NHH

ESG Ratings and Stock Performance

An Empirical Investigation of the Link Between ESG Ratings and Stock Performance of European Large Cap Firms

Erik B. Wilhelmsen & Erik Woods

Supervisor: Carsten G. Bienz

Master Thesis in Financial Economics

NORWEGIAN SCHOOL OF ECONOMICS

Acknowledgments

We wish to express our sincere gratitude to our supervisor Carsten G. Bienz for assisting and giving us encouragement throughout the process of writing this thesis. His insight and enthusiasm for our research has been of great importance. We also want to express our gratitude to the Norwegian School of Economics for facilitating access to databases and literature. This have been crucial for our research. Finally, we thank each other for a rewarding and fulfilling partnership.

Norwegian School of Economics

Bergen, May 31st, 2021

Erik B. Wilhelmsen

Erik Woods

Abstract

This thesis investigates the link between ESG ratings and stock performance of European large capitalization firms. Using ESG ratings from three independent providers – Thomson Reuters, Bloomberg and Sustainalytics – we examine differences in returns of zero-investment portfolios with a long position in the quartile consisting of top ESG-rated firms and a short position the quartile consisting of low ESG-rated firms. We find differences in return based on the choice of rating provider. Only Thomson Reuters portfolios earn negative annual abnormal returns of 6.0%-8.4%, suggesting that investors pay a premium for better ESG-rated firms. Given the observed differences in return and composition of the constructed providers' portfolios, our findings suggest ESG ratings are subjective and may cause confusion when implementing ESG features in a portfolio.

Contents

1. Introduction

Sustainable investing has seen a remarkable rise over the past years and has become one of the mainstays of investing. It involves incorporating environmental, social, and governance (ESG) measures into investment decisions ("MSCI", 2020). The Global Sustainable Investment Alliance has estimated that more than $30 trillion are invested applying ESG criteria ("GSI", 2018). Still, for investors, key questions remain: *How do you incorporate ESG into investment decisions, and what is the risk-return trade-off of such an investment?* As ESG investing is a relatively new trend, investors might be concerned about how incorporating ESG into a portfolio will affect its performance. The increased focus on sustainable investing has led to an emergence of ESG rating providers. These providers analyze firms and assess their corporate sustainability performance by applying their individual methodologies. Consequently, ratings are shown to be subjective (Gibson et al., 2020), which may lead to further confusion for investors that want to incorporate ESG features into their portfolio.

Empirical studies have investigated the effect of corporate sustainability performance on different asset classes, including funds, stocks, and indices (Lima, 2017). On stock level, most research finds a positive relationship between corporate sustainability performance and stock performance (Lima, 2017). This implies that "good" ESG performers are expected to yield higher stock returns than "bad" ESG performers. However, more recent research suggests otherwise, and concludes that the expected return of a company with good ESG practices may be lower than of a company with bad ESG practices (Peng and Yang, 2014; Wang and Sargins, 2020). Firms that are "bad" ESG performers may be considered as riskier. If this is the case, recent findings would be consistent with the risk-return trade-off which states that potential return rises with increased risk. Hence, we should expect to observe a negative relationship between ESG performance and stock performance.

Against this background, we use ESG ratings as a measure of corporate sustainability performance and study its impact on stock performance. Existing literature has mainly used ESG ratings from one provider to assess corporate sustainability performance (Kempf and Osthoff, 2007). We complement existing literature by analyzing ratings from three independent providers in our research to examine differences in portfolio return based on the choice of provider. Additionally, we seek to explain the potential differences in return by investigating

the portfolio variances through breaking down the different portfolios and rating methodologies.

To capture differences in ratings, we include the following rating providers in our analysis: Thomson Reuters[1], Bloomberg, and Sustainalytics. These are leading rating providers thoroughly covering the multidimensional field of corporate sustainability. Selecting these providers, also ensures available ratings for the firms of interest, namely those included on the STOXX Europe 600 Index, as we narrow our analysis to the European market. The index consists of 600 large capitalization firms and provides us with a consistent data sample with ESG ratings available across firms and sectors.

We sort the firms into quartiles and deciles based on ESG ratings from the specific rating providers. The top and bottom quartiles and deciles are of interest, as they include top-rated and low-rated firms in terms of ESG. This allows us to isolate the top and bottom groups containing the best and worst firms in corporate sustainability performance. We create a zero-investment portfolio with a long position in the top group and a short position bottom group. This strategy aims to capture differences in abnormal return between good and bad ESG performers and capture a possible ESG premium or discount. We do this separately for each rating provider and use both equal- and value-weighted portfolio returns in our analysis. Since ESG ratings are updated annually, we rebalance the portfolios at the beginning of each year, as we track monthly stock returns for six years from 2014 to 2019. This approach enables us to examine which portfolios that yield significant abnormal returns using this investment strategy. We aim to point out which providers that treat ESG as a risk factor. Specifically, we examine alphas of the ESG factor portfolios by applying the Fama French framework to account for possible differences in portfolio risk exposures.

Our analysis' descriptive results show that bad ESG performers tend to outperform top ESG performers. All portfolios have a downward trend in terms of cumulative return over the sample period. The most considerable outperformance of bad ESG performers is observed for Thomson Reuters portfolios, followed by Bloomberg. Sustainalytics' portfolios yield a smaller negative outperformance than the portfolio strategy constructed from the other two rating providers. The

---

[1] Thomson Reuters Financial and Risk unit was renamed Refinitiv at the end of 2018 (Reuters, 2018). As our data sample mainly consists of years prior to the name change, we use the name Thomson Reuters in this thesis.

outperformance is greater for the decile portfolios than for the quartile portfolios. These results are the first sign of an inverse relationship between ESG ratings and stock return.

The Fama French regressions show a negative alpha of the ESG factor, confirming the results of our descriptive analysis. Only the constructed Thomson Reuters portfolios are consistently statistically significant for decile and quartile portfolios with a negative average monthly abnormal return ranging from 0.5% to 0.6%. Bloomberg's decile portfolio shows similar evidence. Lastly, the regression results demonstrate that good ESG performers are larger firms in terms of market capitalization than the firms with low ESG rating.

Our findings are interesting given the observed variation in performance of the different portfolios, suggesting ESG ratings are subjective. Only Thomson Reuters consistently picks up the negative abnormal return of the long-short portfolio which may imply that it is the only rating provider that treat ESG as a risk factor. Based on this, we examine the impact of variance of the Thomson Reuters portfolios in relation to the STOXX Europe 600 index as a benchmark. Our findings show significantly less variance in the Thomson Reuters portfolios, which implies that investors pay a premium for better rated ESG firms.

The rest of the thesis is structured as follows. The second section reviews existing literature on corporate sustainability, ESG rating disagreement, and the link between corporate sustainability and stock performance. Section 3 covers our hypothesis. Section 4 describes our data sample and adjustments made. Section 5 describes empirical methodology for the analysis, and section 6 presents our results and discussion. Finally, section 7 concludes the thesis.

2. Literature Review

The second section aims to provide an overview on the field of study and existing literature on the topic. The section is divided into three subsections: a brief introduction to corporate sustainability and investing, ESG rating disagreement, and previous research on the link between corporate sustainability and stock performance.

2.1. A Brief Introduction to Corporate Sustainability

Governments worldwide have enacted more than 500 regulations that attempt to incentivize the implementation of ESG features into investment decisions (KPMG, 2020). This has resulted in a positive shift where investors, shareholders, and governments now require more accurate information on environmental, social, and governance aspects. The Global Sustainable Investment Alliance (GSIA) estimated that more than $30 trillion are invested applying ESG criteria to analyze and select investments (GSIA, 2020). Socially responsible investing ("SRI") and corporate sustainability have thus received increasing attention in financial research over the last two decades (Lima, 2017).

Academic studies have examined the association between ESG ratings with different variables of interest, such as stock market performance, accounting performance, financial constraints, and governance characteristics (Gibson et al., 2020). Lima (2017) covers a thorough analysis of more than 80 empirical papers from the SRI field and further divides previous research into empirical analysis on three different asset classes: funds, stocks, and indices. Funds have received the most attention by comparing SRI funds to conventional benchmarks. The second category, stocks, mainly focuses on constructing different portfolios of individual stocks based on their ESG ratings and comparing them. The third category, indices, has received the least amount of attention in the research. Typically, indices' research is based on comparing the performance of different ESG indices with their respective benchmark. Overall, most previous research choses one specific rating provider and measure the effect of corporate sustainability on the different asset classes (Lima, 2017). We add to this literature by including ratings from three independent providers.

2.2. ESG Rating Disagreement – Empirical Evidence

ESG rating providers have the objective of measuring firms' corporate sustainability performance. Although providers base their ESG ratings on the same pillars (environmental, social, and governance), there is often substantial disagreement as they use different rating methodologies (Christensen et al., 2020). In recent years, financial journalists (Mackintosh, 2018) and academics (Chatterji et al., 2016; Gibson et al., 2020) have drawn attention to this topic. Further, the Securities and Exchange Commission (SEC) commissioner, Hester M. Pierce, has expressed apprehensions regarding the considerable dispersions in ESG ratings. In 2019 she said, *"the different ratings available can vary so widely, and provide such bizarre results that it is difficult to see how they can effectively guide investment decisions".* (Pierce 2019). Hence, investors seeking to implement ESG investment strategies could be deceived by ESG ratings. This issue is further discussed in an article published by the Financial Times, highlighting the importance of ESG providers' methodology. The journalist discusses that in the absence of agreements among providers, investors do not necessarily comprehend how the rating is calculated (Allen, 2018).

Moreover, a working paper published by Harvard Business School argues that, for instance, a greater disclosure score results in a higher disagreement among providers (Christensen et al., 2020). They find a greater disagreement in the case of firms with high- or low ESG ratings, than firms with a medium rating. Other researchers study the impact of ESG rating disagreement on stock returns. Gibson et al. (2020) document a wide dispersion of six prominent rating providers with an average correlation 0.46. Disagreement regarding the firm's financials can result in prices primarily reflecting optimistic investors as pessimistic investors usually are not involved in the market due to high short-sales constraints (Miller 1977). Based on this premise, Gibson et al. (2020) found that ESG rating disagreement results in overvaluation and hence lower financial returns.

Berg et al. (2019) outline three sources of disagreement among rating providers; the use of different categories (scope divergence), the measurement of identical categories in different ways (measurement divergence) and using different weights of the various categories (weight divergence). They argue that the different ratings mainly derive from scope and measurement divergence. To improve ESG data, Douglas et al. (2017) suggests the following improvements: First, they argue that there is a need for consistency in the material data tracked and published

by companies. Secondly, the corporate world needs a widely adopted and verified data provider standard that sets guidelines for a credible approach and reduces discrepancies across the same firm's ratings. They argue that investors should choose different approaches to suit their objectives, but the lack of transparency and basic guidelines create industry-wide credibility challenges.

2.3. Corporate Sustainability and Stock Performance – Empirical Evidence

Empirical studies have investigated the link between corporate sustainability and stock performance with inconclusive results. While several studies find this link to be positive (Kempf and Osthoff, 2007), others find it to be negative (Vance, 1975; Griffin and Mahon, 1997). To investigate the relationship, ESG screening process are implemented to construct portfolios that meet certain sustainability requirements. Positive and negative sustainability screens have allowed researchers to investigate the difference in return (Amel-Zadeh and Serafeim, 2018). Previous research has shown to be limited to investigating certain aspects of the multi-dimensional corporate sustainability field, e.g., only applying an environmental screen. This may lead to wrong inferences being drawn from the papers. More recent studies address this issue by investigating the link by using ESG ratings to capture the overall corporate sustainability performance (Lima, 2017).

Kempf and Osthoff (2007) were among the first researchers to apply the Fama French framework to examine the relationship between corporate sustainability and portfolio performance. They find that the performance of the socially responsible portfolios is never significantly negative, implying that socially responsible investors do not suffer a performance loss by reaching their sustainability goals. Eccles et al. (2014) made similar findings that socially responsible investing does not harm portfolio performance. The approach of this thesis is motivated by these researchers' methodology and findings.

3.   Hypothesis

This section presents the main hypothesis for our thesis and a supplementary research question.

3.1. Main Hypothesis

Investors pay a premium for holding top ESG rated firms – and therefore the return of a firm with good ESG practices is expected to be lower than that of a firm with bad ESG practices.

We expect firms with a low ESG rating to be considered riskier and therefore compensate this risk through higher potential returns. The hypothesis aims to answer if ESG can be considered a risk factor in the financial markets by explaining systematic risk and returns patterns.

3.2. Supplementary research questions

How does the return of the long-short ESG portfolios depend on the choice of rating provider – does it matter which provider an investor decides to use, and what are the drivers of potential differences between ratings?

This thesis aims to explicitly address the first part of the supplementary research question to cover potential differences in return based on the choice of rating providers. Corporate sustainable investors may use ratings as criteria to be included in a portfolio. If there exist large rating differences of the same firms, investors might use these differences to their advantage. This may give misleading information about the corporate sustainability level of their investments.

Specific drivers of potential differences between ratings of the same firm are a more complex matter, which we do not aim to answer in this thesis. Instead, this will be the basis for our discussion as we aim to provide insights on ESG ratings and its role in the financial markets.

4. Data

This section covers the data used in our analysis. This includes data sources and descriptive statistics of the final sample used in our analysis.

4.1. Data Sources

We obtain monthly adjusted stock prices, financial fundamentals, and Thomson Reuters ESG ratings from Thomson Reuters Datastream. Further, we obtain Bloomberg ESG ratings and Sustainalytics ESG ratings from Bloomberg Professional Services. We retrieve the Fama French three factors (Fama & French, 1993), the momentum factor (Carhart, 1997), and the Fama French five factors (Fama & French, 2015) from Kenneth R. French's data library (2020). Furthermore, we obtain the one-month European Central Bank government rate (ECB, 2020) and historical data on the STOXX Europe 600 index (Yahoo, 2020).

4.2. ESG Data and ESG Rating Providers

ESG ratings are obtained from three independent providers, presented in Table 1. Description of each rating provider is included in the following subsections.

Table 1: ESG Ratings – Data Providers

This table outlines the ESG rating providers used in this analysis. We list the providers (Data Provider), the origin of each provider (Origin), the rating scale (Rating Scale), the number of firms in our sample that are rated by the provider (Number of Stocks), and the individual rating pillars, e.g., environmental, social, governance (Pillars).

Data Provider	Origin	Rating Scale	Number of Stocks (Sample)	Pillars[2]
Thomson Reuters	CH	0 – 100	552	E, S, G
Bloomberg	US	0 – 100	488	E, S, G
Sustainalytics	NL	0 – 100	432	E, S, G

---

[2] The data sample includes overall ESG scores, in addition to individual pillar scores: Environmental (E), Social (S) and Governance (G).

4.2.1. Thomson Reuters

Thomson Reuters provides ESG data on more than 6,000 firms dating back to 2002. The provider acquired ASSET4 and is the first agency to offer raw ESG-data to investors (Polk and Llp, 2017). In measuring ESG-ratings, Thomson Reuters employs more than 400 metrics and rate companies on a scale from 0-100. These are sorted into ten groups: resource use, emissions, innovation, management, shareholders, CSR strategy, workforce, human rights, community, and product responsibility – that is then accounting towards a pillar score (Reuters, 2018).

4.2.2. Bloomberg

Bloomberg currently offers ESG data on more than 11,000 firms (Bloomberg, 2020). The firms are assessed on annual basis, on which data is gathered through CSR reports and annual reports. The data covers approximately 120 indicators but differs from other providers with a focus on penalizing companies that are missing ESG data. The ESG rating scale is from 0 to 100. (Bloomberg, 2019)

4.2.3. Sustainalytics

Sustainalytics covers more than 6,500 firms spanning over 42 sectors (Sustainalytics, 2020). It originates from the consolidation of DSR (Netherlands), Scoris (Germany), and AIS (Spain) in 2008. Their methodology is based on dividing issues into three pillars: environmental, social, and governance and rates on a scale 0 to 100. Companies' ratings are based on an analysis of 70 indicators for the respective sectors.

4.3. Sample Selection

To test our hypothesis, we construct a representative and homogenous sample. ESG ratings are often restricted to cross-section and time-series, as it is less available for smaller firms and normally only for recent years (Gibson et al., 2020). We account for this issue by restricting our data sample to firms included on the STOXX Europe 600 index. The index includes 600 firms categorized into large, middle and small capitalization firms across 17 European countries[3] (STOXX, 2020). Overall, this increases the ESG rating availability for the firms of interest.

---

[3] The index covers approximately 90% of the market capitalization of the European stock market (STOXX, 2020), implying the index in general consists of large capitalization firms.

4.3.1. Screening

ESG ratings from Sustainalytics are only available from 2013. We, therefore, restrict our sample from 2014 to 2019[4]. Moreover, we require each firm-observation to have an ESG rating from minimum two of the three providers. By not requiring ESG ratings from all three providers we exclude the possibility for a firm not rated by a specific rating provider to be included in the respective portfolio. This may potentially create differences in the composition of the portfolios. However, we do consider this a realistic approach from an investor's perspective. An investor will likely choose one rating provider for his investment decision, thus limiting the investor to only consider the firms rated by that specific provider.

In summary, we construct a representative sample suitable to our approach that is not overly restrictive. The final data sample includes 552 (Thomson Reuters), 488 (Bloomberg) and 432 (Sustainalytics) firms, as illustrated in Table 1.

4.3.2. Portfolio Selection

This subsection details how we construct the portfolios based on the ESG ratings from the three rating providers.

Our sample is restricted to only include observations in the period from 2014 to 2019, including six years of monthly stock data. ESG ratings are provided annually, and we lag ESG ratings to the following year to make sure the information on corporate sustainability is available to the market when calculating portfolio return.

At the end of year $t - 1$, the individual rating providers report their ratings of the firms. Based on these ESG ratings, we divide the companies into quartiles. Following a long-short zero-investment strategy, we go long in the top quartile consisting of the top-rated firms and short the quartile consisting of low-rated firms. This creates an ESG portfolio that examines the difference in abnormal return between high-rated and low-rated firms. The ESG portfolio is formed at the beginning of year t and held unchanged until the end of year t, as the portfolios rebalance annually. To examine whether our results are sensitive to different portfolio weighting schemes, we construct both equal-weighted and value-weighted portfolios[5].

---

[4] Portfolios are constructed based on the one-year lagged ESG ratings. Described in detail under section 4.3.2.
[5] Calculation of the value-weighted portfolios is described in detail in section 4.4.1.

The portfolio selection process described above is performed for both quartile and decile portfolios and repeated for each rating provider. At the end of year t, we rank the firms based on the updated ratings and construct the portfolios to be held in year $t + 1$. This constitutes a to a time series of six years of monthly returns spanning from 2014 to 2019.

In summary, we construct a total of 72 quartile and decile portfolios[6] based on the ESG ratings from the different providers. The top and bottom quartiles and deciles are used to construct the long-short zero-investment portfolio (ESG portfolio), henceforth "ESG TMB" (ESG Top Minus Bottom). In total, the analysis covers 36 long-short zero-investment portfolios.

4.4. Variable Descriptions

4.4.1. ESG Factor

The return of the ESG TMB factor portfolio is the dependent variable when performing regression analysis. This is a factor investment strategy like the Fama French factors, forming a hedged long-short portfolio to capture the ESG premium or discount. We use both the equal-weighted return and the value-weighted return of the long-short portfolios to capture potential differences. The equal-weighted return is calculated by averaging returns from the top ESG firms minus the returns of the bottom ESG firms. The value-weighted return is calculated as follows:

<div align="center">

Equation 1

$$r_{p_t} = \sum_{i=1}^{N} (w_{it} * r_{it})$$

$$w_{i_t} = \frac{r_{it}}{\sum_{i=1}^{N} r_{it}}$$

</div>

Where r_{p_t} *Value weighted portfolio return at time t*

w_{i_t} *Return on stock i at time t*

---

[6] We explain the total number of portfolios constructed as follows: As the firms are sorted by ESG ratings provided by for example Thomson Reuters, we are interested in the top and bottom quartiles and deciles – in total four portfolios. Annual rebalancing of the portfolios leads to six new sets of these four portfolios – in total 24 portfolios. The same procedure is followed for Bloomberg and Sustainalytics – and that makes up 72 portfolios in total.

If the ESG risk premium exists, we should expect the return on the ESG TMB factor to be negative, thus a discount in factor returns.

4.4.2. Fama French Factors

We obtain the Fama French factors from the Kenneth R. French Library. More specifically, we obtain the Western European market's risk factors applied in the regression analysis. These factors are representative for our analysis as the markets used to construct the factors show great overlap with markets represented in our data sample. The factors will be further explained in section 5.1.

4.5. Descriptive Statistics

Table 2: Descriptive Statistics: Data Sample ESG Attributes

This table presents the ESG attributes of the firms included in the data sample. We include the total number of firms rated (N), mean score (Mean), median (Median) and standard deviation (StdDev) of firm's ESG rating, environmental rating, social rating and governance rating, separately. Correlation of the different providers are also included (Pearson Correlations).

	N	Mean	Median	StdDev		Pearson Correlations	
						Thomson Reut.	Sustainalytics
ESG Rating							
Thomson Reut.	35,241	63.637	66.000	17.669			
Sustainalytics	25,929	75.476	82.609	22.501		0.535	
Bloomberg	31,689	60.215	62.097	14.500		0.592	0.454
					Avg	0.527	
Environmental Rating							
Thomson Reut.	35,241	61.714	66.330	25.031			
Sustainalytics	25,929	73.043	79.464	23.207		0.496	
Bloomberg	31,689	47.414	50.000	26.693		0.529	0.409
					Avg	0.478	
Social Rating							
Thomson Reut.	35,241	67.363	71.144	20.441			
Sustainalytics	25,929	73.571	80.000	23.440		0.475	
Bloomberg	31,689	62.520	65.427	16.152		0.514	0.352
					Avg	0.447	
Governance Rating							
Thomson Reut.	35,241	59.150	61.901	21.659			
Sustainalytics	25,929	70.259	77.612	25.514		0.318	
Bloomberg	31,689	70.711	72.113	13.218		0.327	0.158
					Avg	0.268	

In the following, we present a more detailed overview of the sample with descriptive statistics presented in Table 2. The table shows more observations for Thomson Reuters (35,241) than Bloomberg (31,689) and Sustainalytics (25,929) for the sample. Despite that our sample consists of 35,241 observations, Bloomberg and Sustainalytics have fewer observations, as we only require each firm to have a rating from a minimum of two rating providers. This method corresponds with previous research on ESG rating dispersion (Christensen et al., 2020) and analyst disagreement (Sadka and Scherbina, 2007; Barinov, 2013).

While Thomson Reuters and Bloomberg tend to issue overall ESG ratings of around 60 points (63.6 and 60.2 on average, respectively), Sustainalytics tends to issue higher ratings with an average of 75.4. Additionally, Sustainalytics has the greatest variation among the providers (standard deviation of 22.5), while Bloomberg has the least variance (standard deviation of 14.5). Regarding the three pillars, these trends seem to be resembling. The exception is the environmental pillar, where Bloomberg issues relatively low ratings with a median of 50.0. In comparison, Thomson Reuters and Sustainalytics have a median of 61.7 and 73.0, respectively.

To provide intuition for the providers' dispersion, we include a correlation matrix. The matrix yields interesting results and shows that for total ESG rating the providers have an average correlation of 0.52. These results correspond with the findings of Berg et al. (2019). Furthermore, the correlation is lower among the providers for the pillars than the overall rating. Not surprisingly, the average correlation is highest for the environmental pillar (0.47), and lowest for the governance pillar (0.26). These findings are consistent with previous ESG research (Berg et al., 2020). In general, descriptive evidence suggests that Thomson Reuters and Bloomberg are more coherent than Sustainalytics. Furthermore, Sustainalytics issues higher individual ratings compared to Thomson Reuters and Bloomberg, as evidenced by the mean scores.

5.   Methodology

We construct portfolios following a hedged long-short investment strategy with a long position in top ESG performers and a short position in bottom ESG performers. This allows us to measure differences in performance between the top and bottom portfolios with respect to ESG ratings from the independent rating providers. This will in turn allow us to investigate the historical returns of the ESG TMB portfolio by applying the Fama French framework. More specifically, we use Fama French three-factor, Carhart and five-factor with- and without momentum models. These models are more suitable to estimate portfolio returns than the CAPM (Bartholdy and Pearce, 2005), which has for many applications been invalidated (Fama and French, 2003).

This chapter outlines the approach for our analysis and is divided into two subsections. The first subsection presents the Fama French models' specifications, while the second discusses the assumptions to justify the interpretation of our regression results.

5.1. Model Specification – Fama French Framework

The Fama French models attempt to explains variation in stock returns by including different factors that explains excess return (Womack and Zhang, 2003). The models' goal is to capture all variations in stock prices through the included factors. An intercept of zero implies that the investment strategy does not create any abnormal return, (i.e., no alpha). If an estimate is not statistically significant, there is no difference in exposure to this specific risk factor between the top and the bottom portfolios in the long-short portfolio. For the time-series analysis, we use Ordinary Least Squares (OLS) regressions. The following subsections will briefly explain the different Fama French models and factors.

5.1.1. Fama French Three-Factor Model

The Fama French three-factor model extends the CAPM (Sharpe, 1964; Lintner, 1965; Mossin, 1966) by combining the original market factor and two company-specific risk factors: SMB ("Small Minus Big") and HML ("High Minus Low"). SMB is designed to measure the additional return investors have historically received by investing in stocks of companies with relatively small market capitalization, often referred to as the "size premium". HML is

constructed to measure the "value premium" provided to investors for investing in companies with high book-to-market value.

Equation 2

$$ESG\ TMB_t = \alpha + \beta_{MKT} \times (MKT_t - rf_t) + \beta_{SMB} \times SMB_t + \beta_{HML} \times HML_t + u_t$$

$ESG\ TMB_t$	*ESG factor: Excess return on top minus bottom portfolio at time t*
α	*Intercept/abnormal return/alpha*
β_{MKT}	*Exposure to the market factor (market beta)*
$MKT_t - rf_t$	*Excess return in the market at time t*
β_{SMB}	*Exposure to the size factor*
SMB_t	*The size factor at time t*
β_{HML}	*Exposure to the value factor*
HML_t	*The value factor at time t*
u_t	*Error term at time t*

5.1.2. Carhart Four-Factor Model

The Carhart model is a refinement of the three-factor model, as it adds a factor that accounts for performance tendency (Carhart, 1997). The additional factor is called "momentum" and measures a portfolio's exposure to previous winners and losers.

Equation 3

$$ESG\ TMB_t = \alpha + \beta_{MKT} \times (MKT_t - rf_t) + \beta_{SMB} \times SMB_t + \beta_{HML} \times HML_t$$
$$+ \beta_{MOM} \times MOM_t + u_t$$

β_{MOM}	*Exposure to the momentum factor*
MOM_t	*The momentum factor at time t*

5.1.3. Fama French Five-Factor Model

The five-factor model adds two new factors, RMW ("Robust Minus Weak") and CMA ("Conservative Minus Aggressive"). RMW represents the difference in returns between a portfolio comprising of firms with robust profitability and a portfolio comprising of firms with weak profitability. CMA represents the investment strategy approach. The factor covers the

difference in returns between a portfolio of low investment stocks and a portfolio of high investment stocks (Fama French, 2014).

<div align="center">Equation 4</div>

$$ESG\ TMB_t = \alpha + \beta_{MKT} \times (MKT_t - rf_t) + \beta_{SMB} \times SMB_t + \beta_{HML} \times HML_t$$
$$+ \beta_{RMW} \times RMW_t + \beta_{CMA} \times CMA_t + u_t$$

β_{RMW}	*Exposure to the profitability factor*
RMW_t	*The profitability factor at time t*
β_{CMA}	*Exposure to the investment factor*
CMA_t	*The investment factor at time t*

5.1.4. Fama French Five-Factor Model with Momentum

This model adds the momentum factor presented in section 5.1.2. to the five-factor model presented in the previous subsection.

<div align="center">Equation 5</div>

$$ESG\ TMB_t = \alpha + \beta_{MKT} \times (MKT_t - rf_t) + \beta_{SMB} \times SMB_t + \beta_{HML} \times HML_t$$
$$+ \beta_{RMW} \times RMW_t + \beta_{CMA} \times CMA_t + \beta_{MOM} \times MOM_t + u_t$$

All variables for this model are explained in the previous sub-subsections.

5.2. Model Testing

The data must satisfy certain assumptions to justify the results of our analysis and to be valid for interpretation. First, we run a Breusch-Pagan test to see if we have an issue with heteroscedasticity. The test outputs are presented in the Appendix A1 Model Testing and show that we do have a problem with heteroscedasticity. Furthermore, we run a Breusch-Godfrey test to check for autocorrelation. The regression outputs show that we do have an issue for the equal-weighted Bloomberg quartile portfolio, only. To address heteroscedasticity and autocorrelation problems, we run all regressions with robust standard errors, according to Sarzosa (2012). Model testing, output tables and adjustments made are thoroughly covered in Appendix A1 Model Testing.

6. Analysis

In this section, we present the results of our analysis. More specifically, we test whether a zero-investment portfolio that goes long in the top rated ESG quartiles and shorts the low rated ESG quartile earns significant abnormal returns. We do the same for decile portfolios.

As described in section 4, for every year, we sort firms into quartiles and deciles based on their ESG rating. This is done separately for the three independent rating providers. We then calculate both the equal-weighted and value-weighted monthly returns of each long-short portfolio. Lastly, we test how these portfolios perform relative to the risk-free rate and when controlling for the Fama French factors described in section 5.

6.1. Descriptive Results

6.1.1. Portfolios Attributes

We start off by conducting an initial analysis where we compare the attributes of the constructed top and bottom portfolios. Descriptive results in Table 3 shows that the top ESG portfolios include firms that have higher market capitalization relative to the bottom ESG portfolios. This is consistent for decile and quartile portfolios, but the average market capitalization is greatest for the deciles. Furthermore, the low-rated ESG portfolios have higher average return than the top-rated ESG portfolios. The largest differences are observed for portfolios constructed from Thomson Reuters and Bloomberg ratings. We also observe that the differences between the top and bottom portfolios are larger for the deciles. In summary, the portfolios' attributes provide initial evidence of an inverse relationship between ESG score and average monthly stock return. Thus, our findings may indicate that investors pay a premium for holding high-rated ESG firms in their portfolios.

Table 3: Portfolio Attributes

This table presents the total market capitalization (Market Cap) of the firms included in the constructed portfolios as well as the average monthly return (Monthly Return). The table is split into quartile and decile portfolios.

Portfolio Attributes: Quartile Portfolios

Portfolio	Market Cap (mEUR)	Monthly Return
Thomson Reuters Top Quartile	55,980	0.84%
Bloomberg Top Quartile	42,122	0.89%
Sustainalytics Top Quartile	28,071	0.97%
Thomson Reuters Bottom Quartile	6,117	1.68%
Bloomberg Bottom Quartile	7,166	1.75%
Sustainalytics Bottom Quartile	12,576	1.16%

Portfolio Attributes: Decile Portfolios

Portfolio	Market Cap (mEUR)	Monthly Return
Thomson Reuters Top Decile	40,772	0.80%
Bloomberg Top Decile	34,440	0.96%
Sustainalytics Top Decile	28,008	0.81%
Thomson Reuters Bottom Decile	6,671	1.55%
Bloomberg Bottom Decile	8,627	1.38%
Sustainalytics Bottom Decile	15,856	1.03%

6.1.2. Portfolios Risk and Returns

Next, we examine the cumulative returns of the hedged long-short portfolios by plotting the outcome of investing EUR 1 in the portfolios. Over the sample period, the long-short investments have a negative return. While Thomson Reuters and Bloomberg portfolios are relatively similar with a negative return of almost 50%, Sustainalytics portfolios have a flatter trend with a negative return of approximately 10%. The cumulative returns indicate that the bad ESG portfolios yield higher returns than the good ESG portfolios over the sample period, regardless of rating provider.

Figure 1: Monthly Cumulative Returns of the Decile Long-Short ESG Portfolio

This figure presents the cumulative performance (in EUR) of a EUR 1 investment in the decile long-short portfolios from Jan. 2014 to Dec. 2019.
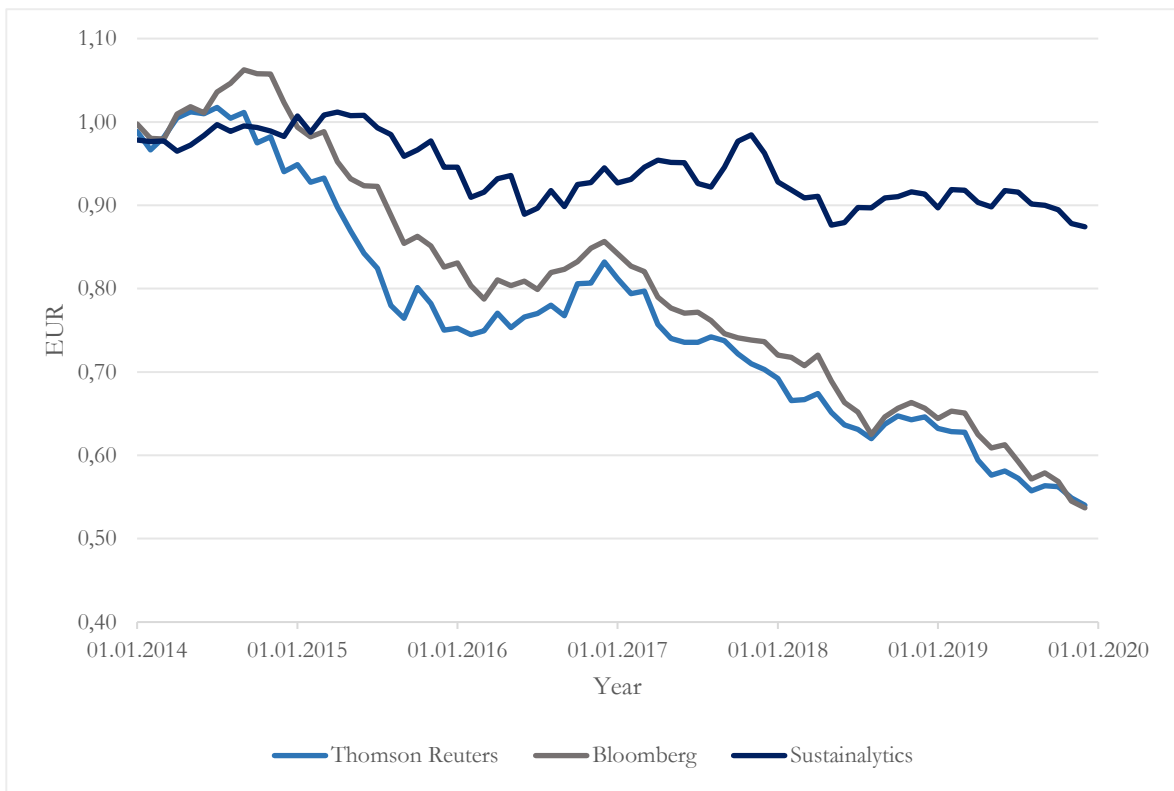
Next, we examine the cumulative returns for the quartile portfolios. The negative trends are not as dramatic as for the decile portfolios, but the downward trends are still evident. There are also larger differences between the three providers than evidenced for the decile portfolios. Once again, Sustainalytics performs best with a negative return of roughly 15%. Both Bloomberg and Thomson Reuters perform better than for the decile portfolios. Bloomberg falls around 25%, while Thomson Reuters clearly shows the worst performance with a negative return of more than 40%.

Figure 2: Monthly Cumulative Returns of the Quartile Long-Short ESG Portfolio

This figure presents the cumulative performance (in EUR) of a EUR 1 investment in the quartiles long-short portfolios from Jan. 2014 to Dec. 2019.
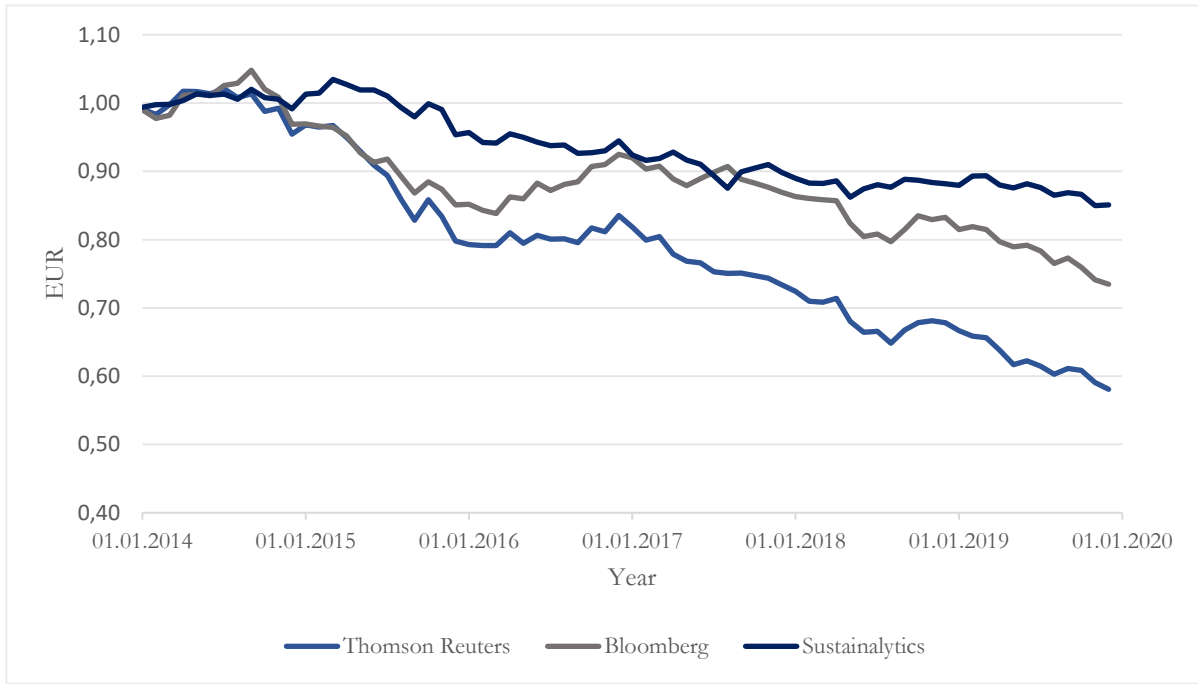
The descriptive statistics, including the cumulative returns, provides us with two key takeaways. First, it indicates which provider you should choose if you are an investor using ESG as an investment criterium. Secondly, it indicates that all portfolios seemingly pick up the negative relationship between ESG performance and stock performance. In the next section we investigate whether this relationship is significant when we control for the Fama French factors.

6.2. Regression Results

This section presents the results from the regression analysis. We test how the portfolios perform relative to the risk-free rate when controlling for the Fama French factors described in section 5.

6.2.1. Fama French Three-Factor Model

Table 4: Fama French Three-Factor Model Regression Outputs

This table presents the results from the Fama French Three-Factor Model. It presents the factor loadings and abnormal returns (in percentage) of both equal-weighted (EW) and value-weighted (VW) zero-investment portfolios with a long position in good ESG performers and short position in bad ESG performers. This first section shows the quartile portfolios picked on top/bottom quartiles, while the second section shows the decile portfolios picked on top/bottom deciles, for each independent provider. MktRf is the market factor, SMB is the small-minus-big size factor, HML is the high-minus-low value factor.

Fama French Three-Factor Model: Regression Output Quartile Portfolios

	Thomson Reuters ESG		Bloomberg ESG		Sustainalytics ESG	
	EW	VW	EW	VW	EW	VW
Intercept	−0.505***	−0.377**	−0.177	−0.067	−0.165	−0.056
	(−4.481)	(−2.344)	(−1.455)	(−0.389)	(−1.531)	(−0.365)
MktRf	0.061	−0.085	−0.050	−0.181***	0.100***	0.142***
	(1.558)	(−1.519)	(−1.327)	(−3.291)	(3.022)	(3.196)
SMB	−0.705***	−0.812***	−0.596***	−0.628***	−0.354***	−0.307***
	(−9.544)	(−7.071)	(−7.134)	(−5.610)	(−5.009)	(−3.138)
HML	0.537***	0.437***	0.476***	0.488***	0.159***	0.102
	(8.540)	(5.199)	(7.407)	(4.264)	(2.658)	(1.307)
N	72	72	72	72	72	72
R²	0.756	0.611	0.646	0.517	0.445	0.298
Adjusted R²	0.746	0.594	0.630	0.495	0.420	0.267

Fama French Three-Factor Model: Regression Output Decile (Small) Portfolios

	Thomson Reuters ESG		Bloomberg ESG		Sustainalytics ESG	
	EW	VW	EW	VW	EW	VW
Intercept	−0.554***	−0.147	−0.592***	−0.372	−0.117	−0.058
	(−3.322)	(−0.694)	(−3.177)	(−1.586)	(−0.584)	(−0.212)
MktRf	0.026	−0.157**	−0.026	−0.138**	0.113	0.175*
	(0.430)	(−2.260)	(−0.623)	(−2.025)	(1.514)	(1.698)
SMB	−0.811***	−1.034***	−0.644***	−0.786***	−0.333**	−0.356
	(−7.170)	(−7.527)	(−5.975)	(−5.851)	(−2.135)	(−1.596)
HML	0.586***	0.275***	0.519***	0.387***	0.195*	−0.074
	(6.931)	(2.706)	(5.218)	(2.987)	(1.810)	(−0.655)
N	72	72	72	72	72	72
R²	0.670	0.470	0.496	0.386	0.230	0.185
Adjusted R²	0.589	0.446	0.474	0.358	0.196	0.149

t statistics based on robust standard errors in parentheses
* p < 0.1, ** p < 0.05, *** p < 0.01

In terms of abnormal return, only Thomson Reuters shows a significant negative relationship. Excess return among the bottom ESG performers is on average around 0.5% higher for both the equally-weighted decile and quartile portfolios. As for the value-weighted portfolios, only the quartile portfolios show significance at the 5% level with the bottom quartile yielding 0.38% higher returns than the top quartile. Bloomberg also shows significantly higher returns for the bottom quartile of about 0.6%. None of the intercepts are significant for the Sustainalytics portfolios.

Results for the systematic risk factor differ among the independent providers' portfolios. The bottom ESG performers rated by Bloomberg shows higher volatility than the top performers when value-weighted, but no difference when equal-weighted. On the other hand, Sustainalytics shows the opposite effect where the top performers are associated with higher volatility compared to the bottom performers, both when equal- and value-weighted. The value-weighted decile portfolios of Thomson Reuters and Bloomberg show similar results, consistent with the high volatility of the quartile Bloomberg portfolios.

The SMB factor shows that all quartile portfolios consisting of bad ESG performers have higher exposure to smaller capitalization firms than the portfolios of top ESG performers. As for the decile portfolios, Thomson Reuters and Bloomberg show similar evidence, while for Sustainalytics only the equal-weighted portfolio yields a significant factor.

All equal-weighted quartile portfolios show a positive exposure to the HML factor, indicating that the top ESG portfolios are more exposed towards high value stocks than the bottom ESG portfolios. The relationship is weaker for Sustainalytics than for Thomson Reuters and Bloomberg, and the latter two also show similar results for the value-weighted portfolios. These results are also valid for the decile portfolios of Thomson Reuters and Bloomberg, both equal- and value-weighted. However, Sustainalytics only shows significance at 10% level for its equal-weighted portfolio.

6.2.2. Carhart Model

Table 5: Carhart Four-Factor Model Regression Outputs

This table presents the results from the Carhart Four-Factor Model. It presents the factor loadings and abnormal returns (in percentage) of both equal-weighted (EW) and value-weighted (VW) zero-investment portfolios with a long position in good ESG performers and short position in bad ESG performers. This first section shows the quartile portfolios picked on top/bottom quartiles, while the second section shows the decile portfolios picked on top/bottom deciles, for each independent provider. MktRf is the market factor, SMB is the small-minus-big size factor, HML is the high-minus-low value factor, UMD is the momentum factor.

Carhart Four-Factor Model: Regression Output Quartile Portfolios

	Thomson Reuters ESG		Bloomberg ESG		Sustainalytics ESG	
	EW	VW	EW	VW	EW	VW
Intercept	−0.515***	−0.424**	−0.233*	−0.141	−0.168	−0.042
	(−4.406)	(−2.481)	(−1.820)	(−0.779)	(−1.464)	(−0.262)
MktRf	0.065	−0.066	−0.027	−0.150**	0.101***	0.136***
	(1.523)	(−1.146)	(−0.652)	(−2.596)	(2.782)	(2.701)
SMB	−0.706***	−0.818***	−0.602***	−0.636***	−0.354***	−0.305***
	(−9.291)	(−7.063)	(−7.395)	(−5.646)	(−4.836)	(−3.005)
HML	0.545***	0.475***	0.522***	0.509***	0.162**	0.090
	(7.624)	(4.941)	(7.744)	(4.300)	(2.597)	(1.043)
UMD	0.017	0.075	0.091	0.120	0.005	−0.023
	(0.244)	(0.846)	(1.418)	(1.208)	(0.093)	(−0.305)
N	72	72	72	72	72	72
R^2	0.757	0.618	0.659	0.535	0.445	0.299
Adjusted R^2	0.742	0.595	0.639	0.507	0.411	0.257

Carhart Four-Factor Model: Regression Output Decile Portfolios

	Thomson Reuters ESG		Bloomberg ESG		Sustainalytics ESG	
	EW	VW	EW	VW	EW	VW
Intercept	−0.604***	−1.888	−0.586***	−0.385	−0.069	0.002
	(−3.144)	(−0.794)	(−2.867)	(−1.504)	(−0.361)	(0.006)
MktRf	0.047	−0.140*	−0.029	−0.132	0.092	0.150
	(0.763)	(−1.906)	(−0.584)	(−1.709)	(1.352)	(1.585)
SMB	−0.816***	−1.038***	−0.643**	−0.788***	−0.326**	−0.349
	(−7.294)	(−7.524)	(−5.844)	(−5.678)	(−2.015)	(−1.498)
HML	0.627***	0.309**	0.514***	0.398***	0.155	−0.123
	(6.334)	(2.386)	(4.575)	(2.735)	(1.256)	(−0.950)
UMD	0.081	0.066	−0.010	0.022	−0.078	−0.096
	(0.837)	(0.513)	(−0.180)	(0.165)	(−0.603)	(−0.531)
N	72	72	72	72	72	72
R^2	0.612	0.473	0.497	0.386	0.238	0.193
Adjusted R^2	0.589	0.441	0.466	0.349	0.193	0.145

t statistics based on robust standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

When applying the Carhart model, the excess return of the different portfolios follows similar patterns as for the Three-factor model. The coefficients are similar estimates. For the equal-weighted quartile portfolios, only Thomson Reuters shows significant results below the 5% level. However, for the equal-weighted decile portfolios, Bloomberg also shows a significant negative relationship. In summary, the excess return of the bad ESG performers is about 0.5% and on average 0.6% for the quartile and decile portfolios, respectively.

Considering the systematic risk of the portfolios, only Sustainalytics shows a significant coefficient. As for the Three-factor model the coefficient is positive, implying that the top ESG performers have higher volatility than the bad performers. However, this is only valid for the quartile portfolio, whereas none of the coefficients are significant for the decile portfolios.

The SMB factor has the same characteristics when applying the Three-factor model with the bad ESG firms of the quartile portfolios being smaller in market capitalization. When sorted into decile portfolios, Thomson Reuters and Bloomberg show similar results, while only the equal-weighted portfolio is significant at the 5% level for Sustainalytics.

Like the Three-factor model, all equal-weighted quartile portfolios show a positive exposure to the HML factor. Still, there is a weaker significance for Sustainalytics than the two other providers. Only, Thomson Reuters and Bloomberg yield significant results for the value-weighted quartile portfolios, as well for their decile portfolios.

Lastly, the Carhart model includes the momentum factor, which is not significant for any of the regressions.

6.2.3. Fama French Five Factor Model

Table 6: Fama French Five-Factor Model Regression Outputs

This table presents the results from the Fama French Five Four-Factor Model. It presents the factor loadings and abnormal returns (in percentage) of both equal-weighted (EW) and value-weighted (VW) zero-investment portfolios with a long position in good ESG performers and short position in bad ESG performers. This first section shows the quartile portfolios picked on top/bottom quartiles, while the second section shows the decile portfolios picked on top/bottom deciles, for each independent provider. MktRf is the market factor, SMB is the small-minus-big size factor, HML is the high-minus-low value factor, RMW is the robust-minus-weak profitability factor, CMA is the conservative-minus-aggressive investment factor.

Fama French Five-Factor Model: Regression Output Quartile Portfolios

	Thomson Reuters ESG		Bloomberg ESG		Sustainalytics ESG	
	EW	VW	EW	VW	EW	VW
Intercept	−0.596***	−0.443**	−0.189	−0.030	−0.170	0.027
	(−5.526)	(−2.630)	(−1.418)	(−0.166)	(−1.448)	(0.171)
MktRf	0.082**	−0.059	−0.038	−0.176***	0.103***	0.128***
	(2.051)	(−1.082)	(−0.979)	(−3.309)	(2.968)	(2.840)
SMB	−0.619***	−0.728***	−0.561***	−0.633***	−0.342***	−0.371***
	(−8.797)	(−6.811)	(−5.979)	(−5.769)	(−4.771)	(−3.662)
HML	0.741***	0.537***	0.478***	0.346**	0.177*	−0.047
	(6.597)	(3.870)	(3.992)	(2.092)	(1.775)	(−0.339)
RMW	0.470***	0.349*	0.097	−0.103	0.057	−0.317*
	(3.333)	(1.915)	(0.646)	(−0.445)	(0.432)	(−1.866)
CML	0.232*	0.327*	0.182	0.151	0.062	−0.083
	(1.873)	(1.709)	(0.956)	(0.554)	(0.422)	(−0.460)
N	72	72	72	72	72	72
R^2	0.804	0.650	0.653	0.532	0.445	0.337
Adjusted R^2	0.789	0.624	0.627	0.497	0.403	0.286

Fama French Five-Factor Model: Regression Output Decile Portfolios

	Thomson Reuters ESG		Bloomberg ESG		Sustainalytics ESG	
	EW	VW	EW	VW	EW	VW
Intercept	−0.651***	−0.242	−0.601***	−0.351	−0.077	0.099
	(−3.884)	(−1.070)	(−2.693)	(−1.323)	(−0.375)	(0.389)
MktRf	0.042	−0.142*	0.001	−0.116	0.072	0.124
	(0.677)	(−1.993)	(0.028)	(−1.550)	(1.101)	(1.468)
SMB	−0.730***	−0.961***	−0.571***	−0.743***	−0.433***	−0.523***
	(−6.336)	(−6.752)	(−3.809)	(−4.791)	(−3.349)	(−2.961)
HML	0.843***	0.549***	0.432**	0.249	0.355*	−0.229
	(5.782)	(2.942)	(2.420)	(1.088)	(1.836)	(−1.036)
RMW	0.519***	0.550**	0.047	−0.058	−0.045	−0.579
	(2.822)	(2.192)	(0.203)	(−0.188)	(−0.151)	(−1.479)
CML	0.170	0.190	0.347	0.326	−0.434	−0.480
	(0.961)	(0.904)	(1.526)	(1.145)	(−1.038)	(−0.864)
N	72	72	72	72	72	72
R^2	0.637	0.505	0.507	0.399	0.262	0.262
Adjusted R^2	0.609	0.468	0.469	0.353	0.206	0.206

t statistics based on robust standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The Fama French Five-factor model provides similar evidence as the previous models on the excess return of our long-short zero investment strategy. Only Thomson Reuters quartile portfolios show a significant difference on the returns of the top and bottom ESG performers. This is true for both the equal- and value-weighted portfolio. Excess return among the bottom ESG performers is on average 0.6% and 0.45% higher than for the top ESG performers, for the equal- and value-weighted quartiles, respectively. Bloomberg is significant for the equal-weighted decile portfolio, showing 0.6% higher excess return for the bottom ESG performers than for the top performers. This relationship is also valid for the equal-weighted Thomson Reuters decile portfolio with a 0.65% higher return of the bottom ESG performers compared to the top performers.

Sustainalytics shows significant coefficients for systematic risk of both its quartile portfolios, which is in line with the Carhart model. Similar coefficients are observed for the Thomson Reuters equal-weighted quartile portfolio and Bloomberg's value-weighted quartile portfolio. In contrast to Sustainalytics and Thomson Reuters, Bloomberg shows a negative coefficient for systematic risk, implying that the bottom ESG performers have higher volatility than the top performers.

The SMB factor has the same characteristics as we control for the different factors. All portfolios show a significant negative coefficient. The HML factor show similar characteristics as the previous models with Thomson Reuters and Bloomberg demonstrating significant coefficients of their portfolios.

The five-factor model introduces two new factors, RMW and CMA. Only the former is of interest, showing significant positive coefficients for the Thomson Reuters decile portfolios, as well as for the Thomson Reuters equal-weighted quartile portfolio. This implies that the top ESG performers constitute firms with more robust profitability than the bottom ESG performers. The CMA factor is not significant for any portfolios.

6.2.4. Fama French Five Factor with Momentum

Table 7: Fama French Five-Factor Model with Momentum Regression Outputs

This table presents the results from the Fama French Five-Factor Model with momentum. It presents the factor loadings and abnormal returns (in percentage) of both equal-weighted (EW) and value-weighted (VW) zero-investment portfolios with a long position in good ESG performers and short position in bad ESG performers. This first section shows the quartile portfolios picked on top/bottom quartiles, while the second section shows the decile portfolios picked on top/bottom deciles, for each independent provider. MktRf is the market factor, SMB is the small-minus-big size factor, HML is the high-minus-low value factor, RMW is the robust-minus-weak profitability factor, CMA is the conservative-minus-aggressive investment factor, UMD is the momentum factor.

Fama French Five-Factor Model with Momentum: Regression Output Quartile Portfolios

	Thomson Reuters ESG		Bloomberg ESG		Sustainalytics ESG	
	EW	VW	EW	VW	EW	VW
Intercept	−0.605***	−0.481**	−0.241	−0.099	−0.169	0.049
	(−4.916)	(−2.576)	(−1.644)	(−0.504)	(−1.393)	(0.295)
MktRf	0.084*	−0.047	−0.022	−0.155**	0.103***	0.121***
	(1.907)	(−0.810)	(−0.489)	(−2.670)	(2.706)	(2.606)
SMB	−0.622***	−0.741***	−0.579***	−0.656***	−0.342***	−0.363***
	(−8.641)	(−6.700)	(−6.150)	(−5.654)	(−4.655)	(−3.485)
HML	0.754***	0.596***	0.559***	0.451***	0.177	−0.081
	(5.373)	(4.005)	(4.355)	(2.776)	(1.364)	(-0.468)
RMW	0.474***	0.367**	0.122	−0.071	0.057	−0.328*
	(3.181)	(2.006)	(0.792)	(−0.312)	(0.410)	(−1.906)
CMA	0.221	0.279	0.115	0.064	0.062	−0.055
	(1.593)	(1.503)	(0.697)	(0.274)	(0.383)	(−0.274)
UMD	0.014	0.059	0.082	0.107	−0.001	−0.035
	(0.210)	(0.673)	(1.319)	(1.057)	(−0.007)	(−0.422)
N	72	72	72	72	72	72
R²	0.804	0.654	0.663	0.545	0.445	0.339
Adjusted R²	0.786	0.622	0.632	0.503	0.394	0.278

Fama French Five-Factor Model with Momentum: Regression Output Decile Portfolios

	Thomson Reuters ESG		Bloomberg ESG		Sustainalytics ESG	
	EW	VW	EW	VW	EW	VW
Intercept	−0.712***	−0.292	−0.566**	−0.336	−0.054	0.152
	(−3.487)	(−1.121)	(−2.270)	(−1.135)	(−0.251)	(0.586)
MktRf	0.060	−0.127*	−0.009	−0.121	0.065	0.108
	(0.934)	(−1.682)	(−0.171)	(−1.498)	(0.942)	(1.144)
SMB	−0.750***	−0.978***	−0.559***	−0.738***	−0.425***	−0.505***
	(−6.218)	(−6.664)	(−3.624)	(−4.508)	(−3.082)	(−2.759)
HML	0.938***	0.625***	0.378	0.227	0.298	−0.311
	(4.586)	(2.724)	(1.650)	(0.847)	(1.483)	(−1.349)
RMW	0.548***	0.574**	0.031	−0.065	−0.057	−0.604
	(2.693)	(2.188)	(0.127)	(−0.202)	(−0.179)	(−1.485)
CML	0.092	0.127	0.392	0.344	−0.404	−0.413
	(0.439)	(0.564)	(1.630)	(1.114)	(−1.021)	(−0.806)
UMD	0.096	0.078	−0.055	−0.023	−0.037	−0.083
	(0.930)	(0.574)	(−0.495)	(−0.152)	(−0.301)	(−0.529)
N	72	72	72	72	72	72
R²	0.644	0.509	0.509	0.399	0.264	0.267
Adjusted R²	0.611	0.464	0.464	0.344	0.196	0.200

t statistics based on robust standard errors in parentheses
* p < 0.1, ** p < 0.05, *** p < 0.01

The final model adds the momentum factor to the specifications of the Fama French five-factor model. As for excess return, we register similar results for the two models. Only Thomson Reuters' portfolios are significant. This is true for both quartile portfolios, as well as the equal-weighted decile portfolio. The estimates are consistent with the five-factor model without momentum. Once again, Bloomberg is statistically significant at the 5 percent level for the equal-weighted decile portfolio.

Regarding systematic risk, Bloomberg and Sustainalytics show significant coefficients of their quartile portfolios. However, the results are contradicting; while the coefficient is negative for the value-weighted Bloomberg portfolio, it is positive for the Sustainalytics portfolio with same specifications.

The SMB and HML factors show consistent results with the previous models and therefore we do not comment these factors further.

In line with the five-factor model without momentum, the RMW coefficient on Thomson Reuters is positive for all four portfolios. Neither the CMA factor nor the momentum factor yield significant coefficients.

6.3. F-Test Results

In this section, we conduct an F-test to further investigate the variance between the constructed ESG portfolios and the benchmark. The purpose of the test is to examine whether the variance of the constructed portfolios is significantly lower than the STOXX 600 Europe Index, which we use as a benchmark. We propose the following hypothesis:

$$\boldsymbol{H_0}: \sigma_A^2 \geq \sigma_B^2$$

$$\boldsymbol{H_A}: \sigma_A^2 < \sigma_B^2$$

Where A denotes the variance of the Thomson Reuters portfolios, and B denotes the variance of the benchmark.

Note that since only the results from the Thomson Reuters portfolios consistently showed a negative significant return in the previous section, we only conduct the test on these portfolios.

When comparing the Thomson Reuters value-weighted quartile and decile portfolios, the F-test output shows a p-value of approx. zero in both cases[7]. The equally weighted portfolios show similar results with low p-values for both the quartile and decile portfolio[8].

The test output indicates that the variance of the constructed Thomson Reuters ESG portfolios are lower for all cases on a 5% significance level. Hence, we reject the null hypothesis and conclude that the variance of the Thomson Reuters portfolios is lower than the benchmark. This is coherent with the risk-return trade-off theory and the findings suggests that since the portfolios have significantly lower return, they have significantly lower risk (Sharpe, 1964; Lintner, 1965; Mossin, 1966)

---

[7] P-values of Thom. Reuters <u>value-weighted</u> quartile and decile portfolios: 1.32×10^{-6} and 2.51×10^{-3}, respectively.
[8] P-values of Thom. Reuters <u>equal-weighted</u> quartile and decile portfolios: 2.75×10-7 and 3.55×10-4, respectively.

6.4. Further Discussion

In this section, we will discuss the implications of our findings.

6.4.1. Excess Returns

Existing literature have established that firms with higher ESG ratings yield higher return (Kempf and Osthoff, 2007). However, our findings suggest the opposite, in particular portfolios based on ratings from Thomson Reuters. Most Thomson Reuters portfolios provide similar evidence that a portfolio comprising of bottom ESG performers outperform a portfolio comprising of top ESG performers. For the decile portfolios, the monthly difference is around 0.6% to 0.7% basis points and for the quartile portfolios the difference is around 0.5% to 0.6% basis points. Our findings from the Bloomberg portfolios are only significant for decile portfolios, both equal- and value-weighted. Sustainalytics is the only provider that does not show a significant negative outperformance. This is consistent with the results from our initial analysis; the flat cumulative trend of the long-short portfolio from our descriptive analysis substantiates this observation. We would expect Bloomberg to be more coherent with Thomson Reuters, however, with limited insight into how the methodologies differ we do not know exactly why this is the case.

As stated in our hypothesis, if we consider ESG as a risk factor, firms with low ESG scores should compensate their owners with a higher return. Thomson Reuters is the only provider that has yielded evidence consistent with this hypothesis. The Thomson Reuters portfolios generate a slightly negative abnormal return, which suggest that investors pay a premium in the stock market for firms with higher ESG ratings. However, this may change in the future, as government intervention may lead to changes in regulations or the providers' approach.

6.4.2. Exposure to Size

The long-short portfolios have a consistent negative exposure to the SMB factor, indicating that large firms have better ESG ratings. A possible explanation for this result is that larger companies may have more resources available than smaller firms, which enables them to invest more time and capital into better ESG practices and certification. The findings are somewhat intuitive as ESG often is non-quantitative data and can be difficult to measure properly. Furthermore, according to Vermeir et al. (2005) larger companies often have greater social and

corporate responsibilities and therefore experience higher demands from the media, governments, and society in general to adapt more sustainable practices.

However, from another point of view, in terms of an investor seeking to invest in the best ESG companies, this implies that the portfolio will be skewed towards companies with larger market capitalization and not necessarily the companies with the best ESG practices. A concern that arises from these findings is that socially responsible investors do not receive the correct information, as ESG ratings are potentially biased towards large capitalization companies. Drempetic et al. (2020) document a significant positive relationship between firm size and available resources for providing ESG data with the ESG score. These findings once again sheds light on the important discussion on what a ESG rating measures, the current methodologies and whether the goal of sustainable finance is improved by these measures.

6.4.3. Differences Among ESG Rating Providers

Our results show differences in abnormal returns associated with the different providers' portfolios. We further explore and discuss these differences by examining the composition of the portfolios. This includes the overlap between the portfolios, the portfolios' sector breakdown, and the portfolios' ESG attributes. We seek to provide insights as to why Thompson Reuters is the only provider who manages to capture the ESG risk premium and is consistent with our hypothesis.

6.4.3.1. Portfolio Overlap

First, we examine portfolio overlap with results shown in Table 8. The first part shows the large portfolios, while the second part covers the small portfolios. Overlap percentages are calculated annually when each portfolio is rebalanced.

<div align="center">

Table 8: Portfolio Overlap

</div>

This table shows the average overlap percentages between the constructed portfolios for the total sample period. The left side of the table shows the top quartile consisting of top ESG-rated firms, while the right side shows the bottom quartile consisting of low ESG-rated firms.

Large Portfolios – Overlap

	Top Quartile (Long)			Bottom Quartile (Short)	
	Bloomberg	Sustainaly.		Bloomberg	Sustainaly.
Thom. Reut.	42%	43%	Thom. Reut.	59%	35%
Sustainaly.	29%		Sustainaly.	31%	

Small Portfolios – Overlap

	Top Decile (Long)			Bottom Decile (Short)	
	Bloomberg	Sustainaly.		Bloomberg	Sustainaly.
Thom. Reut.	23%	18%	Thom. Reut.	54%	26%
Sustainaly.	14%		Sustainaly.	24%	

We observe greater overlap for the quartile portfolios than for the decile portfolios, which suggests there is a certain rating consensus among the rating providers. The overlap percentages decrease as we sort for the decile portfolios. This indicates that the agreement among ESG ratings diverges as we examine the extreme observations of our sample.

Furthermore, the overlap percentages are greater for the long portfolios with top-rated firms than for the short portfolios consisting of low-rated firms, with the exemption of the quartile portfolios of Thomson Reuters and Bloomberg. These two providers have a slightly greater overlap for the long portfolios than for the short portfolios. Interestingly, the overlap is greatest between Thomson Reuters and Bloomberg. This is consistent with the cumulative return figures presented in section 6.1.2. where the two providers follow similar return patterns, particularly for the decile portfolios.

Moreover, the overlap between Sustainalytics and the other providers is generally low. This might be explained by a significant difference in methodology. While Sustainalytics rewards

firms for their sustainability efforts, Thomson Reuters and Bloomberg are constrained by a stricter rating methodology. We discuss this further in section 6.4.3.4.

6.4.3.2. Portfolio Sector Breakdown

Secondly, we examine the portfolio's sector composition, which is presented in Table 9 and Table 10.

Table 9: Portfolio Sector Breakdown

This table presents the sectors represented in the quartile portfolios (in percentage). The sector composition is calculated as an average over the sample period, based on the composition after the annual rebalance of portfolios. The table includes the long, the short, and the long-short equal-weighted portfolios. The shaded areas indicate the three sectors with the largest weights in the portfolios.

Sector Breakdown – Quartile Portfolios

Sector	Long-Short			Long			Short		
	TR	BB	SU	TR	BB	SU	TR	BB	SU
Communications	7%	6%	8%	5%	5%	6%	10%	6%	10%
Consumer Discretionary	12%	8%	12%	13%	2%	16%	12%	14%	9%
Consumer Staples	9%	10%	9%	10%	10%	7%	9%	9%	12%
Energy	3%	3%	4%	5%	4%	7%	2%	2%	1%
Financials	21%	30%	18%	17%	42%	14%	25%	18%	22%
Health Care	8%	15%	8%	8%	19%	9%	8%	12%	7%
Industrials	18%	18%	19%	13%	8%	17%	23%	29%	21%
Materials	11%	10%	9%	16%	10%	10%	7%	9%	9%
Real Estate	2%	0%	3%	2%	0%	3%	2%	0%	3%
Technology	3%	0%	5%	5%	0%	6%	2%	0%	3%
Utilities	3%	0%	5%	6%	0%	7%	0%	0%	4%

For the quartile portfolios we observe several patterns. First, the portfolios consisting of low-rated firms have the greatest exposure towards industrials. These firms are generally given a low environmental score, as evidenced by a low average environmental pillar score (see Table 11). This is intuitive, as these firms have difficulties of fulfilling certain environmental requirements and are consequently penalized by Thomson Reuters and Bloomberg. On the other

hand, Sustainalytics reward companies for their corporate sustainability efforts, evidenced by a greater average score of the industrial sector (see Table 11).

All portfolios have a significant exposure to financial firms, both the for the long position and the short position. Interestingly, financial firms with similar business models are rated differently based on their ESG efforts. This indicates that the rating providers pick up a difference among the firms, and reward firms with good ESG practices while punishing firms with poor ESG practices. Similar evidence is shown for industrial firms, which are also represented in the long portfolios of Thomson Reuters and Sustainalytics. This demonstrates that firms which typically have a low environmental rating may obtain a high overall ESG rating. As expected, the percentage of industrial firms included in the top-rated portfolios decreases reduced when we sort into decile portfolios.

Table 10: Portfolio Sector Breakdown

This table presents the sectors represented in the decile portfolios (in percentage). The sector composition is calculated as an average over the sample period, based on the composition after the annual rebalance of portfolios. The table includes the long, the short, and the long-short equal-weighted portfolios. The shaded areas indicate the three sectors with the largest weights in the portfolios.

Sector Breakdown – Small Portfolios

Sector	Long-Short			Long			Short		
	TR	BB	SU	TR	BB	SU	TR	BB	SU
Communications	6%	4%	7%	1%	0%	4%	12%	8%	9%
Consumer Discretionary	15%	7%	13%	16%	0%	18%	14%	13%	9%
Consumer Staples	10%	6%	8%	11%	6%	8%	8%	6%	8%
Energy	3%	1%	3%	5%	1%	6%	1%	1%	1%
Financials	19%	40%	18%	16%	57%	10%	22%	23%	27%
Health Care	9%	23%	10%	12%	30%	10%	7%	15%	10%
Industrials	18%	16%	18%	8%	4%	12%	27%	27%	18%
Materials	11%	4%	11%	16%	3%	11%	6%	6%	10%
Real Estate	2%	0%	4%	2%	0%	2%	1%	0%	5%
Technology	5%	0%	5%	8%	0%	7%	2%	0%	2%
Utilities	2%	0%	3%	4%	0%	6%	0%	0%	0%

In summary, we observe similar patterns in the sectors represented in the portfolios. This reveals a certain consensus among the rating providers. In the next subsection we further examine the ESG attributes of firms included in the sample.

In addition to the possible rating limitations described in section 4.3., a drawback of our analysis is that Bloomberg do not rate firms in the following sectors: Real Estate, Technology and Utilities. However, these three sectors are the least represented sectors in the overall data sample and the other providers' exposure to these sectors is also low. This makes the overall implications less significant.

6.4.3.3. ESG Attributes

Table 11 presents the average ESG ratings of firms within the same sector included in the dataset.

Table 11: Average ESG Ratings of Sectors

This table shows the average ESG ratings of firms within the same sector included in the dataset. The shaded areas represent the three sectors with the best ratings within each ESG pillar for each rating provider.

Sector	ESG			ENVIROMENTAL			SOCIAL			GOVERNANCE		
	TR	BB	SU	TR	BB	SU	TR	BB	SU	TR	BB	SU
Communications	58.1	59.7	70.6	52.5	42.5	68.8	61.2	66.0	70.8	54.6	70.5	65.4
Consumer Discretionary	63.3	55.9	78.2	65.0	42.5	75.7	66.7	54.8	77.3	56.4	70.3	72.9
Consumer Staples	64.9	61.2	72.7	65.1	50.7	71.2	69.1	61.4	71.5	57.4	71.3	62.3
Energy	69.2	62.5	87.1	67.9	55.7	82.6	72.2	66.4	84.3	67.0	65.5	84.0
Financials	61.0	64.6	72.2	56.5	58.0	71.0	61.9	63.4	69.0	62.4	72.2	70.6
Health Care	63.6	63.4	75.8	55.0	54.3	73.8	71.0	67.1	75.3	58.7	68.9	69.1
Industrials	60.2	55.3	73.9	58.0	35.0	71.8	65.4	60.5	71.7	55.0	70.4	67.3
Materials	68.5	61.8	76.3	69.5	48.1	69.4	70.4	65.4	73.7	64.8	71.8	71.9
Real Estate	66.3	-	73.4	74.3	-	71.4	69.6	-	75.3	53.1	-	69.1
Technology	68.6	-	82.0	63.3	-	81.5	77.6	-	77.5	62.7	-	80.7
Utilities	73.5	-	82.3	78.6	-	80.2	77.6	-	80.6	60.3	-	73.7

The average total ESG rating of the firms rated by Sustainalytics is 76.8, which is higher than the average rating for Thomson Reuters (65.2) and Bloomberg (60.5). Sustainalytics is also the

provider that rates the fewest firms included on the STOXX Europe 600 index. The high ratings combined with the fewest rated firms may indicate that they exclude firms that do not meet certain ESG standards or firms that do not disclose certain information. Hence, it may appear that they exclude firms that would receive low ratings.

Sustainalytics also differ from the other providers in its rating methodology. While Thomson Reuters and Bloomberg weight the three pillars equally, Sustainalytics performs an overall assessment of firms' ESG risks and how the firms manage these risks. We discuss this in detail in the next subsection.

6.4.3.4. Discussion on Rating Methodlogies

Given the differences in ratings shown in our analysis, we will briefly discuss why this is the case by examining the different methodologies implemented by the rating providers. As previously mentioned, limitations of this thesis include lack of detailed data on each rating provider's methodology. Nevertheless, Thomson Reuters, Bloomberg, and Sustainalytics provides insight on calculation of the ratings.

According to Thomson Reuters (Refinitiv, 2020), the rating is derived from three equally weighted pillars, which all have different sub-criteria. Further, all scores are normalized by z-scoring and benchmarking the rating to all companies in the database. Similarly, the ratings composed by Bloomberg is computed through the same equally weighted pillars, and similar sub-criteria as Thomson Reuters. In summary, Thomson Reuters and Bloomberg define corporate sustainability using comparable rating methodologies. This confirms the descriptive statistics and analysis, which indicate that the ratings by Thomson Reuters and Bloomberg are more correlated.

In contrast with Thomson Reuters and Bloomberg, Sustainalytics' rating is based on firms' exposure or vulnerability to the risks associated with a particular ESG issue (Sustainalytics, 2020). These risks are further divided into risks that are manageable and those that are not. For instance, an airline company cannot fully manage their emissions through their daily business operations. An admittedly extreme and concrete example from our analysis is the car manufacturer Porsche which is subject to the largest rating dispersion among the providers. Porsche's ratings are displayed in Table 12.

Table 12: Ratings of Porsche

This table shows the rating providers' ratings of Porsche. It includes overall ESG ratings, environmental ratings, social ratings, and governance ratings from the respective providers, Thomson Reuters (TR), Bloomberg (BB) and Sustainalytics (SU).

Company	ESG			ENVIROMENTAL			SOCIAL			GOVERNANCE		
	TR	BB	SU	TR	BB	SU	TR	BB	SU	TR	BB	SU
Porsche	8.5	27.6	88.1	0.8	0	88.1	5.2	18.5	94.9	24.4	64.5	8.5

The overall ESG rating of Porsche ranges from 88.1 (Sustainalytics) to 8.5 (Thomson Reuters). Porsche is included in Thomson Reuters' and Bloomberg's bottom quartile and decile portfolios. Sustainalytics has a substantially higher rating of 88.1, close to being included in the top decile and quartile. The large difference between the ratings of Porsche is difficult to explain but based on available information one reason is the rating approach, previously described.

Sustainalytics identifies the management gap, which is the amount of manageable ESG risks that is not being managed. Sustainalytics argues that this creates a more realistic picture of how a company is exposed to ESG risks and how they are managed (Sustainalytics, 2020). This provides insight and may explain why Sustainalytics' ratings are less correlated than the other providers. Furthermore, we argue that the approach where they distinguish between manageable and unmanageable risks generates relatively higher scores. While we do not deem such approach to be wrong, it is important for investors to be aware of this, because it can potentially have a large impact when investing based on ESG criteria. If an investor seeks to invest in low emission companies and uses Sustainalytics' environmental ratings, this may produce conflicting information, as oil and transportation firms will not be penalized by Sustainalytics. Nevertheless, we argue that the increased focus on sustainability, there is doubt that Sustainalytics can continue to issue relatively high ratings to firms that harm the environment through their operations.

Thus, we emphasize the importance for investors to understand how the different ESG ratings are calculated and which metrics they are based, as this is crucial to fully understand which provider that matches their own assessments of ESG risks.

7.   Conclusion

In this thesis, we investigate whether there is a premium on firms with good ESG performance by examining historical stock returns. We find that a long-short ESG investment strategy achieves a negative monthly abnormal return in the case of ratings provided by Thomson Reuters. The returns range from 0.50% to 0.60%. Our findings point out that there may exist a negative relationship between responsible firms and returns. Further, we find significant negative abnormal returns using ratings from Bloomberg and Sustainalytics in some cases. However, these are only significant for the decile portfolios. Hence, Thomson Reuters is the only provider that is consistent with the hypothesis that ESG may be a risk factor that captures the negative abnormal return of the ESG portfolio. We argue that this is a result of the current discrepancies in rating methodologies. Ratings are clearly subjective, illustrated by low rating correlation, and the choice of rating provider impacts ESG investment decisions and portfolio performance. This issue is amplified as providers quantify qualitative information. Considering our findings, we therefore recognize the necessity of implementing a common framework for rating corporate sustainability.

The results of our F-tests showed clear evidence that the constructed Thomson Reuters portfolios are significantly less risky compared to the STOXX Europe 600. These results imply that ESG is priced into the market, as good ESG performers yield lower returns because they are less volatile. In other words, there exists a premium in return, urging investors to hold poor ESG-rated firms. This premium compensates investors for moral condemnation and various ESG related risks, such as environmental disasters, corporate scandals due to governance failures or government regulations.

In summary, we find that there is an ESG premium when we use data from the European stock market when implementing a long-short strategy using Thomson Reuters ESG ratings. Regarding the vast differences in ratings and results when comparing the different providers, we do not have any clear evidence on why this is the case. This remains a question for future research.

References

Allen, K. (2018). Lies, damned lies and ESG rating methodologies. *Financial Times*. Retrieved October 11, 2020, from https://ftalphaville.ft.com/2018/12/06/1544076001000/Lies--damned-lies-and-ESGrating-methodologies/.

Amel-Zadeh, A., & Serafeim, G. (2018). Why and how investors use ESG information: Evidence from a global survey. *Financial Analysts Journal 74*(3): 87-103.

Barinov, A. (2013). Analyst disagreement and aggregate volatility risk. *Journal of Financial and Quantitative Analysis*, 1877-1900.

Bartholdy, J., & Peare, P. (2005): Estimation of Expected Return: Capm Vs. Fama and French. *International Review of Financial Analysis, 14(4):*407–427.

Berg, F., Kölbel, J., & Rigobon, R. (2020) Aggregate Confusion: The Divergence of ESG Ratings. Retrieved September 15, 2020, from: https://ssrn.com/abstract=3438533.

Bloomberg. (2019). Sustainable Finance Bloomberg Professional Services. Retrived October 14, 2020, from: https://www.bloomberg.com/professional/solution/sustainable-finance/.

Bloomberg. (2020). Sustainable Finance Bloomberg Professional Services. Retrived October 14, 2020, from: https://www.bloomberg.com/professional/solution/sustainable-finance/.

Carhart, M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52 (1), 57–82.

Chatterji, A., Durand, R., Levine, D. I., & Touboul, S. (2016). Do ratings of firms converge? Implications for managers, investors, and strategy researchers. *Strategic Management Journal* 37(8), 1597–1614.

Christensen, D., Serafeim, G., & Sikochi, A. (2020). Why is Corporate Virtue in the Eye of The Beholder? The Case of ESG Ratings. *Harvard Business School Working Paper.* No. 20-084.

Douglas, E., Holt, T.V., & Whelan, T. (2017) Responsible Investing: Guide to ESG. Data Providers and Relevant Trends. *Journal of Environmental Investing 8, no 1.*

Drempetic, S., Klein, C., & Zwergel, B. (2020). The Influence of Firm Size on the ESG Score: Corporate Sustainability Ratings Under Review. *Journal of Business Ethics.* 167. 10.1007/s10551-019-04164-1.

Eccles, R. G., Ioannou, I., & Serafeim, G. (2014). The impact of corporate sustainability on organizational processes and performance. *Management Science*, *60*(11), 2835-2857.

European Central Bank. (2020). Key ECB Interest Rates. Retrived November 15, 2020, from: https://www.ecb.europa.eu/stats/policy_and_exchange_rates/key_ecb_interest_rates/html/index.en.html.

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

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

French, K. (2020). Kenneth R. French – Data Library. Retrieved October 25, 2020, from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Gibson, R., Krueger, P., Riand, N., & Scmidt, P. (2020). ESG rating disagreement and stock returns. *Swiss Finance Institute Research Paper.* No. 19-67.

Global Sustainable Investment (2018). *Global sustainable investment review*. Retrieved October 29, 2020, from: http://www.gsi-alliance.org/

Griffin, J. J., & Mahon, J. F. (1997). The corporate social performance and corporate financial performance debate: Twenty-five years of incomparable research. *Business & society*, *36*(1), 5-31.

Huber, B. M., Comstock, M., Polk, D., & LLP, W. (2017). ESG reports and ratings: What they are, why they matter. In *Harvard Law School Forum on Corporate Governance and Financial Regulation* (Vol. 44).

Kempf, A., & Osthoff, P. (2007): The Effect of Socially Responsible Investing on Portfolio Performance, European Financial Management, 13 (5)

KPMG. (2020). Sustainable Investing: fast-forwarding its evolution. Retrieved September 30, 2020, from: https://home.kpmg/xx/en/home/insights/2020/02/sustainable-investing.html.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A. & Vishny, R. W. (1998). Law and finance. *Journal of Political Economy* 106(6), 1113–1155.

Lima, Pedro (2017). Fama-French Models Application to the analysis of FTSE4GOOD, MSCI ESG and STOXX ESG Indices: Do SRI Indices have biases? Tech. Retrieved, October 28, 2020, from https//repositorio- aberto.up.pt/bitstream/10216/108203/2/224407.pdf

Litner, J. (1965). The Valuation of Risk Assets and the Selection of Risk Investments in Stock Portfolios and Capital Budgets, *Review of Economic Statictics 47*.

Lins, K. V., Servaes, H. & Tamayo, A. (2017). Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis. *The Journal of Finance,* 72(4), 1785–1824.

Mackintosh, J. (2018). Is Tesla or Exxon more sustainable? It depends whom you ask. *Wall Street Journal.*

Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica: Journal of the econometric society*, 768-783.

MSCI (Morgan Stanley Capital International). (2020). "Index Solutions." Retrieved November 11, 2020, from https://www.msci.com/index-solutions.

Miller, E. M. (1977). Risk, uncertainty, and divergence of opinion. *The Journal of Finance* 32(4), 1151–1168

Peng, C.W. & Yang, M.L. (2014). The effect of corporate social performance on financial performance: the moderating effect of ownership concentration. *Journal of Business Ethics, Vol. 123 No. 1*, pp. 171-182.

Peirce, M. H. (2019). Scarlet Letters: Remarks before the American Enterprise Institute. Retrieved September 30, 2020, from: https://www.sec.gov/news/speech/speech-peirce-061819?fbclid=IwAR1M9Kqv3YRQxFEN4syU8yKyc6O_xoZKhwxhbMv9Os8l37znCLzUvYt23o.

Perold, A. (2004). The Capital Asset Pricing Model. *The Journal of Economic Perspectives*, 18(3), 3-24. Retrieved October 18, 2020, from http://www.jstor.org/stable/3216804.

Reuters. (2020). Thomson Reuters unit to be renamed Refinitiv after Blackstone deal. Retrieved September 12, 2020, from: https://www.reuters.com/article/us-thomsonreuters-m-a-blackstone-group-idUSKBN1KH15H.

Refinitiv. (2020). An Overview of Environmental, Social and Corporate Governance – ESG. Retrieved October 8, 2020, from: https://www.refinitiv.com/en/financial-data/company-data/esg-data.

Sadka, R., & Scherbina, A. (2007). Analyst Disagreement, Mispricing, and Liquidity*. *Journal of Finance, 62*, 2367–2403.

Sarzosa, M. (2012). Introduction to Robust and Clustered Standard Errors. Retrieved December 11, 2020, from: http://econweb.umd.edu/~sarzosa/teach/2/Disc2_Cluster_handout.pdf.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, *19*(3), 425-442.

Sustainalytics. (2020). Sustainalytics Company ESG Risk Ratings. Retrieved October 17, 2020, from: https://www.sustainalytics.com/esg-ratings/.

STOXX Europe 600. (2020). Retrieved December 4, 2020, from: https://www.stoxx.com/index-details?symbol=SXXP.

Vance, S. C. (1975) Are socially responsible corporations good investment risks? *Management Review, 1975, 64 (8)*, 18-24.

Vermeir, W., Corten, F., & Van de Velde, E.. (2005). Corporate social responsibility and financial performance. *The international journal of business in society 5.3*, 129–138.

Wang, P., & Sargins, M. (2020). Better Minus Worse – Evaluating ESG Effects on Risk and Return. Moringstar Quantitative Research. Retrieved September 23, 2020, from: https://www.morningstar.com//content/dam/marketing/shared/pdfs/Research/963558_better_worse_final.pdf?utm_source=eloqua&utm_medium=email&utm_campaign=&utm_content=20815.

Womack, K. L., & Zhang, Y. (2003). Understanding risk and return, the CAPM, and the Fama-French three-factor model. *Lean Thinking*.

Yahoo Finance. (2020). STOXX 600 Europe. Retrieved November 17, 2020, from: https://finance.yahoo.com/quote/%5Estoxx/.

Appendix

A1 Model Testing

Table 8: Model Testing: Breusch-Pagan Test for Heteroscedasticity

This table shows the results of the Breusch-Pagan test for heteroscedasticity. The first section shows F-statistics and associated P-values of the quartile portfolios, while the second table covers the decile portfolios. The null hypothesis of this test is homoscedasticity, i.e., the variance of the error term is constant. In the case of a high P-value we do not reject the H_0, and we do not have an issue. In the case of a low P-value, we account for the issue by running regressions with robust standard errors.

Breusch-Pagan Test: Quartile Portfolios

	Thomson Reuters ESG		Bloomberg ESG		Sustainalytics ESG	
	EW	VW	EW	VW	EW	VW
FF3F	8.829	5.900	6.996	5.822	1.413	2.089
	(0.032)	(0.117)	(0.072)	(0.121)	(0.703)	(0.554)
FF4F	9.101	5.578	5.968	3.818	2.330	2.241
	(0.058)	(0.218)	(0.202)	(0.431)	(0.675)	(0.692)
FF5F	10.862	7.157	6.803	12.807	1.847	2.815
	(0.054)	(0.209)	(0.236)	(0.025)	(0.870)	(0.729)
FF5FM	10.591	8.312	5.913	8.068	2.643	3.127
	(0.102)	(0.216)	(0.433)	(0.233)	(0.852)	(0.793)

F-statistics. P-value in parentheses.

Breusch-Pagan Test: Decile Portfolios

	Thomson Reuters ESG		Bloomberg ESG		Sustainalytics ESG	
	EW	VW	EW	VW	EW	VW
FF3F	5.733	0.295	2.105	1.477	0.323	4.878
	(0.125)	(0.961)	(0.551)	(0.688)	(0.956)	(0.181)
FF4F	10.468	1.290	3.170	1.968	2.879	7.150
	(0.033)	(0.863)	(0.530)	(0.742)	(0.578)	(0.128)
FF5F	5.918	3.103	14.353	8.143	3.705	8.671
	(0.314)	(0.684)	(0.013)	(0.149)	(0.593)	(0.123)
FF5FM	11.707	5.124	15.473	10.165	3.760	7.721
	(0.069)	(0.528)	(0.017)	(0.118)	(0.710)	(0.259)

F-statistics. P-value in parentheses.

Table 9: Model testing: Breusch-Godfrey Test for Autocorrelation

This table shows the results of the Breusch-Godfrey test for autocorrelation. The first section shows Chi squared-values and associated P-values of the quartile portfolios, while the second table covers the decile portfolios. The null hypothesis of the test is no autocorrelation. Hence, we do have an issue if we observe a large Chi-value and a corresponding low P-value. The test results only show an issue for the equally weighted quartile Bloomberg portfolio. We account for this issue by lagging the regression of this specific portfolio.

Breusch-Godfrey Test: Quartile Portfolios

	Thomson Reuters ESG		Bloomberg ESG		Sustainalytics ESG	
	EW	VW	EW	VW	EW	VW
FF3F	0.900	1.451	4.551	0.078	3.019	1.277
	(0.343)	(0.228)	(0.033)	(0.780)	(0.082)	(0.259)
FF4F	1.035	1.785	4.870	0.087	2.943	1.415
	(0.309)	(0.182)	(0.027)	(0.768)	(0.086)	(0.234)
FF5F	0.031	0.314	4.574	0.160	2.891	2.268
	(0.860)	(0.576)	(0.032)	(0.690)	(0.089)	(0.132)
FF5FM	0.010	0.511	4.568	0.170	2.898	2.427
	(0.918)	(0.475)	(0.032)	(0.680)	(0.089)	(0.120)

Chi squared value. P-value in parentheses.

Breusch-Godfrey Test: Decile Portfolios

	Thomson Reuters ESG		Bloomberg ESG		Sustainalytics ESG	
	EW	VW	EW	VW	EW	VW
FF3F	1.570	0.713	2.331	0.213	1.179	0.253
	(0.211)	(0.399)	(0.127)	(0.644)	(0.278)	(0.615)
FF4F	2.333	0.953	2.318	0.193	1.290	0.334
	(0.127)	(0.329)	(0.128)	(0.660)	(0.256)	(0.563)
FF5F	0.014	0.078	3.150	0.076	1.848	1.274
	(0.905)	(0.780)	(0.076)	(0.782)	(0.174)	(0.259)
FF5FM	0.176	0.202	3.084	0.088	1.842	1.305
	(0.675)	(0.653)	(0.079)	(0.767)	(0.175)	(0.253)

Chi squared value. P-value in parentheses.

Table 10: Model testing: Shapiro-Wilk Normality Test

This table shows the results of the Shapiro-Wilk test for normality. The null hypothesis is that the sample comes from a normally distributed population. The normality test is only performed for Thomson Reuters portfolios in relation to the F-tests.

Shapiro-Wilk Normality Test: Quartile and Decile Portfolios

| | Thomson Reuters ESG | | | |
| | EW | | VW | |
	W-Stat	P-Value	W-Stat	P-Value
Quartile	0.991	0.902	0.977	0.207
Decile	0.984	0.470	0.989	0.799

A2 Cumulative Returns of the Value-Weighted Portfolios

Figure 3: Monthly Cumulative Returns of the Decile Long-Short ESG Portfolio

This figure presents the cumulative performance (in EUR) of a EUR 1 investment in the decile long-short portfolios from Jan. 2014 to Dec. 2019.
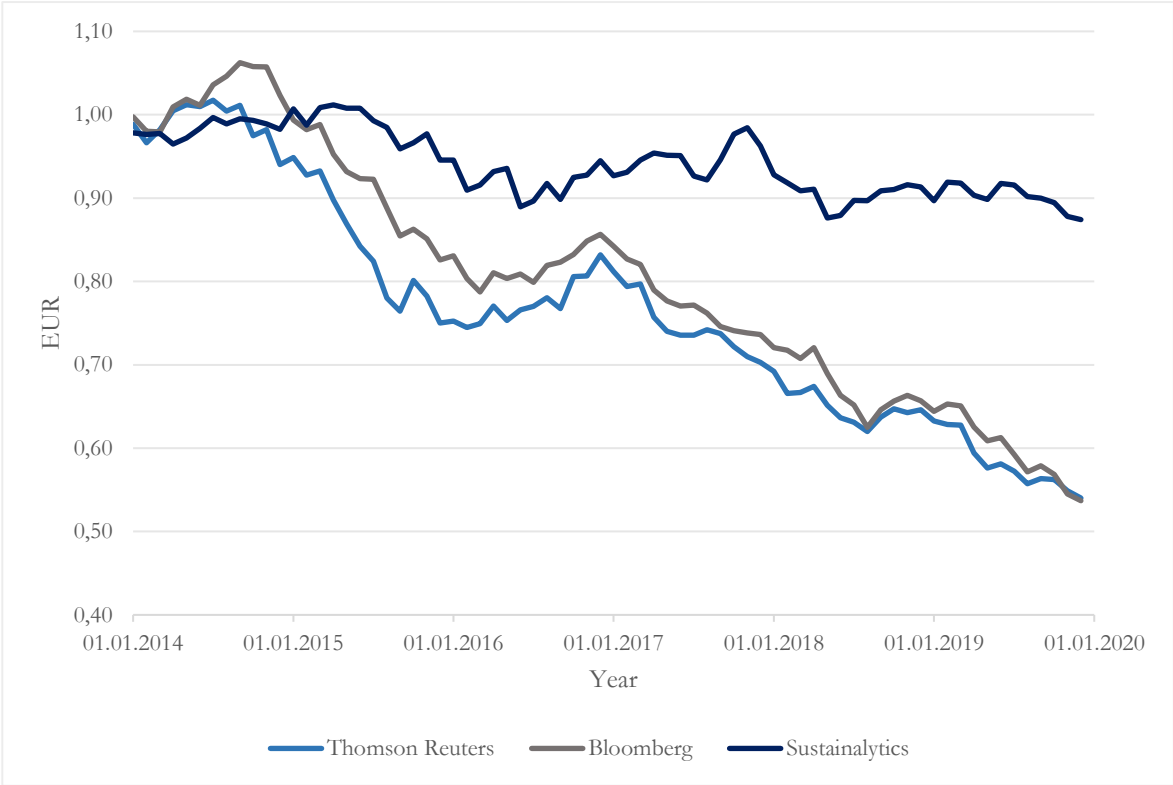
Figure 4: Monthly Cumulative Returns of the Quartile Long-Short ESG Portfolio

This figure presents the cumulative performance (in EUR) of a EUR 1 investment in the quartile long-short portfolios from Jan. 2014 to Dec. 2019.