ERASMUS UNIVERSITY ROTTERDAM

FINANCIAL ECONOMICS

THE EFFECT OF ESG ON ETF RETURNS

AUTHOR

CHRISTIAAN VAN ECK

STUDENT NUMBER: 510075

SUPERVISOR

DR. JAN LEMMEN

SECOND ASSESSOR

DR. ONNO STEENBEEK

DATE

August 25, 2024

Abstract

This study investigates the relationship between Environmental, Social, and Governance (ESG) factors and the returns of Exchange-Traded Funds (ETFs). Time series regressions using sorted portfolios based on the ESG risk score are performed to find the potential existence of a risk premium. Then cross-sectional regressions using the Fama and Macbeth (1973) method are conducted to find the direct effect of the ESG risk score on the returns, followed by the introduction of an ESG dummy to account for ESG terminology in the name of the ETF. Data on the ETFs of investment regions Europe, Japan, North America, and the USA from October 2019 until November 2024 are retrieved from Morningstar and the Fama and French website.

As ETF and ESG investing continues to grow in popularity, this study provides insights into the factors driving returns of ETFs and explores the potential incorporation of an ESG risk premium to the Fama and French models.

The sorted portfolios method found significant positive differences between the alphas in three out of the four regions, suggesting a positive risk premium. However, different cutoff points and a provider specific analysis lead to inconsistencies in the results. The Fama and Macbeth (1973) method found a significant negative direct effect of the ESG risk score on the returns in North America but not in any other region. In addition, ESG terminology in the name of the ETF suggests lower returns of the ETF in Japan but did not show an effect in the other regions. These findings indicate the complexity of the relation between ESG and the return on ETFs, suggesting that investors need to carefully investigate the ESG ETFs before investing.

Keywords: ESG risk score, ETFs, Fama and French, sorted portfolios, cross-sectional regression

JEL Classification: M14, G11, C13, C21, C22

TABLE OF CONTENTS

1 Introduction

Environmental, Social, and Governance (ESG) investing has gained a lot of traction in recent years. ESG refers to a set of criteria that companies use to demonstrate their commitment to protecting the environment (E), practicing social responsibility (S), and maintaining transparent governance (G). ESG Exchange Traded Funds (ETFs) enable investors to invest in a broad range of high ESG-rated companies at once.

Bloomberg states that $30 trillion in 2022 can be attributed to Global ESG assets and is expected to expand to $40 trillion by 2030. In addition, the Institute for Sustainable Investing of Morgan Stanley stated that in 2023 sustainable funds' assets under management (AUM) were up 15% from 2022, reaching $3.4 trillion, which was 7.2% of total AUM. This means that the popularity of sustainable investing is also present in funds and therefore ETFs.

In addition, investing in ETFs is becoming increasingly popular. PWC (2024) estimated that the AUM of the global ETFs achieved a Compound Annual Growth Rate (CAGR) of 18.9% in the past five years.

Furthermore, there is an increasing demand for sustainability reporting. KPMG International (2022) states that in 2022, 79% of the N100 group, which are the leading 100 companies in every country, report on sustainability. In 2024, the EU's Corporate Sustainability Reporting Directive (CSRD) subjects 50.000 companies to report on their sustainability. This increased transparency for investors makes it easier for investors with ESG preferences to invest accordingly. Investors can combine their financial objectives with their concerns about ESG risks.

The question that arises is if these investors are willing to receive a smaller compensation for holding these investments. And if so, does a higher ESG rating result in a negative ESG risk premium for ETFs? Or do investors hold these investments as they expect a higher return relative to other financial instruments, resulting in a positive ESG risk premium? This leads to the following research question:

What is the effect of ESG risk scores on the returns of ETFs?

Thorough literature research found few studies regarding ESG investing in ETFs. Previous research on ETFs was mainly focused on individual ETFs, such as the work of Rompotis (2021), or used ETFs which the researcher condemned as sustainable (Pavlova and Boyrie, 2022). This study used the entire universe of ETFs in the investment regions. Additionally, a cross-sectional approach is performed to study the direct effect of ESG risk on the returns of ETFs, using proxies for the Fama and French factors which has not been done in previous research on ETFs.

This leads to the potential incorporation of the ESG risk premium in the Fama and French models, giving more insight into the factors driving returns in ETFs. This can increase the models' explanatory power. It aligns with work of Halbritter & Dorfleitner (2015), who have explored the integration of ESG factors into the Fama and French models using stocks but did not extend this to ETFs. Furthermore, it adds to the theory of portfolio optimization and diversification showing the potential opportunities for implicating ESG factors, building on the work of Fieberg et al. (2022) who investigated portfolio optimization for sustainable investments focused on stocks. For behavioral finance, this study can reflect the preferences of investors. Are these preferences driven by returns or is there an altruistic ethical or sustainable objective?

A potential risk premium for ESG investing gives investors a comprehensive view of the risks and returns associated with such an investment. This leads to better decision-making based on the preferences of the investors. In addition, institutional investors may be able to trade on this information generating higher returns. If there is a risk premium, then investing, by going long or by going short dependent on the sign of the premium, in ESG ETFs can enhance the long-term risk-adjusted returns.

The ESG subject has been widely studied for multiple financial instruments and geographical scopes. However, the studies have not reached a consensus.

One side argues that firms with better ESG scores perform better in the long run due to factors such as better downside protection and higher operating efficiencies (Kotsantonis et al., 2016; Kassam et al.,2016). The other side argues that ETFs that are not ESG need to be compensated with higher returns to account for the additional risk (Hübel & Scholz, 2019) and that investors are willing to receive lower returns due to their ESG concerns (Pastor et al., 2022).

Friede et al. (2015) conducted a meta-analysis and found that 90% of the research tilted toward a nonnegative risk premium. However, on a portfolio basis, the results are more ambiguous. On this portfolio basis, Pollard et al. (2018) and Kassam et al. (2016) find a positive risk premium, Luo (2022) and Ciciretti et al. (2023) find a negative risk premium, and no relation is found by Jacobsen et al. (2019) Halbritter and Dorfleitner (2015) and Galema et al. (2008).

This study uses Morningstar and the Fama and French website data from October 2019 until April 2024 to construct portfolios based on the ESG risk score in the investment areas of Europe, Japan, North America, and the United States of America (USA). Then the Fama-French Three-Factor-, Fama-French Five-Factor, Carhart (1997)- and a Fama-French Five-Factor plus momentum models are used to find the potential difference in alpha between the low and high portfolios. A cross-sectional regression is performed using the Fama and Macbeth (1973) method to find if the ESG risk score has a direct effect on the performance of the ETFs. Lastly, a dummy variable is added to see if ESG terminology in the name of the ETF affects the performance.

The portfolio analysis found predominantly significantly positive results for Japan, North America, and the USA which suggests a positive risk premium. Europe had mostly insignificant negative results. However, when considering multiple cutoff points for robustness considerations, it is shown that there are inconsistencies in the existence and sign of the risk premium. This suggests that the results are dependent on the portfolios. Furthermore, the analysis on the specific provider using the iShares ETFs suggests that the potential premium may be reliant on the provider.

The cross-sectional regression shows little significance across the models and investment regions. Only North America shows significant negative results in two of its models at the 10% level. This means that if the ESG risk score increases, the returns decrease. Additionally, when the dummy is added, mostly insignificant results are generated. There is some evidence in Japan that ESG terminology in the name leads to lower returns.

Chapter 2 reviews the existing literature on the potential ESG premium in financial instruments, followed by the data collection of the ESG scores, the Fama and French factors, and the financials on ETF level in Chapter 3. Then the four models of the time-series portfolios and the four models of the cross-sectional analysis are discussed in Chapter 4. In Chapter 5, the results are discussed followed by the conclusion, discussion and limitations of this study in Chapter 6.

2 Literature Review

Meta Table 1 shows the findings of multiple studies on the existence of a risk premium in the financial world. It focuses on stocks, bonds, and funds including ETFs. The meta-table reveals ambiguous results, showing mostly positive risk premia from studies published between 2006 and 2018, followed by predominantly negative risk premia in studies published after 2018. There are also over the entire time frame multiple studies that did not find a significant relationship between ESG scores and expected return. However, the periods of the samples of the entire base of studies, no relation, negative and positive risk premia, are overlapping. In addition, the geographical scope of the studies did not change significantly, keeping the focus mainly global and on the US and Europe. Therefore, there is no consensus on the impact of the ESG scores on the expected returns. The studies are split based on the findings, it has a positive risk premium, no relation or a negative risk premium. The studies on Sustainable Responsible Investing (SRI) and ESG are analyzed interchangeably as their objectives and methodologies are overlapping.

2.1 Positive risk premium

Friede et al. (2015) conducted a meta-analysis of over 2200 studies on the relation between the Corporate Financial Performance (CFP) and the ESG criteria. They find that approximately 90% of the studies find a non-negative relationship between CFP and ESG. Next, they divided their research into subgroups. They found that on a portfolio basis, the results were more ambiguous with more mixed findings relative to nonportfolio studies, showing both positive and negative results. Friede et al. (2015) state that this may be due to the overlapping effects of systematic and idiosyncratic risks, construction constraints, and portfolio costs and fees. These factors diminish and potentially erase the ESG alpha. In addition, Friede et al. (2015) looked at regional differences of the ESG-CFP relation. Developed markets, excluding North America, see a smaller percentage of around 50% of positive findings while emerging markets find approximately 70% of the studies to be positive.

Screens are a method employed by Kempf and Osthoff (2007) and Auer (2016). It consists of negative screens, where companies are excluded if they operate in controversial business areas,positive screens, where companies are chosen based on their ESG rating, and best-in-class screens, which is the same method as positive screening but balances the portfolio across industries. Screens improve the overall ESG scores of the portfolio. Kempf and Osthoff (2007) use the KLD ratings on social responsibility. They find that high-rated portfolios outperform the low-rated portfolios. Auer (2016) uses the ESG scores of Sustainalytics. He finds that negative screens allow investors to significantly outperform the passive benchmark. However, positive screens can result in underperformance of the benchmark.

Pollard et al. (2018) and Kassam et al. (2016) construct a portfolio based on the ESG-rating of the individual stocks and compare this with a benchmark portfolio. Pollard et al. (2018) construct two portfolios, a ESG portfolio and a benchmark portfolio. Initially, the portfolios comprise of the same stocks. However, each quarter, high ESG rated stocks replace the worst performing stocks in the ESG portfolio. Kassam et al. (2016) make two portfolios, one that overweights high ESG rated stocks and one that overweights stocks that increased their ESG rating the most, which is known as the momentum strategy. They find that the ESG premium provides a positive alpha. In addition, Pollard et al. (2018) find that it is independent of other risk premia. Kassam et al. (2016) found that the momentum strategy, relative to the MSCI World Index, produced an annual alpha of 2.2%. A significant part of this outperformance came from stock-specific sources which could be indirectly attributed to ESG-related factors.

Hartzmark and Sussman (2019) examine the behavior of mutual fund investors. As of 2016, Morningstar started displaying sustainability rankings. This means that the visibility and accessibility of the sustainability of mutual funds increased. Hartzmark and Sussman (2019) state this increase in visibility and accessibility as a large shock. The worst and best 10% of funds based on sustainability rating saw a significant change in flows while the middle funds saw insignificant results. The worst 10% of funds saw their inflows decrease and the best 10% of funds saw it increase. This means that mutual fund investors see sustainability as a positive attribute.

Kanamura (2021) focused on the financial instrument bonds instead of stocks. He studied the effect of ESG factors on high-yield bond ETF investments during the COVID-19 shocks in March 2020 using a conventional high-yield bond ETF and an ESG high-yield bond ETF. It is found that ESG hedges the downside risk of the price of the ETF and reduces risk overall. In addition, during the shock, the ESG high-yield bond outperformed the conventional high-yield bond in terms of return.

2.2 No relation or negative risk premium

Mutual funds are closely related to ETFs. This makes the insights of the studies on mutual funds helpful. SRI or ethical funds underperform their benchmarks in the studies of Bauer et al. (2006), Renneboog et al. (2008), and Cao et al. (2018). However, Renneboog et al. (2008) find that when they compare the risk-adjusted returns they find no significant difference in performance between the SRI funds and their conventional counterparts. In addition, Bauer et al. (2006) find that when they split up their sample (1992-1996 and 1996-2003), the funds catch up to their benchmarks in terms of return in the second period of their sample and produced similar returns.

Bauer et al. (2007) and Cao et al. (2018) focus on the trading possibilities and strategies of funds. Bauer et al. (2007) find no significant difference in returns between ethical and conventional funds. Moreover, their analysis reveals no evidence that ethical funds employ different screening processes, as there are no significant differences in investment styles. Cao et al. (2018) find that underpriced stocks with bad ESG performance have the highest positive alpha and overpriced stocks with good ESG performance have the highest negative alpha. They state that this is due to the institutional limitations to trade bad ESG stocks because of the preferences of its investors and therefore limiting the full possibility of the efficient market hypothesis. This contradicts the findings of Bauer et al. (2007).

In terms of ETFs, the relation remains uncertain. On one hand, Kanuri (2020) finds that ESG ETFs have slightly underperformed their benchmarks. On the other hand, Rompotis (2022) and Pavlova and Boyrie (2022) find no evidence of a significant relation between ETFs and returns over the whole sample. Rompotis (2022) compared ESG ETFS in the UK from the inception of these ETFs to their benchmarks and the FTSE 100 Index. He finds no evidence of the ESG ETFs outperforming their benchmarks and generating a positive alpha but finds some evidence of ESG ETFs outperforming the FTSE 100 Index. Pavlova and Boyrie (2022) looked at the return of different rated ETFs based on their ESG score before and during the COVID-19 market crash. They find that before the crash, lower-rated ETFs outperformed the market and during the crash the premium was negative and insignificant.

Galema et al. (2008), Halbritter and Dorfleitner (2015), Jacobsen (2019), Luo (2022), and Ciciretti et al. (2023) form portfolios of stocks based on their ESG or SRI ratings. Luo (2022) constructs five value-weighted ESG portfolios and rebalance them yearly. Ciciretti et al. (2023) form two market capitalization weighted portfolios based on ESG score, one with high ESG rated stocks and one with low ESG rated stocks and rebalance them monthly. Then they take the difference between the two portfolios. Luo (2022) and Ciciretti et al. (2023) find a negative ESG risk premium. However, Luo (2022) states that after adding liquidity the effect becomes less pronounced. The results show a significant ESG premium for low-liquidity stocks but insignificant results for high-liquidity stocks. Ciciretti et al. (2023) state that ESG mainly drives the risk premium as a characteristic and not by their betas.

Jacobsen et al. (2019), Halbritter and Dorfleitner (2015) and Galema et al. (2008) find no significant difference in the return performance between the best and worst ESG or SRI portfolios. Jacobsen et al. (2019) form two portfolios, one ESG portfolio and one non-ESG portfolio and were rebalanced monthly. Halbritter and Dorfleitner (2015), form two market capitalization weighted portfolios for three ESG providers and rebalance them yearly. They find a significant influence of several ESG variables. However, investors are not able to exploit this relationship and will therefore not lead to alpha generation. Galema et al. (2008) form two equally weighted portfolios, one who scores good and one who scores bad, for the six SRI dimensions and rebalance them yearly. They find that SRI scores have a significant negative relationship with the book-to-market ratio. As a result, alpha does not capture the SRI effect. They conclude that SRI has a significant effect on stock returns.

Pastor et al. (2022) find that green stocks (best ESG rated stocks) strongly outperform brown stocks (worst ESG rated). By using a media index, which measures the growing concerns about the climate, they find that these shocks have a significant positive relation to the relative performance of green stocks. However, if the shocks are set to zero, Pastor et al. (2022) find that the relative performance of green stocks to brown stocks becomes negative. Positive realized returns are followed by lower expected performance of green stocks. This implies that there is an inverse relation between realized returns and expected returns.

Chava (2014) uses the implied cost of capital to proxy for the expected stock returns of stocks controlling for the environmental profile of a company. He finds that investors expect higher returns for firms with net environmental concerns. However, he finds no significant effect of the number of environmental strengths of a company on the expected returns.

Kumar (2019) focuses on ESG Indexes. He finds significant factors but no significant alpha. In addition, most of the average return of the MSCI USA ESG indexes are explained by the market betas.

Zerbib (2019), Wang et al. (2020) and MacAskill et al. (2021) focus on bonds. MacAskill et al. (2021) conduct a systematic literature review of green premium of bonds in the primary and secondary markets. Only studies that have quantitative results are included which results in a total of 15 papers. In the primary market, MacAskill et al. (2021) find mixed results with spreads of -84 to +213 basis points. The secondary market had more consensus. The spreads were consistently between the -1 and -9 basis points. In line with the systematic literature review of MacAskill et al. (2021), Zerbib (2019) finds a small significant negative green bond premium in the secondary market. However, he states that this difference is so small that it cannot be seen as a substantial discrepancy in valuation between these two bond groups. Wang et al. (2020) find a significantly lower yield of green bonds relative to their synthetically created conventional bonds in the primary market.

2.3 Hypotheses

Arguments for a positive risk premium are shown in the work of Kostantonis et al. (2016) and Kassam et al. (2016). They state that companies with better ESG scores perform better in the long run. Kotsantonis et al. (2016) state that this is due to lower cost of capital, higher level of operating efficiency, and better stakeholder relations which may lead to better financial performance. Kassam et al. (2016) argue that this is due to the downside protection to ESG issues, such as environmental fines and lawsuits. Also, ESG opportunities, such as clean technologies, can swiftly be undertaken.

Arguments for a negative risk premium are mostly based on investor preferences and risk mitigation. Investors may be willing to pay more for their ESG preferences (Pastor et al., 2022; Ciciretti et al., 2023). Therefore, they are willing to forego higher returns as they are socially conscious. In addition, it is expected that investors care about ESG risk. Companies that have high ESG exposures are related to high idiosyncratic risk, which can be explained by the ESG risk factors (Hübel and Scholz, 2019), and are more exposed to systematic market shocks and therefore have higher systematic risk. This means that these companies require a higher rate of return to compensate investors for the increased risk.

As shown, no consensus exists regarding the sign and possibility of an ESG risk premium. The more recent literature finds mostly a negative risk premium as shown in Meta Table 1 Therefore, this study also expects a negative risk premium which leads to the following hypothesis:

Hypothesis 1: ESG ETFs have a negative risk premium.

Furthermore, it is expected that investors will give up more returns if ESG risk is better managed. Simultaneously, investors demand a higher return if there is a higher degree of unmanaged ESG risk.

Hypothesis 2: A higher ESG risk score leads to more expected returns.

Lastly, Investors have limited attention. When buying a stock, investors do not systematically search through the thousands of listed shares until they find a good "buy", but they typically choose from a set of stocks that have caught their attention (Barber and Odean, 2008). This paper assumes this is also the case for ETFs. Investors who want to invest in ESG ETFs search for and invest in ETFs with an annotation to the principles of ESG investing and not focus on the ESG risk score directly. It is expected that adding ESG terminology in the name of the ETF will capture part of the ESG preferences of the investors. This results in that there is a smaller direct effect of the ESG risk score on the returns but rather via the name of the ETF.

Hypothesis 3: Including ESG terminology in the name of the ETF will lead to lower returns.

Meta Table 1: Overview of existing literature

Author(s) (Publication year)	Time period	Region(s)	Method(s)	Financial instrument	ESG-provider	Control variables	Results
Bauer et al. (2006)	November 1992 - April 2003	Germany, UK, US	Carhart (1997) four-factor model	Funds	Morningstar, EIRIS, Ecoreporter	SMB, HML, WML	No relation/negative risk premium: ethical funds first underperformed their benchmarks but caught up and produced similar returns.
Kempf and Osthoff (2007)	1992-2004	US	Portfolio construction on Carhart (1997) four-factor model	Stocks	KLD	SMB, HML, MOM and transaction costs	Positive risk premium: abnormal returns of up to 8.7% per year
Bauer et al. (2007)	January 1994 - January 2003	Canada	Carhart (1997) four-factor model and conditional modelling approach	Funds	Globefund	SMB, HML, WML, Canadian economic information variables,	No relation: no significant difference in returns between the ethical and conventional funds
Renneboog et al. (2008)	January 1991-December 2003	Europe, North America and Asia-Pacific	CAPM, Carhart (1997) four-factor model	Funds	Label of the fund	SMB, HML, WML, Country	Negative risk premium: SRI funds strongly underperform their benchmarks by approximately -2,2 to -6,5%
Galema et al(2008)	June 1992 - July 2006	US	Carhart (1997) four-factor model, Fama and French 3-factor model	Stocks	KLD	SMB, HML, WML, post-ranking beta, BM, SRI score	No relation: no alpha significant different from zero
Auer(2014)	June 2004 - October 2012	Europe	Portfolio construction basde on negative screens and Sharpe ratio analyses	Stocks	Sustainalytics	Firm size, BM ratio, momentum, leverage, industry, and country	Positive risk premium: monthly Sharpe ratio of 0.16 of the ESG screen portfolio and 0.14 of the passive benchmark
Chava (2014)	1992-2007	US	ICC regression	Stocks	KLD	environmental concern, environmental strenght, loan specific features, firms book assets, BM, leverage, volatility, return	No relation/negative risk premium: investors expect higher returns for firms with net environmental concerns but no significant effect of the number of environmental strengths of a company on the expected returns
Friede et al. (2015)	- December 2014	Global	Vote count and meta-analysis study both on portfolio and non-portfolio samples	Equities, bonds, real estate	-	-	Positive risk premium/no relation: 90% of the studies find a non-negative relationship between CFP and ESG.
Halbritter and Dorfleitner (2015)	1991-2012	US	Carhart 4-factor model and the Fama-and-French 3-factor model	Stocks	ASSET4, Bloomberg, and KLD	SMB, HML, WML, post-ranking beta, BM, MOM, ESG pillars score	No relation: no significant difference in the return performance between the best and worst ESG portfolios

Kassam et al. (2016)	February 2007 - March 2015	Global	Portfolio construction based on ESG tilt and ESG momentum strategies	Stocks	MSCI	-	Positive risk premium: annualized alpha 1,1% and 2,2%
Pollard et al. (2018)	January 2007 - January 2017	Global	Portfolio case study	Stocks	MSCI	-	Positive risk premium: consistent postive alpha generation
Cao et al.(2018)	January 2004 - December 2013	US	CAPM, Carhart (1997) four-factor model, Fama and French 3-factor model	Stocks	MSCI	Mispricing measures, ESG scores, size, turnover, idiosyncratic volatility, illiquidity, analyst coverage, stock lending fee, institutional ownership, churn ratio	Negative risk premium: underpriced stocks with bad ESG performance have the highest positive alpha and overpriced stocks with good ESG performance have the highest negative alpha
Hartzmark and Sussman (2019)	March 2016 - January 2017	US	novel natural experiment	Funds	Morningstar	Flow, visits, size, rating	net outflows of $12 billion when categorized as low sustainability and net inflows of $24 billion when categorized as high sustainability
Kumar (2019)	- September 2018	US	CAPM, Carhart (1997) four-factor model, Fama and French 3-factor, 5-factor, and 5-factor plus momentum factor	Index	MSCI	SMB, HML, WML, RMW, CMA	No relation: no alpha significant
Jacobsen et al. (2019)	April 2014 - November 2018	US	Fama-and-French-five-factor plus momentum model	Stocks	MSCI	SMB, HML, WML, RMW, CMA	No relation: risk and return profiles of the ESG and non-ESG stocks are comparable.
Zerbib (2019)	July 2013 - December 2017	Europe and US	matching method, followed by a two-step regression procedure	Bonds	Bloomberg	Liquidiity	Negative risk premium: significant negative green bond premium of -2 basis points
Kanuri (2020)	February 2005 - July 2019	US	CAPM, Carhart (1997) four-factor model, Fama and French 5-factor model	ETFs	Morningstar	SMB, HML, WML, RMW, CMA	Negative risk premium: ESG ETFs slightly underperformed their benchmarks
Wang et al. (2020)	January 2016 - July 2019	China	Univariate and multivariate analysis	Bonds	Bloomberg, Climate Bond Initiative (CBI), Thomson Reuter's Eikon	Rating, size, maturity, callable, puttable, enhancement, enterprice, leverage, ROA, Tangibility, listed, SOE	Negative risk premium: significantly lower yield spread of 34 basis points of green bonds than their synthetically created conventional bonds
Kanamura (2021)	October 2018 - May 2020	Europe/US	PCV and DCC-model	Bonds	Label of the bond	Price, volume, volatility	Positive risk premium: outperformance of ESG high-yield bond to conventional peer

Rompotis (2021)	June 2007 - Dember 2021	UK	Single- and multifactor analysis of the performance of individual ETFs	ETfFs	MSCI	SMB, HML, RMW, CMA	Negative risk premium: underperformance against benchmark but outperformance against FTSE 100 Index
MacAskill et al. (2021)	2007-2019	Global	Systematic literature review	Bonds	-	-	In the primary market mixed results with spreads of -84 to +213 basis points. The secondary market consistently between the -1 and -9 basis points
Pavlova and Boyrie(2022)	November 2019 - May 2020	US	CAPM, Carhart (1997) four-factor model, Fama and French 3-factor, 5-factor, and 5-factor plus momentum factor	ETFs	MSCI	SMB, HML, WML, RMW, CMA, COVID crash	Before COVID crash, lower-rated ETFs outperformed the market and during the crash the premium was negative and insignificant.
Pastor et al. (2022)	November 2012 - December 2020	US	PST model, Fama and French 3-factor, 5-factor, and 5-factor plus momentum factor	Stocks	MSCI	HML, WML, SMB, RMW, CMA, liquidiity, ROE, ME, I/A, climate shocks, environmental score, earnings news	if climate shocks are set to zero, relative performance of green stocks to brown stocks becomes negative
Avramov et al.(2022)	2002-2019	US	CAPM and Fama and French model	Stocks	MSCI KLD, MSCI IVA, Bloomberg, Sustainalytics and RobecoSAM,	ESG score & uncertainty, Size, BM, MOM, illiquidity, gross profitability, corporate investment, leverage, analyst coverage & dispersion	Negative risk premium
Luo (2022)	July 2003 - December 2020	UK	Carhart (1997) four-factor model and multiple Fama-and-French three-factor based models	Stocks	Thomson Reuters	ESG & pillar scores, investment rate, WWindex, PVGO, liquidity	Negative risk premium: lowest quintile outperform the firms in the highest quintile by 0.513% per month for value-weighted returns.
Ciciretti et al. (2023)	July 2003 - June 2020	Asia Pacific ex-Japan, Europe, Japan, and North America	time-series regressions and cross-sectional regression	Stocks	Asset4	SMB, HML, WML, RMW, CMA	Negative risk premium: one standard deviation increase in the ESG scores decreases expected return by 2.73% annually

Note: This table summarizes the related literature concerning the existence of an ESG risk premium. It states the author(s), the publication year, the time period, the region, the method, the financial instruments, the ESG provider, the control variables and the study results.

3 Data

This section describes the sources of the data, as well as the criteria and definition of the variables used in the time series and cross-sectional regressions.

Morningstar offers a comprehensive overview of all ETFs, listed and delisted. Both listed and delisted ETFs have been incorporated to account for survivorship bias, which would have lead to overestimation of the historical performance of the fund. The focus of this study is specifically on passive, equity and long-only ETFs. The Fama and French factors on size, market-to-book ratio, investment, profitability, and momentum are collected from the site of Kenneth French which provides updated data for the following regions: Europe, Japan, Asia Pacific ex Japan, North America, Global developed markets, the USA, and Emerging markets. In addition, it provides information on the risk-free rate and the return of the the market. This information is needed to perform the time series regression.

Emerging markets are not included in this sample. This is because of the lack of ESG Risk Scores in Morningstar. Morningstar computes an ESG Risk Smart Score based on a comparable company analysis and machine learning which is an initial indicator for ESG risk. This leads to a lack of comparability between the different regions in our sample.

Morningstar Direct is used to retrieve the ESG risk scores as well as fund returns, fund size, the price-to-book (P/B) ratio, the investment growth, the Return on Equity (ROE), and inception date of the funds.

This study uses the Sustainalytics' ESG Risk Ratings in the database of Morningstar Direct. These ratings measure the degree of unmanaged ESG risks. This means that it gives insights into the magnitude of the exposure of investors' investments to ESG risks that companies do not fully or sufficiently manage. For ETFs, the Portfolio Corporate Sustainability Score is calculated by taking the asset-weighted averages of Sustainalytics' ESG Risk Ratings. Then, the weighted average of the trailing 12 months of the Portfolio Corporate Sustainability Score is taken. Here portfolios more recently are given more weight than the portfolios further in the past. This leads to more stability while still valuing the current investment decisions more. These steps combined lead to the Historical Corporate Sustainability Score. This is a score between 0 and 100 in theory but most scores are between 0-50 in practice with a lower score meaning less ESG risk. A score of 100 means that all the underlying companies have no ESG risk managed which is not realistic in practice.

The individual ESG pillars, Environmental, Social and Governance, are only available on Portfolio basis. This means that the weighted average of the trailing 12 months is not available. Therefore, this study focuses on the total ESG risk.

Using the Historical Corporate Sustainability Score, which offers an absolute ESG risk score, enables the comparison of ETFs across different dimensions such as industries and countries. These ratings started in October 2018. However, in September 2019, Morningstar switched to the Sustainalytics' ESG Risk Ratings and added buffer rules for increased stability. This makes the old scores incomparable with the new scores. Therefore, and due to the requirement of lagged control variables, this study spans from October 2019 until April 2024. For simplicity, the Historical Corporate Sustainability Score is called the ESG risk score throughout this study.

The monthly market returns of the funds are calculated by taking the end-of-month close price, adjusting this for reinvesting all income and capital-gains distributions during the period, such as dividend, and then taking the difference in the end-of-month close prices. A benefit of this calculation is that it incorporates the fees for the fund as these are deducted from the income the fund generates, before it is distributed to the shareholders.

For the cross-sectional regression ETF level information is needed. The size of the ETF is calculated as the natural logarithm of the market capitalization. The natural logarithm of the price-to-book ratio is used, which is the inverse of the book-to-market ratio. The momentum effect (MOM) is calculated as the 12-month rolling average return of the previous month.

In addition, investment growth is needed. Fama and French (2015) calculate investment growth as the growth in total assets from time *t-1* to *t*. Since this calculation is performed at the stock level, a proxy is required to adjust for factors that may influence total assets beyond organic growth. In this study, total assets at time *t* are calculated as the net asset value (NAV) minus cash inflows plus cash outflows. This adjustment is made to control for investment flows and to capture the changes in the underlying securities' assets. The growth in total assets is then defined as the difference in total assets between t and $t-1$.

Furthermore, operating profitability is needed. Fama and French (2015) use operating profitability to capture the RMW factor. This is calculated as the annual revenues minus the cost of goods sold (COGS) and expenses such as interest, selling, general, and administrative expenses. This is then divided by book equity at the end of year *t-1*. Since ETFs do not generate revenue or incur expenses in the same way individual companies do, a proxy for operating profitability is needed. This study uses the Return on Equity. While the RMW factor used by Fama and French (2015) captures the operating activities of a company, the ROE focuses on the entire company. It is calculated as the net income divided by end-of-year net worth. The value of the previous month is used to prevent look-ahead bias and therefore control for collinearity between the control variables and the monthly returns.

This study uses the ESG terminology used in the paper of Candelon et al. (2021). Candelon et al.

(2021) use lists of words related to SRI terminology from USSIF (2018) and extend this list using the lexical databases from Miller (1995) and Fellbaum (1998). The full list of the terminology is listed in Table 15 in the Appendix.

The steps in the data filtering Table 1 below are shown to compute the final sample. First, the universe of ETFs is filtered on the investment regions and the global broad category equity. Then the duplicates are removed and ETFs that do not have at least one data point in one of the variables of the chosen period. Lastly, the ETFs are required to have continuous data for at least 12 months. These ETFs are not spread evenly over the different regions. To conduct the initial portfolio analysis, at least 5 ETFs are needed per region (and for the additional robustness checks at least 20 ETFs per region). A sample size of 5 ETFs is considered too small as it may not be representative of the overall universe of ETFs in a certain region. Asia Pacific ex Japan, Global Developed, and Global ex USA have 1, 2, and 5 ETFs respectively. Therefore, this study does not cover these regions. This leads to a total sample of 368 ETFs.

Table 1: Data filtering table

Criteria	**# of unique ETFs**
Equity	15.255
Investment Area	3.359
Unique ISINs	1.218
Missing data	1.073
Time period alignment	1.065
Continuous data for at least 12 months	374
Investment regions: Europe, Japan, North America and the USA	368

Notes: This table shows the criteria of this study and the number of unique ETFs that are still in the sample after filtering on this criterion

3.1 Summary Statistics

In Table 2, the summary statistics are shown covering the age of the fund, the ESG risk score, the fund size, the Price-to-Book ratio, the investment growth, the Return on Equity, and the momentum.

The USA has the most ETFs in this sample. It has approximately five times the number of ETFs as Europe in this sample. This can enhance the internal validity of this research as the results as it is a better representation of the market. Therefore, it is expected that the research will have higher (adjusted) R2 in the USA than the other regions.

The mean and the median of the age of the ETFs is the highest in the USA and the lowest in North American. A possible explanation is that North American ETFs consist mostly of stocks from the USA. Therefore, investors will typically choose the conventional established USA ETF given the lack of distinctions between the two, showing low demand for the coverage of other regions in North America. Furthermore, it shows that the North American ETF market is still in its early stages. The funds in the four regions have a similar minimum age. This makes sense as the ETFs are restricted to having data for 12 consecutive months.

ETFs in Europe have on average a lower ESG risk score than the other regions. In addition, the standard deviation of this score is the lowest, showing that there is less spread between the ESG risk scores of the ETFs. On average, Japan and the USA are comparable with North America performing slightly better. This means that the ETFs in Europe manage their ESG risk better on average than the other regions.

The average fund size of the USA and Japan is bigger than that of Europe and North America. In perspective, the average fund size of the ETFs in Europe is approximately 11 times smaller than the average fund in the USA. The biggest funds in the USA and Japan are even approximately 20 and 44 times bigger respectively than the largest fund in Europe. However, using the average can give a distorted image due to the influence of exceptionally large funds. The difference between the regions becomes smaller when using the median. The median of the fund size in Europe and Japan is comparable. However, the USA is still approximately four times larger than the funds in Europe and Japan and seven times larger than the funds in North America.

The Price-to-Book ratio in all regions is on average above one. This ratio is bigger in North America and the USA than in Europe and Japan. This implies that the ETFs in all regions are valued above their book value and even more so in North America and the USA. The investment growth is on average negative in Europe and Japan and positive in North America and the USA. However, the median is positive for all regions. This implies that the underlying companies of the funds grow their assets more often. The Return on Equity is higher in North America and the USA than in Europe and Japan which suggests that the ETFs in these regions are more profitable. The average and the median of the momentum was positive in all regions. In addition, it suggests that the ETFs in North America and the USA had slightly higher returns on average over the past 12 months at a given time than the ETFs in Europe and Japan.

Table 2: Summary statistics of the ETFs per investment region

	Mean	SD	Min	Max	Median	Skew	Kurt	Obs.
Europe								
Age	8.563	6.372	1.144	23.767	6.185	1.263	3.311	34
ESG risk score	20.333	3.082	13.075	31.490	20.376	0.893	7.072	34
Fund Size (million)	402	1,040	0.886	5,880	73.8	4.537	24.173	34
PB	1.976	0.915	0.847	5.030	1.827	1.895	7.019	34
Invest	-0.061	1.385	-4.462	5.556	0.024	0.812	12.255	34
ROE	18.95	5.407	7.624	30.845	18.908	0.160	2.745	34
MOM	0.829	0.386	0.1635	2.009	0.785	0.979	4.541	34
Japan								
Age	9.814	5.723	1.859	22.812	8.457	0.709	2.597	124
ESG risk score	23.089	4.389	12.953	42.195	23.334	0.706	7.346	124
Fund Size (million)	3390	13500	0.908	115000	81.9	5.884	42.189	124
PB	1.417	0.503	0.473	4.329	1.363	2.497	15.259	124
Invest	-0.217	2.252	-18.782	3.997	0.015	-6.193	47.575	124
ROE	10.945	3.581	4.887	32.970	11.277	2.055	14.420	124
MOM	0.698	0.725	-2.297	3.165	0.676	-0.524	6.550	124
North America								
Age	5.513	3.482	1.374	13.544	3.860	0.765	2.425	34
ESG risk score	22.272	5.516	14.286	46.209	21.498	2.568	11.826	34
Fund Size (million)	179	312	3.832	1150	39.6	2.274	7.030	34
PB	4.743	1.996	0.755	9.75	4.199	0.381	2.648	34
Invest	0.144	0.200	-0.028	0.706	0.069	1.814	4.984	34
ROE	23.440	13.477	-21.75	39.646	28.827	-1.575	5.447	34
MOM	1.224	0.856	-0.517	3.702	1.086	0.547	3.760	34
USA								
Age	12.544	6.952	1.747	28.994	11.998	0.097	1.580	174
ESG risk score	23.146	3.645	14.783	42.638	22.266	2.075	10.718	174
Fund Size (million)	4600	23800	0.924	261000	292	8.704	86.554	174
PB	3.808	1.760	0.852	10.238	3.74	1.019	4.251	174
Invest	0.046	1.239	-7.694	6.097	0.037	-1.33	22.608	174
ROE	24.864	8.788	-5.340	45.144	26.87	-0.904	3.658	174
MOM	1.170	0.574	-0.648	3.330	1.163	0.348	4.532	174

Notes: This table presents the mean, standard deviation, minimum, maximum, skewness, kurtosis and the number of observations of the ETFs per investment region. First, the means per ETFs are taken over the entire period of the ETF. Then the ETFs are grouped by region with each ETF weighing the same. The PB is the Price-to-Book ratio, Invest is the investment growth, ROE is the Return on Equity and MOM is the momentum.

4 Methodology

This section describes the different methods that are used, how the variables are constructed, and which assumptions are made. First, the time series model using sorted portfolios is explained and then the cross-sectional regression of Fama and Macbeth (1973) with and without ESG dummy.

4.1 Sorted portfolios: Time series

Sorted portfolios are a method commonly used to magnify the potential effects of a possible anomaly, such as the ESG risk premium. The method of Halbritter and Dorfleitner (2015) is followed but instead of yearly rebalancing, this study, given the availability of monthly ESG risk scores, will rebalance each month.

In each month in the sample, two market capitalization-weighted portfolios are constructed based on the ESG risk scores. These portfolios are the 20% worst and 20% best-performing ETFs based on the ESG risk score. The ETFs are held for one month. Then rebalancing of the portfolios is done at the start of each month. This is done for each region.

Since the inception dates of the ETFs differ, ETFs are added to the sample if the ETF is listed for more than one month, as the returns can only be calculated if one month has passed.

To evaluate the performance of the portfolios, the following models will be made:1) Fama-French Three-Factor Model, 2) Carhart (1997) Model, 3) Fama-French Five-Factor Model, and 4) Fama-French Five-Factor Model plus momentum factor. The excess return of the portfolio is taken by subtracting the risk free rate from the return of the portfolio

$$R_{i,t} - R_{f,t} = \alpha_i + b_i(R_{m,t} - R_{f,t}) + s_i\text{SMB}_t + h_i\text{HML}_t \tag{1}$$

$$R_{i,t} - R_{f,t} = \alpha_i + b_i(R_{m,t} - R_{f,t}) + s_i\text{SMB}_t + h_i\text{HML}_t + w_i\text{WML}_t \tag{2}$$

$$R_{i,t} - R_{f,t} = \alpha_i + b_i(R_{m,t} - R_{f,t}) + s_i\text{SMB}_t + h_i\text{HML}_t + r_i\text{RMW}_t + c_i\text{CMA}_t \tag{3}$$

$$R_{i,t} - R_{f,t} = \alpha_i + b_i(R_{m,t} - R_{f,t}) + s_i\text{SMB}_t + h_i\text{HML}_t + r_i\text{RMW}_t + c_i\text{CMA}_t + w_i\text{WML}_t \tag{4}$$

$R_{i,t} - R_{f,t}$ is the excess return of the portfolio *i* over the risk-free rate at month *t*. $R_{m,t} - R_{f,t}$ is the market's excess return over the risk-free rate at month *t*. The α_i , or the alpha, is the excess return above what is expected using the independent variables in the models. SMB equals small minus big. This is the average return of small stocks in a diversified portfolio minus the average return of big stocks in a diversified portfolio. HML equals high minus low. This is the difference between the average returns of diversified portfolios of high book-to-market stocks and low book-to-market stocks. CMA equals conservative minus aggressive. It is the difference between the average returns of diversified portfolios of stocks with high and low investments. RMW equals robust minus weak. It is the difference between the average returns of diversified portfolios with robust and weak profitability stocks. WML denotes the momentum factor. This captures the difference between the average return of diversified portfolios with past winners and past losers.

The differences in performance between the low and high portfolios in each of the models are taken to evaluate the possible out- or underperformance of the low (best ESG risk score) portfolios relative to the high (worst ESG risk score) portfolios. This is captured by the α_i. Both the α_i and the independent variables are tested using a two-sided t-test.

4.2 Cross-sectional regression: Fama and Macbeth (1973) model

Then a cross-sectional regression is performed. Panel data is used over the whole sample. So not only over the top and worst 20% in terms of ESG risk score. This is done to assert the direct effect of the ESG risk scores on the returns. The Fama and French factors are taken as control variables.

The two-step procedure of Fama and MacBeth (1973) is followed. In the first step, an Ordinary Least Squares (OLS) regression is performed which estimates N cross-sectional regressions. In the second step, the time-series averages of the N cross-sectional-regressions are taken. This approach rules out intra-temporal dependencies by taking the averages over the time periods. However, when both cross-sectional and time series (inter-temporal) dependence is suspected Newey-West standard errors need to be used. The independent variables are tested using a two-sided t-test.

Following Fama and French (1992) the market beta is included to capture the effect of the systematic risk factor on average returns. Since the ETFs are already portfolios, the steps to calculate post-ranking betas based on portfolios sorted on size and pre-ranking betas are not needed. Fama and French (1992) follow these steps as estimates of market betas are more precise for portfolios than individual-level stocks.

As some of the ETFs in the sample are relatively young, the rolling market beta cannot be used as a rolling market beta needs additional data points. Therefore, the market beta of each ETF is calculated by regressing the excess returns over the market excess returns. This market beta is calculated per ETF based on the available data, inception dates, and delisting dates. The calculation period starts no earlier than October 2019 and ends no later than April 2024 given the time frame of this study.

To test hypothesis 3, a dummy is added to each equation which takes on the value of one if the ETF has one of the words included in the ESG terminology of Candelon et al. (2021). This shows if there is an effect of adding these words to the name of the ETF or if investors only care about the ESG rating.

This leads to the following formulas:

$$R_{i,t} - R_{f,t} = \gamma_{0,t} + \gamma_{1,t}\hat{\beta}_{\text{market}} + \gamma_{2,t}\ln(\text{SIZE}_{t-1}) + \gamma_{3,t}\ln(\text{PB}_{t-1}) + \gamma_{4,t}\text{ESG}_{t-1} + (\gamma_{5,t}\text{D.ESG}) + \nu_{i,t} \tag{5}$$

$$\begin{aligned} R_{i,t} - R_{f,t} = \gamma_{0,t} &+ \gamma_{1,t}\hat{\beta}_{\text{market}} + \gamma_{2,t}\ln(\text{SIZE}_{t-1}) + \gamma_{3,t}\ln(\text{PB}_{t-1}) \\ &+ \gamma_{4,t}\text{MOM}_{t-1} + \gamma_{5,t}\text{ESG}_{t-1} + (\gamma_{6,t}\text{D.ESG}) + \nu_{i,t} \end{aligned} \tag{6}$$

$$\begin{aligned} R_{i,t} - R_{f,t} = \gamma_{0,t} &+ \gamma_{1,t}\hat{\beta}_{\text{market}} + \gamma_{2,t}\ln(\text{SIZE}_{t-1}) + \gamma_{3,t}\ln(\text{PB}_{t-1}) \\ &+ \gamma_{4,t}\text{INVEST}_{t-1} + \gamma_{5,t}\text{ROE}_{t-1} + \gamma_{6,t}\text{ESG}_{t-1} + (\gamma_{7,t}\text{D.ESG}) + \nu_{i,t} \end{aligned} \tag{7}$$

$$\begin{aligned} R_{i,t} - R_{f,t} = \gamma_{0,t} &+ \gamma_{1,t}\hat{\beta}_{\text{market}} + \gamma_{2,t}\ln(\text{SIZE}_{t-1}) + \gamma_{3,t}\ln(\text{PB}_{t-1}) + \gamma_{4,t}\text{MOM}_{t-1} \\ &+ \gamma_{5,t}\text{INVEST}_{t-1} + \gamma_{6,t}\text{ROE}_{t-1} + \gamma_{7,t}\text{ESG}_{t-1} + (\gamma_{8,t}\text{D.ESG}) + \nu_{i,t} \end{aligned} \tag{8}$$

$R_{i,t} - R_{f,t}$ is the excess return of the individual ETF over the risk-free rate at month *t*. $R_{m,t} - R_{f,t}$ is the market's excess return over the risk-free rate at month *t*. The term $\hat{\beta}_{\text{market}}$ represents the estimated market beta of the individual ETF. The other variables include SIZE which is the size of the ETF and PB, which refers to the the price-to-book ratio. MOM denotes the momentum of the ETF, and INVEST is the investment growth. ROE is the Return On Equity and ESG is the ESG risk score. D.ESG is the Dummy for the ESG terminology in the name of the ETF.

4.3 Assumptions

Since both models follow an Ordinary Least Squares model. The assumptions are the same. Stated below are problems that may arise in the models, how it is tested for and possible ways to mitigate the problems.

Autocorrelation refers to the similarity between the error term of a variable and the lagged error term of this variable. This violates the assumption that the errors should be independently distributed and not correlated. To test for autocorrelation, the Woolridge test for serial correlation (2002) is performed for the cross-sectional regression, and the Durbin Watson test for the time series regression. The null hypothesis states that there is no first-order autocorrelation. The null hypothesis is rejected in all models and regions for both the time-series and the cross-sectional regressions.

Heteroskedasticity is the change in the spread of the residuals which means that the error variances are not constant. This violates the assumption that the residuals have a constant variance. To test for this, three versions of the Breusch-Pagan (1979) and Cook-Weisberg (1983) are performed. The null hypothesis in the three tests state that there is no heteroskedasticity. For completeness, heteroskedasticity is tested in an OLS regression for both the time series and the cross-sectional. The null hypothesis is predominantly rejected in the time series and cross-sectional regression.

To account for the heteroskedasticity and autocorrelation, the Newey-West standard errors are used for both the time series and the cross-sectional regression. To calculate the number of lags needed, the rule of thumb is used. This is m= $T^{\frac{1}{4}}$, where T is the number of months. The data consists of 55 months, resulting in 3 lags.

Multicollinearity is the situation where the variables in the model are correlated. This makes it hard to isolate and estimate the effect of each variable. To test this the variance inflation factors (VIF) are calculated. All VIF values are well below the threshold of 10 across the models and regions, indicating that multicollinearity is not an issue in this research. VIF values above 10 are considered problematic (James et al. 2013).

5 Results

This section describes the results of the sorted portfolios time series, cross-sectional and the cross-sectional plus the ESG dummy regressions.

5.1 ESG Portfolio's: Premium

In Table 3 the results for alpha of the Fama and French portfolio regression are shown. In Table 10 and 11 in the appendix the results including the factors are shown.

The first hypothesis states that a negative risk premium is expected. This means that it is expected that the high portfolios, which are the worst 20% ETFs in terms of ESG risk score in the sample, outperform the low portfolios, which are the best 20% ETFs in terms of ESG risk score.

In Europe, the difference between the alpha is insignificant for all four models. This means that the hypothesis cannot be rejected that the difference between the alpha is different from zero. There is some evidence in models 1 and 2 of outperformance of the market at the 1% level for both the low and high portfolios. The model's adjusted R-squared is on average 60% for model 1 and 61% for model 4 which shows that adding extra factors does not lead to a big increase in R-squared.

In Japan, the opposite is the case. There is a significantly positive difference of the alphas between the low and high portfolios. This suggests a positive risk premium. The results are significant at a 1% level for all four models. The premium is as high as 0.106% per month (1.28% yearly) and as low as 0.089% per month (1.07% yearly). The difference between the model's adjusted R-squared was 4% as it increased from 59% to 63% between model 1 and model 4.

In North America, more disparity between the different models can be seen. Models 1 and 2 are significantly positive at the 5% level and models 3 and 4 are significantly positive at the 1% level, suggesting a positive risk premium. The positive risk premium differs in magnitude between models 1 and 2 and models 3 and 4. The effect is as small as 0.357% per month (4.294% yearly) and as big as 0.835% per month (10.024% yearly). This effect is between 4.0 and 9.4 times as large in North America as in Japan. Adding more factors to the model increases the model's adjusted R-squared from 28 to 37%. This shows that the model is not as effective in explaining the returns as it does for other regions.

In the USA, significant results are found in all 4 models at the 1% level. Significant outperformance of the low portfolio relative to the high portfolio is seen which supports a positive risk premium. The difference between the alphas of the low and high portfolios varies between 0.147% (1.766% yearly) and 0.208% (2.501% yearly). The model's adjusted R-squared was around the same percentage of 66% throughout the models.

This shows that the results are not in line with the hypothesis as predominantly significantly positive results are found across the investment regions.

In addition, both the 20% worst and the 20% best portfolios in all regions predominantly outperform the market, as defined by the Fama and French website, in all four of the models. This shows superior performance of these ETFs relative to the market, regardless of the ESG risk score in this sample period. This advocates for investors to invest in the ETFs. It should be noted that there is a possibility that the benchmarks from the Fama and French website do not accurately reflect the benchmarks of the individual ETFs.

Table 3: Alpha of the ESG risk portfolios: time-series

		(1) FF3	(2) Carhart	(3) FF5	(4) FF5 + MOM
Europe					
	Low	0.268***	0.412***	0.094	0.234**
		(0.102)	(0.102)	(0.104)	(0.112)
	High	0.383***	0.498 ***	0.117	0.215*
		(0.102)	(0.115)	(0.108)	(0.113)
	Dif	-0.116	-0.086	-0.022	0.019
		(0.110)	(0.120)	(0.124)	(0.131)
Japan					
	Low	0.188***	0.222***	0.180***	0.212***
		(0.038)	(0.034)	(0.034)	(0.032)
	High	0.082	0.117**	0.077	0.123**
		(0.050)	(0.047)	(0.050)	(0.048)
	Dif	0.106***	0.105***	0.103***	0.089***
		(0.031)	(0.032)	(0.031)	(0.032)
North America					
	Low	0.475**	0.473**	0.735***	0.705***
		(0.200)	(0.202)	(0.185)	(0.189)
	High	0.117	0.091	-0.086	-0.130
		(0.225)	(0.240)	(0.267)	(0.285)
	Dif	0.358**	0.381**	0.821***	0.835***
		(0.159)	(0.153)	(0.167)	(0.180)
USA					
	Low	0.495***	0.495***	0.520***	0.513***
		(0.029)	(0.029)	(0.025)	(0.025)
	High	0.348***	0.345***	0.317***	0.305***
		(0.047)	(0.047)	(0.049)	(0.049)
	Dif	0.147***	0.149***	0.209***	0.208***
		(0.045)	(0.045)	(0.050)	(0.049)

*Notes: This table presents the results of the Fama-French Three-Factor Model (1), the Carhart (1997) Model (2), the Fama-French Five-Factor Model (3) and the Fama-French Five-Factor plus momentum Model (4) focused on the monthly alpha over the period October 2019 to April 2024 of the investment regions: Europe, Japan, North America and the USA. The low (high) portfolio consists of the 20% best (worst) performing ETFs in terms of ESG risk score. The portfolios are weighted based on the market capitalization of the ETFs. "Dif" states the difference in alpha between the two portfolios. The excess return is the dependent variable and is in percentage points. The standard errors are estimated using the Newey and West (1987) procedure. *p<0.1, **p<0.05, ***p<0.01*

5.2 ESG Portfolio's: Factor returns

The market betas of both the low and high portfolios are positive and significant on a 1% level for all regions. In addition, the differences between the low and high portfolios are significant on a 1% level across the regions, except for North America where no significance is seen. In this investment area, the market beta is not significantly different from zero which indicates no differences between the market betas of the low and high portfolios. In Europe and the USA, the low portfolios have higher market betas than the high portfolios but are not greater than one. This means that the low portfolios are more volatile than the high portfolios but not as volatile as the underlying market. In Japan, the low portfolio has a lower market beta than the high portfolio and is therefore less volatile.

The signs of the difference of the betas of the size effect (SMB) are significantly negative in all models across all regions except Europe (showing no significance at all), at a 5% or a 1% level. This indicates that the low portfolios have more exposure to large-cap stocks than the high portfolios. Europe shows no significant size betas for the low and high portfolios. In Japan, the size betas for both the low and high portfolios are negative, which means that they both are tilted more toward large-cap stocks. In the USA, the size betas of the low portfolios are negative, and those of the high portfolios are positive. So not only are high portfolios less tilted to large-cap stocks than low portfolios but they are even predominantly tilted toward small-cap stocks. North America gives significantly positive results for the high portfolio but no significance for the low portfolio.

The signs of the difference of the betas of the value effect (HML) are predominantly significantly negative across the regions at a 1% level. This implies that the low portfolios are more exposed to low book-to-market stocks than high portfolios. The low portfolios are predominantly negative, and the high portfolios are predominantly positive. So, the low portfolios are more exposed to low book-to-market or growth stocks and high portfolios are more exposed to high book-to-market or value stocks.

The investment factor (CMA) shows ambiguous results. Positive differences between the betas are found in Europe and Japan. This is significant on the 1% level for Japan but not significant for Europe. This means that the betas of the low and high portfolios do not significantly differ from each other in Europe. North America and the USA give significantly negative results between the betas on a 1% level (except for model 4 for the USA which is at the 5% level). So, in Europe, although insignifican, and Japan, the low portfolios are more exposed to conservative firms than aggressive firms relative to the high portfolios, and in North America and the USA the opposite. Both the low and high portfolios are predominantly significantly negative except for North America. Here the high portfolio is significantly positive. This means that the ETFs in this portfolio are more exposed to conservative firms.

In Europe, North America, and the USA, the difference between the betas of the operating profitability factor (RMW) is significantly negative at a 1% level. This means that the low portfolios are more exposed to 'weak' firms with low profitability than the high portfolios. Japan shows the opposite results. The betas are significantly positive at a 1% level which indicates that the low portfolios are more exposed to 'robust' profitability with high profitability than the high portfolios. Europe, Japan, and the USA show predominantly positive results for both the low and high portfolios. Europe is the only region that shows significant results for both portfolios at a 1% level. The low portfolio in North America shows significant negative results at the 5% level but significant positive results for the high portfolio at a 1% level. Overall, there is a strong indication that both low and high portfolios are tilted towards ETFs with stocks generating robust profitability.

In addition to the CMA factor, the momentum factor (WML) reveals ambiguous results. Japan shows a significantly positive difference between the betas of the low and high portfolios at a 1% level for model 4. This indicates that the low portfolios load up more on past winners relative to high portfolios. Europe, North America, and the USA show a negative difference which means that the high portfolios load up more on past winners. However, there are no significant results found. Both the portfolios in Europe and Japan are significantly negative at a 1% level. Therefore, both portfolios tilt more towards past losers. The USA shows significantly positive betas at a 1% level for model 4 but not for model 2. These portfolios tilt more toward past winners. North America shows no significant results for both portfolios.

5.3 Robustness check: Other cutoff points

To test for the robustness of the results, several new cutoff points are used: 5%, 10%, 25%, and 50%, which can be seen in Table 4. The expectation is that the effect between the low and high portfolios becomes less pronounced when the cutoff points become larger as the differences between the portfolios decrease. In addition, it is expected that there is no significant sign change (from plus to minus or minus to plus).

However, the cutoff points show ambiguous results. In Japan, the difference between the alphas of the low and high portfolios is significantly negative at the 5%, 10%, and 25% cutoff points but becomes significantly positive when the cutoff point is 50%.

In Europe, the models show significantly negative results at the 5% and 10% level for models 2,3, and 4 difference but not significant in any other case, both in different models and different cutoff points.

In North America, the difference is significantly positive for the 50% cutoff point at the 5% and 1% level for models 1 and 2 and models 3 and 4 respectively. On the other hand, model 4 at the 25% cutoff point shows a significantly negative difference at the 1% level. All other models and cutoff points show insignificant results.

In the USA, the difference is significantly negative at the 5% cutoff point at varying levels for all models but becomes significantly positive at the 25% cutoff points at the 5% levels for models 3 and 4. The 50% cutoff point has significant positive results for all models at the 1% level.

These inconsistencies between models and cutoff points questions the robustness of the models and the possibility of an effect of the ESG risk Score on the returns.

Table 4: ESG portfolios: Difference in alphas dependent on the cut-off points

	Cutoff point	(1) FF3	(2) Carhart	(3) FF5	(4) FF5 + Mom
Europe					
	5%	-0.496 (0.385)	-0.466 (0.454)	-0.317 (0.441)	-0.259 (0.513)
	10%	-0.247 (0.252)	-0.307 (0.279)	-0.087 (0.264)	-0.137 (0.286)
	25%	0.010 (0.081)	0.036 (0.089)	0.077 (0.090)	0.109 (0.097)
	50%	-0.058** (0.026)	-0.061** (0.029)	-0.057** (0.029)	-0.059* (0.030)
Japan					
	5%	-0.887*** (0.148)	-0.837*** (0.156)	-0.881*** (0.146)	-0.852*** (0.152)
	10%	-0.613*** (0.112)	-0.590*** (0.113)	-0.611*** (0.114)	-0.617*** (0.112)
	25%	-0.066*** (0.021)	-0.079*** (0.021)	-0.067*** (0.022)	-0.091*** (0.022)
	50%	0.329*** (0.016)	0.331*** (0.016)	0.323*** (0.016)	0.325*** (0.017)
North America					
	5%	0.024 (0.616)	0.018 (0.561)	0.432 (0.517)	0.324 (0.503)
	10%	0.192 (0.415)	0.234 (0.415)	0.696 (0.447)	0.729 (0.455)
	25%	-0.185 (0.125)	0.197 (0.128)	0.042 (0.123)	-0.066*** (0.021)
	50%	0.177** (0.074)	0.184** (0.074)	0.193*** (0.059)	0.173*** (0.057)
USA					
	5%	-0.343* (0.176)	-0.381** (0.176)	-0.344* (0.199)	-0.416** (0.198)
	10%	-0.089 (0.090)	-0.082 (0.087)	-0.078 (0.101)	-0.081 (0.100)
	25%	0.043 (0.040)	0.056 (0.038)	0.083** (0.041)	0.103** (0.040)
	50%	0.062*** (0.014)	0.067*** (0.013)	0.115*** (0.014)	0.120*** (0.013)

*Notes: This table presents the results of the Fama-French Three-Factor Model (1), the Carhart (1997) Model (2), the Fama-French Five-Factor Model (3) and the Fama-French Five-Factor plus momentum Model (4) focused on the difference in monthly alpha between the low (best) and high (worst) portfolio over the period October 2019 to April 2024 of the investment regions: Europe, Japan, North America and the USA. The excess return is the dependent variable and is in percentage points. Multiple cutoff points are introduced: 5%, 10%, 25% and 50%. The low (high) portfolio consists of a certain percentage of the best (worst) performing ETFs in terms of ESG risk score based on the cutoff points. The standard errors are estimated using the Newey and West (1987) procedure.*p<0.1, **p<0.05, ***p<0.01*

5.4 Robustness check: iShares ETFs

Additional analysis is needed to test if the results are region or provider specific. Blackrock is the biggest provider of ETFs under the brand name of iShares and has ETFs in Europe, the USA and Japan. However, given the small sample size in Japan, only Europe and the USA will be studied. In this section, the sorted portfolio analysis and a region analysis are performed using only the ETFs of iShares in Europe and the USA.

First, a region analysis is performed. This is to test if the ETFs differ in performance between the investment areas without considering ESG factors. Then, the sorted portfolios method, as explained in section 4.1, is performed. In the sample, there are 10 ETFs in Europe and 15 ETFs in the USA provided by iShares. The 50% cutoff point is used, as the sample of iShares ETFs is relatively small. This increases the statistical power of the analysis and reduces the probability of the results being biased by individual ETF-specific characteristics.

The iShares ETFs in Europe have a median ESG risk score of 20.189 and a mean of 20.431. The ETFs in the USA have slightly worse ESG risk scores with a median of 21.534 and 22.293. This means that it is expected that the ETFs in the USA will have higher returns than the ETFs in Europe, ceteris paribus.

The region analysis, in Table 5, shows that the iShares ETFs in the USA have a significant positive alpha between 0.286% monthly (3,432% yearly) and 0.316% monthly (3,792% yearly) on average at the 1% significance level. Europe has varying levels of significance and has a positive alpha between 0.281% monthly (3,372% yearly) and 0.488% monthly (5,856% yearly) on average. This shows discrepancies between the regions showing that iShares ETFs in Europe outperform the iShares ETFs in the USA on a risk-adjusted basis. This difference in alpha between the two regions may be due to the worse ESG risk scores in the USA than in Europe suggesting a possible positive risk premium.

When looking at the sorted portfolios of the two investment areas in Table 6, the alpha is negative in all four models in both regions showing significance at the 1% level for all models in Europe and models 3 and 4 in the USA. The alpha is between -0.396% monthly (-4,746% yearly) and -0.541% monthly (-6,492% yearly) on average in Europe and between -0.123% monthly (-1,476% yearly) and - 0.205% monthly (-2,456% yearly) on average in the USA. This suggests a negative risk premium which is in line with hypothesis 1.

When comparing the results of the iShares ETFs with the results of the total sample, so including the other ETF providers, in Europe the negative risk premium is larger and significant at the 1% level for the iShares ETFs. In the USA the premium turns negative and significant for models 3 and 4 when using iShares ETFs. This suggests that the choice of a certain provider of ETFs can impact the returns when choosing an ESG investing strategy or that ESG is not the driver of the alpha.

Table 5: iShares: Total alpha of Europe and the USA

		(1) FF3	(2) Carhart	(3) FF5	(4) FF5 + Mom
Alpha	**Europe**	0.481***	0.488***	0.282*	0.280
		(0.160)	(0.180)	(0.161)	(0.180)
	USA	0.288***	0.286***	0.316***	0.305***
		(0.088)	(0.088)	(0.086)	(0.085)
Observations	Europe	453	453	453	453
	USA	670	670	670	670
R2	Europe	0.698	0.698	0.717	0.717
	USA	0.805	0.806	0.811	0.812
Adjusted R2	Europe	0.698	0.696	0.714	0.713
	USA	0.804	0.805	0.809	0.810

*Note: This table presents a summary of the results of the Fama-French Three-Factor Model (1), the Carhart (1997) Model (2), the Fama-French Five-Factor Model (3) and the Fama-French Five-Factor plus momentum Model (4) over the period October 2019 to April 2024 of the investment regions: Europe and the USA using only iShares ETFs. The excess return is the dependent variable and is in percentage points. The observations (Obs.) as well as the R2 and adjusted (Adj. R2 are included per portfolio and for the difference between the portfolios. The standard errors are estimated using the Newey and West (1987) procedure. *p<0.1, **p<0.05, ***p<0.01*

Table 6: iShares ESG Portfolios: Difference in alpha in Europe and in the USA

		(1) FF3	(2) Carhart	(3) FF5	(4) FF5 + Mom
Alpha	**Europe**	-0.040***	-0.396***	-0.541***	-0.541***
		(0.135)	0.147	(0.137)	(0.146)
	USA	-0.123	-0.125	-0.201***	-0.205***
		(0.085)	(0.085)	(0.073)	(0.073)
Observations	Europe	453	453	453	453
	USA	670	670	670	670
R2	Europe	0.624	0.624	0.640	0.640
	USA	0.286	0.289	0.376	0.377
Adjusted R2	Europe	0.621	0.621	0.636	0.636
	USA	0.282	0.284	0.372	0.372

*Note: This table presents a summary of the results of the Fama-French Three-Factor Model (1), the Carhart (1997) Model (2), the Fama-French Five-Factor Model (3) and the Fama-French Five-Factor plus momentum Model (4) over the period October 2019 to April 2024 of the investment regions: Europe and the USA using only iShares ETFs. The difference of the portfolios within the two regions are used to calculate the alpha. The portfolios are weighted based on the market capitalization of the ETFs. The excess return is the dependent variable and is in percentage points. The observations (Obs.) as well as the R2 and adjusted (Adj. R2 are included per portfolio and for the difference between the portfolios. The standard errors are estimated using the Newey and West (1987) procedure. *p<0.1, **p<0.05, ***p<0.01*

5.5 Fama and Macbeth (1973) model: direct ESG effect

Table 7 presents the estimates for the cross-sectional regression. This is done to test for hypothesis 2 that a higher ESG risk score leads to more returns.

In line with the literature, there is not one consensus. Japan exhibits positive and negative betas for the ESG risk score. Europe and the USA find only positive betas for all models. North America shows significantly negative coefficients for models 5 and 6 at the 10% level. It shows that an increase of the ESG risk score by one leads on average to a decrease in returns of 0.096% per month (or 1.16% yearly) for the North American ETFs. This means that a higher (worse) ESG risk score leads to lower returns which is not in line with the hypothesis. This is also the investment region with the highest adjusted R-squared on average with 68% in model 8. The closest following region is Europe with an adjusted R-squared average of 47%.

The overall results suggest that there is little to no evidence of the ESG risk score affecting the returns, both positive and negative. The size factor is negative in Europe and North America but shows no significance. Although insignificant, this is in line with the theory of Fama and French (1993) who state that smaller companies have higher average returns, which translates to a negative relation. Japan and the USA have negative coefficients with Japan showing significance at the 1% level. This means that bigger ETFs have higher average returns.

The price-to-book effect has mixed results. Japan has negative betas with significance varying between the 5% and 10% level. Europe, North America, and the USA have insignificantly positive betas. A lower price-to-book ratio is the same as a higher book-to-market ratio. A negative beta is therefore in line with the work of Fama and French (1993) as it suggests that high book-to-market companies (value stocks) or low price-to-book companies outperform low book-to-market (growth stocks) or high price-to-book. A negative beta, in the sample of this research, shows that ETFs with higher book-to-market ratios exhibit higher returns than ETFs with lower book-to-market ratios.

The betas for investment are predominantly positive, except for Japan which shows negative betas. This suggests that higher investments lead to higher returns. It is the opposite of the expectations following Fama and French (1993) who state that conservative investing leads to higher returns than aggressive investing. However, the betas are mostly insignificant.

The betas for ROE are insignificantly negative in Europe. This suggests that higher returns accompany a lower ROE. Japan, North America, and the USA show positive coefficients, with Japan showing significance at the 1% level. This is in line with Fama and French (1993). They state that companies with robust profitability outperform companies with weak profitability.

The momentum factor is positive for Japan, North America, and the USA. Japan and the USA exhibit significant results at mostly the 1% level. North America shows no significance. Past winners tend to outperform past losers. This follows the framework of Fama and French (1993). Europe has negative betas across the two models but are insignificant.

Table 7: Fama and MacBeth (1973) cross-sectional regression

	Europe				**Japan**			
	5	6	7	8	5	6	7	8
Constant	-1.015	-1.131	-0.102	-0.356	-2.119 **	-2.077 **	-1.635 *	-1.633*
	(1.759)	(2.440)	(1.808)	(1.784)	(0.862)	(0.847)	(0.821)	(0.815)
Beta	0.165	0.261	0.023	0.189	0.702	0.685	0.440	0.431
	(0.901)	(0.865)	(0.947)	(1.007)	(0.647)	(0.644)	(0.671)	(0.672)
LnSIZE	-0.003	-0.017	-0.075	-0.088	0.243***	0.230***	0.194***	0.187 ***
	(0.189)	(0.295)	(0.223)	(0.236)	(0.060)	(0.058)	(0.059)	(0.058)
LnPB	0.385	0.344	0.566	0.561	-0.759*	-0.704 **	-1.180**	-1.103 **
	(0.303)	(0.427)	(0.440)	(0.442)	(0.403)	(0.396)	(0.471)	(0.464)
Invest			0.259*	0.259*			-0.026	-0.032
			(0.152)	(0.137)			(0.028)	(0.025)
ROE			-0.016	-0.017			0.067***	0.062***
			(0.022)	(0.022)			(0.021)	(0.021)
MOM		-0.033		-0.035		0.097***		0.080**
		(0.072)		(0.074)		(0.035)		(0.034)
ESG	0.066	0.078	0.073	0.088	0.003	0.004	-0.018	-0.016
	(0.046)	(0.059)	(0.050)	(0.054)	(0.018)	(0.018)	(0.016)	(0.016)
Observations	1,332	1,332	1,332	1,332	5,449	5,449	5,449	5,449
Avg. R2	0.520	0.556	0.604	0.634	0.353	0.368	0.388	0.400
Avg. Adj. R2	0.419	0.433	0.463	0.473	0.325	0.333	0.347	0.353

	North America				**USA**			
	5	6	7	8	5	6	7	8
Constant	1.728	1.792	1.686	3.262	-1.602	-1.545	-1.381	-1.323
	(2.357)	(2.325)	(2.951)	(2.9390	(1.939)	(1.949)	(1.902)	(1.908)
Beta	0.227	0.220	0.019	-0.431	0.696	0.637	0.629	0.561
	(0.914)	(0.927)	(0.974)	(1.129)	(0.562)	(0.559)	(0.550)	(0.547)
LnSIZE	-0.029	-0.034	-0.153	-0.253	0.073	0.067	0.071	0.063
	(0.084)	(0.090)	(0.186)	(0.215)	(0.114)	(0.115)	(0.108)	(0.109)
LnPB	1.223	1.203	1.281	0.746	0.235	0.233	0.240	0.235
	(0.988)	(1.013)	(0.910)	(0.925)	(0.429)	(0.429)	(0.432)	(0.431)
Invest			0.696*	0.822			0.015	0.012
			(0.399)	(0.676)			(0.015)	(0.015)
ROE			0.027	0.036			0.001	0.002
			(0.031)	(0.037)			(0.011)	(0.012)
MOM		0.048		0.193		0.083***		0.084***
		(0.046)		(0.124)		(0.022)		(0.024)
ESG	-0.096*	-0.098*	-0.063	-0.061	0.047	0.045	0.039	0.038
	(0.055)	(0.054)	(0.043)	(0.043)	(0.047)	(0.047)	(0.047)	(0.047)
Observations	988	988	988	988	8,042	8,042	8,042	8,042
Avg. R2	0.716	0.740	0.787	0.804	0.407	0.413	0.438	0.444
Avg. Adj. R2	0.632	0.634	0.677	0.676	0.390	0.392	0.413	0.416

*Notes: This table presents the results of the monthly Fama and MacBeth (1973) regressions using the factors of the the Fama-French Three-Factor Model (1), the Carhart (1997) Model (2), the Fama-French Five-Factor Model (3) and the Fama-French Five-Factor plus momentum Model (4) focused on the individual ETFs over the period October 2019 to April 2024 of the investment regions: Europe, Japan, North America, and the USA.BETA, lnSIZE, lnPB, Invest, ROE and MOM are control variables with regard to beta, market capitalization, price-to-book ratio, investment growth, return on equity, and average returns of the last 12 months. The independent variable is the ESG which is the ESG risk score. The excess return is the dependent variable and is in percentage points. The total number of observations as well as the R2 and adjusted R2 are given per model. The standard errors are estimated using the Newey and West (1987) procedure. *p<0.1, **p<0.05, ***p<0.01.*

5.6 ESG Dummy

Table 8 shows a summary of the results of adding a dummy variable to the models. The full model is in the appendix under Table 10. This is done to test for hypothesis 3 which states that including ESG terminology in the name of the ETF will lead to less returns.

In the North American sample, no ETFs contain any of the terms specified by the SRI terminology. Therefore, this region is excluded from the regression including the "ESG Dummy".

Europe and the USA show insignificant results. In addition, Europe has both positive and negative betas while the USA only has positive betas. A positive beta is the opposite of what is expected. Japan shows significant negative betas for all models at the 1% level. By adding ESG terminology in the name of the ETF the monthly excess returns decrease by 0.285% (3,420% yearly) in model 1 and this effect decreases to 0.192% (2,304% yearly) in model 4. This suggests that adding ESG terminology in the name leads to fewer returns, supporting hypothesis 3.

The models in all regions have a relatively low adjusted R-squared with all models showing adjusted R-squared of below the 50%. This means that the model does not fit the data well.

Table 8: Fama and MacBeth (1973) cross-sectional regression including ESG dummy

		5	6	7	8
Europe	ESG score	0.073 (0.048)	0.083 (0.060)	0.084* (0.050)	0.096* (0.051)
	ESG dummy	-0.018 (0.121)	-0.001 (0.190)	0.034 (0.115)	0.006 (0.136)
Japan	ESG score	0.002 (0.018)	0.003 (0.018)	-0.019 (0.015)	-0.017 (0.016)
	ESG dummy	-0.285*** (0.072)	-0.252*** (0.0651)	-0.208*** (0.070)	-0.192*** (0.068)
USA	ESG score	0.047 (0.047)	0.046 (0.047)	0.039 (0.047)	0.039 (0.048)
	ESG dummy	0.108 (0.166)	0.133 (0.169)	0.105 (0.165)	0.130 (0.169)

*Notes: This table presents a summary of the results of the monthly Fama-French Three-Factor Model (1), the Carhart (1997) Model (2), the Fama-French Five-Factor Model (3) and the Fama-French Five-Factor plus momentum Model (4) focused on the ESG score and ESG dummy of the individual ETFs over the period October 2019 to April 2024 of the investment regions: Europe, Japan, and the USA. The independent variable is the ESG which is the ESG risk score. The excess return is the dependent variable and is in percentage points. The dummy for the ESG terminology in the name of the ETF is denoted ESG Dummy. The standard errors are estimated using the Newey and West (1987) procedure. *p<0.1, **p<0.05, ***p<0.01*

6 Discussion & Conclusions

This section concludes the research and puts it into the existing literature showing papers that are in line or oppose the findings of this study. Also, the limitations of this study and suggestions for future research are discussed.

This study reviews the effect of the ESG risk scores on the excess returns of ETFs. Previous research on this effect on financial instruments has not reached one consensus on the existence or sign of a potential effect. In addition, the research on the effect of ESG risk scores on ETFs is limited, although both ESG and ETFs are becoming increasingly popular. It was mostly focused on individual ETFs or ETFs labeled as 'sustainable'. In addition, a cross-sectional regression to see the direct effect of the ESG score was missing. Studying this effect leads to more insights into the factors driving returns in ETFs. Therefore, the question that was studied in this thesis was: *"What is the effect of ESG scores on the returns of ETFs?"*

Multiple hypotheses are examined to answer this question. It is expected that the ESG ETFs have a negative risk premium. In addition, a higher ESG risk score leads to more expected returns. Lastly, including ESG terminology in the name of the ETF will lead to lower returns. To test these hypotheses, a time series regression using sorted portfolios and a cross-sectional regression were conducted across four regions: Europe, Japan, North America, and the USA. The time series regression followed multiple variations of the Fama and French model: Fama-French Three-Factor Model, the Carhart (1997) Model, the Fama-French Five-Factor Model, and the Fama-French Five-Factor plus momentum Model. The cross-sectional analysis followed the method of Fama and Macbeth (1973).

The findings of this study show a disparity in the effect of the ESG risk score on the returns between regions and cutoff points and therefore the possible existence of a premium. This aligns with the diverse results observed in the current literature on the subject.

A positive risk premium at the 20% cutoff point used in this study for Japan, North America, and the USA was found, In addition, The cross-sectional analysis shows limited evidence of a negative relation between the ESG risk score and returns in North America. These findings suggest that the low ESG risk portfolios outperform the high ESG risk portfolios, and that a better ESG score leads to higher returns. These results align with the findings of Kassam et al. (2016) and Kotsantonis et al. (2016). They argue that companies with better ESG scores can outperform companies with low ESG scores due to factors like lower costs of capital and better stakeholder relations. Kassam et al. (2016) used the Fama and French factors and state that a significant part of this outperformance found was not explained by these factors, suggesting that this may be attributable to the ESG factors. Kassam et al. (2016) also construct portfolios that overweight either high ESG rated stocks or stocks that increased their ESG rating the most. However, Kassam et al. (2016) compare it to an index benchmark.

In addition, Kanamura (2021) found that ESG factors can hedge downside risks and improve returns during market shocks, which could explain the observed positive premium. Since our study includes the COVID-19 crisis and therefore a market shock this reflection is highly relevant. This shows some possible similarities between the bond market and the ETF market. Also, it is in line with the work of Pollard et al. (2018). The method used by Pollard et al. (2018), which replaces the worst performing stocks with the best ESG rated stocks and comparing it to an index benchmark, is different than the method of this study. However, the idea is the same, namely constructing an ESG portfolio, although not initially, as the portfolios start with the same stocks, and looking at the relative performance of this ESG portfolio. The findings are also in line with the general outcome of the meta-analysis of Friede et al. (2015), who found a nonnegative relation between ESG and performance in 90% of the literature. Investors should, therefore, consider investing more in the best ESG rated ETFs as it results in higher returns and leads to better business practices on ethical, sustainable and social levels. Based on these findings, the value of the ESG factor in the ETFs is not fully captured by the Fama and French frameworks in these regions. To increase the model's explanatory power, incorporating the ESG factors is recommended.

On the other hand, Europe does not show a significant difference in alpha between the low and high ESG risk portfolios. When considering different cutoff points as a robustness check, the premium is inconsistent showing no relation, a positive relation, or a negative relationship depending on the cutoff. However, when a single provider of ETFs namely iShares is used the sorted portfolios find a negative risk premium in all four models. This implies that the sign of the difference between the high and low portfolios is provider specific which undermines the ESG risk score as a driver of the returns. Furthermore, the cross-sectional regressions predominantly show no significant relation between the ESG risk score and the excess returns. This aligns with the study of Rompotis (2022), who also focused on ETFs and found no evidence of the ESG ETFs outperforming their benchmarks However, the research of Rompotis (2022) is based on individual ESG ETFs in the United Kingdom compared to their underlying benchmark, which makes the comparability and generalizability of the results difficult.

Halbritter and Dorfleitner (2015) also find no significant alpha generation even when incorporating multiple ESG scores providers. Although, Halbritter and Dorfleitner (2015) focus on stocks and use a different ESG provider than this study, the methodological framework is mostly the same. The time span of their study is from 1992 to 2012 which is in the early stages of ESG investing and focused on the USA only. So, although ESG investing is becoming increasingly popular, the underlying drivers of returns seems to not be influenced. This suggests that there is no difference in returns between the best rated ESG ETFs and the worst rated ESG ETFs. ESG may not be a strong determinant for financial performance. So, while ESG factors are inherently good by promoting business practices that are ethical, sustainable and socially responsible, it does not necessarily result in superior performance. However, investors with ESG preferences can do so without giving up or receiving more returns

The results contradict the findings of Ciciretti et al. (2023) who find a negative risk premium. They also form two market capitalization weighted portfolios, but with stocks, based on the ESG score and rebalance them monthly. However, the ESG score provider is Asset4 and the study spans over a longer time frame namely from 2003 until 2020. In addition, it accounts for the ESG bias that result in positive realized returns that are generated by an increase in demand.

There is evidence in Japan that adding ESG terminology to the name of an ETF leads to lower returns, suggesting a potential negative risk premium associated with ESG investing in that region. Consequently, ETF providers for Japan need to be cautious when branding as it leads to possible overvaluation of the ETF. Investors can use this information when choosing an investment strategy and potentially short the ETFs with ESG terminology in the name to generate a profit. The other regions show no relation. This makes the potential effect specific to Japan and, therefore, not generalizable.

Certain limitations of this study should be taken into consideration. Meta-table 1 shows that several ESG providers were used in the studies, all having their unique rating system. This questions the comparability and reliability across the different providers and the different studies. In addition, it is important to know which rating agencies are commonly used by investors to have reliable results. With the EU's CSRD introduction in 2024, ESG reporting in Europe will be generalized. This makes it easier to study the effect of ESG on the returns. Data availability led to a significant decrease in sample size which can result in a decrease in representativeness making generalizability harder. In addition, the period of this study, due to the availability of the ESG risk score of Morningstar from October 2019, covers the COVID-19 crisis, which leads to a higher chance of extremes in the sample. The proxies capturing the Fama and French factors might not sufficiently align which can potentially lead to measurement errors. For example, the ROE as a proxy for the RMW factor may not fully capture the operational aspects of profitability targeted by Fama and French, as the ROE focuses on the whole company. In addition, the dummy variable for the SRI terminology may not cover the full impact of marketing and branding of ETFs with the ESG terminology. Multicollinearity can also be still a problem. Although the VIF value was below 10, there are multiple advocates of setting the VIF value below 5 or even lower. Johnston et al. (2018) state that a VIF greater or equal to 2.5 already indicates considerable collinearity. Also, increasing investor demand may drive up the prices and therefore the returns in the short run. This can influence the results of the effect of the ESG risk score on the returns and not capture the effect of the

ESG risk score in the long run. Therefore, there needs to be differentiated between short term realized returns and long term expected returns.

Future research should aim to use ESG scores over longer periods from different ESG providers. This should help reduce the inconsistencies that arise from the different ratings methods that the different providers use and therefore increase the explanatory power of the model. The study of Pastor et al. (2022) can also be used to give more insights in the relation between ESG risk and returns. By using a media index to account for climate concerns, it may be possible to control for the increasing investors' demand which drives up the prices of the ETFs. Additionally, it is needed to test if the proxies for the Fama and French factors are representable. If not, this could lead to biased and misleading results, either showing a possible relation that does not exist or ignoring a possible insightful relation.

To conclude, significantly positive differences between the alphas are found in three out of the four regions suggesting a positive risk premium. However, inconsistencies are found when considering different cutoff points and focusing on a single ETF provider. In addition, there is little evidence of a direct effect of the ESG risk score on the returns. ESG terminology in the name of the ETF suggests lower returns of the ETF in Japan but did not show an effect in the other regions. Additional research regarding ESG ETFs is needed to get better insights in the relation between ESG and the returns of ETFs.

REFERENCES

Auer, B. R. (2016). Do socially responsible investment policies add or destroy European stock portfolio value? *Journal of Business Ethics*, 135, 381-397.

Avramov, D., Cheng, S., Lioui, A., & Tarelli, A. (2022). Sustainable investing with ESG rating uncertainty. *Journal of Financial Economics*, 145(2), 642-664.

Barber, B. M., Odean, T., & Zhu, N. (2008). Do retail trades move markets?. *The Review of Financial Studies*, 22(1), 151-186.

Bauer, R., Koedijk, K., & Otten, R. (2005). International evidence on ethical mutual fund performance and investment style. *Journal of banking & finance*, 29(7), 1751-1767.

Bauer, R., Derwall, J., & Otten, R. (2007). The ethical mutual fund performance debate: New evidence from Canada. *Journal of Business Ethics*, 70, 111-124.

Bauer, R., Otten, R., & Rad, A. T. (2006). Ethical investing in Australia: Is there a financial penalty? *Pacific-Basin Finance Journal*, 14(1), 33-48.

Berg, F., Koelbel, J. F., & Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, 26(6), 1315-1344.

Candelon, B., Hasse, J. B., & Lajaunie, Q. (2021). Esg-washing in the mutual funds industry? from information asymmetry to regulation. *Risks*, 9(11), 199.

Cao, J., Titman, S., Zhan, X., & Zhang, W. E. (2019). ESG preference and market efficiency: Evidence from mispricing and institutional trading. *SSRN Electronic Journal.*

Chava, S. (2014). Environmental externalities and cost of capital. *Management science*, 60(9), 2223-2247.

Ciciretti, R., Dalò, A., & Dam, L. (2023). The contributions of betas versus characteristics to the ESG premium. *Journal of Empirical Finance*, 71, 104-124.

Engle, R. F., Giglio, S., Kelly, B., Lee, H., & Stroebel, J. (2020). Hedging climate change news. *The Review of Financial Studies*, 33(3), 1184-1216.

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

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

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3), 607-636.

Fieberg, C., Poddig, T. &Varmaz, A. (2022). Portfolio optimization for sustainable investments. *Annals of Operations Research*, 1-26.

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.

Galema, R., Plantinga, A., & Scholtens, B. (2008). The stocks at stake: Return and risk in socially responsible investment. *Journal of Banking & Finance*, 32(12), 2646-2654.

Giese, G., Lee, L. E., Melas, D., Nagy, Z., & Nishikawa, L. (2019). Foundations of ESG investing: How ESG affects equity valuation, risk, and performance. *Journal of Portfolio Management*, 45(5), 69-83.

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

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

Hübel, B., & Scholz, H. (2020). Integrating sustainability risks in asset management: the role of ESG exposures and ESG ratings. *Journal of Asset Management*, 21(1), 52-69.

Jacobsen, B., Lee, W., & Ma, C. (2019). The alpha, beta, and sigma of ESG: Better beta, additional alpha? *Journal of Portfolio Management*, 45(6), 6-15.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). New York: springer.

Johnston, R., Jones, K., & Manley, D. (2018). Confounding and collinearity in regression analysis: a cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. *Quality & Quantity*, 52, 1957-1976.

Kanamura, T. (2021). Risk mitigation and return resilience for high yield bond ETFs with ESG components. *Finance Research Letters*, 41, 101866.

Kanuri, S. (2020). Risk and return characteristics of environmental, social, and governance (ESG) equity ETFs. *Journal of Index Investing*, 11(2), 66-75.

Kempf, A., & Osthoff, P. (2007). The effect of socially responsible investing on portfolio performance. *European Financial Management*, 13(5), 908-922.

Kotsantonis, S., Pinney, C., & Serafeim, G. (2016). ESG integration in investment management: Myths and realities. *Journal of Applied Corporate Finance*, 28(2), 10-16.

KPMG International. (2022). *Survey of Sustainability Reporting 2022: Big Shifts, Small steps*. https://www.morganstanley.com/content/dam/msdotcom/en/assets/pdfs/MSInstituteforSustainableInvesting-SustainableRealityFY2023-Final.pdf

Kumar, R. (2019). ESG: Alpha or Duty?. *Journal of Index Investing*, 9(4), 58-66.

Luo, D. (2022). ESG, liquidity, and stock returns. *Journal of International Financial Markets, Institutions and Money*

MacAskill, S., Roca, E., Liu, B., Stewart, R. A., & Sahin, O. (2021). Is there a green premium in the green bond market? Systematic literature review revealing premium determinants. *Journal of Cleaner Production*, 280, 124491.

Morgan Stanley Institute for Sustainable Investing. (2024). *Sustainable Reality*. https://www.morganstanley.com/content/dam/msdotcom/en/assets/pdfs/MSInstituteforSustainableInvesting-SustainableRealityFY2023-Final.pdf

Morningstar (2023). *Measuring ESG risk in 3 undervalued small-cap and emerging-markets stocks*. https://www.morningstar.com/sustainable-investing/measuring-esg-risk-3-undervalued-small-cap-emerging-markets-stocks

Muñoz, F., Vargas, M., & Marco, I. (2014). Environmental mutual funds: Financial performance and managerial abilities. *Journal of Business Ethics*, 124, 551-569.

Nagy, Z., Kassam, A., & Lee, L. E. (2016). Can ESG add alpha? An analysis of ESG tilt and momentum strategies. *Journal of Investing*, 25(2), 113-124.

Pástor, Ľ., Stambaugh, R. F., & Taylor, L. A. (2022). Dissecting green returns. *Journal of Financial Economics*, 146(2), 403-424.

Pavlova, I., & de Boyrie, M. E. (2022). ESG ETFs and the COVID-19 stock market crash of 2020: Did clean funds fare better? *Finance Research Letters*, 44, 102051.

Pollard, J. L., Sherwood, M. W., & Klobus, R. G. (2018). Establishing ESG as risk premia. *Journal of Investment Management*, 16(1), 32-43.

PWC (2024). *ETFs 2028: Shaping the future.* https://www.pwc.com/gx/en/industries/financial-services/publications/etfs-2028-shaping-the-future.html

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

Rompotis, G. G. (2022). The ESG ETFs in the UK. *Journal of Asset Management*, 23(2), 114-129.

Wang, J., Chen, X., Li, X., Yu, J., & Zhong, R. (2020). The market reaction to green bond issuance: Evidence from China. *Pacific-Basin Finance Journal*, 60, 101294.

Zerbib, O. D. (2019). The effect of pro-environmental preferences on bond prices: Evidence from green bonds. *Journal of banking & Finance*, 98, 39-60.

APPENDIX

This appendix shows the full models of the time series of Fama-French Three-Factor Model (1), the Carhart (1997) Model (2), the Fama-French Five-Factor Model (3) and the Fama-French Five-Factor plus momentum Model. In addition it shows the full models of the cross sectional regression including the ESG Dummy following the Fama and Macbeth (1973) model. Furthermore, the terminology is listed that is used to construct the ESG Dummy.

Table 10: ESG risk portfolios: time-series: Europe and Japan

		Europe				Japan			
		(1) FF3	(2) Carhart	(3) FF5	(4) FF5 + Mom	(1) FF3	(2) Carhart	(3) FF5	(4) FF5 + Mom
Alpha	Low	0.268***	0.412***	0.094	0.234**	0.188***	0.222***	0.180***	0.212***
		(0.102)	(0.102)	(0.104)	(0.112)	(0.038)	(0.034)	(0.034)	(0.032)
	High	0.383***	0.498 ***	0.117	0.215*	0.082	0.117**	0.077	0.123**
		(0.102)	(0.115)	(0.108)	(0.113)	(0.050)	(0.047)	(0.050)	(0.048)
	Dif	-0.116	-0.086	-0.022	0.019	0.106***	0.105***	0.103***	0.089***
		(0.110)	(0.120)	(0.124)	(0.131)	(0.031)	(0.032)	(0.031)	(0.032)
MKT	Low	0.835***	0.775***	0.781***	0.730***	0.701***	0.677***	0.664***	0.641***
		(0.020)	(0.026)	(0.019)	(0.026)	(0.011)	(0.013)	(0.012)	(0.014)
	High	0.668***	0.621***	0.589 ***	0.553***	0.785***	0.760***	0.739***	0.706***
		(0.024)	(0.024)	(0.019)	(0.025)	(0.017)	(0.018)	(0.019)	(0.020)
	Dif	0.166***	0.154***	0.191***	0.176***	-0.084***	-0.083***	-0.075***	-0.065***
		(0.021)	(0.023)	(0.022)	(0.0252)	(0.008)	(0.008)	(0.010)	(0.009)
SMB	Low	-0.075	-0.049	-0.095	-0.071	-0.199***	-0.217***	-0.145***	-0.155***
		(0.064)	(0.065)	(0.061)	(0.062)	(0.022)	(0.023)	(0.022)	(0.023)
	High	0.005	0.025	-0.011	0.005	-0.159***	-0.178***	-0.066**	-0.081**
		(0.082)	(0.077)	(0.071)	(0.065)	(0.033)	(0.033)	(0.033)	(0.033)
	Dif	-0.079	-0.074	-0.084	-0.077	-0.041**	-0.040**	-0.079***	-0.075***
		(0.065)	(0.062)	(0.065)	(0.063)	(0.018)	(0.018)	(0.019)	(0.019)
HML	Low	-0.175***	-0.244***	0.146**	0.067	-0.207***	-0.228***	-0.058***	-0.075***
		(0.030)	(0.036)	(0.065)	(0.065)	(0.012)	(0.012)	(0.016)	(0.014)
	High	0.582***	0.527***	1.077***	1.022***	0.263***	0.241***	0.431 ***	0.405***
		(0.031)	(0.046)	(0.067)	(0.072)	(0.016)	(0.015)	(0.023)	(0.021)
	Dif	-0.757***	-0.771***	-0.931***	-0.955***	-0.470***	-0.469***	-0.489***	-0.481***
		(0.035)	(0.047)	(0.063)	(0.069)	(0.008)	(0.007)	(0.015)	(0.015)
CMA	Low			-0.347***	-0.333***			0.010	-0.057*
				(0.086)	(0.082)			(0.032)	(0.029)
	High			-0.471***	-0.461***			-0.348***	-0.445***
				(0.121)	(0.120)			(0.040)	(0.034)
	Dif			0.124	0.128			0.359***	0.389***
				(0.109)	(0.109)			(0.030)	(0.031)
RMW	Low			0.459***	0.431***			0.387***	0.320***
				(0.099)	(0.102)			(0.038)	(0.033)
	High			0.778***	0.758***			0.061	-0.036
				(0.110)	(0.111)			(0.065)	(0.061)
	Dif			-0.319 ***	-0.327***			0.326***	0.356***
				(0.093)	(0.093)			(0.036)	(0.038)
WML	Low		-0.165***		-0.149***		-0.126 ***		-0.115***
			(0.040)		(0.039)		(0.018)		(0.018)
	High		-0.131**		-0.105*		-0.132***		-0.166***
			(0.060)		(0.055)		(0.024)		(0.023)
	Dif		-0.034		-0.044		0.006		0.051***
			(0.055)		(0.054)		(0.012)		(0.012)
Obs.		527	527	527	527	2,154	2,154	2,154	2,154
R2	Low	0.794	0.801	0.806	0.812	0.810	0.818	0.818	0.824
	High	0.740	0.745	0.772	0.774	0.665	0.674	0.675	0.687
	Dif	0.603	0.604	0.611	0.612	0.590	0.590	0.623	0.626
Adj. R2	Low	0.792	0.799	0.804	0.810	0.810	0.817	0.818	0.824
	High	0.739	0.743	0.769	0.772	0.665	0.673	0.674	0.687
	Dif	0.601	0.601	0.607	0.608	0.589	0.589	0.622	0.625

*Note:This table presents the results of the Fama-French Three-Factor Model (1), the Carhart (1997) Model (2), the Fama-French Five-Factor Model (3) and the Fama-French Five-Factor plus momentum Model (4) over the period October 2019 to April 2024 of the investment regions: Europe and Japan. The low (high) portfolio consists of the 20% best (worst) performing ETFs in terms of ESG risk score. The portfolios are weighted based on the market capitalization of the ETFs. "Dif" states the difference in alpha between the two portfolios. The excess return is the dependent variable and is in percentage points. The observations (Obs.) as well as the R2 and adjusted (Adj.) R2 are included per portfolio and for the difference between the portfolios. The standard errors are estimated using the Newey and West (1987) procedure. *p<0.1, **p<0.05, ***p<0.01*

Table 11: ESG risk portfolios: time-series: North America and the USA

		North America				USA			
		(1) FF3	(2) Carhart	(3) FF5	(4) FF5 + Mom	(1) FF3	(2) Carhart	(3) FF5	(4) FF5 + Mom
Alpha	Low	0.475**	0.473**	0.735***	0.705***	0.495***	0.495***	0.520***	0.513***
		(0.200)	(0.202)	(0.185)	(0.189)	(0.029)	(0.029)	(0.025)	(0.025)
	High	0.117	0.091	-0.086	-0.130	0.348***	0.345***	0.317***	0.305***
		(0.225)	(0.240)	(0.267)	(0.285)	(0.047)	(0.047)	(0.049)	(0.049)
	Dif	0.358**	0.381**	0.821***	0.835***	0.147***	0.149***	0.209***	0.208***
		(0.159)	(0.153)	(0.167)	(0.180)	(0.045)	(0.045)	(0.050)	(0.049)
MKT	Low	0.827***	0.828***	0.833***	0.838***	0.843***	0.843***	0.835***	0.838***
		(0.031)	(0.034)	(0.034)	(0.034)	(0.005)	(0.006)	(0.006)	(0.006)
	High	0.854***	0.863***	0.821***	0.830***	0.650***	0.652 ***	0.632***	0.639***
		(0.077)	(0.082)	(0.064)	(0.067)	(0.013)	(0.014)	(0.012)	(0.013)
	Dif	-0.027	-0.035	0.011	0.009	0.193***	0.191***	0.203***	0.199***
		(0.078)	(0.084)	(0.065)	(0.067)	(0.012)	(0.013)	(0.011)	(0.012)
SMB	Low	0.121***	0.122**	-0.144**	-0.117	-0.038***	-0.038***	-0.063***	-0.047***
		(0.041)	(0.047)	(0.073)	(0.078)	(0.010)	(0.010)	(0.010)	(0.011)
	High	0.655***	0.676***	0.880***	0.922***	0.136***	0.1422***	0.179***	0.208***
		(0.095)	(0.094)	(0.167)	(0.172)	(0.012)	(0.012)	(0.016)	(0.014)
	Dif	-0.535***	-0.554***	-1.025***	-1.038***	-0.174***	-0.180***	-0.241 ***	-0.255***
		(0.109)	(0.117)	(0.162)	(0.175)	(0.0137)	(0.014)	(0.016)	(0.017)
HML	Low	-0.408***	-0.408***	-0.056	-0.029	-0.270***	-0.270***	-0.212 ***	-0.208***
		(0.047)	(0.0472)	(0.088)	(0.092)	(0.007)	(0.007)	(0.008)	(0.008)
	High	0.035	0.038	-0.083	-0.043	0.387***	0.388***	0.394 ***	0.401***
		(0.057	(0.059	(0.089)	(0.092)	(0.008)	(0.008)	(0.015)	(0.015)
	Dif	-0.443***	-0.446***	0.028	0.014	-0.657***	-0.658***	-0.606***	-0.609***
		(0.038)	(0.038)	(0.105)	(0.095)	(0.007)	(0.007)	(0.013)	(0.014)
CMA	Low			-0.546***	-0.584 ***			-0.119***	-0.131***
				(0.122)	(0.128)			(0.018)	(0.019)
	High			0.174	0.117			-0.062***	-0.083***
				(0.147)	(0.140)			(0.023)	(0.025)
	Dif			-0.720***	-0.701***			-0.057***	-0.047**
				(0.138)	(0.121)			(0.020)	(0.020)
RMW	Low			-0.246*	-0.217			-0.006	0.003
				(0.138)	(0.144)			(0.017	(0.016)
	High			0.373**	0.417**			0.105***	0.123***
				(0.151)	(0.162)			(0.026)	(0.026)
	Dif			-0.619***	-0.634***			-0.112***	-0.120***
				(0.174)	(0.189)			(0.030)	(0.031)
WML	Low		0.004		0.058		0.001		0.024***
			(0.046)		(0.049)		(0.007)		(0.008)
	High		0.051		0.087		0.011		0.045***
			(0.076)		(0.078)		(0.013)		(0.013)
	Dif		-0.047		-0.029		-0.011		-0.020*
			(0.065)		(0.061)		(0.012)		(0.012)
Obs.		396	396	396	396	3,229	3,229	3,229	3,229
R2	Low	0.629	0.629	0.655	0.656	0.874	0.874	0.877	0.877
	High	0.572	0.572	0.582	0.584	0.738	0.738	0.740	0.741
	Dif	0.285	0.286	0.382	0.382	0.656	0.656	0.660	0.656
Adj. R2	Low	0.626	0.625	0.651	0.651	0.874	0.874	0.876	0.877
	High	0.568	0.568	0.577	0.577	0.737	0.737	0.740	0.741
	Dif	0.279	0.278	0.374	0.373	0.655	0.655	0.659	0.659

*Note:This table presents the results of the Fama-French Three-Factor Model (1), the Carhart (1997) Model (2), the Fama-French Five-Factor Model (3) and the Fama-French Five-Factor plus momentum Model (4) over the period October 2019 to April 2024 of the investment regions: North America and the USA. The low (high) portfolio consists of the 20% best (worst) performing ETFs in terms of ESG risk score. The portfolios are weighted based on the market capitalization of the ETFs. "Dif" states the difference in alpha between the two portfolios. The excess return is the dependent variable and is in percentage points. The observations (Obs.) as well as the R2 and adjusted (Adj.) R2 are included per portfolio and for the difference between the portfolios. The standard errors are estimated using the Newey and West (1987) procedure. *p<0.1, **p<0.05, ***p<0.01*

Table 12: iShares ESG Portfolios: Difference in alpha in Europe and in the USA

		(1) FF3	(2) Carhart	(3) FF5	(4) FF5 + Mom
Alpha	**Europe**	-0.040*** (0.135)	-0.396*** 0.147	-0.541*** (0.137)	-0.541*** (0.146)
	USA	-0.123 (0.085)	-0.125 (0.085)	-0.201*** (0.073)	-0.205*** (0.073)
MKT	**Europe**	0.097*** (0.023)	0.096*** (0.027)	0.072** (0.025)	0.072*** (0.026)
	USA	0.131*** (0.016)	0.138*** (0.017)	0.131*** (0.017)	0.134*** (0.017)
SMB	**Europe**	-0.124 (0.095)	-0.123 (0.095)	-0.135 (0.096)	-0.135 (0.097)
	USA	0.062** (0.025)	0.077*** (0.027)	0.180*** (0.038)	0.192*** (0.046)
HML	**Europe**	-0.628*** (0.038)	-0.630*** (0.042)	-0.427*** (0.076)	-0.427*** 0.080
	USA	-0.145*** (0.017)	-0.146*** (0.018)	-0.291*** (0.028)	-0.289*** (0.027)
CMA	**Europe**			-0.159 (0.124)	-0.159 (0.123)
	USA			0.224*** (0.034)	0.216*** (0.035)
RMW	**Europe**			0.346*** (0.120)	0.346*** 0.120
	USA			0.146*** (0.048)	0.153*** (0.051)
WML	**Europe**		-0.005 (0.049)		0.000 (0.048)
	USA		0.028* (0.017)		0.018 (0.021)
Observations	Europe	453	453	453	453
	USA	670	670	670	670
R2	Europe	0.624	0.624	0.640	0.640
	USA	0.286	0.289	0.376	0.377
Adjusted R2	Europe	0.621	0.621	0.636	0.636
	USA	0.282	0.284	0.372	0.372

*Note:This table presents the results of the Fama-French Three-Factor Model (1), the Carhart (1997) Model (2), the Fama-French Five-Factor Model (3) and the Fama-French Five-Factor plus momentum Model (4) over the period October 2019 to April 2024 of the investment regions: Europe and the USA using only iShares ETFs. The low (high) portfolio consists of the 20% best (worst) performing ETFs in terms of ESG risk score. The portfolios are weighted based on the market capitalization of the ETFs. "Dif" states the difference in alpha between the two portfolios. The excess return is the dependent variable and is in percentage points. The observations (Obs.) as well as the R2 and adjusted (Adj.) R2 are included per portfolio and for the difference between the portfolios. The standard errors are estimated using the Newey and West (1987) procedure. *p<0.1, **p<0.05, ***p<0.01*

Table 13: iShares: Total alpha of Europe and the USA

		(1) FF3	(2) Carhart	(3) FF5	(4) FF5 + Mom
Alpha	**Europe**	0.481***	0.488***	0.282*	0.280
		(0.160)	(0.180)	(0.161)	(0.180)
	USA	0.288***	0.286***	0.316***	0.305***
		(0.088)	(0.088)	(0.086)	(0.085)
MKT	**Europe**	0.610***	0.607***	0.585***	0.586***
		(0.030)	(0.035)	(0.030)	(0.033)
	USA	0.698***	0.705***	0.693***	0.701***
		(0.018)	(0.018)	(0.018)	(0.018)
SMB	**Europe**	0.230**	0.231**	0.255**	0.255**
		(0.107)	(0.106)	(0.104)	(0.103)
	USA	0.001	0.015	-0.048	-0.015
		(0.033)	(0.037)	(0.040)	(0.041)
HML	**Europe**	0.217***	0.214***	0.479***	0.480***
		(0.057)	(0.071)	(0.088)	(0.097)
	USA	-0.150***	-0.151***	-0.067**	-0.061**
		(0.022)	(0.022)	(0.031)	(0.030)
CMA	**Europe**			-0.083	-0.083
				(0.130)	(0.130)
	USA			-0.145***	-0.168***
				(0.042)	(0.044)
RMW	**Europe**			0.586***	0.586***
				(0.134)	(0.135)
	USA			-0.039	-0.020
				(0.052)	(0.050)
WML	**Europe**		-0.009		0.001
			(0.067)		(0.063)
	USA		0.026		0.049**
			(0.022)		(0.021)
Observations	Europe	453	453	453	453
	USA	670	670	670	670
R2	Europe	0.698	0.698	0.717	0.717
	USA	0.805	0.806	0.811	0.812
Adjusted R2	Europe	0.698	0.696	0.714	0.713
	USA	0.804	0.805	0.809	0.810

*Note:This table presents the results of the Fama-French Three-Factor Model (1), the Carhart (1997) Model (2), the Fama-French Five-Factor Model (3) and the Fama-French Five-Factor plus momentum Model (4) over the period October 2019 to April 2024 of the investment regions: Europe and the USA using only iShares ETFs. The excess return is the dependent variable and is in percentage points. The observations (Obs.) as well as the R2 and adjusted (Adj.) R2 are included per portfolio and for the difference between the portfolios. The standard errors are estimated using the Newey and West (1987) procedure. *p<0.1, **p<0.05, ***p<0.01*

Table 14: Fama and MacBeth (1973) cross-sectional regression

	Europe				Japan			
	5	6	7	8	5	6	7	8
Constant	-0.667	-1.045	0.080	-0.527	-2.287**	-2.233 **	-1.745**	-1.736**
	(1.797)	(2.437)	(1.817)	(1.772)	(0.870)	(0.853)	(0.829)	(0.824)
Beta	0.186	0.328	0.119	0.287	0.645	0.630	0.402	0.393
	(0.959)	(0.876)	(1.008)	(1.043)	(0.646)	(0.645)	(0.670)	(0.673)
LnSIZE	-0.053	-0.043	-0.117	-0.091	0.270***	0.256***	0.213***	0.206***
	(0.196)	(0.282)	(0.227)	(0.224)	(0.063)	(0.061)	(0.063)	(0.062)
LnPB	0.444	0.382	0.713	0.651	-0.762*	-0.709*	-1.169**	-1.100**
	(0.317)	(0.423)	(0.448)	(0.425)	(0.403)	(0.396)	(0.470)	(0.463)
Invest			0.256	0.264*			-0.027	-0.033
			(0.154)	(0.141)			(0.028)	(0.026)
ROE			0.023	-0.021			0.065***	0.061***
			(0.022)	(0.021)			(0.021)	(0.021)
MOM		-0.008		-0.012		0.094***		0.077**
		(0.086)		(0.087)		(0.034)		(0.033)
ESG	0.073	0.083	0.084*	0.096*	0.002	0.003	-0.019	-0.017
	(0.048)	(0.060)	(0.050)	(0.051)	(0.018)	(0.018)	(0.015)	(0.016)
ESG Dummy	-0.018	-0.001	0.034	0.006	-0.285***	-0.252***	-0.208***	-0.192***
	(0.121)	(0.190)	(0.115)	(0.136)	(0.072)	(0.0651)	(0.070)	(0.068)
Observations	1,332	1,332	1,332	1,332	5,499	5,499	5,499	5,499
Avg.R2	0.537	0.573	0.621	0.650	0.360	0.374	0.393	0.405
Avg. Adjusted R2	0.408	0.422	0.454	0.463	0.325	0.332	0.346	0.351

	USA			
	5	6	7	8
Constant	-1.592	-1.545	-1.364	-1.314
	(1.939)	(1.949)	(1.901)	(1.907)
Beta	0.693	0.636	0.626	0.559
	(0.563)	(0.561)	(0.550)	(0.548)
LnSIZE	0.070	0.064	0.067	0.058
	(0.116)	(0.117)	(0.109)	(0.110)
LnPB	0.237	0.236	0.242	0.239
	(0.426)	(0.427)	(0.430)	(0.430)
Invest			0.013	0.011
			(0.015)	(0.015)
ROE			0.001	0.002
			(0.011)	(0.012)
MOM		0.083***		0.084***
		(0.022)		(0.024)
ESG	0.047	0.046	0.039	0.039
	(0.047)	(0.047)	(0.047)	(0.048)
ESG Dummy	0.108	0.133	0.105	0.130
	(0.166)	(0.169)	(0.165)	(0.169)
Observations	8,042	8,042	8,042	8,042
Avg.R2	0.412	0.419	0.443	0.450
Avg. Adjusted R2	0.391	0.393	0.415	0.418

*Note:This table presents the results of the monthly Fama and MacBeth (1973) regressions using the factors of the the Fama-French Three-Factor Model (1), the Carhart (1997) Model (2), the Fama-French Five-Factor Model (3) and the Fama-French Five-Factor plus momentum Model (4) focused on the individual ETFs over the period October 2019 to April 2024 of the investment regions: Europe, Japan, and the USA. BETA, lnSIZE, lnPB, Invest, ROE and MOM are control variables with regard to beta, market capitalization, price-to-book ratio, investment growth, return on equity, and average returns of the last 12 months. The independent variable is ESG which is the ESG risk score. The dependent variable is the excess return. The dummy for the ESG terminology in the name of the ETF is denoted ESG Dummy. The total number of observations as well as the R2 and adjusted R2 are given per model. The standard errors are estimated using the Newey and West (1987) procedure. *p<0.1, **p<0.05, ***p<0.01*

Table 15: ESG Terminology

ESG Terminology
Baptist
blue
carbon
Catholic
Christian
climate
community
durable
environment
ESG
ethical
faith
governance
green
human
rights
impact
Islam
Lutheran
mission
moral
peace
philosophy
religion
responsible
social
solidarity
subsidiarity
sustainable
sustainability
values

Note: This table presents the terminology used to construct the ESG Dummy following the work of Candelon et al. (2021).