Handelshøyskolen BI

GRA 19703 Master Thesis

Thesis Master of Science 100% - W

Predefinert informasjon

Startdato:	09-01-2023 09:00 CET	**Termin:**	202310
Sluttdato:	03-07-2023 12:00 CEST	**Vurderingsform:**	Norsk 6-trinns skala (A-F)
Eksamensform:	T		
Flowkode:	202310||11184||IN00||W||T		
Intern sensor:	(Anonymisert)		

Deltaker

Navn:	Tobias Wike Gurrik og Cornelia Fjeld Haagensen

Informasjon fra deltaker

Tittel *:	ESG Ratings and Stock Returns in the US Equity Market
Navn på veileder *:	Paul Ehling

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

Gruppe

Gruppenavn:	(Anonymisert)
Gruppenummer:	197
Andre medlemmer i gruppen:	

ESG Ratings and Stock Returns in the US Equity Market

An Empirical Study of the Relationship Between Stock Returns on the U.S. Stock Market and ESG Scores

Master Thesis

By

Tobias Gurrik & Cornelia Haagensen

MSc in Business with Major Finance/Sustainable Finance

Supervisor:

Paul Ehling

Oslo, June 26th, 2023

Abstract

This thesis investigates the relationship between ESG ratings and stock returns in the US equity market. Using the companies listed on the S&P 500 as a benchmark for the American stock market we perform a long-short zero investment strategy based on the companies ESG scores, with a short position in the lowest, and a long position in the highest ESG rated companies. We use three different ESG ratings to construct our portfolios, including Bloomberg ESG Disclosure score, Refinitiv ESG score, and Refinitiv ESG Controversies score. In order to provide robust results and account for possible differences in risk exposures of the portfolios we use multivariate regression analysis, applying the Fama-French (Fama & French 1993 & 2015) and Carhart (1997) framework. Our study returns negative significant alphas for both the Bloomberg ESG Disclosure, and the Refinitiv ESG study, which indicates a negative relationship between high ESG performers and stock returns. However, when controlling for ESG controversies, the ESG portfolios do not state a significant return difference between the top and bottom ESG portfolios. Our findings indicate that investors should adjust their expectations and no longer anticipate abnormal returns by trading a difference portfolio of high and low ESG rated companies using ESG as the sole portfolio screen.

Acknowledgements

We would like to thank our supervisor, Professor Paul Ehling for providing us with guidance and feedback throughout the semester. Second, we would like to thank BI Norwegian Business School for facilitating access to the necessary databases. Lastly, we want to thank our friends and family for being encouraging and keeping us motivated.

Table of Contents

List of Tables

List of Figures

1. Introduction and Motivation

The thesis objective is to investigate whether ESG integration in S&P 500 companies can deliver superior financial returns. The topic of interest was discovered as the relationship between a company's financial performance and its commitment to environmental, social, and governance (ESG) principles has been a subject of increasing attention in recent years.

While some argue that ESG considerations are incompatible with maximizing shareholder value, others claim that companies with strong ESG practices tend to outperform their peers. The ESG terminology has moved to be a more distinct investment strategy and away from being a socially responsible philosophy. To achieve long-term financial performance and make a positive impact on the environment, demand across the financial ecosystem has changed, leading to the development of ESG investment. Assets managed with an eye on sustainability have grown rapidly over the recent years, and the trend seems poised to grow further. The Global Sustainable Investment Review reports over USD 35 trillion invested with explicit ESG goals as of 2020, compared to USD 13.6 trillion in 2012 (GSIA, 2020 & GSIA, 2012).

Through our paper we conduct an empirical study examining the effects of ESG ratings on the stock performance of publicly traded companies in the US stock market. We investigate whether there is a difference between the best and the worst ESG performers. By dividing our sample returns into ten percentiles based on lagged ESG scores and performing top-bottom regression analysis, with a short position in the bottom portfolio and a long position in the top portfolio we examine risk-adjusted stock returns and present the following research question:

Is there a positive statistically significant relationship between the corporate financial performance of the companies listed on the S&P 500 and their corresponding Environmental, Social, and Governance (ESG) scores?

After conducting different portfolios consisting of top to bottom ESG rated companies we examine whether the choice of ESG score providers affects the performance of the different portfolios. As the number of rating agencies, such as Sustainalytics, Bloomberg, S&P, and Eikon Refinitiv have increased as sustainable investing has gained attention in the financial universe, the agencies are all working to develop a compounded measure of ESG ratings. Despite their experience and financial understanding, it adds to the problem of information uncertainty, as there is no standardized framework. As Berg et al., (2022) states, ESG scores are a result of deviating opinions, different performance measurement and different methodologies, which all cause a discrepancy in the results and add to the lack of consensus. We confirm this finding in our dataset, where the correlation between ESG ratings range from 0.59 to 0.84. This is based on ESG ratings from two different providers: Bloomberg and Refinitiv, also including Refinitiv ESG Controversies score.

Our answer to the research question is based on the performance of our respective top and bottom portfolios, over the period from 2014 to 2021, consisting of the companies listed at the S&P 500 index. We measure their financial performance by regressing our top-bottom strategy on the Fama-French and Carhart risk factors and obtain the following results when using the Bloomberg ESG Disclosure score as a screen: Our long-short zero investment strategy produces negative significant alphas when using ESG as a sole screen for stock performance. The findings imply that the best ESG performers (top portfolio) exhibit poorer performance than the least successful ESG performers (bottom portfolio). We obtain the same results when regressing the Refinitiv ESG screened portfolios, overserving a negative relationship between ESG and financial performance. The findings from the Bloomberg and Refinitiv ESG studies contradicts with the findings of Kempf & Osthoff (2007), which indicates a positive relationship between high ESG performers and stock returns and are more aligned with research suggesting that one cannot expect positive abnormal returns when trading a difference portfolio of high and low rated ESG firms (Halbritter & Dorfleitner, 2015).

Our study reveals significant disparities when employing the Refinitiv ESG Controversies score as a portfolio screen. This study produces positive alphas when regressing the top-bottom portfolios on the risk factors, implying a positive relationship between ESG and financial performance. Lack of statistical significance makes it impossible to conclude on this relationship, leaving us with a neutral relationship between ESG and financial performance.
Our results highlight previous research suggesting that the magnitude of the impact of ESG investing is substantially dependent on the rating provider (Halbritter and Dorfleitner, 2015).

The thesis contributes to the existing empirical literature on the relationship between ESG and financial performance. By utilizing data from the US market, our research expands upon majority of literature, which focuses primarily on the US market. By focusing our study on companies listed at the S&P 500 index, we have restricted our sample to include larger cap companies, exemplifying a significant proportion of large corporations and the overall US economy.

The remainder of the thesis is structured as follows: The second section examines the empirical links between ESG and stock performance, as well as ESG rating weaknesses. In the third section we present our hypothesis and additional research question. Section 4 describes the empirical approach and methodology used in the analysis. Section 5 describes data sources, ESG providers, sample selection and descriptive statistics of our sample. Section 6 presents our analysis and additional discussion, prior to concluding the thesis with the final section, section 7, with a summary and conclusion of the research findings and results.

2. Literature Review

2.1 The importance of corporate ESG performance

Over the last decade the growth of sustainable investment in financial markets has drawn considerable academic attention. According to research, a firm's value and ESG performance are interconnected, and the degree to which the firm exposes its ESG-related data affects this relationship. Research suggests organizations can lessen the detrimental effect that concerns about their ESG performance have on their valuation by disclosing ESG-related information (Fatemi et al., 2018). Their research suggests that the association between ESG performance and firm value is moderated by ESG transparency. There has also been demonstrated how ESG disclosure level affects firm value, which suggests that improved transparency and accountability, as well as enhanced stakeholder trust play a significant role in boosting firm value (Li et al., 2018).

Studies indicate that there is still a lack of consensus regarding the economic implications of ESG factors. Just until recently the conventional wisdom has been that ESG investors would have to accept a lower expected return from holding green portfolios. This assumption is no longer universally accepted, and although the empirical evidence is mixed, it seems in most cases to support the contention that ESG investing does not have to underperform relative to traditional portfolios (Hill, J., 2020). Although recent studies produce more conclusive results, it is worth acknowledging the challenges with inconsistent and unreliable data on ESG scores, inconsistent terminology, and confusion regarding different investment strategies (Whelan, T., Atz, U., Van Holt, T., & Clark, C., 2021).

Despite some limitations, research on the relationship between ESG scores and stock performance can still provide valuable insights for investors and companies interested in the role of ESG in the financial markets. It is important to carefully consider these limitations when interpreting the results of such research.

2.2 Empirical studies of ESG and stock performance

Prior research on our topic examines the link between ESG practices and financial success, but most of those studies tended to concentrate on just one component of ESG (Smith *et al.,* 2007; Ponnu, 2008; Han *et al.,* 2016a). It is essential to highlight that ESG issues are interconnected, thus, focusing on one single subcomponent of ESG could be problematical (Galbreath, 2013).

Alareeni, B. A., & Hamdan, A. (2020) investigates whether there are relationships among corporate disclosure of ESG and S&P 500 firm's operational (ROA), financial (ROE) and market performance, and whether these relationships are positives, negatives or even neutral. The study conducted showed that ESG disclosures positively affect a firm's performance measures. Measuring ESG subcomponents separately showed that the environmental and corporate social responsibility disclosures are negatively associated with ROA and ROE.

Several studies in the past have examined the relationship between ESG scores and stock performance by using ESG-screened portfolios. Eccles et al. (2014) investigate the effect of corporate sustainability on organizational processes and performance. Their study is highly relevant for our topic of interest, as they investigate 180 U.S companies. By combining ESG scores from ASSET4 and Sustainable Asset Management, together with personal research and interviews they divided their sample into high or low ESG companies and found that high ESG score companies outperformed low ESG score companies, using a high-low strategy.

Corresponding with this, the study by Kempf and Osthoff (2007) use a sample period ranging from 1992 to 2004. While employing a long-short investment strategy by purchasing high SRI-rated (socially responsible investing) companies and selling low SRI-rated companies, they discovered that it produces high abnormal returns using ESG ratings from KLD Research & Analytics.

Borgers et al. (2013) provide evidence to suggest that the superior performance of enterprises with high Environmental, Social, and Governance (ESG) scores,

which was previously observed by Kempf and Osthoff (2007), diminishes after the initial sample period of 2004. Borgers et al. attribute this finding to the underreaction of the market, which resulted in anomalous returns that were initially favorable.

In the Journal of Sustainable Finance & Investment, Gunnar Friede, Timo Busch, and Alexander Bassen conducted a review study in 2015 that included more than 2000 global individual empirical publications. The authors aimed to overcome the shortcomings of previous studies, which made findings difficult to generalize. Thus, knowledge on the financial effects of ESG criteria remained fragmented. Their study extracted all provided primary and secondary data of previous academic review studies, resulting in a meta study combining the findings of 2200 individual studies. Evidence from their study shows that roughly 90% of studies find a nonnegative ESG-Corporate Financial Performance (CPI) relation. The key takeaway from their study is that most studies report positive findings. Notable is also the finding that close to $60 trillion in assets under management are managed by Principles of Responsible Investment (PRI) signatories, demonstrating a commitment to incorporating ESG performance within investment strategies.

In accordance with recent studies, a global survey, by PwC's Asset and Wealth Management Survey issued in 2022, with 250 respondent asset managers and institutional investors shows that nine out of ten asset managers believe that integrating ESG into their investment strategy will improve overall returns. Another result shows that 60 % of institutional investors reported that ESG investing has already resulted in higher performance yields, compared to non-ESG equivalents.

In contrast to these findings, Halbritter and Dorfleitner (2015) find that the magnitude and direction of the impact of ESG investing, using high ESG portfolios and low ESG portfolios, are substantially dependent on the rating provider, company sample, and subperiod. They suggest that investors cannot expect abnormal returns when trading a difference portfolio of high and low ESG

rated firms. Berg et al., (2022) underscore these findings, as they found that the ESG rating diverges significantly across different rating providers.

Our study aims to delve deeper into the results indicating a positive impact of ESG on a company's performance metrics, resulting in improved risk-adjusted returns. To achieve this, we intend to utilize two ESG rating agencies, enabling investors to make more informed decisions, and assessing whether variations in ratings between agencies might offer a higher anticipated abnormal return.

2.3 ESG rating weaknesses

Environmental, social, and governance (ESG) ratings have become an increasingly popular tool for investors seeking to assess the sustainability and social impact of companies. While ESG ratings can be useful in identifying companies that prioritize sustainability and social responsibility, there are several potential weaknesses or limitations of ESG ratings that investors should be aware of. Academic literature on ESG scores has brought attention to a significant issue, namely the absence of reporting guidelines and consensus among service providers on the most appropriate rating methodology. As a result, the choice of rating provider can have substantial implications on the results, underscoring the need for greater uniformity in the field.

A study by Berg, Kolbel & Rigobon (2022) investigates the divergencies in different ESG ratings. According to their research, the market for ESG scores has an inefficient flow of information, which creates an economic issue when investors do not share a common understanding and valuation of a company's ESG performance. Because it is difficult to pinpoint outperformers due to the variance amongst agencies, the authors contend that investors dismiss ESG performance as a component in corporate stock value.

Complementary to the study of Berg, Kolbel and Rigobon (2022), is the study by Gibson, Krueger, and Schmidt (2021), which is one of the most thorough data coverages on the subject. By investigating several different data providers in the period 2010 to 2017 using S&P 500 firms, their findings suggest that stock returns are positively related to ESG rating disagreement, which suggests a risk premium for firms with higher ratings of ESG disagreements.

There is a notable difference of opinion among financial experts when it comes to assessing ESG performance ratings. According to a study conducted by Billio et al. (2021), the disagreement among rating agencies regarding ESG ratings leads to a dispersal of the effect of ESG investors' preferences on asset prices. In some cases, even when there is an agreement, the impact of ESG investors' preferences is weak, resulting in no discernible effect on the financial performance of ESG portfolios. These findings have important implications for ESG investors, as they highlight the need for increased transparency and standardization in ESG rating methodologies to avoid the dispersion of ESG preferences' impact on asset prices. As such, policymakers and stakeholders in the financial industry should focus on addressing these concerns to ensure that ESG investments can contribute positively to sustainable development.

3. Hypothesis

3.1 Hypothesis and research question

It is common knowledge in finance that most investors seek to maximize their returns at a preferred level of risk. At the same time an increasing number of investors seek to incorporate environmental, social, and governance (ESG) when considering their investments. Even though sustainable investing and ESG have received increased attention over the last years, there is still some lack of knowledge on how to incorporate these practices as well as disagreements around their impact on financial returns.

In this thesis we present two portfolios with respectively low and high ESG risk, which will be used to perform a long-short investment strategy, in order to examine whether high ESG scores can generate higher abnormal returns in the US stock market. Our null and alternative hypothesis is presented below.

H_0: High ESG rated companies in the U.S stock market do not provide higher abnormal returns compared to low rated ESG companies.
H_A: High ESG rated companies in the U.S. stock market provide higher abnormal returns compared to low rated ESG companies.

When running our regressions, we expect to observe higher abnormal return for the portfolio consisting of companies with good ESG practices compared to the portfolio with bad ESG practices. Hence, we expect to achieve positive abnormal returns from our long-short zero investment strategy.

As previously discussed, it's worth noting that ESG scores can vary depending on the methodology and criteria used by the organization providing the scores. Given that there are several providers we would also like to investigate whether we will obtain different outcomes based on which provider we use. Hence, we will answer the additional research question:

Does the choice of ESG score provider affect the abnormal return on the US stock market?

In the following we will outline a research methodology describing how we aim to answer these questions. Further the thesis will present the data collection method, results, and analysis, and finally a conclusion.

4. Research Methodology

4.1 Portfolio construction

Our empirical analysis consists of a dataset of 497 US listed companies, which are the companies listed on the S&P 500 index. To construct our portfolios, these companies are divided into ten percentiles based on their ESG rating in the previous year, t-1. To perform our zero investment long-short strategy we subtract the bottom portfolio from the top portfolio, where the top portfolio consists of the companies with the highest ESG scores and the bottom portfolio consists of the companies with the lowest ESG scores.

For our analysis of the company's financial performance, we have conducted the daily closing prices of all the S&P 500 companies over the period from 2014 to 2022. The prices are collected from Yahoo Finance, which is a leading provider of stock quotes and financial data worldwide. Further, we have transformed these prices into daily returns, which serves as a base for our portfolio construction and regression analysis.

To divide our dataset into subgroups we have collected ESG ratings of the respective companies from 2013 to 2021, which is one year prior to the stock returns. The reason for this is that the ESG ratings must be lagged to construct an “accurate” investment strategy and to avoid look ahead bias. The top and bottom portfolios are further aggregated into both average-weight portfolios and value-weight portfolios, where the latter is based on the company's market capitalization.

To answer our additional research question, on whether the source provider of ESG score has an impact on the abnormal stock returns, we have included two providers, and three different ESG ratings for our analysis, to compare the portfolio performance. The paper will include the ESG ratings from Bloomberg, and both ESG and ESG Combined ratings from Refinitiv going forward with the analysis.

4.2 Model specification

The last step in our analysis consists of performing multiple linear regression models (OLS) using both providers and their respective rating methodologies. In our examination we investigate whether the companies' ESG scores have a significant impact on their returns, and if the ESG provider and rating methodology will impact these results.

For our regression analysis we have applied the Fama & French three-factor model, Carhart four-factor model, and Fama & French five-factor model with and without momentum. By incorporating models that encompass multiple factors to explain stock returns, we increase the robustness of our results. The excess return of the respective portfolios will be the dependent variable in our models. (Brooks, 2019).

4.2.1 Fama & French Three-Factor Model (1993)

As an extension of the CAPM model, which only considers market risk, Fama & French three-factor model introduces two other factors to explain stock returns. Small minus big (SMB) and high minus low (HML), which accounts for that small market capitalization firms and companies with high book-to-market ratio can generate higher returns because of an earned risk premium.

$$r_{ESG_{top}} - r_{ESG_{bottom}} = \alpha + \beta_{MKT} * (MKT_t - rf_t) + \beta_{SMB} * SMB_t + \beta_{HML} * HML_t + \varepsilon_t$$

Where,

$$\alpha = Alpha\ (Abnormal\ return)$$
$$\beta_{MKT} = Market\ risk\ exposure$$
$$MKT_t - rf_t = Market\ excess\ return\ (market\ risk\ premium)$$
$$\beta_{SMB} = Exposure\ to\ size\ factor$$
$$SMB_t = Size\ factor\ at\ time\ t$$
$$\beta_{HML} = Exposure\ to\ value\ factor$$
$$HML_t = Value\ factor\ at\ time\ t$$
$$\varepsilon_t = Error\ term$$

4.1.2 Carhart Four-Factor Model (1997)

The Carhart four-factor model extends the Fama & French three-factor model my adding another variable, momentum. Winners minus losers (WML) is a performance factor which measures previous winners and losers in the stock market.

$$r_{ESG_{top}} - r_{ESG_{bottom}} = \alpha + \beta_{MKT} * (MKT_t - rf_t) + \beta_{SMB} * SMB_t + \beta_{HML} * HML_t + \beta_{WML} * WML_t + \varepsilon_t$$

Where,

$$\beta_{WML} = Exposure\ to\ momentum\ factor$$
$$WML_t = Momentum\ factor\ at\ time\ t$$

4.2.3 Fama & French Five-Factor Model (2015)

Fama & French five-factor model extend the Fama & French three-factor model by adding two new variables: conservative minus aggressive (CMA) and robust minus week (RMW). Conservative minus aggressive refers to the company's investment strategy and robust minus weak refers to the company's profitability.

$$r_{ESG_{top}} - r_{ESG_{bottom}} = \alpha + \beta_{MKT} * (MKT_t - rf_t) + \beta_{SMB} * SMB_t + \beta_{HML} * HML_t$$

Where,

$$+\beta_{CMA} * CMA_t + \beta_{RMW} * RMW_t + \varepsilon_t$$
$$\beta_{CMA} = Exposure\ to\ investment\ policy\ factor$$
$$CMA_t = Investment\ policy\ factor\ at\ time\ t$$
$$\beta_{RMW} = Exposure\ to\ profitability\ policy\ factor$$
$$RMW_t = Profitability\ factor\ at\ time\ t$$

4.2.4 Fama & French Five-Factor Model with Momentum

Finally, the Fama & French five-factor with momentum is an extension of the previous model but with an additional momentum factor.

$$r_{ESG_{top}} - r_{ESG_{bottom}} = \alpha + \beta_{MKT} * (MKT_t - rf_t) + \beta_{SMB} * SMB_t + \beta_{HML} * HML_t + \beta_{CMA} * CMA_t + \beta_{RMW} * RMW_t + \beta_{WML} * WML_t + \varepsilon_t$$

4.3 Model testing

For the results and interpretation of our analysis to be valid when using OLS, the data must justify the assumptions of heteroscedasticity and autocorrelation (Brooks, 2019). To test for heteroscedasticity, we ran both a Whites test and a Breusch-Pegan test. Our findings, which are reported in Table 9 in the Appendix, show that heteroscedasticity is present in several of our models. To cope with this, we ran the regressions with heteroscedasticity again, this time with Whites Heteroscedastic-Consistent (HC) standard errors (White, 1980). To test for autocorrelation, we ran a Durbin-Watson test. The results of the tests show that there is no significant autocorrelation in any of our models, hence no further action is required (Appendix A, Table 10). Another important assumption for time-series inference is that the data must be stationary (Brooks, 2019). We ran an Augmented Dickey-Fuller test to test for stationarity and conclude from our findings that our data is stationary (Appendix A, Table 11). After testing and correcting our data for heteroscedasticity we conclude that the data is valid and ready for interpretation.

5. Data

5.1 Data sources

We obtained our monthly adjusted closing stock prices and returns for the S&P 500 companies and index over the period from 2014 to 2022 from Yahoo Finance. Yahoo is one of the world's largest providers of financial data and gives easy access to specific extraction of time series stock quotes (Yahoo, 2023). Further we obtained Bloomberg ESG ratings, and the market capitalization of all the sample companies from Bloomberg Professional Services, and the Refinitiv ESG and ESGC ratings from Thomson Reuters Eikon. The Fama-French three-factors (Fama & French, 1993), momentum factor (Carhart, 1997), Fama-French five factors (Fama & French, 2015), and the risk-free rate over the period from 2014 to 2022 are retrieved from Kenneth R. French data library (French, 2023).

5.2 ESG providers

For our study we will focus on the aggregated score of the three pillars of Environmental, Social and Governance (ESG). ESG ratings are obtained from two independent providers to be able to study how different ESG providers might affect investment decisions and returns related to ESG ratings. We have chosen Bloomberg and Refinitiv, which are both superior data providers within the field of finance research and data, including ESG ratings. The scores used for the analysis are Bloomberg ESG disclosure score, Refinitiv ESG score, and Refinitiv ESGC score.

The selection of rating providers is chosen because of their different levels of comprising ESG data. The Refinitiv ESG and ESGC scores are the most comprehensive and consists of a deeper analysis compared to the one provided by Bloomberg. While Bloomberg do not account for the quality of the data and only measure a company's transparency and disclosure of ESG data, Refinitiv compare a company to its industry peers on ESG performance metrics, before providing a score dependent on their performance relative to its peers.

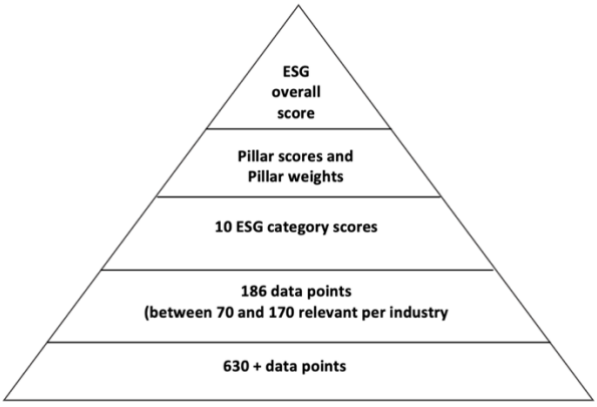
5.2.1 Bloomberg ESG Disclosure Score

Bloomberg's environmental, social, and governance (ESG) offers ESG scores for more than 15 000 companies in more than 100 countries. Their database dates to 2006 in addition to extensive ongoing data coverage. The ESG data is organized into more than 2000 fields that span several key sustainability topics such as air quality, climate change water and energy management, materials and waste, health and safety, audit risk and oversight, compensation, diversity, board independence, structure and tenure, and shareholder rights. Their ESG rating scale ranges from 1 to 100 (Bloomberg, 2023).

5.2.2 Refinitiv ESG Score

Refinitiv provides ESG ratings on more than 12 000 public and private companies globally, with time series data dating back to 2002. In measuring ESG, Refinitiv captures and calculates more than 630 different company-level ESG datapoints which breaks down to 186 of the most comparable and material per industry. These 186 datapoints are further categorized into 10 main themes including emission, innovation, resource use, human rights, product responsibility, workforce, community, management, stakeholders, and CSR strategy. The themes are then categorized under the three pillars environmental, social, and governance. The ESG pillar scores are determined as a relative sum of the category weights. The social and environmental weights vary for different industries, while the governance weight stays the same across the industries. Their ratings are available both in percentages (1 to 100) and in letter grades (from D- to A+) (Refinitiv, 2022).

Figure 1: Refinitiv ESG Scoring Methodology



5.2.3 Refinitiv ESGC Score

In addition to the ESG score presented above, Refinitiv also offers an additional ESGC score. This score is calculated as an average of the ESG and the ESG Controversies score when there are controversies during the fiscal year. However, if the ESG Controversies score is greater than the ESG score, the ESGC score will be equal to the ESG score. The ESG Controversies score is calculated based on 23 ESG controversy topics, which are benchmarked on industry group, and accounts for market capitalization bias from which large-cap companies are more likely to attract more media attention than smaller-cap companies. In practice this is accounted for by applying a sensitivity rate based on benchmark market capitalization intervals. If a company has no controversies within the fiscal year, they will get an ESG Controversies score of 100 (Refinitiv, 2022).

Table 1: ESG Controversies score calculation example

Scenario	ESG Controversies Score	ESG Score	ESGC Score
If controversies score is >= ESG score, then ESG score = ESGC score	100	89	89
If controversies score < ESG score, then ESG score = average of ESG and controversies score	48	49	48.5

Table 1 illustrates how the ESGC score is calculated based on the initial ESG score and additional ESG Controversies score. (Refinitiv, 2022)

All the chosen ESG data providers can be accessed through BI Norwegian Business School´s computers on campus.

5.3 Sample selection and screening

Since our study of ESG rating´s impact on financial return focuses on the US stock market, we decided to use the companies listed on the S&P 500 as our selected sample. The reasoning behind this consists of two main objectives, availability, and representability. One of the advantages of choosing such a big and established index, which mainly consists of larger-cap companies, is the data availability. To conduct a robust time series analysis, it is important to be able to include enough data, both regarding ESG ratings data, financial return data, and market capitalization data. We consider S&P 500 a representative sample for the US equity markets, compared to other indexes and samples, due to its size and its sector span. Compared to the Dow Jones Industrial Average index, which is also commonly used as a benchmark for the US stock market, S&P 500 comprises more stocks across all sectors, approximately 500 compared to Dow´s 30 (S&P Global, 2023).

Our sample period covers the period from January 2014 to December 2022. The returns are conducted monthly, which leaves us with nine years of monthly returns, resulting in 108 observations for our regression analysis. The ESG ratings are conducted from 2013 to 2021 which is the last available rating at the time of the analysis. The ESG ratings are lagged by one period to avoid look ahead bias, and we rebalance the portfolios yearly based on ESG_{t-1}.

Initially, we gathered data for both financial returns, ESG ratings, and market capitalization when the S&P 500 consisted of 503 companies. Even though the actual number of companies have not been a subject to change over the research period, there have been some changes in the composition of companies over the past year. During the study period we have made some changes which made it necessary to collect new data for the Refinitiv samples. In order to keep our sample consistent among the providers, we decided to remove certain companies from our sample, which includes Lumen Technologies Inc, Signature Bank,

Silicon Valley Bank, Constellation Energy Corp, GE HealthCare Technologies Inc, and Linde Group. Where Lumen, Signature Bank, and Silicon Valley Bank have been removed from the index during the research period, and Constellation Energy, GE HealthCare, and Linde Group made the index at recent point in time which results in lack of data. This leaves us with a sample of 497 companies moving forward with our analysis.

5.4 Descriptive statistics

5.4.1 Sector Breakdown

Table 2: S&P 500 Sector Breakdown

S&P 500 Sector Breakdown	
Sector	**Weight**
Healthcare	12.72%
Industrials	13.72%
Consumer Cyclicals	13.72%
Technology	17.50%
Financials	11.63%
Consumer Non Cyclicals	7.75%
Utilities	5.96%
Basic Materials	5.57%
Real Estate	5.96%
Energy	5.17%

Table 2: Presents the allocation of companies among economic sectors. The companies are sorted by the TRBC economic sector identifier. (Refinitiv, 2023)

As previously discussed, the S&P 500 index represents a broad selection of sectors, which makes it a suitable proxy for the U.S. equity market. The sector breakdown for our sample presented in Table 2 above outlines the ten economic sectors sorted by the TRBC[1] economic sector identifier. The data reveals that technology comprises the largest portion of our sample with a percentage share of 17.5%, while energy comprises the smallest portion with a percentage share

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

of 5.17%. The technology sector includes several growth stocks, which reinvest most of their profits in expansion versus paying dividends. Tech stock investments are cyclical, and usually perform better in strong economies. However, they saw a boost during the coronavirus pandemic, due to increased demand of video-conferencing platforms and cloud storage as more companies adopted remote work. The energy sector is more sensitive to economic movements and supply-demand trends, which could be some explanation for why it comprises the smallest portion with regards to industry.

5.4.2 Rating Discrepancy

Table 3 describes the discrepancy in ESG ratings of our two providers of ESG scores, Refinitiv and Bloomberg. As expected, the table visualizes a significant correlation between the Refinitiv ESG Score and Refinitiv ESGC Score, as they are given from the same provider. However, by incorporating the factor of controversy, a substantial influence on the Refinitiv score is observed. Even though the correlation coefficient of 0.83 between the two Refinitiv scores is relatively high in comparison to the results reported by Berg et al. (2022), the magnitude of the disparity between the scores attributable to a single factor is noteworthy when all other ESG metrics are held constant. This finding highlights the significance of considering the impact of the controversy factor when evaluating ESG performance.

The Pearson correlation between Refinitiv ESG score and Bloomberg disclosure score is 0.74, which suggests a strong positive correlation between the two rating providers. Similarly, the correlation between Refinitiv ESGC and Bloomberg disclosure score is 0.59, which indicates a moderate positive correlation between these two ratings.

It is important to note that correlation does not necessarily imply causation. Therefore, we cannot conclude that one rating provider causes another to change. The correlation matrix highlights the findings of Berg et al. (2022), which suggests that other factors influence the ESG score provided by each agency, such as different methodologies, data sources, and the weighting of various factors within each rating system.

Table 3: Pearson Correlation Matrix

Pearson Correlation			
	Refinitiv ESG Score	Refinitiv ESGC Score	Bloomberg ESG Score
Refinitiv ESG Score	1		
Refinitiv ESGC Score	0.8274	1	
Bloomberg ESG Score	0.7414	0.5873	1

Table 3 represents the average correlation between the different ESG scores throughout our sample period. The correlation is calculated using the Pearson correlation coefficient.

Table 4: Average Economic Sector Scores

Average Economic Sector Scores			
Sector	**Refinitiv ESG Score**	**Refinitiv ESGC Score**	**Bloomberg Score**
Healthcare	60.73	56.11	49.84
Industrials	57.79	51.12	47.72
Consumer Cyclicals	57.94	54.00	45.74
Technology	56.32	53.03	46.11
Financials	57.83	52.53	46.08
Consumer Non Cyclicals	61.01	55.68	48.09
Utilities	60.03	56.62	51.82
Basic Materials	58.63	55.40	49.44
Real Estate	61.68	57.14	49.45
Energy	54.88	49.21	47.17

Table 4 shows the average economic sector score for S&P 500 index constituents (2013-2021). Sectors are based on TRBC Economic Sector Name from Refinitiv.

Table 4 presents the average ESG score when screening for TRBC economic sector, and highlights the discrepancy discussed using the Pearson Correlation Matrix in Table 3. Based on the data presented in the table it can be observed that the highest average ESG score across all sectors is in the utilities sector with an average score of 56.16 between the two providers. We also see the Energy sector averaging the lowest ESG score of 50.42. It appears from the table that Refinitiv ESG scores are higher than the Bloomberg ESG scores.

Table 5: Summary Statistics ESG Rating Samples

Refinitiv ESG Summary Statistics									
Statistic	2013	2014	2015	2016	2017	2018	2019	2020	2021
N	434	445	475	480	489	492	494	496	497
Mean	50.82	51.31	53.48	56.20	58.30	59.94	62.79	65.88	68.70
Median	50.69	51.84	55.10	58.16	60.67	62.65	65.43	68.32	69.90
Std	19.10	18.39	18.57	17.72	17.53	16.93	15.61	14.54	12.69
Min	2.49	2.46	5.81	9.19	10.87	11.79	19.71	6.35	26.60
Max	92.28	92.39	92.84	91.29	91.89	93.18	93.84	93.45	92.90

Refinitiv ESGC Summary Statistics									
Statistic	2013	2014	2015	2016	2017	2018	2019	2020	2021
N	434	445	475	480	489	492	494	496	497
Mean	46.05	46.77	50.64	52.57	55.04	56.18	58.12	58.32	61.48
Median	45.61	45.71	50.85	53.10	55.37	57.01	59.21	58.61	63.21
Std	16.10	16.90	17.43	16.81	16.75	16.52	15.75	15.00	14.85
Min	2.49	2.46	5.81	9.19	10.87	11.79	19.71	6.35	16.06
Max	88.23	89.29	88.02	88.77	90.69	91.97	92.65	89.73	92.10

Bloomberg ESG Summary Statistics									
Statistic	2013	2014	2015	2016	2017	2018	2019	2020	2021
N	468	474	477	479	483	491	496	498	499
Mean	39.76	40.62	44.36	44.79	46.45	49.98	52.07	54.06	55.38
Median	37.03	37.75	42.22	43.23	46.34	51.15	52.74	54.41	56.00
Std	10.60	11.00	11.77	1.86	11.92	12.17	11.74	11.20	10.97
Min	5.09	16.36	27.40	27.40	15.82	28.82	28.82	28.28	29.32
Max	73.84	72.96	78.83	78.83	79.81	82.54	82.54	85.71	85.71

Table 5 presents the summary statistics for the Refinitiv ESG, Refinitiv ESGC, and Bloomberg ESG samples. The table includes the number of observations (N), the average score (Mean), median score (Median), the standard deviation (Std), the smallest observation (Min), and the largest observation (Max), for each year of the ESG sample period (2013 – 2021).

To further visualize the discrepancies between the two agencies and different rating methodologies, Table 5 presents summary statistics for the three ESG rating samples. Both the mean and the median suggest that the Bloomberg ESG score is the lowest among the three scoring methodologies, and that the Refinitiv ESGC score is lower than the Refinitiv ESG score. The latter occurs from the fact that the ESGC score is created as a combination of the ESG and the ESG Controversies score which might decrease, but never increase the original score as described in Table 5. We also observe an increasing trend of the number of observations, the average and median score, the smallest and the largest observation for both providers and rating methodologies over the sample period. The standard deviation decreases over the sample period for both Refinitiv ESG and ESGC, while Bloomberg exhibits greater stability.

5.5 Limitations to data

It is important to acknowledge limitations to our data sample and study. It is possible that our reliance on ESG scores from solely Bloomberg and Refinitiv, with a total of three scores, could be deemed inadequate. It would be preferable to incorporate supplementary scores; however, we have encountered obstacles such as insufficient data coverage and subscription requirements from other

providers, impeding our ability to do so. We acknowledge that the present circumstance poses a constraint to our data, as there exists a vast array of ESG score providers, each yielding different outcomes when conducting top-to-bottom methodology. The work of Berg et al., (2022) underscores the limitations of rating disparities in empirical research, demonstrating that the utilization of the six leading score providers produces a correlation of 0.54.

In addition, there are other biases and challenges with our data and analysis to be taken into consideration. The choice of S&P 500 companies as our data sample limits the study to large-cap companies. This might cause a bias in our study. Having a sample consisting of large-cap companies also mitigates potential lack of datapoints, which is more likely to occur by examining smaller unlisted companies. This makes the sample more robust and suitable for the study conducted in this thesis. Additionally, given the diversification of the S&P 500 index we consider the selected sample highly representable for the U.S stock market.

Despite our data sample being derived from one of the world's largest indexes, containing robust data, it is crucial to note that there are still gaps in the data, including returns, ESG ratings, and market capitalization, for several companies and years during our sample period. These gaps are particularly relevant for the first years of the sample period, as the ESG rating agencies' methodologies and coverage have evolved over time. As a result, it is important to take these limitations into account while interpreting our findings.

6. Analysis

6.1 Cumulative portfolio returns

Figure 2: Cumulative Portfolio Returns

Figure 2 presents the cumulative return of the top and bottom portfolios relative to the cumulative return of the S&P 500 index over the sample period 2014-2022.

To get an overview of the relative portfolio performance Figure 2 presents the cumulative return of both the average and value weighted top and bottom portfolios for the Bloomberg ESG Disclosure score, Refinitiv ESG score, and Refinitiv ESGC score. The respective top and bottom portfolios are displayed together with the cumulative return of the S&P 500 index over the same period.

According to the table`s data the bottom portfolio outperforms the top portfolio and the S&P 500 index for all weightings, providers, and rating metrics.

6.2 Regression results

In this section we will present the regression results for the average and value-weighted portfolios from the Fama-French 5 factor model, with and without momentum. The regression results from the Fama-French 3-factor model and the Carhart four factor model are presented in the Appendix.

6.2.1 Bloomberg

Table 6: Bloomberg ESG Disclosure Score Portfolios Regression Output

Bloomberg

	Fama-French 5		Fama-French 5 + mom	
	AW	**VW**	**AW**	**VW**
α	-0.0076 ***	-0.0084***	-0.0067***	-0.0077***
	(-3.908)	(-3.307)	(-3.424)	(-3.037)
MKTRf	0.0054	-0.0167	-0.0402	-0.0487
	(0.089)	(-0.273)	(-0.828)	(-0.767)
SMB	-0.1330	-0.0147	-0.1870**	-0.0526
	(-1.423)	(-0.132)	(-2.163)	(-0.465)
HML	0.1278*	-0.0707	0.0479	-0.1268
	(1.752)	(-0.736)	(0.623)	(-1.259)
MoM			-0.1969***	-0.1382*
			(-3.192)	(-1.712)
RMW	0.1856	0.4626***	0.1346	0.4268***
	(1.417)	(3.332)	(1.266)	(3.068)
CMA	0.5940***	0.5353***	0.6433***	0.5699***
	(4.352)	(3.583)	(5.637)	(3.815)
Observation	108	108	108	108
R2	0.472	0.289	0.520	0.309
Adjusted R2	0.446	0.254	0.491	0.268

Table 6 presents the regression output from the Bloomberg ESG disclosure study for both average (AW) and value (VW) weighted portfolios using the Fama-French 5 factor model with and without momentum. The alpha represents the returns from the long-short zero investment portfolios sorted with respect to the Refinitiv ESGC factors. MKTRf represents the market risk factor, SMB the small minus big factor, HML the high minus low factor, MoM the winner minus losers' factor, RMW the robust minus weak factor, and the CMA the conservative minus

*aggressive factor. The stars behind the coefficients represent the significance level, where *=10%, **=5%, and ***=1%. The corresponding t-statistic is represented in the brackets under the coeffects.*

The results from the Bloomberg ESG disclosure score study show negative alphas for both models, which is significant on a 1% level. This implies that the long-short investment strategy produces negative abnormal returns, indicating that the top portfolio, which consists of the companies with the highest ESG ratings, underperforms relative to the bottom portfolio, which consists of the companies with the lowest ESG ratings. These results violate our hypothesis, which states that there is a positive relationship between high ESG scores and financial return in the US stock market.

The market risk factor is negative for both portfolios and models, except for the average weighted portfolio in the Fama-French five factor model. A negative market risk factor indicates that the high ESG score portfolio has a lower beta than the low ESG score portfolio, which means that the high ESG rated companies have a lower exposure to systematic risk than the low rated ESG companies. Even if this relationship appears as intuitive, the findings are not significant on any level, and we cannot conclude on the exposure to the market factor.

The SMB factor is only significant in the average weighted portfolio regressed on the Fama-French five factor model with momentum. The coefficients are negative, indicating that the top portfolio contains more companies with higher market capitalization than the bottom portfolio. Equivalent, the HML factor only appears as significant at a 10% level in the average weighted portfolio, regressed on the Fama-French five factor model. The HML factor is positive and significant for the average weighted portfolio under the five-factor model. This implies that the higher ESG rated companies have a higher book-to-market ratio compared to the lower ESG rated companies, hence the top portfolio is relatively more exposed to value companies than the bottom portfolio.

The RMW factor is positive and significant for the value weighted portfolios, but insignificant for the average weighted portfolios in the Fama-French five

factor models with and without momentum. The positive coefficients indicate that the top portfolio, with higher ESG ratings, are more exposed to the profitability premium. Hence the companies in the top portfolio might have better management and operational practices which turns into greater profitability. The CMA factor is also positive significant on a 1% level for both the average and value weighted portfolio in the five and six factor models. The positive CMA factor suggests that the high ESG portfolios consists of companies with more conservative investment and financing policies compared to the low ESG portfolios.

6.2.2 Refinitiv ESG

Table 7: Refinitiv ESG Score Portfolios Regression Output

Refinitiv ESG

	Fama-French 5 factor		Fama-French 5 factor + mom	
	AW	**VW**	**AW**	**VW**
α	-0.0063***	-0.0045*	-0.0056***	-0.0038
	(-3.606)	(-1.924)	(-3.254)	(-1.587)
MKTRf	0.0409	-0.0143	0.0059	-0.0472
	(0.967)	(-0.185)	(0.137)	(-0.557)
SMB	-0.1067	-0.1469	-0.1482*	-0.1860
	(-1.379)	(-1.272)	(-1.940)	(-1.534)
HML	0.1720**	-0.0719	0.1106	-0.1297
	(2.591)	(-0.591)	(1.626)	(-0.998)
MoM			-0.1515***	-0.1424
			(-2.779)	(-1.369)
RMW	0.0395	0.0287	0.0003	-0.0082
	(0.412)	(0.213)	(0.003)	(-0.058)
CMA	0.4959***	0.2910	0.5339***	0.3267
	(4.806)	(1.394)	(5.293)	(1.595)
Observation	108	108	108	108
R2	0.469	0.097	0.507	0.128
Adjusted R2	0.443	0.053	0.478	0.076

Table 7 presents the regression output from the Refinitiv ESG study for both average (AW) and value (VW) weighted portfolios using the Fama-French 5 factor model with and without momentum. The alpha represents the returns from the long-short zero investment portfolios sorted with respect to the Refinitiv ESGC factors. MKTRf represents the market risk factor, SMB the small minus big factor, HML the high minus low factor, MoM the winner minus losers' factor, RMW the robust minus weak factor, and the CMA the conservative minus aggressive factor. The

*stars behind the coefficients represent the significance level, where *=10%, **=5%, and ***=1%. The corresponding t-statistic is represented in the brackets under the coeffects.*

The Refinitiv ESG study produces the same results as the Bloomberg study in terms of abnormal returns. Nevertheless, the coefficients lack significance when the value weighted portfolio is regressed on the six-factor model. Even though the significance disappears in the last regression, the consistency with the Bloomberg study supports the negative relationship between ESG ratings and financial performance.

The market risk factor appears as negative for the value weighted portfolio for both models, which is equivalent to the Bloomberg study. The average weighted portfolio produces the opposite result, indicating a higher beta for the top portfolio compared to the bottom portfolio. The lack of significant coefficients makes it again impossible to conclude on the market risk exposure.

Consistent with the Bloomberg study, the SMB factor only appears as significant when the average weighted portfolio is regressed on the six-factor model, and the HML factor when regressing the average weighted portfolio on the five-factor model. The SMB factor still appears negative, and HML positive, for the average weighted portfolios. This supports previous findings that the top portfolio consists of more companies with higher market capitalization and a higher book-to-market ratio in comparison to the bottom portfolio. The CMA factor is significant and consistent with the Bloomberg study for the average weighted portfolio but fails to show any significance for the value weighted portfolio in both models. Furthermore, the portfolios show no difference in terms of the profitability risk factor, which is consistent across both models and weighting strategies.

6.2.3 Refinitiv ESGC

Table 8: Refinitiv ESGC Score Portfolios Regression Output

Refinitiv ESGC

	Fama-French 5 factor		Fama-French 5 factor + mom	
	AW	**VW**	**AW**	**VW**
α	0.0065 (1.429)	0.0013 (0.430)	0.0058 (1.329)	0.0009 (0.299)
MKTRf	-1.1714*** (-10.750)	-1.1331*** (-15.175)	-1.1396*** (-8.678)	-1.1143*** (-14.239)
SMB	0.2997 (1.505)	0.3700*** (2.711)	0.3373* (1.743)	0.3923*** (2.814)
HML	-0.1965 (-1.149)	0.0071 (0.060)	-0.1407 (-0.764)	0.0401 (0.323)
MoM			0.1374 (0.921)	0.0814 (0.818)
RMW	0.2745 (1.112)	0.2556 (1.511)	0.3101 (1.075)	0.2767 (1.614)
CMA	0.6868** (2.585)	0.3289* (1.807)	0.6523** (0.035)	0.3085* (1.677)
Observation	108	108	108	108
R2	0.615	0.742	0.619	0.743
Adjusted R2	0.597	0.729	0.596	0.728

*Table 8 presents the regression output from the Refinitiv ESGC study for both average (AW) and value (VW) weighted portfolios using the Fama-French 5 factor model with and without momentum. The alpha represents the returns from the long-short zero investment portfolios sorted with respect to the Refinitiv ESGC factors. MKTRf represents the market risk factor, SMB the small minus big factor, HML the high minus low factor, MoM the winner minus losers' factor, RMW the robust minus weak factor, and the CMA the conservative minus aggressive factor. The stars behind the coefficients represent the significance level, where *=10%, **=5%, and ***=1%. The corresponding t-statistic is represented in the brackets under the coeffects.*

The results from the Refinitiv ESGC study differ from the Bloomberg ESG disclosure and Refinitiv ESG score studies. By applying the Refinitiv ESGC scores to our portfolio construction, all models produce positive alphas in both average weighted and value weighted portfolios. These results indicate that higher ESG ratings produce higher abnormal returns, which would support our initial hypothesis. There is lack of significance across all models and weighting

strategies, hence we cannot conclude any positive relationship between high ESGC ratings and stock returns.

The market risk factor is negative and significant at 1% for both the value weighted and average weighted portfolios in both models. This implies that the top ESG portfolio has less exposure to systematic risk than the bottom ESG portfolio. This could suggest that the portfolio has exposure to factors that are not captured by the market risk factor.

The momentum factor and the RMW factor fail to provide any significant results for either average weighted or value weighted portfolios. The SMB coefficients are all positive and significant, with the exception of the average weighted portfolio in the five-factor model. The value factor (HML) fails to provide any significance in both portfolios and models but produces negative coefficients for the average weighted and positive coefficients for the value weighted portfolios. Interestingly, both SMB and HML contradicts with the results from the previous studies, which implies that the ESG Controversies score might create a market capitalization bias in our sample even though Refinitiv aim to control for size bias in their rating methodology. Equivalent to the Refinitiv ESG study, both portfolios and models fail to provide any difference in terms of the profitability factor. On the other hand, the investment factor (CMA) supports previous findings that the top portfolio consists of companies with more conservative investment policies compared to the bottom portfolio.

6.3 Discussion

6.3.1 Abnormal Return

Our study does not provide evidence to support our hypothesis which suggests that there is a positive relationship between high ESG ratings and financial return in the US stock market. In contrast, the Bloomberg ESG disclosure score study produces negative significant alphas when regressing our long-short investment strategy on the risk factors in every model. This indicates a reverse relationship between ESG scores and stock returns on the U.S. equity market. This evidence contradicts with the findings of Kempf and Osthoff (2007) and Eccles et al.

(2014), which suggests that high SRI and ESG rated companies outperform low SRI and ESG companies in terms of abnormal returns. Nonetheless, it is worth noting that both studies examine a different sample and sample period, as well as applying different measures and providers of environmental, social, and governance compared to our study. These discrepancies and the resulting conflicting outcomes align with the study of Berg et al., (2022), which highlights that ESG rating diverges significantly across different rating providers.

The Refinitiv ESG study supports our findings from the Bloomberg ESG Disclosure study in terms of abnormal returns. The study produces negative significant alphas for both average and value weighted portfolios in the five-factor model, and for the average weighted six-factor model.

The Refinitiv ESGC study fails to produce any significant alphas, leaving the study inconclusive in terms of ESGC and financial performance. The results from the Refinitiv ESGC study suggests a neutral relationship between ESGC ratings and stock performance. These results align with the findings of Halbitter and Dorfleitner (2015) which suggests that investors cannot expect abnormal returns when trading a difference portfolio of high and low rated ESG firms.
One potential explanation for a neutral relationship lies in the complexity associated with assessing ESG performance and the limited scope of ESG ratings in capturing the entirety of ESG activities' effects. The majority of ESG score metrics rely on qualitative data, posing challenges in quantifying and aggregating them into a comprehensive score. Furthermore, our analysis reveals significant dissimilarities in the scoring methodology and data collection processes employed by the two rating providers. Given these pronounced disparities, it becomes less probable that the diverse ESG scores effectively reflect the true extent of ESG performance.

Our results align with the shareholder theory, which claims that businesses should refrain from any activities not maximizing value for the firm and leave philanthropy and ESG activities to individuals. According to shareholder theory, ESG activities should negatively affect stock returns and increase the performance of companies that do not focus on ESG activities (Friedman, 1970).

Our study underscores this view as it produces negative alphas for both the Bloomberg ESG Disclosure score and Refinitiv ESG score.

6.3.2 Systematic Risk

We obtain negative market risk coefficients for all providers, portfolios, and models, except for the average weighted Bloomberg portfolio using the five-factor model, and the average weighted Refinitiv portfolio using the five-factor model with and without momentum. These results imply that the high ESG score companies have lower exposure to systematic risk than the low ESG score companies. The lack of statistical significance makes it difficult to confirm this relationship, where the Refinitiv ESGC study is the only one to provide significant coefficients for all portfolios and models. The Refinitiv ESG study does not produce any significant market risk coefficients, and the Bloomberg ESG disclosure study only produces significant coefficient for the average weighted portfolio under the Carhart four-factor model.

As expected, the systematic risk exposure increases and gets more robust when accounting for ESG Controversies. This relationship indicates that high-risk companies receive more media attention than low-risk companies. However, the lack of significance in the Refinitiv ESG study, compared to the Refinitiv ESGC study is somehow surprising.

6.3.3 Factor Risk Exposure

All portfolios and models within the Bloomberg ESG Disclosure and the Refinitiv ESG study produces negative SMB coefficients. Given our long-short investment strategy this indicates a greater exposure to larger market capitalization companies in the top portfolio compared to the bottom portfolio. This composition seems intuitive considering that larger market capitalization companies might possess advantages regarding disclosure of ESG as well as interpretation of ESG practices compared to smaller market capitalization companies. One plausible explanation for the advantage is the availability of superior resources, as the implementation and execution of ESG disclosures and practices may be financially onerous for smaller market capitalization companies, which is also shown in Drempetic et al. (2020). Furthermore,

companies with larger market capitalization may possess a competitive edge attributable to their expertise in the ESG domain, as well as their heightened obligation to satisfy the ESG disclosure and practice requirements of a larger number of stakeholders.

In contrast to the Bloomberg and Refinitiv ESG analysis, the incorporation of ESG controversies alters this pattern and reveals an inverse relationship between company size and ESG ratings. The findings suggest that firms with greater market capitalization are subject to a greater penalty with respect to controversies, relative to their smaller counterparts. An explanation for this occurrence could be found in Dorfleitner et al. (2020), which concludes that having a "clean coat" regarding controversies is especially profitable for smaller companies, as the absence of scandals may be overlooked and incorrectly incorporated in the market prices.

As for the HML factor we observe positive significant factor exposure for the average weighted portfolios in both models. The value weighted portfolios do however appear as negative but insignificant in both the Bloomberg ESG Disclosure and the Refinitiv ESG study. A possible explanation for these differences is that the average weighting strategy assigns more weight to smaller companies compared to the market capitalization weight strategy. A high book-to-market ratio indicates that the market is valuing the company's equity cheaply compared to its book-value. This is often observed for smaller market capitalization companies due to different factors such as risk perception, information asymmetry, growth expectations, and illiquidity.

All studies and portfolios produce positive CMA coefficients in every model. The coefficients are all statistically significant on a 1%, 5%, or 10% level, except for the value weighted Refinitiv ESG portfolios which remains insignificant. The consistent positive coefficients suggest that in general, higher ESG rated companies are more conservative in their investment practices than the lower ESG rated companies. These findings could possibly be correlated with the book-to-market findings, as value companies might be more likely to adopt more conservative investment practices compared to growth companies. However, the

relationship between book-to-market ratio and investment policy conservatism is a topic of ongoing debate within finance research, thus we cannot draw any conclusion on this matter.

7. Conclusion

This thesis has investigated the relationship between ESG integration and financial return on the U.S. stock market, using companies listed on the S&P 500 index as a benchmark. We have created top- and bottom portfolios based on the companies ESG ratings, provided by Bloomberg and Eikon Refinitiv, to perform a long-short zero investment strategy applying the Fama-French and Carhart risk factors to our analysis.

By running multiple linear regressions, for both average weighted and value weighted portfolios, we find a statistically significant negative relationship between ESG ratings and financial performance for both the Bloomberg ESG Disclosure and the Refinitiv ESG studies. These findings support our null hypothesis; *High rated ESG companies in the U.S stock market do not provide higher abnormal returns compared to low rated ESG companies.* The Refinitiv ESGC study produces insignificant positive alphas, which fails to provide any conclusion regarding ESG rating´s impact on financial return, indicating a neutral relationship. Even if our results are tending towards a negative relationship between high ESG rated companies and financial performance the Refinitiv ESGC study prevents us from providing a definitive answer on this matter. Due to the presence of ambiguous and inconclusive results our conclusion aligns with the study of Halbitter and Dorfleitner (2015), which suggests that the relationship between ESG and stock return is dependent on the rating provider, company sample, and subperiod.

For future research it would be interesting to delve deeper into different industries. Industries react differently to the integration of ESG factors and therefore it is not necessarily wise to treat them by the same token. We also see the need for a uniformed framework as today's research in the field often suffers from inconsistent results without any clear answers to the impact of ESG on stock performance.

8. References

Alareeni, B. A., & Hamdan, A. (2020). ESG impact on performance of US S&P 500-listed firms. *Corporate Governance: The International Journal of Business in Society*.

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

Billio, M., Costola, M., Hristova, I., Latino, C., & Pelizzon, L. (2021). Inside the ESG ratings:(Dis) agreement and performance. *Corporat*

Bloomberg Professional Services. *Global Environmental, Social, and Governance – ESG Data. Content and Data.* Retrieved January 3, 2023, from: https://www.bloomberg.com/professional/dataset/global-environmental-social-governance-data/

Brooks, C. (2019) *Introductory to Econometrics for Finance.* (4. Edition). Cambridge university press.

Capelle-Blancard, G., & Petit, A. (2019). Every little helps? ESG news and stock market reaction. *Journal of Business Ethics*, *157*, 543-565.

Carhart, M. (1997). *On Persistence in Mutual Fund Performance.* Journal of Financial Economics.

Dorfleitner, G., Kreuzer, C., & Sparrer, C. (2020). ESG controversies and controversial ESG: about silent saints and small sinners. *Journal of Asset Management*, *21*(5), 393-412.

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*, 333-360.

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

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

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

Fatemi, A., Glaum, M., & Kaiser, S. (2018). ESG performance and firm value: The moderating role of disclosure. *Global finance journal*, *38*, 45-64.

French, K. (2023). Kenneth R. French data library. Retrieved: January 5. 2023, from:
https://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.

Friedman, M. (2007). *The social responsibility of business is to increase its profits* (pp. 173-178). springer berlin heidelberg.

Galbreath, J. (2013). ESG in focus: The Australian evidence. *Journal of business ethics*, *118*(3), 529-541.

Gibson Brandon, R., Krueger, P., & Schmidt, P. S. (2021). ESG rating disagreement and stock returns. *Financial Analysts Journal*, *77*(4), 104-127.

GSIA. (2021). *Global Sustainable Investment Review 2020.* Retrieved January 2, 2023, from: http://www.gsi-alliance.org/.

Hill, J. (2020). *Environmental, Social, and Governance (ESG) investing: A balanced analysis of the theory and practice of a sustainable portfolio*. Academic Press.

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

Han, J.-J., Kim, H.J. and Yu, J. (2016a), "Empirical study on relationship between corporate social responsibility and financial performance in Korea", Asian Journal of Sustainability and Social Responsibility, Vol. 1 No. 1, pp. 1-16.

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

Kim, J., Chung, S., & Park, C. (2013). Corporate social responsibility and financial performance: the impact of the MSCI ESG ratings on Korean firms. *Journal of the Korea Academia-Industrial Cooperation Society*, *14*(11), 5586-5593.

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.

Marketplace, S&P Global. *S&P Global ESG Scores.* Retrieved January 4, 2023, from: https://www.marketplace.spglobal.com/en/datasets/s-p-global-esg-scores-(171)

Ponnu, C. H. (2008). Corporate governance structures and the performance of Malaysian public listed companies. *International Review of Business Research Papers*, *4*(2), 217-230.

Pwc. (2022) Asset and Wealth Management Revolution 2022 – Exponential Expectations for ESG. Retrieved: January 5, 2023, from: https://www.pwc.com/gx/en/financial-services/assets/pdf/pwc-awm-revolution-2022.pdf

Refinitiv. *Environmental, Social, and Governance Sources from Refinitiv May, 2022*. Retrieved January 3, 2023, from: https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/refinitiv-esg-scores-methodology.pdf

Smith, M., Yahya, K., & Amiruddin, A. M. (2007). Environmental disclosure and performance reporting in Malaysia. *Asian Review of Accounting*.

S&P Global. (2023). *S&P Dow Jones Indices*. https://www.spglobal.com/spdji/en/indices/equity/dow-jones-industrial-average/#overview

Whelan, T., Atz, U., Van Holt, T., & Clark, C. (2021). ESG and financial performance. *Uncovering the Relationship by Aggregating Evidence from*, *1*, 2015-2020.

Sustainalytics https://www.sustainalytics.com/esg-data

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica: journal of the Econometric Society, 817-838.

Yahoo Finance. (2023). https://finance.yahoo.com/

9. Appendix

A: Regression results

Table 9: Fama-French 3 factor model regression output

Regression output: Fama-French 3 factor model

	Bloomberg		Refinitiv ESG		Refinitiv ESGC	
	AW	VW	AW	VW	AW	VW
α	-0.0058**	-0.0061**	-0.0051***	-0.0036	0.0087**	0.0026
	(-2.584)	(-2.196)	(-2.639)	(-1.401)	(1.890)	(0.821)
MKTRf	-0.0441	-0.0259	-0.0112	-0.0522	-1.2130***	-1.1296***
	(-0.878)	(-0.416)	(-0.242)	(-1.669)	(-11.842)	(-16.216)
SMB	-0.2359***	-0.2286**	-0.1579*	-0.1413	0.1654	0.2270*
	(-2.717)	(-2.121)	(-1.748)	(-1.308)	(0.934)	(1.885)
HML	0.3644***	0.1931**	0.3582***	0.0223	0.1659	0.2352***
	(6.081)	(2.595)	(6.138)	(0.234)	(1.356)	(2.829)
MoM						
RMW						
CMA						
Observatio	108	108	108	108	108	108
R2	0.308	0.104	0.342	0.042	0.586	0.726
Adjusted R2	0.288	0.078	0.323	0.014	0.574	0.718

*Table 9 presents the regression output from the Bloomberg Disclosure, Refinitiv ESG, and Refinitiv ESGC study for both average (AW) and value (VW) weighted portfolios using the Fama-French 3 factor model. The alpha represents the returns from the long-short zero investment portfolios sorted with respect to the Refinitiv ESGC factors. MKTRf represents the market risk factor, SMB the small minus big factor, HML the high minus low factor, MoM the winner minus losers factor, RMW the robust minus weak factor, and the CMA the conservative minus aggressive factor. The stars behind the coefficients represent the significance level, where *=10%, **=5%, and ***=1%. The corresponding t-statistic is represented in the brackets under the coeffects.*

Table 10: Carhart four factor model regression output

Regression output: Carhart four factor model

	Bloomberg		Refinitiv ESG		Refinitiv ESGC	
	AW	VW	AW	VW	AW	VW
α	-0.0050**	-0.0055*	-0.0046**	-0.0031	0.0079*	0.0022
	(-2.257)	(-1.958)	(-2.360)	(-1.164)	(1.755)	(0.693)
MKTRf	-0.0907*	-0.0640	-0.00443	-0.0852	-1.1660***	-1.1072***
	(-1.699)	(-0.953)	(-0.906)	(-1.004)	(-9.885)	(-14.627)
SMB	-0.2646***	-0.2520**	-0.1782*	-0.1617	0.1943	0.2408*
	(-3.073)	(-2.326)	(-1.790)	(-1.336)	(1.067)	(1.973)
HML	0.3040***	0.1438*	0.3153***	-0.0205	0.2266	0.2642***
	(4.708)	(1.769)	(4.732)	(-0.184)	(1.358)	(2.886)
MoM	-0.1584**	-0.1293	-0.1124	-0.1122	0.1593	0.0760
	(-2.258)	(-1.465)	(-1.464)	(-1.141)	(1.164)	(0.764)
RMW						
CMA						
Observatio	108	108	108	108	108	108
R2	0.340	0.122	0.364	0.062	0.590	0.727
Adjusted R2	0.315	0.088	0.339	0.025	0.574	0.717

*Table 10 presents the regression output from the Bloomberg Disclosure, Refinitiv ESG, and Refinitiv ESGC study for both average (AW) and value (VW) weighted portfolios using the Carhart four factor model. The alpha represents the returns from the long-short zero investment portfolios sorted with respect to the Refinitiv ESGC factors. MKTRf represents the market risk factor, SMB the small minus big factor, HML the high minus low factor, MoM the winner minus losers factor, RMW the robust minus weak factor, and the CMA the conservative minus aggressive factor. The stars behind the coefficients represent the significance level, where *=10%, **=5%, and ***=1%. The corresponding t-statistic is represented in the brackets under the coeffects.*

B: Model testing

Table 11: Whites Test and Breuch-Pegan Test for Heteroscedasticity

Output: Whites Test for Heteroscedasticity

	Bloomberg ESG		Refinitiv ESG		Refinitiv ESGC	
	AW	VW	AW	VW	AW	VW
FF 3	10.7965	4.9766	18.7766**	22.1109***	14.0213	14.1981
	(0.2899)	(0.8363)	(0.0272)	(0.0085)	(0.1216)	(0.1155)
Carhart	20.3819	8.5243	28.4387**	29.4643***	37.1061**	20.2689
	(0.1186)	(0.8603)	(0.0124)	(0.0090)	(0.0113)	(0.1219)
FF 5	24.2493	10.2362	17.0881	42.3695***	17.6042	21.0379
	(0.2317)	(0.9637)	(0.6472)	(0.0025)	(0.2254)	(0.3949)
FF 5 MoM	33.7578	10.9669	26.2706	43.0973**	43.5987**	28.5274
	(0.1732)	(0.9973)	(0.5036)	(0.0256)	(0.0227)	(0.3842)

Output: Breuch-Pegan Test for Heteroscedasticity

	Bloomberg ESG		Refinitiv ESG		Refinitiv ESGC	
	AW	VW	AW	VW	AW	VW
FF 3	4.4632	2.8051	6.6066*	0.4285	4.6692	4.6834
	(0.2156)	(0.4227)	(0.0856)	(0.9343)	(0.1977)	(0.1965)
Carhart	1.0581	2.7953	8.0974*	2.3110	4.4290	5.2038
	(0.9009)	(0.5926)	(0.0881)	(0.6788)	(0.3511)	(0.2670)
FF 5	10.9931*	1.8383	2.4477	12.8643**	5.8988	5.5396
	(0.0515)	(0.8710)	(0.7844)	(0.0247)	(0.3162)	(0.3536)
FF 5 MoM	10.0041	1.8407	5.2751	13.0295**	5.8915	6.0441
	(0.1245)	(0.9338)	(0.5090)	(0.0426)	(0.4355)	(0.4183)

Table 11 presents the results from the Whites test and the Breuch-Pegan test, both testing our models for heteroscedasticity. The null hypothesis in both tests states Homoscedasticity and the alternative hypothesis states presence of heteroscedasticity. The portfolios with significant p-values on either 1%, 5%, or 10% level, we account for heteroscedasticity by running the regressions again applying heteroscedasticity consistent standard errors.

Table 12: Durbin-Watson Test for Autocorrelation

Output: Durbin-Watson Test for Autocorrelation						
	Bloomberg ESG		Refinitiv ESG		Refinitiv ESGC	
	AW	VW	AW	VW	AW	VW
FF 3	2.1477	1.9058	2.4241	2.2506	1.9527	2.1124
Carhart	2.0543	1.8712	2.3877	2.2573	1.9601	2.0966
FF 5	2.2147	2.0580	2.4688	2.2337	2.0534	2.1547
FF 5 MoM	2.0744	2.0281	2.4137	2.2425	2.0483	2.1317

Table 12 presents results from the Durbin-Watson test, testing our models for autocorrelation. The table presents the Durbin-Watson test statistics, where a test statistic between 1.5 and 2.5 indicates no absence of autocorrelation. A test statistic closer to 0 indicates positive autocorrelation, and a test statistic above 2.5 indicates negative autocorrelation. From our results we can conclude that there is no evidence for autocorrelation in any of our models.

Table 13: Augmented Dickey-Fuller Test for Stationarity

Output: Augmented Dickey-Fuller Test for Stationarity						
	Bloomberg ESG		Refinitiv ESG		Refinitiv ESGC	
	AW	VW	AW	VW	AW	VW
FF 3	-3.6157 (0.0055)	-11.1134 (0.0000)	-3.3352 (0.0134)	-11.7841 (0.0000)	-10.0789 (0.0000)	-10.9675 (0.0000)
Carhart	-3.4958 (0.0081)	-6.1838 (0.0000)	-3.0463 (0.0308)	-11.8297 (0.0000)	-10.1151 (0.0000)	-10.8788 (0.0000)
FF 5	-11.4919 (0.0000)	-11.6928 (0.0000)	-13.2286 (0.0000)	-11.6505 (0.0000)	-10.5631 (0.0000)	-11.1372 (0.0000)
FF 5 MoM	-10.7208 (0.0000)	-11.5867 (0.0000)	-12.8867 (0.0000)	-11.7156 (0.0000)	-10.5336 (0.0000)	-11.0049 (0.0000)

Table 13 presents results from the Augmented Dickey-Fuller test, testing our models for stationarity. The table presents the test statistics and the p-values in parenthesis. The null hypothesis in the test states non-stationarity and the alternative hypothesis states stationarity. We can reject the null hypothesis for all our models based on their p-value and conclude stationarity in all models.