Master Thesis U.S.E.

Quantifying sustainability risk from stock returns using multi-factor models – A comparison of different sustainability (risk) measures[1]

Lilla Mónika Mátrai (4451708) l.m.matrai@students.uu.nl

Thesis Supervisor: PD Dr. Martin Rohleder, martin.rohleder@wiwi.uni-augsburg.de

Co-reader: Evgenia Zhivotova, e.zhivotova@uva.nl

Abstract

This paper examines how different aspects of sustainability are incorporated into stock prices and compares their relative importance from the financial markets' perspective, both in Europe and in the US. This study explores whether individual ESG issues have sufficient informational power for investors, and in particular compares whether environmental risk is more relevant for investors compared to social risk, and whether more clear-cut issues such as corporate emissions and employee relations are more important than resource use and community relations. To study the pricing effects, sustainability factors are constructed in the form of long-short portfolios, based on ESG scores and are added to the Carhart common factor model. Formal asset pricing tests are performed to test for the models' explanatory power and for the risk premia associated with the factors. Overall, results show that investors in the US require higher compensation for bearing sustainability risk, suggesting higher market efficiency. In line with expectations, environmental risk is proved to be more material for investors than social risk, in Europe requiring more than 2.3 times larger risk premium. Although, compared to expectations, on average, resource use and communities related performance are incorporated into stock prices to a higher extent than emissions and workforce, results are less significant and not consistent across tests, implying that individually they might not carry enough relevant information for investors.

JEL-codes: G12, G15, Q51, Q52

Keywords: Sustainability risk, Factor models, Asset pricing, Europe, US

---

[1] **Acknowledgments**: I would like to thank my family, particularly my parents, for supporting me and making this journey possible in the first place. I am very grateful for my supervisor, PD Dr. Martin Rohleder, for the continuous feedback that he gave me during the last couple of months. Finally, I am thankful for my peers at the University for all the support they gave me.

1. Introduction

In this thesis, I study and compare the pricing effects of different sustainability measures using multi-factor models. I explore and compare how the environmental and social performance of a company is incorporated into equity prices using sustainability scores provided by EIKON. I dive deeper into these two pillar scores and measure the relative importance of two key issues per pillars, namely *Emissions*, *Resource Use*, *Workforce* and *Communities*. For the analysis I construct zero-cost portfolios to mimic the performance of these risk measures and apply formal asset pricing multi-factor models such as the Carhart four factor model. I separate and compare European and US markets to explore in which region sustainability risk is priced to a greater extent.

In recent years, ESG investing has become mainstream, first dominated by Europe and with an exponential growth in the US. Global assets managed under sustainable strategies exceeded $35.3 trillion in 2020 after a steady increase according to the GSIA. ESG integration has been proved to be the most popular sustainable investing strategy, while sustainable-themed investing has experienced the highest growth between 2016-2020. (Global Sustainable Investment Alliance, 2021) A lot has been written and studied about ESG integration and investment returns, although there has been a less focus on the comparison between measures within ESG that could deliver returns or present higher risk for investors. A deeper understanding on the pricing of ESG constituents could help investors make decisions when balancing between sustainability and financial performance.

The pricing of sustainability risk has been studied recently to an increasing extent by adding new long-short portfolios into multi-factor models. Empirical studies have mostly focused on overall sustainability (Bennani et al., 2018) performance, on transition risk exposure in the form of corporate emissions (Huij et al., 2022; In et al., 2019), or authors construct a composite environmental risk measure (Görgen et al., 2020). Results differ: some authors found evidence for a positive risk premium associated with companies with lower carbon efficiency (Gurvich & Creamer, 2022), meaning that financial markets acknowledge sustainability risk and require higher return for bearing additional risk. On the contrary, other studies found no positive risk premium associated with sustainability risk (Gregory et al., 2021) or they did not discover evidence for a significant risk premium at all (Görgen et al., 2020) suggesting that investors do not require a compensation in the form of a premium for bearing sustainability risk.

The relative importance of sustainability issues from a performance-return relationship perspective is less studied. In most cases governance and environmental performance had the most substantial financial impact, whereas social pillar performance is less incorporated into prices (Drei et al., 2020; Giese et al., 2021). Carbon emissions, labour management and health & safety were found to be the most influential sustainability issues within the pillars by Giese et al. (2021). A gap between European and US markets were also identified, sustainability generally delivering higher returns in the Eurozone (Drei et al., 2020; Bennani et al., 2018).

The gap in existing literature is straightforward and further investigation is needed on: (1) whether the separate environmental and social pillar performance is incorporated into equity prices and to see if one has a more substantial financial impact than the other; (2) whether sustainability issues independently have enough informational power to influence prices and also to observe if more clear-cut issues, such as Emissions and Workforce have a more substantial effect compared to less tangible aspects such as Resource Use and Communities.

I examine these research questions by using multi-factor models. First, I compute zero-investment portfolios both for the European and US markets based on each sustainability measure by longing 30% of firms with the lowest performance and shorting 30% of firms with the highest performance. I include a time lag to avoid look-ahead bias. After looking at the low-minus-high (LMH) portfolios' performance, I perform a formal asset pricing test first on a portfolio-level and then on a single-stock level: I regress monthly returns on the Carhart four factors and the sustainability factors to compare the extent to which variation in stock returns are explained by these risk factors: I analyse the magnitudes and the significance of the coefficients, as well as compare the different values of the adjusted R^2. This is followed by a Fama and Macbeth two-pass regression to test for risk premium.

I show that all long-short portfolio generated a positive mean monthly return suggesting that firms with low sustainability scores outperformed their sustainable peers. Sustainability factor returns, and risk premiums are higher in the US than in Europe, implying that sustainability risk is incorporated into prices to a greater extent and investors do require extra return, a risk premium for bearing additional risk. For instance, the risk premium on the social factor is more than 3.7 times higher in the US than in Europe. On average, coefficients on the environmental performance-based factor are higher and have a larger level of significance, as well as there is a greater risk premia associated with it compared to the social risk-based factor. For instance, in Europe, the risk premium associated with environmental performance is more than 2.3 times larger than the one associated with social performance. This suggests that environmental performance is considered to be more material for investors, and they are more likely to recognize environmental risk as a form of financial risk. Compared to expectations, Resource use and Communities are proved to be more important for the financial markets than Corporate emissions and Workforce, respectively. However, these issues, compared to broader risk measures such as overall ESG, environmental and social performance, exhibit less significant and less consistent results across the different asset pricing tests, suggesting that they might not hold enough informational power for investors in none of the regions. Overall, the LMH portfolios' positive performance and the evidence for the positive and significant risk premium are not fully in line with previous academic findings. On the contrary, the larger relative importance of environmental risk is proved in other previous research papers, as well.

The paper is organized as follows. Chapter 2 reviews the related literature on the inclusion of sustainability into asset pricing tests and multi-factor models. Theoretical background and hypotheses are explained in Chapter 3. Chapter 4 describes the data sources, data sample, as well as elaborates on the methodology for the empirical study. This is followed by the presentation and interpretation of the results in Chapter 5. Discussion and conclusion, including limitations are presented in Chapter 6.

2. Literature Review

Factor models are widely used in asset pricing research to investigate the systematic drivers behind the cross section of stock returns. First, Fama and French (1993) extended the classic CAPM model of Sharpe (1964), Lintner (1965) and Black (1972) and identified two additional common stock-market risk factors to explain variation in stock returns besides the overall market factor: a factor standing for size (SMB) and for value (HML). In the following years the FF 3 factor model was extended several times, improving the explanatory power of the original model. Carhart (1997) introduced another factor to the FF3 model, namely Momentum, while in 2015 Fama and French extended the model with two factors standing for investment and profitability.

As Venturini (2022) states, it is still highly debated whether sustainability risk present another systematic risk, or so-called anomaly, in contrast of a single firm-specific characteristic, that could be eliminated through diversification. To address this question, several sustainability risk factors were constructed in the form of zero-cost portfolios and were added to multi-factor models in recent years. Some focus on overall ESG performance taking firms' ESG score (Bennani et al., 2018; Giese et al., 2021; Roberodo et al., 2022) or comparing the performance of the constituents of a sustainability index with those of companies in the polluting industries through a long-short portfolio (Gregory et al., 2021). However, a large part of research investigates the financial implications only of the E (environmental) pillar with ambivalent results on which issues within E have the most substantial financial impact. Studied climate factors are ranging from proxies of drought measures (Hong et al., 2019) and heat stress measures (Gostlow, 2022) to corporate carbon emissions (Bolton et al., 2021; 2022).

Görgen et al. (2020) proxied companies' carbon risk by the combination of three subscores: value chain, public perception and adaptability. The authors constructed a BMG (Brown-Minus-Green) carbon risk factor-mimicking portfolio and experienced that the cumulative return of the BMG factor had dropped to around -0.23%. They found that the average monthly return of the BMG factor was -0.11% and its correlation with the Carhart model's factors was relatively low, implying that it is capable to enhance the explanatory power of the factor model. They found insignificant but negative risk premium of -0.097% for the BMG factor indicating that investors did not require a compensation for taking carbon risk, as opposed to classical finance theory expectations. This is also supported by the negative average monthly return of the BMG portfolio, i.e. the outperformance of green firms.

The mostly used proxy for transition risk is carbon emissions. Pollutive-Minus-Clean (PMC) portfolio was constructed by Huij et al. (2022) by taking a long position in the heaviest-emitting firms and a short position in the least-emitting companies using relative GHG emissions data. They found that the mean return of the PMC portfolio was considerably negative, -2.8% in the period of 2007-2021. The EMI (efficient-minus-inefficient) portfolio constructed by In et al. (2019) also justified the outperformance of least-polluting firms with a 3.5-5.4% annual abnormal return. However, different results were found by Hsu et al. (2023) when they measured the emission intensity of firms and took a long position in companies with the highest emission intensity, while shorting firms with the lowest emission intensity. Compared to Görgen et al. (2020) and Huij et al. (2022), they found a positive annualized excess return of 4.42% for the long-short portfolio. They argue that the significant positive risk premium might be driven by investor preferences, political connectivity or by the fact that low prices are driven up by bidders.

As elaborated earlier, in most cases carbon emissions are studied as a proxy for transition risk, but there are some examples for studying the effects of physical risk measures, too. The effect of physical risk through the lens of drought trend measures was studied by Hong et al. (2019). After sorting countries' food industry portfolio based on the exposure to drought trends, an annual 6.72% spread was found between negative trending and positive trending countries' portfolio of food stocks implying that stock markets indeed underreact to this physical risk measure. Gostlow (2022) also studied the pricing of physical climate risk and associated a statistically significant 0.39% risk premium with hurricane risk and a negative premium of -0.59% with heat stress, while no significant risk premium was found for sea-level rise and neither for extreme rainfall. Hain et al. (2022) provided a comprehensive study on the comparison of physical risk measures and their implication for financial markets. They found that firm-level physical climate risk measures diverge substantially among service providers making it ambiguous whether investors can actually reflect and incorporate physical risk in equity prices.

The social pillar of sustainability has been explored to a lesser extent so far, especially in the form of long-short portfolios. Hong et al. (2009) published one of the most pioneering studies on the effects of social norms while exploring 'sin stocks'. They constructed a portfolio of longing sin stocks and short selling their comparable and found a 30-bps monthly outperformance of sin stocks after adjusting for the Carhart four-factors. Edmans (2010) studied another aspect of social sustainability, namely employee satisfaction by analysing the performance of a value-weighted portfolio of the “100 Best Companies to Work For in America” and found an annual 3.5% four-factor alpha when measuring long-run drift, reflecting that companies with higher employee satisfaction outperform their comparable. Drei et al. (2020) highlights that the social pillar is less straightforward from a risk management perspective, while Bennai et al. (2018) points out that the S factor has only started to be rewarded recently, from 2016.

There are only a few existing examples for comparing the pricing effects of the three ESG pillars, i.e. their relative importance from the financial markets’ perspective. Giese et al. (2021) performed a study on the stock-price performance of the E, S and G pillar scores, as well as of the key issues within the pillar scores provided by MSCI by constructing long-short portfolios. All pillar score-based Q5-Q1 portfolios (best-rated minus worst-rated companies) yielded a positive return, showing the highest spread for the G score-based portfolios, followed by the E and S score-based portfolios respectively. Regional differences were identified: in Europe, financial markets responded the most for the G score, followed by the E and S scores, whereas in the US the E score gained the most attention. Looking at key issues, carbon emission score-based portfolios showed the highest outperformance, followed by labour management and health & safety scores-based portfolios. After observing the results, Giese et al. (2021) argue that “correlations between ESG indicators and style factors may explain the performance contribution”. They also highlight the importance of the time horizon in studying the financial implications, indicating that in the short-term G score has a more substantial effect, while the E and S scores developed gradually but had an impact in the long-run.

Drei et al. (2020) likewise identified a divide between North American markets and the Eurozone. The 2018-2019 period was marked by the reduction in the performance of the long-short portfolios (long in best-performing and short in worst-performing firms) based on the S and G pillars in the US, experiencing the first negative return for the portfolio based on E score, implying that firms with worse environmental performance outperformed their peers. At the same time, a positive and increasing trend was found for all three pillars in the Eurozone. Interestingly, in the US the long-short portfolio based on E showed the highest annualized return among all ESG scores in the period of 2014-2017 (exceeding 4%) before reaching negative returns in the following period, probably reflecting the changing public policy environment of the Trump administration.

Similarly, Bennani et al. (2018) found a gap between financial markets' reaction in the US and in Europe: in the latter region, ESG investors were rewarded with neutral or positive returns (reaching an annualized excess return of 6.6% in 2014-2017 when buying best performing stocks and selling worst performing stocks in terms of ESG), while it meant rather a drawback for investors in the US. When studying the relationship using the FF5 model, like Giese et al. (2021), the authors also note when introducing more factors in the time-series regression, the impact of the ESG factor is reduced, demonstrating an interaction with traditional risk factors.

Overall, based on previous literature, one could argue that further evidence is still needed on (1) which aspects of ESG are priced (more), (2) which issues are priced more within the pillars, (3) whether ESG and separate pillar-based performance actually present a risk premium for investors especially in case of a high correlation with traditional risk factors. Furthermore, regional differences have been less studied so far, thus a comparative analysis provides a new perspective.

3. Theoretical Framework

3.1. Theoretical Background

3.1.1. Measuring systematic risk using factor models

Multi-factor models are used to provide an explanation to security returns by offering investors a method to measure their exposure to different systematic risks. The benchmark or factor portfolios of the time-series multi-factor models are zero-investment portfolios whose returns track the development of one source of risk but are not correlated with other sources of risk. (Bodie et al., 2021) One of the most well-recognized multi-factor models are the Fama and French three-factor model (FF3) (1993), as well as its extensions: the Carhart (1997) four-factor model and the Fama and French five-factor model (FF5) (2015). These models follow a time-series regression approach: investors regress stock returns on factor mimicking portfolios. These zero-investment portfolios have a long position in stocks with high exposure to the risk factor while simultaneously shorting those with low exposure reflecting the difference in returns between the two diversified portfolios. The factor loadings show to what extent stock returns can be explained by the exposure to the risk factors, representing the trading component of the risk. (Fama & French, 1993; 2015) When including another factor to the FF3, in most cases researchers control for the size factor and in some cases for the value factor as well by creating double-sorting portfolios from the intersection of the two (three) factor groups (terciles/quintiles), to relieve the effects arising from size (value) on the examined risk factor.

In most asset pricing studies, time series multi-factor regressions are followed by the Fama and MacBeth (1973) multi-factor model to test if there is a risk premium associated with a certain factor. They look at the relationship between returns and estimated risk factor betas, as well as including a vector of firm-specific characteristics. This regression is a two-pass regression: the first phase includes the estimation of factor sensitivities / betas by regressing rate of returns on the given factor(s), followed by a regression of the annualized rate of returns on the estimated betas, i.e. obtaining the risk premium for each factor in the second pass. For the first pass regression, a rolling-window time series method is often used to relieve us from the time series regressions' common assumption of having constant coefficients over time. (Zivot & Wang, 2007)

3.1.2. Integrating sustainability risk in factor models

A prevalent approach in empirical asset pricing literature to study if a type of sustainability risk is priced by financial markets and/or associated with a risk premium is constructing high-and low risk portfolios mimicking the performance related to the given sustainability factor. Researchers take a long position in a portfolio of companies with low performance in the considered sustainability measure(s) and a short position in firms with high performance. The constructed zero-investment portfolio then tracks the evolution of the source of sustainability risk: if the return of the low-minus-high is positive, investors do require compensation for bearing sustainability risk. Bennani et al. (2018) argue that there are two conditions for ESG to be considered as a new risk factor: first, it has to achieve extra performance or decrease risk for investors and second, it has to complement other traditional risk factors.

Usually the sustainability factors are constructed following the Fama and French methodology: researchers first divide stocks into two groups based on their market capitalization: small and big, taking the median value as the breakpoint; then they further sort the two groups into subgroups based on the terciles of the performance in the given sustainability score, usually acquiring six double-sorted portfolios per factor (Görgen et al., 2020; Gurvich & Creamer, 2022; Huij et al., 2022). Given the size of this study's sample and the selection method (indexes representing mostly companies with large market capitalization), adjustment is not necessary, so I do not enforce size neutrality.

3.2. Hypotheses Development

Based on the literature review, I argue that firms with high sustainability performance outperform those with low sustainability scores, reflecting their business resilience against systematic and local market shocks. This is line with the doing-well-by-doing-good hypothesis of Bolton et al. (2022) and with the assumption that financial markets are not (yet) efficient in pricing sustainability related risk, meaning that investors do not require extra return for bearing additional risk. Consequently, I expect portfolios having a long position in companies with low sustainability score while shorting those with high score, to generate negative monthly returns on average.

Hypothesis 1: *Each sustainability score-based long-short portfolio generates a negative average monthly rate of return.*

I expect the findings for all long-short portfolios to be more significant in the European markets based on the literature review.

Hypothesis 2: *The sustainability score-based long-short portfolios generate a larger absolute mean monthly return in Europe than in the US.*

When comparing the environmental and social performance, I expect a higher return spread between companies based on environmental performance, since environmental performance is considered more relevant by the financial markets due to strengthening regulations and amplifying general vigilance about climate change. It is also in line with the findings of Giese et al. (2021) and based on the reasoning of Drei et al. (2020) that environmental aspects are more straightforward from a risk management perspective.

Hypothesis 3a: *Environmental performance is priced to a higher extent by investors than social performance, i.e. the Environmental score-based long-short portfolio generates a larger absolute mean monthly return than the Social score-based long-short portfolio.*

When diving within the environmental key issues, I expect the long-short portfolio based on Emissions to yield a higher absolute return on average compared to the one based on Resource use. This is because most sustainability regulations target corporate emissions, which (will) result(s) in carbon taxes, i.e. direct, additional expenses for companies. Moreover, emission reduction is among the most common sustainability targets set by companies and recognized by markets. Some people even argue that instead of ESG, investors should only focus on firms' emissions. (The Economist, 2022) Whereas when exploring the key social issues, I expect Workforce performance to have a more substantial relationship with stock returns than Communities, since employee relations and employee satisfaction have a direct impact on company operations. These expectations are also in line with the findings of Giese et al. (2021).

Hypothesis 3b: *Corporate emission related performance is priced to a higher extent by investors than resource use performance, i.e. the Emissions score-based long-short portfolio generates a larger absolute mean monthly return than the Resource use score-based long-short portfolio.*

Hypothesis 3c: *Workforce related performance is priced to a higher extent by investors than communities related performance, i.e. the Workforce score-based long-short portfolio generates a larger absolute mean monthly return than the Communities score-based long-short portfolio.*

When comparing the returns of companies with low and high sustainability score, I expect high performing companies' returns to move to the opposite direction as the return of the sustainability factor. On the other hand, companies with low sustainability score are expected to have a positive beta, representing their sustainability risk exposure.

Hypothesis 4a: *High-sustainability firms' returns are negatively correlated with the sustainability factor, i.e. they have a negative beta.*

Hypothesis 4b: *Low-sustainability firms' returns are positively correlated with the sustainability factor, i.e. they have a positive beta.*

Following the findings of Görgen et al. (2020) and Giese et al. (2021) I expect the common factor model's explanatory power to increase when adding the sustainability factors to the Carhart model.

Hypothesis 5: *Overall ESG and pillar-score based factors are able to increase the explanatory power of the common factor model.*

In a formalized test for priced risk premia (Fama and Macbeth, 1973), following the findings of Görgen et al. (2020) and Gregory et al. (2021) I expect sustainability factors to not be associated with a significantly positive risk premium, assuming that markets are not efficient, and investors do not require a compensation for bearing sustainability risk. This is in line with Hypothesis 1, expecting that more sustainable companies generate higher returns than their low sustainable peers. I expect a statistically significant (negative) premium for the overall ESG and pillar score-based factors.

Hypothesis 6: *The sustainability factors are not associated with a positive risk premium, i.e. investors do not require compensation for bearing sustainability risk.*

Following the reasoning explained before, I expect investors to be more vigilant about environmental risk, emissions related risk and workforce related risk compared to social risk, risk associated with resource use and with communities, respectively.

Hypothesis 7a: *The Environmental pillar score-based factor is associated with a larger absolute risk premium than the Social pillar score-based factor.*

Hypothesis 7b: *The Corporate emissions score-based factor is associated with a larger absolute risk premium than the Resource use score-based factor.*

Hypothesis 7c: *The Workforce score-based factor is associated with a larger absolute risk premium than the Communities score-based factor.*

In the following chapters, these hypotheses are tested, and results are presented.

4. Data and Empirical Strategy

4.1. Data

The observed time period is 2011-2022. Data sample includes the constituents of the Stoxx Europe 600 index and the S&P 500 index excluding companies from the financial sector and companies that did not exist or did not have an ESG score throughout the whole sample period. Overall, 582 companies remained in the final sample, including 307 firms in the European sample and 272 firms in the US sample. Industry and country representation are presented in Table 1 and Table 2, separately for the companies in the European and US samples.

Industry name	# of Firms	
	Europe	**US**
Communication Services	22	11
Consumer Discretionary	39	36
Consumer Staples	29	30
Energy	13	17
Health Care	27	37
Industrials	79	44
Information Technology	15	36
Materials	39	19
Real Estate	17	18
Utilities	27	24
Total	**307**	**272**

Table 1 - Industry representation by number of firms

Country of Headquarters	# of Firms
PANEL A: Europe	
Austria	5
Belgium	7
Cyprus	1
Denmark	11
Finland	11
France	46
Germany	34
Republic of Ireland	8
Italy	9
Luxembourg	4
Netherlands	12
Norway	7
Portugal	3
Spain	14
Sweden	27
Switzerland	25
United Kingdom	83
Total	**307**
PANEL B: US	
Republic of Ireland	6
Switzerland	2
United Kingdom	1
United States of America	263
Total	**272**

Table 2 - Country representation by number of firms

4.1.1. Data on sustainability performance

To measure the companies' sustainability performance, ESG data from EIKON is retrieved. Data is obtained for each company i in the sample for all year t between 2010-2021:

ESG score name	**ESG score explanation**
Overall ESG score ($ESG_{i,t}$)	measures the company's overall ESG performance
Environmental pillar score ($E_{i,t}$):	measures the company's environmental performance
Social pillar score ($S_{i,t}$):	measures the company's social performance
(Corporate) Emissions score ($CE_{i,t}$):	measures the company's commitment and effectiveness towards reducing environmental emission in the production and operational processes
Resource use score ($RU_{i,t}$):	reflects a company's performance and capacity to reduce the use of materials, energy or water, and to find more eco-efficient solutions by improving supply chain management
Workforce score ($WF_{i,t}$):	measures a company's effectiveness towards job satisfaction, healthy and safe workplace, maintaining diversity and equal opportunities, and development opportunities for its workforce
Community score ($CO_{i,t}$):	measures the company's commitment towards being a good citizen, protecting public health and respecting business ethics

Table 3 - ESG score list

Governance related scores are not obtained since governance is mostly associated with financial stability and compliance. The scores are measured on a scale of 0-100, higher scores representing better performance.

Overall, as Table 4 depicts, when comparing the ESG performance of the companies in the final sample in Europe and in the US, it is clear that all ESG scores in Europe have a higher mean, higher median and smaller standard deviation except for the Communities score.

Variable	Mean	Median	Std. Dev.
PANEL A: Europe			
ESG score	65.865	68.63	16.972
Environmental pillar score	66.865	71.56	21.250
Social pillar score	68.658	73.29	20.704
Emissions score	72.671	79.55	23.720
Resource use score	74.145	81.35	24.073
Workforce score	79.912	85.2	18.810
Communities score	65.337	71.41	27.817
PANEL B: US			
ESG score	59.844	63.16	17.993
Environmental pillar score	55.288	60.27	25.638
Social pillar score	61.349	63.63	20.454
Emissions score	60.222	66.07	29.512
Resource use score	62.720	70.94	30.210
Workforce score	65.993	69.69	23.424
Communities score	78.7705	84.60	18.928

Table 4 - ESG score summary statistics

4.1.2. Stock market and corporate data

Stock market and corporate data is retrieved from EIKON: monthly total returns are downloaded, and market capitalization data is obtained annually to create the value-weighted portfolios. Additional financial firm-specific characteristics are retrieved for each company annually: total assets, total liabilities, return on assets (ROA), return on equity (ROE), leverage ratio (total debt to total equity (%)) and cash ratio (cash & short-term investments/total assets) is calculated.

Variable	N	Mean	Std. Dev.	Min	Max
PANEL A: EUROPE					
Total monthly return	44,208	.879%	.009	-77.801%	115.103%
Total assets (EUR)	3684	3.455e+10	5.894e+10	63813513	6.071e+11
Total liabilities (EUR)	3684	2.31e+10	4.22e+10	3.57e+07	4.52e+11
ROA	3684	.065	.118	-.393	2.368
ROE	3684	.186	.761	-3.285	24.099
Leverage	3684	.957	1.401	0	39.933
Cash	3684	.104	.094	0	.928
PANEL B: US					
Total monthly return	39,168	1.218%	.081	-83.2%	219.2%
Total assets (USD)	3264	3.929e+10	6.230e+10	1.043e+09	7.478e+11
Total liabilities (USD)	3264	2.59e+10	4.44e+10	1.75e+08	6.29e+11
ROA	3264	.404	7.878	-248.5	315.6
ROE	3264	.07	.07	-.41	.457
Leverage	3264	1.633	11.068	0	422.1
Cash	3264	.116	.133	0	.843

Table 5 - Descriptive statistics of financial data

4.1.3. Factor data

Monthly Fama and French 3 factors plus the Momentum factor for the European and the US market for the sample period are retrieved from Kenneth French's website.

4.2. Methodology

4.2.1. Portfolio sorting and return calculation for the zero-cost portfolios

As an observable proxy for sustainability risk, in each year I construct seven zero-cost portfolios separately for the European and US markets that show the return spread between a portfolio of low performing firms relative to a portfolio of high performing firms in terms of sustainability. They take a long (short) position in the worst (best) performing 30% of stocks based on the previous year's sustainability scores to avoid look-ahead bias, using the 30th and 70th percentiles. As mentioned in 3.1.2., adjustment for the size factor is not needed. As an example, the return on a zero-cost based on overall ESG performance in a given month is given by:

$$r_{LMHESG,t} = r_{LESG,t} - r_{HESG,t}$$

,where $r_{LESG,t}$ and $r_{HESG,t}$ are the returns on the single-sorting value-weighted portfolios in month t. The long-short portfolios are summarized in Table 6.

LMH portfolio name	Explanation
$LMHESG_t$	long in stocks with low overall ESG performance in the previous year and short in stocks with high overall ESG performance in the previous year
$LMHE_t$	long in stocks with low environmental performance in the previous year and short in stocks with high environmental performance in the previous year
$LMHS_t$	long in stocks with low social performance in the previous year and short in stocks with high social performance in the previous year
$LMHCE_t$	long in stocks with low corporate emissions score (high emission) in the previous year and short in stocks with high corporate emissions score in the previous year
$LMHRU_t$	long in stocks with low resource use score in the previous year and short in stocks with high resource use score in the previous year
$LMHWF_t$	long in stocks with low workforce score in the previous year and short in stocks with high workforce score in the previous year
$LMHCO_t$	long in stocks with low communities score in the previous year and short in stocks with high communities score in the previous year

Table 6 - LMH portfolio list

4.2.2. Low and high equally-weighted portfolio analysis

I sort firms into two annually rebalanced equally-weighted portfolios based on the median sustainability score and thus obtain a portfolio of low and high performing firms per each score. Then I run a time-series regression of the equally-weighted portfolios' monthly excess returns on the Carhart model augmented with the respective LMH portfolio. A two-tailed t-test is performed to see if there is a significant difference between the low and high groups' coefficients. I do not include two sustainability factors in the same regression since grouping is different for each score. This how I am able (1) to compare the betas of the low and high performing firms, (2) compare the betas' magnitude and significance between the different sustainability scores, (3) compare the explanatory power of the models.

4.2.3. Time-series regression on a firm level using the Carhart four-factor model

To estimate and compare the effect of the firms' sustainability performance on stock performance, return sensitivities to the sustainability factors are calculated by running a time-series regression of the firm's monthly excess returns on Carhart model while also including (1) the ESG-based long-short portfolio; (2) the two pillar-score-based long-short portfolios;

(3) the two environmental subscore-based long-short portfolios; and (4) the two social subscore-based long-short portfolios. I estimate the factor loadings:

Model name	Explanation
Model 1: Carhart four factors	$R_{i,t} = \alpha_i + \beta_{iRM}RM_t + \beta_{iSMB}SMB_t + \beta_{iHML}HML_t + \beta_{iMOM}WML_t + \varepsilon_{i,t}$
Model 2: Carhart four factors + LMHESG	$R_{i,t} = \alpha_i + \beta_{iRM}RM_t + \beta_{iSMB}SMB_t + \beta_{iHML}HML_t + \beta_{iMOM}WML_t + \beta_{iLMHESG}LMHESG_t + \varepsilon_{i,t}$
Model 3: Carhart four factors + LMHE + LMHS Comparison of environmental and social pillar scores	$R_{i,t} = \alpha_i + \beta_{iRM}RM_t + \beta_{iSMB}SMB_t + \beta_{iHML}HML_t + \beta_{iMOM}WML_t + \beta_{iHMLE}LMHE_t + \beta_{iHMLS}LMHS_t + \varepsilon_{i,t}$
Model 4: Carhart four factors + LMHCE + LMHRU Comparison of environmental subscores	$R_{i,t} = \alpha_i + \beta_{iRM}RM_t + \beta_{iSMB}SMB_t + \beta_{iHML}HML_t + \beta_{iMOM}WML_t + \beta_{iHMLCE}LMHCE_t + \beta_{iHMLEI}LMHRU_t + \varepsilon_{i,t}$
Model 5: Carhart four factors + LMHWF + LMHCO Comparison of social subscores	$R_{i,t} = \alpha_i + \beta_{iRM}RM_t + \beta_{iSMB}SMB_t + \beta_{iHML}HML_t + \beta_{iMOM}WML_t + \beta_{iHMLHR}LMHWF_t + \beta_{iHMLWF}LMHCO_t + \varepsilon_{i,t}$

Table 7 - Model list

Our coefficients of interest are $\beta_{iLMHESG}$, β_{iLMHE}, β_{iLMHS}, β_{iLMHCE}, β_{iLMHRU}, β_{iLMHWF} and β_{iLMHCO}: I compare their average, sign and significance level. Standard errors are calculated using Newey and West (1987), which allows for $\varepsilon_{i,t}$ to be heteroskedastic and serially correlated.

4.2.4. Fama and MacBeth regression

This section tests whether the constructed factors are priced by the market for which I use the methodology of Fama and MacBeth (1973) on a single stock level. This procedure allows me (1) to see if there is a risk premium associated with the sustainability factors and (2) to compare the risk premiums in each model listed in Table 7. For each model I follow the following steps: First, I estimate 24-months rolling window coefficients for each factor and I obtain the average return for each window. Then I run a 24-months rolling-window regression of the monthly excess returns on the average returns. Finally, I perform the Fama and Macbeth regression including factor betas and firm-specific control variables as independent variables. Control variables are listed in Table 8.

Control variable	**Explanation**
ROA	Return on total assets
ROE	Return on total equity
Leverage	Total debt to total equity
Cash	Cash & short-term investments / Total assets
Size	Ln of total assets

Table 8 - Control variables list

5. Results and interpretation

5.1. Sustainability factors – zero-cost portfolio performance

The average monthly return of every sustainability score-based long-short portfolio is positive in both regions suggesting that companies with low sustainability scores perform better than more sustainable firms. Thus, *Hypothesis 1 is rejected* since results are inconsistent with the expectations as none of the long-short portfolios have realized a negative average monthly return. Consequently, evidence for the doing-well-by-doing good hypothesis was not found.

Demonstrated in Table 9, returns are larger in the US than in Europe for all sustainability portfolios, having the LMHS portfolio reaching the highest mean monthly return of 0.668%, while in Europe the LMHE portfolio realized the highest mean monthly return (0.292%). In both regions the LMHWF portfolio generated the smallest monthly return on average, 0.135% in Europe and 0.167% in the US. Since all of the zero-cost portfolios yielded a higher return in the US than in Europe, results are inconsistent with the expectations, thus *I reject Hypothesis 2*. This larger performance dispersion between companies with low and high performance could suggest that investors in the US incorporate sustainability considerations into their decision making to a higher extent, since they require a higher return when bearing additional risk.

In the US, the return spread between the sustainability portfolios is also higher: the LMHS portfolio outperformed the LMHESG portfolio by 0.354 pp. and the LMHE portfolio by 0.265 pp. On the contrary, in Europe, there is only a minor difference when comparing the performance of the best-performing portfolio (LMHE) portfolio to the LMHESG and LMHS portfolios: the difference is 0.014 pp. and 0.036 pp. respectively. Since the LMHE portfolio only generates a larger absolute mean monthly return in Europe, *I do not reject Hypothesis 3a in Europe, but I do reject in the US.* The larger difference between the portfolios' performance in the US could indicate that investors differentiate between environmental and social performance to a higher extent than investors in Europe.

Looking at the within pillar score-based portfolios, the LMHRU portfolio outperforms the LMHCE portfolio in Europe, while the LMHCE portfolio generates the higher mean monthly return in the US. Thus, *Hypothesis 3b is not rejected in the US*, suggesting that corporate emissions are priced to a higher extent than resource use performance. When comparing the social subscore-based portfolios, in both regions the LMHCO portfolio outperforms the LMHWF portfolio. Hence, *Hypothesis 3c is rejected in both regions.* This implies that compared to expectations, relationship with communities is considered more material by investors than relationship with employees.

Factor	Europe		US	
	Mean monthly return (%)	**Standard deviation**	**Mean monthly return (%)**	**Standard deviation**
ERM	0..491	.049	0.907	.044
SMB	0.073	.017	-0.225	.023
HML	-0.013	.029	-0.121	.032
WML	0.759	.033	0.461	.030
LMHESG	0.278	.021	0.314	.022
LMHE	0.292	.018	0.403	.012
LMHS	0.256	.016	0.668	.022
LMHCE	0.157	.016	0.398	.021
LMHRU	0.222	.015	0.338	.020

LMHWF	0.135	.015	0.167	.025
LMHCO	0.203	.019	0.399	.019

Table 9 - LMH portfolio returns

The correlations between the sustainability factors and the factors of the Carhart model are relatively low, as shown in Table 10. As Görgen et al. (2020) argues, this suggests that the sustainability factors are able to improve the explanatory power of the Carhart model. When it comes to the correlation between the sustainability factors themselves, the correlation between the LMHESG, LMHE and LMHS factors are relatively large in both regions. Since in Model 3, both pillar score-based factors are included in the regression, resulting from the high correlation, it might be challenging to isolate the effects of social and environmental performance. On the other hand, within pillar score-based factors correlate with each other to a smaller extent; especially when looking at the correlation between the LMHWF and LMHCO portfolios. Consequently, separating the factors' influence is more explicit.

PANEL A: EUROPE											
Factors	RM	SMB	HML	WML	LMH ESG	LMHE	LMHS	LMH CE	LMH RU	LMH WF	LMH CO
RM	1.000										
SMB	0.090	1.000									
HML	0.282	-0.132	1.000								
WML	-0.465	0.044	-0.532	1.000							
LMHESG	0.144	0.358	-0.404	0.047	1.000						
LMHE	0.083	0.466	-0.404	0.135	0.876	1.000					
LMHS	0.072	0.322	-0.393	0.081	0.878	0.820	1.000				
LMHCE	0.146	0.506	-0.327	0.098	0.640	0.774	0.620	1.000			
LMHRU	-0.089	0.291	-0.227	0.275	0.536	0.616	0.639	0.591	1.000		
LMHWF	-0.142	0.303	-0.304	0.245	0.594	0.586	0.646	0.550	0.746	1.000	
LMHCO	0.367	0.305	0.088	-0.276	0.666	0.586	0.645	0.412	0.329	0.225	1.000
PANEL B: US											
Factors	RM	SMB	HML	WML	LMH ESG	LMHE	LMHS	LMH CE	LMH RU	LMH WF	LMH CO
RM	1.000										
SMB	0.367	1.000									
HML	-0.026	0.004	1.000								
WML	-0.295	-0.167	-0.398	1.000							
LMHESG	0.360	0.370	0.240	-0.160	1.000						
LMHE	0.400	0.321	-0.002	-0.116	0.799	1.000					
LMHS	0.465	0.335	0.063	-0.174	0.784	0.664	1.000				
LMHCE	0.235	0.239	0.234	-0.134	0.837	0.846	0.674	1.000			
LMHRU	0.039	0.188	0.034	0.069	0.637	0.639	0.485	0.674	1.000		
LMHWF	0.203	0.252	0.184	-0.275	0.639	0.484	0.570	0.569	0.227	1.000	
LMHCO	0.148	0.030	-0.290	0.143	0.229	0.334	0.396	0.369	0.256	0.105	1.000

Table 10 - Factor correlation matrix

5.2. Low and high equally-weighted portfolio analysis

Table 11 presents the factor loadings on the sustainability factors for each low and high equally-weighted portfolio, as well as the adjusted R^2 separately for Europe and for the US.[2] A comparison of the adjusted R^2 confirms that the sustainability factors are able to enhance the explanatory power of the Carhart model, both for the high and low portfolios. The adjusted R^2 is higher in the US in all cases, exceeding at least 92%. In the US, the model including the LMHCE portfolio has the highest explanatory power while it is the case in Europe only for the high portfolio.

As expected, the high portfolio has a negative loading, though insignificant, while the low portfolio has a positive and significant loading on the LMHESG factor. The two-tailed t-test results suggest that the difference between the two betas is significant at 10% in Europe and at 1% in the US. Thus, *in case of the overall ESG score, I do not reject Hypothesis 4a and 4b,* suggesting that indeed firms with low ESG score positively, while firms with high ESG score are negatively correlated with the sustainability factor.

Similarly, for both of the pillar score-based groups in the US, high performing firms have a negative but not significant coefficient, while low performing companies have a positive and significant coefficient on the LMHE and LMHS factors. The difference is significant at 1% in both cases. Hence, *in the case of the pillar score-based groups, I do not reject Hypothesis 4a and 4b in the US.* However, in Europe I reject Hypothesis 4a since the portfolio of high performing companies do not have a negative coefficient on the LMHE and LMHS factors. In both regions, the absolute value of coefficients on the LMHE factor are higher than of the LMHS factor, suggesting that environmental performance is more important from a financial perspective.

[2] The complete regression results including the coefficients on the Fama and French and Carhart factors can be found in the Appendix.

For all within pillar score-based groups, I do not reject Hypothesis 4b since low performing firms' loading on the sustainability factors are positive and significant at 1% and 5%. However, similarly to the European pillar score-based portfolios, high performing firms' loadings are also positive thus *I reject Hypothesis 4a.* As opposed to expectations, portfolios have a higher absolute beta on the resource use and communities factors compared to the corporate emissions and workforce factors respectively.

The results suggest again that American financial markets price sustainability risk more efficiently (at least when proxied by overall ESG, environmental and social performance).

	Coefficient						**Adj. R^2 (%)**			
	Europe			**US**			**Europe**		**US**	
Factor	Low	High	Diff.	Low	High	Diff.	Low	High	Low	High
Carhart 4	-	-	-	-	-	-	76.6	78.9	92.6	93.2
LMHESG	0.227** (1.92)	-0.006 (-0.05)	0.233*	0.276*** (4.77)	-0.024 (-0.41)	0.300***	77.4	78.8	93.8	93.2
LMHE	0.334** (2.30)	0.068 (0.46)	0.266	0.271*** (4.50)	-0.026 (-0.39))	0.297***	77.1	79.5	94.0	92.6
LMHS	0.321** (2.15)	0.045 (0.30)	0.276	0.239*** (5.00)	-0.007 (-0.17)	0.247***	78.3	78.4	93.4	93.3
LMHCE	0.419*** (3.19)	0.053 (0.43)	0.366**	0.303*** (5.88)	0.069 (1.39)	0.234***	76.8	80.2	94.3	93.8
LMHRU	0.439*** (2.93)	0.095 (0.64)	0.344	0.212*** (3.19)	0.022 (0.43)	0.190**	79.2	78.5	93.3	93.6
LMHWF	0.305** (2.02)	0.088 (0.54)	0.217	0.230*** (4.54)	0.008 (0.18)	0.222***	78.2	78.5	93.4	93.8
LMHCO	0.293** (2.55)	0.112 (0.99)	0.181	0.196** (2.24)	0.036 (0.61)	0.160	79.4	78.0	92.9	93.8

Table 11 - Sustainability low and high portfolio performance

5.3. Stock-level time series regression

To reinforce the results of the portfolio-level regression, I compare the results with the stock-level regression, following the models described in 4.2.3. The results are robust because the average adjusted R^2 is larger again when sustainability factors are included in the regressions. It is the highest in the US when the pillar-score based factors are added to the Carhart model, while in Europe the inclusion of the social subscore-based factors resulted in the highest average adjusted R^2. Based on the portfolio-and single-stock level time-series regressions, *I do not reject Hypothesis 5*: the LMHESG, LMHE and LMHS factors are able to increase the explanatory power of the common factor model. Overall, since in this case the regressions are based on a single stock-level, the explanatory power of the models is lower since the effect of idiosyncratic factors are not diversified away. Interestingly, compared to the portfolio-level regressions, the models' explanatory power is higher in Europe than in the US when running the regressions on a single stock level.

For a detailed assessment of the sustainability factors' effect on company stock performance, the number of significant betas for each factor are presented in Table 12. Based on two-sided t-tests 98.7% of firms show a significant LMHESG beta on a 10% level in Europe, while 88.3% in the US. This is comparable to the % of significant betas of the common factors: in Europe it's even higher than SMB betas (37.5%), HML betas (41.0%) and WML betas (22.8%).

In the US, the LMHE and LMHS factors also perform well in explaining variation in stock returns: 81.6% of firms have a significant LMHE beta and 49.5% have a significant LMHS beta on a 10% level. In Europe, the ratio of significant betas is lower for the pillar score-based factors: it's 15.6% and 12.4% respectively. The results suggest that (1) again, investors in the US consider sustainability risk more material, (2) environmental performance is more important from the financial markets' perspective.

Within pillar score-based factors' impact are negligible in both regions, meaning that they do not carry relevant information for investors.

Panel A. Significance tests for explanatory power of various models in Europe vs. US

	Avg. adj. R2	
	Europe	US
(1) Carhart 4 factors	0.296	0.232
(2) Carhart 4 factors + LMHESG	0.296	0.232
(3) Carhart 4 factors + LMHE + LMHS	0.298	0.258
(4) Carhart 4 factors + LMHCE + LMHRU	0.298	0.230
(5) Carhart 4 factors + LMHWF + LMHCO	0.302	0.228

Panel B. Significance tests for factor betas in Europe

	Coefficient								
		Sign		10% level		5% level		1% level	
	Avg. coeff.	+	-	#	%	#	%	#	%
Alpha	0.005	265	42	85	27.7	55	17.9	15	4.9
ERM	0.712	307	0	303	98.7	302	98.4	294	95.8
SMB	-0.084	128	179	115	37.5	92	30.0	43	14.0
HML	-0.095	128	179	126	41.0	101	32.9	61	19.9
WML	-0.143	88	219	70	22.8	44	14.3	19	6.2
LMH ESG	0.114	181	126	303	98.7	62	20.2	28	9.1
LMHE	0.136	178	129	48	15.6	36	11.7	12	3.9
LMHS	0.084	176	131	38	12.4	22	7.2	4	1.3
LMH CE	0.109	169	138	77	25.1	52	17	16	5.2
LMH RU	0.216	200	107	59	19.2	35	11.4	10	3.3
LMH WF	0.152	189	127	67	21.8	48	15.6	20	6.5
LMH CO	0.167	180	127	91	29.6	68	22.1	30	9.8

Panel C. Significance tests for factor betas in the US

	Coefficient								
		Sign		10% level		5% level		1% level	
	Avg. coeff.	+	-	#	%	#	%	#	%
Alpha	-0.047	0	272	272	100	272	100	272	100
ERM	1.111	272	0	271	99.6	271	99.6	266	97.8
SMB	-0.026	113	159	37	13.6	17	6.25	4	1.5
HML	0.271	217	55	271	99.6	43	15.8	22	8.1
WML	-0.064	123	149	271	99.6	271	99.6	1	1
LMH ESG	0.344	225	47	271	88.3	23	7.5	2	0.7
LMHE	1.278	271	1	222	81.6	191	70.2	107	39.3
LMHS	-0.843	6	266	152	49.5	110	35.8	31	10.1
LMH CE	0.229	186	86	22	7.2	12	3.9	1	0.3
LMH RU	0.058	149	123	13	4.2	6	2.0	3	1.0
LMH WF	0.201	220	52	21	6.8	10	3.3	3	1.0
LMH CO	-0.186	75	197	15	4.9	8	2.6	2	0.7

Table 12 - Time series regression on single stock level

5.4. Test for risk premium - Fama and MacBeth regression

Table 13-15 present the results of the Fama and Macbeth regressions. When comparing the common factor risk premiums across models, they are similar: the SMB factor lacks significant premia in almost all cases. The HML factor is associated with a significant and negative premium in all models, while the WML factor is significantly positive. The tables show that all sustainability factors have a significant (1%) and positive coefficient, except for the LMHCE factor in the US which is negative but insignificant. Consequently, results are inconsistent with the expectations that low sustainability firms do not command a positive risk premium; thus *I reject Hypothesis 6*. However, results are in line with the previous findings of this paper, that low sustainability firms outperform high sustainability firms. In this case, one could argue that investors are aware of the financial risks associated with lower sustainability performance, and they require a premium for bearing additional risk.

When comparing the risk premia required by investors in Europe vs. in the US, again, consistent with the previous findings, the risk premium is almost twice as high in the US than in Europe for the LMHESG portfolio. This indicates that in the US investors are more vigilant when taking additional overall sustainability risk and they are more aware of the financial risk associated with it.

Variable	**Europe**		**US**	
	Carhart 4	Carhart 4 + LMHESG	Carhart 4	Carhart 4 + LMHESG
β_{RM}	-0.006 (-0.12)	-0.014 (-0.27)	0.284*** (6.93)	0.267*** (6.31)
β_{SMB}	0.013 (0.66)	0.004 (0.22)	0.052* (1.86)	0.028 (1.06)
β_{HML}	-0.165*** (-4.40)	-0.161*** (-4.23)	-0.418*** (-9.49)	-0.408*** (-9.11)
β_{WML}	0.421*** (13.18)	0.427*** (13.56)	0.694*** (16.08)	0.712*** (16.65)
β_{LMHESG}	-	0.062*** (5.55)	-	0.122*** (4.37)
ROA	0.847*** (4.79)	0.823*** (4.73)	-0.010** (-2.38)	-0.114** (-2.39)
ROE	0.102* (1.71)	0.105* (1.72)	1.367*** (7.00)	1.295*** (6.65)

Leverage	-0.027*** (-2.70)	0.025** (-2.52)	0.004** (2.04)	0.004* (1.83)
Cash	1.072*** (11.42)	1.054*** (11.52)	0.482*** (3.36)	0.460*** (3.32)
Total assets	-0.131*** (-10.51)	0.132*** (-10.53)	0.051*** (-7.06)	-0.048*** (-6.74)
Constant	3.908*** (13.23)	3.936*** (13.33)	-0.027*** (-5.67)	-2.784*** (-5.85)
R^2(in %)	21.27	21.57	26.44	27.54
N	37,147	37,147	32,912	32,912

Table 13 - Fama and MacBeth regression - Model 1 vs. Model 2

Similarly to the overall ESG risk, American investors require higher compensation for bearing additional risk when looking at the environmental and social pillar-score based performance; in the case of the LMHS portfolio the premia is 3.7 times higher in the US than in Europe. In both regions, the risk premium associated with environmental risk is higher than the social risk premium. For instance, the environmental premia is more than 2.3 times higher than the social premia. Hence, *I do not reject Hypothesis 7a*, indicating that investors are indeed more vigilant about environmental risk. This is in line with the results of the time-series regressions.

Variable	**Europe**	**US**
β_{RM}	-0.023 (-0.47)	0.247*** (5.74)
β_{SMB}	-0.006 (-0.31)	0.043*** (1.41)
β_{HML}	-0.175*** (-4.71)	-0.415*** (-9.56)
β_{WML}	0.420*** (13.91)	0.710*** (16.54)
β_{LMHE}	0.121*** (9.01)	0.199*** (7.81)
β_{LMHS}	0.051*** (3.87)	0.192*** (5.77)
ROA	0.887*** (5.13)	-0.009*** (-1.77)
ROE	0.085 (1.37)	1.467*** (7.45)
Leverage	-0.022** (-2.10)	0.005** (2.50)
Cash	3.67*** (11.96)	0.436*** (3.12)
Total assets	-0.121*** (-9.12)	-0.045*** (-6.18)
Constant	3.675*** (11.87)	-2.873*** (-6.05)
R^2(in %)	23.22	28.50
N	37,147	32,912

Table 14 - Fama and Macbeth regression - Model 3

When comparing the risk premiums associated with within pillar score performance, consistent with the overall, environmental and social risk associated premiums, risk premia is higher in the US when coefficients are significant. As opposed to expectations, but in line with the findings of the time series regressions, resource use related risk is associated with a higher premium than corporate emissions, thus *I reject Hypothesis 7b*. Investors do not require a higher compensation when taking workforce related risk than when bearing communities related risk, again, inconsistent with the expectations. Thus, *I reject Hypothesis 7c*.

Variable	Europe	US	Variable	Europe	US
β_{RM}	-0.019 (-0.38)	0.243*** (5.82)	β_{RM}	-0.009 (-0.18)	-0.273*** (6.63)
β_{SMB}	-0.0082 (-0.44)	0.038 (1.36)	β_{SMB}	0.012 (0.61)	0.031 (1.17)
β_{HML}	-0.164*** (-4.37)	-0.413*** (-9.02)	β_{HML}	-0.170*** (-4.51)	-0.410*** (-9.41)
β_{WML}	0.420*** (13.77)	0.714*** (16.74)	β_{WML}	0.433*** (14.99)	0.691*** (15.95)
β_{LMHCE}	0.052*** (4.10)	0.109*** (3.48)	β_{LMHWF}	0.035*** (2.84)	-0.038 (-1.10)
β_{LMHRU}	0.120*** (11.85)	0.253*** (10.78)	β_{LMHCO}	0.047*** (3.14)	0.150 (6.48)***
ROA	0.738*** (4.41)	-0.012** (-2.55)	ROA	0.874*** (5.33)	-0.005 (-1.20)
ROE	0.108* (1.86)	1.339*** (6.77)	ROE	0.064 (1.14)	1.233*** (6.08)
Leverage	-0.029*** (-3.04)	0.005** (2.42)	Leverage	-0.023** (-2.33)	0.004* (1.84)
Cash	1.161*** (11.33)	0.494*** (3.79)	Cash	0.011*** (12.20)	0.520*** (3.68)
Total assets	-0.127*** (-10.49)	-0.043*** (-6.16)	Total assets	-0.001*** (-10.29)	-0.053*** (-7.65)
Constant	3.824*** (13.29)	-2.877*** (-5.95)	Constant	0.038*** (13.08)	-2.680*** (-5.59)
R^2(in %)	22.66	28.44	R^2(in %)	23.6	27.2
N	37,147	32,912	N	37,147	32,912

Table 15 - Fama and MacBeth regression - Model 4 vs. Model 5

Overall, following the reasoning of Bolton et al. (2022), the positive and significant premiums for the overall and pillar score-based factors suggest that financial markets price the risk associated with sustainability in the US more than in Europe. The risk premia associated with overall ESG, environmental performance, social performance is higher in the US than in Europe: respectively by 0.06, 0.078 and 0.141.

6. Discussion and conclusion

Overall, the different regressions have consistent results, i.e. empirical results are robust in most cases. According to the findings, low sustainable firms outperformed those with high sustainability scores since all zero cost portfolios generated a positive average monthly return, as opposed to expectations. This is in line with the results of Hong et al. (2009) and Hsu et al. (2022) who found a positive average return for the long-short portfolio, while it is not evidenced by Görgen et al.'s (2020) BMG factor, nor by the PMC factor of Huij et al. (2022).

As expected, the constructed factors are able to increase the explanatory power of the common factor model, as suggested by the relatively low correlation with the FF and Carhart factors and by the increase in the adjusted R^2 both in the case of the portfolio-and single stock level regressions. This increase was also found in the study of Görgen et al. (2020).

Financial markets in the US are proved to be more efficient than in Europe and investors are more aware of the financial risk associated with sustainability as they require a higher return when taking additional risk. This is implied by (1) the LMH portfolios' larger return, i.e. return spread between firms with low and high sustainability scores, (2) the larger return spread between the different sustainability factors, indicating that investors differentiate between sustainability measures to a higher extent, (3) the larger absolute value of betas and higher ratio of significant betas, and (4) the higher risk premia they require when bearing sustainability risk. Even though only a few academic papers compare the North American and European markets in terms of pricing sustainability, this gap in market efficiency was not found by Drei et al. (2020) nor Bennani et al (2018).

The relative higher importance of environmental issues compared to social performance is consistent across regions, except for the fact that the LMHS portfolio on average outperformed the LMHE portfolio in the US. The number of significant betas is higher on the LMHE factor than on the LMHS factor, as well as investors require a higher risk premium when bearing environmental risk, as depicted by the Fama and MacBeth regression results. Overall, this is in line with the expectations and also reported by Giese et al. (2021): environmental issues gained more attention than social issues both in Europe and North-America.

Results on the within pillar score-based portfolios are inconsistent with the expectations and they are contradictory when comparing them between the different regressions. Moreover, results are less significant compared to the overall and pillar score-based portfolio regressions, especially in the case of number of significant betas when looking at the results of the stock-level time series regression. These imply that individually they might not hold enough relevant information for investors, i.e. they are not incorporated into stock prices to the same extent as broader sustainability issues. In spite of the inconsistency, when comparing the portfolio performances, significant betas and risk premiums, one could say that opposed to expectations, resource use and communities related performance are more relevant than emissions and workforce related issues. This is not in line with the findings of Giese et al. (2021).

It is clear from the results that investors do require a (significant) premium in both regions when bearing sustainability risk, consistent with the positive risk premium associated with hurricane by Gostlow (2022) and opposed to findings of Görgen et al. (2020). Risk premia is higher for each factor in the US, and it is the highest for overall environmental performance and within environmental pillar related issues. Consistent results across the different asset pricing tests suggest that most results are indeed robust.

Limitations can be divided into data and methodology related ones. The applied sample size can be considered as one of the main limitations of this study since having 307 companies in the European sample and less than 300 firms in the American one questions the robustness of the results. A larger sample with different data cleaning methods might bring differing results. Also, due to limited access to ESG data, scores from only one provider were obtained, which might have caused bias. Ideally, the combination of ESG scores obtained from different rating agencies or ESG data providers can be used as a proxy for ESG risk. Concerning the empirical strategy, the LMH portfolios were created based on their previous year's performance although taking the difference in performance, i.e. the change in scores for two consecutive years, might provide different perspectives. Moreover, the fact that companies were divided into only two groups for the equally-weighted portfolio-based time series regression, even though they are typically divided into quartiles or quintiles, might have not created a gap significant enough between the two groups. Again, this could have been overcome with a bigger sample size, so that the portfolios can still remain well-diversified. When looking at the correlation between the sustainability factors, the high correlation between the LMHE and LMHS portfolios might have made it difficult to isolate the effects of environmental and social performance. All these limitations provide an opportunity for further academic research.

Results found in this paper provide an insight for investors when making investment decisions while balancing between sustainability and financial performance. Since markets are proved to be efficient especially in the US, on average, higher returns can be achieved when bearing additional sustainability risk. However, since results are not consistent with all previous findings, financial markets are still in development and have just started adjusting to this new type of information. Looking at the results from a sustainability perspective, since the doing-well-by-doing-good hypothesis is rejected in this paper, financing the sustainability transition might not be financially beneficial for investors (yet), but is proved to be an advantage for companies as they face lower cost of capital. Consequently, this could provide an incentive for companies to engage in sustainability and make more socially responsible decisions.

The aim of this thesis has been to determine whether the Environmental and Social performance of companies are incorporated into stock prices and to see whether one has a more substantial impact than the other one. Furthermore, individual sustainability key issues were also investigated whether they have enough informational power for investors, while also looking at their relative importance. Results suggest that sustainability related information is relevant for investors and markets are efficient especially in the US, as investors require higher return for bearing additional risk. Overall, in line with the expectations, investors require a higher return for environmental risk compared to social risk. When it comes to information related to less broad sustainability issues such as performance associated with corporate emissions, resource use, workforce and communities, stock markets do not include the information to the same extent as they incorporate the overall, environmental and social risk related data. However, sustainability related information has only become relevant recently, hence their pricing impact might change in the coming years. In addition, sustainability performance might become more appreciated by financial markets, resulting in the outperformance of more sustainable firms and supporting the doing-well-by-doing hypothesis in the future.

7. References

Bennani, L., Le Guenedal T., Lepetit, F., Ly, L., Mortier, V., Roncally, T. & Sekine, T. (2018). How ESG Investing Has Impacted the Asset Pricing in the Equity Market. *Social Science Research Network*, http://dx.doi.org/10.2139/ssrn.3316862

Black, F. (1972). Capital Market Equilibrium with Restricted Borrowing. *The Journal of Business*, *45*(3), 444. http://doi.org/10.1086/295472

Bodie, Z., Kane, A. & Marcus, A. J. (2020). *Investments.* McGraw-Hill Education

Bolton, P. & Kacperczyk, M. (2021). Do investors care about climate risk*? Journal of Financial Economics*, *142*(2), 517-549. https://doi.org/10.1016/j.jfineco.2021.05.008

Bolton, P., Lazard, Z. H. & Kacperczyk, M. (2022). The Financial Cost of Carbon. *Journal of Applied Corporate Finance*, *34*(2), 17-29. https://doi-org/10.1111/jacf.12502

Carhart, M. M. (1997). On Persistence in mutual fund performance. *The Journal of Finance*, *52*(1), 57-82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x.

Drei, A., Le Guenedal, T., Lepetit, F., Mortier, V., Roncalli, T. & Sekine, T. (2020). ESG Investing in Recent Years: New Insight from Old Challenges. *Social Science Research Network*, http://dx.doi.org/10.2139/ssrn.3683469

Edmans, A. (2011). Does the stock market fully value intangibles? Employee satisfaction and equity prices. *Journal of Financial Economics*, *101*(3), 621-640. https://doi.org/10.1016/j.jfineco.2011.03.021

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. https://doi.org/10.1016/0304-405X(93)90023-5

Fama, E. F. & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1-22. https://doi.org/10.1016/j.jfineco.2014.10.010

Fama, E. F. & MacBeth, J. D. (1973). Risk, Return and Equilibrium: Empirical Tests. *Journal of Political Economy*, *81*(3), 607-636. https://doi.org/10.1086/260061

Giese, G., Nagy, Z. & Lee, L-E. (2021). Deconstructing ESG Ratings Performance: Risk and Return for E, S, and G by Time Horizon, Sector, and Weighting. *The Journal of Portfolio Management*, *47*(3), 94-111. https://doi.org/10.3905/jpm.2020.1.198

Global Sustainable Investment Alliance (2021). Global Sustainable Investment Review 2020. https://www.gsi-alliance.org/wp-content/uploads/2021/08/GSIR-20201.pdf

Gostlow, G. (2022). Pricing Physical Climate Risk in the Cross-Section of Returns. *Social Science Research Network*, http://dx.doi.org/10.2139/ssrn.3501013

Görgen, M., Jacob, A., Nerlinger, M., Riordan, R., Rohleder, M. & Wilkens, M. (2020). Carbon Risk. *Social Science Research Network*, https://doi.org/10.2139/ssrn.2930897

Görgen, M., Jacob, A. & Nerlinger, M. (2021). Get Green or Die Trying? Carbon Risk Integration into Portfolio Management. *The Journal of Portfolio Management*, *47*(3), 77-93. https://doi.org/10.3905/jpm.2020.1.200

Gregory, R. P., Stead, J. G. & Stead, E. (2021). The global pricing of environmental, social and governance (ESG) criteria. *Journal of Finance & Sustainable Investment*, *11*(4), 310-329. https://doi.org/10.1080/20430795.2020.1731786

Gurvich, A. & Creamer, G. G. (2022). Carbon Risk Factor Framework. *The Journal of Portfolio Management Novel Risks*, *48*(10), 148-164. https://doi.org/10.3905/jpm.2022.1.416

Hain, L. I., Kölbel, J. F. & Leippold, M. (2022). Let's get physical: Comparing metrics of physical climate risk. *Finance Research Letters*, *46*, 102406. https://doi.org/10.1016/j.frl.2021.102406

Hong, H. & Kacperczyk, M. (2009). The price of sin: The effects of social norms on markets. *Journal of Financial Economics*, *93*(1), 15-36. https://doi.org/10.1016/j.jfineco.2008.09.001

Hong, H., Weikai Li, F. & Xu, J. (2019). Climate risks and market efficiency. *Journal of Econometrics*, *208*(1), 265-281. https://doi.org/10.1016/j.jeconom.2018.09.015

Hsu, P-H., Li, K. & Tsou C-Y. (2023). The Pollution Premium. *The Journal of Finance*, https://doi.org/10.1111/jofi.13217

Huij, J., Laurs, D., Stork, P. & Zwinkels, R. C.J. (2022). Carbon Beta: A Market-Based Measure of Climate Risk. *Social Science Research Network*, http://dx.doi.org/10.2139/ssrn.3957900

In, S. Y., Park, K. Y. & Monk, A. (2019). Is 'Being Green' Rewarded in the Market?: An Empirical Investigation of Decarbonization and Stock Returns. *Social Science Research Network*, https://ssrn.com/abstract=3020304

Kan, R., Robotti, C., Shanken, J. (2013). Pricing Model Performance and the Two-Pass Cross-Sectional Regression Methodology. *The Journal of Finance*, *68*(6), 2617-2649. https://doi.org/10.1111/jofi.12035

Lintner, J. (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics*, *47*(1), 13. https://doi.org/10.2307/1924119

Lioui, A. (2018). Is ESG Risk Priced? Social Science Research Network, https://ssrn.com/abstract=3285091

Pástor, L., Stambaugh, R. F. & Taylor, L. A. (2021). Sustainable investing in equilibrium. *Journal of Financial Economics*, *142*(2), 550-571. https://doi.org/10.1016/j.jfineco.2020.12.011

Pástor, L., Stambaugh, R. F. & Taylor, L. A. (2022). Dissecting green returns. *Journal of Financial Economics*, *146*(2), 403-424. https://doi.org/10.1016/j.jfineco.2022.07.007

Roberedo, J. C. & Ugolini, A. (2022). Climate transition risk, profitability and stock prices. *International Review of Financial Analysis*, *83*, 102271. https://doi.org/10.1016/j.irfa.2022.102271

Sharpe, W. F. (1964). Capital asset prices: a theory of market equilibrium under conditions of risk. *Journal of Finance*, *19*(3), 425-442. https://doi-org./10.1111/j.1540-6261.1964.tb02865.x

The Economist. (2022, July 21). ESG should be boiled down to one simple measure: emissions. *The Economist*. https://www.economist.com/leaders/2022/07/21/esg-should-be-boiled-down-to-one-simple-measure-emissions

Venturini, A. (2022). Climate change, risk factors and stock returns: A review of literature. *International Review of Financial Analysis*, *79*, 101934. https://doi.org/10.1016/j.irfa.2021.101934

Zivot, E., & Wang, J. (2007). Rolling Analysis of Time Series. In *Springer eBooks* (pp. 313-360). Springer Nature. https://doi.org/10.1007/978-0-387-32348-0_9

8. Appendix

Low and high equally weighted portfolio regression in Europe

Coefficient

ESG low and high performance

Group	Alpha	ERM	SMB	HML	WML	Adj. R2
Low	0.006***	0.717***	0.119	-0.193**	-0.157**	0.766
	(3.87)	(13.34)	(0.83)	(-2.16)	(-2.41)	
High	0.006***	0.724***	-0.224*	-0.078	-0.15**	0.789
	(3.32)	(13.85)	(-1.73)	(-1.04)	(-2.35)	

Group	Alpha	ERM	SMB	HML	WML	LMHESG	Adj. R2
Low	0.006***	0.699***	0.044	-0.114	-0.136**	0.228**	0.774
	(3.60)	(13.68)	(0.30)	(-1.17)	(-2.13)	(1.92)	
High	0.006***	0.725***	-0.223*	-0.081	-0.149**	-0.04	0.788
	(3.18)	(14.43)	(-1.70)	(-0.89)	(-2.24)	(-0.05)	

E and S low and high performance

Group	Alpha	ERM	SMB	HML	WML	LMHE	Adj. R2
Low	0.006***	0.681***	-0.073	-0.135	-0.133**	0.334**	0.771
	(3.60)	(14.03)	(-0.58)	(-1.55)	(-2.20)	(2.30)	
High	0.006***	0.739***	-0.198	-0.037	-0.167**	0.068	0.795
	(3.12)	(14.50)	(-1.49)	(-0.40)	(-2.46)	(0.46)	

Group	Alpha	ERM	SMB	HML	WML	LMHS	Adj. R2
Low	0.006***	0.722***	0.05	-0.105	-0.129**	0.321**	0.783
	(3.40)	(14.14)	(0.39)	(-1.07)	(-1.99)	(2.15)	
High	0.006***	0.704***	-0.244**	-0.073	-0.157**	0.045	0.784
	(3.19)	(14.44)	(-2.01)	(-0.83)	(-2.15)	(0.30)	

CE and RU low and high portfolio performance

Group	Alpha	ERM	SMB	HML	WML	LMHCE	Adj. R2
Low	0.007***	0.678***	-0.108	-0.16*	-0.125*	0.419***	0.768
	(3.69)	(12.67)	(-0.83)	(-1.72)	(-1.94)	(3.19)	
High	0.006***	0.732***	-0.198	-0.025	-0.181***	0.053	0.802
	(3.35)	(15.05)	(-1.48)	(-0.32)	(-2.78)	(0.43)	

Group	Alpha	ERM	SMB	HML	WML	LMHRU	Adj. R2
Low	0.006***	0.712***	0.02	-0.119	-0.186***	0.439***	0.792
	(3.78)	(14.67)	(0.17)	(-1.59)	(-3.19)	(2.93)	
High	0.006***	0.727***	-0.263**	-0.132*	-0.184***	0.095	0.785
	(3.37)	(13.81)	(-2.05)	(-1.77)	(-2.68)	(0.64)	

WF and CO low and high portfolio performance

Group	Alpha	ERM	SMB	HML	WML	LMHWF	Adj. R2
Low	0.006***	0.707***	-0.002	-0.145*	-0.133**	0.305**	0.782
	(3.77)	(14.17)	(-0.01)	(-1.92)	(-2.27)	(2.02)	
High	0.006***	0.742***	-0.198	-0.086	-0.192***	0.088	0.785
	(3.37)	(13.71)	(-1.52)	(-0.99)	(-2.80)	(0.54)	

Group	Alpha	ERM	SMB	HML	WML	LMHCO	Adj. R2
Low	0.006***	0.717***	-0.024	-0.154*	-0.123*	0.293**	0.794
	(3.47)	(14.03)	(-0.17)	(-1.97)	(-1.93)	(2.55)	
High	0.006***	0.683***	-0.205	-0.103	-0.139**	0.112	0.780
	(3.40)	(14.02)	(-1.64)	(-1.39)	(-2.03)	(0.99)	

*This table shows coefficients of the Carhart + respective sustainability factors model. T-statistics are included in the brackets. ***, **, * denote significance at 1%, 5% and 10% respectively. For alphas and beta coefficients, significance statistics are based on two-sided t-tests. Groups are rebalanced annually and are based on the median sustainability scores. Regressions are run with Newey-West standard errors with a lag of 1.*

Low and high equally weighted portfolio regression in the US

Coefficient

ESG low and high performance

Group	Alpha	ERM	SMB	HML	WML	Adj. R2
Low	0.003**	1.013***	0.032	0.186***	-0.012	0.926
	(2.44)	(33.39)	(0.66)	(4.38)	(-0.22)	
High	0.002**	0.954***	-0.088**	0.153***	0.019	0.932
	(2.16)	(37.13)	(-2.29)	(4.64)	(0.49)	

Group	Alpha	ERM	SMB	HML	WML	LMHESG	Adj. R2
Low	0.002**	0.973***	-0.039	0.136***	-0.026	0.276***	0.938
	(1.98)	(35.73)	(-0.80)	(4.04)	(-0.67)	(4.77)	
High	0.002**	0.957***	-0.081*	0.155***	0.018	-0.024	0.932
	(2.12)	(35.49)	(-1.84)	(5.03)	(0.53)	(-0.41)	

E and S low and high performance

Group	Alpha	ERM	SMB	HML	WML	LMHE	Adj. R2
Low	0.002**	0.97***	0.004	0.174***	-0.001	0.271***	0.94
	(2.05)	(32.80)	(0.08)	(5.36)	(-0.01)	(4.50)	
High	0.002*	0.961***	-0.1**	0.163***	0.006	-0.026	0.926
	(1.84)	(34.34)	(-2.24)	(4.88)	(0.17)	(-0.39)	

Group	Alpha	ERM	SMB	HML	WML	LMHS	Adj. R2
Low	0.002	0.967***	-0.029	0.153***	-0.013	0.239***	0.934
	(1.54)	(31.20)	(-0.61)	(4.66)	(-0.24)	(5.00)	
High	0.002**	0.954***	-0.071*	0.171***	0.016	-0.007	0.933
	(2.04)	(37.23)	(-1.79)	(5.05)	(0.48)	(-0.17)	

CE and RU low and high portfolio performance

Group	Alpha	ERM	SMB	HML	WML	LMHCE	Adj. R2
Low	0.002*	0.99***	-0.027	0.136***	-0.005	0.303***	0.943
	(1.77)	(38.91)	(-0.62)	(5.05)	(-0.14)	(5.88)	
High	0.002*	0.942***	-0.086**	0.143***	-0.002	0.069	0.938
	(0.001)	(41.44)	(-2.16)	(5.39)	(0.997)	(1.39)	

Group	Alpha	ERM	SMB	HML	WML	LMHRU	Adj. R2
Low	0.002**	0.993***	-0.024	0.157***	-0.02	0.212***	0.933
	(1.76)	(33.32)	(-0.49)	(3.96)	(-0.44)	(3.19)	
High	0.002**	0.973***	-0.075*	0.167***	0.003	0.022	0.936
	(2.15)	(36.65)	(-1.84)	(5.26)	(0.14)	(0.43)	

WF and CO low and high portfolio performance

Group	Alpha	ERM	SMB	HML	WML	LMHWF	Adj. R2
Low	0.002**	0.997***	-0.036	0.161***	0.028*	0.23***	0.934
	(2.12)	(38.77)	(-0.76)	(4.85)	(0.71)	(4.54)	
High	0.002**	0.958***	-0.067*	0.159***	0.015	0.008	0.938
	(2.20)	(40.85)	(-1.78)	(5.16)	(0.40)	(0.18)	

Group	Alpha	ERM	SMB	HML	WML	LMHCO	Adj. R2
Low	0.002*	0.986***	-0.012	0.181***	0.01	0.196**	0.929
	(1.79)	(32.03)	(-0.25)	(4.73)	(0.27)	(2.24)	
High	0.002**	0.964***	-0.04*	0.194***	-0.014	0.036	0.938
	(2.34)	(47.44)	(-1.06)	(6.00)	(-0.38)	(0.61)	

*This table shows coefficients of the Carhart + respective sustainability factors model. T-statistics are included in the brackets. ***, **, * denote significance at 1%, 5% and 10% respectively. For alphas and beta coefficients, significance statistics are based on two-sided t-tests. Groups are rebalanced annually and are based on the median sustainability scores. Regressions are run with Newey-West standard errors with a lag of 1.*