha and not vice versa. The cross-sectional model will be assessed for endogeneity, or reverse causality, between ESG scores and alpha. This will highlight whether ESG scores depend on past abnormal returns in our sample.

6.5 Endogeneity Test

The interpretation of the ESG variables' coefficients depends on the causal direction. If there is reverse causality between ESG and alpha, then the ESG coefficients are not useful in making inferences. ESG studies are generally plagued by issues of endogeneity. As emphasized by Hermalin and Weisbach, empirical work on governance is plagued by endogeneity because all governance variables have a potentially endogenous relationship with firm performance (Hermalin & Weisbach, 2003). Therefore, an indicative test for reverse causality between significant ESG variables and alpha is conducted, inspired by Hirigoyen & Poulain-Rehm (2015). Firms' past five-year alpha is regressed on their ESG performance. The ESG performance is measured in the final year of the five-year period as ESG is measured year-end. This test will indicate whether ESG scores depend on firms' past abnormal returns.

The control variables from the cross-sectional regression could be included as independent variables along with alpha depending on the multicollinearity tests. The rationale is that the control variables might be correlated with the ESG variables and omitting them could thus induce omitted variable bias of the alpha coefficient. Multicollinearity tests such as VIF tests and correlation matrices between alpha and the control variables will be conducted to highlight any multicollinearity among alpha and the control variables.

6.6 Limitations of Methodology

In this section, the choices made in creating the methodology will be scrutinized and the choices' impact on the results will be evaluated. Four key limitations will be discussed in turn. First, the choice to conduct a cross-sectional rather than panel regression. Second, the separation of US and European firms. Third, potentially important variables omitted from the cross-sectional regression. Lastly, the

potential issue of endogeneity or reserve causality between total shareholder performance and ESG scores.

Cross-sectional model

Our methodology has been constructed with the purpose of capturing the long-term effects of ESG scores on abnormal returns. Another method of capturing this relationship would be conducting a panel regression, which essentially conducts the same cross-sectional regression at several points in time (Stock & Watson, 2011). The main advantage of using panel regression, as opposed to cross-sectional regression, would be the ability of the model to capture the effects of both the initial level and change in ESG scores and not either or. In our sample, however, the impact of not capturing the change in ESG scores appears to be limited. This is because the median percentage-point change in ESG scores is low compared to the initial level, which is highlighted for US firms in one of the time periods in Appendix 8. Hence, it is a limitation that our model does not capture the change in ESG scores but, in our sample, the impact of this limitation seems limited. Additionally, panel data would not increase the number of firms in the sample. If a panel data regression was conducted across the time periods from 2003 to 2017, it would result in an unbalanced panel (Wooldridge, 2012). This is because the number of firms with ESG data increases over time. If the same three time periods were used in panel regression as in our cross-sectional regressions, the number of firms in the sample would be unchanged. Another method than panel data of capturing change over time is a pooled cross-sectional regression, where cross-sectional regressions on new, randomly sampled entities each period are pooled (Wooldridge, 2012). However, because this method requires random sampling, it would severely reduce our sample size due to the limited number of firms with ESG data. Therefore, the cross-sectional method with non-random sampling was used to increase the sample size.

Geographic segmentation

The inclusion of the European firms in addition to US firms introduces two main limitations. The first limitation is that it is highly likely that the country-specific sample of firms is not representative of their country of origin. Therefore, broad generalizations about the country-effect of ESG score on alpha cannot be credibly asserted. Additionally, the firms' characteristics might differ by country in ways not accounted for by the country dummy variables or control variables, which would bias the coefficient estimates. Another limitation is that we have categorized firms by country of origin and not by firms' investors' countries of origin. As highlighted in the literature review, investor

perceptions about ESG and firm performance seem to differ across countries. It would be more sensible to categorize firms by investors' countries of origin to analyze differing investor perceptions. However, data on investors' country of origin is unlikely to be available for all investors for all firms. It follows that we would need to either exclude firms or only focus on firms' biggest investors, which is a source of bias, as the biggest investors might have different countries of origin that smaller investors. Additionally, the investors' listed country of origin might not be an accurate reflection of where the investor is actually located. For these reasons, the firms, and not their investors, were categorized by country of origin. As there is likely a correlation between firm and investor country of origin, we can still analyze differing investor perceptions across countries, but the analysis is limited by using firms' country of origin.

Omitted variable bias

Omitting important variables in the regression can induce omitted variable bias, which occurs when left-out variables are correlated with independent variables and important in explaining the variation in the dependent variable (Wooldridge, 2012). This leads to bias in the coefficient estimates for the included independent variables. A potentially important omitted variable is R&D expense. Previous research has found R&D to be important to include in studies investigating the link between CSR and firm performance (McWilliams & Siegel, 2000). This is likely to be the case for our study as well, especially due to our focus on long-term abnormal returns. Regarding omitted variable bias, R&D expense might be correlated with some ESG variables that have a positive influence on abnormal returns. Thus, R&D expense might be the real driver behind increased abnormal returns. R&D expense was not included as a control variable because data was only available for a small subset of firms, which would reduce the sample significantly. An alternative method would be obtaining R&D expense from another database than Thomson Reuter's DataStream such as Bloomberg. However, there is a risk of errors-in-variables in combining data from different databases, where the R&D expense of one firm is attributed to another firm. This is because the firm identifiers, or tickers, are not always compatible across databases. Translating tickers to another database is an inevitable source of error. In addition, R&D expense is treated differently accounting-wise in the US and Europe. In the US under GAAP accounting rules, R&D is expensed and in Europe under IFRS accounting rules, R&D can be capitalized under certain conditions (KPMG, n.d.). Therefore, the R&D expense would not capture all of the R&D conducted by European firms, which would be a source of bias. Due to

these complications, R&D expense was not included as a control variable. There are likely other control variables that could have been valuable to include, which is a limitation of this study.

Endogeneity

Lastly, endogeneity or reverse causality between firm performance and ESG scores is likely to impact our results. In other words, past abnormal returns might have affected the ESG scores. Though we test for endogeneity in our data set, the remedies for potential endogeneity are complex to implement in the context of our study. There are two main ways of mitigating endogeneity in the regression: creating a natural experiment or using one or more instrumental variables. These methods will be briefly explained and the reason for not applying them will be highlighted.

A natural experiment uses an exogenous event, such as a government policy, that affects the independent variable(s) but only affects the dependent variable through the effect on the independent variable(s) (Wooldridge, 2012). Because the change in the independent variable(s) is caused by the policy, and thus not the dependent variable, the observed effect on the dependent variable is more likely to be causal. However, natural experiments are rare regarding ESG because ESG actions are often voluntary firm actions, which is thus often driven by the objective of increased firm performance. Moreover, recall that the purpose of this thesis is to understand which underlying ESG variables are significantly related to abnormal returns. A natural experiment would likely affect only a few ESG variables and not allow for comparison of the effect on abnormal returns across different ESG variables.

An instrumental variable is a variable that is correlated with the independent variable(s) but not correlated with the error term (Wooldridge, 2012). Once this variable has been found, a two-stage least squares (2SLS) regression can be conducted to use the instrumental variable to "cleanse" the independent variable of correlation with the error term or endogeneity (Wooldridge, 2012). However, due to the intertwined relationship between ESG and firm performance, it is notoriously difficult to find instrumental variables for ESG variables. By extension, finding reasonable instrumental variables for all our 15 ESG variables would be infeasible. Therefore, the instrumental variable approach was not applied. Future research in identifying instrumental variables for specific ESG variables would be highly useful. This leaves the issue of endogeneity unresolved. Direct causality between the ESG variables and firm performance cannot be inferred regardless of the conducted

endogeneity test. Past abnormal returns may influence ESG performance in many ways not accounted for by the endogeneity test. Still, investigating the relationship between ESG and abnormal returns can bring valuable insights. The next section will highlight the results of our regressions, followed by a discussion of the plausible drivers of the results.

7. Results

The key insights from the analysis will be conveyed in this section. The section begins with descriptive statistics to provide an overview of the data set and highlight important trends in the data. Subsequently, the results of the cross-sectional and time-series regressions will be highlighted. The four related hypotheses will be tested and collectively they will provide the answer to our research question. Sensitivity tests will then be conducted to assess the robustness of the regression results. The endogeneity test will highlight the prevalence of reverse causality among significant ESG variables. Lastly, the key insights from the analysis will be summarized.

7.1 Descriptive Statistics

The key insights from analyzing the descriptive statistics will be highlighted in this section. The ESG values are from 2002, 2007 and 2012 because it is the ESG values from those years that are used in the regressions. The full descriptive statistics are referred to Appendix 9 and 10 for the US and European summary statistics, respectively.

There are three systemic differences between the ESG scores for US and European firms. The first is that US firms on average score higher on governance variables than European firms in four out of the five governance variables as illustrated below in Figure 6.

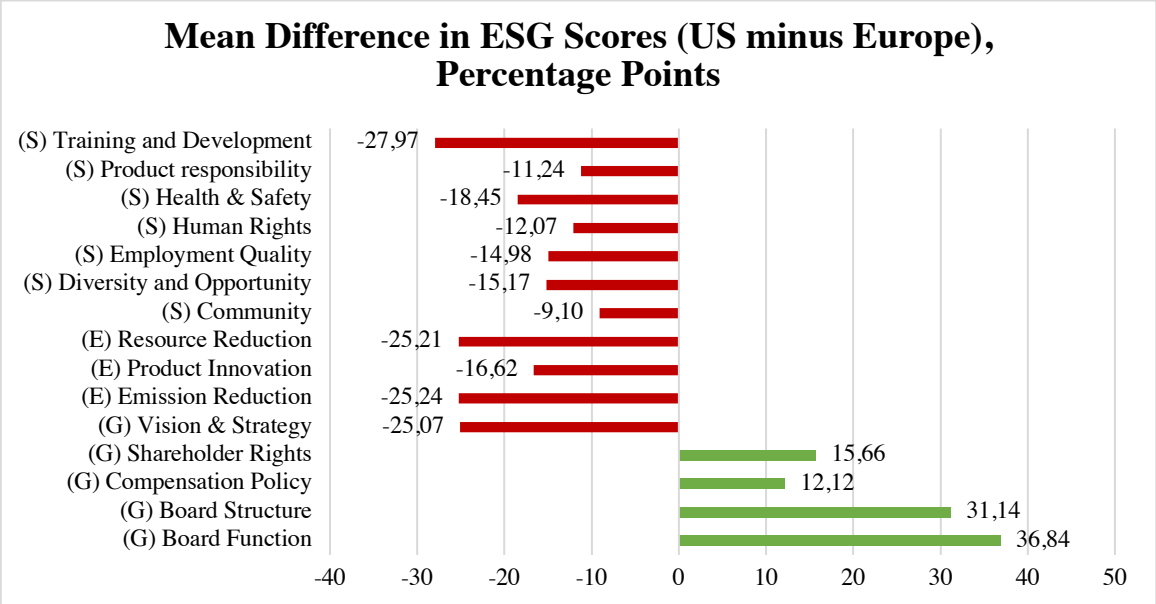


Figure 6

The mean scores, used to compute the mean difference scores in Figure 6, are averages of the firms' mean ESG scores for firms in each respective region in 2002, 2007 and 2012. The mean difference is subsequently computed by taking the mean for US firms and subtracting the mean of European firms, so US firms have higher ESG scores on average on ESG variables with green bars. As evident from Figure 6, European firms score higher on both social and environmental variables, whereas US firms score higher on four of five governance variables. This regional difference is pervasive across time. European firms have higher mean social and environmental scores across all respective variables and in all time periods, and the difference is substantial. On environmental variables, European firms mean score is more than 13 percentage points higher in any time period. The difference between the US and Europe is smaller for social variables with the exception of the 'Training and Development' variable, where European firms mean score is at least 26 percentage points higher than their US counterparts in any time period. In sum, several interesting patterns emerge when comparing the ESG performance of US and European firms. Possible causes for the divergence will be explored in the 'Discussion' section.

Another noticeable trend across the US and Europe is that mean ESG scores generally increase from 2002 to 2007 and further from 2007 to 2012 (Appendix 9 and 10). For the vast majority of ESG variables, the mean ESG scores increases in each subsequent time period. Specifically, in the US 11 of 15 ESG variables increase in each subsequent period and in Europe the trend is even more pronounced with 13 out of 15 ESG variables rising monotonically. All variables in the US and Europe are higher in 2012 than in 2002 except for the variable 'Shareholder Rights' in the US.

The firms' mean alphas develop differently over time in the US and Europe. The mean five-year accumulated alpha for US firms drops from 36% to 23% to 16% in the three consecutive time periods (Appendix 9). European firms' mean alpha rises from 10% to 29% to 64% in each subsequent five-year time period (Appendix 10). However, there is also a common trend among the distribution of firm alphas across the US and Europe. The mean is higher than the median in both regions and in all time periods. This indicates a general positive skew in the alpha distribution, which was confirmed by observing histograms. A positive skew generally implies many smaller values and fewer very large, positive values, illustrated by a long tail to the right (Wooldridge, 2012). This makes sense given there is a lower bound for stock returns but no upper bound. The large, positive values were not erroneous outliers, however, as outlined in the 'Methodology' section. The descriptive statistics

have highlighted the most important trends observed in the data. The section below will explain the key findings of the cross-sectional regressions.

7.2 Cross-sectional regression results

The purpose of this section is to utilize the regression results to answer the stated hypotheses. The regression results can be found in Table 3 below or in a regression table with coefficient standard errors in Appendix 11. The key insights from the cross-sectional regression will be presented next. Afterward, the overall regression model performance will be scrutinized, followed by an evaluation of each hypothesis.

Cross-sectional Regression

	(1) **US Alpha 03-07**	(2) **US Alpha 08-12**	(3) **US Alpha 13-17**	(4) **Europe Alpha 03-07**	(5) **Europe Alpha 08-12**	(6) **Europe Alpha 13-17**
Board Function (G)	-0.34	-0.20	0.52*	1.26**	-0.09	0.26
Board Structure (G)	0.15	-0.11	-0.52	-1.13**	0.50***	-0.12
Compensation Policy (G)	-0.12	0.20	0.32	-0.24	0.39**	0.21
Shareholder Rights (G)	-0.01	-0.03	-0.18	-0.42	-0.03	-0.11
Vision and Strategy (G)	0.32	-0.21	0.55	-0.38	-0.38	0.60
Emission Reduction (E)	-0.99	0.18	0.26	-0.06	-0.16	-0.01
Product Innovation (E)	1.29**	-0.21*	-0.37*	0.97*	0.05	-0.34
Resource Reduction (E)	-0.24	-0.24	-0.39	0.76	0.40	-0.24
Community (S)	-0.16	-0.19	-0.03	0.92*	-0.09	-0.32
Diversity and Opportunity (S)	0.97**	-0.08	0.41*	0.52	0.17	0.85
Employment Quality (S)	-0.09	0.25**	0.20	1.18**	0.04	0.17
Human Rights (S)	0.41	0.16	-0.23	-0.35	-0.05	-0.19
Health and Safety (S)	0.09	0.03	-0.56**	-0.93*	0.04	-0.04
Product Responsibility (S)	-0.02	-0.07	-0.21	-0.03	-0.39***	0.03
Training and Development (S)	0.14	0.36**	0.14	-0.20	0.07	-0.03
Leverage Ratio	0.06	-0.09	-0.31	-0.90	-0.34	-0.43
Market Capitalization (log)	-22.40***	-1.25	10.14**	-28.42**	-1.75	-23.36*
Price-To-Book Ratio (log)	-15.98	-0.70	-4.47	-35.65**	3.89	-8.16
Constant	435.59**	52.53	-103.86	407.65**	72.00	450.23***
Obs.	290	480	681	219	419	478
R-squared	0.45	0.30	0.19	0.39	0.27	0.17
Adjusted R-squared	0.35	0.24	0.14	0.25	0.19	0.09
Root mean square error	133.67	67.85	110.92	155.00	68.63	131.66

*** p<0.01, ** p<0.05, * p<0.1
Industry and European country dummy variables omitted for brevity

Table 3

Three interesting findings emerge when analyzing Table 3, which each relate to the hypotheses 1-3. First, there are several ESG variables associated with alpha at the 5% significance level. In total there are 11 ESG variables are significant at the 5% level and eight of the variables have a positive association. The second pattern is that no variables are significant across time periods. Lastly, there is a divergence between the US and Europe in which ESG variables are significant in a given time period. In fact, no ESG variables are significant in the same time period in the US and Europe. These patterns will be analyzed in greater detail for each related hypothesis. Prior to assessing the hypothesis, the regression model performance will be assessed.

Regression model performance

The overall cross-sectional model performance will be assessed on its ability to explain variation in alpha and its precision in predicting alpha. The explained variation in alpha is expressed by the adjusted R^2 value. Unlike R^2, adjusted R^2 penalizes having insignificant regressors (Wooldridge, 2012). This feature makes adjusted R^2 more meaningful than R^2 when the model contains many regressors. In the US, the adjusted R^2 declines from 35% to 24% to 14% in the three consecutive time periods as evident in Table 3. A similar pattern of declining adjusted R^2 as a function of time emerges in Europe. In this region, the adjusted R^2 declines from 25% to 19% to 9% in the three consecutive time periods. This pattern indicates that ESG variables explain less variation in alpha in later time periods, although control variables could also be contributing to the lower adjusted R^2.

The precision of the cross-sectional regression model in predicting alpha is assessed using the root mean square error (RMSE). RMSE is the standard deviation of the forecasting errors (Wooldridge, 2012). This metric is scaled according to the dependent variable and thus yields valuable insight into how precise the model is. The RMSE ranges from 134% to 68% to 111% in the three consecutive time periods in the US (Table 3). In Europe, a similar pattern is evident with a "U-shaped" relationship across the three time periods. The RMSE in Europe declines from 155% to 69% and then rises to 132% in the last time period. The regression model is thus better at predicting alpha in the middle time period from 2008-2012. For investment purposes, the RMSE is quite high compared to the mean alpha in each regression. The mean alpha in the US ranges from 35.85% to 16.82% and in Europe the mean alpha value ranges from 10.45% to 63.92% (Appendix 9 and 10). Comparing the RMSE to the mean alpha suggests that the predictive power of the model is low. However, the primary purpose of

constructing the model was exploring how ESG relates to alpha and not predicting alpha as precisely as possible.

The above sections have provided an overview of the most important findings. Now a more detailed analysis will be presented under each of the three hypotheses related to the cross-sectional analyses.

Hypothesis 1

There is a positive relationship between one or more ESG variables and abnormal returns, alpha, at the 5% significance level.

The first hypothesis is not rejected. Eight variables were significantly positive at the 5% significance level across time periods and regions, which is illustrated in Table 3. Only three variables were negatively associated with alpha at the 5% significance level. In addition to significance levels, it is important to examine the magnitude of the significant coefficients. A significant variable's associated coefficient can be so small that it loses economic significance. The magnitude of the significant, positive coefficients ranges from a maximum of 1.29 for 'Product Innovation' in the US from 2003-2007 to a minimum of 0.25 for 'Employment Quality' in the US from 2008-2012. The magnitude of negative, significant coefficients ranges from -1.13 to -0.39 for 'Board Structure' in Europe from 2003-2007 and 'Product Responsibility' in Europe from 2008-2012, respectively. A coefficient of 1.29 implies that one additional percentage point of the 'Product Innovation' score in 2002 was associated with a 1.29 percentage point increase in the accumulated alpha from 2003-2007. In other words, a US firm increasing its 'Product Innovation' score in 2002 by ten percentage points would on average increase its accumulated alpha from 2003-2007 by 12.9 percentage points. That corresponds to an annualized alpha of 2.5%. The economic significance of this annualized alpha can be assessed by comparing the return to the risk-free rate. The risk-free rate is an imperfect but meaningful benchmark return because neither alpha nor the risk-free rate is exposed to systemic risk. The annualized alpha return of 2.5% was above the annual risk-free rate in 13 of the 17 years from 2001-2017 (French, 2019a). The annual risk-free rates are based on the return from compounded 1-month T-bills. In the five-year period from 2003-2007, the risk-free rate was higher in three of the five years due to the relatively high interest rates prior to the Great Financial Crisis. Still, the observed effect of 'Product Innovation' is arguably still sufficiently large to hold economic significance for investors.

It should be cautioned that the precision of the coefficient estimates decreases with multicollinearity. It was established earlier that there is significant multicollinearity among the ESG variables, so the precision of the coefficient is mitigated by the degree of multicollinearity. However, the degree of multicollinearity among the 11 significant variables is limited, which can be observed in Table 2. The VIF of the significant variables ranges from 1.46 to 2.10, which is equivalent to a squared VIF range of 1.21 to 1.45. It follows that the standard errors of the significant variables are inflated by between 21% and 45% as a result of multicollinearity. Larger standard errors expand the confidence intervals that contain the true value of the coefficients (Wooldridge, 2012). The low squared VIF range indicates that multicollinearity is not a major issue among of the significant variables, which increases the reliability of the coefficient estimates.

Hypothesis 2

No ESG variable is significant at the 5% level in both the US and Europe in the same time period.

This hypothesis cannot be rejected either. 'Product Innovation' during 2003-2007 nearly rejected this hypothesis as it was significant at the 5% level in the US but only at the 10% level in Europe. The coefficient in Europe was not statistically different from zero, as its 95% confidence interval contained zero. Only the variable 'Employment Quality' was significant at the 5% level in both the US and Europe, albeit in different time periods. The sign of the coefficient was positive in both regions.

An interesting pattern emerged regarding the significant variables in each region. Recall from the descriptive statistics section that European firms scored substantially lower than US firms on most governance variables and European firms scored substantially higher on all social and environmental variables. The emerging pattern is that in nine out of the eleven significant variables, the significant variable's region had a comparatively lower mean score than the other region on the given variable. It is evident from Figure 6 that European firms score lower on four of the five governance variables and all significant governance variables are concentrated in Europe. Further, Figure 6 showed that US firms score lower on all social and environmental variables and all significant environmental and social variables are in the US. The exceptions are the two social variables 'Employment Quality' and 'Product Responsibility'. The variable 'Employment Quality' is significant in both regions and

'Product Responsibility' is significant in Europe. The interpretation of this interesting finding is referred to the 'Discussion' section.

Hypothesis 3

No ESG variable is consistently either positive or negative at the 5% significance level across consecutive time periods in the same region.

This hypothesis could not be rejected. Two variables showed some temporal persistence but not enough to reject the hypothesis. 'Product Innovation' in the US was significantly positive at the 5% level from 2003-2007 but negatively associated with alpha at the 10% level in the following period. Likewise, 'Board Structure' was negatively related to alpha at the 5% significance level from 2003-2007 but positively related to alpha at the 1% level in the subsequent period. This illustrates that significant variables were highly inconsistent across both regions and time periods. The following section will highlight the key findings of the time-series regressions.

7.3 Time-series regression results

The time-series regressions resulted in only a small fraction of portfolios with a significant, positive alpha. Recall that the portfolios were formed based on the significant ESG variables in the cross-sectional regressions. Two portfolios were formed for each significant ESG variable: one with the firms scoring in the top 30% on the ESG variable and the another of the firms in the bottom 30%. These portfolios were then tested for a positive alpha, indicated by the intercept, using both the CAPM and Fama and French's five-factor model. Only two portfolios had significant alphas, which indicates that the suggested trading strategy was unsuccessful. These two portfolios have highlighted alphas in Table 4 and 5, respectively. The tables for US portfolios are not shown as no US portfolio had a significant alpha.

Europe Time-series CAPM

ESG variable portfolios	Market premium	Alpha	Adjusted R2
Board Function (G) 2002 Top 30%	0.65***	0.00	0.33
Board Function (G) 2002 Bottom 30%	0.81***	0.00	0.36
Board Structure (G) 2002 Top 30%	0.64***	-0.00	0.31
Board Structure (G) 2002 Bottom 30%	0.78***	0.00	0.36
Employment Quality (S) 2002 Top 30%	0.63***	0.00	0.30
Employment Quality (S) 2002 Bottom 30%	0.77***	-0.00	0.40
Board Structure (G) 2007 Top 30%	**0.70*** **	**0.01** **	**0.45**
Board Structure (G) 2007 Bottom 30%	0.78***	0.00	0.47
Compensation Policy (G) 2007 Top 30%	0.74***	0.01*	0.44
Compensation Policy (G) 2007 Bottom 30%	0.74***	0.00	0.49
Product Responsibility (S) 2007 Top 30%	0.69***	0.00	0.36
Product Responsibility (S) 2007 Bottom 30%	0.80***	0.01	0.49

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

Table 4

Europe Time-series Fama-French Five-factor

ESG variable portfolios	Market Premium	SMB	HML	RMW	CMA	Alpha	Adjusted R2
Board Function (G) 2002 Top 30%	0.55***	0.02	-0.32	-0.83***	-0.38	0.01	0.31
Board Function (G) 2002 Bottom 30%	0.73***	0.01	0.05	-0.56	0.05	0.00	0.34
Board Structure (G) 2002 Top 30%	0.52***	0.05	-0.24	-0.96**	-0.44	0.00	0.31
Board Structure (G) 2002 Bottom 30%	0.72***	0.01	0.07	-0.44	0.10	0.00	0.33
Employment Quality (S) 2002 Top 30%	0.56***	-0.16	-0.07	-0.54	-0.15	0.01	0.27
Employment Quality (S) 2002 Bottom 30%	0.66***	0.16	-0.24	-0.90***	-0.07	0.00	0.40
Board Structure (G) 2007 Top 30%	0.45***	-0.91**	0.76	-0.17	-0.78	0.01*	0.51
Board Structure (G) 2007 Bottom 30%	0.40**	-1.12**	0.96*	-0.45	-1.40*	0.01	0.56
Compensation Policy (G) 2007 Top 30%	**0.43****	**-0.83****	**0.87**	**-0.29**	**-0.95**	**0.02****	**0.49**
Compensation Policy (G) 2007 Bottom 30%	0.42***	-1.05**	0.87	-0.33	-1.12	0.01	0.58
Product Responsibility (S) 2007 Top 30%	0.35**	-1.19***	0.83	-0.51	-1.21*	0.01	0.55
Product Responsibility (S) 2007 Bottom 30%	0.47***	-0.78*	1.02	-0.14	-1.11	0.01*	0.55

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

Table 5

The one significant portfolio is the European firms scoring in the top 30% of 'Board Structure' in 2007 as highlighted in Table 4. The observed CAPM alpha was significant at the 5% level over the 2008-2012 period, but only at the 10% level in the Fama French five-factor regression (Table 5). The other significant portfolio is the top 30% 'Compensation Policy' portfolio consisting of European firms (Table 5). The observed alpha at the 5% level was based on Fama and French's five-factor model, but only at the 10% level in the CAPM regression (Table 4). Hence, the significance levels shift depending on the regression model. This illustrates the importance of using several regression models to ensure that the results are robust across models. The following section will explain the results of the time-series portfolio regressions in relation to the fourth hypothesis of this thesis.

Hypothesis 4

Positive alpha portfolios can be formed by using significant ESG variables as a screening criterion.

This hypothesis cannot be rejected. Two portfolios registered a positive alpha at the 5% level, but as a trading strategy, the success rate is low given that 22 portfolios were formed. These portfolios were tested for a significant alpha using both the CAPM and the five-factor model, resulting in 44 regressions. That two regressions out of 44 resulted in a significant alpha implies a rather low success rate of 4.5%. The magnitude of the alpha in both regressions was small from a practical trading perspective. For 'Board Structure' the alpha coefficient was 0.0096, implying that investing in this portfolio would yield a monthly abnormal return of 0.0096%. This corresponds to a compounded annual abnormal return of 0.1225%. The alpha for 'Compensation Policy' was larger but still small in magnitude. 'Compensation Policy's alpha was 0.0158, which corresponds to a monthly abnormal return of 0.0158% and an annual compounded abnormal return of 0.2064%. The relatively low portfolio alphas imply that a large investment is needed to capitalize on the opportunity. However, there is a reason to be skeptical about the observed alphas. A large number of conducted time-series regressions implies a high risk of false positives. Specifically, the probability of obtaining a false positive result, or type 1 error, is given by:

$$Probability\ of\ one\ false\ positive\ test = 1 - (1 - \alpha)^{\gamma}$$

$$\alpha = significance\ level$$

$$\gamma = number\ of\ tests\ conducted$$

Equation 21 (Goldman, 2008)

Applying the above formula with our 44 regressions, the probability of one intercept, or alpha, being falsely positive is 89.5% with a 5% significance level and 35.7% with a 1% significance level. It follows that there is a considerable risk that the observed significant, positive alpha portfolios are due to random chance.

In sum, our trading strategy of filtering firms on their ESG performance on particular ESG variables associated with alpha did not turn out to be effective. The four hypotheses have been tested and neither could be rejected. The focus will shift towards testing the sensitivity of our results by tweaking essential elements in our regressions.

7.4 Sensitivity tests

The sensitivity test will assess the robustness of the results. The sensitivity tests for the cross-sectional regression will be conducted first, followed by a sensitivity test for the time-series regressions. Two central aspects of the cross-sectional regressions are the chosen time-horizon of five years and the decision to use 15 ESG variables rather than aggregated E, S and G scores or an overall ESG score. The five-year time period will be shortened to three years and the ESG variables will be aggregated to E, S and G scores and subsequently to an ESG score. The sensitivity test of the time-series regression results will be assessed by changing the cut-off value for portfolios from the top and bottom 30% to the top and bottom 10%.

Three-year time period

Shortening the time period changed the significant variables considerably. Of the 11 significant variables in the cross-sectional regressions, only three were still significant at the 5% level when shortening the time period to three years (Appendix 12). This illustrates that the regression results are highly sensitive to the chosen time period. There was a total of ten significant ESG variables at the 5% significance level, so the number of significant variables was slightly lower with the shortened time period. Though the kind of significant ESG variables changed, the three earlier established patterns persisted. First, for nine out of the ten significant variables, the significant variable's region had a comparatively lower mean score than the other region on the given variable. Second, no ESG variables were significant in the US and Europe in the same time period, supporting that Hypothesis

2 could not be rejected. Third, no ESG variables were significant at the 5% level in consecutive time periods, which supports the failure to reject Hypothesis 3. The robustness of these patterns across both five-year and three-year time periods is striking and warrants further examination in the 'Discussion' section.

Aggregation of ESG variables

Four variables were significant when grouping ESG variables into E, S and G scores (Appendix 13). It is insightful to compare the coefficient sizes of the higher-level E, S and G scores and their related lower-level ESG variables. Only two of the significant variables had significant lower-level ESG variables in the same time period. For example, the 'Social' score was significant in the US from 2003-2007 and in the primary cross-sectional regression the social variable 'Diversity and Opportunity' was also significant in the US from 2003-2007 (Appendix 11). This allows for a comparison of the coefficient sizes of the lower-level 'Diversity and Opportunity' score and the higher-level 'Social' score. The 'Social' score's coefficient was 33% larger than the 'Diversity and Opportunity' score's coefficient. The 'Social' score in Europe from 2003-2007 was also significant and along with the lower-level social variable 'Employment Quality'. The 'Social' score's coefficient was 19% larger than 'Employment Quality's coefficient. The larger coefficient of the 'Social' score is logical because it contains more information than the lower-level social variables. However, the Social score also contains a lot of noise as it averages seven social lower-level variables, where most often none or only one of the lower-level variables are statistically significant. Therefore, the Social score contains more information but much of this additional information is partly unrelated to alpha. The higher-level E, S and G variables coefficients are larger, which indicates potential synergetic effects of combining ESG variables. The drawback is that higher-level E, S and G variables also contain more noise as they average several lower-level ESG variables that are not significantly related to alpha.

The explanatory power of the regression model with E, S and G scores is slightly poorer than its comparable regression model with lower-level ESG scores. The adjusted R^2 is slightly lower and root mean square error is slightly higher of the regression model with E, S and G compared to the regression model with lower-level ESG variables, but the deviations do not exceed two percentage points. This shows that decomposing the E, S, and G scores in lower-level scores explains slightly more variation in alpha and enables a marginally more precise prediction of alpha.

Averaging the E, S and G scores into an ESG score results in the ESG score being positively significant in a single time period, which is in Europe from 2003-2007 (Appendix 13). It makes sense that this particular time period is significant given that both the 'Social' and 'Environmental' variable were positively significant in the same time period. The significant ESG score's coefficient size of 2.62 is almost twice as large as either the 'Social' or 'Environmental' coefficient sizes in the same period. However, the ESG score also incorporates the 'Governance' score, which was not statistically significant in the period but had a negative coefficient of -1.62. This highlights that using an aggregate ESG score has the benefit of capturing some synergetic effect among ESG variables, resulting in a larger coefficient. The drawback is that the ESG score bundles elements that have both positive and negative effects on alpha. Without breaking the ESG score down into its components, it is unknowable which subcomponent is really driving the effect on alpha.

The adjusted R^2 is marginally higher and the root mean square error is marginally lower on average in the regression with ESG scores than the regression with E, S and G scores. This supports the earlier established pattern that breaking down the ESG score enables the regression model to explain slightly more variation in alpha and predict alpha a bit more precisely.

Top and bottom 10% portfolios

In the top 10% portfolios, a single portfolio had a significant, positive alpha at the 5% level (Appendix 14). The portfolio consisted of the firms scoring in the top 10% on the governance variable 'Board Structure' in Europe from 2008-2012. Interestingly, the corresponding top 30% portfolio for the same variable, time period and region also registered a positive alpha at the 5% significance level in the primary time-series regression. In other words, the variable is significant at the 5% level whether the cut-off value is top 30% or top 10%. The variable was associated with alpha at the 1% level in the cross-sectional regression. The robustness of this particular variable's association with alpha is noteworthy and will be explored further in the 'Discussion' section.

7.5 Endogeneity Test

The chosen ESG variables to test for endogeneity were the 11 significant ESG variables in the cross-sectional regressions (Appendix 11). It is of particular interest to explore whether these significant

variables are affected by endogeneity because we make inferences based on these variables' coefficients. Unfortunately, only six out of the 11 significant variables could be tested as data on 'Total Shareholder Return' was unavailable from 1998-2002. It follows that the alpha from 1998-2002 could not be computed and, by extension, that the ESG variables from 2002 showing significant association with alpha over the 2003-2007 period could not be tested for endogeneity. Thus, the six ESG variables tested for endogeneity were significantly associated with alpha either over the period from 2008-2012 or from 2013-2017.

The endogeneity tests showed that the null hypothesis of no endogeneity between ESG scores and prior five-year alpha could not be rejected (Appendix 15). No alpha values were significantly associated with either of six ESG variables at the 5% level. This test does not provide conclusive evidence about endogeneity. However, the test is useful in highlighting that current ESG performance does not seem to depend on past abnormal returns in our sample. The control variables from the cross-sectional regression were included because they did not induce multicollinearity, evident by VIF values for the alpha variables close to one.

7.6 Summary of Results

This section will summarize the key insights of the 'Results' section. The descriptive statistics revealed a stark contrast in ESG scores between the American and European firms. European firms scored higher on all environmental and social variables along with the governance variable 'Vision and Strategy'. It follows that US firms scored higher on the remaining governance variables.

Four hypotheses were tested, and no hypothesis could be rejected. The results of the cross-sectional regression illustrated that 11 ESG variables were significantly correlated with alpha in Europe and the US across the three examined time periods. The first hypothesis stated a positive relationship between ESG variables and alpha. This could not be rejected as eight ESG variables across the US and Europe were positively correlated with alpha. The second hypothesis stated that no ESG variable would be significant in the US and Europe during the same time period, which could not be rejected either. The third hypothesis stated no ESG variables to be significant in consecutive time periods. This hypothesis could likewise not be rejected. Lastly, the fourth hypothesis stated that portfolios based on significant ESG variables would have a significant alpha. The results of the time-series

regression illustrated that two of the portfolios based on significant ESG variables from the previous cross-sectional analysis yielded a significant positive alpha. Specifically, the portfolios based on the top 30% of 'Compensation Policy' and 'Board Structure' provided a significant alpha from 2008-2012 at the 5% level. The fourth hypothesis could thus not be rejected. However, considering the large number of portfolios and regressions, there is a considerable risk of obtaining false positives, meaning that the observed positive portfolio alphas could be due to random chance.

Several sensitivity tests were also conducted to assess the robustness of the results. The first sensitivity test showed that the significance of ESG variables are highly sensitive to shortening the time period from five years to three years. Interestingly, the three-year sensitivity test supported our conclusion of hypothesis 2 and 3. No ESG variables were significant in both regions in the same time period and no ESG variables were significant in consecutive time periods. The second sensitivity test showed that the ESG variable's coefficients grow larger when aggregating ESG variables to higher-level E, S and G scores as well as a combined ESG score. The drawback is that the aggregated ESG scores use lower-level ESG variables not related to alpha. Additionally, the underlying drivers of the relationship between an ESG score and alpha are unclear unless the ESG score is broken into components. Thirdly, the cut-off value for forming the portfolios was adjusted from the top and bottom 30% to the top and bottom 10%. One portfolio, the top firms on 'Board Structure' in Europe from 2008 to 2012, recorded a significant, positive alpha both with the top 30% and top 10% of firms. An endogeneity test was conducted for the significant ESG variables and could not reject the null hypothesis of no endogeneity. However, this was only an indicative test, so the results might still be affected by endogeneity. The plausible theoretical reasons for the results of this thesis will be discussed in the following section.

8. Discussion

The results section highlighted the outcomes of the two regression analyses, which enabled us to comment on each of the four hypotheses of this thesis. The results of the cross-sectional analysis illustrated that some ESG variables were significantly correlated with firm alpha. It was further shown that significant variables were conditional on the region and time period examined. The subsequent time-series regression highlighted that the significant relationships between ESG variables and firm alpha could be utilized in investor strategies to form portfolios with significant alphas on two occasions. Specifically, 'Board Structure' and 'Compensation Policy' could be utilized to create significant alpha portfolios in Europe from 2008-2012. The following sections will examine the results of the regression analyses and its implications for the hypotheses by applying relevant theories. Firstly, the general difference between ESG scores in Europe and the United States will be discussed. This will allow a subsequent discussion of the drivers of significant ESG variables in each region. Thereafter, potential explanations for why significant variables change over time will be presented. Finally, the results of the portfolio analysis will be discussed and interpreted by applying different theoretical perspectives. The following discussion of results will provide valuable perspectives regarding the causes and implications of our findings.

8.1 Impact of Institutional Differences

The descriptive statistics section illustrates that there are some significant differences between the ESG scores of European and American firms. Generally, American firms score higher on governance variables across all examined time periods, while European firms score significantly higher on environmental and social factors across all examined time periods (Appendix 9 and 10). The clear trends in ESG scores among the two regions can be partly explained by institutional theory, which argues that organizations are significantly influenced by the institutional setting in which they operate (Doh & Guay, 2006). Accordingly, the actions of firms in Europe and the United States will be affected by the isomorphic pressures in each region. Doh and Guay (2006) argue that institutional differences between Europe and the United States have resulted in different perceptions regarding the validity, relevance and acceptance of ESG initiatives. The institutional differences between the United States and Europe emanate from a range of different political, social and economic experiences. Europe has been highly influenced by the two World Wars in the previous century, the

desire for political and social assimilation among countries, the formation of the European Union and a range of other factors too numerous to describe in detail in this thesis (ibid.). Institutional differences do exist within Europe due to country-specific disparities, but for the purpose of this thesis, the region will be treated as one. This implies that certain arguments in the following discussion will be more applicable to specific European countries than others. The institutional environment in the United States has been influenced by the early political history of the country, which emphasized a rather decentralized political structure in order to limit the power and dominance of the central government (ibid.). The fundamental institutional differences between the United States and Europe impact the difference in both social, environmental and governance scores of companies, which will be examined in the next couple of sections.

Differences in social scores

The higher social scores in Europe can be attributed to a number of region-specific institutional differences between Europe and the United States. Doh and Guay (2006) emphasize that interest groups in the United States have no formal or traditional standing in the regulative process. The creators of the American Constitution specifically sought to design a system that limited the ability of interest groups to dominate the political process (ibid.). Certain interest groups exercise power by lobbying politicians to shape specific public policies. However, the opportunities to influence public policy are generally diffused because of the country's federal structure and the separation of powers among three national branches (ibid.). In contrast, European interest groups often have an institutionalized place in the regulative process. This implies that a wide range of views is considered in the political process, which generally results in more generous welfare contributions and the inclusion of more social issues. In the United Kingdom and Germany, funds and charities are required to disclose social and ethical investment policies (Renneboog, Horst, & Zhang, 2008a). The "New Economic Regulations" in France requires listed companies to publish specific social information in their annual reports. Such regulation is not established in the United States, which suggests that there is less regulative isomorphic pressure for companies to assume an active social role. Section 406 of the Sarbanes-Oxley Act of 2002 does require listed companies to disclose a written code of ethics by the CEO and CFO (ibid.). However, this is less extensive than the comparative regulation in most European countries.

The Chartered Financial Analysts Institute (2015) performed a survey on their members with the purpose of inspecting the attitude towards ESG among global investors and asset managers. The survey illustrates that 57% of investors from EMEA take social issues into account in their investment analysis and decisions. Comparatively, only 46% of American investors include social issues in their strategies. The survey indicates that American firms are less pressured by investors to score high on social ESG scores, which supports that the difference in isomorphic pressures can help explain the general difference in social ESG scores.

Lastly, Hofstede Insights (2019) illustrates that the United States scores significantly higher on 'Individualism' compared to European countries. The site states that in individualistic societies, people are only supposed to look after themselves and their direct family. Contrarily, in more collectivistic societies people belong to groups, such as organizations, that take care of its members in exchange for a high degree of loyalty (ibid.). This illustrates the impact of the cognitive-cultural pillar, as the cultural difference between the United States and Europe affects the actions and processes of companies, which results in a general difference in social scores. In combination, these general institutional differences can somewhat explain why European companies score significantly higher on social factors across all examined time periods.

Differences in environmental scores

The lower environmental scores of American companies can also be explained by institutional differences. The aforementioned survey by the Chartered Financial Analysts Institute (2015) also illustrates that 45% of American investors consider environmental issues in their investment strategies, while the equivalent number is 58% for investors from EMEA. Hofstede Insights (2019) further illustrates that the United States scores significantly lower on 'Long-term Orientation' compared to European countries. The environmental scores are highly concerned with reducing the impact of climate change, which is a long-term issue. Accordingly, improving environmental scores will likely sacrifice short-term profits to achieve a sustainable and long-term future, which is not considered as much in the United States as in Europe. Markman (2018) argues that Americans often overvalue the short-term benefits relative to long-term benefits. The article states that Americans tend to downplay the impact of climate issues in order to maintain their living standards and habits, while corporations tend to keep manufacturing costs low by avoiding investments in new processes that limit carbon emissions. This is an example of normative isomorphism. The regulatory pressures on

environmental issues are further less significant in the United States compared to Europe, as corporations face fewer transparency and reporting requirements. In Germany, the "Renewable Energy Act" from 2001 provides tax advantages when investing in renewable energy projects (Renneboog, Horst, & Zhang, 2008a). In contrast, the United States has withdrawn from the Paris Agreement, which requires countries to report on emission goals and implementation efforts (Mooney, 2018).

Differences in governance scores

The higher governance scores of American companies can be explained by institutional differences as well. The above section stated that the United States scores low on 'Long-term Orientation' compared to European countries. This cultural difference results in different isomorphic pressures. The difference in long-term focus is exemplified by the fact that the only governance factor where European firms score higher than American firms is 'Vision and Strategy'. The description of the 'Vision and Strategy' factor emphasizes the commitment and effectiveness of firms in integrating ESG aspects into both the overarching strategy and decision-making processes, which is long-term in nature. The difference in long-term orientation is further illustrated by the difference in sources of capital. American companies derive major sources of capital from the stock market, where investors often require a short-term profit. In contrast, European companies typically have a small number of large institutional investors, which have a more long-term focus (Danko, Goldberg, Goldberg, & Grant, 2008). The relatively high corporate governance scores in the United States can be partly explained by the difference in ownership structures. Dharmastuti and Wahyudi (2013) argue that the ownership structure can have a significant impact on corporate governance. Specifically, the article argues that institutional owners tend to do a better job at monitoring management compared to individual investors. Institutional investors usually own a larger proportion of shares and are thus better incentivized to monitor compared to an individual investor, which can reduce agency problems (ibid.). As a result, a fragmented ownership structure necessitates stronger governance mechanisms to ensure owners are protected from potential opportunism of management.

Thompson Reuters database scores companies based on five different governance variables, none of which take ownership structure into account in the scoring system (Appendix 6). Thompson Reuters further value specific governance mechanisms, such as having a large proportion of non-executive directors on the board and key committees, as being best-practice. In the US, most large, listed firms

have a majority of non-executive directors on the board (Davies, 2000). Accordingly, American companies score significantly higher on corporate governance factors compared to European companies in all examined periods, as the scoring criteria favor American corporate governance practices. However, this does not necessarily imply a lower level of agency problems in the United States due to the significant effect of governance exerted by owners in Europe.

In conclusion, the clear difference in ESG scores between Europe and the United States can be largely explained by significant institutional differences. These differences further indicate that the value-drivers of ESG initiatives will likely differ between the two regions resulting in the dissimilar significant variables that were observed in the 'Results' section. The following sections will discuss the significant variables in each region to identify the potential drivers of firm performance, which affect the subsequent investor return.

8.2 Drivers of Abnormal Return

The 'Theoretical Overview' highlighted that both a negative and a positive relationship between ESG factors and abnormal returns can be theoretically supported. Shareholder theory is useful to explain a negative correlation between ESG factors and investor return, as the theory argues that ESG initiatives are likely to decrease firm profits and create a diverted business focus (Renneboog, Horst, & Zhang, 2008a). Contrarily, stakeholder theory holds great explanatory power about a positive correlation between ESG factors and investor return. Stakeholder theory argues that ESG initiatives can improve the relationship with key stakeholders, which can drive firm value and subsequently impact investor return. The following two sections will assess the significant variables in the United States and Europe with the aim of establishing how the specific variable help drive or diminish investor return.

Significant variables in the United States

The cross-sectional analysis illustrated that five different ESG variables are correlated with firm alpha on a 5% significance level across the three time periods (Table 3). In the period from 2003-2007, 'Product Innovation' and 'Diversity & Opportunity' were both positively correlated with alpha, which implies that both of these ESG factors are value-drivers for American companies. The 'Product Innovation' factor measures a firm's commitment and effectiveness in developing eco-efficient

products (Appendix 6). Developing eco-efficient products is likely associated with high production input costs, which has to be offset by either increased revenue or cost reductions to be a value-driver for companies. Focusing on the revenue side, Berman et al. (1999) argue that ESG factors, such as 'Product Innovation', can be used to differentiate products and services of companies to enhance value creation. The positive correlation between 'Product Innovation' and alpha in the period from 2003-2007 suggests that many consumers were willing to pay a premium for products that were environmentally friendly, and this ESG factor thus improved the relationship with the consumer stakeholder group. The 'Diversity & Opportunity' factor reflects a firm's capacity to increase its workforce's loyalty and productivity by promoting good work-life balance, a friendly environment and equal career advancement opportunities (Appendix 6). This ESG factor is thus related to the employees of a company, and a high score is likely to be correlated with high employee satisfaction. Turban and Greening (1997) state that signals of a good working environment will help attract talented workers, and these workers are further likely to stay for longer. The positive correlation between 'Diversity & Opportunity' and firm alpha can thus be caused by increased productivity due to the attraction and retention of a talented workforce.

In the period from 2008-2012, 'Employment Quality' and 'Training & Development' were positively correlated with firm alpha. Both of these variables are also related to the employees of the firm. 'Employment Quality' measures a firm's commitment to and effectiveness towards providing high-quality employment benefits and job conditions (Appendix 6). This factor is thus highly related to the motivation of employees, and the positive correlation with alpha indicates that employee benefits may drive firm performance. The 'Training & Development' factor measures a firm's ability to provide training and education for its workforce (ibid.). In comparison to the previous factor, this is more related to the abilities of the employees and the subsequent effect on productivity. Common for these two factors is that both concern employee relations, but they may drive firm value in different ways.

'Health & Safety' was the only variable that was correlated with alpha in the period from 2013-2017. This variable was negatively correlated with alpha at a 5% significance level, in contrast to the variables in the previous periods that were all positively correlated with alpha. This ESG factor measures a firm's ability to create a healthy and safe workplace and is thus directed towards the employees of the firm (Appendix 6). This factor is arguably less influential on employee productivity

and motivation than the previously discussed ESG variables, which could explain the negative correlation. The negative correlation suggests that initiatives that improve 'Health & Safety' scores decrease firm value, and firm managers that invest in 'Health & Safety' initiatives thus spend money on a social interest that compromises the interests of the shareholders (Friedman, 1970). Accordingly, the shareholder would benefit if resources were repositioned to other initiatives that increase firm value. A contrasting view is that 'Health & Safety' initiatives are proactive in nature and thus help avoid costs of negative reactions from potential employee accidents and subsequent corporate scandals. Berman et al. (1999) argue that certain ESG variables can help mitigate risks in times of crisis as a form of insurance instead of drive value, which is arguably the case with this ESG variable. 'Health & Safety' initiatives can thus help firms to avoid the negative impacts that a potential employee accident would cause.

The significant variables in the three periods in the United States are positively correlated with alpha with the exception of 'Health & Safety' in the last examined period. Investing in these ESG factors can thus potentially drive value for companies. Agle and Wood (1997) presented a stakeholder identification model, which helps determine the stakeholder groups that have a significant impact on firm performance and thus requires considerable attention from company management. Four of the five significant variables in the United States belong to the social ESG category by relating to the employee relationship. This suggests that American firms may benefit from paying high attention to employee relations, as it can be a significant value driver. Employees often possess both power, legitimacy and urgency, which is required to be a stakeholder with a significant impact (ibid.). The following section will discuss the significant variables in Europe similarly to access the potential value drivers in this region.

Significant variables in Europe

The cross-sectional analysis of European companies resulted in six ESG variables that were significantly correlated with alpha at a 5% significance level across the three time periods. In the period from 2003-2007, 'Board Functions' and 'Employment Quality' were both positively correlated with firm alpha, while 'Board Structure' was negatively correlated with alpha. The 'Board Functions' factor measures a firm's effectiveness towards following best practice corporate governance principles related to board functions and activities (Appendix 6). Having the appropriate board functions and activities help ensure that the board performs the main tasks of monitoring and

servicing the firm, which can help create value by diminishing agency problems. Both shareholder and stakeholder theory emphasize that diminishing agency problems can create value for firms by ensuring that shareholder interests are considered (Freeman & McVea, 2001). 'Employment Quality' was the second factor that was positively correlated with alpha, and it is further the only factor that is significant in both Europe and the United States, albeit in different time periods. As previously stated, the 'Employment Quality' factor measures a firm's ability to provide appropriate benefits and job conditions, which is related to employee motivation (Appendix 6). Employee motivation is an influential factor in employee productivity, which can drive value for companies.

'Board Structure' measures a firm's effectiveness in creating a diverse and independent board (ibid.). The corporate governance literature generally highlights the presence of independent directors as a necessity for appropriate monitoring. A board of dependent directors, who have a connection to the company, would likely result in an entrenched board, which would not monitor management appropriately (Johnson, Daily, & Ellstrand, 1996). However, the monitoring role is not the only role of a board, as servicing and providing access to important resources have also been identified as important value-enhancing roles of a company board (ibid.). Duchin et al. (2010) state that independent directors may not possess the same level of information about a company compared to dependent directors, which can impact the servicing role of the board. The negative relation between 'Board Structure' and alpha can thus be explained by the disadvantages of having a diverse and independent board. High board diversity and independence will likely result in increased monitoring, which reduces agency problems, but it can have a negative impact on the other board roles that influence firm value.

In the following period from 2008-2012, there were three ESG variables correlated with alpha. Interestingly, 'Board Structure' is positively correlated with alpha in this time period as opposed to the previous period. The period from 2008-2012 consists of the years following the Great Financial Crisis, where firms may benefit from other governance mechanisms than before the financial crisis. Scoring high on 'Board Structure' means that a firm has a diverse and highly independent board, which can increase board monitoring. The positive correlation may thus be attributed to an increased need for board monitoring during unstable times (Switzer, Tu, & Wang, 2018). The 'Compensation Policy' factor was also positively related to alpha in this time period. 'Compensation Policy' measures a firm's ability to establish appropriate compensation packages for company management

that are related to company targets (Appendix 6). As highlighted in the 'Theoretical Overview' section, appropriate compensation structures are a way to align the incentives of shareholders and management (Nyberg, Fulmer, Gerhart, & Carpenter, 2010). Agency problems are thus likely to diminish as a consequence of appropriate compensation structures, which will create value for firms. This can explain the positive correlation between this ESG factor and firm alpha. The last significant variable is 'Product Responsibility', which was negatively correlated with alpha in this time period. 'Product Responsibility' measures a firm's effectiveness towards creating value-added products that uphold the customer's security. The negative correlation suggests that initiatives that improve 'Product Responsibility' may diminish profits and investor return in accordance with shareholder theory. Shareholder theory argues that companies should not invest in ESG initiatives if it decreases firm value, which the negative correlation between 'Product Responsibility' and abnormal return suggests. However, similar to the 'Health & Safety' factor, this ESG factor is likely proactive in nature. Not investing in 'Product Responsibility' may thus diminish profitability due to the negative effects of potential corporate scandals. Berman et al. (1999) support that certain ESG variables can help mitigate risks in times of crisis instead of drive value.

None of the ESG variables were significantly correlated with alpha in the last period from 2012-2017, and no ESG variable can thus be identified to drive firm value in this period. Across the three time periods, four out of six significant variables belong to the governance ESG category, which contrasts the findings in the United States. This suggests that governance variables are essential to consider for European firms in order to increase firm value. This trend is in accordance with both shareholder and stakeholder theory, as both theories emphasize the value of lessening agency problems. The above sections have illustrated the differences in significant variables and value-drivers in the United States and Europe, which is likely caused by different institutional pressures as examined previously. However, value-drivers are not the only potential explanation of a relationship between ESG variables and abnormal return, as share prices are affected by other factors than firm performance.

8.3 Mispricing

The above sections illustrate how ESG can potentially drive value for companies. However, it is not possible to earn abnormal returns by utilizing ESG information, if the market already takes all relevant ESG information into account (Derwall, Guenster, Bauer, & Koedijk, 2005). In this case, ESG scores will be reflected in stock prices, and its effect on firm performance in the following years

will be accounted for. The results section illustrated several relationships between ESG variables and abnormal returns at a 5% significance level, which indicates that stocks were mispriced to a certain degree. In these cases, investors did not account for the positive net benefit of specific ESG variables, and the firms that scored high on these variables outperformed expectations and provided an abnormal return.

The results section further identified an interesting pattern in significant variables, which supports the mispricing argument. The five significant ESG variables in the United States belong either to the 'Environment' or 'Social' category of ESG, where US firms scored substantially lower than their European counterparts. Similarly, four out of six significant variables in Europe belong to the 'Governance' category, where European firms scored significantly lower than US firms. Generally, high regional scores of ESG factors suggest that these factors receive a significant amount of attention in this region. Companies may be induced to increase their ESG efforts due to isomorphic pressures, and investors implement these factors in their investment strategies, accordingly. The ESG efforts may have a value-enhancing effect, but investors account for this, resulting in a relatively efficient market. Investors may thus be better able to earn an abnormal return by utilizing specific ESG information in regional areas, where this ESG factor receives comparatively less attention. This could explain the identified pattern in results.

Another trend in the cross-sectional regression is that the number of significant variables decreases in the last period from 2013-2017. 'Health & Safety' is the only significant ESG variable in the United States at a 5% significance level, while there are no significant variables in Europe. Consistent with the argument above, this could suggest that an increasing number of investors take the value-driving ESG factors into account, which makes markets more efficient and decreases the possibility of making a significant abnormal return. The increased focus on ESG investing, which was highlighted previously, could have a significant impact on this trend (SIF, 2018). In conclusion, the results of the cross-sectional regression indicate that ESG factors can drive company value in several ways, and this can result in abnormal returns due to mispricing. The following section will discuss the impermanence of significant variables over time.

8.4 The Impermanence in Significance of ESG Variables

The results of the cross-sectional regression analysis illustrate that most significant ESG variables do not continue to be significant across time periods. Furthermore, the few variables that remain significant over time change from either a positive to a negative coefficient or vice versa. In the United States, 'Product Innovation' is positively associated with alpha in the period from 2003-2007 at a 5% significance level, while it becomes negatively correlated with alpha in the subsequent period from 2008-2012 at a 10% significance level. In Europe, 'Board Structure' is negatively correlated with alpha from 2003-2007, while it is positively correlated with alpha in the following period. This illustrates that significant variables fluctuate considerably over time, which influences the applicability and interpretation of the results. The impermanence of significant variables over time is further supported by the sensitivity analysis of the regression analysis, where the time-period was shortened from a five-year to a three-year period. This analysis also showed impermanence in variables, as no single ESG variable was significantly correlated with alpha for two consecutive periods.

Bebchuk et al. (2013) find that the disappearance of a correlation between abnormal returns and ESG variables can be caused by a "learning effect" in the market. The paper replicates the methodology of an influential paper by Gompers et al. (2003), which finds a positive correlation between corporate governance provisions and abnormal return. However, the article by Bebchuk et al. (2013) examines the subsequent period and finds that the positive correlation disappeared, which is similar to the results in this thesis. The argumentation is the same as the mispricing argument in the above section. The paper argues that market participants started to appreciate the difference between companies that score well and poorly on governance indices and began to take this into account in their investment strategies. Consequently, the market started to react less significant to positive earnings announcements of firms with good corporate governance practices, which resulted in less positive abnormal stock returns. The article argues that the market change was caused by an increased focus on corporate governance in both the literature and the public debate following a series of corporate scandals in the early 2000s (ibid.). The findings by Bebchuk et al. (2013) is consistent with The Law of One Price, which states that the price of equivalent investment opportunities should be the same (Berk & DeMarzo, 2017). The companies that scored high on corporate governance indicators performed better than companies that scored low on corporate governance. This resulted in abnormal returns for investors, as the shares of good and poor governance performers were priced similarly.

Accordingly, investors could earn abnormal profits by including governance factors in their investment analysis. However, as the share price of superior corporate governance companies started to increase, an increasing number of investors started to include governance indicators in their analysis. Investors could thus no longer earn abnormal profits because firms with good governance performance started being priced fairly. Recall Figure 1 below, where the effect is illustrated by the security market line (SML) that represents the fair value of stocks according to the CAPM.

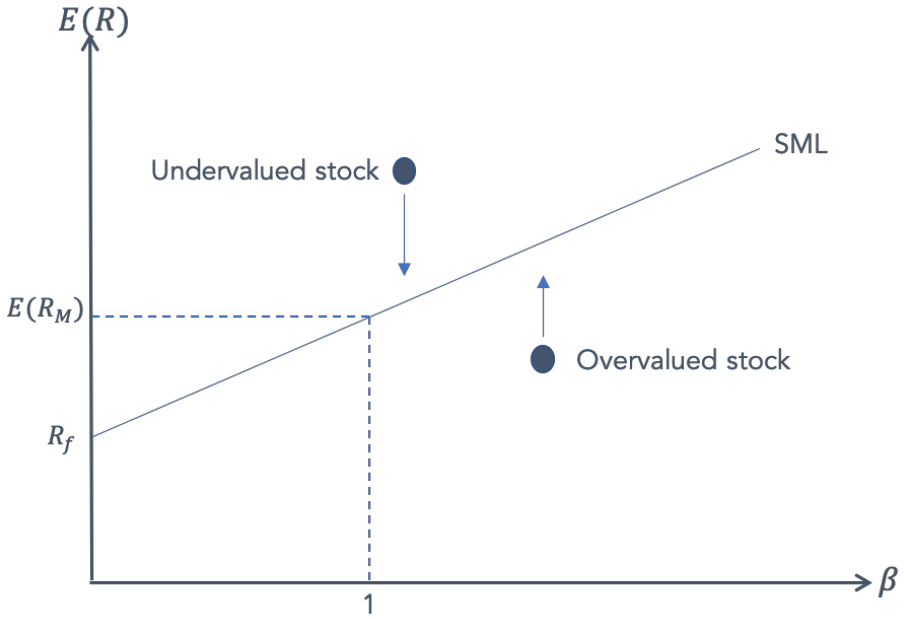


Figure 1; Source: (Sharpe, 1964)

The variables that are significantly correlated with alpha in this thesis could be utilized by investors to earn abnormal returns. However, as soon as the impact of ESG variables is recognized by the market, the opportunity for abnormal returns begins diminishing. This mechanism can potentially explain the impermanence in the significance of ESG variables in this thesis, which makes it difficult to earn continuous abnormal returns from ESG information. The impermanence of the results suggests that investors will not know ex-ante which variables that will be significantly correlated with abnormal returns in the following period. Investors cannot determine a permanent relationship between the ESG variables and abnormal returns due to the temporal variability in abnormal returns. Figure 1 can further exemplify why several variables in this thesis change coefficient from a positive to a negative with alpha or vice versa. When investors realize that a firm scoring highly on a specific ESG variable is undervalued, they start purchasing the stock. This increases the stock price and moves it towards the SML line. However, this may eventually result in an overvalued stock, in case investors

overvalue the net benefit of the ESG variable, resulting in investors paying a premium for these stocks. The learning effect is one explanation of the observed impermanence of significant ESG variables over time. Another explanation is that ESG variables affect alpha differently depending on the examined time period. This thesis studies three different time periods, which differ in several aspects.

Firstly, the time period from 2008-2012 includes the years following the Great Financial Crisis, which may influence the correlation between specific ESG factors and abnormal returns. The regression models in this thesis do account for market movements by including stock betas, which means that company alphas should not be affected by either market increases or downturns. However, the effect of ESG variables may depend on general market conditions. Switzer et al. (2018) find that an increased level of corporate governance will reduce risk significantly during financial crises and thus affect differently compared to more stable market environments. Accordingly, governance variables may have a significant positive impact on firm alpha during recessions. The results of this thesis support the argument to a certain degree. In Europe, the only ESG variables that had a significant positive correlation with alpha during 2008-2012 were 'Board Structure' and 'Compensation Policy'. However, the trend was different in the United States where only social ESG variables turned out to have a significant positive correlation with alpha in this time period.

Secondly, changes in the salience of different stakeholders can further impact the impermanence of significant ESG variables. The theory section highlights that institutional pressures may change over time, as the values and priorities of stakeholders may change (Dimaggio & Powell, 1983). For example, consumers could start to value certain environmental factors more highly than previously due to increased coverage of global warming in the media. Shifting isomorphic pressures will affect the value-drivers of companies, as the effect of ESG initiatives will change accordingly. This has an impact on both the impermanence in significance of variables, and it helps explain why dissimilar variables are significant in different time periods. The theory of stakeholder identification and salience argues that management should prioritize the stakeholder groups that possess both legitimacy, power and urgency, as these stakeholders will have the largest impact on firm performance. The theory further states that all of these three attributes are transitory, which supports the argument of this section (Mitchell, Agle, & Wood, 1997). The impermanence of significant ESG

variables over time can thus be explained by shifting stakeholder values, which affect the value of specific ESG initiatives.

8.5 An Optimal Level of ESG

The previous sections have aimed at explaining the results of the cross-sectional regression. However, a basic assumption in the regression analysis is a linear relationship between ESG variables and abnormal return. The aim has thus been to examine which significant linear relations that exist and further how these relations can be exploited by investors. However, a fully linear relationship implies that investors would yield the highest return by investing in firms that score either highest or lowest on specific ESG factors. McWilliams and Siegel (2001) argue that there are diminishing marginal returns of ESG initiatives, which suggests that an optimal level exists. When firms introduce ESG initiatives they can improve the relationship with certain stakeholder groups by establishing increased legitimacy. However, continuing to invest in specific ESG initiatives may not necessarily result in increased legitimacy and improved firm performance. In fact, if an ESG initiative does not improve the relationship with a stakeholder group, then the net benefit of the initiative might be negative, which supports the diminishing returns argument by McWilliams and Siegel (2001). Supposing there are diminishing returns of ESG initiatives, the optimal score for each ESG variable must be somewhere between 0 and 100, which is illustrated in Figure 7 below:

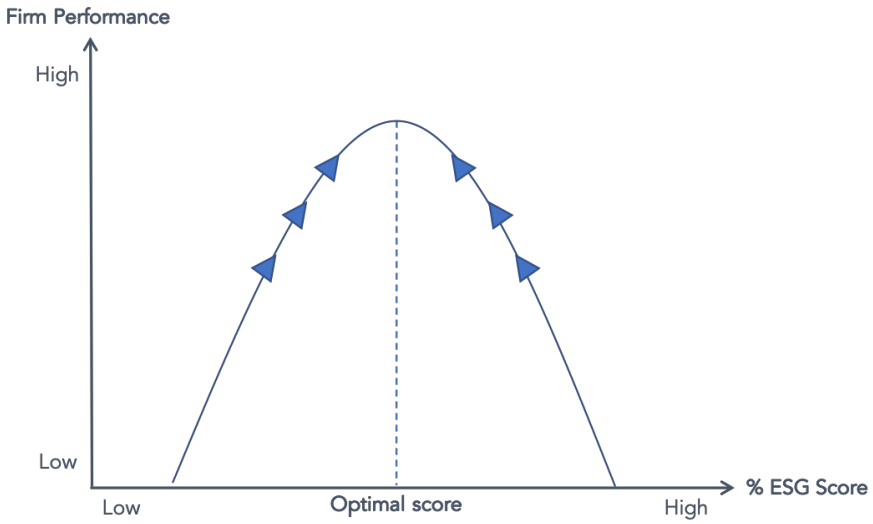


Figure 7; Source: own creation

Figure 7 can provide valuable insights about the interpretation of our results. This thesis hypothesizes a significant positive relationship between one or more ESG variables and abnormal return, which was identified in several time periods in the analysis. According to the model, the relationship between ESG variables and investor return is positive, if firms are located to the left of the optimal point. In this case, increasing the level of ESG initiatives will likely improve firm performance and result in increased investor returns. In contrast, a negative relation between ESG variables and investor return is likely to be found, if most companies score relatively high on the ESG variable and are located to the right of the optimal point. Lastly, if firms are equally distributed on either side of the optimal point, then the result could be an insignificant variable. This does not imply that a given ESG variable does not have a significant relationship with investor return, but that the relationship is not possible to determine using an ordinary least squares (OLS) regression because of its curvilinearity. In conclusion, the linearity assumption of our research design may not necessarily hold, as a firm's ESG score is not necessarily linearly correlated with firm performance, and this can highly influence the result of our analysis.

8.6 Investment Strategy

All of the above discussion sections have concerned the results of the cross-sectional regressions, whereas this section will discuss the results of the subsequent time-series regressions. The time-series regression analyses intended to utilize the significant relationships between ESG variables and firm alphas. The aim was thus to use significant relationships to build portfolios that resulted in abnormal returns for investors. However, only two of the portfolios provided a significant alpha, even though eleven variables were correlated with alpha at the 5% significance level across time periods and regions. This thesis hypothesized that positive alpha portfolios could be formed using significant ESG variables as screening criteria, and the hypothesis is not rejected due to the two significant portfolios. However, the low success rate indicates that investors will have difficulties in achieving a consistent alpha by utilizing our proposed investment strategy. The following section will examine the results of the time-series regression in further detail.

Significant and insignificant portfolios

The two portfolios that obtained a significant alpha were both European governance portfolios in the period from 2008-2012. The first portfolio was based on the top 30% companies that scored highest

on ‘Board Structure’, while the second portfolio was based on the top 30% companies that scored highest on ‘Compensation Policy’. Both of these variables are governance factors, which can help diminish agency problems and thus generate value for companies. However, the portfolios only provided significant alphas in a single period, which suggests that improved governance created significant value in the period from 2008-2012 but not in other periods. As previously described, Switzer et al. (2018) find that an increased level of corporate governance will reduce risk significantly during financial crises and thus affect differently compared to more stable market periods. This could explain why these portfolios are the only ones that yield a significant alpha. Knowing that specific governance variables can result in abnormal returns during financial crises is important to investors if the results are consistent in future periods.

The result section highlighted that the vast majority of portfolios did not provide a significant alpha. Moreover, the two significant portfolios described above may be significant due to a high risk of false positives, as a large number of regressions were conducted (Goldman, 2008). This indicates that ESG information may be difficult to utilize by investors in order to obtain a consistent alpha. The efficient market hypothesis argues that it is impossible for investors to consistently outperform the market since all relevant information is already incorporated into the share price (Berk & DeMarzo, 2017). Company shares will thus trade at the fairest price, which means that they cannot be purchased undervalued or sold overvalued. Efficient market theory is presented in three different forms, as outlined in the ‘Theoretical Overview’. The semi-strong form of market efficiency theory argues that share prices fully reflect all publicly available market data, which means that arbitrage opportunities will not arise (ibid.). The ESG data used in this study is publicly available and all investors can thus utilize the information. This can explain the failure to create portfolios that consistently obtain a positive alpha.

Optimizing the investment strategy

Grossman and Stiglitz (1980) argue that to believe that investors have the same information available and make full use of this information is incorrect. The article states that if markets were perfectly efficient, there would be no profit from gathering information or making trades and markets would eventually collapse. Investors could thus potentially earn abnormal returns by utilizing ESG information if an appropriate investment strategy is chosen, which would contradict the argument of efficient market theory in the above section. This study created portfolios of the top and bottom 30%

scoring companies within each significant ESG variable. This simplistic portfolio formation approach is one method of how ESG information can potentially be utilized, but it is not necessarily best practice. The primary aim of this thesis is to examine the relationship between ESG and abnormal returns and not to establish the most appropriate investment strategy to "beat the market". Accordingly, other strategies could potentially provide a higher proportion of portfolios with significant alphas and establish a more definite relationship between ESG and abnormal returns.

Barnett and Salomon (2006) argue that a curvilinear relation exists when building ESG portfolios. Investors that screen on ESG criteria will benefit from an improved selection of investment targets, as ESG can have a positive impact on firm performance. However, applying screening criteria by only picking the companies with the best ESG scores limits the investment universe, which affects negatively according to the mean-variance analysis by Markowitz (1952). These two mechanisms push in opposite directions and the effect on investor return will depend on the dominating mechanism. Barnett and Salomon (2006) find that investor return is affected positively by either a very high or very low degree of ESG screening in the investment strategy. Investors that are "stuck in the middle" may bear all the costs of either strategy without gaining the benefits. This raises the question of when an investor is considered to be "stuck in the middle".

This thesis applied an investment strategy where the top and bottom 30% of companies on significant ESG variables were included in the portfolios. However, only two portfolios provided a significant alpha in the following five-year period. In the sensitivity analysis, we applied a stricter screening strategy by only including firms in the top and bottom 10%. An investment strategy that applies stricter ESG screening could result in improved results, according to the article by Barnett and Salomon (2006). However, this change in strategy did not increase the number of significant alphas. In fact, only the portfolio that included the top 10% on 'Board Structure' provided an alpha that was significant at the 5% level. A possible explanation is that the ESG screening was still not appropriately high, and as a result, the portfolios were "stuck in the middle" with insignificant alphas.

Another explanation is that filtering on a single ESG factor may not be sufficient to achieve a consistent alpha, as other factors should be included in the analysis as well. Other ESG variables could influence the alpha of the portfolios, despite being insignificant in the cross-sectional regression. Accordingly, if companies in the portfolios score equivalently high on ESG factors with

a negative correlation with alpha, then the effect will cancel out. For example, a portfolio could consist of the top 30% of firms scoring high on a variable that is significantly positively correlated with alpha, such as 'Product Innovation'. These firms could potentially also score high on a variable that is significantly negatively correlated with alpha, such as 'Health & Safety. In this case, the effect of the positive variable could be canceled out. This suggests that additional factors could be included in the portfolio formation process in order to optimize portfolios and get enhanced results. Algorithms could be applied to utilize patterns in ESG variables associated with a significant alpha. For example, the asset management firm Arabesque uses over 200 ESG metrics to screen firms (The Economist, 2017). However, such extensive an analysis was out of scope for this thesis.

Additional information could also be used in combination with ESG information. The cross-sectional regression analyses in this thesis did not capture all variation in alpha, as the adjusted R^2 ranges from 9-35%, and additional factors could thus have been useful in the portfolio formation to obtain abnormal returns. Including financial information in investment strategies may result in better-performing portfolios and is standard for investment managers. This can include fundamental analysis that attempts to measure a security's intrinsic value by examining related financial and economic factors, which can be of both qualitative and quantitative nature. Jegadeesh and Titman (1993) find that including information about past stock performance may result in significant abnormal returns. The article finds that portfolios based on stocks that have performed well the last 6 months against their peers realized a significant excess return. Including information about past stock performance could thus be another example of additional factors to consider in the portfolio formation process. This thesis did not find strong evidence that the significant ESG variables could be utilized by investors to earn abnormal returns, but the results are highly affected by the choice of investment strategy. Choosing alternative strategies that included additional factors in the portfolio formation process could alter the findings.

8.7 Causal Inference and Endogeneity

In the previous discussion sections, the results of the regression analyses have been discussed and interpreted by applying different theoretical perspectives. The results illustrated a significant relationship between several ESG variables and abnormal return in both Europe and the US in different time periods. However, statistical significance does not imply causality. The following section will discuss the validity of causal inference between ESG variables and abnormal return.

Causality can be inferred if there is a clear and unambiguous connection between cause and effect. Causality between ESG variables and abnormal return can thus be inferred in case it is clear that the variation in abnormal return is caused by ESG variables and not a confounding variable. Hill (1965) presents a number of aspects that are useful to consider when determining the causality of a study. The most relevant aspects, in the context of this study, will be discussed in the sections below.

Firstly, Hill (1965) argues that it is useful to assess the plausibility of causation. He argues that causation is more likely if there is a theoretically bound logic between cause and effect. The 'Theoretical Overview' highlighted several mechanisms that could explain how ESG variables could impact abnormal return. In essence, ESG initiatives affect the relationship with the stakeholders of a company. The performance of a company is highly dependent on the relationship with key stakeholder groups, as an improved relationship can drive value in several ways (Renneboog, Horst, & Zhang, 2008a). This suggests a high level of plausibility between ESG variables and firm performance. Abnormal returns are directly linked to firm performance, which supports that causation is highly plausible. However, stock prices are affected by many factors, and the relationship between ESG variables and abnormal return could thus be caused by confounding factors despite high plausibility.

Secondly, the consistency of results is considered a significant indicator of causality by Hill (1965). The paper argues that the causality of a study is more likely if an observed association has been repeatedly observed by different studies, in different places, at different times. The 'Empirical Findings' section illustrated that studies often have contradictive findings regarding the relationship between ESG and abnormal return. Findings within this research area are thus highly inconsistent. The inconsistency of results is further illustrated by the results of this thesis, which shows a high level of impermanence in significant variables in both Europe and the US across time periods. The efficient market theory can somewhat explain the impermanence of results. In efficient markets, the opportunities of earning an abnormal return will be exploited by investors and quickly disappear (Berk & DeMarzo, 2017). Accordingly, the relationship between a specific ESG variable and abnormal return may be causal but merely under certain circumstances and for a brief period of time.

Thirdly, specificity is an aspect of causality formulated by Hill (1965). The probability of a causal link is higher if the examined sample and setting are highly specific. This thesis studies the

relationship between ESG variables and abnormal returns across many types of firms, which operate in different industries and in many different geographical areas. This suggests that the specificity of the sample in this study is relatively low. European and American companies are analyzed in independent regression analyses, which increases the specificity of each of these regressions. However, the European sample consists of companies from five different countries, which implies low specificity. Hill (1965) states that there is a trade-off between specificity and generalizability of results. The low specificity of this study decreases the probability of a causal link, but the findings are more likely to be generalizable.

Lastly, Hill (1965) emphasizes the importance of endogeneity. As previously described, studies examining ESG and financial performance often suffer from the problem of endogeneity, which makes the results difficult to interpret. Specifically, it is not possible to determine whether ESG variables have an impact on abnormal returns or vice versa. This study does not mitigate the prevalence of endogeneity in the data set, but it was tested whether endogeneity was a problem. The test illustrated that the null hypothesis of no endogeneity between significant ESG variables and prior alpha could not be rejected (Appendix 15). No alpha values were significantly associated with either of the significant ESG variables at a 5% significance level. However, the test does not provide conclusive evidence about the endogeneity and it might be a factor that affects the results of this study. The aspects presented by Hill (1965) illustrate that problems of causality may be an issue in this thesis for several reasons. This affects the interpretation and implications of the results, which will be assessed further in the following sections.

8.8 Summary of Discussion

The discussion illustrates that the results of each regression analysis can be caused by several factors. Institutional differences between the United States and Europe are likely to explain the clear difference in ESG scores between American and European companies. These differences might influence the value-drivers of ESG variables in each region, as social and environmental variables are more significant in the United States and governance variables are more significant in Europe. The mispricing argument can possibly explain why all value-increasing ESG variables are not significant, as it is not possible for investors to earn an abnormal return if the value of ESG variables are appropriately incorporated in share prices. The arbitrage theory and the learning effect can further explain the impermanence of significant variables over time. The high level of insignificance in the

time-series regression can potentially be explained by the market efficiency theory. However, the choice of investment strategy could also imply that the ESG information was not utilized optimally to make appropriate portfolios. The benefits of including firms with high ESG scores of significant variables may further be offset by the negative effects of limiting the investable universe of stocks. Lastly, the results of the thesis may be influenced by the problems of causality and endogeneity, which impact the interpretation of the results.

9. Conclusion

Firms are increasingly pressured to act responsibly by customers, governments and investors. This raises the question of whether responsible business initiatives are associated with abnormal investor returns. The aim of this thesis was to examine the relationship between ESG variables and abnormal returns in Europe and the US, and further whether investors could have exploited this relationship to construct portfolios with abnormal returns.

The results of the analysis showed a positive association between several ESG variables and alpha in different time periods from 2002-2017. Specifically, 11 ESG variables were significantly associated with alpha, of which eight were positively associated. The analysis of the coefficient sizes for significant variables showed that the magnitudes of the coefficients were large enough to hold economic significance for investors. Moreover, several interesting patterns emerged from analyzing the ESG data and regression results. On average, European firms scored much higher than US firms on environmental and social variables. By contrast, US firms generally scored higher on governance variables. These structural differences between US and European firms had implications for the regression results. Nine of the 11 significant ESG variables were in a region that on average scored comparatively worse on the given ESG variable. A plausible reason is that the effect of ESG performance on alpha is more pronounced when firms are outperforming relative to the mean ESG score in the region. Another insight derived from the regression analysis was that no ESG variables were significant in the US and Europe in the same time period. Moreover, no ESG variable remained significant over consecutive time periods. This indicates that significant associations between ESG variables and abnormal returns are relatively impermanent and highly dependent on the institutional environment.

In addition to analyzing ESG at the firm-level, portfolios were also formed with the aim of exploiting ESG information to obtain positive, abnormal returns. Based on the ESG variables significantly associated with alpha in the firm-level regressions, portfolios were formed of the best and worst scoring firms on the given ESG variables. Two out of 22 portfolios turned out to have significant, positive alphas. This illustrates that investors could have exploited the significant relationships between ESG variables and abnormal to construct portfolios with abnormal returns. However, the success rate is too low for the trading strategy to be practically implementable. Our study thus

provides empirical evidence that ESG variables are often associated with positive abnormal returns, but the information is difficult to utilize by investors.

9.1 Implications of Results

There are three main groups who benefit from the insights uncovered in this thesis. The benefits for each group will be discussed in turn. Academic researchers benefit from several insights. The significant variables add to the existing literature and support the current view that the relationship between ESG and abnormal return is highly impermanent and dependent on the setting. Furthermore, our thesis has two methodological implications that are valuable for academic researchers to consider. Breaking down ESG scores can yield valuable insights that would otherwise be missed. High-level ESG scores obscure countervailing effects on an alpha of different underlying ESG factors. The link between ESG and alpha becomes clearer when examining the underlying ESG factors separately. The second methodological insight is that examining geographical differences regarding the relationship between ESG and abnormal returns can nuance the ESG research. There are large differences in ESG performance, and association with alpha, between the US and Europe, which has implications for future research within this research area.

This thesis applied an investor perspective and the results thus have significant implications for investors. The cross-sectional regression analysis illustrated 11 different ESG variables that had a significant association with abnormal returns across Europe and the US. This knowledge can support investors by potentially improving their investment strategies. Like academic researches, investors also benefit from the insight that decomposing ESG scores establishes a clearer link between each ESG component and alpha. In other words, breaking down ESG scores can filter out underlying ESG variables that have no association with alpha. A deeper understanding of which underlying ESG variables are important drivers of alpha, and why that is the case, could lead to more sophisticated trading strategies that utilize the link between ESG performance and alpha. The identified pattern of significant ESG variables is further valuable to investors. The results of this thesis indicate that the effect of ESG is more pronounced when firms outperform relative to the firms in their region. This is also useful information for investors to optimize their investment strategies.

Lastly, firms can benefit from this thesis' reflections on the salience of different stakeholders in relation to alpha. It is highly circumstantial for each firm which stakeholders to prioritize. However,

by understanding how relationships to different stakeholder groups affect abnormal returns, firms can obtain information that is useful in prioritizing stakeholders. It is important to emphasize that the findings in this thesis is subject to potential problems of causation and endogeneity. The 'Discussion' section highlighted that the exact cause and effect cannot be decisively concluded, which limits the usability of the findings for both academic researchers, investors and firms.

9.2 Suggestions for Further Research

The findings of this thesis have both yielded interesting insights and posed new questions worthy of further investigation. The insights of this study need to be corroborated by other studies to ensure that the findings are robust and correct. Additionally, new studies can extend our analysis in ways that would further enlighten the relationship between ESG and alpha and how this relationship can be exploited. Our research points towards three fruitful avenues of future research. The first is a deeper understanding of how ESG factors influence investor returns. In this thesis, we tested which ESG variables that were significantly associated with alpha and then sought to explain the significant ESG variables' link with alpha theoretically. Different theories were applied to explain different ESG variables, but a more holistic theoretical framework could be useful to explain the link between ESG and alpha. This could potentially establish a deeper understanding of the mechanisms through which ESG drives value for shareholders and stakeholders.

The second suggested avenue of future research is exploring how the link between ESG and alpha depends on the circumstances. This thesis explored how ESG factors' influence on alpha depended on region and time period. However, there are numerous other comparative studies that could be insightful. This thesis controlled for the impact of industry on the alpha when estimating the association between ESG variables and alpha. However, a more industry-specific study could make the findings more practically applicable to firms in a given industry. Additionally, we only examined developed countries, but in emerging markets, the relationship between ESG and alpha might be different due to different institutional environments. Another interesting comparative study could be between public and private firms. The increased scrutiny from being a public firm could influence how ESG affects abnormal returns. Another performance measure than alpha would be needed as private firms do not have an observable stock price. More broadly, research on how a variety of different ownership structures affect the link between ESG and alpha could also be useful. There is

already a wealth of research on how ownership structures affect governance in particular. However, it could be interesting to explore if the ESG performance of firms depends on whether owners are more short-term oriented, such as hedge funds, and long-term oriented, such as pension funds.

The third promising avenue of research is to develop strategies to exploit ESG information better. Our thesis has indicated a relationship between ESG and the long-term abnormal return of firms. However, our portfolio formation method for exploiting this information was simplistic, as it was not the primary focus of this thesis. More sophisticated trading strategies might reveal patterns between ESG and abnormal returns. It is a likely impediment that the profitability of trading strategies often diminishes once they become public knowledge.

In conclusion, there is a wealth of literature on ESG and financial performance already, which has yielded numerous valuable insights. However, a deeper understanding of the contextual nature of the relationship between ESG and abnormal returns would make the findings more implementable and relevant for firms. This would help firms prioritize ESG investments to the benefit of its shareholders, other stakeholders and the world.

10. Bibliography

Barnett, M. L., & Salomon, R. M. (2006). Beyond Dichotomy: The Curvilinear Relationship between Social Responsibility and Financial Performance. *Strategic Management Journal, 27*(11), 1101-1122.

Baumol, W. J., & Blackman, S. A. (1991). *Perfect Markets and Easy Virtue: Business Ethics and the Invisible Hand.* Oxford: Wiley-Blackwell.

Bebchuk, L. A., Cohen, A., & Ferrell, A. (2009). What matters in Corporate Governance? *The Review of Financial Studies, 22(2),* 783-827.

Bebchuk, L., Cohen, A., & Wang, C. (2013). Learning and the Disappearing Association Between Governance and Returns. *Journal of Financial Economics, 108*(2), 323-348.

Berk, j., & DeMarzo, P. (2017). *Corporate Finance.* Harlow, England: Pearson Education.

Berman, S., Wicks, A., Kotha, S., & jones, T. (1999). Does Stakeholder Orientation Matter? The Relationship between Stakeholder Management Models and Firm Financial Performance. *The Academy of Management Journal, 42*(5), 488-506.

Bloomberg. (2017). *Bloomberg for Sustainable Finance Analysis.* Retrieved from Bloomberg for Sustainable Finance Analysis: https://data.bloomberglp.com/professional/sites/10/ESG-Brochure1.pdf

Bodie, Z., Kane, A., & Marcus, A. J. (2014). *Investments.* New York, NY, USA: McGraw-Hill Education.

Brown, L. D., & Caylor, M. L. (2006). Corporate Governance and Firm Valuation. *Journal of Accounting and Public Policy, 25*(4), 409-434.

Carhart, M. (1997). On Persistence in Mutual Fund Performance. *Journal of Finance, 52*(1), 57-82.

CFA. (2015, May 26). *Environmental, Social and Governance (ESG) Survey.* Retrieved from CFA Institute: https://www.cfainstitute.org/research/survey-reports/esg-survey-2015

Cheng, B., Ioannou, I., & Serafeim, G. (2014). Corporate Social Responsibility and Access to Finance. *Strategic Management Journal, 35*(1), 1-23.

Cohen, M., Fenn, S., & Naimon, J. (1997). Environmental and Financial Performance: Are They Related? *Working paper, Vanderbild University, 18*(6), 417-420.

Cook, R. D. (1977). Detection of Influential Observation in Linear Regression. *Technometrics, 10*(1), 15-18.

Cremers, M., & Nair, V. (2005). Governance mechanisms and equity prices. *Journal of Finance, 60*(6), 2859-2894.

Damodaran, A. (2002). *Investment Valuation* (Vol. 2nd edition). Hoboken, NJ, USA: Wiley.

Damodaran, A. (n.d.). *Estimating Beta.* Retrieved from Stern Business School: http://people.stern.nyu.edu/adamodar/podcasts/cfspr17/session7.pdf

Danko, D., Goldberg, J., Goldberg, S., & Grant, R. (2008). Corporate social responsibility: The United States vs. Europe. *Journal of Corporate Accounting & Finance, 19*(6), 41-47.

Davies, P. L. (2000, December). *Company Law Reform in OECD Countries: A Comparative Outlook of Current Trends.* Retrieved from OECD: https://www.oecd.org/daf/ca/corporategovernanceprinciples/1857291.pdf

Derwall, J., Guenster, N., Bauer, R., & koedijk, K. (2005). The Eco-Efficiency Premium Puzzle. *Financial Analysts Journal, 61*(2), 51-63.

Dharmastuti, C., & Wahyudi, S. (2013). The Effectivity of Internal and External Corporate Governance Mechanisms Towards Corporate Performance. *Research Journal of Finance and Accounting, 4*(4), 132-140.

Dimaggio, P., & Powell, W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review, 48*(2), 147-160.

Doh, J., & Guay, T. (2006). Corporate Social Resopnsibility, Public Policy, and NGO Activism in Europe and the United states: An Institutional-Stakeholder Perspective. *Journal of Management Studies, 43*(1), 47-73.

Duchin, R., Matsusaka, J., & Ozbas, O. (2010). When are outside directors effective? *Journal of Financial Economics, 96*(2), 195-214.

Duuren, E. v., Plantinga, A., & Scholtens, B. (2016). ESG Integration and the Investment Management Process: Fundamental Investing Reinvented. *Journal of Business Ethics, 138*(3), 525-533.

Eurosif. (2018, November 19). *European SRI Study 2018.* Retrieved from Eurosif: http://www.eurosif.org/eurosif-2018-sri-study-launch-event/

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance, 25*(2), 383-417.

Fama, E. F. (1991). Efficient Capital Markets: II. *The Journal of Finance*, *46*(5), 1575-1617.

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

Fama, E. F., & French, K. R. (1996). Multifactor Explanations of Asset Pricing Anomalies. *Journal of Finance, 51*(1), 55-84.

Fama, E. F., & French, K. R. (2015). A Five-Factor Asset Pricing Model. *Journal of Financial Economics, 116*(1), 1-22.

Fama, E. F., & Jensen, M. (1983). Separation of Ownership and Control. *Journal of Law and Economics, 26*(2), 301-325.

Freeman, E., & McVea, J. (2001). *A Stakeholder Approach to Strategic Management.* Oxford: Blackwell Publish.

French, K. (2019a). *Current Research Returns*. Retrieved from Kenneth French: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

French, K. (2019b). *Description of Fama/French 5 Factors (2x3)* . Retrieved from Kenneth French: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_5_factors_2x3.html

French, K. (2019c). *Description of Fama/French 5 Factors for Developed Markets*. Retrieved from Kenneth French: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_5developed.html

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

Friedman, M. (1970, September 13). The Social Responsibility of Business is to Increase Profits. *The New York Times Magazine*.

Goldman, M. (2008). *Berkeley.* Retrieved from Why is multiple testing a problem? Retrieved from: https://www.stat.berkeley.edu/~mgoldman/Section0402.pdf

Gompers, P., Ishii, J., & Metrick, A. (2003). Corporate Governance and Equity Prices. *Quarterly Journal of Economics, 118*(1), 107-155.

Green, T., & Peloza, J. (2011). How does corporate social responsibility create value for consumers? *Journal of Consumer Marketing, 28*(1), 48-56.

Grossman, S., & Stiglitz, J. (1980). On the Impossibility of Informationally Efficient Markets. *The American Economic Review, 70*(3), 393-408.

Halbritter, G., & Dorfleitner, G. (2015, April 2). The wages of social responsibility — where are they? A critical review of ESG investing. *Review of Financial Economics, 26*, 25-35.

Harrington, M. (2017, January 16). *Survey: People's Trust Has Declined in Business, Media, Government, and NGOs*. Retrieved from Harvard Business Review: https://hbr.org/2017/01/survey-peoples-trust-has-declined-in-business-media-government-and-ngos

Hermalin, B. E., & Weisbach, M. S. (2003). Boards of Directors as an Endogenously Determined Institution: A Survey of the Economic Literature. *Economic Policy Review, 9*(1), 7-20.

Hill, A. (1965). The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine, 58*, 295-300.

Hillman, A. J., & Keim, G. (2001). Shareholder Value, Stakeholder Management, and Social Issues: What's the Bottom Line? *Strategic Management Journal, 22*(2), 125-139.

Huber, B. M., Comstock, M., & Polk, D. (2017). *ESG Reports and Ratings: What They Are, Why They Matter*. Retrieved from Harvard Law School Forum on Corporate Governance and Financial Regulation: https://corpgov.law.harvard.edu/2017/07/27/esg-reports-and-ratings-what-they-are-why-they-matter/

Hofstede Insights, (2019, April 16). *Country Comparison*. Retrieved from Hofstede Insights: https://www.hofstede-insights.com/country-comparison/

Jegadeesh, N., & Titman, S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *Journal of Finance, 48*(1), 65-91.

Jensen, M. (2002). Value Maximization, Stakeholder Theory, and the Corporate Objective Function. *Business Ethics Quarterly, 12*(2), 235-256.

Jensen, M. C. (1967). The Performance Of Mutual Funds In The Period 1945-1964. *Journal of Finance, 23*(2), 389-416.

Johnson, J., Daily, C., & Ellstrand, A. (1996). Boards of directors: A review and research agenda. *Journal of Management, 22*(3), 409-438.

Kell, G. (2018, July 11). *The Remarkable Rise of ESG*. Retrieved from Forbes: https://www.forbes.com/sites/georgkell/2018/07/11/the-remarkable-rise-of-esg/#16c02df11695

KPMG. (2018). *ESG, risk, and return.* Retrieved from KPMG: https://assets.kpmg/content/dam/kpmg/be/pdf/2018/05/esg-risk-and-return.pdf

KPMG. (N.d.). *IFRS vs. US GAAP: R&D costs*. Retrieved from KPMG Advisory: https://advisory.kpmg.us/articles/2017/ifrs-vs-us-gaap-rd-costs.html

Luther, R. G., Matatko, J., & Corner, D. C. (1992). The Investment Performance of UK “Ethical” Unit Trusts. *Accounting, Auditing & Accountability Journal, 5*(4), 57-70.

Markman, A. (2018, October 11). *Why People Aren't Motivated to Address Climate Change*. Retrieved from Harvard Business Review: https://hbr.org/2018/10/why-people-arent-motivated-to-address-climate-change

Markowitz, H. (1952, March). Portfolio Selection. *The Journal of Finance, 7*(1), 77-91.

Mănescu, C. (2011). Stock returns in relation to environmental, social and governance performance: Mispricing or compensation for risk? *Sustainable Development, 19*(2), 95-118.

McWilliams, A., & Siegel, D. (2000). Corporate social responsibility and firm performance: correlation or misspecification? *Strategic Management Journal, 21*(5), 603-609.

McWilliams, A., & Siegel, D. (2001). Corporate Social Rersponsibility: A Theory of the Fim Perspective. *The Academy of Management Review,, 26*(1), 117-127.

Mitchell, R., Agle, B., & Wood, D. (1997). Toward a Theory of Stakeholder Identificayion and Salience: Defining the Principle of Who and What Really Counts. *Academy of Management Review, 22*(4), 853-886.

Moneyland. (n.d.). Retrieved from LIBOR definition: https://www.moneyland.ch/en/libor-definition

Mooney, C. (2018, June 1). *Trump withdrew from the Paris climate deal a year ago. Here's what has changed.* Retrieved from Washignton Post: https://www.washingtonpost.com/news/energy-environment/wp/2018/06/01/trump-withdrew-from-the-paris-climate-plan-a-year-ago-heres-what-has-changed/?noredirect=on&utm_term=.150dfafec8fb

Mourougan, S., & Sethuraman, D. K. (2017). Hypothesis Development and Testing. *Journal of Business and Management, 19*(5), 34-40.

Munk, C. (2017, September 2). *Financial Markets and Investments.*

Nofsinger, J., & Varma, A. (2014). Socially responsible funds and market crises. *Journal of Banking & Finance, 48*(C), 180-193.

Nollet, J., Filis, G., & Mitrokostas, E. (2016). Corporate social responsibility and financial performance: A non-linear and disaggregated approach. *Economic Modelling, 52*(B), 400-407.

Nyberg, A., Fulmer, I., Gerhart, B., & Carpenter, M. (2010). Agency Theory Revisited: CEO Return and Shareholder Interest Alignment. *Academy of Management Journal, 53*(5), 1029-1049.

Ortas, E., Álvarez, I., & Garayar, A. (2015). The Environmental, Social, Governance, and Financial Performance Effects on Companies that Adopt the United Nations Global Compact. *Sustainability, 7*(2), 1932-1956.

PennState. (2018). *Identifying Influential Data Points*. Retrieved from Stat 501: regression methods: https://newonlinecourses.science.psu.edu/stat501/node/340/

Prasad, S., Rao, A., & Rehani, E. (2001, September 18). *Developing hypothesis and research questions.* Retrieved from Arizona State University: https://www.public.asu.edu/~kroel/www500/hypothesis.pdf

Renneboog, L., Horst, J. T., & Zhang, C. (2008a). Socially responsible investments: Institutional aspects, performace, and investor behavior. *Journal of Banking & Finance, 32*, 1723-1742.

Renneboog, L., Horst, J. T., & Zhang, C. (2008b). The price of ethics and stakeholder governance: The performance of socially responsible mutual funds. *Journal of Corporate Finance, 14*(3), 302-322.

Reuters, T. (2012). *Thomson Reuters Business Classification.* Retrieved from Thomson Reuters: https://www.thomsonreuters.com/content/dam/openweb/documents/pdf/financial/trbc-quick-guide.pdf

Roll, R., & Ross, S. (1980). An Empirical Investigation of the Arbitrage Pricing Theory. *Journal of Finance, 35*(5), 1073-1103.

Ross, S. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory, 13*(3), 341-360.

Sassen, R., Hinze, A.-K., & Hardeck, I. (2016, April 23). Impact of ESG factors on firm risk in Europe. *Journal of Business Economics, 86*(8), 867-904.

Scott, W. (1995). *Institutions and Organizations. Ideas, Interests and Identities* (4th ed). London: Sage.

Shank, T., Manullang, D., & Hill, R. (2005). "Doing Well While Doing Good" Revisited: A Study of Socially Responsible Firms' Short-Term versus Long-term Performance. *Managerial Finance, 31*(8), 33-46.

Sharpe, W. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance, 19*(3), 425-442.

SIF. (2018, October 31). *US SIF Trends Report.* Retrieved from US SIF: https://www.ussif.org/trends

St. Louis Fed. (2019). *LIBOR rates*. Retrieved from Federal Reserve Bank of St. Louis: https://fred.stlouisfed.org/categories/33003?t=nation&rt=nation&ob=pv&od=desc

Stock, J. H., & Watson, M. W. (2011). *Introduction to Econometrics.* Boston: Pearson Education Inc.

Suchman, M. (1995). Managing Legitimacy: Strategic and Institutional Approaches. *Academy of Management Review, 20*(3), 571-610.

Switzer, L., Tu, Q., & Wang, J. (2018). Corporate governance and default risk in financial firms over the post-financial crisis period: International evidence. *Journal of International Financial Markets, institutions & Money, 52*, 196-210.

The Economist. (2017, November 25). *Sustainable investment joins the mainstream*. Retrieved from The Economist: https://www.economist.com/finance-and-economics/2017/11/25/sustainable-investment-joins-the-mainstream

Thomson Reuters. (2011). *Asset4 ESG Data.* Retrieved from Thompson Reuters: http://zeerovery.nl/blogfiles/esg-scores-methodology.pdf

Thomson Reuters. (2015, February). *Asset4 ESG Data Glossary*. Retrieved from Thomson Reuters: zeerovery.nl/blogfiles/DataStream_ESG_Glossary_Extranet.xlsx

Tirole, J. (2001). Corporate Governance. *Econometrica, 69*(1), 1-35.

Turban, D., & Greening, D. (1997). Corporate Social Performance and Organizational Attractiveness to Prospective Employees. *The Academy of Management Journal, 40*(3), 658-672.

United Nations. (2016, September 6). *Principles for Responsible Investment.* Retrieved from United Nations https://www.unpri.org/about-the-pri/annual-report-2016/711.article

Wooldridge, J. M. (2012). *Introductory Econometrics - A Modern Approach* (Vol. 5th edition). Mason, Ohia: South-Western.

11. Appendices

Appendix 1

The six Principles for Responsible Investment (PRI)

1 We will incorporate ESG issues into investment analysis and decision-making processes.

2 We will be active owners and incorporate ESG issues into our ownership policies and practices.

3 We will seek appropriate disclosure on ESG issues by the entities in which we invest.

4 We will promote acceptance and implementation of the Principles within the investment industry.

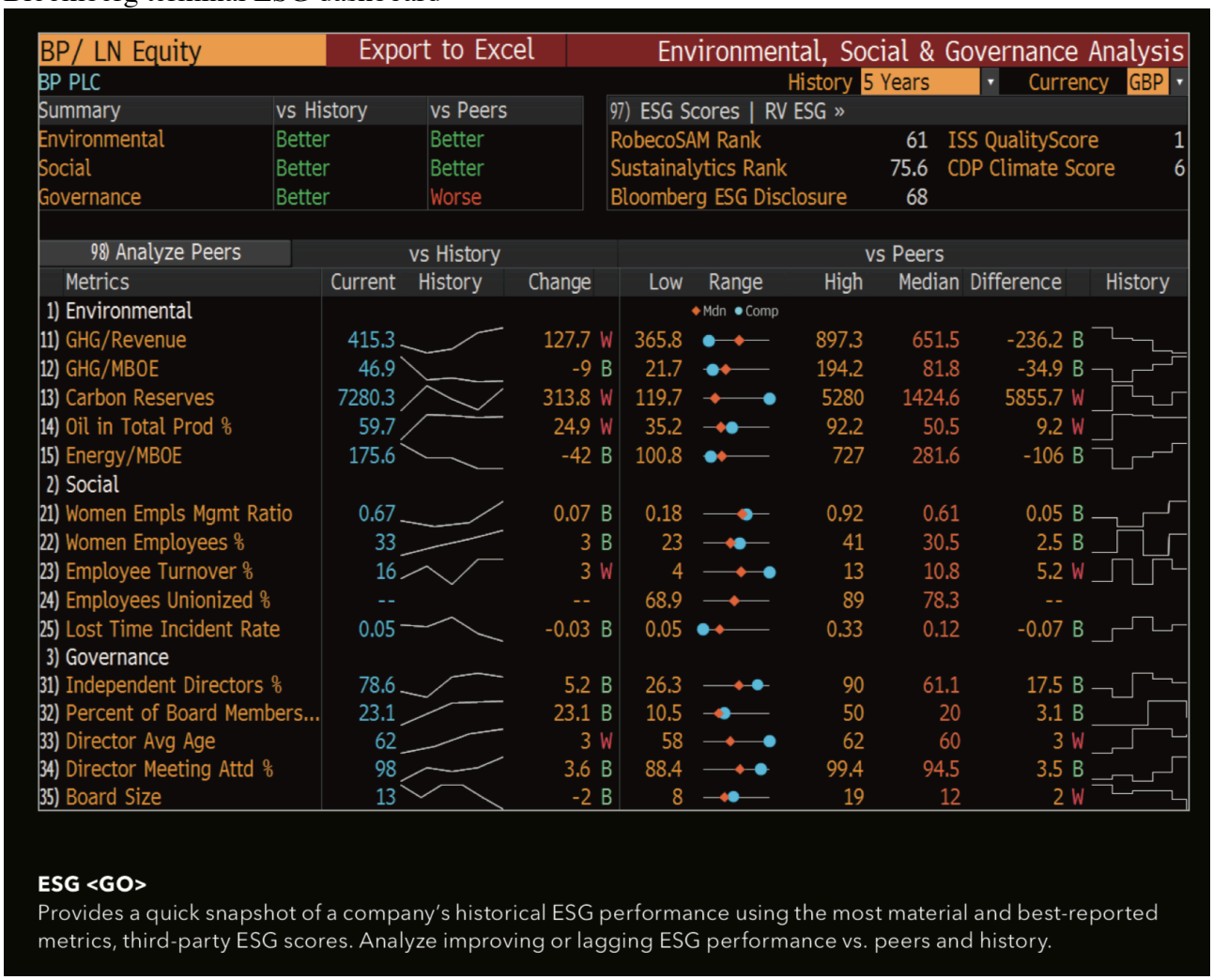
5 We will work together to enhance our effectiveness in implementing the Principles.

6 We will each report on our activities and progress towards implementing the Principles.

Source: (United Nations, 2016)

Appendix 2

Bloomberg terminal ESG dashboard



Source: (Bloomberg, 2017)

Appendix 3

THE CLASSICAL LINEAR MODEL ASSUMPTIONS

Now is a good time to review the full set of classical linear model (CLM) assumptions for cross-sectional regression. Following each assumption is a comment about its role in multiple regression analysis.

Assumption MLR.1 (Linear in Parameters)

The model in the population can be written as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u,$$

where $\beta_0, \beta_1, \ldots, \beta_k$ are the unknown parameters (constants) of interest and u is an unobserved random error or disturbance term.

Assumption MLR.1 describes the population relationship we hope to estimate, and explicitly sets out the β_j—the ceteris paribus population effects of the x_j on y—as the parameters of interest.

Assumption MLR.2 (Random Sampling)

We have a random sample of n observations, $\{(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i): i = 1, \ldots, n\}$, following the population model in Assumption MLR.1.

This random sampling assumption means that we have data that can be used to estimate the β_j, and that the data have been chosen to be representative of the population described in Assumption MLR.1.

Assumption MLR.3 (No Perfect Collinearity)

In the sample (and therefore in the population), none of the independent variables is constant, and there are no exact *linear* relationships among the independent variables.

Once we have a sample of data, we need to know that we can use the data to compute the OLS estimates, the $\hat{\beta}_j$. This is the role of Assumption MLR.3: if we have sample variation in each independent variable and no exact linear relationships among the independent variables, we can compute the $\hat{\beta}_j$.

Assumption MLR.4 (Zero Conditional Mean)

The error u has an expected value of zero given any values of the explanatory variables. In other words, $E(u|x_1, x_2, \ldots, x_k) = 0$.

As we discussed in the text, assuming that the unobserved factors are, on average, unrelated to the explanatory variables is key to deriving the first statistical property of each OLS estimator: its unbiasedness for the corresponding population parameter. Of course, all of the previous assumptions are used to show unbiasedness.

Assumption MLR.5 (Homoskedasticity)

The error u has the same variance given any values of the explanatory variables. In other words,

$$\text{Var}(u|x_1, x_2, \ldots, x_k) = \sigma^2.$$

Compared with Assumption MLR.4, the homoskedasticity assumption is of secondary importance; in particular, Assumption MLR.5 has no bearing on the unbiasedness of the $\hat{\beta}_j$. Still, homoskedasticity has two important implications: (1) We can derive formulas for the sampling variances whose components are easy to characterize; (2) We can conclude, under the Gauss-Markov assumptions MLR.1 to MLR.5, that the OLS estimators have smallest variance among *all* linear, unbiased estimators.

Assumption MLR.6 (Normality)

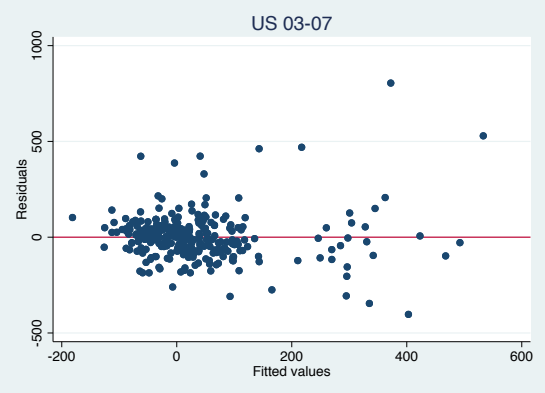
The population error u is *independent* of the explanatory variables $x_1, x_2, \ldots, x_k$ and is normally distributed with zero mean and variance σ^2: $u \sim \text{Normal}(0, \sigma^2)$.

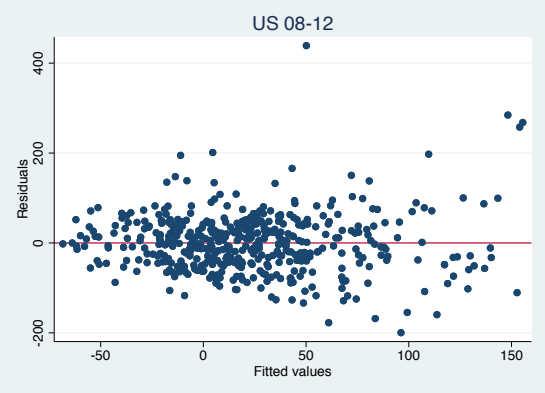
In this chapter, we added Assumption MLR.6 to obtain the exact sampling distributions of t statistics and F statistics, so that we can carry out exact hypotheses tests. In the next chapter, we will see that MLR.6 can be dropped if we have a reasonably large sample size. Assumption MLR.6 does imply a stronger efficiency property of OLS: the OLS estimators have smallest variance among *all* unbiased estimators; the comparison group is no longer restricted to estimators linear in the $\{y_i: i = 1, 2, \ldots, n\}$.

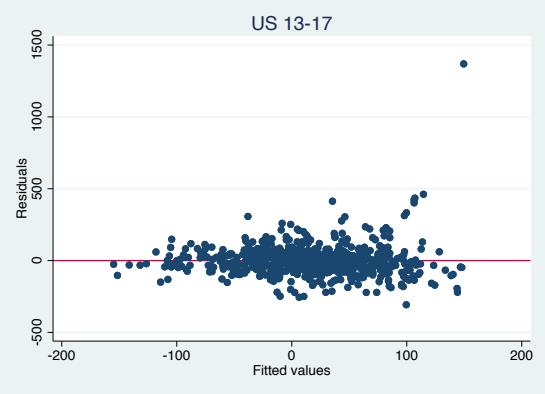
Source: (Wooldridge, 2012)

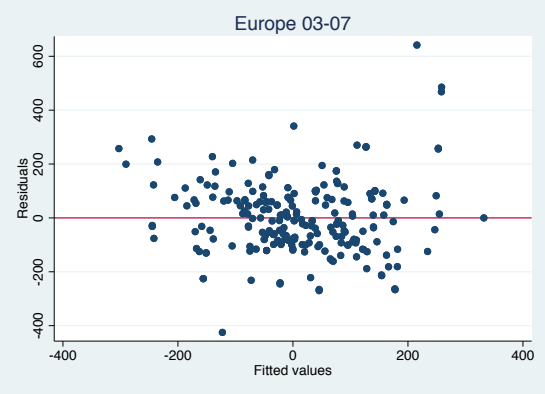
Appendix 4

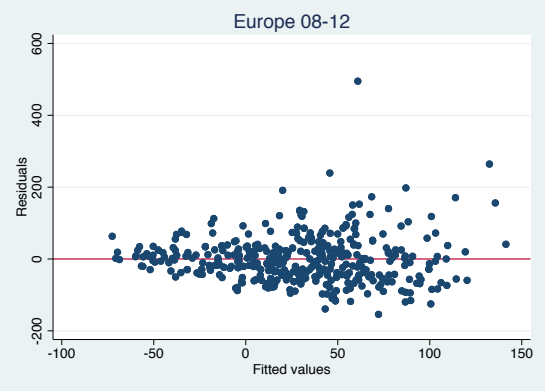
Heteroskedasticity plots for cross sectional regressions

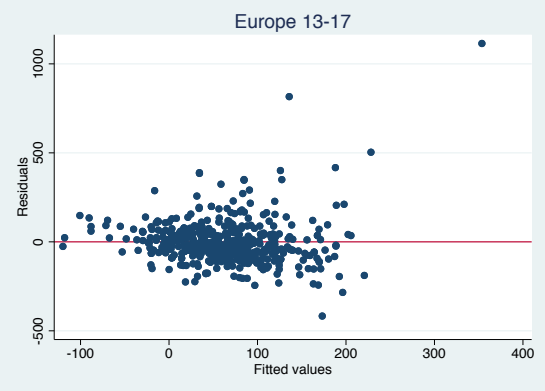












Appendix 5

Fama and French market coverage

Stocks: Rm–Rf for July of year t to June of t+1 include all stocks for which we have market equity data for June of t. SMB, HML, RMW, and CMA for July of year t to June of t+1 include all stocks for which we have market equity data for December of t-1 and June of t, (positive) book equity data for t-1 (for SMB, HML, and RMW), non-missing revenues and at least one of the following: cost of goods sold, selling, general and administrative expenses, or interest expense for t-1 (for SMB and RMW), and total assets data for t-2 and t-1 (for SMB and CMA).

Country	Global	Global ex US	Europe	Japan	Asia Pacific ex Japan	North America
Australia	✔	✔			✔	
Austria	✔	✔	✔			
Belgium	✔	✔	✔			
Canada	✔	✔				✔
Switzerland	✔	✔	✔			
Germany	✔	✔	✔			
Denmark	✔	✔	✔			
Spain	✔	✔	✔			
Finland	✔	✔	✔			
France	✔	✔	✔			
Great Britain	✔	✔	✔			
Greece	✔	✔	✔			
Hong Kong	✔	✔			✔	
Ireland	✔	✔	✔			
Italy	✔	✔	✔			
Japan	✔	✔		✔		
Netherlands	✔	✔	✔			
Norway	✔	✔	✔			
New Zealand	✔	✔			✔	
Portugal	✔	✔	✔			
Sweden	✔	✔	✔			
Singapore	✔	✔			✔	
United States	✔					✔

Source: (French, 2019c)

Appendix 6

ESG variable	Asset4 description
Board Functions (G)	The board of directors/board functions category measures a company's management commitment and effectiveness towards following best practice corporate governance principles related to board activities and functions. It reflects a company's capacity to have an effective board by setting up the essential board committees with allocated tasks and responsibilities.
Board Structure (G)	The board of directors/board structure category measures a company's management commitment and effectiveness towards following best practice corporate governance principles related to a well balanced membership of the board. It reflects a company's capacity to ensure a critical exchange of ideas and an independent decision-making process through an experienced, diverse and independent board.
Compensation Policy (G)	The board of directors/compensation policy category measures a company's management commitment and effectiveness towards following best practice corporate governance principles related to competitive and proportionate management compensation. It reflects a company's capacity to attract and retain executives and board members with the necessary skills by linking their compensation to individual or company-wide financial or extra-financial targets.
Vision and Strategy (G)	The integration/vision and strategy category measures a company's management commitment and effectiveness towards the creation of an overarching vision and strategy integrating financial and extra-financial aspects. It reflects a company's capacity to convincingly show and communicate that it integrates the economic (financial), social and environmental dimensions into its day-to-day decision-making processes.
Shareholder Rights (G)	The shareholders/shareholder rights category measures a company's management commitment and effectiveness towards following best practice corporate governance principles related to a shareholder policy and equal treatment of shareholders. It reflects a company's capacity to be attractive to minority shareholders by ensuring them equal rights and privileges and by limiting the use of anti-takeover devices.
Emission Reduction (E)	The emission reduction category measures a company's management commitment and effectiveness towards reducing environmental emission in the production and operational processes. It reflects a company's capacity to reduce air emissions (greenhouse gases, F-gases, ozone-depleting substances, NOx and SOx, etc.), waste, hazardous waste, water discharges, spills or its impacts on biodiversity and to partner with environmental organisations to reduce the environmental impact of the company in the local or broader community.
Product Innovation (E)	The product innovation category measures a company's management commitment and effectiveness towards supporting the research and development of eco-efficient products or services. It reflects a company's capacity to reduce the environmental costs and burdens for its customers, and thereby creating new market opportunities through new environmental technologies and processes or eco-designed, dematerialized products with extended durability.

Resource Reduction (E)	The resource reduction category measures a company's management commitment and effectiveness towards achieving an efficient use of natural resources in the production process. It reflects a company's capacity to reduce the use of materials, energy or water, and to find more eco-efficient solutions by improving supply chain management.
Product Responsibility (S)	The customer/product responsibility category measures a company's management commitment and effectiveness towards creating value-added products and services upholding the customer's security. It reflects a company's capacity to maintain its license to operate by producing quality goods and services integrating the customer's health and safety, and preserving its integrity and privacy also through accurate product information and labelling.
Community (S)	The society/community category measures a company's management commitment and effectiveness towards maintaining the company's reputation within the general community (local, national and global). It reflects a company's capacity to maintain its license to operate by being a good citizen (donations of cash, goods or staff time, etc.), protecting public health (avoidance of industrial accidents, etc.) and respecting business ethics (avoiding bribery and corruption, etc.).
Human Rights (S)	The society/human rights category measures a company's management commitment and effectiveness towards respecting the fundamental human rights conventions. It reflects a company's capacity to maintain its license to operate by guaranteeing the freedom of association and excluding child, forced or compulsory labour.
Diversity and Opportunity (S)	The workforce/diversity and opportunity category measures a company's management commitment and effectiveness towards maintaining diversity and equal opportunities in its workforce. It reflects a company's capacity to increase its workforce loyalty and productivity by promoting an effective life-work balance, a family friendly environment and equal opportunities regardless of gender, age, ethnicity, religion or sexual orientation.
Employment Quality (S)	The workforce/employment quality category measures a company's management commitment and effectiveness towards providing high-quality employment benefits and job conditions. It reflects a company's capacity to increase its workforce loyalty and productivity by distributing rewarding and fair employment benefits, and by focusing on long-term employment growth and stability by promoting from within, avoiding lay-offs and maintaining relations with trade unions.
Health & Safety (S)	The workforce/health & safety category measures a company's management commitment and effectiveness towards providing a healthy and safe workplace. It reflects a company's capacity to increase its workforce loyalty and productivity by integrating into its day-to-day operations a concern for the physical and mental health, well-being and stress level of all employees.
Training and Development (S)	The workforce/training and development category measures a company's management commitment and effectiveness towards providing training and development (education) for its workforce. It reflects a company's capacity to increase its intellectual capital, workforce loyalty and productivity by developing the workforce's skills, competences, employability and careers in an entrepreneurial environment.

Source: (Thomson Reuters, 2015)

Appendix 7

THOMSON REUTERS BUSINESS CLASSIFICATION

CLASSIFICATION STRUCTURE 2012

ECONOMIC SECTOR: 50 ENERGY

BUSINESS SECTOR	INDUSTRY GROUP	INDUSTRY
5010 Energy - Fossil Fuels	501010 Coal	50101010 Coal
	501020 Oil & Gas	50102010 Integrated Oil & Gas
		50102020 Oil & Gas Exploration and Production
		50102030 Oil & Gas Refining and Marketing
	501030 Oil & Gas Related Equipment and Services	50103010 Oil & Gas Drilling
		50103020 Oil-related Services and Equipment
		50103030 Oil & Gas Transportation Services
5020 Renewable Energy	502010 Renewable Energy	50201010 Renewable Energy Equipment & Services
		50201020 Renewable Fuels
5030 Uranium	503010 Uranium	50301010 Uranium

ECONOMIC SECTOR: 51 BASIC MATERIALS

BUSINESS SECTOR	INDUSTRY GROUP	INDUSTRY
5110 Chemicals	511010 Chemicals	51101010 Commodity Chemicals
		51101020 Agricultural Chemicals
		51101030 Specialty Chemicals
		51101090 Diversified Chemicals
5120 Mineral Resources	512010 Metals & Mining	51201010 Precious Metals & Minerals
		51201060 Gold
		51201070 Mining Support Services & Equipment
		51201080 Integrated Mining
		51201020 Steel
		51201030 Aluminum
		51201050 Specialty Mining & Metals
	512020 Construction Materials	51202010 Construction Materials
5130 Applied Resources	513010 Paper & Forest Products	51301010 Forest & Wood Products
		51301020 Paper Products
	513020 Containers & Packaging	51302010 Non-Paper Containers & Packaging
		51302020 Paper Packaging

Source: (Reuters, 2012)

Appendix 8

US firms' ESG scores 2007-2012

ESG variable	Initial average score 2007 (pct.)	Average change 2007-2012 (pct. points)	Median change 2007-2012 (pct. points)
Board Function (G)	80.82	2.13	1.23
Board Structure (G)	77.86	3.36	1.43
Compensation Policy (G)	64.41	1.24	0.35
Shareholder Rights (G)	71.68	-15.72	-14.91
Vision and Strategy (G)	39.31	12.07	1.69
Emission Reduction (E)	42.53	12.80	3.26
Product Innovation (E)	44.33	8.07	-0.3
Resource Reduction (E)	44.41	14.25	4.85
Community (S)	55.77	2.33	-0.09
Diversity and Opportunity (S)	52.62	6.95	2.74
Employment Quality (S)	53.93	4.96	4.25
Human Rights (S)	42.22	12.06	-6.11
Health and Safety (S)	45.95	3.55	-1.64
Product Responsibility (S)	50.11	1.43	-2.2
Training and Development (S)	44.73	7.47	2.19

Appendix 9

Descriptive statistics tables for US firms

United States 2002, %

	N	Mean	Median	St.Dev	min	max
Alpha 2003-2007	290	35.85	-.01	166.35	-267	1176.81
Board Function (G)	290	70.12	86.12	26.08	6.61	93.44
Board Structure (G)	290	68.07	73.6	21.61	6.88	96.17
Compensation Policy (G)	290	67.61	73.73	20.25	3.27	94.34
Shareholder Rights (G)	290	66.63	71.84	26.53	2.49	96.85
Vision and Strategy (G)	290	36.61	21.56	24.57	21.56	98.63
Emission Reduction (E)	290	36.14	21.17	26.98	10.23	97.67
Product Innovation (E)	290	36.85	27.04	21.67	19.22	99.6
Resource Reduction (E)	290	35.51	21.17	26.27	20.07	96.83
Community (S)	290	47.38	43.95	31.45	5.56	96.23
Diversity and Opportunity (S)	290	44.84	21.36	29.16	7.03	98.64
Employment Quality (S)	290	41.73	36.78	28.9	4.9	98.32
Human Rights (S)	290	41.98	37.43	17.38	2.13	100
Health and Safety (S)	290	39.23	23.18	25.97	4.35	99.54
Product Responsibility (S)	290	40.62	32.79	28.15	3.86	99.07
Training and Development (S)	290	35.91	19.78	25.81	15.79	97.17
Leverage Ratio	290	41.1	40.51	22.86	0	98.12
Market Capitalization (log)	290	15.92	15.72	1.31	10.81	19.44
Price-To-Book Ratio (log)	290	1.19	1.12	.66	-.24	4.44

United States 2007, %

	N	Mean	Median	St.Dev	min	max
Alpha 2008-2012	480	22.58	16.14	77.6	-130.67	489.08
Board Function (G)	480	80.79	83.65	10.37	10.3	93.21
Board Structure (G)	480	77.83	80.77	13.79	15.72	92.9
Compensation Policy (G)	480	64.36	68.83	19.15	11.3	90.85
Shareholder Rights (G)	480	71.48	82.16	23.64	4.56	97
Vision and Strategy (G)	480	39.27	23.11	29.91	12.57	95.32
Emission Reduction (E)	480	42.44	25.39	32.44	8.76	95.25
Product Innovation (E)	480	44.27	28.58	31.28	14.84	96.76
Resource Reduction (E)	480	44.4	39.42	32.74	9.74	94.7
Community (S)	480	55.79	62.56	31.02	4.1	96.9
Diversity and Opportunity (S)	480	52.67	48.32	28.66	9.07	96.23
Employment Quality (S)	480	53.95	55.14	28.54	3.77	96.04
Human Rights (S)	480	42.47	24.87	28.56	9.85	98.56
Health and Safety (S)	480	46.12	32.94	30.91	7.15	97.88
Product Responsibility (S)	480	50.21	49.93	28.71	3.69	97.86
Training and Development (S)	480	44.78	40.23	29.89	7.74	95.06
Leverage Ratio	480	37.7	35.41	23.98	0	99.04
Market Capitalization (log)	480	16.25	16.06	1.17	12.07	20.04
Price-To-Book Ratio (log)	480	1.05	1.01	.65	-1.71	4.79

United States 2012, %

	N	Mean	Median	St.Dev	min	max
Alpha 2013-2017	681	16.82	11.43	119.48	-264.02	1518.44
Board Function (G)	681	81.4	84.63	11.74	11.69	93.22
Board Structure (G)	681	79.85	84.75	12.85	8.96	91.08
Compensation Policy (G)	681	68.2	72.46	17.11	10.21	89.67
Shareholder Rights (G)	681	54.63	56.69	24.84	3.83	98.63
Vision and Strategy (G)	681	43.06	31.25	32.38	7.74	93.67
Emission Reduction (E)	681	46.48	41.03	32.37	7.3	94.41
Product Innovation (E)	681	48.54	39.51	31.76	13.33	96.18
Resource Reduction (E)	681	49.85	53.12	33.07	7.21	93.2
Community (S)	681	50.65	52.42	30.26	3.79	96.34
Diversity and Opportunity (S)	681	53.01	52.1	27.97	4.63	94.9
Employment Quality (S)	681	53.31	53.35	25.26	2.98	97.11
Human Rights (S)	681	49.68	20.71	33.92	11.21	95.31
Health and Safety (S)	681	44.1	35.63	30.92	3.58	97.09
Product Responsibility (S)	681	50.48	40.31	27.81	2.53	97.37
Training and Development (S)	681	45.93	44.31	29.95	5.71	93.98
Leverage Ratio	681	37.74	37.06	21.96	0	99.4
Market Capitalization (log)	681	15.87	15.74	1.22	11.38	20.03
Price-To-Book Ratio (log)	681	.84	.78	.79	-.94	6.59

Appendix 10

Descriptive statistics tables for European firms

Europe 2002, %

	N	Mean	Median	St.Dev	min	max
Alpha 2003-2007	219	10.45	-5.9	179.44	-547.77	857.35
Board Function (G)	219	30.06	21.33	21.38	5.53	88.58
Board Structure (G)	219	33.33	31.37	27.35	1.76	90.71
Compensation Policy (G)	219	43.17	39.33	31.4	2.08	94.48
Shareholder Rights (G)	219	36.55	34.49	24.78	2.43	91.09
Vision and Strategy (G)	219	57.93	53.74	29.62	21.56	98.63
Emission Reduction (E)	219	64.02	74.4	28.86	19.71	97.68
Product Innovation (E)	219	57.44	57.19	29.74	20.74	99.6
Resource Reduction (E)	219	67.01	84.03	29.94	20.07	96.82
Community (S)	219	59.92	67.47	29.33	5.34	96.84
Diversity and Opportunity (S)	219	60.44	68.86	31.53	20.37	98.6
Employment Quality (S)	219	60.88	66.38	30.3	4.9	98.28
Human Rights (S)	219	48.8	37.43	24.63	2.13	100
Health and Safety (S)	219	57.59	72.39	30.82	6.71	99.57
Product Responsibility (S)	219	56.76	48.02	30.5	3.92	99.19
Training and Development (S)	219	66.6	67.32	29.28	19.78	97.73
Leverage Ratio	219	45.96	44.14	22.5	0	95.56
Market Capitalization (log)	219	15.35	15.26	1.35	11.82	18.79
Price-To-Book Ratio (log)	219	.9	.82	.73	-.97	3.73

Europe 2007, %

	N	Mean	Median	St.Dev	min	max
Alpha 2008-2012	419	28.51	14.44	76.42	-96.04	555.91
Board Function (G)	419	43.62	46.52	24.72	4.57	93.17
Board Structure (G)	419	46.03	43.3	25.4	2.37	91.16
Compensation Policy (G)	419	59.61	68.27	26.44	3.36	90.92
Shareholder Rights (G)	419	51.13	51.71	27.05	1.06	96.06
Vision and Strategy (G)	419	66.11	83.16	30.91	12.63	95.33
Emission Reduction (E)	419	66.2	79.6	29.07	10.69	95.14
Product Innovation (E)	419	59.9	66.71	32.5	14.87	96.76
Resource Reduction (E)	419	67.47	80.71	28.53	9.38	94.74
Community (S)	419	58.24	66.88	28.28	2.96	97.04
Diversity and Opportunity (S)	419	64.26	74.25	30.16	5.46	96.23
Employment Quality (S)	419	64.78	75.7	29.15	3.84	96.82
Human Rights (S)	419	57.89	56.78	32.62	17.33	98.56
Health and Safety (S)	419	62.53	65.17	28.78	5.56	97.88
Product Responsibility (S)	419	56.63	57.23	31.25	6.06	97.7
Training and Development (S)	419	70.87	83.64	26.22	7.85	94.64
Leverage Ratio	419	42.68	42.39	23.39	0	97.29
Market Capitalization (log)	419	15.48	15.41	1.5	12.3	19
Price-To-Book Ratio (log)	419	.93	.85	.59	-.24	3.95

Europe 2012, %

	N	Mean	Median	St.Dev	min	max
Alpha 2013-2017	478	63.92	44.81	137.79	-243.85	1468.23
Board Function (G)	478	48.12	53.31	26.28	3.28	93.32
Board Structure (G)	478	52.98	52.24	29.2	4.03	90.97
Compensation Policy (G)	478	61.02	65.16	24.96	2.34	89.75
Shareholder Rights (G)	478	58.09	61.35	28.19	.44	98.92
Vision and Strategy (G)	478	70.1	86.02	27.36	8.73	94.55
Emission Reduction (E)	478	70.56	83.56	27.07	9.05	94.45
Product Innovation (E)	478	62.17	72.29	31.89	13.65	96.19
Resource Reduction (E)	478	70.91	83.39	25.35	7.21	93.22
Community (S)	478	62.94	72.37	28.56	3.39	96.22
Diversity and Opportunity (S)	478	71.33	84.84	27.24	7.34	94.9
Employment Quality (S)	478	68.29	79.66	27.92	3.29	97.02
Human Rights (S)	478	63.64	88.89	32.86	18.74	95.32
Health and Safety (S)	478	64.69	66.16	26.56	10.52	97.08
Product Responsibility (S)	478	61.65	66.15	29.5	4.78	97.53
Training and Development (S)	478	73.08	81.52	22.21	5.88	94.48
Leverage Ratio	478	39.28	38.38	24.19	0	94.53
Market Capitalization (log)	478	15.05	14.98	1.62	8.11	18.91
Price-To-Book Ratio (log)	478	.49	.46	.77	-2.21	5.25

Appendix 11

Cross-sectional Regression

	(1) **US Alpha 03-07**	(2) **US Alpha 08-12**	(3) **US Alpha 13-17**	(4) **Europe Alpha 03-07**	(5) **Europe Alpha 08-12**	(6) **Europe Alpha 13-17**
Board Function (G)	-0.34	-0.20	0.52*	1.26**	-0.09	0.26
	(0.44)	(0.56)	(0.31)	(0.58)	(0.17)	(0.37)
Board Structure (G)	0.15	-0.11	-0.52	-1.13**	0.50***	-0.12
	(0.49)	(0.28)	(0.38)	(0.47)	(0.17)	(0.30)
Compensation Policy (G)	-0.12	0.20	0.32	-0.24	0.39**	0.21
	(0.34)	(0.18)	(0.26)	(0.42)	(0.18)	(0.33)
Shareholder Rights (G)	-0.01	-0.03	-0.18	-0.42	-0.03	-0.11
	(0.39)	(0.15)	(0.18)	(0.52)	(0.12)	(0.26)
Vision and Strategy (G)	0.32	-0.21	0.55	-0.38	-0.38	0.60
	(0.49)	(0.19)	(0.37)	(0.47)	(0.24)	(0.38)
Emission Reduction (E)	-0.99	0.18	0.26	-0.06	-0.16	-0.01
	(0.69)	(0.23)	(0.40)	(0.71)	(0.28)	(0.51)
Product Innovation (E)	1.29**	-0.21*	-0.37*	0.97*	0.05	-0.34
	(0.52)	(0.12)	(0.22)	(0.53)	(0.17)	(0.30)
Resource Reduction (E)	-0.24	-0.24	-0.39	0.76	0.40	-0.24
	(0.63)	(0.19)	(0.26)	(0.58)	(0.27)	(0.50)
Community (S)	-0.16	-0.19	-0.03	0.92*	-0.09	-0.32
	(0.28)	(0.15)	(0.20)	(0.55)	(0.16)	(0.28)
Diversity and Opportunity (S)	0.97**	-0.08	0.41*	0.52	0.17	0.85
	(0.38)	(0.14)	(0.21)	(0.57)	(0.17)	(0.52)
Employment Quality (S)	-0.09	0.25**	0.20	1.18**	0.04	0.17
	(0.29)	(0.12)	(0.18)	(0.51)	(0.16)	(0.34)
Human Rights (S)	0.41	0.16	-0.23	-0.35	-0.05	-0.19
	(0.35)	(0.16)	(0.17)	(0.53)	(0.16)	(0.25)
Health and Safety (S)	0.09	0.03	-0.56**	-0.93*	0.04	-0.04
	(0.56)	(0.15)	(0.26)	(0.50)	(0.17)	(0.31)
Product Responsibility (S)	-0.02	-0.07	-0.21	-0.03	-0.39***	0.03
	(0.35)	(0.12)	(0.18)	(0.42)	(0.15)	(0.31)
Training and Development (S)	0.14	0.36**	0.14	-0.20	0.07	-0.03
	(0.35)	(0.15)	(0.22)	(0.44)	(0.19)	(0.41)
Leverage Ratio	0.06	-0.09	-0.31	-0.90	-0.34	-0.43
	(0.52)	(0.17)	(0.26)	(0.75)	(0.23)	(0.45)
Market Capitalization (log)	-22.40***	-1.25	10.14**	-28.42**	-1.75	-23.36*
	(7.80)	(4.10)	(4.89)	(11.18)	(4.02)	(12.04)
Price-To-Book Ratio (log)	-15.98	-0.70	-4.47	-35.65**	3.89	-8.16
	(15.33)	(5.67)	(7.44)	(17.97)	(7.38)	(8.50)
Constant	435.59**	52.53	-103.86	407.65**	72.00	450.23***
	(195.12)	(93.57)	(94.38)	(170.28)	(50.84)	(154.80)
Obs.	290	480	681	219	419	478
R-squared	0.45	0.30	0.19	0.39	0.27	0.17
Adjusted R-squared	0.35	0.24	0.14	0.25	0.19	0.09
Root mean square error	133.67	67.85	110.92	155.00	68.63	131.66

Standard errors are in parenthesis
*** p<0.01, ** p<0.05, * p<0.1
Industry and European country dummy variables omitted for brevity

Appendix 12

Cross-sectional Three-year Regression

	(1) **US Alpha 03-05**	(2) **US Alpha 08-10**	(3) **US Alpha 13-15**	(4) **Europe Alpha 03-05**	(5) **Europe Alpha 08-10**	(6) **Europe Alpha 13-15**
Board Function (G)	-0.08	0.16	0.29*	0.27	-0.03	0.48
	(0.28)	(0.24)	(0.17)	(0.41)	(0.10)	(0.38)
Board Structure (G)	-0.14	0.02	-0.22	-0.70**	0.12	-0.20
	(0.35)	(0.15)	(0.17)	(0.30)	(0.13)	(0.30)
Compensation Policy (G)	-0.29	-0.01	0.34**	-0.68**	0.20	0.12
	(0.28)	(0.10)	(0.13)	(0.33)	(0.13)	(0.30)
Shareholder Rights (G)	0.06	0.06	-0.10	-0.72**	-0.11	-0.06
	(0.27)	(0.09)	(0.09)	(0.28)	(0.09)	(0.20)
Vision and Strategy (G)	0.14	-0.09	0.11	-0.27	-0.11	0.25
	(0.34)	(0.10)	(0.13)	(0.30)	(0.17)	(0.25)
Emission Reduction (E)	-0.37	0.24**	0.07	-0.30	0.02	-0.03
	(0.46)	(0.11)	(0.16)	(0.47)	(0.17)	(0.42)
Product Innovation (E)	0.55	-0.11	-0.21**	0.41	-0.06	-0.05
	(0.37)	(0.07)	(0.10)	(0.32)	(0.11)	(0.24)
Resource Reduction (E)	-0.06	-0.15	-0.12	0.22	-0.03	-0.39
	(0.45)	(0.10)	(0.12)	(0.40)	(0.17)	(0.45)
Community (S)	-0.14	-0.08	-0.12	0.41	-0.12	-0.17
	(0.21)	(0.09)	(0.11)	(0.29)	(0.11)	(0.19)
Diversity and Opportunity (S)	0.34	-0.16*	0.31***	0.05	0.07	0.64
	(0.27)	(0.08)	(0.12)	(0.31)	(0.12)	(0.48)
Employment Quality (S)	0.31	0.13*	0.26***	0.40	0.06	0.43
	(0.22)	(0.07)	(0.09)	(0.34)	(0.10)	(0.30)
Human Rights (S)	-0.29	0.05	-0.14	-0.56*	-0.01	-0.03
	(0.26)	(0.08)	(0.09)	(0.30)	(0.11)	(0.20)
Health and Safety (S)	0.28	0.02	-0.24**	-0.09	0.13	-0.27
	(0.29)	(0.08)	(0.10)	(0.29)	(0.11)	(0.24)
Product Responsibility (S)	0.08	-0.03	-0.12	-0.11	-0.13	0.04
	(0.21)	(0.07)	(0.09)	(0.27)	(0.09)	(0.20)
Training and Development (S)	0.22	0.22***	0.05	0.28	0.09	0.03
	(0.24)	(0.08)	(0.10)	(0.33)	(0.12)	(0.31)
Leverage Ratio	0.12	-0.05	0.02	0.29	-0.27**	0.19
	(0.35)	(0.10)	(0.12)	(0.50)	(0.12)	(0.46)
Market Capitalization (log)	-33.45***	-7.31***	8.86***	-12.02	-0.79	-24.42**
	(10.68)	(2.13)	(2.46)	(8.21)	(2.63)	(12.36)
Price-To-Book Ratio (log)	-18.10	1.72	-3.64	-33.65***	6.63	-0.31
	(14.40)	(3.48)	(4.21)	(10.04)	(4.32)	(7.16)
Constant	520.77***	116.92***	-147.97***	318.38***	13.13	400.88***
	(152.20)	(43.50)	(45.78)	(112.71)	(34.51)	(142.30)
Obs.	290	480	681	219	419	478
R-squared	0.35	0.30	0.30	0.43	0.32	0.25
Adjusted R-squared	0.24	0.24	0.26	0.28	0.24	0.17
Root mean square error	94.17	37.59	52.43	87.52	42.78	98.50

Standard errors are in parenthesis
*** p<0.01, ** p<0.05, * p<0.1
Industry and European country dummy variables omitted for brevity

Appendix 13

Cross-sectional Regression E, S and G Scores

	(1) **US Alpha 03-07**	(2) **US Alpha 08-12**	(3) **US Alpha 13-17**	(4) **Europe Alpha 03-07**	(5) **Europe Alpha 08-12**	(6) **Europe Alpha 13-17**
Environmental Score	0.14	-0.41**	-0.33	1.59**	-0.04	-0.39
	(0.48)	(0.20)	(0.27)	(0.68)	(0.26)	(0.62)
Social Score	1.29**	0.42	0.18	1.41**	-0.12	0.60
	(0.55)	(0.28)	(0.34)	(0.66)	(0.33)	(0.76)
Governance Score	-0.58	-0.20	0.44	-1.67	0.07	0.41
	(0.69)	(0.63)	(0.49)	(1.07)	(0.33)	(0.65)
Market Capitalization (log)	-23.62***	-2.97	9.30**	-18.36*	-3.11	-23.22**
	(7.67)	(3.81)	(4.36)	(10.82)	(3.79)	(9.67)
Price-To-Book Ratio (log)	-15.31	2.08	-6.42	-31.92	3.73	-8.54
	(15.37)	(5.36)	(7.46)	(19.73)	(7.45)	(9.00)
Constant	494.52**	85.33	-130.76*	528.07***	106.66**	431.09***
	(197.53)	(77.93)	(75.60)	(136.42)	(43.78)	(125.72)
Obs.	290	480	681	219	419	478
R-squared	0.42	0.28	0.16	0.36	0.26	0.16
Adjusted R-squared	0.35	0.23	0.12	0.24	0.19	0.09
Root mean square error	133.80	68.08	112.19	156.03	68.71	131.20

Standard errors are in parenthesis
*** p<0.01, ** p<0.05, * p<0.1
Industry and European country dummy variables omitted for brevity

Cross-sectional Regression ESG Scores

	(1) **US Alpha 03-07**	(2) **US Alpha 08-12**	(3) **US Alpha 13-17**	(4) **Europe Alpha 03-07**	(5) **Europe Alpha 08-12**	(6) **Europe Alpha 13-17**
ESG Score	0.99	-0.29	-0.08	2.62**	-0.13	0.35
	(0.62)	(0.27)	(0.33)	(1.07)	(0.30)	(0.54)
Leverage Ratio	0.06	-0.16	-0.25	-0.32	-0.26	-0.36
	(0.49)	(0.16)	(0.26)	(0.75)	(0.22)	(0.47)
Market Capitalization (log)	-20.17***	-3.00	8.44**	-18.08	-3.28	-23.25**
	(7.59)	(3.74)	(4.26)	(11.26)	(3.77)	(9.58)
Price-To-Book Ratio (log)	-14.61	2.57	-5.75	-40.47**	3.64	-6.79
	(15.30)	(5.47)	(7.36)	(19.69)	(7.27)	(8.63)
Constant	335.69***	87.49	-92.08	564.47***	109.37**	434.13***
	(123.45)	(59.23)	(66.99)	(144.22)	(43.00)	(118.08)
Obs.	290	480	681	219	419	478
R-squared	0.41	0.27	0.16	0.33	0.26	0.16
Adjusted R-squared	0.35	0.23	0.12	0.22	0.20	0.09
Root mean square error	134.07	68.25	112.15	158.14	68.54	131.09

Standard errors are in parenthesis
*** p<0.01, ** p<0.05, * p<0.1
Industry and European country dummy variables omitted for brevity

Appendix 14

Time-series table with positive alpha portfolio (top and bottom 10% portfolios)

Europe Time-series CAPM Top and Bottom 10%

ESG variable portfolios	Market premium	Alpha	Adjusted R2
Board Function (G) 2002 Top 10%	0.60***	0.01	0.32
t-value	(6.12)	(1.48)	
Board Function (G) 2002 Bottom 10%	0.75***	0.01	0.27
t-value	(4.84)	(1.08)	
Board Structure (G) 2002 Top 10%	0.61***	0.00	0.24
t-value	(5.13)	(0.05)	
Board Structure (G) 2002 Bottom 10%	0.84***	0.01	0.31
t-value	(5.82)	(0.97)	
Employment Quality (S) 2002 Top 10%	0.73***	0.00	0.32
t-value	(6.10)	(0.64)	
Employment Quality (S) 2002 Bottom 10%	0.94***	-0.01	0.39
t-value	(5.42)	(-1.06)	
Board Structure (G) 2007 Top 10%	**0.72*** **	**0.01** **	**0.44**
t-value	**(5.83)**	**(2.01)**	
Board Structure (G) 2007 Bottom 10%	0.63***	-0.00	0.42
t-value	(5.87)	(-0.02)	
Compensation Policy (G) 2007 Top 10%	0.73***	0.01	0.42
t-value	(6.17)	(1.59)	
Compensation Policy (G) 2007 Bottom 10%	0.65***	-0.00	0.43
t-value	(5.95)	(-0.62)	
Product Responsibility (S) 2007 Top 10%	0.66***	0.00	0.36
t-value	(4.52)	(0.09)	
Product Responsibility (S) 2007 Bottom 10%	0.76***	0.01	0.44
t-value	(6.49)	(1.47)	

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

Appendix 15

Endogeneity test US

	(1) **Employment Quality 2007**	(2) **Training & Development 2007**	(3) **Health & Safety 2012**
Alpha 2003-2007	0.01	0.01	
	(0.01)	(0.01)	
Alpha 2008-2012			-0.01
			(0.01)
Leverage Ratio	-0.01	0.01	0.15***
	(0.09)	(0.08)	(0.05)
Market Capitalization (log)	7.70***	9.07***	12.36***
	(1.34)	(1.29)	(0.96)
Price-To-Book Ratio (log)	8.15***	2.56	-2.50
	(2.59)	(2.31)	(1.84)
Constant	-116.74***	-125.13***	-162.47***
	(30.07)	(24.42)	(21.88)
Obs.	290	290	480
R-squared	0.28	0.27	0.46
Adjusted R-squared	0.21	0.19	0.43
Root mean square error	25.73	23.18	23.36

Standard errors are in parenthesis
*** p<0.01, ** p<0.05, * p<0.1
Industry and European country dummy variables omitted for brevity

Endogeneity test Europe

	(1) **Board Structure 2007**	(2) **Compensation Policy 2007**	(3) **Product Responsibility 2007**
Alpha 2003-2007	-0.01	-0.00	0.02*
	(0.01)	(0.02)	(0.01)
Leverage Ratio	-0.29***	-0.16	-0.12
	(0.10)	(0.12)	(0.11)
Market Capitalization (log)	-2.39	-0.84	3.65**
	(1.85)	(1.89)	(1.72)
Price-To-Book Ratio (log)	-1.41	-0.11	1.49
	(3.45)	(4.39)	(3.77)
Constant	72.88***	32.32	45.70*
	(26.19)	(28.37)	(24.85)
Obs.	219	219	219
R-squared	0.18	0.13	0.17
Adjusted R-squared	0.07	0.01	0.05
Root mean square error	26.44	31.26	29.74

Standard errors are in parenthesis
*** p<0.01, ** p<0.05, * p<0.1
Industry and European country dummy variables omitted for brevity