The ESG valuation curve: Are ESG score downgrades punished less than upgrades are being rewarded?

Master Thesis

October 2021

Erasmus School of Economics, Erasmus University Rotterdam

ERASMUS UNIVERSITY ROTTERDAM

Author: Sheda Delfan

Student number: 451826

Msc. Financial Economics

Supervisor: Dr. Ronald Huisman

Co-supervisor:

Keywords: Sustainability, ESG investing, The prospect theory, ESG valuation curve, Sustainable return, Financial return, Social return

Table of contents

1. Introduction **4**

1.1. Thesis outline 6

2. Literature review and hypotheses **7**

2.1. Sustainability reflected in ESG ratings 7

2.2. The concept of social return 9

2.3. The prospect theory in the light of sustainable return 11

2.4. Hypotheses 12

3. Data and methodology **13**

3.1. Data on sustainability 13

3.2. Data on valuation 14

3.3. Other factors 15

3.4. Construction of the final dataset 15

3.5. Methodology 16

3.6. Descriptive statistics 19

4. Results **19**

4.1. Main results 19

4.2. Robustness test 24

4.3. Interpretation of results 25

5. Conclusion **27**

5.1. Overall conclusion 27

5.2. Limitations and further research 28

References **30**

Appendix **32**

Abstract

This paper examines the financial performance effect of ESG score changes, in response to the MSCI study by Giese and Nagy (2018). Sustainable returns, measured by the year-on-year changes of ESG scores, are discussed in the light of the prospect theory by Kahneman and Tversky (1979). Differences between financial return and sustainable return are assessed by means of a qualitative and quantitative study. In contrast to my expectations, I have found that investors place more weight on losses in sustainability than on equivalent gains, although this result is only significant for extreme ESG score changes (upgrades and downgrades of more than 10%). Another finding is that high ESG score upgrades have a lower financial performance effect than low ESG score upgrades. This finding can be explained when making a distinction between two parts of sustainable return: the part that materializes into financial value ("financial return") and the part that does not ("social return"). In this paper, I will elaborate on these components of sustainable return. My aim with this study is to provide new insights on how investors value gains and losses of sustainability.

1. Introduction

Over the last years, we experience ongoing debate about the financial benefits of sustainable investing. A survey of KPMG (2020) among hedge fund managers shows that institutional investors are currently the main drivers of demand for sustainable investments. From the perspective of investors, allocations to socially oriented hedge funds were mainly driven by opportunities to generate alpha and by the argument that sustainability issues are material to the financial performance of their portfolio companies (KPMG, 2020). Many studies have examined the materiality of sustainability issues by looking at the relationship between Environmental, Social and Governance (ESG) concerns and corporate financial performance (Friede et al, 2015; Revelli & Viviani, 2015). However, a clear consensus among researchers is lacking, partly because studies differ significantly in ESG rating methodologies and financial metrics used to assess financial performance (Berg et al., 2019). Another potential reason for the insignificant relationship could be the difficulty in making a distinction between sustainability issues that materialize into financial value versus issues that do not. I will refer to this unmaterialized part of sustainable return as "social return" throughout this research. Prior academic literature has failed to distinguish between the two concepts, and this may affect the results of those studies.

In addition, most existing studies fail to prove causation since it is a kind of chicken and egg situation. Companies with a strong ESG proposition are often accompanied with decent growth opportunities, strong balance sheets and good management. It is not unlikely that those companies, who are often in better financial shape, invest more in ESG concerns (Koller et al., 2020, Chapter 6). On behalf of MSCI, Giese and Nagy (2018) go further than the existing literature by assessing the question of causality between ESG scores and stock prices. Does the improvement of a companies' ESG proposition directly drive share price performance or is it merely coincidental? The MSCI study focused on the pricing of ESG Momentum (the financial value of changes in companies' ESG profiles) by equity markets. In contrast to ESG scores, ESG Momentum scores historically show uncorrelation to any of the equity style factors used in the study (See Appendix 1). Giese and Nagy (2018) use year-on-year changes of MSCI ESG scores as a proxy for ESG Momentum. As their results show, changes in ESG scores had an impact on equity prices that could not be explained by the general market or other factors. More specifically, their results indicate that stocks that exhibited the greatest positive change in ESG scores financially performed best (in both developed and emerging markets).

One of the outcomes of Giese and Nagy (2018) relates to the question whether the financial impact of ESG score upgrades or downgrades depends on the initial ESG proposition of the company. To understand this mechanism, the companies were divided into three groups of initial ESG score levels: low, medium, and high. For each tertile, long-short performance analyses were conducted. The portfolios consisted of long-positions in the top quintile of ESG Momentum score and short-positions in the bottom quintile (See Appendix 2). In that way, the historical performance of the top versus the bottom ESG Momentum scores were calculated per tertile. The study found that the strongest ESG Momentum performance was generated in the medium tertile, meaning that companies in this tertile experienced the strongest performance effect when their ESG score changed. This strong performance effect could not be explained by higher levels of risk since the realized volatilities of the portfolios were similar (around 4%). In addition, the study suggests that the ESG valuation curve is non-linear for developed markets (See Appendix 3). The curve shows that ESG score changes have the strongest financial performance effect when curve is steepest and thus for companies in the middle tertile of initial ESG score. At the top and bottom end of initial ESG score profiles, a change of ESG score had less of a financial performance effect. A last finding of the study that I find remarkable is that ESG score downgrades are seemed to be punished less than ESG score upgrades are being rewarded. The curve in Appendix 3 shows that valuation is affected to a lesser extent when the ESG score goes down than when it rises.

The results of the MSCI study raise many questions. The curve in Appendix 3 suggests that the perception of changes in valuation because of adjustments in ESG scores is greater for companies with a medium initial ESG score than for companies in the low and high tertile. Thus, the volatility costs of ESG score changes seem to be greater for companies in the medium tertile. This contradicts to the authors' claim that the difference in financial performance effect between low, medium, and high score companies does not follow from different levels of risk. If volatility would be the same for every tertile, the ESG valuation curve would be either step wise or linear. A non-linear ESG valuation curve with very similar realized volatilities for the three tertiles is questionable. How could the curve be non-linear when volatility remains the same and still be a gradual curve?

According to the curve, companies that initially score low on ESG concerns seem to have the most potential since their option value of becoming sustainable is the most significant. However, this option value materializes slowly due to the small performance effect at the beginning of the curve. Investors with a short-term investment horizon would therefore prefer to invest in companies in the medium tertile that are taking steps in the field of ESG. Because those companies have the greatest incentive to quickly improve their ESG proposition —and thereby their valuation. If this is true, its implication might not be desirable from a society point of view. From that point of view, it might be better if low-scoring companies have the greatest incentives to boost their ESG scores in the short term.

Finally, the outcome that ESG score downgrades are punished less than upgrades are being rewarded may not be preferred from a society point of view. This result is very interesting because it reminds me of the prospect theory of Kahneman and Tversky (1979). According to this well-known theory, investors value gains and losses differently by placing higher weight on losses. When applying the prospect theory to the concept of ESG valuation, I would expect a decline in sustainability to hurt more than an increase in sustainability is being rewarded. But MSCI's ESG valuation curve shows the opposite.

To test to which extent the ESG valuation curve of Giese and Nagy (2018) holds, I will focus on examining the following question:

"Are ESG score downgrades punished less than upgrades are being rewarded?"

Even though there are several other aspects of MSCI's ESG valuation curve that ask for further investigation, I believe answering the question above is most valuable due to its implications. If the performance effect indeed differs between ESG upgrades and downgrades, how does this relate to Kahneman and Tversky's prospect theory? Perhaps different outcomes can be attributed to behaviour of sustainable return versus financial return. Assuming that there is a part of sustainable return that does not materialize into financial value, to which extent is it then valid to incorporate sustainability into traditional valuation models?

1.1. Thesis outline

This thesis is organized as follows. In the following section, I will summarize important literature related to concepts that are relevant for answering the question brought up above. At the end of the following section, the three hypotheses that will be tested in this study are presented. Information on the data and methodology of the quantitative part of this study can be found in Section 3 and the results of the analysis are presented in Section 4. Finally, I will discuss the results and their implications in Section 5.

2. Literature review and hypotheses

As discussed in the previous section, the question that underlies this literature review is how investors value "gains and losses" of sustainability. Regarding financial return, the prospect theory tells us that investors dislike financial losses more than they reward equivalent gains. Suppose you apply the prospect theory to the concept of sustainable return: You would expect that investors value a decline in sustainable return more negatively than they would value the same increase positively. MSCI's ESG valuation curve shows us the opposite, however. To which extent do sustainable and financial return then overlap? Does sustainable return also contain some kind of "social return" that must be considered when examining the performance effect of sustainable return?

To address the issue above, this section will touch on the following topics: ESG ratings to measure sustainability, the concept of social return, and the application of Kahneman and Tversky's prospect theory on sustainable return.

2.1. Sustainability reflected in ESG ratings

In the field of finance, sustainability can be best defined based on environmental, social and governance (ESG) concerns. In certain industries, for example in the transportation industry, the environmental factor may be of greater weight, whereas in other industries the focus lies on the social aspect of ESG. Think of the pharmaceutical industry, where inequality due to the wealth gap makes affordable prices of medicines the most pressing issue at the time. In addition, the maturity level of the transition towards sustainability does also vary heavily between industries and regions. For instance, when looking at the power and utilities industry, we see that the journey towards sustainability started around fifteen years ago. Whereas in the agricultural industry, sustainability has become a core issue only recently (Koller et al., 2020, Chapter 6). In terms of geography, studies show that Western companies have more recognition for sustainability as a core strategic question than companies in non-Western markets do (Kaplan & Montiel, 2017). These regional divergences could be driven by different levels of regulatory pressure (Nehrt, 1996).

The use of ESG ratings to integrate sustainability in the investment decision-making process is relatively new. Over the last fifteen years, different providers (Asset4, Bloomberg, MSCI) have entered the ESG rating market and started offering investors a broad range of products and services in the field of ESG. Also, institutional investors like hedge funds are building internal models to integrate sustainability into their investment decisions. Publicly available company data as well as data provided by NGOs, governmental institutions, or trade unions are used to calculate ESG scores. One of the problems that ESG rating agencies face is that they are highly dependent on the data consistency of ESG performance data reported by companies. On average, larger firms tend to have better ESG scores; that's probably because larger companies have more resources to develop their ESG policies and to report on their ESG activities. Kotsantonis and Serafeim (2019) show how differences in the way that companies report their ESG data can lead to very different rating outcomes. Another problem in this field is that the definition of a companies' peer group is crucial in determining its ESG score. However, there exists a lack of transparency among ESG rating agencies about how those peer groups are composed. Due to different methodologies, ESG ratings seem to vary greatly between rating agencies (Kotsantonis & Serafeim, 2019). A recent MIT study by Berg et al. (2020) confirms that the correlation between ESG scores from different rating agencies is limited.

Despite the problems with ESG ratings as described above, plenty of studies have used ESG scores to examine the link between sustainability and financial returns. Henisz et al. (2019) state five ways that ESG may be linked to cash flows and thereby may affect the valuation of companies. First, the study argues that a strong ESG proposition can create growth opportunities due to consumer preferences and the perceptions of stakeholders. Second, the study claims that cost reductions can be achieved when for example lowering the energy consumption or water intake. Improving the ESG proposition would also be beneficial when trying to avoid regulatory and legal interventions or when attempting to increase employee productivity. Fifth, on the long term, a strong ESG proposition can enhance investment returns due to better capital allocation (think of more sustainable plant and equipment). Consistent with what is mentioned before, Henisz et al. (2019) point out that these links apply to a different degree for certain industries, sectors, and geographies. In a meta-analysis, Friede et al. (2015) study over 2,200 papers on the relationship between ESG and financial performance and find that around 90% of the papers indicates a non-negative relationship between the two variables, which basically means that ESG stocks do not underperform on average.

Most studies that examined the link between financial performance and sustainability fail to provide clear evidence on which fundamental value drivers ensure that ESG stocks will not underperform on the long term. A logical explanation for the popularity of ESG stocks among investors nowadays could also be that investors agree to a lower financial return when they invest in ESG stocks. In addition to financial motives, I would expect social motives to also play a role in investment decisions. As this “social return” is not included in most of such studies, it is an interesting concept to look further into.

2.2. The concept of social return

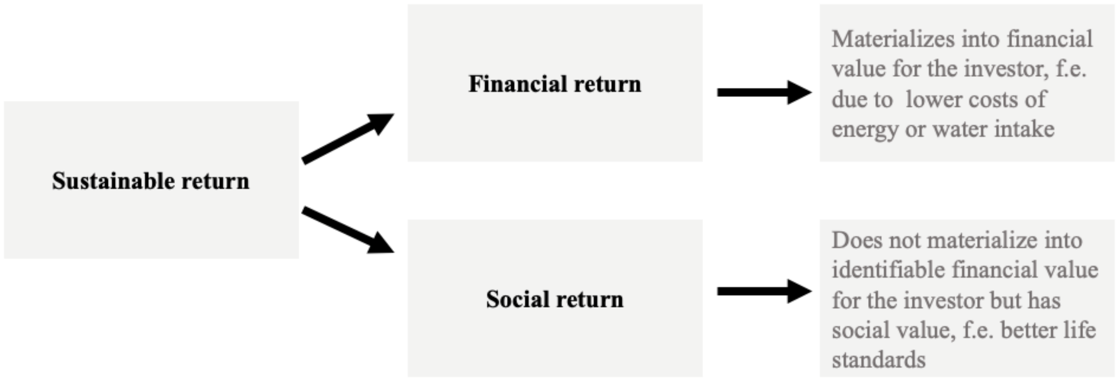
When assessing the performance effect of sustainable investments, it does not seem inconceivable to me that investors also appreciate pure social benefits their investments. A survey among 600 institutional investors shows that most respondents would consider investing with a lower rate of return if it meant investing in ESG stocks (Edelman, 2019). When looking at the bigger picture, consumers are also willing to pay a premium when they know that their purchase is considered sustainable (Miremadi et. al, 2012; Zhong & Chen, 2019; Koller et al., 2020). Social return may thus play a role in the decision to invest, although it’s not always easy to express its value in monetary terms. Studies that simply equate the “excess” social return of an investment with the premium that investors are willing to pay above the price of a comparable non-ESG stock, fail to distinguish between the part of sustainable return that does materialize into financial value and the part that does not.

Definitions do matter in this case. If you define the premium paid for ESG stocks as *the price of sustainability*, you could argue that the price of sustainability consists out of two parts as illustrated in Figure 1. First, a price paid for sustainability today to obtain possible financial returns in the future. And second, a price paid for sustainability now for possible social returns in the future. The part of sustainable return that will realize in terms of financial value refers to the part that can be directly linked to cash flows. Think of the previously discussed study by Henisz et al. (2019), who states five ways how ESG improvements may lead to the generation of cash flows. The second part refers to the concept of social return. This part does not materialize into identifiable financial value for the investor but has a certain social impact. This social impact has value today since it influences investor sentiment.

As an example of this framework, think of an investor who is deciding on whether to participate in a new funding round of an early-stage med tech company. The investor decides based on several factors, such as his or her perception of the company’s valuation, confidence in the management, etcetera. But what also might play a role in the decision to invest, is the feeling that investor gets from investing. If the investor attaches a certain positive value to investing in socially desirable projects, this will probably play a role in the investment decision process. However, this value does not materialize into identifiable financial return for the investor and therefore might not behave as such.

Figure 1

Framework sustainable return.



To come back to the MSCI study by Giese and Nagy (2018): By focusing on the impact of upgrades and downgrades in sustainability on stock prices, the study implicitly included both financial and social returns –assuming that possible social returns are included in the excess return of ESG stocks through investor sentiment (See Baker & Wurgler, 2007; Piñeiro-Chousa et al., 2021). Subsequently, the two concepts were lumped together when the authors examined the overall performance effect of sustainable returns, ignoring the distinction between a financial performance effect and a social performance effect. Most studies do not make this distinction, which has its consequences for the interpretation of their results since it could be that these two types of returns behave differently.

Khan et al. (2016) do make a distinction between ESG issues that are material and issues that are immaterial. They found that firms with good performance on material ESG issues and concurrently poor performance on immaterial issues outperformed other companies in their data sample, which means that those firms also outperformed companies that did well on both material and immaterial ESG issues. The results of this study tell us that if the only goal of a firm would be to financially outperform other firms, it can best focus on the part of sustainable return that does materialize into financial value and neglect the part that does not. Khan et al (2016) thus point out that a distinction must be made between the two parts of sustainable return.

I will explain this point further in the next chapter by shedding a light on the results of Giese and Nagy (2018) regarding the difference in the performance effect of ESG score upgrades versus downgrades.

2.3. *The prospect theory in the light of sustainable return*

As mentioned, one of the outcomes of Giese and Nagy (2018) particularly caught my attention. Their results show that the performance effect of ESG score changes is less strong for ESG score downgrades than for upgrades. Since this outcome reminded me of the prospect theory of Kahneman and Tversky (1979), I started thinking about the behaviour of investors towards sustainable return. Why is this result at odds with the prospect theory and what does that tell us about characteristics of sustainable return compared to financial return?

According to the prospect theory, individuals are subject to biases and violations when it comes to the standard economic theory of expected utility. Kahneman and Tversky (1979) propose that these inconsistencies are mainly caused by loss-aversion and the phenomenon that people evaluate a specific prospect based on the utility they receive from deviating from their personal reference point. While the utility function is being concave for gains and convex for losses, the function is around two times steeper for losses than for gains. This implies that a loss reduces the utility of an individual two times more than an equivalent gain increases it. It also means that utility is less affected with every extra gain or loss (Tversky & Kahneman, 1992).

At first sight, the result of Giese and Nagy (2018) on excess stock returns caused by ESG score fluctuations cannot be reconciled with the prospect theory as such. Figure 2 shows that the utility curve of the prospect theory (left) and MSCI's ESG valuation curve (right) are horizontally symmetric.

Figure 2

Prospect theory versus MSCI's ESG valuation curve

What are the differences? The prospect theory tells us something about the utility of financial gains and losses, and thus the impact of financial returns on the utility of an individual. The ESG valuation curve shows how changes in ESG ratings affect the excess price that investors are willing to pay for a stock, and thus the impact of sustainable returns on stock prices. The price of a stock is determined by the price that investors are willing to pay for the stock, and excess stock returns that can be linked to ESG score fluctuations can therefore be defined as the price investors are willing to pay for sustainability. Following the prospect theory, an investor is loss-averse since he or she places more weight on financial losses than on equivalent gains. From that point of view, the question is if investors are as much loss averse when it concerns sustainable returns. In other words, when the ESG score of a stock goes down, do investors face the same loss-aversion regarding the excess price they are willing to pay for sustainability? The results of Giese and Nagy (2018) tell us otherwise: ESG upgrades seem to outweigh downgrades from an investor perspective.

If the results of Giese and Nagy (2018) are true, what can we say about the way investors perceive sustainable returns versus financial returns? To what extent are studies that link sustainability to corporate financial performance correct? As discussed in the previous chapter, Khan et al. (2016) already touch upon the differences between material and immaterial ESG issues. In this study, my aim is to further explore these differences in the behaviour of sustainable return versus financial return. I will do this by testing the four hypotheses that are stated in the following chapter.

2.4. Hypotheses

In response to the ESG valuation curve of Giese and Nagy (2018) and the application of Kahneman and Tversky's (1979) prospect theory on the concept of sustainable return, I will test the following four hypotheses:

Hypothesis (1)

H_0: There is no relationship between ESG score changes and excess stock returns.

H_a: ESG score changes are positively related to excess stock returns.

Hypothesis (2)

H_0: Investors do not place less weight on ESG score downgrades than they do on equivalent upgrades.

H_a: Investors value gains and losses of sustainability differently, placing less weight on ESG score downgrades than they do on equivalent upgrades.

The first step is to determine if there is a relationship between ESG score changes and expected excess stock returns. A positive relationship means that investors reward upgrades in sustainability and punish downgrades. Giese and Nagy (2018) show that on average, ESG score downgrades are punished less than equivalent gains are being rewarded. This outcome seems to contradict the prospect theory of Kahneman and Tversky (1979), in which investors place more weight on financial losses. By testing the second hypothesis, I will test the validity of the ESG valuation curve of Giese and Nagy (2018).

Hypothesis (3)

H_0: The ESG valuation curve is not concave for gains in sustainability, with stable ESG scores as the reference point.

H_a: The ESG valuation curve is concave for gains in sustainability, with stable ESG scores as the reference point.

Hypothesis (4)

H_0: The ESG valuation curve is not convex for losses in sustainability, with stable ESG scores as the reference point.

H_a: The ESG valuation curve is convex for losses in sustainability, with stable ESG scores as the reference point.

By testing the third and fourth hypothesis, I will go a step further in providing more clarity about how changes in sustainability are perceived by the investors. The prospect theory of Kahneman and Tversky (1979) shows that an individual's utility function is convex for financial losses and concave for financial gains, which means that every extra gain or loss has less effect on its utility. Do investors react the same with regard to sustainable returns?

3. Data and methodology

3.1. Data on sustainability

The main variable of interest in this study is ESG Momentum. ESG Momentum is defined as the year-on-year change of ESG scores and has historically shown to be uncorrelated with traditional factors like size, book-to-market values, and excess returns (See Appendix 1).

This study uses ESG data on companies in developed markets from the beginning of 2008 until the end of 2018. The ESG data sample used is retrieved out of the Refinitiv Asset4 database (Thomson Reuters). The Refinitiv database is one of the most comprehensive databases in the field of ESG ratings (Drempetic, Klein & Zwergel, 2019). Asset4 covers more than 80% of the global market capitalization and contains data on more than 10,000 companies across 76 countries. Another database that is frequently used among scholars and investors is the MSCI ESG database. Since I want to test the extent to which the ESG valuation curve of the MSCI study by Giese and Nagy (2018) holds, I have chosen to use the Refinitiv database. Besides, recent concerns about the methodology of MSCI ESG ratings are causing the Asset4 database to rise in popularity. Bouten et al. (2017) explain these concerns, noting the fact that MSCI fails to make a valid distinction between industries when assigning weights to ESG factors. In contrast, Asset4 customizes ESG metrics per firm, by which the rating agency accounts for industry differences. Moreover, Refinitiv ESG scores have a range from 0-100, with higher scores reflecting better sustainability performances. Potential changes in ESG ratings can therefore be investigated quite precisely, which is an advantage of Asset4 in the light of this research. I have used Refinitiv's overall company ESG scores to calculate ESG Momentum. The overall scores are based on self-reported information on the environmental, social, and corporate governance pillars.

As a next step, I narrow the data sample down to companies from developed countries: *Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Hong Kong, Ireland, Israel, Italy, Japan, Luxembourg, Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, the UK,* and *the USA*. Companies from emerging markets are left out because of the scarcity of ESG data from emerging countries before 2012. In addition to that, the ESG valuation curve of Giese and Nagy (2018) to which I respond with this study is based on developed countries only. Furthermore, I exclude companies from the financial industry (SIC codes 6,000-6,999) and the utilities industry (SIC codes 4,900-4,999). These industries are subject to exceptional regulations, which makes companies in these industries less comparable to other firms. After narrowing down, the data sample consists out of 6940 equities.

3.2. Data on valuation

Next, I cross reference these 6940 equities with data on the other variables used in the analysis. Year-on-year changes of stock prices are used to measure how a company is valued by investors. Expected excess returns (the dependent variable in the regression) are calculated by subtracting a risk-free rate (assumed to be 1%) from the year-on-year changes of equity prices. Yearly equity prices from 2009-2019 are used, by which my model allows for a time lag of one year between ESG score changes and the price reaction of the market. This is in line with Giese and Nagy (2018), who prove a one-year horizon to be the optimal time frame for stable and robust performance results of ESG Momentum. The MSCI study shows that a minimum of one year is required for the market to price new ESG information, but that the price signal becomes weaker beyond a one-year horizon.

3.3. Other factors

Other independent variables are included as control variables. Based on the three-factor model of Fama and French (1992), I add three controlling factors representing the size of firms (SMB), book-to-market values (HML), and the excess return on the market (Rm-Rf). The factors for developed markets are collected from *Kenneth R. French's data library* (2021). By including small minus big (SMB) as a variable, the model controls for the size effect referring to the phenomenon that small stocks on average outperform large stocks over time. With the inclusion of high minus low (HML), the model accounts for the spread in returns between value stocks and growth stocks. Excess market returns (Rm-Rf) are included in the model by subtracting the one-month U.S. Treasury bill rate from the value-weighted return on all NYSE, AMEX and NASDAQ stocks (from CRSP).

3.4. Construction of the final dataset

To examine the performance effect of upgrades and downgrades of ESG scores, I subdivide the variable ESG Momentum into several groups of upgrades and downgrades, which I include in the equations as dummy variables. An elaboration on these dummy variables can be found in the next chapter, in which the methodology of this study is described. An overview of all the used variables and their sources can be found in Appendix 4. Furthermore, outliers (1^{st} and 99^{th} percentile) and all individual observations for which for one of the variables in the model there was no data available are dropped. I also narrow the sample down so that the variable ESG Momentum only includes scores between minus 50% (ESG score downgrades) and 50% (ESG score upgrades). I do this to reduce problems around the skewed distribution of the variable ESG Momentum, which is a result of the fact that there are many more outliers on the ESG upgrade side than there are on the downgrade side at the time. After performing these operations, the final dataset consists of 19,336 observations, of which 8,235 ESG downgrades and 11,077 ESG upgrades.

3.5. Methodology

To understand how changes in ESG scores are perceived by investors, excess stock returns of 6940 equities throughout the years 2009-2019 are assessed by using the three-factor model introduced by Fama and French (1992):

(1) $$R_i - R_f = a_i + \beta_1 \times (R_m - R_f) + \beta_2 \times SMB + \beta_3 \times HML + e_i$$

The three-factor model is an extended version of the capital asset pricing model (CAPM), in which expected excess returns $(R_i - R_f)$ are explained not only by the excess market return, but also by firm size and the book-to-market ratio. In the equation above, $R_m - R_f$ is the excess market return calculated by subtracting the risk-free return (R_f) from the return of the market portfolio (R_m). SMB measures the historic excess returns of small cap companies over big cap companies, and thus controls for firm size. The HML factor controls for the outperformance of value stocks (high book-to-market ratios) compared to growth stocks (low book-to-market ratios). The three factor-model is used in this research since it has higher explanatory power than the CAPM and is more useful in practice (Fama & French, 1992; Griffin, 2002).

First, the following equation is used to determine the relationship between ESG Momentum and expected excess stock returns:

(2) $$R_{i,t+1} - R_{f,t+1} = a_i + \beta_1 \times (R_{m,t+1} - R_{f,t+1}) + \beta_2 \times SMB_{t+1} + \beta_3 \times HML_{t+1} + \beta_M \times M_t + e_{i,t+1}$$

with M_t representing ESG Momentum at time *t* and $R_{i,t+1} - R_{f,t+1}$ being the expected excess stock returns at time *t+1*. Following Giese and Nagy (2018), the model allows for a one-year time lag between ESG score changes and the price reaction of the market. Based on the previous literature, I expect that $\beta_M > 0$, indicating a positive relationship between ESG score changes and expected excess stock returns. A positive relationship means that ESG upgrades are being rewarded by investors and downgrades are being punished. Via OLS-regression, the beta coefficients of the model above will be estimated to test the first hypothesis.

To test if the performance effect of ESG upgrades and downgrades differs, individual historical data points are divided into two groups: Upgrades and Downgrades. These groups are included in the model as dummy variables. For example: If company (Company A) had an ESG upgrade of 40% over 2008-2009 (Period 1), but its ESG score decreased by 30% over 2009-2010 (Period 2), *Company A/Period 1* is included in the "Upgrades" group and *Company A/Period 2* is included in the "Downgrades" group. The dummy variables representing each group are the main independent variables of interest. Upgrades (U) is a dummy variable that equals 1 when a company falls within the group of ESG score upgrades above 5% or 0 when it does not. The dummy variable *D* stands for Downgrades and equals 1 when a company falls withing the group of ESG downgrades lower than 5% or 0 otherwise. By taking ESG score changes around 0% as a reference point, the expectation is to find a more significant difference.

When including the ESG Momentum dummies in equation (1), the model obtained is:

(3) $$R_{i,t+1} - R_{f,t+1} = a_i + \beta_1 \times \left(R_{m,t+1} - R_{f,t+1}\right) + \beta_2 \times SMB_{t+1} + \beta_3 \times HML_{t+1} + \beta_U \times U_t + \beta_D \times D_t + e_{i,t+1}$$

with U_t and D_t being dummy variables representing groups of ESG score upgrades and downgrades at time *t* and $R_{i,t+1} - R_{f,t+1}$ being the expected excess stock returns at time *t+1*. I expect that $\beta_D < -\beta_U$, meaning that investors put less weight on ESG downgrades than they do on equivalent upgrades. I also expect that $\beta_D < 0$ and $\beta_U > 0$, in line with my expectations for the estimated coefficients of equation (2). An additional Wald must be conducted to assess whether the difference between the beta coefficient of the upgrades group and the coefficient of the downgrades group is significant. The following hypothesis is tested in the Wald test: H_0: $\beta_D + \beta_U = 0$ and H_a: $\beta_D + \beta_U > 0$.

In the next regression model, I further subdivide the group of highest upgrades and downgrades. My initial plan was to create dummy variables based on the percentiles of ESG momentum as can be seen in Appendix 5. However, as the distribution of the variable ESG Momentum is slightly skewed, I split the companies within *U* and *D* into smaller groups of ESG score changes. By this, four groups are created with different levels of ESG Momentum: *UH*, *UL*, *DL*, *DH*. In detail, Upgrades High (*UH)* represents the group of high upgrades between 10% and 50%; Upgrades Low (*UL)* represents the group of lower upgrades between 5% and

10%. The other way around, *DH* represents the group of downgrades between 10% and 50%; *DL* represents the group of downgrades between 5% and 10%. The used percentages are based on the distribution the variable ESG Momentum (See Appendix 5).

When including the four dummy variables, I obtain the following equation:

(4) $$R_{i,t+1} - R_{f,t+1} = a_i + \beta_1 \times (R_{m,t+1} - R_{f,t+1}) + \beta_2 \times SMB_{t+1} + \beta_3 \times HML_{t+1} + \beta_4 \times UH_t + \beta_5 \times UL_t + \beta_6 \times DH_t + \beta_7 \times DL_t + e_{i,t+1}$$

with UH_t, UL_t, DH_t and DL_t being dummy variables representing groups of ESG score upgrades and downgrades at time *t* and $R_{i,t+1} - R_{f,t+1}$ being the expected excess stock returns at time *t+1*. My expectation is that $\beta_4 > \beta_5$ and $\beta_6 < \beta_7$. Moreover, I expect that $\beta_6 < -\beta_4$ and $\beta_7 < -\beta_5$, which is consistent with the second hypothesis. By means of OLS regression, the beta coefficients of this model will be estimated. The beta coefficients show how groups of ESG score upgrades and downgrades are valued by investors. Significant differences between the coefficients will be confirmed with a Wald test.

For a better understanding of the curvature, I subdivide the groups once more into eight groups of upgrades and downgrades: *UHH*, *UHL*, *ULH*, *ULL*, *DLL*, *DLH*, *DHL*, *DHH*. Again, the groups are split by percentages of ESG Momentum. *UHH* represents the group of upgrades between 15% and 50%; *UHL* represents the group of upgrades between 10% and 15%; *ULH* represents the group of upgrades between 7.5% and 10%; *ULL* represents the group of upgrades between 5% and 7.5%. The groups representing downgrades are constructed in the same manner. An overview of all groups can be found in Appendix 6.

The eight groups are included as dummy variables in the last equation:

(5) $$R_{i,t+1} - R_{f,t+1} = a_i + \beta_1 \times (R_{m,t+1} - R_{f,t+1}) + \beta_2 \times SMB_{t+1} + \beta_3 \times HML_{t+1} + \beta_4 \times UHH_t + \beta_5 \times UHL_t + \beta_6 \times ULH_t + \beta_7 \times ULL_t + \beta_8 \times DHH_t + \beta_9 \times DHL_t + \beta_{10} \times DLH_t + \beta_{11} \times DLL_t + e_{i,t+1}$$

The estimated beta coefficients via OLS regression will be used to plot a graph that shows how ESG upgrades and downgrades are historically priced by the market. When applying Kahneman and Tversky's (1979) prospect theory on the concept of sustainable returns, I expect the ESG valuation curve to be convex for losses and concave for gains. In other words, relative to the group of "Stable" ESG Momentum, my expectation is that investors place less weight on every extra gain or loss in ESG Momentum.

3.6. Descriptive statistics

When examining the excess stock returns of the 6940 companies included in the final dataset, Appendix 7 shows that the distribution of the dependent variable is relatively normal but somewhat positively skewed (0.5730). This skewness can also be identified from the asymmetry of the data when plotting a histogram (Appendix 8). The kurtosis of the variable is 4.0390, and thus one point greater than the kurtosis of a normal distribution (3.0000). In the final dataset, the lowest excess return is minus 74.30%, whereas the highest excess return is 157.04%. The excess return of the 6940 equities over the years 2009-2019 was 8.98%, meaning that on average, the average return of the stocks in the data sample exceeded the risk-free rate by 8.98%. The descriptive statistics of the main variable of interest, ESG Momentum, are presented in Appendix 9. As can be derived from the table, the distribution of this variable is skewed to the right by 0.6490. The method by which ESG Momentum is included in the equations, namely via dummy variables representing groups of ESG Momentum based on actual percentage changes, prevents the skewness to have impact on the validity of the results (See Appendix 5 and 6). The variable ESG Momentum has a kurtosis of 4.4039, which is about one and a half point greater than the kurtosis of a normal distribution. In Appendix 10, this is illustrated by the sharpness of the peak in the distribution.

The final data sample consists out of companies within 539 different industries (SIC 4-digit codes). More generally, eight different industry divisions can be distinguished on which descriptive statistics can be found in Appendix 11. The highest frequency of observations (8,879) was in the *Manufacturing industry* (SIC codes 2,000-3,999). Regarding sustainable returns, the industries *Wholesale trade* (SIC codes 5,000-5,199) and *Services* (SIC codes 7,000-8,999) performed above the weighted average ESG Momentum of 3.3% over the years 2008-2018. The industry *Construction* (SIC codes 1,500-1,799) performed equal to the weighted average of sustainable returns; all other industries underperformed.

4. Results

4.1. Main results

First, the described data on developed markets is used to determine the relationship between ESG Momentum and expected excess stock returns. The results of regressing equation (2) can be found in Appendix 12. As expected, the regression output shows a positive coefficient of the variable ESG Momentum (0.092). Null hypothesis (1) can thus be rejected at a 99% confidence level. Based on the data sample, every unit of increase in ESG Momentum has a 9.2% positive effect on the expected excess stock returns.

To examine my second hypothesis that ESG score downgrades are less heavily punished by investors than equivalent gains are being rewarded, equation (3) is estimated by OLS regression and the results are reported in Table 1.

Table 1

Results of estimating equation (3) using data on developed markets.

(3) $R_{i,t+1} - R_{f,t+1} = a_i + \beta_1 \times (R_{m,t+1} - R_{f,t+1}) + \beta_2 \times SMB_{t+1} + \beta_3 \times HML_{t+1} + \beta_U \times U_t + \beta_D \times D_t + e_{i,t+1}$

Eq. (2) estimated with OLS	Coefficient	P-value	Std. error
a_i	0.003	(0.380)	0.004
$R_{m,t+1} - R_{f,t+1}$	0.109***	(0.000)	0.002
SMB_{t+1}	0.071***	(0.000)	0.006
HML_{t+1}	0.017***	(0.000)	0.004
$\boldsymbol{U_t}$	**0.015*** **	(0.003)	0.005
$\boldsymbol{D_t}$	**-0.017*** **	(0.005)	0.006
Nb. observations	19,336		
R^2	0.263		
F Statistic	1381.34		

*Significant at 10% level **Significant at 5% level *** Significant at 1% level

According to the regression output presented in Table 1, the upgrade factor (U_t) has a beta coefficient of 0.015. With a p-value of 0.003, this is significantly different than zero at a 99% confidence level. Furthermore, the coefficient of the downgrade factor (D_t) is -0.017, which is significantly different than zero with a one-sided p-value of 0.005.

Now that the coefficients of the two groups are estimated, the question is whether the sum of the two coefficients significantly differs from zero. The main expectation of this paper is that ESG score downgrades are punished more heavily than equivalent gains are being rewarded, $\beta_D < -\beta_U$ (or $\beta_D + \beta_U > 0$). However, the estimated coefficients presented in Table 1 point to the opposite of my expectation since the estimated beta coefficient of the downgrades group is more negative than the beta coefficient of the upgrades group is positive (0.017 > 0.015). To prove a difference in performance effect, a Wald test is needed to confirm that either $\beta_D < -\beta_U$ (or $\beta_D + \beta_U > 0$) or $\beta_D > -\beta_U$ (or $\beta_D + \beta_U < 0$).

Table 2

Results of Wald test for the difference between the estimated coefficients of equation (3)

H0: $\beta_D + \beta_U = 0$		
	F Statistic	0.03
Ha: $\beta_D + \beta_U > 0$	P-value (one-sided)	0.7829
Ha: $\beta_D + \beta_U < 0$	P-value (one-sided)	0.2171

The results of the Wald test are presented in Table 2. The main hypothesis of this paper, that ESG score downgrades are punished less than equivalent upgrades, can be tested by means of the p-value presented in the second row of the table. The null hypothesis of the performed Wald test is $\beta_D + \beta_U = 0$ and the one-sided alternative hypothesis is $\beta_D + \beta_U > 0$. Based on the p-value of 0.7829, the null hypothesis of the Wald test cannot be rejected. The sample does not provide enough evidence to confirm my expectation that investors place less weight on ESG score downgrades than on equivalent upgrades. The regression output reported in Table 1 indicates that the opposite might be true since 0.017 > 0.015. However, the results of the Wald test do not confirm this as the one-sided p-value for the left tailed test is 0.2171. Thus, the 0.2% difference in weight between the estimated beta coefficients of U_t and D_t is not statistically significant.

Table 3

Results of estimating equation (4) using data on developed markets.

(4) $R_{i,t+1} - R_{f,t+1} = a_i + \beta_1 \times (R_{m,t+1} - R_{f,t+1}) + \beta_2 \times SMB_{t+1} + \beta_3 \times HML_{t+1} + \beta_{UH} \times UH_t + \beta_{UL} \times UL_t + \beta_{DH} \times DH_t + \beta_{DL} \times DL_t + e_{i,t+1}$

Eq. (4) estimated with OLS	Coefficient	P-value	Std. error
a_i	0.003	(0.375)	0.004
$R_{m,t+1} - R_{f,t+1}$	0.109***	(0.000)	0.002
SMB_{t+1}	0.071***	(0.000)	0.006
HML_{t+1}	0.017***	(0.000)	0.004
$\boldsymbol{UH_t}$	**0.012****	(0.049)	0.006
$\boldsymbol{UL_t}$	**0.022*****	(0.002)	0.007
$\boldsymbol{DH_t}$	**-0.025*****	(0.001)	0.008
$\boldsymbol{DL_t}$	**-0.009**	(0.186)	0.007
Nb. observations	19,336		
R^2	0.263		
F Statistic	987.4		
*Significant at 10% level **Significant at 5% level *** Significant at 1% level			

The OLS estimation results of regression equations (4) and (5) are reported in Table 3 and 5. When considering the model estimates of equation (4), the estimate of β_{UH} is 0.012 and the estimate of β_{UL} is 0.022. Both estimates are significantly different from zero. The estimates of β_{DH} and β_{DL} are -0.025 and -0.009, respectively. However, the estimated beta coefficient of *DL* is the only test result that is not significant. Table 3 thus implies that my first expectation that $\beta_{DH} < \beta_{DL}$ is correct. My second expectation that $\beta_{UH} > \beta_{UL}$ does not follow from the table. In other words, it seems that within the group of downgrades, high downgrades are indeed punished more severely (β_{DH} = -0.025) than low downgrades (β_{DL} = -0.009). But for ESG score upgrades the opposite seems to be true. The performed Wald tests as reported in Appendices 13 and 14 confirm that $\beta_{DH} < \beta_{DL}$ and $\beta_{UH} < \beta_{UL}$.

An explanation for the fact that $\beta_{UH} < \beta_{UL}$ could be that high upgrades are more common for companies that previously performed poorly regarding sustainability issues. When such companies experience a high upgrade, the performance effect of it could be smaller since it might take time to improve their bad image regarding sustainability. In relative terms, an ESG upgrade may be significant while in absolute terms the new ESG score is still low. This is especially true for low companies with low absolute ESG scores. This explanation is in line with Giese and Nagy (2018), who found that the option value for companies that initially score low on ESG concerns materializes very slowly.

Table 4

Results of Wald test for the difference between the estimated coefficients of equation (4)

H0: $\beta_{DH} + \beta_{UH} = 0$		
	F Statistic	1.45
Ha: $\beta_{DH} + \beta_{UH} > 0$	P-value (one-sided)	0.9421
Ha: $\beta_{DH} + \beta_{UH} < 0$	P-value (one-sided)	0.0571

The further distinction in groups of upgrades and downgrades as in equation (3) allows to test hypothesis (2) of this paper again, but with a focus on the more extreme ESG score changes. When taking the group of highest downgrades and upgrades (*DH* and *UH*) into consideration, the results of the Wald test reported in the second row of Table 4 show that my initial expectation cannot be confirmed based on this data sample. It cannot be said that investors place less weight on ESG score downgrades than on equivalent upgrades, since the p-value of this test is 0.9421. However, the null hypothesis can be rejected when testing for the alternative hypothesis that investors place more weight on ESG downgrades. This is in line with the results in both Table 1 and 2. With a one-sided p-value of 0.0571, the sum of β_{DH} and β_{UH} significantly differs from zero when performing a left tailed test. Therefore, the data sample provides evidence that $\beta_{DH} + \beta_{UH} < 0$. Thus, at least for high ESG score changes (>10%), investors place more weight on ESG downgrades than on equivalent upgrades. For low ESG score changes, it cannot be said that the sum of β_{DL} and β_{UL} significantly differs from zero (See Appendix 15).

Based on the estimated coefficients of equation (3) and (4), hypothesis (2) cannot be rejected. This implies that the findings of Giese and Nagy (2018) cannot be confirmed based on this data sample. Part of the reason why the results are at odds with the MSCI study of Giese and Nagy (2018) may lie in the database that is used. I will therefore test the robustness of the results by using MSCI ESG data in the next chapter. The Refinitiv ESG Database and the MSCI ESG Database differ in terms of the methodology by which ESG ratings are calculated. Kotsantonis and Serafeim (2019) show that such differences may significantly impact the results.

Table 5

Results of estimating equation (5) using data on developed markets.

(5) $$R_{i,t+1} - R_{f,t+1} = a_i + \beta_1 \times \left(R_{m,t+1} - R_{f,t+1}\right) + \beta_2 \times SMB_{t+1} + \beta_3 \times HML_{t+1} + \beta_{UHH} \times UHH_t + \beta_{UHL} \times UHL_t + \beta_{ULH} \times ULH_t + \beta_{ULL} \times ULL_t + \beta_{DHH} \times DHH_t + \beta_{DHL} \times DHL_t + \beta_{DLH} \times DLH_t + \beta_{DLL} \times DLL_t + e_{i,t+1}$$

Eq. (3) estimated with OLS	Coefficient	P-value	Std. error
a_i	0.003	(0.374)	0.004
$R_{m,t+1} - R_{f,t+1}$	0.109***	(0.000)	0.002
SMB_{t+1}	0.071***	(0.000)	0.006
HML_{t+1}	0.017***	(0.000)	0.004
$\boldsymbol{UHH_t}$	**0.010**	(0.148)	0.007
$\boldsymbol{UHL_t}$	**0.015***	(0.090)	0.009
$\boldsymbol{ULH_t}$	**0.020****	(0.042)	0.010
$\boldsymbol{ULL_t}$	**0.023*****	(0.010)	0.009
$\boldsymbol{DHH_t}$	**-0.028*****	(0.009)	0.011
$\boldsymbol{DHL_t}$	**-0.022****	(0.030)	0.010
$\boldsymbol{DLH_t}$	**-0.012**	(0.238)	0.010
$\boldsymbol{DLL_t}$	**-0.008**	(0.405)	0.009
Nb. observations	19,336		
R^2	0.263		
F Statistic	628.3		
*Significant at 10% level **Significant at 5% level *** Significant at 1% level			

Table 5 presents the results of estimating equation (5). Using the estimated coefficients of the three tables above, the results suggest that the ESG valuation curve is non-linear for developed markets. This non-linearity is in line with my expectation and can be seen when plotting a graph using interpolation (See Appendix 16). The estimated coefficients of equation (5) are used to predict the value of the points in the graph that are unknown. The interpolated graph gives two new insights on the valuation of ESG issues by investors in addition to the results of Table 1 and 3.

First, the curve in Appendix 16 illustrates that for high downgrades of at least 10%, the performance effect seems to be stronger than for equivalent high upgrades. For example, a 30% downgrade in ESG score is negatively valued by the market by 2.5%, while a 30% upgrade is positively valued around 1%. However, this is different for the group of small upgrades and downgrades within a range of 0%-10%. Within this range of ESG momentum, upgrades are appreciated more relative to downgrades. The difference in performance effect can also be derived from the estimated values presented in Table 5, however β_{DLH} and β_{DLL} are not significant.

Second, hypothesis (3) and (4) can both be rejected based on the ESG momentum curve. The curve shows no evidence that the ESG valuation curve is concave for gains in sustainability or convex for losses. When looking at the upgrades side, the curve rises sharply at first until a positive 10% ESG momentum, but for further upgrades it keeps descending. The curve makes more sense when looking at the downgrades side, although based on the results it cannot be proven that marginal gains and losses have less effect on the difference in valuation. The prospect theory of Kahneman and Tversky (1979) can thus not be used to assess how investors value extra units of sustainable gains and losses.

4.2. Robustness test

Different methodologies of ESG rating agencies may lead to different outcomes (Berg et al., 2020. In order to not rely too heavily on the Refinitiv ESG scores, I have performed a robustness test using ESG data from the MSCI database. The results of this robustness test are presented in Appendix 17. The sample includes MSCI ESG scores from the three pillars: *Environmental, Social* and *Governance*. In the data sample, all three pillars are subdivided into categories of either ‘strengths’ or ‘concerns’. If a concern exists it is quantified by a score of 1 and 0 if no concern exists. The same applies regarding the ‘strengths’ category. Total scores are calculated by adding or subtracting the total number of strengths or concerns. With the yearly total scores, ESG momentum is calculated by taking the year-on-year change of MSCI ESG scores. The ESG company data is cross referenced with financial data from the Refinitiv Eikon database and the controlling variables from the Kenneth R. French Database.

The original data sample (with Refinitiv ESG data) consists out of 6940 companies in developed markets with ESG data over the years 2008-2018 and financial data over the years 2009-2019. The dataset for the robustness test with MSCI ESG data consists out of 1107 companies in developed markets over the years 2010-2013. In total, the dataset consists out of 3,303 observations, which is considerably less than the 19,336 observations in the original sample. Out of the 3,303 observations, 1,450 observations were ESG upgrades, and 1,411 observations were ESG downgrades. The other observations (442) represent stable ESG momentum. Since there were less observations, it was not possible to find significant results when using several smaller groups of ESG Momentum due to the small sample size of the groups. Therefore, the robustness test only covers the results regarding hypothesis (2). All observations are divided into two groups: Upgrades or Downgrades. The methodology of the robustness test is the same as the methodology of the main results described in chapter 3.5.

Appendix 17 shows the results of estimating equation (3) are different when using the MSCI database for developed markets. The performance effect of ESG score changes now seems stronger for upgrades (0.0386) than for downgrades (-0.0369). The difference is small (0.2%) but statistically significant as confirmed by the Wald test in Appendix 18. The estimations of equation (3) with MSCI ESG data are in line with findings of the MSCI study by Giese and Nagy (2018).

This robustness test underlines the concerns about the divergence in ESG data and their implications, as pointed out in several studies and discussed in chapter 2.1.

4.3. Interpretation of results

The results of estimating the equations (3), (4) and (5) can be interpreted as follows. First, the estimations of the main data sample using Refinitiv ESG scores of companies in developed markets show that investors seem to place more weight on high ESG score downgrades than on equivalent upgrades. Thereby, hypothesis (1) cannot be rejected when using this data sample. The dependent variable of this study, $R_{m,t+1} - R_{f,t+1}$, represents the expected excess returns of stocks. The positive estimated beta coefficient of the upgrades dummy variable may be explained by the positive factors associated with a strong ESG proposition, such as growth opportunities or better financing terms. These factors relate to the part of sustainable return that materializes into financial value. The negative estimated beta coefficient of the downgrades dummy variable, representing a stronger performance effect in comparison with the group of high upgrades, could possibly be explained by risk-aversity of investors towards sustainable return. Kahneman and Tversky (1979) examined the behaviour of investors towards gains and losses of financial returns and proved that investors place more weight on financial losses. Based on the results of estimating equation (4), it could be argued that risk-aversity of investors towards towards gains and losses in either sustainable or financial return is quite similar, as long as it concerns more extreme cases (>10% ESG Momentum). However, the robustness test in the previous chapter shows that the ESG rating methodology used can have a significant impact on the regression results. With Refinitiv ESG scores, more weight is put on sustainable losses instead of gains. But with the MSCI ESG data base, it seems that investors place more weight on sustainable gains, which is in line with Giese and Nagy (2018). Therefore, another finding of this study is that the choice of ESG rating agency could significantly affect results, confirming Berg et al. (2019).

An interesting outcome of estimating equation (4) is that for the group of high ESG upgrades, the performance effect is smaller than for the group of low upgrades. For example, the regression output shows that 50% ESG upgrades are valued less positively than 5% ESG upgrades. This result contradicts my expectations, since I expected the ESG valuation curve to be concave for gains in sustainability. This result, however, confirms the results of Giese and Nagy (2018), who explain that ESG upgrades have less of a financial performance effect for companies that initially score low on ESG issues. This could be an explanation for my results since it could be that high ESG upgrades are more common for initially poor scoring companies.

When making a distinction between the part of sustainable return that materializes (financial return) and the part that does not (social return), it can be argued that the low overall performance effect of the high ESG upgrade group can be explained by this distinction. If sustainable return would one-on-one translate into financial value and the prospect theory of Kahneman and Tversky (1979) would be applied, I would expect concavity in the ESG valuation curve for sustainable gains. However, the ESG valuation curve presented in Appendix 16 shows that this is not the case. If the overall performance effect of sustainable returns can be divided into a part of financial performance effect and a part of social performance effect, it can be argued that the part representing the financial performance effect is small relative to the part of social performance effect. For companies within the group of small ESG upgrades the social performance effect is strong, since the group may consist out of companies with an initial "sustainable image" due to higher previous ESG scores. The companies that fall within the group of high ESG upgrades may not experience an equally strong social performance effect since they are more likely to have a poor initial image with regard to sustainability issues. The "social premium" paid by investors for a positive change in the ESG scores of initially low scoring companies could therefore be lower, which is in line with the results of Giese and Nagy (2018).

Final, the ESG valuation curve of Appendix 16 based on estimating equation (4) shows no concavity or convexity for gains or losses of sustainability. This means that both hypothesis (2) and (3) can be rejected based on the regression output of this data sample. It cannot be said that extra gains or losses in sustainability have less effect on the excess return that investors are willing to pay. Again, this outcome contradicts with the prospect theory of Kahneman and Tversky (1979) and therefore indicates that sustainable return behaves differently compared to financial return.

5. Conclusion

5.1. Overall conclusion

This study provides several new insights concerning the valuation of ESG issues by investors. The main conclusion of this paper is that with the use of Refinitiv ESG data, the average performance effect of ESG score downgrades is stronger than the average performance effect of ESG score upgrades for ESG score changes greater than 10%. This means that investors place more weight on downgrades in sustainability than they do on equivalent downgrades. This result is in line with the prospect theory of Kahneman and Tversky (1979) and thus points out that investors face a similar risk-aversion regarding sustainable return as they do when it concerns financial return.

Moreover, the results of this study show that on average, ESG upgrades above approximately 10% have a lower performance effect than ESG upgrades below 10%. A possible explanation for this could be that companies that initially score low on ESG concerns experience higher percentual ESG upgrades than companies that already have a good ESG proposition. Therefore, the group of high ESG upgrades may mostly consist out of initially low scoring companies. This is in line with the findings of Giese and Nagy (2018), who found that the option value for initially low scoring companies is huge but materializes very slowly (See Appendix 3). The social performance effect of ESG upgrades of those companies may be less strong, since investors are willing to pay less of a premium for ESG upgrades of companies that initially score low. Further research is needed to confirm this explanation.

With regard to the expected concavity of the ESG curve for upgrades, the results of this paper do not confirm that the ESG valuation curve is concave for sustainable gains. Likewise, convexity of losses cannot be confirmed by the results. Overall, it seems that the investors face risk-aversion regarding sustainability, similar to what is argued with regard to financial return in the prospect theory of Kahneman and Tversky (1979). When examining the ESG valuation curve, however, the non-concavity of gains and the non-convexity of losses show that sustainable return is not valued on the same manner as pure financial returns are valued by investors as the prospect theory shows. My expectation is that the differences between sustainable and financial return can be explained by making a distinction between the two parts of sustainable return, namely the part that materializes into financial value and the part that does not.

A last finding of this study has to do with the differences between the methodologies of ESG rating agencies and their implications. In the robustness test, I estimated the same equation with a different data sample (MSCI ESG data instead of Refinitiv ESG data). The findings of this test were the exact opposite, which implies that the choice of ESG rating agency has a significant impact on the results of this study. This last finding is in line with many concerns scholars have on diverging ESG ratings (Kotsantonis & Serafeim, 2019; Berg et al., 2019).

My results support the view that performing bad on ESG issues is punished by investors. The finding that investors place more weight on ESG score downgrades than on equivalent upgrades is an outcome which is desired from a society point of view. It gives companies incentives to progress on sustainability issues, since upgrades are positively rewarded by investors and downgrades are punished more severely.

5.2. Limitations and further research

This study has some limitations regarding its reach and methodology. One limitation of this study is that I have solely focused on developed markets. As mentioned in the literature review, outcomes could be different when focusing on other geographies as well. It could be valuable to study the effects of ESG score changes for a wider range of companies, including companies emerging markets. It could also be interesting to study the differences between the results per geography. Kaiser (2020) argues that the adoption rate of ESG in Europe is higher, which leads to a more efficient pricing of ESG score changes by the market. This could possibly affect the significance of the results in a positive manner.

Another limitation is the measurement of how ESG score changes are valued by investors. Data from stock markets is used to measure valuation, which may paint a distorted picture. Increased attention by investors represented in stock prices are not the best indicator of long-term value creation. However, since the questions that this paper aims to answer are related to short-term valuation by investors, this limitation is not a problem as such within this study. In further research, it could be valuable to focus on how investors value the part of sustainable returns that materialize into long-term value creation. Another measurement method for the valuation by investors would possibly be needed then.

The last limitation of this paper is that at the time, the data sample based on Refinitiv ESG data consists out of much more ESG upgrades than downgrades. Therefore, the effects of ESG upgrades can on average be much better analyzed than the effects of ESG score downgrades since the number of observations are much lower in the last group. This is a problem that researchers in the field of ESG must deal with. The trend is that companies are increasingly strengthen their ESG proposition, limiting possible concerns.

To conclude, an interesting field for further research could be the following. Overall, we see that if profits of companies are higher than expected, the stock price of goes up. An interesting question to examine would be if this is also the case if ESG scores turn out higher than expected.

References

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of economic perspectives, 21*(2), 129-152.

Berg, F., Kölbel, J. F., & Rigobon, R. (2019). Aggregate confusion: The divergence of ESG ratings. *MIT Sloan School of Management.*

Bouten, L., Cho, C. H., Michelon, G., & Roberts, R. W. (2017). CSR Performance Proxies in Large-Sample Studies:'Umbrella Advocates', Construct Clarity and the 'Validity Police'. *Construct Clarity and the 'Validity Police' (August 2017)*.

Drempetic, S., Klein, C., & Zwergel, B. (2020). The influence of firm size on the ESG score: Corporate sustainability ratings under review. *Journal of Business Ethics*, *167*(2), 333 360.

Edelman (2019). An overwhelming majority of investors expect companies to implement effective environmental, social-impact and governance (ESG) practices. Retrieved from: https://www.edelman.com/news-awards/2019-institutional-investor-trust.

Fama, E. F., & French, K. R. (1992). *The cross-section of expected stock returns* (pp. 349-391). University of Chicago Press.

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

Giese, G., & Nagy, Z. (2018). How Markets Price ESG: Have Changes in ESG Scores Affected Stock Prices?. *MSCI.* Retrieved from https://www. msci. com/www/research-paper/how-markets-price-esg-have/01159646451.

Griffin, J. M. (2002). Are the Fama and French factors global or country specific?. *The Review of Financial Studies*, *15*(3), 783-803.

Henisz, W., Koller, T., & Nuttall, R. (2019). Five ways that ESG creates value.

Kahneman, D., & Tversky, A. (1979). Prospect Theory of Decisions Under Risk. *Econometrica*, *47*(2), 1156-67.

Kaplan, J., & Montiel, I. (2017). East vs. West Approaches to Reporting Corporate Sustainability Strategies to the World: Corporate Sustainability Reporting: East vs. West. In *Comparative Perspectives on Global Corporate Social Responsibility* (pp. 49-68). IGI Global.

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

KMPG (2020). Sustainable investing: Fast-forwarding its evolution. Retrieved from https://home.kpmg/xx/en/home/insights/2020/02/sustainable-investing.html.

Kotsantonis, S., & Serafeim, G. (2019). Four things no one will tell you about ESG data. Journal of Applied Corporate Finance, 31(2), 50-58.

McKinsey & Company Inc., Koller, T., Goedhart, M., & Wessels, D. (2020). Valuation: Measuring and Managing the Value of Companies *(Wiley Finance*) *(7th ed.)*. Wiley.

Miremadi, M., Musso, C., & Weihe, U. (2012). How much will consumers pay to go green? *McKinsey Quarterly*, *4*, 14-14.

Nehrt, C. (1996). Timing and intensity effects of environmental investments. *Strategic Management Journal*, *17*(7), 535-547.

Piñeiro-Chousa, J., López-Cabarcos, M. Á., Caby, J., & Šević, A. (2021). The influence of investor sentiment on the green bond market. *Technological Forecasting and Social Change*, *162*, 120351.

Revelli, C., & Viviani, J. L. (2015). Financial performance of socially responsible investing (SRI): what have we learned? A meta‐analysis. *Business Ethics: A European Review*, *24*(2), 158-185.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, *5*(4), 297-323.

Yoon, B., Lee, J. H., & Byun, R. (2018). Does ESG performance enhance firm value? Evidence from Korea. *Sustainability*, *10*(10), 3635.

Zhong, S., & Chen, J. (2019). How environmental beliefs affect consumer willingness to pay for the greenness premium of low-carbon agricultural products in china: Theoretical model and survey-based evidence.

Appendix

Appendix 1

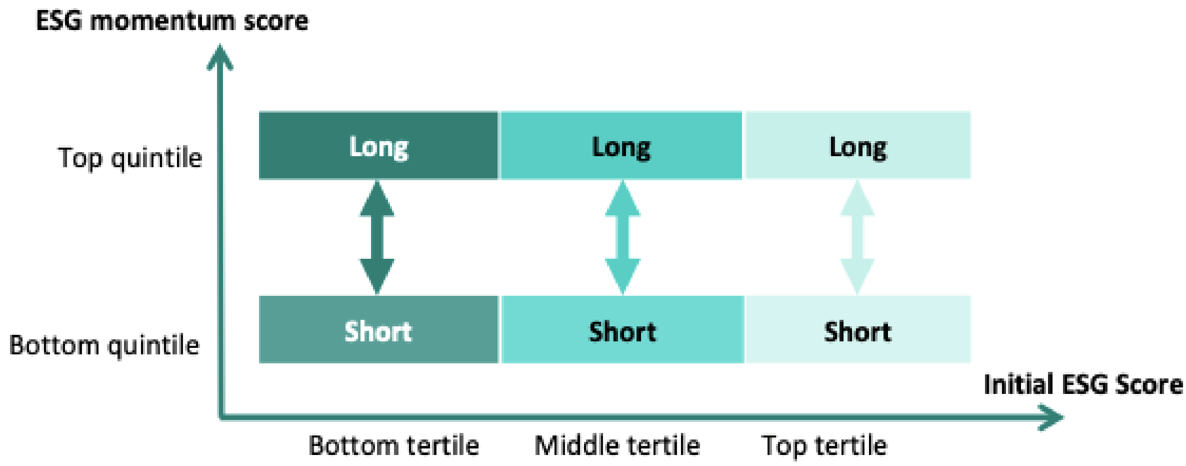
Correlation of equity style factors with ESG and ESG Momentum.

Factor	Correlation with ESG	Correlation with ESG Momentum
Mid Capitalization	-0.14	0.02
Earnings Variability	-0.13	-0.03
Residual Volatility	-0.09	-0.04
Book-to-Price Ratio	-0.08	-0.03
Liquidity	-0.04	-0.01
Leverage	-0.03	-0.01
Growth	-0.02	0.00
Beta	-0.02	-0.01
Earnings Yield	0.00	0.01
Momentum	0.00	0.02
Earnings Quality	0.02	-0.04
Long-Term Reversal	0.05	-0.02
Profitability	0.05	0.02
Dividend Yield	0.08	0.01
Investment Quality	0.09	0.00
Size	0.14	-0.02

Source: Giese and Navy (2018), p.8

Appendix 2

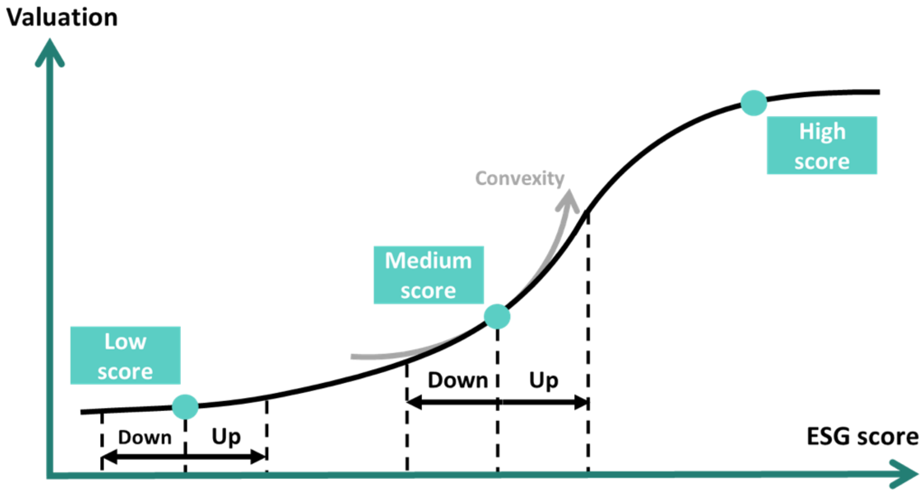
Calculation historical performance of top versus bottom ESG Momentum score.



Source: Giese and Navy (2018), p.15

Appendix 3

Non-linear stylized ESG valuation curve.



Source: Giese and Navy (2018), p.16

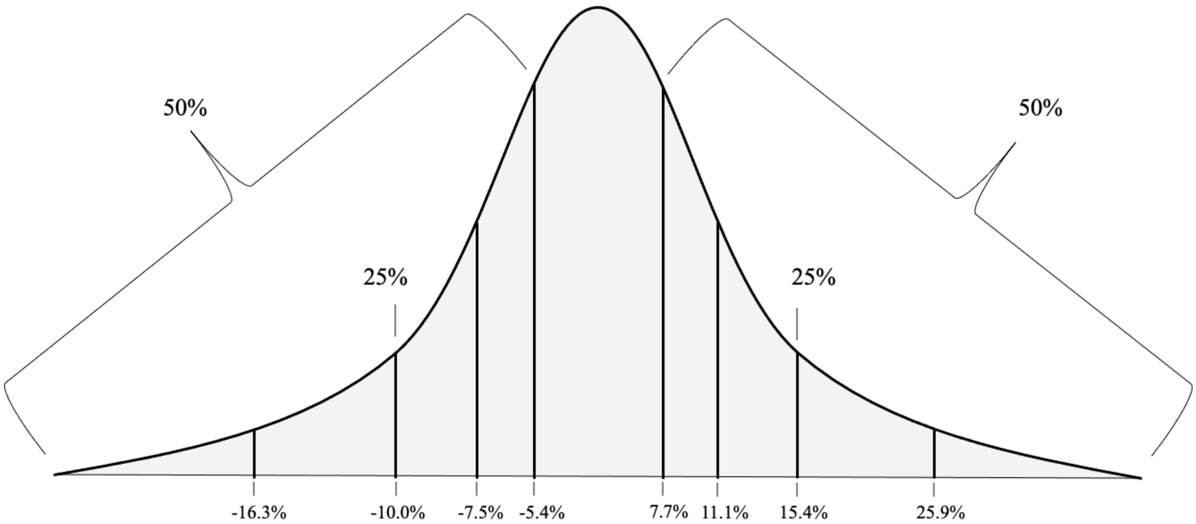
Appendix 4

Description of the variables used.

Variable name	Note	Unit	Source
ESG Momentum *(M)*	Year-on-year changes of Refinitiv ESG Scores from 2008-2018; Refinitiv ESG Scores are overall company scores based on self-reported information in the environmental, social, and corporate governance pillars	%	Refinitiv ESG Database
Stock price changes (Ri)	Year-on-year stock price changes from 2009-2019	%	Refinitiv Eikon
Excess return on the market (Rm-Rf)	The return on a developed market region's value-weight market portfolio minus the U.S. one month treasury bill rate	$	Kenneth R. French
Firm size (SMB)	SMB is the equal-weight average of the returns on the three small stock portfolios for developed markets minus the average of the returns on the three bi stock portfolios	$	Kenneth R. French
Book-to-market values *(HML)*	HML is the equal-weight average of the returns for the two high B/M portfolios for a region minus the average of the returns for the two low B/M portfolios	$	Kenneth R. French

Appendix 5

Visual overview of the initial concept to create dummy variables representing each group of ESG Momentum based on percentiles of distribution.



Note: I have decided not to use the method shown above since the results would be harder to interpret. As can be seen, the highest 12.5% upgrades start at 25.9%, while the highest 12.5% downgrades start at -16.3% due to the skewed distribution of the variable ESG Momentum.

Appendix 6

Overview of final dummy variables representing each group of ESG Momentum.

Dummy variable	ESG Momentum range		Eq. (2)	Eq. (3)	Eq. (4)
	Min	Max			
U	5.0%	50.0%	x		
D	-50.0%	-5.0%	x		
UH	10.0%	50.0%		x	
UL	5.0%	10.0%		x	
DL	-10.0%	-5.0%		x	
DH	-50.0%	-10.0%		x	
UHH	15.0%	50.0%			x
UHL	10.0%	15.0%			x
ULH	7.5%	10.0%			x
ULL	5.0%	7.5%			x
DLL	-7.5%	-5.0%			x
DLH	-10.0%	-7.5%			x
DHL	-15.0%	-10.0%			x
DHH	-50.0%	-15.0%			x

Appendix 7

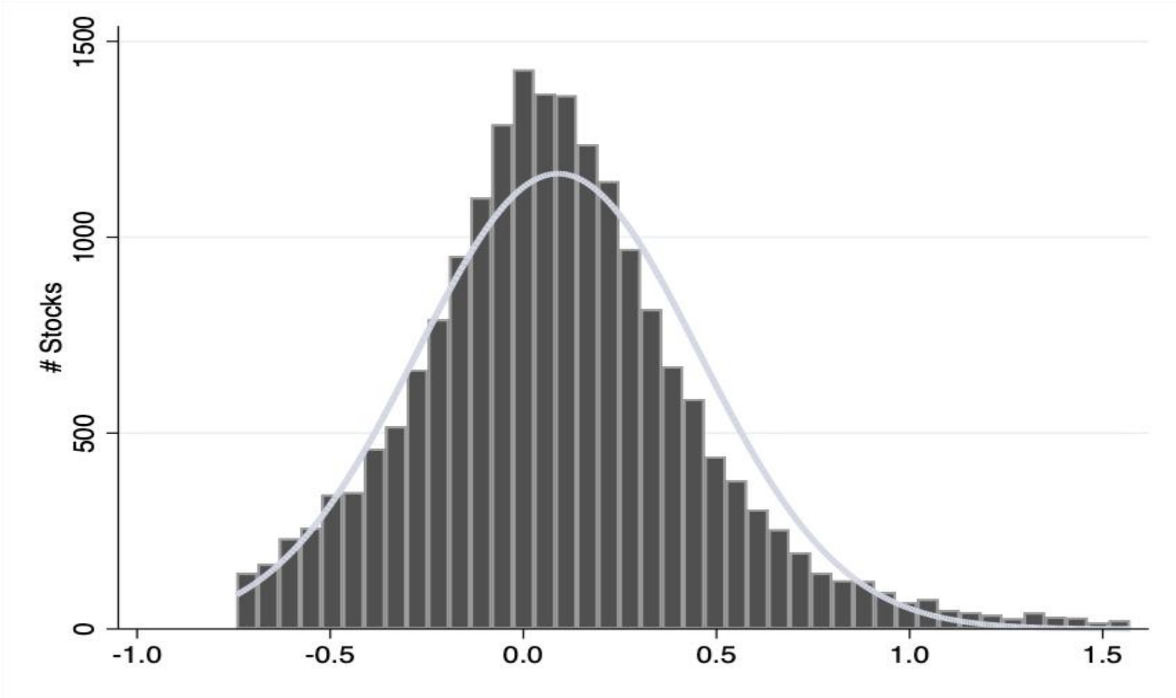
Statistics of the variable excess stock returns in the final dataset.

Mean	Min	Max	St. deviation	Skewness	Kurtosis
0.0898	-0.7430	1.5704	0.3656	0.5730	4.0390

Data source: Refintiv Eikon Database

Appendix 8

Distribution of the variable excess stock returns in the final dataset.



Data source: Refinitiv Eikon Database

Appendix 9

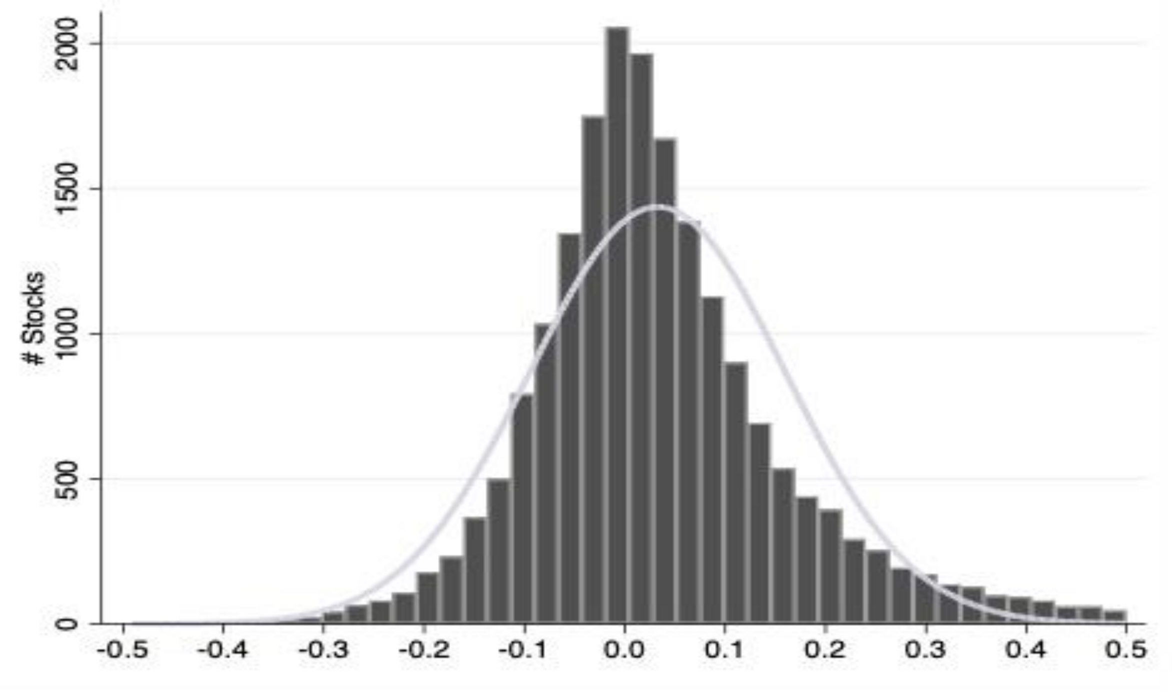
Statistics of the variable ESG Momentum in the final dataset.

Mean	Min	Max	St. deviation	Skewness	Kurtosis
0.0326	-0.4907	0.4992	0.1267	0.6490	4.4039

Data source: Refintiv Eikon Database

Appendix 10

Distribution of the variable ESG Momentum Scores in the final dataset.



Data source: Refinitiv Eikon Database

Appendix 11

Overview of the different industries included.

Industry	SIC-codes	N	Mean M	Mean Ri-Rf
Agriculture, forestry, and fishing	100-900	78	3.1%	10.0%
Mining	1,000-1,499	1,536	2.6%	1.0%
Construction	1,500-1,799	891	3.3%	10.1%
Manufacturing	2,000-3,999	8,879	3.2%	11.7%
Transportation	4,000-4,899	2,115	2.4%	6.5%
Wholesale trade	5,000-5,199	733	4.4%	9.7%
Retail trade	5,200-5,999	1,831	2.7%	9.1%
Services	7,000-8,999	3,271	4.2%	13.1%

Data source: Refintiv Eikon Database

Appendix 12

Results of estimating equation (2) using data on developed markets.

(2) $$R_{i,t+1} - R_{f,t+1} = a_i + \beta_1 \times (R_{m,t+1} - R_{f,t+1}) + \beta_2 \times SMB_{t+1} + \beta_3 \times HML_{t+1} + \beta_M \times M_t + e_{i,t+1}$$

Eq. (2) estimated with OLS	Coefficient	P-value	Std. error
a_i	0.002	(0.201)	0.003
$R_{m,t+1} - R_{f,t+1}$	0.109***	(0.000)	0.002
SMB_{t+1}	0.070***	(0.000)	0.006
HML_{t+1}	0.018***	(0.000)	0.004
$\boldsymbol{M_t}$	**0.092*** **	(0.000)	0.018
Nb. observations	19,336		
R^2	0.263		
F Statistic	1725.95		

*Significant at 10% level **Significant at 5% level *** Significant at 1% level

Appendix 13

Results of Wald test for the difference between the estimated coefficients β_{UH} and β_{UL}.

H0: $\beta_{UH} - \beta_{UL} = 0$		
	F Statistic	1.69
Ha: $\beta_{UH} > \beta_{UL}$	P-value (one-sided)	0.9030
Ha: $\beta_{UH} < \beta_{UL}$	P-value (one-sided)	0.0970

Appendix 14

Results of Wald test for the difference between the estimated coefficients β_{DH} and β_{DL}.

H0: $\beta_{DH} - \beta_{DL} = 0$		
	F Statistic	1.69
Ha: $\beta_{DH} > \beta_{DL}$	P-value (one-sided)	0.9423
Ha: $\beta_{DH} < \beta_{DL}$	P-value (one-sided)	0.0289

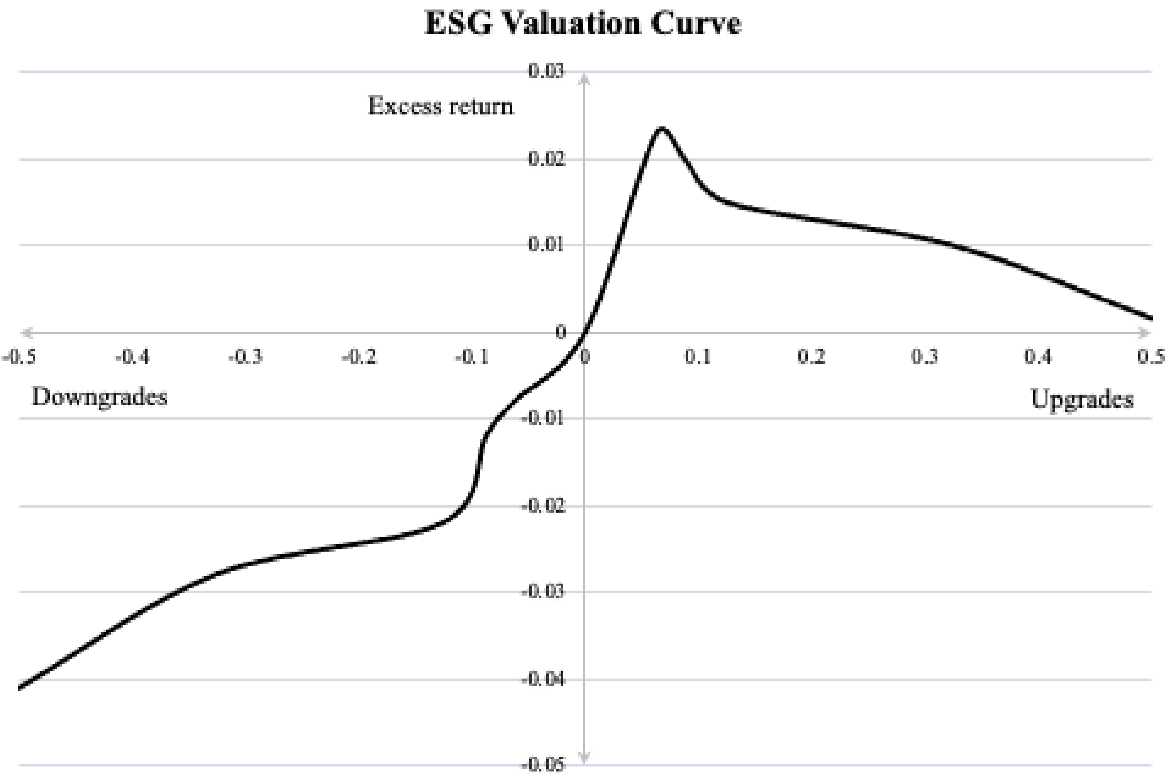
Appendix 15

Results of Wald test for the difference between the sum of estimated coefficients β_{DL} and β_{UL}.

H0: $\beta_{DL} + \beta_{UL} = 0$		
	F Statistic	1.05
Ha: $\beta_{DL} + \beta_{UL} > 0$	P-value (one-sided)	0.5761
Ha: $\beta_{DL} + \beta_{UL} < 0$	P-value (one-sided)	0.4239

Appendix 16

Interpolated ESG valuation curve.



Data source: Refinitiv Eikon Database

Appendix 17

Results robustness test estimating equation (3) using the MSCI ESG Database.

(3) $$R_{i,t+1} - R_{f,t+1} = a_i + \beta_1 \times (R_{m,t+1} - R_{f,t+1}) + \beta_2 \times SMB_{t+1} + \beta_3 \times HML_{t+1} + \beta_U \times U_t + \beta_D \times D_t + e_{i,t+1}$$

Eq. (2) estimated with OLS	Coefficient	P-value	Std. error
a_i	-0.0139	(0.203)	0.016
$R_{m,t+1} - R_{f,t+1}$	0.0138***	(0.000)	0.001
SMB_{t+1}	0.0173***	(0.000)	0.001
HML_{t+1}	0.011***	(0.000)	0.002
$\boldsymbol{U_t}$	**0.0386****	(0.021)	0.019
$\boldsymbol{D_t}$	**-0.0369****	(0.018)	0.018
Nb. observations	3,303		
R^2	0.215		
F Statistic	226.07		
*Significant at 10% level **Significant at 5% level *** Significant at 1% level			

Appendix 18

Results of Wald test for the difference between the sum of estimated coefficients β_U and β_D.

H0: $\beta_U + \beta_D = 0$		
	F Statistic	1.05
Ha: $\beta_U + \beta_D > 0$	P-value (one-sided)	0.0117