



Handelshøyskolen BI

GRA 19703 Master Thesis

Thesis Master of Science 100% - W

Predefinert informasjon

Startdato:	16-01-2022 09:00	Termin:	202210
Sluttdato:	01-07-2022 12:00	Vurderingsform:	Norsk 6-trinns skala (A-F)
Eksamensform:	T		
Flowkode:	202210 10936 IN00 W T		
Intern sensor:	(Anonymisert)		

Deltaker

Navn: Tobias Warne Drake og Valera Kristiansen

Informasjon fra deltaker

Tittel *: The importance of ESG ratings on risk-adjusted stock performance

Navn på veileder *: Salvatore Miglietta

Inneholder besvarelsen konfidensielt materiale? Nei Ja
Kan besvarelsen offentliggjøres? Ja Nei

Gruppe

Gruppenavn: (Anonymisert)
Gruppenummer: 209
Andre medlemmer i gruppen:

The importance of ESG ratings on risk-adjusted stock performance

Master Thesis

by

Tobias Warne Drake and Valera Kristiansen

MSc in Business with Major Finance

Supervisor:

Salvatore Miglietta

Oslo, July 1st, 2022

Abstract

In this thesis, we conduct an empirical study of whether Environmental, Social, and Governance (ESG) ratings impact risk-adjusted stock performance and its ability to be used as a signal in investment strategies. We form top and bottom decile portfolios based on three different ESG ratings and regress the portfolio excess returns on Fama-French risk factors. The study returned positive significant monthly 5-factor alphas of up to 0.57% for top portfolios in the European market and indicates that aggregated ESG scores can provide valuable information to private investors; however, the US study failed to produce the same evidence. The inconsistent results prevent us from providing a decisive answer on the impact of ESG ratings on risk-adjusted stock performance.

This thesis is a part of the MSc programme at BI Norwegian Business School. The school takes no responsibility for the methods used, results found, or conclusions drawn.

Acknowledgments

We would like to thank our supervisor Associate Professor Salvatore Miglietta for guiding us through the process of writing this thesis with honest and helpful feedback. We would also like to thank the library at BI for facilitating access to the necessary databases. Lastly, we would like to thank our friends and families for their support.

Contents

<i>List of Tables</i>	<i>II</i>
<i>List of Figures</i>	<i>II</i>
<i>List of Abbreviations</i>	<i>III</i>
<i>List of Symbols</i>	<i>III</i>
1. Introduction and motivation	1
2. Literature review	4
2.1 The growing importance of corporate ESG performance	4
2.2 Discrepancy in agency ratings	5
2.3 Conflicting research on ESG performance	6
3. Hypothesis	8
3.1 Research question	8
3.2 Hypothesis	8
4. Research methodology	9
4.1 Portfolio construction	9
4.2 Portfolio theory	9
4.3 Validity of results	10
5. Data collection	12
5.1 Databases	12
5.2 The ESG scores and agencies	12
5.2.1 Choice of scores	12
5.2.2 Refinitiv ESG score.....	13
5.2.3 Refinitiv ESG Combined score	13
5.2.4 Bloomberg ESG Disclosure Score	14
5.3 Sample selection	14
5.4 Descriptive Statistics	16
5.4.1 Rating discrepancy	16
5.1.2 Portfolio characteristics.....	17
6. Analysis	20
6.1 Portfolio Returns	20
6.2 Regression results	21
6.2.1 Introductory analysis	22
6.2.2 Analysis of Fama-French 5-factor model regression	23
6.3 Discussion	26
6.3.1 Abnormal returns.....	26
6.3.2 Factor exposure	27
6.3.3 Limitations of the study.....	27
7. Conclusion	29
8. References	30
9. Appendix	33
A: Regression outputs	33
B: Model testing	37

List of Tables

Table 1: Rating coverage of ESG.....	15
Table 2: Correlation Matrices.....	16
Table 3: Average economic sector score US.....	17
Table 4: Average economic sector score Europe	17
Table 5: Average economic sector composition US	18
Table 6: Average economic sector composition Europe.....	18
Table 7: Fama-French 5 regression results (Value-weighted)	23
Table 8: Fama-French 5 with momentum regression results (Value-weighted).....	33
Table 9: Carhart 4 factor model regression results (Value-weighted)	33
Table 10: Fama-French 3 factor model regression results (Value-weighted)	34
Table 11: Fama-French 5 regression results (Equally-weighted).....	34
Table 12: Fama-French 5 with momentum regression results (Equally-weighted)	35
Table 13: Carhart 4 factor model regression results (Equally-weighted)	35
Table 14: Fama-French 3 factor model regression results (Equally-weighted)	35
Table 15: Subperiod (2011-2016) Fama-French 5 regression results (Value-weighted)..	36
Table 16: Subperiod (2016-2019) Fama-French 5 regression results (Value-weighted)..	36
Table 17: White Test on value-weighted portfolios	37
Table 18: White Test on equally-weighted portfolios.....	37
Table 19: Breusch-Godfrey test on value-weighted portfolios	38
Table 20: Breusch-Godfrey test on equally-weighted portfolios	38
Table 21: Augmented Dickey-Fuller test	38

List of Figures

Figure 1: Plot of cumulative return of portfolios	20
---------------------------------------------------------	----

List of Abbreviations

ADF	Augmented Dickey-Fuller
B/M	Book to market
CMA	Conservative minus aggressive
CSR	Corporate Social Responsibility
ESG	Environmental, Social and Governance
EU	European Union
EW	Equally-Weighted
FF	Fama and French
GSIA	Global Sustainable Investment Alliance
HML	High minus low
ISIN	International Securities Identification Number
IFS	International Finance Corporation
MKT	Market factor
RF	Risk-free
RMW	Robust minus weak
R&D	Research and development
SMB	Small minus big
S&P	Standard & Poor
SRI	Socially Responsible Investing
TRBC	The Refinitiv Business Classification
US	United States
VW	Value-Weighted
WML	Winner minus loser

List of Symbols

α	Alpha (Abnormal return)
β	Beta (Exposure to factors)
H_0	Null hypothesis
H_A	Alternative hypothesis
R^2_{adj}	Adjusted R-Squared
u	Error term

1. Introduction and motivation

The recent increase in the number of green companies, their price developments, and broader media coverage has sparked the ESG interest of investors worldwide. Large institutional investors have adapted to the new environmental discussion, and the increased focus on ESG is becoming a mainstream phenomenon. Socially responsible investing continues to gain momentum, and index funds focusing on sustainability have more than doubled, both in the number of funds and total assets. The sustainability factor has made its mark on the financial world, and with sustainable funds currently overseeing 20% of the total asset under management in the US (Stevens, 2020), its impact can no longer be neglected.

The Global Sustainable Investment Alliance showed in their 2020 report the continuing prevalence of sustainable investment across the global investment industry. They provide evidence for the statement with assets under management reaching USD 35.3 trillion (GSIA, 2020) compared to USD 13.6 trillion in 2012 (GSIA, 2012). Here, sustainable investment is referred to as an investment approach that considers ESG factors in portfolio selection and management (GSIA, 2020). Sustainability has become a notable part of financial markets, and it seems like sustainability will be a key factor for future investment strategies. Earlier studies on the subject provide conflicting results regarding the impact of ESG performance on stock returns. Some studies provide evidence that suggests that high-rated ESG firms outperform low-rated ESG firms (Kempf and Osthoff, 2007). Others conclude with no long-term difference in returns between environmentally friendly firms and regular firms (Atz et al., 2019), or even some presence of an underperformance (Renneboog et al., 2008).

We want to conduct an empirical study examining the effects of ESG ratings on the performance of publicly traded stocks. We investigate whether there is a difference in risk-adjusted stock performance between the top ESG performers relative to bottom ESG performers. We will be constructing top and bottom portfolios based on a selection of lagged ESG scores and examine the returns using both Fama-French and Carhart risk factors to make inferences on performance and potential abnormal returns. We therefore present the following research question:

Does ESG performance have an impact on risk-adjusted stock performance?

The analysis will explore both the American stock market with the S&P 500 Index and the European stock market with the STOXX Europe 600 index as our asset universe. The inclusion of two separate markets enables us to make inferences on differences between American and European companies on ESG performance. While the financial world seems more intertwined than ever, national restrictions and laws on the topic of ESG differ between geographical markets and companies are obliged to act in accordance, enabling geographical differences to be uncovered. The sample period is set from the beginning of 2011 to the end of 2019. The period was chosen to include the latest trends and market movements, and sufficiently cover company ESG scores in our asset universe.

We will also examine whether the choice of ESG score-providers (agencies) affects the performance of portfolios consisting of high-rated ESG companies and low-rated ESG companies. Existing literature, such as Kempf and Osthoff (2007), has mainly used ESG ratings from only one provider to assess corporate sustainability performance. As a result of the rapid increase in the importance of ESG, numerous rating agencies have tried to incorporate ESG performance into a quantifiable, measurable score. Financial agencies such as Eikon Refinitiv, MSCI, S&P, and Bloomberg have all attempted to create a compounded measure of ESG performance. Despite their experience on the topic and financial understanding, there seems to be disagreement between agencies regarding ESG performance. Deviating opinions, performance measurements and methodologies cause a discrepancy in results and a lack of consensus (Berg et al., 2019; Gibson et al., 2021).

The lack of consensus adds to the problem of information uncertainty. When investors try to quantify a debatable and qualitative measurement in different ways, it decreases information flow, and the individual opinion of a company's ESG performance will be heavily affected by the choice of rating provider. Therefore, our analysis will explore multiple measurements of aggregated ESG performance provided by Refinitiv and Bloomberg. The thesis will explore the characteristic differences between rating methods and unveil the consequences of using the different scores as the only sorting characteristic in portfolios.

We want to extend existing research that has documented differences between top and bottom ESG performers. While many papers such as Friede, Busch, and Bassen (2015) have studied the relation between stock returns and average ESG ratings, our paper will systematically test whether ESG portfolios, constructed using ratings from different providers and methodologies, can be used to obtain abnormal returns by firms and investors. This has also been studied recently, by Gibson, Krueger and Schmidt (2021). However, their paper focused on whether the disagreement about a firm's ESG performance can be utilized as a trading signal, which is not a part of our study.

Our contribution to the existing literature is to make an inference on the long-term performance of the two categories of companies and how it could potentially be utilized in discovering future investment opportunities. Our findings suggest that it is possible to obtain significant abnormal returns using ESG scores as an investment signal in portfolio construction. However, with ambiguous and inconsistent results across markets, the effect of ESG on global stock performance has yet to be determined. The study will unveil how sensitive ESG research is to sample period, asset universe, and that the inclusion of ESG in investment decisions should be done with caution by considering both score provider and methodologies used.

2. Literature review

2.1 The growing importance of corporate ESG performance

The term ESG as integration of environmental, social, and governance was first introduced in 2005 (IFC, 2005). However, the practice of ESG began in the 1960s as “Social Responsible Investing” (SRI), with investors excluding stocks or entire industries from their portfolios based on business activities (MSCI, 2021). SRI is a term used to describe “*Investments that are considered socially responsible due to the business the company conducts.*” (Chen, 2021). The most usual form of SRI strategy consists of excluding so-called “sin companies”¹ from portfolios. The exclusion strategy is the most common strategy of SRI because it gives the investor the option to choose which sectors or companies they will exclude from their portfolios based on their own criteria. The exclusion of certain companies seems to come at a price, exclusion makes optimization of portfolios harder, and the SRI strategy appears to limit the potential returns of investors (Hong & Kacperczyk, 2009).

The somewhat vague definition of SRI leads to potentially different interpretations across investors and, therefore, different implementations and actions in the markets. To better understand what SRI is, we can look at the definition from Eurosif², the leading European SRI membership organization. The newfound interest in environmental-friendly companies, the increasing popularity of ESG performance, and the appearance of ESG scores from third-party agencies have fundamentally changed our view on ESG importance and blurred our perception of it regarding pricing and stock performance.

Recent research on why investors use ESG information has shown that 82% of investment professionals use ESG information because it is financially material to investment performance, suggesting that it is primarily financial rather than ethical motives (Amel-Zadeh & Serafeim, 2018). The study was conducted as a survey of 413 senior investment professionals, where the respondents, on a value-weighted

¹ “Sin companies” are defined as publicly traded companies involved in the production of alcohol, tobacco, and gaming (Hong & Kacperczyk, 2009)

² “Sustainable and responsible investment (SRI) is a long-term oriented investment approach which integrates ESG factors in the research, analysis, and selection process of securities within an investment portfolio. It combines fundamental analysis and engagement with an evaluation of ESG factors in order to better capture long-term returns for investors, and to benefit society by influencing the behaviour of companies.” (Eurosif, 2021).

basis, comprise 43% of the global institutional assets under management and reflect the views of largely mainstream investment professionals. Furthermore, the study found a significant difference between US and European investors regarding the use of ESG information in investment decisions, with 75% of the US investors compared to 84% of the European investors. They also found evidence that US investors are less likely to use ESG data in their investment process than European investors because of more substantial concerns about data reliability (Amel-Zadeh & Serafeim, 2018).

2.2 Discrepancy in agency ratings

Major rating agencies have conflicting results when evaluating the ESG performance of different firms. The study by Berg, Kölbel, and Rigobon (2019) investigated this divergence in ESG ratings. The paper applies data from six leading rating agencies and finds an average correlation of 0.54, ranging from 0.38 to 0.71. These results suggest an inefficient information flow in the market and cause an economic problem when investors do not have a unified understanding and valuation of ESG performance. The authors believe that the discrepancy between agencies causes investors to disregard ESG performance as a valuation factor of corporate stocks because identifying outperformers becomes clouded and challenging.

The rating discrepancy also poses a challenge for empirical research, as the agencies used in a study would heavily impact the results and conclusions. (Berg et al., 2019). The incentives for businesses to enhance ESG performance are slowed down, as an improvement would receive mixed signals from agencies. As a result, the expected value of the improvement would be difficult to assess.

Another study of ESG rating disagreement is the paper by Gibson, Krueger, and Schmidt (2021). Using ESG ratings from seven different data providers for a sample of S&P 500 firms between 2010 and 2017, this is one of the most comprehensive data coverages on the topic. They find the average pairwise correlation between the ESG ratings of the providers to be 0.45 (Gibson et. al, 2021). This underscores the problem regarding the use of different providers. Further, they study whether the disagreement varies along with observable firm-level financial and accounting characteristics. They provide evidence that disagreement tends to be higher for the

largest firms in the S&P 500, where the driving factor could be the complexity of such firms (Gibson et. al, 2021)

2.3 Conflicting research on ESG performance

The effect of ESG performance on company performance is also a disputed topic. ESG-optimized portfolios may face opposition regarding performance compared to ESG-neutral portfolios. The general market opinion suggests that a company is exposed to two types of risks, systematic risk, and idiosyncratic risk. While systematic risk incorporates risk factors that affect the entirety of the market, idiosyncratic risk focuses only on factors that affect an individual company or an industry. Modern portfolio theory argues that an investor can eliminate all idiosyncratic risks through diversification (Markowitz, 1952). When investors set a high ESG score as necessary for investment opportunities, they introduce a negative screening approach that could hinder full portfolio diversification and infuse more investment risk.

In contrast to Markowitz theory, arguments can be made for ESG performance in the long term. Increased focus and engagement on sustainable solutions, R&D towards a greener production, increased focus on corporate governance, and improvement on a social level are all factors that will improve risk management and limit risk (Hoepner et al., 2022). The statement is also supported by Eccles et al. (2019), who also suggests that higher sustainability contributes to performance through improved human capital, better relationships with stakeholders and a more satisfied consumer base.

Kempf and Osthoff (2007) examined the relationship between corporate sustainability and portfolio performance of US stocks with a sample from the period 1992 to 2004. Based on KLD's SRI ratings, they used a screening approach by forming a high portfolio of the top 10% stocks and a low portfolio of the bottom 10% stocks. They find that the performance of the socially responsible portfolios is never significantly negative and conclude that there is a positive relationship between SRI and financial performance. In addition, their long-short strategy (long in the high-rated stocks, short in the low-rated stock) yields a positive alpha up to 8.7% a year with the Carhart four-factor model.

One of the most comprehensive aggregated studies regarding ESG and financial performance was conducted by Friede, Busch, and Bassen (2015), where they looked at 2200 individual studies, which allows for more generalizable statements. They found that 90% of the studies had a nonnegative ESG and corporate financial performance relationship. The sub-effects in regions were large, where the contrast is most apparent between North America (42.7% positive studies and 7.1% negative) and developed Europe (26.1% positive studies and 8.0% negative). This suggests that ESG outperformance opportunities exist in many market areas but is highly dependent on how ESG criteria are integrated into investment processes.

3. Hypothesis

3.1 Research question

As our literature review displays, there are conflicting theories on the effect of ESG on stock performance. The increased acknowledgement of ESG as a potential factor in the financial world has become obvious by the number of sustainable funds that have risen and the capital invested by investors. However, the question remains on whether ESG performance impacts risk-adjusted stock performance. The lack of consistent effect and importance of ESG on financial performance prevents investors from strategically implementing ESG in their strategy and leaves researchers puzzled. Therefore, we want to have an empirical investigation of the companies listed on the S&P 500 index and the STOXX Europe 600 index and the link between ESG ratings and their stock performance. We therefore propose the following research question:

Does ESG performance have an impact on risk-adjusted stock performance?

3.2 Hypothesis

The hypothesis focuses on the performance of high-rated ESG companies versus low-rated ESG companies in the index. The experiment will let us formally test performance differences between the two types of companies. The portfolios will be based on ratings from the different agencies, and actual performance differences between agency portfolios are expected. Our null and alternative hypotheses are displayed below.

H_0 : *Companies with low ESG scores will not be outperformed by companies with high ESG scores.*

H_A : *Companies with low ESG scores will be outperformed by companies with high ESG scores.*

4. Research methodology

4.1 Portfolio construction

To study the relationship between ESG scores and financial performance, we constructed portfolios from the different rating providers using a sample period from the beginning of 2011 to the end of 2019. The companies in our asset universe were divided into annual deciles portfolios based on their overall ESG score from the last available rating of each agency in the previous calendar year. The highest ESG-performing decile was used as our top portfolio, and the lowest ESG-performing decile was used as the bottom portfolio. The portfolios were formed annually, resulting in a total of 24 new portfolios every year ³.

4.2 Portfolio theory

By constructing different portfolios based on ESG ratings and agencies, we can use the theoretical models of both Fama & French (1993; 2015) and Carhart (1997). When we use multiple models, we ensure a higher robustness of our results. Our dependent variable is the excess return of the constructed ESG portfolio in all models. Furthermore, we have used the S&P 500 and the STOXX Europe 600 as our market proxies, depending on which market we investigate.

The Fama-French 3-factor model, displayed in Equation 1, utilizes the market factor in combination with SMB and HML (Fama & French, 1993).

(Equation 1)

$$r_{ESG_t} - rf_t = \alpha + \beta_{MKT} * (r_{mkt_t} - rf_t) + \beta_{SMB} * SMB_t + \beta_{HML} * HML_t + u_t$$

The Carhart 4-factor model (Equation 2) tries to enhance the findings of Fama (1993), and the model introduces another factor that incorporates performance tendency (Carhart, 1997) and measures the portfolio's exposure to recent market winners and losers.

³ At the end of year t-1 we use the last available ESG score to form top and bottom decile portfolios that are held throughout year t. This is done for both markets, for each of the three scores. This results in 12 portfolios yearly. As we form value-weighted and equally-weighted portfolios this result in a total of 24 portfolios that are formed yearly.

(Equation 2)

$$r_{ESG_t} - rf_t = a + \beta_{MKT} * (r_{mkt} - rf_t) + \beta_{SMB} * SMB_t + \beta_{HML} * HML_t + \beta_{WML} * WML_t + u_t$$

The Fama-French 5-factor model (equation 3) is their latest addition to stock performance theory (Fama&French, 2015). The model is based on the original three-factor model mentioned above but includes two more factors, RMW and CMA.

(Equation 3)

$$r_{ESG_t} - rf_t = a + \beta_{MKT} * (r_{mkt} - rf_t) + \beta_{SMB} * SMB_t + \beta_{HML} * HML_t + \beta_{RMW} * RMW_t + \beta_{CMA} * CMA_t + u_t$$

The final model we use in our study combines the Fama-French 5-factor model with the performance tendency factor from Carhart (equation 4).

(Equation 4)

$$r_{ESG_t} - rf_t = a + \beta_{MKT} * (r_{mkt} - rf_t) + \beta_{SMB} * SMB_t + \beta_{HML} * HML_t + \beta_{RMW} * RMW_t + \beta_{CMA} * CMA_t + \beta_{WML} * WML_t + u_t$$

4.3 Validity of results

The results can only be considered valid if the data satisfies a list of assumptions. To check our data for heteroscedasticity, we run a White test (White, 1980). The test revealed that a small portion of our portfolios had heteroscedasticity, thereby requiring additional calculations. Our solution was to run these regressions with robust standard errors using heteroscedasticity consistent (HC 3) standard errors. Additionally, we checked our data for autocorrelation. This was done by running a Breusch-Godfrey test that reported no presence of autocorrelation in any of the portfolios. We also tested the data for stationarity by running an ADF test. The test revealed that the data was stationary, and no further action was required. In summary, after these tests, the data can be considered valid for interpretation (Results of the three tests on all portfolios are presented in Appendix B).

Our study is also exposed to different biases. By examining the S&P 500 index and the STOXX Europe 600 index, the study focuses on mainly large-cap companies with solid data, including ESG scores. Smaller companies will likely have less public information, and potential ESG scores could be deficient or even non-existent, making the study less suited for smaller samples. The chosen sample is also exposed to survivorship bias as the indexes consist of the 500 largest American companies and the 600 largest European companies. However, we do not consider this an important aspect of the study because we only examine the differences between the listed companies, and any survivorship bias will be present in all portfolios.

Further, the providers of ESG-related information are usually the companies themselves, which could also cause implications. Yu et al. (2020) study mechanisms to holistically lessen firms' greenwashing behavior in ESG dimensions. Greenwashing occurs if firms make misleading ESG disclosures according to Yu et al. (2020). By creating a peer-relative greenwashing score, they measure the magnitude of a firm's greenwashing behavior in ESG issues and identify large companies that engage in greenwashing. The findings suggest that firms exposed to adequate supervision under institutional investors are less likely to engage in ESG greenwashing. By examining companies listed on the S&P 500 and STOXX Europe 600 Indexes, the companies in our sample are under strict supervision from both authorities and large investors. This infers that the presence of greenwashing is minimized in our study.

5. Data collection

5.1 Databases

From Refinitiv, we extract both the market capitalization and two different ESG scores for each company in our asset universe, in addition to matched monthly total return adjusted for stock dividends and splits. The third ESG score is extracted from the Bloomberg Terminal. We have matched the databases from Refinitiv and Bloomberg annually using ticker and name string matching techniques, with ISIN as the common stock identifier, and manual matching where this has been needed. Furthermore, we extract the necessary data for all Fama-French and Carhart models from the Kenneth R. French data library (French, 2022), including information regarding risk premium and risk-free rates for both the European and American markets.

5.2 The ESG scores and agencies

5.2.1 Choice of scores

We will in our study only focus on the aggregated score of the three pillars E, S, and G. Refinitiv provides two such scores, both an ESG score, and an ESG Combined score which also account for controversy further explained in subsection 5.2.3. From Bloomberg, we will use the Bloomberg ESG Disclosure score which focuses on how much ESG information a company discloses, i.e., the transparency of a company. Both providers use a scale from 0 to 100 for their assessments.

The scores have been selected because they offer different levels of incorporating ESG Data. The Bloomberg ESG Disclosure score is the most simplistic score out of the chosen ones. The score is only a measurement of companies' transparency and disclosure of ESG Data and does not account for the quality of the data provided. The Refinitiv ESG score consists of a deeper analysis, companies are compared to industry peers on ESG performance metrics and given a score depending on their performance relative to their industry peers. The last score, Refinitiv ESG Combined is an extension of the Refinitiv ESG score that accounts for a company's controversy in addition to the original metrics of the Refinitiv ESG score.

When we apply the different scores in our portfolio formation and measure their long-term performance, it will be interesting to uncover potential differences between the scores and if the inclusion of more ESG metrics and deeper analysis on the topic increases the performance of top ESG portfolios.

5.2.2 Refinitiv ESG score

Refinitiv designs ESG scores to transparently and objectively measure a company's relative ESG performance. It captures and calculates over 450 company-level measures, of which a subset of 186 of the most comparable and material per industry power the overall company assessment and scoring process. The underlying measures are based on considerations around comparability, impact, data availability, and industry relevance that varies across each industry group. These are grouped into ten categories that form the three pillar scores and the final ESG score, reflecting the company's ESG performance, commitment, and effectiveness based on publicly reported information. The score ranges from 0 for companies that are performing worst, to 100 for those who perform the best. Scores are continuously being updated, but larger updates regularly appear after the release of annual reports. Refinitiv currently offers ESG scores for 10,000 companies around the world. (Refinitiv, n.d.)

5.2.3 Refinitiv ESG Combined score

The Refinitiv ESG Combined Score is an extension of the original ESG score mentioned above. The original ESG score is used as the base score for each company, but the score also takes ESG controversy into account, and any controversy will lower the combined score. The controversies score is calculated based on 23 ESG controversy topics, reflecting recent controversies in the latest complete period. (Refinitiv, 2022). The controversy score is the sum of all controversies published in media linked to the 23 topics, this includes any controversies surrounding topics such as tax fraud, human rights, responsible R&D, environmental controversies, wages and working conditions. The controversies score also addresses the market cap bias from which large-cap companies suffer, as they attract more media attention than smaller-cap companies. (Refinitiv, 2022). The final combined score is equal to the original ESG score only if the company has not experienced any controversy in the prior year. However, if there is a

presence of controversy then the ESG Combined score is calculated as the average of the ESG score and the controversy score.

5.2.4 Bloomberg ESG Disclosure Score

The Bloomberg score methodology differs from the Refinitiv scores as it only measures the amount of ESG data a company reports publicly and does not measure the company's performance on any data point. Hence, it is only based on the extent of a company's Environmental, Social, and Governance disclosure. These three pillars are equally weighted to form the Bloomberg ESG Disclosure score. The score ranges from 0 for companies that do not disclose any of the ESG data, to 100 for those that disclose every data point. The scoring methodology was updated in 2022 due to the evolution of corporate ESG data reporting, which has been applied across all companies and years with available scores (Bloomberg, n.d.).

5.3 Sample selection

Research concerning ESG generally faces the challenge that the availability of ESG data is restricted in both the cross-section and the time-series (Gibson et. al, 2021). To maximize the number of available ESG ratings per firm and the time-series dimension of the data, we will use members of the S&P 500 Index and STOXX Europe 600 Index. The markets were chosen because satisfactory and complete ESG Data is in general only available for larger companies, and using the largest indexes for each market provides us with the highest market coverage. The study will focus on a sample period from the beginning of 2011 to the end of 2019. This will incorporate the latest market changes and trends without additional noise from the Covid 19 period that could dilute our findings, while still obtaining a large enough sample for a meaningful and credible analysis. Although Refinitiv began providing ESG data as early as 2002, we started our sample period in 2011 because Bloomberg terminals integrated ESG data in 2010 and we have restricted the portfolios to include scores from both providers.

The S&P 500 index, consisting of large-cap companies, covers approximately 80% of available market capitalization in the US (S&P500, n.d.). STOXX Europe 600 is an index representing large, mid, and small-capitalization companies across 17 countries of the European region (Stoxx, n.d.). By including both markets in our

study, we examine potential geographical differences and unveil how geographical restrictions and differences could affect the otherwise global financial market.

The observations look at monthly total returns over nine years to infer on long-term performance. The agencies update their ESG ratings with different frequencies. To cope with this problem, we will use the latest available rating of each agency in the calendar year.

Table 1: Rating coverage of ESG

Provider Coverage In Europe					Provider Coverage in US				
Year	Bloomberg	ESG Score	ESGC Score	Data pool	Year	Bloomberg	ESG Score	ESGC Score	Data pool
2011	554	592	592	541	2011	496	495	495	483
2012	560	593	593	548	2012	498	496	496	485
2013	576	594	594	565	2013	496	495	495	482
2014	587	592	592	575	2014	494	495	495	480
2015	593	594	594	582	2015	495	495	495	481
2016	594	597	597	585	2016	495	497	497	482
2017	587	592	592	583	2017	498	500	500	486
2018	594	598	598	587	2018	501	502	502	491
2019	595	598	598	590	2019	500	503	503	494
Mean	582	594	594	573	Mean	497	498	498	485

Table 1 presents the coverage of each of the used scores, both for the STOXX 600 and the S&P 500 throughout our sample period. The average coverage for the period is presented in the last row. The STOXX 600 consist of 600 stocks, while the S&P 500 consist of 505 stocks.

Differences in coverage between ratings from Refinitiv and Bloomberg created limitations in our data pool. Table 1 summarizes the data coverage of each provider in both markets, and the final data pool is restricted to only include stocks with all ESG scores and at least the first observation of monthly total return in the given year of portfolio formation. The criteria were necessary to provide the most feasible strategy and results, leaving the analysis without bias by excluding mergers, private acquisitions, or bankruptcies that happened beyond the portfolio constructions for a given year. We consider the data loss we endured to be minimal and do not believe that this affected the analysis significantly.

Table 1 also shows that the data coverage increases throughout the study period and further highlights the increased focus and interest in ESG performance. Refinitiv provides more ESG data in both markets, but the differences are more prominent in Europe.

5.4 Descriptive Statistics

5.4.1 Rating discrepancy

Table 2 visualizes the discrepancy in ratings of the two providers and the different scores. The low correlation between the Bloomberg ESG Disclosure score and each of the Refinitiv ESG scores shows little correlation between how much a firm discloses on ESG performance and how third-party agencies evaluate their actual performance with developed ESG metrics. The findings suggest that a high score in one of the Refinitiv scores in no way can be utilized to forecast a good score from another provider or vice versa, supporting the findings of (Berg et al., 2019). Although we have used different scores, we obtain an average correlation of 0.36 between the different scores in Europe and an average correlation of 0.51 in the US.

The inclusion of the controversy factor has a significant impact on the Refinitiv score. Even though a correlation between the two Refinitiv scores of 0.71 in the US and 0.68 in Europe are high compared to the findings of Berg et al. (2019), the differences between the scores caused by a sole factor are notable when they share every other ESG metric.

Table 2: Correlation Matrices

Pearson correlation US - Pairwise			
	ESG Combined Score	ESG Score	Bloomberg Disclosure Score
ESG Combined Score	1		
ESG Score	0.7109	1	
Bloomberg Disclosure Score	0.3123	0.5174	1
Pearson correlation Europe - Pairwise			
	ESG Combined Score	ESG Score	Bloomberg Disclosure Score
ESG Combined Score	1		
ESG Score	0.6755	1	
Bloomberg Disclosure Score	0.2528	0.5262	1

Table 2 represents the average correlation between the different scores throughout our sample period. The correlation is calculated using the Pearson correlation coefficient.

Further analysis revealed that the discrepancy discussed in 5.4.1 affected all TRBC⁴ economic sectors and that little of the discrepancy in ratings are caused by certain sectors. Our findings in Tables 3 and 4 suggest that Refinitiv scores, in general, are higher than the Bloomberg ESG Disclosure scores and that sector means of

⁴ The Refinitiv Business Classification (TRBC) is a global, comprehensive, industry classification system owned and operated by Refinitiv, categorized into 13 economic sectors (Refinitiv, Nd.)

Refinitiv ESG and ESG combined scores have an average correlation of 0.95, much higher than for individual companies. The difference in the two Refinitiv scores also paints a picture of which sectors are most affected by the controversy. With an average drop in the score of 6.3 percentage points across both markets when accounting for controversy, the sector for consumer non-cyclicals has the largest decrease, followed by consumer cyclicals, healthcare. On the other hand, Real Estate and Academic & Educational Services seem to be unaffected by controversy.

Table 3: Average economic sector score US

TRBC Average Economic Sector Score: Output US			
	Refintiv ESG Combined Score	Refintiv ESG Score	Bloomberg Disclosure Score
Academic & Educational Services	22.60	22.60	16.10
Basic Materials	60.70	63.94	45.80
Consumer Cyclicals	53.23	61.88	37.07
Consumer Non-Cyclicals	55.75	67.99	44.39
Energy	52.69	58.46	39.32
Financials	56.06	61.44	35.73
Healthcare	57.94	67.87	39.67
Industrials	59.62	63.99	38.85
Real Estate	69.61	69.66	36.40
Technology	55.17	63.18	40.49
Utilities	55.38	61.56	45.91

Table 3 shows the average sector score for S&P 500 index constituents (2011-2019). Sectors are based on TRBC Economic Sector Name.

Table 4: Average economic sector score Europe

TRBC Average Economic Sector Score: Output Europe			
	Refintiv ESG Combined Score	Refintiv ESG Score	Bloomberg Disclosure Score
Academic & Educational Services	-	-	-
Basic Materials	66.07	73.16	46.87
Consumer Cyclicals	62.05	70.18	41.04
Consumer Non-Cyclicals	64.29	73.30	43.85
Energy	62.04	70.64	44.12
Financials	60.24	67.62	37.40
Healthcare	66.13	73.74	37.58
Industrials	64.33	69.59	40.55
Real Estate	68.80	68.81	37.25
Technology	61.26	66.84	39.19
Utilities	64.30	72.86	51.53

Table 4 shows the average sector score for STOXX 600 index constituents (2011-2019). Sectors are based on TRBC Economic Sector Name.

5.1.2 Portfolio characteristics

In the US study, presented in table 5, the top portfolios are heavily exposed to consumer non-cyclicals and technology while the bottom portfolios are exposed to consumer cyclicals and consumer non-cyclicals. The large exposure to consumer non-cyclicals of top portfolios suggests that the portfolios would be less volatile

and exposed during financial turmoil periods. Even though there is a large presence of consumer non-cyclicals in the bottom portfolios as well, the bottom portfolios' exposure to consumer cyclicals is even larger and will increase portfolio volatility.

Table 5: Average economic sector composition US

TRBC Average Economic Sector Composition VW Portfolios: Output US						
	Refintiv ESG Combined Score		Refintiv ESG Score		Bloomberg Disclosure Score	
	Top	Bottom	Top	Bottom	Top	Bottom
Academic & Educational Services	0.00%	8.11%	0.00%	1.39%	0.00%	0.71%
Basic Materials	8.33%	0.93%	6.25%	5.79%	14.59%	2.32%
Consumer Cyclicals	8.79%	5.81%	12.28%	16.19%	4.62%	20.04%
Consumer Non-Cyclicals	9.74%	21.27%	20.84%	11.34%	16.67%	8.47%
Energy	2.08%	13.89%	6.71%	13.21%	7.4%	6.82%
Financials	9.49%	10.44%	8.1%	6.47%	5.1%	18.82%
Healthcare	10.42%	5.78%	15.28%	6.72%	13.18%	11.33%
Industrials	14.35%	9.04%	9.72%	11.32%	5.56%	8.44%
Real Estate	15.27%	4.85%	6.25%	3.00%	2.31%	3.31%
Technology	16.9%	0.68%	11.1%	16.21%	18.07%	16.94%
Utilities	4.62%	19.21%	3.47%	8.35%	12.49%	2.8%

Table 5 shows the average sector composition of TRBC sectors present in the different portfolios for the US study. Sectors that are not included in either market during the sample period have been excluded from the table.

In the European portfolios, presented in table 6 the top portfolios are exposed to Industrials and basic materials, while the bottom portfolios are more exposed to the financial sector and consumer cyclicals. In contrast to the US market, both top and bottom portfolios seem to largely consist of cyclical sectors, making the portfolios more sensitive to financial cycles.

Table 6: Average economic sector composition Europe

TRBC Average Economic Sector Composition VW Portfolios: Output Europe						
	Refintiv ESG Combined Score		Refintiv ESG Score		Bloomberg Disclosure Score	
	Top	Bottom	Top	Bottom	Top	Bottom
Academic & Educational Services	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Basic Materials	12.88%	10.78%	11.28%	6.76%	22.02%	5.32%
Consumer Cyclicals	13.91%	23.06%	10.78%	13.44%	8.37%	14.96%
Consumer Non-Cyclicals	6.17%	2.77%	14.67%	5.46%	14.6%	6.37%
Energy	5.95%	3.66%	5.83%	4.45%	9.42%	2.83%
Financials	8.4%	18.83%	16.4%	24.6%	5.39%	24.01%
Healthcare	8.4%	4.05%	12.00%	5.05%	8.28%	13.36%
Industrials	23.66%	15.41%	13.96%	17.94%	10.27%	14.86%
Real Estate	10.07%	3.39%	4.41%	4.58%	1.44%	7.28%
Technology	9.48%	14.41%	7.89%	16.05%	5.21%	9.58%
Utilities	1.08%	3.63%	2.8%	1.67%	14.99%	1.42%

Table 6 shows the average sector composition of TRBC sectors present in the different portfolios for the European study. Sectors that are not included in either market during the sample period have been excluded from the table.

The controversy score has a large impact on the portfolio formation using Refinitiv scores. The findings are in line with the findings on average sector scores from tables 3 and 4, where consumer non-cyclical, consumer cyclical and technology were the sectors that were most negatively affected by the controversy. Both in the US and Europe, the top portfolios had a massive decline in the consumer non-cyclical sectors when taking the controversy into account. Similarly, the US bottom portfolios had a decline of 15.5%-points in technology and the European bottom portfolio had an increase in consumer cyclical when accounting for controversy.

6. Analysis

6.1 Portfolio Returns

To better understand how the top and bottom portfolios perform over the sample period, we have plotted the cumulative returns for the three different ESG scores, both for the US and European markets. The cumulative returns for each portfolio can be seen in Figure 1.

Figure 1: Plot of cumulative returns of portfolios

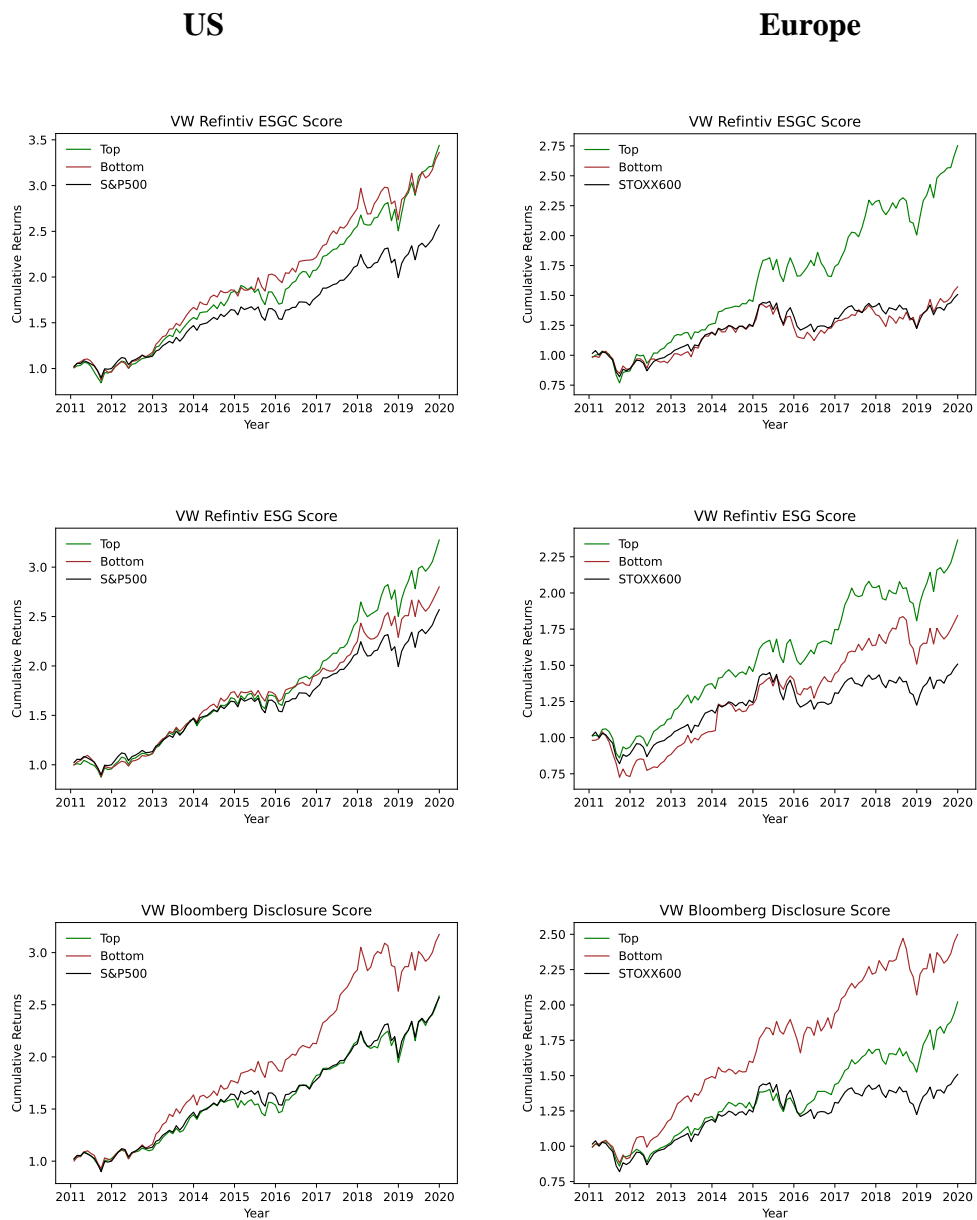


Figure 1 shows the cumulative returns of a 1-dollar investment for each portfolio throughout the sample period. The plots include the associated index as the benchmark and both top and bottom portfolios for each score. All plots are made using the value-weighted portfolios.

We see that we get conflicting results in the US market. For the Bloomberg ESG Disclosure score, the bottom portfolio outperforms both the index and the top portfolio. For both Refinitiv ESG combined score and the Refinitiv ESG score, the top portfolio outperforms the index and bottom portfolio.

Our findings suggest that all US portfolios follow the same trends, as seen in Figure 1, especially the Refinitiv portfolios. The Bloomberg portfolios also have similar patterns, but the bottom portfolio's cumulative growth rate is higher. We also observe that the performance of the ESG score portfolios without the controversy factor outperforms the portfolios including it.

The performance in the European market draws much resemblance to the US market. Using the Bloomberg ESG Disclosure score, the bottom portfolio outperforms both the top portfolio and the index. For both the Refinitiv scores, the top portfolios outperform the index and bottom portfolios, just as in the US study. However, the differences are larger in the European market than in the US. The patterns are once again strikingly similar, but the ESG combined score portfolios clearly distinguish themselves in the European market, and the ESG top performers yield a much higher cumulative return than the other portfolios.

The difference between the two Refinitiv scores and the Bloomberg ESG Disclosure score becomes clear in both markets. The performance of the top and bottom portfolios are opposites and visualize that the fundamental differences in rating methodology impact the performance of portfolios.

6.2 Regression results

The analysis will focus on the value-weighted portfolios for both the US and European markets. Our regressions include the Fama-French 3-factor model, the Carhart 4-factor model, and the Fama-French 5-factor model, with and without momentum factor.

The overview section will cover the factor loading differences between all the models, while the factor loadings will be examined closer by a deeper analysis of the Fama-French 5-factor model. We also ran all the regressions with equally weighted portfolios as a robustness check (Appendix A), however, these results will

not be commented on further as they provide similar results to the value-weighted portfolios.

6.2.1 Introductory analysis

Focusing on the alpha, the different factor models affect the significance of the results for some of our portfolios. We obtain a significant monthly alpha of 0.22% for the US portfolios when we use the Fama-French 3-factor model (Appendix A, Table 10), and 0.21% using the Carhart 4-factor model (Appendix A, Table 9), but the significance disappears when we account for RMW and CMA factors. The European study provides significant alphas in all top and some of the bottom portfolios. The inclusion of more factors does impact coefficients, but the significance of our findings remains unchanged.

The inclusion of additional factors also affects the SMB factor, where certain portfolios lose significance, using the Fama-French 5-factor model. This also applies to the HML factor. The inclusion of RMW does not provide any significant factor loadings, while CMA has explanatory power only for ESG score portfolios. The momentum factor provides explanatory power in the Carhart 4-factor model, but this vanishes when the remaining factors are accounted for.

The following section provides a deeper analysis using the Fama-French 5-factor model. The Fama-French 5-factor model was chosen because it offered the highest R_{adj}^2 of all regressions and includes the CMA risk factor, which showed significant results for our data sample.

6.2.2 Analysis of Fama-French 5-factor model regression

Table 7: Fama-French 5 regression results (Value-weighted)

Fama French Five Factor Model: Regressions output VW												
	Refinitiv ESG Combined Score				Refinitiv ESG Score				Bloomberg Disclosure Score			
	US		Europe		US		Europe		US		Europe	
	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom
Intercept	0.0017 (1.538)	0.0011 (0.702)	0.0057** (2.250)	0.0030 (1.114)	0.0003 (0.308)	0.0006 (0.451)	0.0047*** (2.594)	0.0016 (0.680)	-0.0009 (-0.898)	0.0013 (0.906)	0.0032* (1.907)	0.0056** (2.245)
Mkt-Rf	0.9707*** (29.659)	0.9994*** (21.720)	0.7536*** (11.194)	0.6307*** (9.169)	1.0278*** (36.810)	0.9156*** (23.787)	0.6985*** (12.061)	0.8541*** (13.691)	0.9995*** (33.419)	0.9371*** (21.854)	0.6305*** (14.173)	0.6431*** (9.418)
SMB	-0.0384 (-0.590)	-0.3375*** (-3.694)	-0.3961** (-2.518)	-0.5202** (-3.237)	-0.3888*** (-7.012)	-0.1088 (-1.424)	-0.6139*** (-6.153)	0.2831*** (1.942)	-0.1821*** (-3.067)	-0.1192 (-1.399)	-0.5776*** (-5.556)	-0.1015 (-0.819)
HML	-0.3047*** (-4.226)	-0.1133 (-1.118)	-0.5733** (-2.344)	-0.2456 (-0.983)	-0.0387 (-0.630)	0.0115 (0.135)	-0.3140 (-1.619)	-0.5312** (-2.344)	0.1177** (1.786)	-0.0720 (-0.762)	-0.0430 (-0.266)	-0.3560* (-1.653)
RMW	-0.1290 (-1.388)	-0.2000 (-1.531)	-0.2654 (-0.866)	-0.4900 (0.1221)	0.0049 (0.062)	0.0625 (0.571)	-0.0919 (-0.393)	-0.4640 (-1.633)	0.1222 (1.438)	-0.0103 (-0.085)	0.1347 (0.665)	-0.3236 (-1.228)
CMA	0.4874*** (4.383)	0.0041 (0.026)	-0.2788 (-0.976)	0.0814 (0.781)	0.1213 (1.279)	0.2351* (1.798)	0.0420 (0.207)	-0.2791 (-1.054)	0.1439 (1.416)	-0.0724 (-0.497)	0.0705 (0.373)	-0.2368 (-0.957)
R^2_{adj}	0.910	0.836	0.631	0.581	0.932	0.857	0.767	0.726	0.921	0.843	0.768	0.647

Table 7 shows the summary of regression outputs using the Fama-French 5-factor model. We regress the excess returns from our value-weighted portfolios from 20011-2019 on the Fama-French 5 risk factors using robust standard errors. The table shows t-values in parenthesis and * are used to visualize significant p-values; ***= $p < 0.01$, **= $p < 0.05$, *= $p < 0.1$, the adjusted R-squared for each regression are presented in the last row.

Alpha

The US study produces no significant alphas for any of the portfolios. For the top portfolios, the coefficient is positive using both the Refinitiv scores, but negative using the Bloomberg ESG disclosure score. The bottom portfolios provide higher alphas than the top portfolios using the Refinitiv ESG score and Bloomberg ESG Disclosure score. However, the lack of significant results makes it hard to conclude with the same presence of underperformance by top ESG portfolios as Renneboog et al. (2008).

The European study produces more interesting results. The top portfolios all have significant alphas, with the ESG Combined score being the best performing portfolio. This implies that the top ESG performers can produce a significant excess return in the European market but not in the US market. Only the Bloomberg ESG Disclosure score portfolio can produce significant alpha in the bottom portfolios, and it is the only score with a higher alpha for the bottom portfolio relative to the top portfolio. This implies that the findings using the Bloomberg ESG Disclosure score suggest an underperformance of top ESG portfolios, similar to Renneboog et al.(2008), while the results when using the ESG combined score and ESG score from Refinitiv are in line with the results of Kempf and Osthoff (2007). This further

supports the theory by Berg et al. (2019), suggesting that the choice of provider will dramatically affect the performance of portfolios.

Market factor

All portfolios, both in top and bottom, provide significant positive market exposure as expected. However, we find the differences between the two markets interesting. The US study gives the portfolios a market coefficient between 1.02 and 0.92, and only the ESG Combined Score portfolio shows a larger market exposure for the bottom portfolio relative to the top portfolio. The other portfolios have a higher coefficient for the top portfolios, indicating higher volatility and systematic risk in highly rated ESG firms. The average market exposure of all portfolios in the US is 0.98.

The European study produces results that are opposite of the US study. Here, the market coefficients are higher in the bottom portfolios in the Refinitiv ESG portfolio and the Bloomberg portfolio. In contrast, the Refinitiv ESG Combined score shows a larger coefficient for the top portfolios. The European portfolios have an average market exposure of 0.70, notably lower than in the US study.

SMB factor

Most of our portfolios have negative exposure to the SMB factor, indicating exposure to big companies. 7 of the 12 portfolios show significant negative exposure to the factor, whereas only the bottom Refinitiv ESG score portfolio for Europe has significant positive exposure to the SMB factor. The findings are unsurprising considering our asset universe consists of large-cap companies. The coefficient differences between providers suggest that the bottom portfolios consist of smaller companies than the top portfolios for both Refinitiv ESG score and Bloomberg ESG Disclosure score. However, when accounting for the controversy score in Refinitiv ESG Combined, the larger companies have been penalized heavier and Refinitiv's implementation of controversy has a significant impact on the choice of portfolio constituents.

HML factor

The regression results indicate that top portfolios are negatively exposed to the HML factor and therefore consist of an overweight of growth companies. The

findings are in line with previous research such as (Kempf and Osthoff, 2007). The Refinitiv Combined ESG score returns significant negative exposure in both studies, while Refinitiv ESG score only provides significant exposure for the European bottom portfolio. The Bloomberg ESG Disclosure score portfolios stand out, as this is the only provider that has a top portfolio with positive significant exposure to the factor. The results are expected because growth firms are less likely to disclose as much ESG performance as more established firms, due to a lack of resources and organizational efficiency.

RMW factor

The RMW factor does not provide significant results, nor do the coefficients have any consistency when changing the score-provider, the results leave us with little useful information for further analysis on this factor.

CMA factor

By examining the CMA factor in the US, all top portfolios have positive exposure to the factor, suggesting a tendency for top ESG firms to use conservative investment strategies; however, only the ESG Combined score provides significant results for the top portfolios. For the bottom portfolios, both Refinitiv scores suggest a conservative strategy, while the Bloomberg portfolio has a positive exposure that suggests more aggressive investments. Only the Refinitiv ESG score has significant results for the bottom portfolio. In the European study, none of the portfolios has significant exposure to the CMA factor. Both top and bottom portfolios are exposed positively and negatively depending on the score used for portfolio construction. This makes any interpretation of the European factor exposure difficult.

6.2.3 Alternative subperiods and cut-off

To further analyze the robustness of our results we divided the sample period into two subperiods, from the beginning of 2011 to the end of 2015 and from the beginning of 2016 to the end of 2019. When regressing the portfolio returns on Fama-French risk factors (Appendix A, Table 15 and 16), the portfolios still produced positive alphas for a few of our top portfolios. In the first subperiod, we obtained significant and stronger positive alphas in the European study compared to our original sample period. However, the significance disappears in the second subperiod. The US study provided exactly the opposite result. The first subperiod

shows no significant alphas for any portfolio, but the second subperiod reveals significant positive alphas for both Refinitiv ESG and ESG Combined top portfolios. This clearly shows the sensitivity of research on ESG performance and that even small changes in the sample period could affect the results dramatically.

Furthermore, we examined the robustness by expanding our portfolios from decile portfolios to quartile portfolios. The findings were similar to our findings in Table 7, but they provided less interesting values. Neither of the results from subperiods or alternative cut-offs will be commented on any further.

6.3 Discussion

6.3.1 Abnormal returns

Our study does find some evidence that suggests ESG performance has a significant impact on risk-adjusted stock performance using the Fama-French 5-factor model, but the results are ambiguous.

To start with the European market, we get significant abnormal monthly returns in all the top constructed portfolios. Using sorting from Refinitiv ESG and ESG Combined scores, the top portfolios yield 0.57% and 0.47% respectively. The findings contradict our initial null hypothesis; *Companies with low ESG scores will not be outperformed by companies with high ESG scores*. The results are in line with the findings of Kempf and Osthoff (2007) and suggest that the best ESG-performers generate superior abnormal returns compared to ESG laggards. However, the portfolios with sorting from the Bloomberg ESG Disclosure score, yield abnormal monthly returns for both the top and bottom portfolios in Europe, respectively 0.32% for the top and 0.56% for the bottom. This supports our null hypothesis; although both generate abnormal returns, the bottom portfolio outperforms the top portfolio and highlights how sensitive the results are to different ESG ratings. It is important to underscore that the Bloomberg ESG Disclosure score does not have the same rating methodology as the Refinitiv scores and is based on each company's disclosure and transparency. Our findings imply that the amount of ESG disclosure in the European market has little value for investors, but the performance quality scores from Refinitiv can be considered to provide valuable information that generates significant excess return using our approach.

Moving on to the American market, we lack significant results for abnormal returns. We only get significant results on a 10% level for the top portfolio with Refinitiv ESG Combined sorting, yielding 0.22% monthly. This, in turn, only holds with the Fama-French 3-factor model and Carhart four-factor model (Appendix A). It is difficult to provide a clear answer to why this is the case in the American market. Still, by looking at previous research by Amel-Zadeh & Serafeim (2018), it could be a result of US investors being less likely to use ESG data in their investment process than European investors.

6.3.2 Factor exposure

The regressions provided interesting findings regarding the market betas of the portfolios. In the US, both the Refinitiv ESG score and the Bloomberg ESG score find a larger market exposure for the top portfolios compared to the ESG combined score, which finds the opposite. This implies that the controversy score contributes to lowering the portfolio's systematic risk, while the scores that do not account for controversy have higher volatility relative to the market in the top portfolios. However, the European study presents exactly the opposite result of the US study.

The lack of disclosure of ESG data could be linked to the company's life cycle. We believe the more experienced, larger, and more stable firms are more likely to disclose their ESG data due to better resources and a moral obligation from the population. This is a plausible claim, supported by Bloomberg ESG Disclosure score portfolios exposure to SMB in our results. The top portfolios have a larger negative exposure to the factor than the bottom portfolios and only the top portfolios have a significant exposure using the Bloomberg ESG Disclosure score.

6.3.3 Limitations of the study

It is important to discuss the limitations of our study as the applied research methodology and data collection affect the results' generalizability. First, we had to choose an asset universe with sufficient ESG score coverage to get a meaningful analysis and significant results. By using both S&P 500 and STOXX Europe 600, we got the possibility to explore differences between European and American markets and minimize data loss of companies not having scores. However, the study utilizes a small sample of both markets, focusing mostly on large-capitalization

firms. This implies that the results could differ with small- and medium-capitalization firms.

We have only focused on two different rating providers, Refinitiv and Bloomberg, with a total of three different ESG scores. While additional scores would be preferable, the lack of sufficient data coverage and subscription requirements of other scores prevented us from including them. This is problematic as there are over 100 rating providers on ESG scores, which would yield different results regarding the portfolio sorting of top and bottom performers. We know from previous research (Berg et al., 2019) that the six leading rating agencies give an average correlation of 0.54 between ESG scores, which poses a challenge for empirical research. In addition, we have only used aggregated ESG scores and not segmented them into Environmental, Social, and Governance scores or even deeper to sub-segments of the three pillars. Further, our sample period was set to include returns from 2011 to 2019. As mentioned before, the reason for this was two-fold. It allowed us to include Bloomberg ESG Disclosure scores and avoid potential noise regarding the Covid-19 pandemic that started in 2020. With an out-of-sample test, it is highly possible that the results would give other conclusions.

7. Conclusion

In our thesis, we have investigated the relationship between ESG ratings and financial performance. Our goal has been to investigate whether a third-party ESG score could be used as a signal in investment strategy. Our strategy involved creating decile top and bottom portfolios based on the last publicly available Refinitiv ESG score, Refinitiv ESG Combined score, and Bloomberg ESG Disclosure score, where these scores determined the portfolio sorting for the following year. Our sample covers the period from 2011 to 2019, and we examine both the European and the US markets.

Our findings, using the two Refinitiv scores, suggest that in the European market ESG ratings can be used as a signal by investors in investment strategies. However, the Bloomberg ESG Disclosure score revealed that the depth and quality of the ESG score is a decisive factor. By applying the Refinitiv ESG score and Refinitiv ESG Combined score to form our portfolios, we find that companies with high ESG scores outperform companies with low ESG scores. These findings are in line with earlier research of Kempf and Osthoff (2007). In contrast, the application of the more simplistic Bloomberg ESG Disclosure score, returned contradictory results to the Refinitiv scores, and are more in line with research by Renneboog et al. (2008). Additionally, the choice of market also had a significant impact on our results. The US study found no significant difference between top and bottom ESG performers, findings that are similar to the findings of Atz et al. (2019). These ambiguous and inconsistent results prevent us from providing a clear answer on the impact of ESG ratings on risk-adjusted stock performance globally.

For future research it would be interesting to get a deeper understanding of why top ESG performers can generate abnormal returns, and what the drivers behind this are. Our suggestion would therefore be to construct factor replicating portfolios to provide more context. Until a common framework for ESG rating is implemented, research in the field will often suffer from inconsistent results without a clear direction of impact. The recent introduction of EU taxonomy for sustainable activities may diminish the discrepancy between providers, and in the aftermath of Covid-19, new data may be utilized to find conclusive results on the impact of ESG in financial markets and the driving factors behind it.

8. References

- Amel-Zadeh, A., & Serafeim, G. (2017). Why and How Investors Use ESG Information: Evidence from a Global Survey. *Financial Analysts Journal*, 73(3), 87–103. <https://doi.org/10.2139/ssrn.2925310>
- Atz, U., Van Holt, T., & Liu, Z. Z. (2020). Do Corporate Sustainability and Sustainable Finance Generate Better Financial Performance? Review, Meta-analysis, and Propositions. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3708495>
- Berg, F., Kölbel, J., & Rigobon, R. (2019). Aggregate Confusion: The Divergence of ESG Ratings. *Forthcoming Review of Finance*. <https://doi.org/10.2139/ssrn.3438533>
- Bloomberg. (n.d.). *Global Environmental, Social & Governance – ESG Data*. Retrieved March 30, 2022, from <https://www.bloomberg.com/professional/dataset/global-environmental-social-governance-data/>
- Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, 47(5), 12–87. <https://doi.org/10.2307/1911963>
- Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance*, 52(1), 57–82.
- Chen, J. (2019). *Socially Responsible Investment (SRI)*. Investopedia. <https://www.investopedia.com/terms/s/sri.asp>
- Eccles, R. G., Ioannou, I., & Serafeim, G. (2014). The Impact of Corporate Sustainability on Organizational Processes and Performance. *Management Science*, 60(11), 2835–2857. <https://doi.org/10.1287/mnsc.2014.1984>
- Eurosif. (n.d.). *Sustainable investment*. Retrieved February 10, 2022, from <https://www.eurosif.org/sustainable-investment/>
- Fama, E. F., & French, K. R. (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, 33(1), 3–56.
- Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1–22. <https://doi.org/10.1016/j.jfineco.2014.10.010>
- French, K. (2022). *Kenneth R. French - Data Library*. http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

- Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, 5(4), 210–233.
<https://doi.org/10.1080/20430795.2015.1118917>
- Gibson, R., Krueger, P., Riand, N., & Schmidt, P. S. (2019). ESG Rating Disagreement and Stock Returns. *Financial Analyst Journal*, *Forthcoming*. <https://doi.org/10.2139/ssrn.3433728>
- GSIA. (2021). *Global Sustainable Investment Review 2020*. <http://www.gsi-alliance.org/>
- Hoepner, A. G. F., Oikonomou, I., Sautner, Z., Starks, L. T., & Zhou, X. (2018). ESG Shareholder Engagement and Downside Risk. *AFA 2018 Paper, European Corporate Governance Institute – Finance Working Paper No. 671/2020*. <https://doi.org/10.2139/ssrn.2874252>
- Hong, H., & Kacperczyk, M. (2009). The price of sin: The effects of social norms on markets. *Journal of Financial Economics*, 93(1), 15–36.
<https://doi.org/10.1016/j.jfineco.2008.09.001>
- IFC. (2005, November). *Who Cares Wins: One Year On*. www.ifc.org.
https://www.ifc.org/wps/wcm/connect/topics_ext_content/ifc_external_corporate_site/sustainability-at-ifc/publications/publications_report_whocareswins-oneyearon__wci__1319579364327
- Kempf, A., & Osthoff, P. (2007). The Effect of Socially Responsible Investing on Portfolio Performance. *European Financial Management*, 13(5), 908–922.
<https://doi.org/10.1111/j.1468-036x.2007.00402.x>
- Li, Y., Gong, M., Zhang, X.-Y., & Koh, L. (2018). The impact of environmental, social, and governance disclosure on firm value: The role of CEO power. *The British Accounting Review*, 50(1), 60–75.
<https://doi.org/10.1016/j.bar.2017.09.007>
- Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 7(1), 77–91.
<https://doi.org/10.1111/j.1540-6261.1952.tb01525.x>
- MSCI. (n.d.). *The Evolution of ESG Investing*. Retrieved February 24, 2022, from <https://www.msci.com/esg-101-what-is-esg/evolution-of-esg-investing>
- Refinitiv. (n.d.). *ESG Scores*. www.refinitiv.com. Retrieved February 24, 2022, from <https://www.refinitiv.com/en/sustainable-finance/esg-scores>

- Refinitiv. (2022). *ENVIRONMENTAL, SOCIAL AND GOVERNANCE (ESG) SCORES FROM REFINITIV*. Retrieved February 13, 2022, from https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/refinitiv-esg-scores-methodology.pdf
- 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. <https://doi.org/10.1016/j.jcorpfin.2008.03.009>
- S&P Dow Jones Indices. (n.d.). *S&P 500® - S&P Dow Jones Indices*. [Www.spglobal.com](http://www.spglobal.com). Retrieved February 24, 2022, from <https://www.spglobal.com/spdji/en/indices/equity/sp-500/#overview>
- Stevens, P. (2020, September 2). *ESG index funds hit \$250 billion as pandemic accelerates impact investing boom*. CNBC. <https://www.cnbc.com/2020/09/02/esg-index-funds-hit-250-billion-as-us-investor-role-in-boom-grows.html>
- STOXX. (n.d.). *STOXX Digital | STOXX® Europe 600*. [Www.stoxx.com](http://www.stoxx.com). Retrieved February 13, 2022, from <https://www.stoxx.com/index-details?symbol=SXXP>
- Sustainalytics. (n.d.). *ESG Risk Ratings*. Retrieved February 24, 2022, from <https://www.sustainalytics.com/corporate-solutions/esg-risk-ratings>
- White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, 48(4), 817. <https://doi.org/10.2307/1912934>
- Yu, E. P., Luu, B. V., & Chen, C. H. (2020). Greenwashing in environmental, social and governance disclosures. *Research in International Business and Finance*, 52(101192). <https://doi.org/10.1016/j.ribaf.2020.101192>

9. Appendix

A: Regression outputs

The tables below present the summary of regression outputs using the different factor models, the tables show t-values in parentheses and * is used to visualize significant p-values; ***= $p < 0.01$, **= $p < 0.05$, *= $p < 0.1$, the adjusted R-squared for each regression are presented in the last row. Explanations of abbreviations are explained in the table of abbreviations.

Table 8: Fama-French 5 with momentum regression results (Value-weighted)

Fama French Five Factor Model With Momentum: Regressions output VW												
	Refintiv ESG Combined Score				Refintiv ESG Score				Bloomberg Disclosure Score			
	US		Europe		US		Europe		US		Europe	
	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom
Intercept	0.0017 (1.528)	0.0008 (0.520)	0.0056** (2.100)	0.0038 (1.390)	0.0003 (0.309)	0.0003 (0.250)	0.0051*** (2.798)	0.0015 (0.610)	-0.0008 (-0.742)	0.0011 (0.720)	0.0037** (2.103)	0.0061** (2.373)
Mkt-Rf	0.9700*** (28.325)	1.0182*** (21.359)	0.7552*** (11.102)	0.6227*** (9.008)	1.0275*** (35.171)	0.9331*** (23.541)	0.6940*** (14.787)	0.8552*** (13.564)	0.9892*** (31.830)	0.9551*** (21.507)	0.6257*** (13.984)	0.6373*** (9.458)
SMB	-0.0387 (-0.591)	-0.3280*** (-3.598)	-0.3978** (-2.514)	-0.5117*** (-3.183)	-0.3889*** (-6.962)	-0.1000 (-1.314)	-0.6092*** (-5.581)	0.2820* (1.923)	-0.1873*** (-3.152)	-0.1101 (-1.296)	-0.5725*** (-5.501)	-0.0953 (-0.758)
HML	-0.3076*** (-3.757)	-0.0380 (-0.333)	-0.5567** (-2.170)	-0.3268 (-1.254)	-0.0399 (-0.571)	0.0814 (0.855)	-0.3593** (-2.030)	-0.5203** (-2.188)	0.0765 (1.030)	-0.0002 (-0.002)	-0.0915 (-0.542)	-0.4150* (-1.832)
RMW	-0.1301 (-1.377)	-0.1718 (-1.306)	-0.2663 (-0.865)	-0.4853 (-1.551)	0.0045 (0.056)	0.0886 (0.807)	-0.0893 (-0.420)	-0.4646 (-1.628)	0.1068 (1.245)	0.0165 (0.135)	0.1375 (0.679)	-0.3202 (-1.227)
CMA	0.4889*** (4.311)	-0.0340 (-0.215)	-0.2956 (-0.996)	0.1637 (0.543)	0.1219 (1.260)	0.1998 (1.516)	0.0879 (0.430)	-0.2902 (-1.055)	0.1648 (1.601)	-0.1087 (-0.739)	0.1196 (0.613)	-0.1770 (-0.737)
WML	-0.0035 (-0.078)	0.0898 (1.418)	0.0228 (0.225)	-0.1114 (-1.081)	-0.0014 (-0.036)	0.0834 (1.576)	-0.0621 (-0.889)	0.0150 (0.159)	-0.0491 (-1.188)	0.0856 (1.450)	-0.0655 (-0.997)	-0.0809 (-0.801)
R^2_{adj}	0.909	0.838	0.627	0.582	0.931	0.859	0.766	0.710	0.921	0.845	0.757	0.647

Table 9: Carhart 4 factor model regression results (Value-weighted)

Carhart Four Factor Model: Regressions output VW												
	Refintiv ESG Combined Score				Refintiv ESG Score				Bloomberg Disclosure Score			
	US		Europe		US		Europe		US		Europe	
	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom
Intercept	0.0021* (1.745)	0.0004 (0.280)	0.0046* (1.820)	0.0023 (0.894)	0.0003 (0.750)	0.0006 (0.466)	0.0047*** (2.714)	0.0002 (0.103)	-0.0005 (-0.474)	0.0009 (0.595)	0.0039** (2.349)	0.0052** (2.169)
Mkt-Rf	0.9617*** (26.340)	1.0299*** (21.818)	0.7887*** (12.689)	0.6189*** (9.737)	1.0249*** (35.384)	0.9239*** (23.353)	0.6917*** (16.218)	0.8863*** (15.228)	0.9786*** (31.334)	0.9609*** (22.104)	0.6169*** (15.154)	0.6587*** (10.731)
SMB	-0.0064 (-0.105)	-0.2576*** (-3.290)	-0.3246** (-2.201)	-0.4479*** (-2.969)	-0.3796*** (-7.901)	-0.1377** (-2.099)	-0.5912*** (-5.842)	0.3596** (2.604)	-0.2192*** (-4.232)	-0.1208* (-1.675)	-0.5873*** (-6.079)	-0.0352 (-0.286)
HML	-0.0417 (-0.730)	-0.0657 (-0.888)	-0.5380*** (-4.318)	0.0062 (0.048)	-0.0265 (-0.584)	0.1553** (2.506)	-0.2929*** (-3.426)	-0.3585*** (-3.073)	0.1187** (2.426)	-0.0715 (-1.050)	-0.1432* (-1.755)	-0.3046* (-1.844)
WML	0.0324 (0.674)	0.0985 (1.587)	-0.0059 (-0.060)	-0.1023 (-1.021)	0.0064 (0.169)	0.0901* (1.731)	-0.0559 (-0.832)	-0.0145 (-0.159)	-0.0451 (-1.098)	0.0785 (1.373)	-0.0552 (-0.861)	-0.0995 (-0.979)
R^2_{adj}	0.894	0.837	0.628	0.5777	0.930	0.858	0.769	0.704	0.918	0.848	0.759	0.646

Table 10: Fama-French 3 factor model regression results (Value-weighted)

Fama French Three Factor Model: Regressions output VW												
	Refintiv ESG Combined Score				Refintiv ESG Score				Bloomberg Disclosure Score			
	US		Europe		US		Europe		US		Europe	
	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom
Intercept	0.0022* (1.855)	0.0008 (0.491)	0.0045* (1.903)	0.0015 (0.609)	0.0003 (0.347)	0.0009 (0.694)	0.0042** (2.587)	0.0001 (0.057)	-0.0006 (-0.624)	0.0011 (0.780)	0.0034** (2.193)	0.0044** (2.003)
Mkt-Rf	0.9547*** (27.326)	1.0088*** (22.113)	0.7896*** (13.137)	0.6343*** (10.267)	1.0235*** (37.002)	0.9046*** (23.605)	0.7001*** (16.918)	0.8884*** (15.785)	0.9883*** (32.946)	0.9441*** (22.538)	0.6252*** (15.825)	0.6736*** (10.668)
SMB	-0.0067 (-0.111)	-0.2587*** (-3.280)	-0.3245** (-2.211)	-0.4463** (-2.958)	-0.3797*** (-7.940)	-0.1387** (-2.094)	-0.5904*** (-5.843)	0.3598*** (2.618)	-0.2187*** (-4.219)	-0.1216* (-1.680)	0.5864*** (-6.078)	-0.0337 (-0.272)
HML	-0.0614 (-1.252)	-0.1255* (-1.961)	-0.5352*** (-4.655)	0.0548 (0.464)	-0.0304 (-0.782)	0.1005* (1.869)	-0.2663*** (-3.364)	-0.3516*** (-3.266)	0.1461*** (3.471)	-0.1192** (-2.028)	-0.1169 (-1.548)	-0.2573* (-1.689)
R^2_{adj}	0.895	0.835	0.631	0.588	0.931	0.855	0.770	0.707	0.918	0.846	0.760	0.644

Table 11: Fama-French 5 regression results (Equally-weighted)

Fama French Five Factor Model: Regressions output EW												
	Refintiv ESG Combined Score				Refintiv ESG Score				Bloomberg Disclosure Score			
	US		Europe		US		Europe		US		Europe	
	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom
Intercept	0.0017 (1.578)	0.0000 (-0.048)	0.0068*** (2.947)	0.0004 (0.207)	0.0006 (0.727)	-0.0005 (-0.347)	0.0060*** (3.068)	0.0011 (0.539)	-0.0014 (-1.298)	0.0014 (1.187)	0.0026 (1.390)	0.0041** (2.025)
Mkt-Rf	1.0557*** (32.557)	0.9587*** (22.015)	0.7306*** (11.043)	0.7096*** (13.502)	1.0046*** (39.831)	0.9210*** (22.650)	0.7248*** (14.170)	0.7064*** (13.058)	1.0051*** (30.587)	0.9284*** (27.685)	0.7115*** (14.426)	0.7390*** (13.744)
SMB	0.0893 (1.387)	0.1066 (1.155)	-0.2837** (-2.215)	0.0080 (0.065)	-0.0643 (-1.284)	0.2730*** (3.382)	-0.5332*** (-4.462)	0.1807 (1.430)	0.0034 (0.052)	0.2606*** (3.914)	-0.3400*** (-2.950)	0.0401 (0.319)
HML	-0.1371* (-1.919)	0.1038 (0.934)	-0.1221 (-0.557)	-0.0130 (-0.068)	-0.1063* (-1.913)	0.1619* (1.808)	-0.0267 (-0.144)	-0.1103 (0.197)	0.0293 (0.405)	0.0114 (0.154)	0.1689 (0.943)	-0.2193 (-1.123)
RMW	-0.0570 (-0.619)	0.0154 (0.138)	-0.1740 (-0.568)	-0.1492 (-0.623)	0.0604 (0.843)	0.0687 (0.595)	-0.1203 (-0.516)	-0.2392 (-0.971)	0.1613* (1.729)	0.0830 (0.872)	0.2558 (1.139)	-0.3588 (-1.465)
CMA	0.3247*** (2.947)	0.1005 (0.599)	-0.4478 (-1.596)	-0.1172 (-0.526)	0.4321*** (5.042)	0.1156 (0.837)	-0.1150 (-0.530)	-0.2392 (-0.918)	0.2838** (2.542)	0.0010 (0.009)	0.0907 (0.433)	-0.2574 (-1.128)
R^2_{adj}	0.927	0.869	0.714	0.728	0.945	0.872	0.765	0.711	0.911	0.909	0.758	0.731

Table 12: Fama-French 5 with momentum regression results (Equally-weighted)

Fama French Five Factor Model With Momentum: Regressions output EW												
	Refintiv ESG Combined Score				Refintiv ESG Score				Bloomberg Disclosure Score			
	US		Europe		US		Europe		US		Europe	
	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom
Intercept	0.0020* (1.779)	0.0002 (0.178)	0.0080*** (3.614)	0.0012 (0.600)	0.0009 (1.123)	-0.0003 (-0.199)	0.0066*** (3.302)	0.0018 (0.856)	-0.0009 (30.656)	0.0017 (1.465)	0.0033* (1.730)	0.0049** (2.351)
Mkt-Rf	1.0408*** (31.061)	0.9370*** (22.258)	0.7182*** (12.721)	0.7011*** (13.347)	0.9838*** (38.858)	0.9075*** (21.468)	0.7178*** (13.994)	0.6991*** (12.885)	0.9668*** (30.656)	0.9082*** (26.448)	0.7038*** (14.264)	0.7306*** (13.585)
SMB	0.0818 (1.276)	0.0956 (1.188)	-0.2707** (-2.061)	0.0170 (0.139)	-0.0749 (-1.546)	0.2662*** (3.293)	-0.5259*** (-4.408)	0.1885 (1.494)	-0.0160 (-0.265)	0.2504*** (3.814)	-0.3320*** (-2.892)	0.0490 (0.392)
HML	-0.1967** (-2.455)	0.0170 (0.169)	-0.2476 (-1.163)	-0.0993 (-0.501)	-0.1897*** (-3.134)	0.1078 (1.067)	-0.0973 (-0.503)	-0.1849 (-0.904)	-0.1241 (-1.646)	-0.0693 (-0.844)	0.0914 (0.491)	-0.3050 (-1.504)
RMW	-0.0793 (-0.858)	-0.0171 (-0.147)	-0.1667 (-0.652)	-0.1442 (-0.606)	0.0292 (0.418)	0.0485 (0.679)	-0.1162 (-0.500)	-0.2349 (-0.956)	0.1039 (1.194)	0.0528 (0.558)	0.2603 (1.165)	-0.3538 (-1.453)
CMA	0.3549*** (3.198)	0.1444 (1.036)	-0.3207 (-1.302)	-0.0298 (-0.130)	0.4743*** (5.657)	0.1430 (1.022)	-0.0435 (-0.195)	-0.1352 (-0.572)	0.3614*** (3.460)	0.0418 (0.714)	0.1692 (0.786)	-0.1707 (-0.727)
WML	-0.0711 (-1.597)	-0.1035* (-1.850)	-0.1720** (-2.045)	-0.1183 (-1.512)	-0.0995*** (-2.958)	-0.0646 (-1.149)	-0.0967 (-1.265)	-0.1023 (-1.265)	-0.1830*** (-4.366)	-0.0962** (2.109)	-0.1063 (-1.446)	-0.1174 (-1.465)
R^2_{adj}	0.928	0.872	0.723	0.732	0.949	0.872	0.767	0.712	0.925	0.912	0.760	0.734

Table 13: Carhart 4 factor model regression results (Equally-weighted)

Carhart Four Factor Model: Regressions output EW												
	Refintiv ESG Combined Score				Refintiv ESG Score				Bloomberg Disclosure Score			
	US		Europe		US		Europe		US		Europe	
	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom
Intercept	0.0023** (2.011)	0.0004 (0.313)	0.0073*** (3.005)	0.0008 (0.437)	0.0015 (1.540)	0.0000 (0.031)	0.0061*** (3.225)	0.0012 (0.600)	-0.0004 (-0.335)	0.0019 (1.538)	0.0040** (2.178)	0.0039* (1.972)
Mkt-Rf	1.0341*** (30.033)	0.9328*** (22.521)	0.7515*** (11.619)	0.7056*** (14.808)	0.9695*** (33.847)	0.8995*** (21.553)	0.7272*** (15.620)	0.7134*** (14.429)	0.9518*** (28.715)	0.9030*** (19.426)	0.6870*** (15.224)	0.7508*** (15.212)
SMB	0.0940 (1.646)	0.0950 (1.382)	-0.2135* (-1.643)	0.0327 (0.289)	-0.1028** (-2.163)	0.2358*** (3.407)	-0.4932*** (-4.465)	0.2246* (1.915)	-0.0682 (-1.241)	0.2229*** (4.365)	-0.3727*** (-3.480)	0.1058 (0.904)
HML	0.0072 (0.133)	0.1049 (1.618)	-0.2929** (-2.352)	-0.0267 (0.280)	0.0369 (0.823)	0.2046*** (3.131)	-0.0666 (-0.714)	-0.0967 (-0.975)	0.0395 (0.761)	-0.0266 (-0.434)	-0.0029 (-0.032)	-0.1670* (-1.688)
WML	-0.0462 (-1.021)	-0.0944* (-1.732)	-0.2017* (-1.750)	-0.1223 (-1.629)	-0.0735* (-1.951)	-0.0597 (-1.088)	-0.1021 (-1.391)	-0.1162 (-1.491)	-0.1680*** (-3.852)	-0.0975 (-1.617)	-0.0893 (-1.256)	-0.1357* (-1.744)
R^2_{adj}	0.922	0.873	0.722	0.736	0.933	0.873	0.770	0.714	0.915	0.913	0.760	0.732

Table 14: Fama-French 3 factor model regression results (Equally-weighted)

Fama French Three Factor Model: Regressions output EW												
	Refintiv ESG Combined Score				Refintiv ESG Score				Bloomberg Disclosure Score			
	US		Europe		US		Europe		US		Europe	
	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom
Intercept	0.0021* (1.893)	0.0001 (0.074)	0.0057*** (2.792)	-0.0001 (-0.067)	0.0012 (1.276)	-0.0002 (-0.114)	0.0053*** (2.926)	0.0003 (0.147)	-0.0009 (-0.799)	0.0015 (1.367)	0.0033* (1.878)	0.0029 (1.493)
Mkt-Rf	1.0440*** (31.595)	0.9530*** (23.132)	0.7818*** (15.206)	0.7240*** (15.514)	0.9852*** (35.375)	0.9122*** (22.762)	0.7425*** (16.341)	0.7308*** (15.124)	0.9878*** (29.178)	0.9238*** (28.017)	0.7004*** (15.930)	0.7711*** (15.924)
SMB	0.0945 (1.654)	0.0960 (1.216)	-0.2104* (-1.676)	0.0345 (0.303)	-0.1020** (-2.118)	0.2365*** (3.413)	-0.4917*** (-4.431)	0.2264* (1.918)	-0.0664 (-1.135)	0.2239*** (3.929)	-0.3713*** (-3.458)	0.1079 (0.912)
HML	0.0353 (0.761)	0.1623** (2.060)	-0.1969** (-2.002)	0.0315 (0.353)	0.0816** (2.088)	0.2409*** (4.284)	-0.0180 (-0.207)	-0.0414 (-0.448)	0.1416*** (2.980)	0.0327 (0.706)	0.0396 (0.471)	-0.1024 (-1.106)
R^2_{adj}	0.922	0.871	0.709	0.731	0.931	0.872	0.768	0.711	0.904	0.910	0.759	0.727

Table 15: Subperiod (2011-2016) Fama-French 5 regression results (Value-weighted)

Fama French Five Factor Model: Sub-period (2011-2016) Regressions output VW													
	Refinitiv ESG Combined Score				Refinitiv ESG Score				Bloomberg Disclosure Score				
	US		Europe		US		Europe		US		Europe		
	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	
Intercept	0.0006 (0.343)	0.0019 (0.923)	0.0082** (2.003)	0.0025 (0.632)	-0.0008 (-0.594)	0.0012 (0.668)	0.0071*** (2.970)	0.0013 (0.384)	-0.0012 (-0.920)	0.0017 (0.886)	0.0035 (1.590)	0.0100*** (3.376)	
Mkt-Rf	0.9865*** (15.564)	1.018*** (16.447)	0.6990*** (6.014)	0.6763*** (7.751)	1.0068*** (24.386)	0.8799*** (17.211)	0.6201*** (8.656)	0.8885*** (5.460)	0.9363*** (24.455)	1.0097*** (18.323)	0.5427*** (8.816)	0.5668*** (7.892)	
SMB	-0.0219 (-0.223)	-0.1813 (-1.275)	-0.4748** (-2.243)	-0.4657** (-2.324)	-0.3867*** (-1.751)	-0.0053 (-0.052)	-0.6976*** (-5.685)	0.3706 (1.542)	-0.2710*** (-3.591)	-0.0055 (-0.960)	-0.6174*** (-6.261)	-0.1599 (-1.085)	
HML	-0.3433*** (-2.841)	-0.0166 (-0.762)	-0.6112* (-1.770)	-0.5212** (-1.963)	-0.0005 (-0.006)	-0.048 (-0.412)	-0.4445* (-1.873)	-0.5410* (-1.805)	0.1390 (1.593)	-0.1662 (-0.191)	-0.1182 (-0.598)	-0.6469*** (-2.641)	
RMW	-0.2134 (-1.146)	-0.3340* (-1.952)	-0.6435 (-1.330)	-0.4929 (-1.281)	-0.0599 (-0.490)	0.0602 (0.397)	-0.425 (-1.264)	-0.3424 (-0.696)	-0.0354 (-0.312)	0.1674 (0.310)	-0.1789 (-0.768)	-0.8589** (-2.436)	
CMA	0.6504*** (3.326)	0.1762 (0.769)	-0.5095 (-1.252)	0.3686 (1.125)	0.1976 (-1.350)	0.3158* (1.743)	0.0618 (0.235)	-0.305 (-0.592)	0.1946 (1.434)	-0.0224 (-0.901)	-0.0884 (-0.421)	-0.2335 (-0.685)	
R^2_{adj}	0.910	0.875	0.621	0.652	0.919	0.858	0.777	0.670	0.922	0.871	0.805	0.704	

Table 16: Subperiod (2016-2019) Fama-French 5 regression results (Value-weighted)

Fama French Five Factor Model: Sub-period (2016-2019) Regressions output VW													
	Refinitiv ESG Combined Score				Refinitiv ESG Score				Bloomberg Disclosure Score				
	US		Europe		US		Europe		US		Europe		
	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	
Intercept	0.0032** (2.066)	-0.0010 (-0.535)	0.0044 (1.250)	0.0022 (0.512)	0.0020* (1.737)	-0.0002 (-0.075)	0.0030 (1.115)	0.0010 (0.372)	-0.0005 (-0.354)	0.0008 (0.380)	0.0038 (1.296)	0.0015 (0.394)	
Mkt-Rf	0.9303*** (18.880)	0.9649*** (15.907)	0.7887*** (7.285)	0.6158*** (3.630)	1.0347*** (28.325)	0.9771*** (15.136)	0.8006*** (7.133)	0.8562*** (9.844)	1.0589*** (22.16)	0.8663*** (12.436)	0.7898*** (7.223)	0.7584*** (6.168)	
SMB	-0.4001 (-0.096)	-0.4502*** (-3.796)	-0.3035 (-1.338)	-0.6373** (-2.314)	-0.3751*** (-5.253)	-0.2741** (-2.172)	-0.6355*** (-3.784)	0.1542 (0.717)	-0.1042 (-1.115)	-0.2242 (-1.646)	-0.6676*** (-3.337)	-0.2137 (-0.947)	
HML	-0.2463** (-2.467)	-0.0035 (-0.029)	-0.6566 (-1.578)	0.2685 (0.414)	-0.0899 (-1.214)	0.0878 (0.671)	-0.1855 (-0.453)	-0.5029* (-1.731)	0.0782 (-0.807)	0.0638 (0.452)	-0.064 (-0.156)	-0.0242 (-0.056)	
RMW	-0.0324 (-0.247)	0.1016 (-0.629)	0.0271 (0.053)	-0.4254 (-0.603)	0.0364 (0.374)	0.0072 (0.042)	0.1124 (0.263)	-0.5805* (-1.767)	0.2213* (1.738)	-0.1384 (-0.745)	0.3142 (0.665)	0.0412 (0.108)	
CMA	0.3548** (2.257)	0.0013 (-0.058)	0.0909 (0.496)	-0.4471 (-0.695)	0.1362 (1.160)	0.1803 (0.875)	0.0472 (0.117)	-0.2659 (-0.735)	0.1876 (1.230)	0.2658 (1.196)	0.4114 (0.874)	-0.2696 (-0.655)	
R^2_{adj}	0.910	0.861	0.631	0.474	0.953	0.849	0.754	0.764	0.929	0.814	0.716	0.612	

B: Model testing

White Test, test for Heteroscedasticity:

The tables below present the original outputs from the test for heteroscedasticity both on our value-weighted portfolios and equally weighted portfolios. The tables present both test statistics (chi-squared values) and p-values for each portfolio with p-values in parentheses. The test has a null hypothesis that the variance in the error term is constant. In portfolios where the outputs show significant p-values, we have accounted for heteroscedasticity by running heteroscedasticity consistent standard errors in the final code.

Table 17: White Test on value-weighted portfolios

White Test For Heteroscedasticity: Output VW												
	Refintiv ESG Combined Score				Refintiv ESG Score				Bloomberg Disclosure Score			
	US		Europe		US		Europe		US		Europe	
	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom
FF3	5.4985 (0.7888)	12.9447 (0.1651)	10.1864 (0.3356)	8.7081 (0.4646)	9.1780 (0.4210)	11.6573 (0.2333)	16.1215 (0.0643)	6.8923 (0.6483)	6.8531 (0.6524)	8.5748 (0.4774)	4.9092 (0.8421)	28.2723 (0.0008)
Carhart	5.5901 (0.9757)	14.9582 (0.3810)	14.7422 (0.3959)	12.4021 (0.5740)	14.8043 (0.3916)	11.6139 (0.6372)	22.9002 (0.0619)	7.6570 (0.9464)	8.8129 (0.8428)	11.3229 (0.6604)	11.0455 (0.6824)	38.7316 (0.0004)
FF5	17.4039 (0.6266)	20.3882 (0.4338)	20.7524 (0.4118)	15.4983 (0.7472)	13.5926 (0.8505)	16.4167 (0.6904)	35.0128 (0.0200)	31.0245 (0.0548)	11.6586 (0.9273)	16.0056 (0.7162)	19.8057 (0.4701)	40.1716 (0.0047)
FF5+M	19.2384 (0.8619)	24.0062 (0.6299)	26.8191 (0.4735)	18.9670 (0.8711)	20.8655 (0.7925)	19.1795 (0.8633)	39-2108 (0.0606)	33.3490 (0.1857)	15.2266 (0.9661)	19.5016 (0.8510)	25.0843 (0.5697)	44.3099 (0.0192)

Table 18: White Test on equally-weighted portfolios

White Test For Heteroscedasticity: Regressions output EW												
EW	Refintiv ESG Combined Score				Refintiv ESG Score				Bloomberg Disclosure Score			
	US		Europe		US		Europe		US		Europe	
	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom
FF3	4.7369 (0.8566)	19.5541 (0.0008)	13.9085 (0.1256)	12.4283 (0.1902)	5.6833 (0.7711)	9.0169 (0.4357)	15.0321 (0.0900)	16.4020 (0.0589)	1.7990 (0.9942)	3.1476 (0.1560)	8.9464 (0.4422)	10.9277 (0.2806)
Carhart	11.6692 (0.6328)	22.4705 (0.0694)	29.1902 (0.0098)	14.1957 (0.4352)	9.3197 (0.8100)	14.5022 (0.4130)	18.0772 (0.2032)	21.2998 (0.0941)	8.7170 (0.8487)	26.1624 (0.0246)	11.2584 (0.6656)	12.9954 (0.5268)
FF5	145.8267 (0.7862)	34.4430 (0.0232)	32.6349 (0.0369)	23.2515 (0.2766)	21.3601 (0.3762)	18.9952 (0.5221)	24.9810 (0.2021)	29.6113 (0.0764)	16.3327 (0.6957)	24.0531 (0.2400)	25.3655 (0.1878)	13.9985 (0.8305)
FF5+M	26.7602 (0.4767)	35.4167 (0.1286)	39.2488 (0.0601)	24.9564 (0.5769)	23.5512 (0.6550)	24.9748 (0.5758)	27-9384 (0.4141)	36.0946 (0.1132)	27.3664 (0.4441)	34.3766 (0.1553)	28.4139 (0.3898)	16.7986 (0.9359)

Breusch-Godfrey test for autocorrelation

The tables below present the output from the Breusch-Godfrey test for autocorrelation for all portfolios. The tables include both test statistics (chi-squared values) and p-values in parentheses. The null hypothesis of the tests is that there is no autocorrelation in the data and the tests show no significant p-values. We, therefore, conclude that there is no autocorrelation in our data, tests have been conducted with multiple numbers of lags.

Table 19: Breusch-Godfrey test on value-weighted portfolios

Breusch-Godfrey Test For Autocorrelation: Output VW												
	Refintiv ESG Combined Score				Refintiv ESG Score				Bloomberg Disclosure Score			
	US		Europe		US		Europe		US		Europe	
	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom
FF3	0.8924 (0.8272)	5.2610 (0.1536)	1.9660 (0.5794)	1.4676 (0.6897)	2.1104 (0.5498)	6.8911 (0.0754)	1.5069 (0.6806)	4.3689 (0.2242)	3.9079 (0.2715)	0.2048 (0.9768)	0.5971 (0.8970)	2.8879 (0.4092)
Carhart	1.4143 (0.7021)	7.1988 (0.0658)	2.0017 (0.5720)	1.0387 (0.7918)	2.1940 (0.5331)	5.9340 (0.1148)	1.9890 (0.5796)	4.4443 (0.2173)	4.2286 (0.2378)	0.0237 (0.9990)	0.3835 (0.9436)	2.7928 (0.4246)
FF5	1.1531 (0.7642)	4.5080 (0.2115)	3.0331 (0.3865)	1.2828 (0.7332)	1.6295 (0.6527)	4.1507 (0.2456)	1.4929 (0.6839)	2.9446 (0.4002)	2.3670 (0.4997)	0.2077 (0.9763)	0.6018 (0.8960)	5.4675 (0.1405)
FF5+M	1.1848 (0.7566)	6.1355 (0.1051)	2.9563 (0.3983)	0.8711 (0.8323)	1.6533 (0.6473)	3.7176 (0.2936)	1.8454 (0.6051)	3.0243 (0.3878)	2.3240 (0.5079)	0.0392 (0.9979)	0.4103 (0.9380)	4.9146 (0.1785)

Table 20: Breusch-Godfrey test on equally-weighted portfolios

Breusch-Godfrey Test For Autocorrelation: Output EW												
	Refintiv ESG Combined Score				Refintiv ESG Score				Bloomberg Disclosure Score			
	US		Europe		US		Europe		US		Europe	
	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom
FF3	1.3210 (0.7241)	5.6181 (0.1317)	2.8281 (0.4188)	7.2126 (0.0654)	0.4496 (0.9298)	5.4622 (0.1409)	1.9598 (0.5807)	1.3152 (0.7255)	0.1466 (0.9857)	0.4474 (0.9302)	0.8587 (0.8353)	0.7229 (0.8677)
Carhart	1.2820 (0.7333)	6.3777 (0.0946)	4.9857 (0.1728)	5.4215 (0.1434)	0.3993 (0.9403)	6.3301 (0.0966)	3.0578 (0.3827)	1.1883 (0.7558)	0.0317 (0.9985)	0.1758 (0.9813)	0.6965 (0.8740)	0.3696 (0.9464)
FF5	1.3822 (0.7096)	5.1811 (0.1590)	5.5277 (0.2098)	7.5401 (0.0565)	1.3881 (0.7083)	5.2082 (0.1571)	2.1483 (0.5421)	1.4732 (0.6884)	0.6730 (0.8795)	0.7342 (0.8651)	0.9293 (0.8183)	1.0479 (0.7896)
FF5+M	1.9989 (0.5726)	5.9363 (0.1147)	6.2578 (0.0997)	5.6272 (0.1312)	1.9517 (0.5824)	6.0727 (0.1081)	3.0663 (0.3770)	1.2334 (0.7449)	0.7156 (0.8697)	0.1950 (0.9783)	0.7552 (0.8601)	0.3838 (0.9435)

Table 21: Augmented Dickey-Fuller test

The table below presents results from the Augmented Dickey-Fuller test for stationarity. The p-values for all portfolios are presented in parentheses. The null hypothesis for the tests states the presence of non-stationarity in the data, the test unveils significant results for all portfolios, and we conclude that all data is stationary.

Augmented Dickey-Fuller Test For Stationarity: Output												
	Refintiv ESG Combined Score				Refintiv ESG Score				Bloomberg Disclosure Score			
	US		Europe		US		Europe		US		Europe	
	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom
VW	-11.7811 (0.0000)	-3.9140 (0.0019)	-9.3231 (0.0000)	-12.0498 (0.0000)	-6.7816 (0.0000)	-3.4196 (0.0103)	-10.5553 (0.0000)	-10.5013 (0.0000)	-12.1665 (0.0000)	-12.3907 (0.0000)	-3.3279 (0.0136)	-11.0098 (0.0000)
EW	-4.6379 (0.0001)	-3.5700 (0.0063)	-4.4015 (0.0002)	-3.6935 (0.0042)	-4.8195 (0.0000)	-13.2952 (0.0000)	-10.3220 (0.0000)	-5.8473 (0.0000)	-12.6634 (0.0000)	-3.5225 (0.0074)	-10.3334 (0.0000)	-9.9210 (0.0000)