COPENHAGEN BUSINESS SCHOOL
HANDELSHØJSKOLEN

Busting the Myth! Do Higher ESG Rated Firms Really Bring Higher Returns?

by

Hrafn David Hrafnsson | 107484

&

Yuliya Umin | 97311

A thesis submitted in partial fulfilment for the degree of
Master of Science in Economics and Business Administration, Finance and Investments

Supervisor: Kenneth Winther
Number of pages: 115
Number of characters: 235,746
Submitted: May 15, 2018

COPENHAGEN BUSINESS SCHOOL

Abstract

This thesis aims to contribute to the debate of whether environmental, social and governance (ESG) ratings have a positive or negative effect on returns of firms. This study is conducted on ESG-rated stocks in both the US and developed Europe markets over a sample period from December of 2005 to January of 2018, in order to investigate whether higher-rated ESG firms have higher abnormal returns than lower-rated. Equally-weighted decile portfolios were formed based on the level of each ESG rating and multi-factor models were applied in order to study the relationship. A general conclusion based on high and low-rated portfolios was not clearly evident. For the sustainable investor in US stocks, the highest abnormal return was found in not only the highest-rated but also the smallest firms of the entire ESG-rated sample and this was true for all individual ratings with the only exception being Governance. However, when dividing the entire sample by ratings only, in US lower abnormal returns were found when investing in highly-rated firms as opposed to lower-rated. For the European sample, a slight trend of highly-rated firms having more significantly positive abnormal returns was found. Thus, for European stocks, it was generally found that sustainable investors can earn positive abnormal returns while the same trend was not found in the US. Overall, a higher Governance rating appeared to be the most detrimental to returns in US. On the other hand, in Europe, higher abnormal returns were observed for higher ESG Combined ratings. Overall, this paper hopes to add value to both the sustainable and conventional investor in order to pinpoint the differences in returns between ESG-rated stocks in the hopes of helping them make better investing decisions and optimise their abnormal returns.

Contents

Chapter 1

Introduction

1.1 Introduction & Problem Indication

Globally, within the last decade there has been exponential growth in the socially responsible investment (SRI) market, much of it which is contributed to the establishment and rapid growth of environmental, social and governance (ESG) ratings for firms. In the 1990's, less than 20 firms reported ESG data such as sustainability or integrated reports and by 2016, this figure increased to nearly 9,000 [Amel-Zadeh and Serafeim, 2017]. There is increasing pressure these days on firms to be more sustainable from all stakeholders due to the major effect they have on the economy and the world. Moreover, the ones that do not can get shunned by some investors altogether. Increasingly, perhaps more today than ever, the issues of good corporate governance, environmental impact and social aspects such as treatment of employees are seen in the media. This can adversely affect firms and thus their stock prices in many ways and by incorporating ESG issues, investors may not only reduce their risk related to this but also increase their returns by being proactive in their investment decisions by selecting firms that have a positive impact on society. Thus, the central question surrounding the field these days is whether this is true and whether ESG investing adds or detracts value for investors.

Although ESG has substantially increased the market for SRI, there is still much scepticism surrounding the field and what effect the ratings have on the investor's returns. Due to the difficulty in quantifying many of the ESG issues and the vast amount of data providers which all differ in their methodologies, there is no general consensus on the effect that ESG ratings have on the firm's return and whether being proactive about sustainability issues pays off. Furthermore, the lack of standardisation makes a concrete answer hard to achieve. Many studies that study the effects of ESG mention that a major limitation is that their results are only relevant to the specific data provider that was utilised. Thus, the answer to the central question is much more complicated than a simple yes or no. It depends not only on the data providers but also the country the firms operate in and the industry they are in, since some ESG issues are much more important to specific industries and some industries get shunned by investors entirely. Moreover, the effect on returns can be studied by utilising many different methods, making the conclusions found even more distorted and varied between each study.

1.2 Problem Formulation

This thesis seeks to explore further a subset of the classic question of the effect that ESG ratings have on investor's returns. Unlike much of previous research which has focused on the question of whether ESG investing is superior to traditional investing, this paper seeks to examine the ESG-rated universe from a sustainable investor perspective in order to see whether the level of the ESG rating matters for the highest abnormal return. By conducting an empirical analysis of all ESG-rated stocks for both the US and the developed Europe markets between December, 2005 until January, 2018, this paper attempts to add to the debate of whether highly-rated firms outperform the lower-rated and whether there are cross-regional differences in this between the US and Europe. The reasoning behind exploring the US and Europe sample was explained well by Lean et al. [2015]. The first is that the SRI concept originated in Europe and US. As an example, the first countries that required pension funds to report on ethical and environmental concerns of their investments were Italy, Sweden, United Kingdom, and Belgium. The second explanation is related to the large market share that the two regions have. For instance, North America and Europe together have 96% of the global SRI funds and the same goes for ESG investing, where the two regions have the biggest market share of anywhere else in the world.

This is a particularly important question not only for the sustainable investor but also the conventional one. First, if higher-rated firms outperform the lower-rated, then the sustainable investor who limits their universe to firms that are highly rated on ESG wants to know if this hurts or benefits his overall portfolio. Additionally, where on the ESG-rating scale he invests is an important issue. Perhaps the middle-rated firms have the highest returns and thus they are quite sustainable and so then the sustainable investor should put his focus on investing in those firms and not the absolute top-rated. On the other side, perhaps the conventional investor would also like to know what ESG-rated firms bring the highest returns. If it is found that the lowest-rated firms have the highest abnormal returns of any other rating and the conventional investor does not care about the sustainability aspect of his portfolio, then he may choose to invest in those firms. He can further examine if these rated firms bring higher abnormal returns than his conventional portfolio and if so, focus on investing in them.

1.3 Research Question

The impact that the level of the ESG rating, including the individual components, has on returns will be the main focus of this paper as it is important for the sustainable investor in order to see whether where on the rating scale he invests matters for his optimal return. Thus, the Research Question of this thesis is:

Do firms with higher ESG ratings produce higher abnormal returns than firms with low ratings?

In order to provide meaningful answers to this question, the analysis performed in this paper will involve the empirical testing of portfolios of firms which vary by the level of their ESG score. Several multi-factor models are applied in order to examine the difference of the abnormal return of these portfolios for both the US and developed Europe from December, 2005 until January, 2018. Additionally, three hypotheses will be tested that will aid in answering the central question:

Hypothesis 1: A portfolio consisting of higher-rated ESG stocks will outperform a portfolio consisting of the lower-rated ESG stocks.

Hypothesis 2: A portfolio consisting of stocks with the smallest average market capitalisation and the highest-ratings will outperform a portfolio of the smallest and lowest-rated stocks for both regions.

Hypothesis 3: High-rated (low-rated) ESG stocks from the US sample will outperform high-rated (low-rated) ESG stocks from the European sample.

The overall goal is to conduct an unbiased and fine-grained analysis that will either confirm or reject the notion that ESG ratings can be utilised to gain significant abnormal returns. The purpose is thus to contribute valuable findings to the existing ESG literature by putting the new and improved Thomson Reuters ESG data through a series of tests. As such, this paper will not develop or implement any specific practical investment strategy, but can rather be defined as an ambitious effort to provide practising investors with information about what should be considered when formulating an ESG investment strategy.

1.4 Delimitations

This paper utilises Thomson Reuters ESG data and thus it assumes that these ratings reflect the general investor's view on sustainability with regards to ESG. If the investor does not view these rating's as representative of sustainability, then the analysis presented will be of no value. Thus, the results of this paper are only relevant and applicable to the Thomson Reuters ESG scores and are not comparable with other providers of data. It should be noted that this is a limitation to most studies within ESG due to a variety of data providers available. A comparative study of a number of providers would improve the analysis but is too extensive and out of scope for this paper and it was not possible to retrieve data from other providers, which all have payment subscriptions.

It should also be noted that the empirical strategy in this paper involves testing the performances of ESG-rated equities in a limited and defined geographical investment universe which includes only the US and developed Europe. As such, any conclusions, implications or suggestions regarding the findings of this paper are not applicable to any other asset classes or other countries. Additionally, the applied empirical strategy is designed to examine any potential relationship between the financial performance of ESG ratings as reported by Thomson Reuters within a defined time-frame. The analysis does not involve subdividing the analysis into in- and out-of-sample testing for the sake of robustness. The authors recognise the importance of in- and out-of-sample testing, as highlighted by Hagin [1990]. However, any meaningful out-of-sample testing is beyond the scope of this paper, and instead the decision was made to focus on achieving a more fine-grained analysis of the differences between high- and low-rated ESG stocks which is optimal for the Research Question analysed.

Furthermore, transaction costs and tax considerations are not taken into account in the analysis, although the implications of ignoring them will be addressed. In a more practical setting, these costs will affect the overall realised abnormal return, however, they are not critical in answering the Research Question. Additionally, these costs will vary with each type of investor, their trading volume and the country they are located and are therefore hard to generalise precisely. Additionally, the purpose of this paper is not to simulate or implement any specific investment strategy based on ESG ratings. Instead, this paper's purpose is to dissect the ESG-rated stocks in order to gain important insight about the underlying drivers, and reveal any possible evidence of financial performance differences between high and low-rated ESG stocks. The overall goal is thus focused on providing investors with the proper tools and data-backed information that they in turn can use to construct an investment strategy that suits their agenda.

Finally, in this paper, ESG investing, socially responsible investing (SRI), and corporate social responsibility (CSR) are used interchangeably to describe overall sustainable investing. Additionally, when mentioning the lowest/bottom-rated portfolio this refers to the 1^{st} decile portfolio and the top/highest-rated portfolio refers to the 10^{th} decile portfolio, which will be described in more detail in the forthcoming sections.

1.5 Research Structure

First, a Literature Review will give an overview of the SRI and ESG investing market, the various stakeholders, theories and review the existing literature on the relationship between ESG ratings and returns. The following Theory chapter provides a brief overview of the theories that will be discussed and used in this paper. The Data chapter will discuss data that is used in the analysis, including data description, the data collection process and data issues associated with the data sample used in this paper. Next, the Methodology chapter will describe the methods utilised in the analysis including an explanation of the regression methods, the application of the multi-factor models and performance measures. Then, the Analysis of Results chapter will present the findings of this paper and explain these with various graphs and tables and the analysis is described for each region separately. The Discussion of Analysis chapter will then further discuss the implications and give possible explanations for the various findings of this study. Lastly, a brief Conclusion chapter will summarise the overall paper and findings.

Chapter 2

Literature and Background Review

2.1 Background & Concept

2.1.1 History of Socially Responsible Investing

The concept of Socially Responsible Investing (SRI) has barely been around for a hundred years. However, the roots of SRI go far back and ethical investing was first found to be rooted in religious values centuries ago in Jewish, Christian and Islamic traditions [Renneboog et al., 2008]. What is now known as negative screening, avoiding sin stocks, can be traced back to the 17^{th} century when Quakers avoided investments in firms who operated in tobacco, alcohol and weapons industries and which is generally known as negative screening today [Dorfleitner et al., 2016]. The first SRI fund ever, the Pioneer Fund, began in 1928 and centred around negative screening of social values [Kumar et al., 2017]. However, the modern SRI industry as it is known today is fairly new and investors view on social responsibility only started gaining traction during the 1980's [Aw et al., 2017]. The development of the SRI market had increased substantially after 1981, with the founding of the U.S. Forum for Sustainable and Responsible Investments [Dorfleitner et al., 2016].

To put it simply, SRI is an investment process that incorporates not only risk and reward into the investment process but also environmental, social and ethical considerations. It is also frequently interchangeable with ethical or sustainable investing, among many other terms [Renneboog et al., 2008]. To find one clear-cut definition of SRI is hard to do and the actual definition varies depending on the source. Generally, SRI uses social screens in order to select or exclude various investments based on their social aspects. In the last ten years, there has been more of a trend toward a consensus on what being socially responsible means within the SRI context. There are three main ways in which a firm can be socially responsible: operate sustainably and have a positive impact on the environment, offer benefits to the society with its practices and it can incorporate and implement better corporate governance practices. Altogether, these three definitions equate to environmental, social and governance (ESG) factors that are becoming the standard in the SRI industry and a widely accepted method is screening investments based on their ESG scores [Aw et al., 2017].

In more recent times, a distinction has been made between the "values-driven" and "profit-seeking" SRI approach. The values-driven approach, essentially the way SRI started, integrates both social and personal values into the investment approach and implies that investors are not afraid to take losses in exchange for aligning their investment with their values. Many institutional investors committed to ESG principles have, for example, become signatories of the Principles for Responsible Investments (PRI), an independent non-profit organisation that encourages investors to use responsible investment to enhance returns and better manage risks which was launched in 2006. Within ten years, by 2016, about 1,400 financial institutions with over $60 trillion assets in management were PRI signatories [Amel-Zadeh and Serafeim, 2017]. On the other hand, the more recent "profit-seeking" SRI approach integrates SRI information with the sole goal of higher returns while essentially disregarding the values behind the process, so long as it achieves a higher return that will give them a comparative advantage. With the growing rise of SRI, ESG investing has become an integral and growing niche in the investment market. Among the stakeholders of SRI today are governments, pension funds, foundations, universities and simply investors and asset managers [Derwall et al., 2011].

When looking at worldwide SRI assets, the number one region in the world is Europe, followed by the US and then Canada. However, due to recent stricter definitions of SRI in Europe, it is the only market world-wide that did not experience market share growth between 2014 and 2016. Somewhat surprisingly, for the two years between 2014 to 2016, the fastest growing SRI market had been Japan due to high interest and activity in sustainability and also improved reporting standards [Alliance, 2016].

2.1.2 Development of ESG Investing

The institutionalisation and investor interest in ESG data did not fully pick up until the 2000's. A big step forward in 2006 was the UN Principles for Responsible Investing (PRI) and their commitment to including ESG issues in the investment process. There has also been much progress in the field in the last decade to encourage standardisation and adoption of ESG data. Some of the initiatives taken with the aim of setting standards for voluntary ESG reporting include the United Nations Global Compact (UNGC), Global Reporting Initiative (GRI), International Standards Organisation (ISO) and Sustainability Accounting Standards Board (SASB) [Bender et al., 2017]. Nevertheless, since much of the data related to ESG is voluntarily provided by companies, standardisation is difficult to achieve. For example, in 2011 less than 20% of firms listed on the S&P 500 published a sustainability or CSR report. However, that figure has been greatly increasing and was found to be 72% by 2013 [Roselle, 2016].

The worldwide market for ESG ratings has increased substantially in recent years. For the two-year period from 2014 to 2016, SRI investments had increased by 25%. Additionally, as of 2016, more than $22 trillion of assets were being managed under SRI strategies and globally, ESG investing accounted for around 26% of all assets being managed [Alliance, 2016]. A substantial number for a niche market that barely existed 20 years ago. This figure is expected to keep rising in the future, which shows just how increasingly important the topic is becoming. For many investors, ESG data has become more of a risk management tool than a value strategy since it has been shown to provide lower volatility in firms [Kuzmina and Lindemane, 2017]. Back in the 1990's, less than 20 firms actually disclosed ESG data, while this figure has increased to around 9,000 by 2016 [Amel-Zadeh and Serafeim, 2017]. Although ESG has clearly had great growth in the past decade, it is still a fairly new concept that has a lot of development and work to do in order to be used and applied consistently all over the world and there are still many firms who are not assigned any rating and thus put no effort into ESG initiatives.

ESG is also becoming ever-more important nowadays due to the increased awareness of investors to the issues. Bad corporate governance of firms, take for example Facebook, is consistently seen in the media. Other issues such as global warming, the Kyoto Protocol, child labour and the likes have made ordinary investors think twice about investing in whatever gives them the highest returns

[Renneboog et al., 2008]. In fact, in a survey by the Forum for Sustainable and responsible Investment, it was found that 80% of fund managers incorporate ESG issues due to client pressure [Przychodzen et al., 2016]. Since firms are such an integral part of the world, governments have also started to increasingly set standards for countries to comply with ESG-related issues which has advanced the industry substantially.

The three main participants of the ESG and sustainable investing strategies overall are essentially the public firms, portfolio managers and ESG data providers. Perhaps why there is still scepticism about sustainable investing is the way it all started. Back in the 1990's, the sustainable investment universe seemed to constantly underperform the traditional one and as many pointed out, went completely against the Modern Portfolio Theory (MPT), as introduced by Markowitz in 1952. Many investors did not believe that SRI and maximised risk-adjusted returns could go hand-in-hand together and that only one of them had to be chosen. Only in recent times have studies constantly tried to disprove this view, but the scepticism of some remains nevertheless Roselle [2016].

Although ESG investing has come far since its inception, there are still drawbacks and criticism related to the way ESG scores are calculated and the various data providers. There is generally not a consensus on scores and they can vary widely depending on the data provider. As of 2016, there were more than 125 various organisations providing ESG data [Bender et al., 2017]. Some of the biggest and well known ESG data providers are Thomson Reuters, MSCI, Bloomberg, Sustainalytics and Dow Jones, to name a few [Huber and Comstock, 2017]. With such a variety of providers, each having their own proprietary research methods and methodologies, the question arises of which one is the best and how standardised they are across providers.

2.1.3 ESG Investment Strategies

There are many ways of integrating ESG data into the investment process and they all differ by their methodologies. One of those ways is simply ESG integration, which generally means integrating ESG data alongside typical financial analysis when selecting securities. Implementing screening strategies within the ESG universe is another extremely popular way of selecting stocks and it is one of the most widely used strategies in past studies [Kumar et al., 2017].

Screening strategies have been used extensively within SRI and ESG investing in order to form portfolios that fit the views of the investor. Some of the most popular strategies are positive screening, negative screening and best-in-class, which all differ in the way they are implemented. First of all, positive screening rates all firms based on a set of criteria and then chooses firms with the highest rating in each criteria [Kempf and Osthoff, 2007]. Generally, it means selecting stocks with superior CSR standards or, equivalently, high ESG ratings. For the most typical positive screens, investors select firms with high environmental standards (waste reduction, renewable energies, biotechnology), superior corporate governance (auditor independence, executive compensation, voting rights) and firms that have high social company performance (strong labour relations, employment diversity and community involvement) [Auer and Schuhmacher, 2016]. Unlike other types of screens, positive screening is very broad and it may be applied to any firm in the investment universe since every company has a general responsibility to act within some guidelines. Positive screening is generally more time consuming and difficult to do than other screens since it requires the investor to choose selection criteria, determine the right measures for that criteria and create benchmarks for what are the "best" or "good" practices for a firm [Cahan et al., 2017].

Next, best-in-class screening strategies involve selecting or weighing the leading or best performing securities based on their ESG criteria. The selection can be within a certain universe, industry, category or class. One example of a best-in-class strategy would be a mutual fund with an ESG fund that restricts their investment universe to firms among the top 25% within their respective industry based on their ESG rating [Scholtens, 2014]. Kempf and Osthoff [2007] did a study on ratings and found that this screening method generally led to the highest alpha and worked optimally when investors implemented a combination of SRI screens simultaneously and restricted to picking stocks with extreme ratings.

Lastly, the oldest, most basic and most common screening strategy is based on negative screening, which essentially excludes stocks that don't live up to certain environmental, social or ethical standards [Renneboog et al., 2008]. In the context of SRI and ESG, this generally means excluding stocks in sin industries. Classic sin stocks can be classified as those that operate in alcohol, gambling, tobacco and weapons industries, among other controversial areas [Blitz and Fabozzi, 2017]. However, it can also be extended further to exclude stocks with poor workplace conditions, animal testing, violation of human rights and irresponsible foreign operations. It is one of the most popular types of screens because it is fairly simple to do and due to the fact that investors shun certain industries that are not aligned with their ethical values. In fact, more than 80% of SRI funds implement negative screening in their portfolio construction process [Lee et al., 2010].

Besides the popular screening strategies, more and more unique ESG strategies are emerging in recent times. For example, the ESG momentum strategy relies on the assumption that future stock performance is related to the change in the ESG score. The strategy is fairly simple and instead of picking highly-rated ESG stocks, it over-weighs firms that have improved their rating during the last year. This means even the low rated stocks, so long as they have improved their score, are included. [Nagy et al., 2015] evaluated this strategy over a seven year period between 2007 and 2015 with MSCI's ESG ratings and compared it to the traditional counterpart, the MSCI World Index. They found that the strategy outperformed the benchmark by 2.2% annually.

2.2 Effect of ESG on Returns

Historically, the link between ESG performance and returns has been debated. Back in 1970, Milton Friedman wrote an influential article that has been referenced extensively within the SRI context for decades afterwards. The article, titled "The Social Responsibility of Business is to Increase its Profits," essentially stated that the sole responsibility of businesses is to maximise returns and make as much money as possible. Friedman did not believe social responsibility attributed to this responsibility in any way and quite the contrary, took away resources from profit maximisation. Although a rather outdated view on the topic, it was nevertheless extremely influential and led to many critics of SRI and papers in the years to come to agree with his view. Additionally, according to Modern Portfolio Theory [Markowitz, 1952], a restricted investment universe based on screening strategies related to SRI or ESG should essentially result in a lower risk-adjusted return considering that this strategy is a subset of the entire investment universe [Schröder, 2014]. Another major reason why it is difficult to find a consistent link is the difficulty in analysing it. This is mainly due to the multi-dimensionality of the ESG concept itself and the difficulty in quantifying it as a concrete number [Mnescu, 2011]. Above reasoning, among others, has resulted in varied opinions and studies with regards to what the relationship is between returns and ESG. Some claim there is a positive relationship and that investing in ESG-rated companies gives higher returns. On the other hand, some studies say you must sacrifice returns in order to invest ethically. There are also numerous studies that have found no relationship so that you are no better nor worse off while investing ethically rather than unethically. Overall, the literature on performance of SRI and ESG stocks has been mixed and varied and a general consensus has not been reached as of the time of this paper. However, within the last decade, the view is shifting more and more towards a positive or neutral relationship.

The standard in literature of examining the relationship between SRI and financial performance is to examine ESG-rated stocks, sustainability indices or SRI funds and whether they lead to higher abnormal returns as opposed to traditional investments. However, the fund literature cannot control entirely for the differences in management skills and therefore might not give an entirely accurate representation of the relationship [Kurtz and Dibartolomeo, 2011]. Another popular way of studying the relationship is to examine the differences in returns based on low and high-rated ESG stocks. One would expect high-rated stocks to outperform low-rated ones, however, this is not necessarily always the case. The standard in literature to test the relationship usually involves static econometric techniques, such as the Sharpe ratio, or estimating the traditional market models, such as the CAPM or the multi-factor models like Fama-French three-factor and Carhart's four-factor models [Ortas et al., 2014]. Due to the overall varied nature of literature on the topic, it's important to explore each view and the effect of SRI, including ESG, on returns and the reasoning behind it.

2.2.1 Positive Relationship

Although there have been mixed results of past studies on the effect that ESG ratings have on returns, more recently there has been a growing amount of literature that has found a positive link between the two [Humphrey et al., 2012]. This relates to the "errors-in-expectations hypothesis" which predicts that socially responsible stocks lead to higher risk-adjusted returns because the market doesn't immediately recognise the positive impact that good sustainability practices have on the future expected cash flows of firms [Derwall et al., 2011]. There have been many studies in the past that have in fact showed one does not need to give up returns in order to invest ethically. For instance, Kempf and Osthoff [2007] examined whether a trading strategy in SRI-rated stocks leads to abnormal performance in the

US within a 12 year time period (1992-2004) for the S&P 500 and DS 400 indices. They obtained ratings from the KLD database and formed value-weighted portfolios based on the high and low-rated stocks and applied Carhart's four-factor model for performance measurement, which is a common methodology in past studies. Additionally, they applied screening strategies in order to see which one lead to the highest return. They found that going long in highly rated stocks while shorting the low rated ones actually led to a positive four-factor alpha of 8.7% per year. Furthermore, they found that the highest alpha was achieved by the positive or best-in-class screening but not with the negative screening.

Khan et al. [2015] examined 2,307 firms with ratings from the KLD database with a sample period of 1991-2012. They took a different approach to the analysis and the first of its kind by creating firm-specific sustainability ratings by classifying each sustainability issue as being either material or immaterial to the particular industry, using the criteria measures of the Sustainability Accounting Standards Board (SASB). They formed both equal-weighted and value-weighted portfolios based on the level of their ESG performance and then applied Carhart's four-factor as well as Fama-French's three-factor and five-factor model for regressions and also ran firm-level panel regressions. They found that firms with high ratings on material sustainability issues significantly outperformed those with low ratings. However, for immaterial issues, higher-rated firms did not outperform lower-rated ones. Their findings are important because it shows the need to look into the individual components of each rating and thus aggregate scores might not show a clear representation of the link between ratings and returns.

Another study by Ashwin Kumar et al. [2016] analysed companies based on their ESG score and by industry with a small time frame of two years. They found that firms with an ESG ratings have lower volatility than firms that do not in the particular industry, that the effect of ESG factors is dependant on the type of industry and overall, ESG-rated firms have higher returns. Furthermore, across all industries, they found that, on average, return on equity was 6.12% higher for firms with an ESG rating. Interestingly enough, ESG factors had a negative impact in the automobile, banking, insurance and durables industry. Their paper also shows the importance of studying the relationship within industries since each one is affected by ESG factors in different ways and thus might not be best idea to aggregate the scores together.

Verheyden et al. [2016] investigated the effect of ESG screening, including ESG momentum and best-in-class, on portfolio performance. They wanted to investigate whether incorporation of ESG information is beneficial to the regular investor who has no interest in sustainability. Thus, they compared performance of a global unscreened universe to that of two portfolios which had been screened by ESG and furthermore by cut-off thresholds in order to see if there was an opportunity to be exploited or if investors were at a disadvantage. For three out of four universes, they found that ESG screening actually improved risk-adjusted returns while volatility, drawdowns and the conditional value at risk (CVaR) was lowered when compared with the unscreened universe. Additionally, they found an overall positive impact of ESG screening which improved, on average, annual performance by 0.16% when compared with the unscreened universe.

A recent paper by Dorfleitner et al. [2018] looked at the long-term effects of CSR and financial return by decomposing the ESG scores into individual pillars for US and Canadian stocks for the 2002-2014 sample period. They obtained ratings from the Asset4 database and other relevant variables from Datastream, which in the end included a total of 1,308 firms. They further controlled for a long list of variables that could affect stock price returns, including the debt ratio, industry, total assets and book-to-market ratio. For the environmental and social factors, they found positive mid- and long-term effect of an abnormal return up to 3.8% with respect to a one standard deviation change in the ESG score. Further decomposing the factors, they found that emission and resource reduction, society and workforce were particularly profitable investment areas. Furthermore, they decomposed the ESG score into three pillars and then created quintiles based on the level of the ESG score. They found that the weak portfolio, the one with the lowest ratings, had negative abnormal returns while the strong portfolio, with the highest ratings, had positive abnormal returns.

Perhaps one of the most significant papers on the topic was a meta-analysis done by Friede et al. [2015], who aggregated over 2,200 empirical studies on the topic of ESG and financial performance. By including only academic studies with quantitative summaries, they investigated and generalised the findings of years of papers. Somewhat surprisingly, around 90% of all studies found a link between ESG and financial performance that was not negative. Furthermore, a large majority was found to have a positive link while a small percentage had a negative link. They also found that ESG outperformance was particularly found in emerging markets, North America and in non-equity asset classes.

2.2.2 Neutral Relationship

Another viewpoint is that ESG has no influence on returns and you are no better nor worse off investing in ESG as you are with traditional investments. Proponents of this view state that because CSR is not priced, then the factor is not a proxy for risk and thus does not affect the return of a stock Auer and Schuhmacher [2016]. Halbritter and Dorfleitner [2015] challenged the positive link between ESG and financial performance by a comprehensive study of the US market across various ESG data providers (Asset4, Bloomberg and KLD) from 1991 to 2012. By using the aggregated ESG score and also the individual pillars and applying Carhart's four-factor model and the cross-sectional Fama-Macbeth, they actually found no significant relationship between high and low-rated ESG stocks and therefore, investors should not expect abnormal returns by trading strategies based on the level of the ESG rating. Furthermore, they found that the impact on abnormal returns was largely dependent on the specific rating provider, the sub-period and the company sample. Their findings showed the difficulty in generalising studies to come to one comprehensive conclusion and that each one is not directly comparable to another.

The above findings were further confirmed by Revelli and Viviani [2015] who did a meta-analysis of 190 international studies in the field in order to test the relationship between SRI and financial performance and whether it is more profitable to use ESG in the investment process. Their results showed that globally, including CSR in stock market portfolios does not have any effect on performance when compared to conventional investments. They also determined that findings of a link between CSR and returns in prior literature is largely dependent on the specific study and the thematic approach, data comparison method and investment horizon used. Thus, sometimes, depending on the approach, one might see a link but overall there was no meaningful association between a positive or negative relationship between CSR and return.

Humphrey et al. [2012] investigated the link between ESG ratings and return and risk based on a United Kingdom (UK) sample from the SAM database from 2002 to 2010. By investigating Jensen's alpha and applying Carhart's model, they found no significant difference in risk-adjusted performance of low and high-rated ESG portfolios. They found that highly-ranked firms were much larger in size and seemed to have lower betas, but this was not upheld in robustness tests. Thus, investors in UK could incorporate ESG without worrying about a loss in returns and at the same time, should not expect a higher return.

Another paper by Humphrey and Tan [2013] examined the effect of screening on a portfolio's risk and overall performance by utilising the KLD database. They attempted to replicate typical portfolio holdings of US equity mutual funds, value-weighted, in order to conduct their study and compared the screened universe to the unscreened one in order to investigate the differences. A number of performance measures, such as Jensen's and Carhart's alpha, were investigated in order to study the relationship. There was no evidence found that the screening affected performance or risk of the portfolio and thus a neutral relationship was observed. This essentially went against the "shunned stock hypothesis" as will be described later, since negative screening was found to neither hurt nor improve performance of the portfolios.

Auer and Schuhmacher [2016] evaluated a large sample of firms from 2004 to 2012 from the Asia-Pacific, US and Europe regions with ESG scores retrieved from Sustainalytics. To evaluate whether the level of the ESG rating outperformed a passive benchmark, they created a total of 60 portfolios that varied by region, sector, and ESG criteria. They further applied screening methods and cut-off rates for all portfolios and tested performance with the Sharpe ratio. In the screened portfolios, highly rated portfolios showed higher Sharpe ratios than benchmarks in only 15 out of 60 portfolios, while low rated portfolios outperformed the benchmark in 34 out of the total 60 portfolios. Overall, they concluded that actively selecting stocks based on their ESG rating did not provide any superior risk-adjusted performance when compared with the passive benchmark portfolio. Additionally, in the US and Asia-Pacific the performance of ESG was similar to the broad market and thus ESG neither decreased nor increased the financial performance of portfolios.

2.2.3 Negative Relationship

The last viewpoint on financial performance and ESG is that they have a negative relationship and an investor gives up profits in order to invest sustainably. The negative relationship between returns and ethical investing can be linked to the "trade-off theory" which claims that socially responsible activities take away profits and resources from a firm, therefore placing them at a disadvantage from the firms that do not partake in such activities [Auer and Schuhmacher, 2016]. One of the earliest proponents of this view was Milton Friedman, as mentioned previously, who shaped the view that SRI hurts profits and for many years to come this was the general sentiment surrounding the field, with many studies seemingly agreeing with Friedman.

A recent study by Aw et al. [2017] examined all global publicly traded companies with a market capitalisation above $1 billion from August of 2009 to July of 2016, thus covering a significantly large universe and far larger than most previous studies which only focused on one or two regions. To evaluate whether ESG factors were a useful estimator of expected returns, they created quintiles based on the level of ESG score and it's sub-components and then calculated various performance measurement statistics, such as the information coefficient and t-statistic, in order to evaluate the ESG factor. Furthermore, they compared their results with an out-of-sample universe. Somewhat surprisingly, they found that the highest rated stocks (top quintile) under-performed the out of sample universe and this was statistically significant at the 10% level. They also found that the Governance component overperformed while the environmental and social underperformed, however, those were not significant. Overall, they found that filtering investment universe by ESG factors hurts investor's value while also finding a negative impact on returns when excluding low-rated stocks while integrating ESG into an existing portfolio construction.

Auer and Schuhmacher [2016] studied the US, Europe and Asia-Pacific regions and what effect the ESG rating had on returns. They used a variety of cut-off rates, ESG criteria and sectors to evaluate the relationship. In the US, they found that low-rated stocks performed better than high-rated stocks when using a 25% cut-off for the environmental rating and also when using the 10% cut-off for the social rating.

Additionally, Hirigoyen and Poulain-Rehm [2015] examined the relationship between financial performance and CSR. They used a sample of 329 firms in the US, Europe and Asia-Pacific for a one year sample between 2009 and 2010 and applied linear regressions and Granger causality tests in order to test the relationship. They found a negative relationship between CSR and financial performance. Furthermore, they found that especially two aspects of social responsibility, human resources and environment, had a negative impact on financial performance.

2.3 Additional Implications of ESG Investing

Other than only looking at the return aspect, there has been a growing body of literature that has examined the effect ESG investing has on other aspects of a firm. There has been evidence that higher ESG firms have lower volatility, lower cost of capital and thus cheaper equity and debt financing. There is also much debate about the effect ESG investing has on diversification and whether it increases or decreases for the overall portfolio. Lastly, a subset of ESG investing that has been studied is the "shunned stock hypothesis" and the effect that negative screening and excluding sin stocks has on overall performance of ESG portfolios.

2.3.1 Impact of ESG on Risk

Risk in general, including ESG risk, can be systematic risk that affects the market as a whole and is undiversifiable and unsystematic risk, which is diversifiable risk that is specific to that particular firm [Harper Ho, 2010]. Many studies on ESG overlook the risk aspect and simply assume that a high ESG score leads to less risk [Ashwin Kumar et al., 2016]. However, it is an important issue since many investors don't only want the highest return but also the lowest risk in their portfolio formation process. Although often overlooked, there is still a number of papers that have examined the impact that the ESG rating has on risk. Ashwin Kumar et al. [2016] assessed 157 firms listed on Dow Jones Sustainability Index and 809 unlisted ones over a period of 2 years. The ESG-rated stocks used were in the top score category, otherwise known as "best-in-class," equally weighted and then further decomposed into 12 industries. They found that the ESG factors had lower volatility and thus lower risk while also bringing higher risk-adjusted returns when compared to the conventional stocks. Their findings were also contradictory to the traditional view that lower risk brings lower returns.

Sassen et al. [2016] used the Thomson Reuters Asset4 database in order to evaluate the impact of ESG factors on market-based firm risk solely in Europe. For the 2002-2014 period, they used all small, mid and large-cap companies which came from 18 Western European countries. Estimation of firm risk was done by using total risk, systematic risk and idiosyncratic risks while also introducing control variables. Fixed effects regressions were then applied in order to see the relationship between the risk measures as a function of the ESG score. They found that a higher ESG rating lowered both the total and idiosyncratic risk. Additionally, by decomposing the aggregate ESG score they found a significant negative relationship between the social pillar and all risk measures (total, systematic and idiosyncratic).

Another interesting study by Ashwin Kumar et al. [2016] analysed companies based on their ESG score and by industry with a small time frame of two years. They found that firms with ESG ratings had lower volatility than firms without a rating in the particular industry, that the effect of ESG factors was dependant on the type of industry and overall, ESG-rated firms had higher returns. Rather astonishingly, they found that firms in the energy industry who had an ESG rating had 50.75% lower volatility than ones that do not.

Finally, a recent article by Utz [2017] was one of the first to study the risk benefits that come with CSR in international stock markets. He used a broad sample of all public companies in the US, Europe, Japan and Asia-Pacific regions from the Asset4 database. Applying OLS regressions on three specific models, he measured the idiosyncratic and crash risk. Overall, he found that a higher rating was a predictor for low idiosyncratic risk in all regions. However, for risk of crashes, high CSR in US and Europe significantly lowered stock price crash risk while in the Asia-Pacific region, higher CSR increased the risk of crashing. In summary, the studies mentioned have particularly important implications for investors since they are constantly looking for ways to reduce their risk without giving up profits and it seems the integration of ESG can do just that for them.

2.3.2 Impact on Cost of Capital

Cost of capital, which includes both the cost of equity and cost of debt, is directly linked to the firm's risk level and profitability and so it is crucial for a firm's long-term survival and return [Clark et al., 2015]. Therefore, it's interesting to analyse and explore how ESG and sustainability impact the cost of capital of firms. Clark et al. [2015] investigated 29 academic studies of high quality and the effect CSR has on cost of capital. They found that in 90% of the studies, high CSR led to a reduction in cost of capital, which shows a fairly important link between the two.

A study by El Ghoul et al. [2014] examined an international sample of 30 countries and the effects of environmental responsibility on cost of capital. Utilising control variables in order to control for industry, country and firm-level variables, they found that cost of equity was consistently lower for firms with higher level of corporate environmental responsibility. Although this study was conducted on just one aspect of the ESG score, it showed that even the environmental score alone could potentially lead to a lower cost of capital.

Ghoul et al. [2011] was one of the first studies to examine the effect of CSR on ex ante cost of equity capital. They used a large sample of 12,915 US firms over the 1992 to 2007 period and retrieved ratings from the KLD database. By controlling for industry and year fixed effects as well as firm-specific variables and running multivariate regressions, they found that higher-rated firms had significantly lower cost of equity capital. In particular, they found that environmental policies, product strategies and employee relations affected directly the lowering of cost of capital while other aspects, such as community relations, diversity and human rights, were unrelated to cost of equity. In summary, they found that a higher CSR rating strengthened firm value due to a lower costs of equity capital for the firm.

Cost of debt is another important part of cost of capital, since lower debt financing is considered an advantage for firms. According to Clark et al. [2015], environmental externalities put risks on firms which may directly impact the cost of financing and particularly the firm's cost of debt. These risks may be reputational or financial, among others. There have been studies that showed that by using ESG policies to reduce the aforementioned risks, firm's can achieve lower cost of debt. Another paper by Bauer and Hann [2010] investigated the effect of environmental management on credit risk of 582 US firms from 1995 to 2006. They found that proactive environmental practices were linked to a lower cost of debt while environmental concerns were linked to a higher cost of debt and lower credit ratings. This directly relates to the environmental component of ESG and thus an analysis of the individual pillar alone could potentially help investors come to a conclusion about a firm's cost of debt and further valuation.

2.3.3 Impact on Diversification

For a long time, the topic of loss of diversification within ESG investing has been debated. At the start of SRI development, around 30 years ago, Rudd argued that ESG integration worsens portfolio diversification because the ESG universe is a subset of the entire investment universe [Hoepner, 2010]. This is also generally the conclusion one would come to if they look at Modern Portfolio Theory, which implies that it is impossible for an ESG-screened portfolio to be more diversified than a conventional one [Verheyden et al., 2016]. Contrary to this view, Verheyden et al. [2016] explored portfolios with ESG screens applied and then compared them to the unscreened universe. Furthermore, they looked at the effect the screening had on overall portfolio diversification. For three out of four universes, the excess risk-adjusted returns more than offset the amount of risk introduced by the screening when compared with the unscreened universe. Thus, they found that the ESG screening had an overall net positive impact on diversification since the amount of specific risk taken was more than supported by the large alpha value. Thus, on average, they found that the introduction of ESG did not lead to any big diversification losses.

Similarly, Hoepner [2010] studied the effects of diversification within ESG investing and the effect it had on the overall portfolio. He looked at the three drivers of portfolio diversification: average specific risk of stocks, the number of stocks and the correlation of stocks. He found that although integrating ESG worsens diversification through the amount and correlation of stocks, it also increases diversification due to a lower average specific risk. Hence, ESG integration can be a great risk management tool. Additionally, he found that the best-in-class strategy had a particularly high potential of improving diversification as opposed to the other screening strategies.

Bello [2005] took a different approach and compared SRI mutual funds to conventional funds in order to examine the differences in portfolio diversification from 1994 to 2001. He calculated Jensen's alpha, Sharpe information ratio and the excess standard deviation adjusted return in order to compare the differences. He measured portfolio diversification by residual variance and found that the portfolio diversification and the investment performance of the SRI funds was directly in line with that of conventional funds.

2.3.4 Impact of Excluding Sin Stocks

An interesting sub-topics of ESG investing is the discussion of sin stocks and, in particular, the "sin stock anomaly." Although ESG itself does not exclude sin stocks, many investors choose to do so themselves by way of negative screening because these stocks don't align with the investor's personal values or social norms. Due to this reasoning, it has been shown by various studies that this exclusion leads to a negative effect on investor's return. There is a number of studies that have found that sin stocks earn a higher abnormal return, the so-called "sin-stock anomaly." One explanation for this is that they are constantly underpriced due to the lack of investor preference for them. The rather small population of investors who choose not to shun these stocks then can earn a reputation risk premium by investing in them [Blitz and Fabozzi, 2017]. This leads to the reasoning for the "shunned stock hypothesis," which states that controversial stocks have higher returns due to investor's shunning them which leads to their lower prices when compared with responsible investments. As mentioned by Hong and Kacperczyk [2009], the SRI market is quite big nowadays and constantly growing and many SRI managers, such as pension funds, constantly screen out their investment universe to exclude sin stocks. This is a substantial amount of overall investments and thus it potentially creates a rather large effect of SRI on the prices of sin stocks. Moreover, past studies have found that this higher return cannot be explained entirely by the traditional factor models [Derwall et al., 2011], although recently Blitz and Fabozzi [2017] found that it can be explained. Nevertheless, ESG investors that choose to exclude these stocks can be hurting their returns and it can also have an impact on their overall portfolio performance.

A paper by Hong and Kacperczyk [2009], which has been widely referenced since it was published, did one of the most comprehensive studies of sin stocks, in particular ones in the alcohol, tobacco and gaming industries, and the effect social norms had on them and the overall stock market. They used US sin stocks and had a long sample time period of almost 50 years, from 1962-2006, and then further analysed international markets, such as Canada, UK and Spain. To examine the shunning effects of sin stocks on their prices and whether they give higher excess returns, they analysed a time series of returns of sin stock portfolios by running regressions using the CAPM and Carhart's four-factor model. Additionally, they used cross-sectional regressions to see if sin stocks outperformed their comparable counterpart. They found a large, significant price effect from investors shunning sin stocks equal to around 15-20%. However, this price effect was smaller in the international sample than in the US.

The findings above were further confirmed by Derwall et al. [2011], who formed a sin stock portfolio of US firms using the KLD STATS database over the 1992 to 2008 period. They measured the abnormal returns by applying Fama-French's three-factor and Carhart's four-factor model. Their shunned stock portfolio had an annual return around 11.7%. Moreover, these returns were significant and relatively stable over the entire sample period.

Lastly, Scholtens [2014] studied perhaps one of the largest sin stock samples that was at least larger than any previous study done. Coverage included a global sample of 1,634 sin stocks from 1991-2012 from 94 countries and 14 "sin" industries. They created sin stock cluster portfolios based on various industries and then compared them with a conventional benchmark and then further tested their risk-adjusted performance with Carhart's model. They then examined the impact of negatively screening the S&P 500. They found that essentially all of the cluster sin portfolios significantly outperformed the market while the negative screening led to significant underperformance. However, they also found that the type of sin industry had an effect on performance and alcohol, tobacco, animal testing and contraceptives had significant positive returns. On the other hand, adult entertainment and stem cells actually showed slightly significant underperformance.

2.4 ESG Category Differences

ESG ratings can be further decomposed into individual pillars for the environmental, social and governance aspects. Some rating agencies, such as Thomson Reuters, go even further to decomposing each pillar into further quantitative categories so the ratings are transparent and each category can be looked at separately. This has also led to the debate of whether any individual pillar has the biggest influence on return and whether this varies for countries, industries or overall. For instance, one would assume the environmental rating would have a big influence in the oil industry, while governance might be the most important for investing in emerging markets due to the unstable nature of politics and firms in those markets. Hence, it is useful to note the differences of the individual pillars and the influence they have on stock returns. These pillars and their definitions can also vary substantially between data providers and thus is difficult to generalise.

An overview of the effect on financial performance of the individual pillars can be seen in a meta-analysis of over 2,000 aggregated studies, conducted by Friede et al. [2015]. Out of a total 644 studies examined within this aspect, they find governance as being the one with the highest proportion of studies showing a positive link, equal to 62.3% but also the highest proportion of a negative link of 9.2%. The environmental pillar showed a positive relationship in 58.7% of studies and negative in 4.3%. Lastly, the social pillar appeared to have the lowest relation out of the three, with 55.1% finding a positive link and 5.1% a negative one. Thus, just as with the aggregate ESG score, a general consensus on the relationship has yet to be established between categories.

2.4.1 Environmental Pillar

Just as with regular ESG ratings, the question comes of whether a high environmental rating relates to a high return. This topic is continuously becoming more and more important due to the increasing environmental concerns in the world today due to global warming and the impact firms have on the environment. A number of papers have tried to study this relationship in order to see the environmental performance of firms and its effect on stock returns. What makes it hard to study the relationship is that a generally accepted method to evaluate environmental performance is still yet to be reached due to the difficulties in quantifying the various environmental measures [Escrig-Olmedo et al., 2017]. Nevertheless, there are many advantages to doing well environmentally. For instance, high environmental performance can improve financial performance through saving costs, revenue generation, cost or liability avoidance and being an example of "best practices". Another argument for the positive link is that if environmental pollution is an inefficient resource use then lowering it will help not only the environment but also the bottom line of the firm[Hassan and Romilly, 2018]. Generally speaking, the environmental performance can usually be sorted into three categories: environmental impact, regulatory compliance and organisational processes [Delmas and Blass, 2010]. Most of the indicators and sub-categories that comprise the overall environmental score can usually fall under these categories and what's included will vary by each data provider.

Derwall et al. [2005] focused solely on the environmental aspect of ratings and examined the eco-efficiency scores of US firms from 1995 to 2003 and the impact they had on financial return using the Innovest rating database. Eco-efficiency is the ratio of the value a firm adds and the waste it generates from the creation of that value and it is a concept that is used to measure firm performance in the relative sense. A major improvement over other studies is that the scores were based on industry peers and thus were not penalized for operating in environmentally sensitive industries, such as energy and mining. They constructed two mutually exclusive portfolios that differed on their eco-efficiency and tested them with the CAPM and Carhart's four-factor model and then further with screening strategies. They found that the highly-rated portfolio had substantially higher average returns than the lower-rated one and the performance differential was not explained by differences in market sensitivity, industry-specific components or investment style.

An interesting paper by [Delmas and Blass, 2010] used a small sample and implemented a case study by ranking 15 chemical firms using three different environmental indicators (impact, management and compliance). One major finding they found is that companies who score lower on the environmental performance and compliance categories appear to have better environmental reporting and are more likely to implement pollution prevention plans for the future. They also found importance in looking at every environmental indicator since there is a trade-off between the indicators and an investor might miss some important points if they put all their focus on one indicator.

Further extensive analysis was done by Dixon-Fowler et al. [2013], who did a meta-analysis review of the link between environmental and financial performance based on 71 past studies. Consistent with earlier studies, they found a significant positive relationship between the two. Furthermore, they found that specifically some firm characteristics had an effect on the relationship while others did not. They also found that proactive environmental initiatives by firms did not increase financial performance more than reactive initiatives, meaning that firms that do the bare minimum and simply complied with regulations did not necessarily have lower returns than firms that tried to stay ahead of the curve and innovate with regards to their environmental practices. Lastly, they also found that smaller firms benefited from environmental performance even more so or just as much as big firms. This was somewhat surprising, since one can have the argument that smaller firms might not have the money or would like to not waste their limited resources on improving their environmental performance.

The argument of how correlated the environmental scores between data providers are is up for discussion, just as with the regular ESG scores. For instance, Semenova and Hassel [2015] studied the convergence of the environmental performance and risk of firms by using the MSCI, Asset4, and Global Engagement Services (GES) databases. They found that overall, the environmental ratings of the various agencies did not converge although they did share some common dimensions. They also found that industry-specific factors were drivers of environmental performance, which is to be expected and further enhances studies that controlled for the industry factor. Additionally, the positive link between environmental information and portfolio performance is usually explained differently by various authors. According to Renneboog et al. [2008], one explanation can be that the stock market undervalues environmental information, which essentially goes against the principals of the Efficient Market Hypothesis. It can also be that the eco-efficiency premium [Derwall et al., 2004] captures a premium which cannot be explained by the traditional risk factors and models. Whatever the cause, there is still much research to do in order to pin down the exact cause of the positive performance of the environmental aspect of ESG.

2.4.2 Social Pillar

The Social rating relating to corporate social responsibility (CSR) is another increasingly important aspect of sustainable investing. Nowadays, perhaps more than ever, there is pressure on firms to be responsible corporate citizens and have a positive impact on the firm and their employees [Williams, 2015]. Various stakeholders such as governments, media, activists and shareholders are holding firms accountable for the social consequences of their actions and it has become an extremely important part of everyday business operations [Porter and Kramer, 2006]. Generally, the social pillar of the ESG score relates to socially responsible behaviour which may include issues such as employee rights, fair labour practices, business ethics policy and working conditions [Wang and Sarkis, 2017]. It is often the rather overlooked pillar when compared with the environmental and governance and gets far less attention in research. For instance, Pelosi and Adamson [2016] mentioned a review of the boards of 52 gas, oil and mining firms. At the board level, only four of them examined social performance while around half of them did the same with environmental performance.

Statman and Glushkov [2009] analysed returns of firms from 1992-2007 based on their social responsibility ratings. Using the KLD database, social responsibility sub-topics consisted of community, diversity, employee relations, environment, products and human rights. They further applied screening strategies to their analysis. They found that firms that are highly rated on social responsibility have higher returns than conventional firms. However, they also found that applying negative screening to the portfolio resulted in lower returns of socially responsible stocks when compared to the conventional ones. When the two effects are taken into account together, they essentially find no effect and that investing high on social sustainability is essentially equivalent to conventional investments. Lastly, they found that applying the best-in-class screening strategy to include firms rated highly on the social aspect in portfolio construction resulted in higher excess returns while also still being somewhat socially responsible. However, the limitation was to not shun any stocks from that portfolio and thus include sin stocks.

Additionally, Jiao [2010] examined the effects of stakeholder welfare, which relates to social responsibility, and firm value. He constructed his own stakeholder welfare score, which measured the extent to which firms improve or have good operations with stakeholders other than shareholders, such as employees, communities, customers and environment. He essentially created his own score based on the social pillars from the KLD database and used Tobin's q as a performance measure. He discovered that an increase of 0.587 in Tobin's q was found for every 1 point increase in the stakeholder welfare score. Thus, he found that stakeholder welfare, even though it included many intangibles such as human capital and reputation, nevertheless, was still meaningful for firm value and really showed the important of the social pillar, even when it cannot be measured concretely.

Another paper by Brammer et al. [2006] investigated social corporate performance and stock returns within the United Kingdom. They further decomposed the social score into three parts: environment, employment and community activities. They ran cross-sectional regressions and contrary to other studies, found that firms with the lowest levels of social scores outperformed the market while those with the highest scores had lower returns. In particular, they observed a negative correlation between the environmental and employment indicators and returns while the community aspect was the only one with a weak positive correlation to returns. Furthermore, they found that the lower returns for high-rated scores were not explained when investigated further by Fama-French and industry effects. Their findings contradict other studies since the higher level of social performance was found to be detrimental to investor's financial performance.

Although not specifically testing only the social score, Taylor et al. [2018] used Bloomberg ESG data for a US sample from 2009 to 2014 which included 1,688 firms and all industries which narrowed down to 432 firms after data cleaning. They used Tobin's q as the main dependent variable in order to test the effect of ESG scores on firm value and further tested each individual pillar and also the aggregate ESG score. They found that only the social responsibility score had a positive relationship with firm value, as measured by Tobin's q. Additionally, they found no significant link between firm value and the environmental and governance scores.

2.4.3 Governance Pillar

The governance pillar is an important part of the overall ESG rating and it all comes down to the firm's use of good corporate governance practices. This may include ensuring diversity of board members or separating the CEO and chairman roles, among many other governance issues [Husted and Sousa-Filho, 2017]. High standard of corporate governance can lead to lower risk, reduce information asymmetry due to better public disclosures and limits the possibility of managerial entrenchment [Clark et al., 2015]. According to Ammann et al. [2011], good corporate governance can affect firm value positively in two ways. It can affect stock price multiples positively, since investors expect less cash flows to be wasted and more of the company's total profits will be paid out to them as interest or dividends. It can also reduce the expected return on equity which in turn lowers shareholders' costs related to monitoring and auditing and therefore leads to a decrease in the cost of capital for the firm.

The effect of good governance of firms, although not in the ESG aspect, has long been studied. Gompers et al. [2003] did one of the earlier studies on the topic that has been extensively used in literature in the last decade. They created their own "governance index" using 24 governance rules to represent the balance of power between shareholders and managers. Then, applying time-series regressions on around 1,500 firms from 1990 to 1999, they studied the relationship between governance and corporate performance. They concluded that an investment strategy that bought firms in the lowest decile of the index, which equated to the best governance, and sold firms in the highest deciles, worst governance, earned abnormal returns of 8.5% per year. Overall, firms on the index with the best governance had higher firm value, profits and sales growth and lower capital expenditures.

Ammann et al. [2011] investigated the link between corporate governance at the firm-level and firm value within 22 developed countries using the Governance Metrics International (GMI) database for the five year period from 2003 to 2007. Sampling over 2,300 firms, they created three additive governance indices. They found a strong, positive link between corporate governance and firm valuation in all indices which were further upheld in robustness tests.

Akbar et al. [2016] studied the relation between corporate governance compliance and firm performance for the UK market. The sample covered the ten year period from 1999 to 2009 and examined 435 public companies, excluding financial ones. Their analysis was an improved version of previous studies because they applied the robust Generalised Method of Moments (GMM) Estimation which controlled for the effects of heterogeneity and potential endogeneity. When controlling for endogeneity, they found no significant link between corporate governance and performance, contrary to previous studies although only limited to the UK market. They further did a second analysis using OLS regressions and fixed-effects model, like some prior studies, and found that governance compliance had a positive relationship to return on assets (ROA). Thus, they concluded that earlier studies were biased and found a link simply due to the method applied and for not accounting for the dynamics of firm's performance, which was a very important aspect.

Although many studies have found a positive link, Bhagat and Bolton [2008] challenged this view. They examined the relationship between corporate governance and performance. They enhanced prior studies by accounting for the inter-relationship between corporate governance, corporate performance, capital structure and ownership structure. They found that no governance measures were correlated with future stock returns and thus had no impact on financial performance. However, they did find that better governance was significantly and positively correlated with operating performance.

2.5 Cross-Country Differences in ESG

It is important to note that the definition of SRI and use of ESG is not uniform between countries. There are many cross-country differences in the importance of SRI and ESG and the way it is applied and to what extent. The studies that find an association between ESG and financial performance also vary by country. Due to the increasing trend of being socially responsible, it is important to analyse the differences of performance in some of the biggest SRI markets in the world. Nowadays, although SRI markets are increasingly growing around the world, the US market is the most developed SRI market in the world, with Europe right behind and they account for the biggest market share of all SRI investments globally [Van Dijk-de Groot and Nijhof, 2015].

2.5.1 USA and Europe

Lean et al. [2015] states that for overall responsible investments, around two thirds are held by European investors, around a quarter with US investors, and the other 10% with the rest of the world. This is further confirmed by Schröder [2014], who states that investment of SRIs in the US was around 10% of all assets managed in 2011, while in Europe it was 17% with a market volume of EUR 2.5 trillion, compared to USD 3.3 trillion. For instance, Williams [2007] stated that in the US, there is much more diversity among SRI approaches. There, SRI is mainly promoted by independent organisations and specialist fund providers while in Europe formal standards, institutional strategies and tougher government regulation are contributing to the development of SRI.

In their meta-analysis of past studies, Friede et al. [2015] accumulated the results of 402 global studies. They found a positive relationship between ESG and financial performance in 42.7% of studies in North America with 7.1% finding a negative link. Surprisingly, developed Europe had the lowest share of a positive relationship of any other region studied, equal to only 26.1%, while 8% found a negative link.

Auer [2014] used ESG ratings from Sustainalytics to study whether stock selection within SRI destroys or adds value of portfolio performance in the European market with a sample period of 2004-2012. The coverage included 892 stocks that were listed on the European STOXX 600 for at least six months. Implementing various screening strategies on the ESG score and the individual pillars, he sought to answer whether SRI adds or destroys value for European investors. With the main performance measure used being the Sharpe ratio, the conclusion was that European investors can do well financially by investing socially responsibly but it was contingent on the type of ethical screening method. He found that negative screens, which excluded unrated stocks, actually significantly outperformed a passive investment in a European benchmark stock portfolio. Additionally, extra negative screens based on social and environmental scores were found to not have any effect on portfolio value (if cut-off rates weren't too high) and that governance screens in particular significantly improved portfolio performance. In contrast, positive screens were found to hurt portfolio performance and all conclusions were upheld in robustness tests.

Auer and Schuhmacher [2016] analysed the performance of socially responsible investments covering a broad dataset consisting of stocks from the US, Europe and Asia Pacific. They gathered ratings from Sustainalytics for the period from 2004 to 2012 and applied a "state of the art" statistical performance measurement based on total risk. Their results were highly dependant on the geographic and industry focus of the ESG investment strategy and the ESG criteria that was used. In the US (and Asia-Pacific), they found that picking highly-rated ESG stocks generally had neither a positive nor negative effect on financial performance relative to benchmarks and relative to low-rated stocks. In Europe, they found no outperformance of ESG-rated stocks and furthermore, dependant on the industry and ESG criteria, significantly lower risk-adjusted performance was sometimes observed relative to benchmarks.

Lean et al. [2015] analysed a broad sample of European and North American SRI funds from 2001-2011. They found that SRI funds outperformed the market benchmark for both regions and thus investor's did not need to give up returns in order to invest ethically while also showing that the decrease in diversification does not deter from returns. Furthermore, they found that North American SRI funds actually outperformed the European funds, as examined by Carhart's alpha value, and that European funds had higher downside risk when compared with the North American.

Lastly, Dorfleitner et al. [2013] examined the long term effects of holding stocks based on their level of ESG rating, retrieved from Asset4, for the 2002-2011 period. They used a worldwide sample to represent the investment universe of an international and active stock investor which resulted in a total of 4,396 firms. Overall, Asset4 rates firms from 60 countries but they excluded numerous small countries and split the rest into four regions: North America, Europe, Japan and Asia-Pacific. They tested the individual pillars and examined the abnormal returns of portfolios holding the top 20% and bottom 20% of ESG ratings in each region. Their findings found numerous differences between regions. In North America and Europe, they found overall significantly higher long-term financial performance relating to each pillar of the ESG score. However, for Europe, the governance pillar was the only one that did not have a significantly higher abnormal return and the authors state that a contributing factor to this could be the differences in corporate governance systems between the two markets. Lastly, they found that going long in the top-rated portfolios and shorting the low-rated one led to even higher abnormal returns than investing in any portfolio alone.

Chapter 3

Theoretical Background

One of the most important problems of modern financial economics is the quantification of the trade-off between risk and expected return. It may seem intuitive that investors expect risky investments such as those in the stock market to yield higher returns than investments free of risk, however, it was only with the development of the Capital Asset Pricing Model (CAPM) that economists were able to quantify risk and the reward for bearing it. This chapter provides a review of some of the focal theories in financial economics, where the objective is to provide an important insight into the quantitative methods employed in this paper and illuminate some of the most important implications that accompany them.

3.1 Modern Portfolio Theory

In the financial markets, risk plays no lesser role than returns, and it is often said that risk and returns go hand in hand. This trade-off between risk and return was first documented by Harry Markowitz, a young mathematician who had never bought or sold stocks. Markowitz [1952] published his famous article *Portfolio Selection*, in which he introduced the so-called *mean-variance analysis* as a tool for determining optimal portfolios over a given period. In his paper, Markowitz defines formally, for the first time, risk as the standard deviation in the returns of stock prices; better known as volatility. He also argued that all necessary information regarding securities can be extracted from three variables:

the mean of the return, the standard deviation of the mean, and the mean's covariance with other mean returns. These three variables will then suffice in computing the trade-off between risk and return of any portfolio. The efficient combination of standard deviation and mean form a hyperbola called the *efficient frontier* of risky assets [Munk, 2016]. Thus, investors could optimally hold what Markowitz defined as a mean-variance efficient portfolio, which is a portfolio with the highest expected return for a given level of variance.

Markowitz seminal work on portfolio theory laid the groundwork for a whole new field of asset pricing theory, where researchers started to focus more on quantifying risk in the assessment of expected returns. James Tobins [1958] inclusion of risk-free assets was an important contribution to portfolio theory. The concept he described is known as the separation theorem because it separates Markowitz's approach from the completely different decision of dividing up the whole portfolio between risky and risk-free assets. Tobin showed that when combining a risk-free asset with a portfolio on the efficient frontier, it is possible to assemble portfolios whose risk-return profiles are above those of portfolios on the efficient frontier [Tobin, 1958]. Building on Markowitz' and Tobins work, Sharpe [1964] and Lintner [1965] showed that if investors have homogeneous expectations and optimally hold mean-variance efficient portfolios, then in the absence of market frictions, the so-called market portfolio of all invested wealth will itself be a mean-variance efficient portfolio. In equilibrium, capital asset prices have adjusted so that the investor – assuming he follows rational procedures such as diversification – is able to attain any desired point along what they called the capital market line. By describing the way the price of risk results from the basic influences of investor preferences as well as the physical attributes of capital assets, the Sharpe-Lintner capital market line was the first theory to give any real meaning to the relationship between the price of a single asset and its risk.

Sharpe and Lintner's work laid the groundwork for one of the most established model of the equilibrium prices of financial assets: the *Capital Asset Pricing Model* (CAPM), derived by Treynor [1961], Sharpe [1964], Lintner [1965], and Mossin [1966]. The CAPM offers an elegant model of the determinants of the equilibrium expected return of any individual risky asset in the market. According to the model, the risk premium of any risky asset equals the product of the market-beta of the asset and the market risk premium. Hence, the market-beta is the only asset-specific determinant of the asset's risk premium. In a diagram with market-betas along the horizontal axis and expected returns along the vertical axis, the CAPM equation corresponds to a straight line intercepting the vertical axis at R_f and having a slope equal to the market risk premium, E[R_m]-R_f. This line is called the Security Market Line (SML). Assets corresponding to points on the SML are priced in accordance with the CAPM. If the point plots below (above) the line, the asset is overpriced (underpriced) according to the model [Munk, 2016].

3.2 Arbitrage Pricing Theory

Ross [1976] recommended an alternate theory for assess risk called the *Arbitrage Pricing Theory* (APT), which predicts a relationship between the returns of a portfolio and the returns of a single security through a linear combination of several independent macroeconomic variables. The APT is basically a multi-factor version of the CAPM relation between market-betas and expected returns. The CAPM assumes that market risk is captured in the market portfolio, whereas the APT allows for multiple sources of market-wide risk and evaluates the sensitivity of investments to changes in each source. However, the drawback of the APT is that it does not clarify which factors should be used in the model [Damodaran, 2012]. Chen et al. [1986] tested a multi-factor model that replaced the unknown statistical factors with explicit economic factors and found five macroeconomic variables that were highly correlated with the factors that came out of factor analysis.

The best known multi-factor model is the *Fama-French three-factor model* developed by Fama and French [1993]. Instead of building a risk and return model from economic theory, Fama and French analysed how investments are priced by markets and related returns earned to observable variables. The three factors are

1. *market* (R_m-R_f): the return on a broad stock market index in excess of the risk-free rate
2. *small-minus-big* (SMB): the return on a portfolio of stocks in small companies (according to the market value of all stocks issued by the firm) minus the return on a portfolio of stocks in large companies;
3. *high-minus-low* (HML): the return on a portfolio of stocks issued by firms with a high book-to-market value (value stocks) minus the return on portfolio of stocks issued by firms with low book-to-market values (growth stocks)

In their model, the two easily measured variables, size and book-to-market value, are combined in order to capture the cross-sectional variation in average stock returns associated with the market beta, size (market equity, stock price times number of shares), leverage, earning/price (E/P) and book-to-market equity (BE/ME) [Fama and French, 1993]. They also demonstrated that when tests allow for variation in beta that is unrelated to size, the relation between the market beta and the average return is flat, even when beta is the only explanatory variable. Thus, according to their reasoning, market capitalisation and book-to-market ratios were better proxies for risk than betas.

Carhart [1997] suggested an extension to the Fama-French three-factor model by adding the momentum factor. The momentum factor is represented by WML, the return on a winners-minus-losers strategy, or more precisely the difference between the future return on a portfolio of recent "winners" and the future return on a portfolio of recent "losers". This model is known as the Carhart four-factor model (C4F), and is estimated by simply adding the WML factor to the Fama-French three-factor model. Both the Fama-French three-factor model and the Carhart four-factor model remain the standard of performance evaluation in financial literature due to their ability to explain the cross section of returns. As such, they have been chosen as appropriate models to test the arbitrage pricing theory. The mechanics of the models will be further elaborated on in Section 5.3.2.1.

3.3 The Efficient Market Hypothesis

Since the 1970s, one of the most debated question in finance has been whether financial markets are informationally efficient or not. The Efficient Market Hypothesis (EMH), formalised by Fama [1970], states that prices in financial markets consistently reflect all information available to investors. Fama distinguished between three forms of information efficiency, known as the *weak*, *semi-strong*, and *strong*. The weak form efficiency represents a market in which *all historical information* is reflected in the current security price. The semi-strong form efficiency is when prices reflect *all publicly available information.* Finally, strong form efficiency is when prices reflect *all relevant information*, including non-public information from private or insider sources. Another way of describing the EMH is to say that in an efficient market it is impossible for investors to make abnormal excess returns. This would mean that – given the market is efficient – the ESG-rated portfolios constructed in this paper should not be able to yield any positive returns after being adjusted for risk, regardless of the strategy approach. Whether investors rely on "technical analysis" or "fundamental security analysis" does not matter since the theory argues that, in any case, it will be impossible to persistently obtain positive abnormal returns. Only with the arrival of new information will prices change, and thus only analysis leading to insights other investors don't have is rewarded. Furthermore, even if investors were able to find information that generates risk-adjusted abnormal returns, they would still not be able to come to a conclusion about market efficiency because of the fact that the equilibrium model could be wrong; a problem known as the joint-hypothesis problem [Fama, 1991].

There are numerous empirical studies that provide supporting evidence for the EMH. Fama [1970] and Malkiel [2003] point out that there is enough empirical evidence that supports both the weak and the semi-strong form of the EMH. In fact, Jensen [1978] states that the extensive amount of evidence makes the EMH the most reliable empirical assertion of economic science. Although the EMH has some strong intuitive reasoning, it has also received a considerable amount of criticism. Grossman and Stiglitz [1980] simply argue that perfectly efficient markets are impossible. They point out that if information is costly, prices cannot perfectly reflect the information available. Nalebuff and Stiglitz [1983] also make the point that speculative markets cannot be completely efficient at all points in time. Pedersen [2015] claims that markets are inefficient enough that money managers can be compensated for their costs through the profits of their trading strategies and efficient enough that the profits after costs do not encourage additional active investing. However, perhaps the most compelling evidence against the efficient market hypothesis is the consistent association between value investing and positive abnormal returns.

Chapter 4

Data

In this section, data description of the ESG scores used in the analysis will be examined first and an overview given. Secondly, data collection will be discussed and the methods used to retrieve and clean up the dataset used in the analysis. Lastly, a number of issues with using ESG ratings for the analysis will be discussed with a focus on the limitations of using the data.

4.1 Data Description

4.1.1 ESG Scores

The Thomson Reuters ESG database is used for the analysis in this paper. Back in 2009, Thomson Reuters acquired Asset4, which was the first ESG agency ever to provide raw data to investors [Huber and Comstock, 2017]. Very recently at the end of 2017, they improved the Asset4 database with a new and improved ESG database which may be found in the Thomson Reuters Eikon software, with improved rating methodology and more transparency. At the time of this thesis, no known research papers have utilised this database. Although there are many similar data providers for ESG on the market right now, Thomson Reuters is one of the leaders and particularly its new and improved ESG database gives optimal data to work with. Their ESG database is one of the largest, oldest and most comprehensive in the world, with historical data going back to 2002 for some companies. As of end of 2017, it covers over 6,500 companies globally and includes over 400 relevant measures from various public sources such as CSR reports, annual reports, websites and media sources. The overall ESG score is composed of 178 ESG measures deemed most critical relating to the environmental, social and governance factors and then an overlap of the ESG Controversy score is added in order to come to the final ESG Combined score.

The ESG score can be further decomposed into 10 categories: environmental (resource use, emissions, innovation), governance (management, shareholders, CSR strategy) and social (workforce, human rights, community and product responsibility). The Resource Use score evaluates a company's performance and capacity to minimise the use of energy, water and materials and improving supply chain management in order to have more eco-efficient solutions. The Emissions score measures the company's commitment and effectiveness in reducing environmental emissions in their business operations and the Environmental Innovation score evaluates a firm's capacity to reduce environmental costs and thereby innovating new environmental technologies or eco-designed products. The Management score relates to the firm's commitment and success in following the best corporate governance principles. The Shareholders score evaluates company's equal treatment of shareholders and use of anti-takeover tools. While the CSR Strategy score measures a firm's effectiveness in including economic, social and environmental dimensions in its everyday operations. Lastly, part of the Social score, the Workforce score measures the firm's performance in keeping a safe workplace, job satisfaction and development opportunities for employees. The Human Rights score evaluates respectfulness of fundamental human rights and the Community score evaluates the firm's ability to protect public health, business ethics and being a good citizen. Finally, the Product Responsibility score measures the firm's effectiveness in producing quality goods or services while acknowledging the customer's safety, health, data privacy and integrity [Thomson Reuters, 2017].

The new Thomson Reuters ESG ratings have a number of improvements from the former, equal-weighted Asset4 database which has been widely used in past studies within ESG. The major improvements include the Controversy score, which includes controversies of a firm taken from global media sources and incorporated into the final ESG Combined score. The Controversies score is calculated based on 23 controversy topics and the ESG score gets updated accordingly once a controversy occurs. Second improvement is that industry and country benchmarks are already accounted for in the scores in order to have a comparable analysis within peer groups. This is particularly important since many studies in the past have had to account for this manually by adjusting the scores by industry and country in order to have a sound and comparable analysis. The benchmark used for the Environmental, Social and Controversy scores is the Thomson Reuters Business Classification (TRBC) Industry Group because these topics are more comparable within the same industries. For instance, the Environmental score is likely more important and lower in the oil industry than it is in retail. Likewise, the benchmark for the Governance scores is the country of headquarters. Since governance practises are steadier within countries, scores in USA are not directly comparable to, for example, scores in China and therefore also need to be adjusted accordingly. Thus, the ESG scores used in this analysis adjust all scores for industries and country effects so no firm is penalised for simply operating in a less-developed country or industry that is known to be worse in terms of ESG. Lastly, an improved percentile rank scoring methodology is used in Thomson Reuters' new database, compared with Asset4's equal-weighted ratings based on z-scores. The percentile rank score is based on three factors: how many companies are worse off than the current one, how many have the same value and how many have a value at all. The percentile rank score is based on the rank and thus is not as sensitive to outliers. Each category score is then an equally weighted sum of all suitable indicators for each industry used to create it. This ensures the most sound analysis and comparable firms that do not need to be adjusted for their industry and country [Thomson Reuters, 2017].

The rating scale of Thomson Reuters includes both a percentile score, from 0 to 100, and a letter score, from D- to A+. The distribution of this can be found in Table A.1, which may be found in Section A in the Appendix. This paper focuses on the percentile rank only, since it is the numeric and quantifiable score and can be easily converted to alphabetical. The ESG, ESG Combined and the category scores are all given a score from 0 (worst) to 100 (best). The only exception is the ESG Controversies score which is also graded from 0 to 100 but the lower the score, the better. For instance, a firm with a score of 60 would have more controversies than one with a score of 0 and a high controversy score would also contribute negatively to the overall ESG Combined score. Except for management departures, each Controversy score is quantitative and therefore quite reliable. Furthermore, each of the 10 categories pertaining to ESG are weighted proportionally based on the number of measures in each category [Thomson Reuters, 2017]. For instance, the Management score includes more issues than the Human Rights score, such as composition, compensation, independence, diversity and more.

For this reason, the Management Score has a weight of 19% in the final ESG Score while Human Rights only has a weight of 4.5% [Huber and Comstock, 2017]. The exact percentage proportion of each category in the final ESG Score can be seen in Table 4.1. The new ESG scores of Thomson Reuters do not give an aggregate score for each individual pillar of ESG and the scores had to be combined manually. For the Environmental rating, the total number of indicators for each component were summed up and then given an equivalent weight that equated the final weight to 100%. The aggregated Social and Governance scores were constructed in the same way.

TABLE 4.1: Distribution of Weights of Overall ESG Score

	Number of Indicators	**Weights**
Environmental		
Resource Use	20	11%
Emissions	22	12%
Innovation	19	11%
Social		
Workforce	29	16%
Human Rights	8	5%
Community	14	8%
Product Responsibility	12	7%
Governance		
Management	34	19%
Shareholders	12	7%
CSR Strategy	8	5%
TOTAL	178	100%

Using ESG data begs the question of how do companies get rated on ESG and how are they compared to non-rated firms. For the sample used in this analysis, the entire universe of US stocks included 13,488 stocks, while the developed European universe included 8,184 stocks. The ESG-rated sample for US included 2,844 of stocks, equal to around 21% of the total US universe while the European sample had 1,224 stocks, equal to around 14% of the whole universe. Generally, firms get ESG ratings when they release a sufficient amount of information on sustainability in order for Thomson Reuters to be able to rate them accurately on each sustainability metric and measure. If this information is not able to be gathered for a firm, then calculating a concrete rating for them is difficult to do and thus the stocks cannot get an ESG rating assigned to them.

4.2 Data Collection

4.2.1 ESG Data

All ESG data was retrieved from the Thomson Reuters Datastream database. The empirical analysis in this paper will be conducted on two geographical regions, namely the US and developed Europe and data was gathered equivalently for both markets. The collected data sample includes all publicly listed ESG-rated firms in the relevant regions for the 12 year period from December 31st, 2005 to January 31st, 2018. The chosen time frame was deemed optimal as it provides a sufficient amount of observations. The amount of ESG-rated firms has increased substantially throughout the years and only a small amount of firms were present at the start of the database in 2002. Additionally, the time-frame also includes the financial crisis in 2008, which is important in order to account for the global economic downturn. Thus, the data sample of this analysis resulted in a total of 1,224 firms in the European market and 2,844 firms in US. The geographical distribution of stocks that have been included in the developed Europe dataset can be seen in Figure 4.1. The choice of what countries should be included in the European sample was guided by classifications of Thomson Reuters. However, the decision of what countries should be included in the European sample is dependent on the researchers objective and thus is quite arbitrary. The authors of this paper decided to separate developed and emerging European countries since they are not directly comparable and a different economic landscape is seen for ESG emerging markets than for developed markets [see Sherwood and Pollard [2018] for further elaboration].

A total of 12 data sets relating to the various ESG score components was retrieved for both the US and European sample. This included the ESG score, ESG Combined, Resource Use, Emissions, Environmental Innovation, Management, Shareholders, CSR Strategy, Workforce, Human Rights, Community and Product Responsibility scores. Traditionally, ESG ratings are presented as annual scores, and thus most empirical research papers use annual re-balancing to assess their predictability of returns. However, a close examination of the ESG database revealed that, although the ratings changed generally only once on a 12-month basis, the timing of the change depended on the reporting period (fiscal year) of each respective company. In some rare exceptions, the scores also changed more than once in a year for firms where a major ESG-related event occurred. By examining returns based on

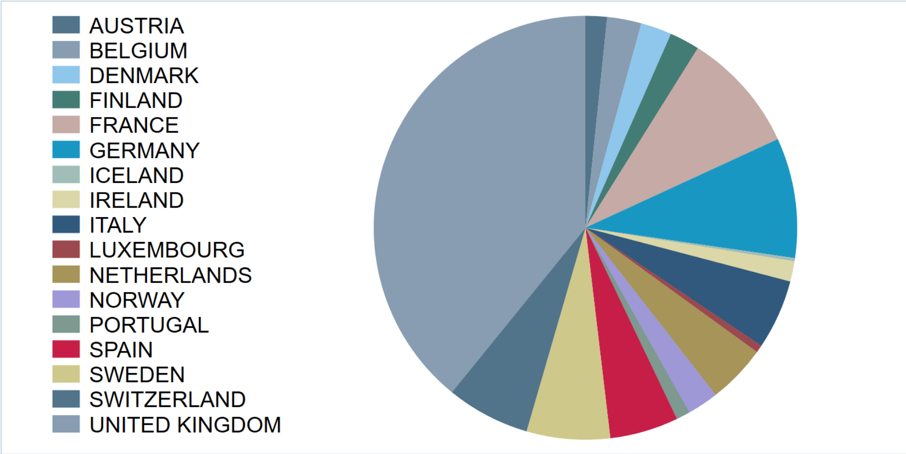


FIGURE 4.1: Geographical Distribution of Stocks In the European Sample

the one month lagged ESG score, the immediate effect that a ESG-related event might have on returns is captured more precisely in this analysis. This is an improvement to many studies which simply use yearly scores, even though the rating could have changed in the first few months of the year and thus the most optimal results are presented in the analysis which are able to capture the change in ratings better.

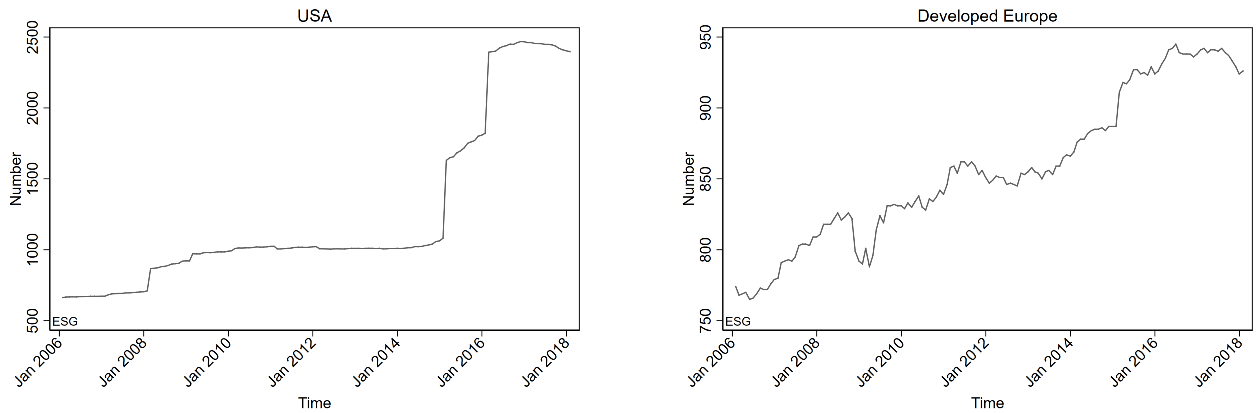


FIGURE 4.2: Historical Development of the Number of Firms Included in the Dataset

4.2.2 Financial Data

After aggregating the ESG data into a single panel dataset, the raw ESG sample contained 131,362 observations for the European sample and 185,644 for the US sample. The next step taken in the data collecting process was to retrieve financial data for the ESG observations at hand, which was all also obtained from Datastream. Although Datastream data is widely used in financial literature, it has been known to contain some significant data errors that could distort any statistical analysis performed on the data. Ince and Porter [2006] found that precautionary steps are needed to be taken in order to screen the data sample for potential errors, which were taken into account throughout the data collecting process.

For each region, both static and time series variables were gathered. Static variables included the company's International Security identification Number (ISIN) in order to uniquely identify each security. For a further check, the ISIN Code was retrieved, which identified whether the security was a primary or secondary listing and each was checked to ensure it was a primary listing. Next, the type of instrument was retrieved in order to check that each security was in fact an equity listing and it was found that several wrong security types were included in the data set, such as American Depository Receipts, and were therefore deleted. Next, the equity status was gathered in order to see whether the company was active or dead and to account for the companies that are no longer listed. In order to avoid survivorship bias, companies were included up to the date of their de-listing that have dropped out of the data sample due to a merger, acquisition or bankruptcy and so the analysis accounted for firms that were not currently trading [Aw et al., 2017]. Thus, the dataset included a total of 426 dead firms for the US sample and 246 firms for the European sample. Additionally, it was observed that in some cases Datastream automatically fills in observations of companies that are dead until the end of the sample period. In order to account for these observations, the inactive date of each security was retrieved in order to identify the day on which a company became privately held, liquidated, merged or otherwise became inactive. The inactive dates were then used to create dummy-variables that accounted for observations for periods after the inactive date. These observations where then dropped out of the sample, which accounted for a total of 17,221 observations in the European sample, and 20,850 observations in the US sample. Lastly, the last static variable retrieved was the geographical classification of company, which gave the home or listing country of each security and was cross-checked to match the regions analysed.

Next, monthly time series variables were retrieved for Unadjusted Prices and the Total Return Index for the sample period. The Unadjusted Price is the closing price which has not been adjusted for rights issues and bonuses and therefore is the actual or "raw" price recorded at end of month. The Total Return Index is the theoretical growth in the value of a share holding over a certain period and it assumes that dividends are re-invested and end-of-month closing prices were also retrieved for this. For both time-series requests, all prices were converted to US dollars in order to have a consistent valuation and comparison throughout the whole analysis and this was done for both the European and US sample. When working with the Datastream database, a number of data issues needed to be manually addressed. Following the framework of Ince and Porter [2006], all return observations were put though specific screens designed to identify critical data errors that would have to be removed from the sample. The first step involved screening for so-called "penny stocks", for which end of previous month's price was below $1. This was important because Datastream rounds prices to the nearest penny and when the stock price is very low, the return can be highly distorted and go above 1,000%. This affected a total of 8,101 observations in the European sample and 2,713 in the US sample. Another data error was found where the return went up significantly but was reversed within one month. Ince and Porter [2006] found that a threshold of 300% was appropriate and so the data was further screened for the error by taking out the returns in R_t and R_{t-1} if R_t or $R_{t-1} > 300\%$ and $(1+R_t)(1+R_{t-1})-1 < 50\%$. Since many of these errors were already accounted for by taking out the penny stocks, it affected no observations in the European sample and two in the US sample. Finally, after screening for data errors, the Total Return Index was used to calculate holding period returns for each stock and each month in the data sample, which will be further elaborated on in Section 5.1. After matching available financial data with the corresponding ESG observations, the summary statistics of the final ESG ratings are summarised in Table 4.2.

4.2.3 Factor Data

The factor data for the analysis has been gathered from Kenneth R. French's Data Library for both regions. The Fama-French five-factor data was retrieved under U.S. Research Returns Data for the

Table 4.2: Summary Statistics of the ESG Ratings for US & Europe

EUR						USA					
variable	mean	sd	min	max	N	variable	mean	sd	min	max	N
ESG	57.9251	16.5144	7.8300	96.7500	124230	ESG	48.4831	17.3233	8.5400	98.0000	179777
ESGC	50.5979	15.7420	7.8300	95.2100	124230	ESGC	41.7002	14.3726	7.1500	95.6700	179741
ENV	62.6379	20.5136	5.0262	99.2479	124230	ENV	45.3387	22.2542	2.9151	98.8861	179741
SOC	59.9144	20.3780	4.9343	99.4056	124230	SOC	49.4294	19.3964	3.7308	98.9373	179741
GOV	50.5946	21.2724	1.0533	98.8878	124230	GOV	50.7290	21.8873	1.8544	99.3185	179741

This table presents the summary statistics for the cleaned ESG ratings. For each region the reported statistics include the mean score, standard deviation of the scores, minimum and maximum of each score, as well as the total number of observations for each respective ESG score.

US region and for the European region, Fama-French European five-factors were retrieved under Developed Market Factors and Returns. For both regions, the risk-free rate (R_f), excess return on the market ($R_m - R_f$), small-minus-big (SMB) and high-minus-low (HML) factors were gathered for the time period analysed. The additional momentum factor (WML), [1] as discovered by Carhart was further retrieved in the same way.

4.3 Data Issues

The pillars of any ESG investment strategy are only as strong as their underlying data. One of the main issues of ESG data can be linked to the global absence of legislation on ESG data standards and reporting requirements. In the US, companies are required to disclose “material information”, i.e. information deemed important to investors for investment decisions. However, since regulatory agencies have not given formal standards as to what ESG data counts as ”material”, companies can essentially determine themselves which ESG factors are material and what information should be disclosed to investors. On the other hand, in Europe, ESG standards are becoming increasingly more formal and in 2014, the Accounting Directive on disclosure of non-financial and diversity information by certain large companies was approved [Bender et al., 2017].

Due to the difficulty in concretely defining sustainability concepts, there is much criticism with regards to ESG data and it’s usage in research and practice [Windolph, 2011]. A major concern with the ESG ratings is that the various data providers use different methodologies to measure the same

[1] The Kenneth R. French Data Library uses MOM to denote the momentum factor. Some literature also denote the momentum factor as UMD for ”Up-Minus-Down”.

sustainability concept. For instance, some rating agencies measure environmental performance with indicators of their environmental processes, like KLD, which attributes a higher score for products that have a positive impact on the environment. On the other hand, other raters concentrate on the firm's environmental impact, such as FTSE4Good, who assess the firm's environmental hazards. Other significant differences can be found between rating providers. For instance, KLD and Asset4 rate companies by their products' safety, which other providers do not do. Moreover, Asset4 and KLD include financial metrics in their rating methodology while others exclude these. This creates some major issues with the data, even if the ratings converge between providers it doesn't necessarily mean they were calculated with the same inputs and thus are not directly comparable [Chatterji et al., 2016]. As further asserted by Hawley [2017], CSR reports of firms that are often the basis for the ratings are "selective, subjective and not comparable." This creates a major challenge with using the scores in research and thus any conclusions that might be made are only relevant to that specific provider and are often mentioned as a limitation in past research. A prime example of this was discussed by Kerber and Flaherty [2017], who stated that Tesla has a top AAA ESG rating from MSCI and only a middle-level rating from Sustainalytics which was below that of Ford and General Motors. This was attributed to the fact that Tesla doesn't release carbon emission data for its manufacturing plants and this aspect is given a significant weight in the total score for Sustainalytics but not for MSCI.

With such a diverse group of data providers, each pertaining to their own ESG rating methodology, the logical question arises of whether these differences in methodologies result in significant differences in their respective ratings and what the correlation of scores is between the various providers. Bender et al. [2017] looked at the cross-sectional correlations of four of the leading data providers' ESG scores, including Sustainalytics, MSCI, RobecoSAM and Bloomberg. They used the MSCI World Index as the coverage universe and found that the correlation ranged from 0.47 go 0.76 across the aggregated scores, thus clearly suggesting a weak association in convergence of ratings across a few of the leading ESG data providers.

Another big issue with the ESG data is lack of transparency of the various rating providers with the way they calculate their ratings. As mentioned by Hawley [2017], there is usually no full disclosure from the data providers of their criteria, methodology and threshold levels. This leads to three biases among the ratings. The first is the geographical bias, where countries that are not as developed as say, the US, get penalised simply because of the standards in their country even though they might have superior ESG when compared to peers within their own country. The second is factor bias, which was mentioned above as the differences in methodologies and criteria of ratings between the providers. Last is the selection bias, which is a bias towards investors or stakeholders. For instance, there can be a bias toward rating larger companies higher simply because they report ESG issues with bigger magnitude and in more detail than smaller firms since they have the resourced to do so [Hawley, 2017]. On the positive side, the data used in this analysis is conducted using Thomson Reuters ESG scores, which was improved from the former Asset4 by creating more transparency and the relevant metrics for the user to be able to calculate each score themselves and see exactly what contributes to the score and with which weight on each metric. Additionally, the biases mentioned are reduced since Thomson Reuters adjusts their ESG scores by country and industry-effects.

Chapter 5

Methodology

This section will present the methodology used for the empirical analysis in order to evaluate and help answer the Research Question of whether higher-rated ESG firms produce higher abnormal returns than lower-rated firms. A number of approaches will be taken that centre around multi-factor models, including the Fama French three-factor model and Carhart's four-factor model.

5.1 Return Properties

In financial economics theory, the attention is essentially focused on developments in returns as opposed to prices. The reason for that is two-fold: on one hand, assuming that financial markets are perfectly competitive, the size of the investment does not affect price changes and thus returns offer a complete and scale-free summary of profitability of an investment opportunity. On the other hand, when working with empirical data, returns provide better statistical properties than prices [Lo et al., 1997].

In a broad definition, the return refers to the total monetary gain from holding an asset (or a portfolio of assets) over a given time period. Since the assets under consideration in this paper are portfolios comprised of stocks, return calculations are adjusted so that the return also captures any dividend payments paid during the holding period in addition to any capital gains achieved from price increases. This method of computing simple returns is often referred to as the *holding period return*

(HPR). To compute the HPR, all returns in this paper are calculated using the Total Return Index (RI)[1], extracted from Thomson Reuters Datastream database. The RI shows a theoretical growth in value of a share holding over a specified period, assuming that any dividends paid are re-invested to purchase additional units of equity. By looking at returns calculated from the RI, a more accurate representation of the stock's performance is displayed. By assuming dividends are reinvested, you effectively account for stocks that do not issue dividends and instead, reinvest their earnings within the underlying company. Including dividends will provide more fair conclusions when exploring the cross-sectional differences of stock returns. For example, ignoring dividends will imply that growth stocks, with large capital gains, will be inappropriately favoured over income stocks (e.g. utilities and mature industries) that pay high dividends [Brooks, 2012]. The holding period return is then simply calculated as:

$$R_t = \frac{RI_t}{RI_{t-1}} - 1 \tag{5.1}$$

The authors of this paper acknowledge that optimal statistical inference generally requires the use of log-returns given the convenient statistical properties[2]. However, the empirical analysis in this paper requires the use of simple returns such as the HPR given in equation (5.1). The reason for this is that the simple return of a portfolio of assets is a weighted average of the simple returns of the individual assets, where the weight of each asset is the share of the portfolio's value invested in that asset. If portfolio p places weight w_{ip} in asset i, then the return on the portfolio at time t, R_{pt}, is related to the returns on individual assets, R_{it}, by the following equation:

$$R_{pt} = \sum_{i=1}^{N} w_{ip} R_{it} \qquad \text{for} \qquad i = 1, 2, \ldots, N \tag{5.2}$$

Unfortunately log-returns do not share this convenient property. Since the log of a sum is not the same as the sum of logs, r_{pt}^{log} does not equal $\sum_{i=1}^{N} w_{ip} R_{it}$ because the operation of taking a log constitutes a non-linear transformation. Therefore all portfolio returns presented in this paper are calculated using equation (5.2), where the individual asset weights for the value- and equal-weighted portfolios respectively are defined as:

[1]For further details about the derivation of Thomson Reuters RI refer to Section B.1 in the Appendix

[2]log-returns are normally distributed, and remain so when added together in a multi-period setting. This simplifies any modelling of statistical behaviour of asset returns over time, as it is far easier to derive the time-series properties of additive processes than of multiplicative processes [Lo et al., 1997]

$$w_{i,t-1} = \frac{\mathrm{MV}_{i,t-1}}{\sum_{i=1}^{N} \mathrm{MV}_{i,t-1}} \quad (5.3) \qquad\qquad w_{i,t-1} = \frac{1}{N} \quad (5.4)$$

where w_i is the weight of the i^{th} security at time $t-1$, MV_i is the market value of the i^{th} security at time $t-1$, calculated as the share price multiplied by the number of ordinary shares in issue, and N is the total number of securities within the portfolio at time $t-1$.

Finally, in financial literature it is customary to express returns on an annualised basis, in order to make a fair comparison of returns on investments with differing time horizons. Since the research conducted in this paper involves monthly returns, all returns reported in this paper have been converted into compounded annualised returns, given by the following equation:

$$R_T = (1 + R_t)^{12} - 1 \tag{5.5}$$

where R_T denotes the return over the annual time period T, and R_t denotes the return over the monthly time period t.

5.2 Portfolio Construction

The portfolio construction method is a useful way of analysing the cross section of returns. For that reason, it is quite common in empirical literature to see the construction of ESG portfolios when investigating the link between sustainability and financial performance of firms. By sorting firms into portfolios based on their respective ESG ratings, it is possible to easily aggregate large panel data sets into a single time-series dimension, thus allowing for the application of basic asset pricing models. Additionally, the portfolio sorting method provides intuitive implications that can be used as key inputs when formulating an investment strategy based on ESG differences. The empirical framework used in this paper is largely in line with that of Kempf and Osthoff [2007] and Statman and Glushkov [2009].

Construction of the decile portfolios for the analysis is as follows. In each month t from December, 2005 to January, 2018, companies are sorted into ten portfolios according to the ESG score available at time $t-1$. The method of lagging the ESG scores by one month is to overcome look-ahead bias and to ensure that fair conclusions can be drawn, since any trading related to changes in ESG scores can only be made after the score has been published. As such, the resulting decile portfolios are ranked from 1 (worst) to 10 (best), where the lowest decile portfolio at any given time t contains the bottom $\sim 10\%$ of ESG-rated companies, and the highest decile portfolio at any given time t contains the top $\sim 10\%$ of ESG-rated companies. A total of 10 equally-weighted portfolios are formed monthly for each ESG, ESGC, and the individual pillar scores and thus there is monthly re-balancing for each portfolio. Furthermore, the firms in each portfolio are matched with the appropriate monthly excess return at time t, which results in a time series of monthly returns for the entire sample period. This results in a total of 50 portfolios, 10 for each specific rating, which all vary by their level of ratings. Finally, each decile portfolio consists of 145 monthly average excess returns, which were obtained by aggregating the individual stock returns.

5.3 Performance Evaluation

In order to assess the performance of portfolios sorted by ESG ratings, the first logical step is to see which portfolio produces the highest mean excess return. However, looking solely at the returns does not provide any meaningful information about the performance relative to the risk taken. Investors who seek a high return on their investment will seek out the highest possible returns accompanied with the lowest possible risk. As such, it is relevant to measure the expected return or risk premium relative to the level of risk taken. In order to compare the mean excess returns to the level of risk taken in terms of volatility (measured by its standard deviation), the paper will report the Sharpe ratio of each portfolio. In order to assess the overall excess abnormal return after being adjusted for common risk-factors related to firm characteristics, this paper mostly focus on looking at the alpha estimate from established multi-factor models. The following sections will elaborate further on these performance evaluation methods.

5.3.1 The Sharpe Ratio

A common method of quantifying the trade-off between risk and returns is the Sharpe ratio, introduced by William F. Sharpe [1966]. The Sharpe ratio measures the expected excess return per unit risk and is thus useful to provide a basis for economic interpretation of the performance of a given ESG-sorted portfolio. For any asset or portfolio p, the Sharpe ratio is defined as the expected excess return divided by the standard deviation of the future return. For the empirical analysis in this paper, the Sharpe ratio of portfolio p is adopted according to the definition by Jobson and Korkie [1981]:

$$\text{SR}_p = \frac{\hat{\mu}_p}{\hat{\sigma}_p} \tag{5.6}$$

where $\hat{\mu}_i$ and $\hat{\sigma}_i$ denote the mean and sample standard deviation of the time-series portfolio returns in excess of the periodic risk-free rate. The mean and standard deviation are then defined as:

$$\hat{\mu}_p = \frac{1}{T}\sum_{t=1}^{T} R_{pt} - R_{ft}, \qquad \hat{\sigma}_p = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(R_{pt} - R_{ft} - \hat{\mu}_i)^2} \tag{5.7}$$

The Sharpe ratio will be used for comparison purposes when looking at the mean excess returns of the different ESG-sorted portfolios. The 'best' portfolio is then taken to be the one with the highest Sharpe ratio.

5.3.2 Factor Models

The most frequently used portfolio performance measure is the alpha of the applied asset pricing model, which was first suggested by Jensen [1968]. In an empirical framework, the alpha refers to the intercept obtained from a linear regression and measures the part of the return that is unexplained by the systematic risk taken. As previously discussed, many researchers have concluded that factors other than the market-beta are relevant for explaining the cross-sections of expected returns. The alpha is therefore estimated with reference to the prevalent equilibrium models where the dependent variables chosen reflects many possible influences on expected returns. Assuming an investor holds an activity-managed portfolio p, then the performance measured by the alpha is given by the following multi-factor model.

$$R_p - R_f = \alpha_i + \sum_{k=1}^{K} \beta_{pk} F_k + \epsilon_p \qquad \epsilon_p \sim \text{IID}\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}) \tag{5.8}$$

where R_p is the return for portfolio p, R_f is the risk-free rate, α_i is the intercept of the factor model, F_k is the realisation of the common factor k, β_{pk} is the sensitivity of portfolio i to that factor, and ϵ_i the firm-specific disturbance. The multi-factor models are essentially an extended version of the CAPM relation between market-betas and expected returns. It is apparent from equation (5.8) that if $\alpha_p = 0$, then the the respective factor model successfully explains the returns of the portfolio. Hence, portfolio p earns a positive *risk-adjusted abnormal return* only if $\alpha_p > 0$. For $\alpha_p < 0$, the portfolio has underperformed relative to the beta risk of the portfolio.

5.3.2.1 Applied Models

The empirical analysis in this paper involves three multi-factor asset pricing models. The first model considered is Fama and French's [1993] classic three-factor model, which is estimated given the following equation:

$$R_{pt} - R_{ft} = \alpha_p + \beta_p \left(R_{mt} - R_{ft}\right) + s_p \mathrm{SMB}_t + h_p \mathrm{HML}_t + \epsilon_{pt} \tag{5.9}$$

where $R_{pt} - R_{ft}$ is the excess return of the portfolio p over the risk-free interest rate in month t, $R_{mt} - R_{ft}$ is the excess return of the value-weighted (VW) market portfolio, SMB_t is the return on a diversified portfolio of small stocks minus the return on a diversified portfolio of big stocks, HML_t is the difference between the returns on diversified portfolios of high and low book-to-market stocks. Fama and French [1996] showed that the model gives a good fit of US stock market data over the period 1963-1993, and later found that the same is true for many other countries [Fama & French, 2012]. These findings are in line with the previous findings from Hawawini and Keim [1995], who also reported that the size and value effects have been shown to be present in a range of international stock returns.

The second model is the Carhart [1997] four-factor model, that was developed to help explain the momentum anomaly[3], first identified by Jegadeesh and Titman [1993]. The prediction of the momentum effect is that recent winners among stocks continue to do well over the next few months, and recent losers continue to do poorly. In cross-sectional regressions, recent returns showed up statistically significant. These findings led Carhart [1997] to suggest the momentum factor as an

[3] In financial literature, the term anomaly refers to a return pattern unexplained by the CAPM

extension to the Fama French three-factor model. The model is given by the following equation:

$$R_{it} - R_{ft} = \alpha_i + \beta_i \left(R_{mt} - R_{ft}\right) + s_i \mathrm{SMB}_t + h_i \mathrm{HML}_t + w_i \mathrm{WML}_t + \epsilon_{it} \tag{5.10}$$

where WML_t is the added momentum factor, given as the return on a winners-minus-losers strategy, or more precisely the difference between the future return on a portfolio of recent "winners" and the future return on a portfolio of recent "losers". The model will thus clarify if any significant differences in returns between high and low-rated ESG portfolios can be explained by investigating the momentum factor. Additionally, it has been the most widely used model in ESG literature.

5.4 Hypothesis Testing and Inference

The alpha, which is the intercept, is calculated by applying the aforementioned models. The hypothesis of this paper is that an abnormal return will be observed based on the level of rating, which means that the intercepts of the regressions have to be statistically significantly different from zero. A positive alpha will indicate outperformance while a negative alpha indicates underperformance of the portfolios. Thus, the main statistical test will be to see if the alpha value is different from zero.

In the hypothesis testing framework, there are always two hypotheses that go together, known as the null hypothesis (denoted H_0) and the alternative hypothesis (denoted H_A). The null hypothesis is the statement or the statistical hypothesis that is actually being tested. The alternative hypothesis represents the remaining outcomes of interest [Brooks, 2012]. This paper will make use of this hypothesis framework to evaluate the statement that the intercept estimate (alpha) obtained from the time-series regression on the factor models is significantly different from zero and this will be applied to all ESG scores. The hypothesis is then formally stated in the following way:

$$H_0 : \alpha_i = 0 \qquad H_A : \alpha_i \neq 0 \qquad \forall_i = 1, \ldots N \tag{5.11}$$

This is known as a two-sided test, since the outcomes of both $\alpha > 0$ and $\alpha < 0$ are both subsumed under the alternative hypothesis. The two-sided hypothesis testing method is of interest in order to test the Arbitrage Pricing Theory, which assumes that after being controlled for risk factors, the intercept should be equal to zero. If the alpha is significantly different from zero, the H_0 is rejected and subsequently the H_A is accepted. Should that be the case, the implication of the factor model theory is that the market is not efficient and an arbitrage opportunity is present [Brooks, 2012].

5.5 Ordinary Least Square

The empirical analysis used in this paper relies heavily on the Ordinary Least Square (OLS) regression method in order to examine the performance of the respective ESG portfolios. The OLS approach is the most common regression method used to fit a line to the data and remains the workhorse of econometric model estimation [Brooks, 2012, p. 31]. The OLS entails that the classical linear regression model that describes the relationship between the dependent variable y and one or more independent variables x_i. For the k factor models used in this paper, the OLS can be written as

$$y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_k x_{kt} + \ldots + \epsilon_t \qquad t = 1, 2, \ldots, T \tag{5.12}$$

where the variables $x_{1t}, x_{2t} \ldots x_{kt}$ represent the set of k explanatory variables from the previously discussed factor models which are thought to influence y (in this case, the excess holding period return). The coefficient estimates $\beta_1, \beta_2 \ldots \beta_k$ are the parameters which quantify the effect of each of these explanatory variables on y. Each coefficient is known as a partial regression coefficient, interpreted as representing the partial effect of the given explanatory variable on the explained variable, after holding constant or eliminating the effect of all other explanatory variables [Brooks, 2012]. For example, in the Fama French three-factor model in (5.9), the estimated coefficient $\bar{s}_p$ measures the effect of the SMB size factor on the excess return $R_p - R_f$ after eliminating the effects of the market $R_m - R_f$ and the HML value factor. Furthermore, the OLS method entails taking each vertical distance from the data point to the line, squaring it and then minimising the total sum of the area of squares, hence the name 'least squares'.

5.5.1 OLS Assumptions

The validity of any obtained OLS estimators is conditional on a set of assumptions regarding the random disturbance term. These assumptions can be summarised as follows:

1. $\mathrm{E}[\epsilon_t] = 0$, i.e. the estimated error terms are assumed to have zero mean
2. $\mathrm{Var}[\epsilon_t] = \sigma^2 < \infty$, i.e. the variance of the errors is constant and finite
3. $\mathrm{Cov}[\epsilon_i, \epsilon_j] = 0$, meaning that the errors are linearly independent of one another.
4. $\mathrm{Cov}[\epsilon_i, x_t] = 0$, i.e. no relationship between the error and corresponding x variate
5. $\epsilon \sim \mathrm{N}(0, \sigma^2)$, i.e. the error is normally distributed.

The fifth OLS assumption is required to make valid inferences about the actual α and β from the sample parameters $\hat{\alpha}$ and $\hat{\beta}$, estimated using a finite amount of data. Should the model be free of any violations of assumptions 1-4, the coefficient estimates obtained from the OLS regression method will provide some desirable properties, referred to as Best Linear Unbiased Estimators (BLUE):

- *Best*, meaning that the OLS estimator $\hat{\beta}$ contains minimum variance between the class of linear unbiased estimators.
- *Linear*, meaning $\hat{\alpha}$ and $\hat{\beta}$ are linear estimators, i.e. the formulae for $\hat{\alpha}$ and $\hat{\beta}$ are linear combinations of the random variables (in this case, y)
- *Unbiased*, meaning that the actual values of $\hat{\alpha}$ and $\hat{\beta}$ will on average be equal to their true values
- *Estimator*, meaning that the $\hat{\alpha}$ and $\hat{\beta}$ are estimators of the true value of α and β

As such, any presence of significant violations of the OLS model would indicate that some of the coefficient estimates should be interpreted with caution. Since all of the regression models include a constant term, the first assumption that the errors are zero can never be violated [Brooks, 2012]. A series of statistical robustness tests will be conducted in order to address the assumptions of the model's used in this paper.

5.5.2 OLS Diagnostic Tests

A number of OLS diagnostic tests will be utilised in order to see if the data sample is appropriate for the analysis and it will further enhance validity of the models used in this paper. The first is a test for heteroscedasticity. The OLS regression model assumes that the variance of the errors is constant: $\mathrm{Var}[\epsilon_t] = \sigma^2 < \infty$. If the errors do not have a constant variance, they are said to be "heteroscedastic". Should we ignore a presence of hederoscedasticity, OLS estimators will still give unbiased and consistent coefficient estimates, but they will no longer be BLUE. This is because the error variance, σ^2, plays no part in the proof that the OLS estimator is consistent and unbiased, but σ^2 does appear in the formulae for the coefficient variances. If the errors are heteroscedastic, the formulae presented for the coefficient standard errors no longer hold [Brooks, 2012]. To test this assumption, the Breusch and Pagan [1979] / Cook and Weisberg [1983] tests for heteroscedasticity are performed. The test assumes that the ϵ_t are normally distributed under the null hypothesis which implies that the score test statistic is equal to the model sum of squares from the augmented regression divided by two. Under H_0, the test statistic has the χ^2 distribution with $p-1$ degree of freedom.

The second is a test for autocorrelation. The OLS regression model assumes that the errors are uncorrelated with one another, and if they are not, it would be stated that they are "autocorrelated." The consequences of ignoring a presence of autocorrelation are similar to those of ignoring heteroscedasticity. The coefficient estimates derived using OLS will still be unbiased, but they are inefficient (i.e. not BLUE), even at large sample sizes, so that the standard error estimates could be wrong. A test of this assumption is therefore required and the Breusch–Godfrey test is applied in this paper, which is a more general test for autocorrelation up to the p^{th} order (p = number of lags of the residuals). The test is a Lagrange Multiplier test of the null hypothesis of no autocorrelation versus the alternative that ϵ_t follows an AR(p) or MA(p) process.

The last test is to check for normality. The OLS regression model assumes that the disturbances are normally distributed: $\epsilon_t \sim N(0, \sigma^2)$. In order to test this assumption the Bera-Jarque [1981] test is conducted, which describes whether the coefficient of skewness and the coefficient of excess kurtosis are jointly zero [Brooks, 2012]. The null hypothesis of normality is rejected in the case the residuals from the model are either significantly skewed or leptokurtic and/or platykurtic.

Chapter 6

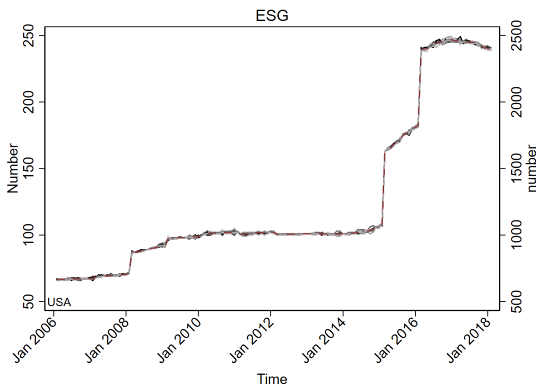
Analysis of Results

This chapter will present a comprehensive analysis of the various test results from the empirical models outlined in the previous sections. The chapter will begin with an overview of the summary statistics of the respective samples, which will serve as the first step of the analysis. The following sections will provide a detailed analysis of the different portfolio constructions, each starting with the Fama French three-factor model before moving on to the Carhart four-factor model. First, the decile portfolios sorted on ESG ratings will be examined. Then, portfolios sorted first on size and then ESG ratings will follow. Finally, portfolios constructed with a long-short strategy approach will conclude the portfolio analysis. Each region will be analysed separately, while comparative discussion will be addressed in Chapter 7. The chapter will end with an analysis of the econometric methods applied, which further investigates the reliability of the test results in the previous sections.

6.1 Descriptive Statistics

6.1.1 Overall Summary Statistics

For illustrative purposes, the descriptive statistics of the decile portfolios are presented for a first impression of the differences between ESG ratings and mean excess returns. The decile portfolios are rebalanced monthly based on each ESG score and the annualised mean, standard deviation, t-statistic



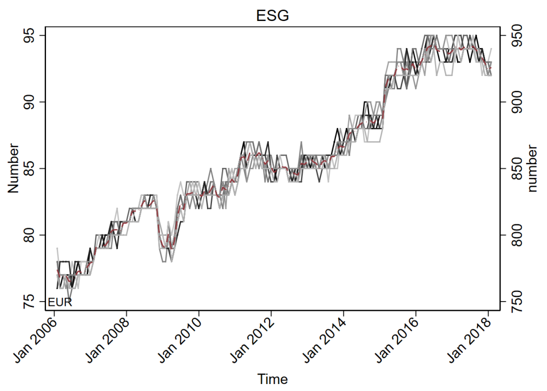


FIGURE 6.1: Historical development in number of stocks within each decile portfolio for US & Europe

and Sharpe ratio are reported. Although a very simple level analysis, it gives a general overview of the portfolios that will be examined in more detail later on.

The distribution of the number of firms in the US sample can be found in Figure 6.1. On the left axis, the number of firms in each decile is shown while the overall total amount of firms in the entire sample can be found on the right axis. There is a line for each decile portfolio but since they are so closely aligned it is difficult to see the difference. Likewise, the same distribution for the European sample can be found in Figure 6.1. It can be seen that at the start of the sample period, the US sample contained around 60 ESG-rated firms on average in each decile, while the European decile portfolios contained around 75 firms, with the numbers steadily increasing throughout the years.

Additionally, the descriptive statistics of all portfolios and ratings are summarised in Table 6.1 for both the US and the European sample. Most of the ratings for both regions are similar in their distributions. However, the European sample does appear to have slightly higher ratings than the US throughout all portfolios, which suggest that on average, European stocks are generally slightly more sustainable. Finally, the average market capitalisation of firms in each decile portfolio is presented in Table 6.2. The table reveals stark differences between the deciles in terms of average company size, suggesting a strong correlation between larger companies and higher ESG ratings. However, this is true for all scores except the ESG Combined, where there seems to be more of a balance in size across the sorted portfolios. This relationship between size and ESG ratings will be taken into consideration in the following sections and further discussed following the analysis.

TABLE 6.1: Descriptive Statistics of Ratings for the US and Europe Sample

Portfolio	**1 (low)**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10 (high)**
					US					
ESG										
mean	23.65	30.60	35.14	39.14	43.40	48.35	53.98	60.80	69.14	80.88
min	8.54	26.61	31.63	35.32	38.59	42.54	46.74	52.82	59.86	70.83
max	32.32	36.62	40.26	43.88	50.05	55.83	61.66	68.26	76.82	98.00
ESG Combined										
mean	21.63	27.66	31.50	34.82	37.80	40.95	44.51	49.38	57.06	71.89
min	7.15	23.62	28.46	32.13	35.34	38.42	41.50	44.67	50.01	58.28
max	27.90	31.56	34.89	37.82	40.43	44.36	48.70	56.11	66.97	95.67
Environmental										
mean	15.95	22.73	27.22	32.15	37.75	44.23	52.04	61.69	72.88	87.07
min	2.92	18.60	23.78	28.36	32.49	36.72	41.17	47.87	58.62	74.00
max	30.27	34.10	37.00	39.31	48.22	57.49	64.49	73.99	83.00	98.89
Social										
mean	20.73	29.09	34.59	39.48	44.55	49.95	56.02	63.09	71.72	85.28
min	3.73	23.56	29.52	33.91	38.13	42.75	48.15	53.81	61.48	72.89
max	30.07	35.71	42.04	46.88	51.52	57.04	64.19	71.92	81.11	98.94
Governance										
mean	14.72	25.48	33.02	40.48	47.67	54.31	61.11	68.21	76.04	86.50
min	1.85	19.67	27.55	35.21	43.08	48.34	55.63	62.87	70.18	77.14
max	23.18	30.84	38.58	45.63	52.33	59.05	65.78	73.76	82.40	99.32
					Europe					
ESG										
mean	28.24	40.00	46.40	51.57	56.53	61.10	65.58	70.45	75.78	83.95
min	7.83	31.98	38.90	43.27	48.19	52.05	56.04	60.67	66.69	72.95
max	39.39	47.51	52.68	57.69	62.70	66.51	70.92	75.39	81.75	96.75
ESG Combined										
mean	25.62	34.34	39.22	42.94	46.63	51.00	56.50	62.42	69.03	78.57
min	7.83	29.34	34.39	38.36	41.79	45.47	49.14	54.45	60.41	67.87
max	32.60	38.65	42.90	47.56	52.72	59.08	65.15	70.26	77.79	95.21
Environmental										
mean	25.79	39.34	48.12	54.98	61.21	67.17	73.29	79.34	85.21	92.39
min	5.03	27.02	36.19	40.53	45.02	51.79	58.40	66.57	73.18	79.06
max	39.50	50.92	57.68	63.52	68.76	74.46	79.56	85.42	91.01	99.25
Social										
mean	22.75	36.98	45.51	52.60	58.86	64.32	69.95	75.65	82.15	90.72
min	4.93	27.01	35.80	43.83	48.82	54.92	59.52	65.52	72.57	81.18
max	36.24	46.97	54.10	60.84	66.24	71.38	76.30	82.24	88.51	99.41
Governance										
mean	15.39	26.27	33.64	40.64	47.48	54.14	60.76	67.62	74.96	85.42
min	1.05	20.37	28.59	34.86	42.13	49.01	55.80	62.28	68.56	76.67
max	22.80	31.64	38.56	45.48	51.87	59.51	65.91	73.15	80.86	98.89

This table reports the mean, minimum and maximum of every ESG rating for each decile portfolio. The 1^{st} decile holds the lowest-rated portfolios and the 10^{th} decile holds the highest-rated portfolios.

Table 6.2: Average Market Capitalisation of Firms in Decile Portfolios for US & Europe

US Sample

Portfolio	**1 (low)**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10 (high)**
					US					
ESG	4,374	4,162	4,480	5,373	5,796	6,844	10,000	14,627	24,554	54,304
ESGC	6,565	6,917	8,365	10,745	16,921	20,255	22,482	16,287	9,683	16,161
ENV	4,074	5,224	4,966	5,220	5,756	7,463	10,479	15,203	25,936	50,171
SOC	3,740	5,011	4,839	5,462	6,553	7,986	10,384	14,785	23,983	51,750
GOV	5,477	6,416	7,328	7,616	9,813	13,288	14,277	16,585	20,322	33,299
					Europe					
ESG	3,187	3,444	3,708	5,137	7,062	8,545	10,498	14,259	22,751	41,488
ESGC	4,840	9,836	16,134	21,394	18,296	10,882	6,867	6,612	9,442	15,615
ENV	3,223	3,742	5,084	6,356	8,493	10,720	13,229	17,233	21,955	29,955
SOC	3,248	3,623	4,418	5,408	6,759	8,452	11,884	16,401	21,969	37,867
GOV	4,646	5,971	7,401	8,292	8,911	10,586	12,895	13,296	21,521	26,432

This table reports the average market capitalisation (share price x ordinary shares issued) within each respective ESG-rated decile portfolio. All values are presented in $ millions.

6.1.2 US Sample

Table 6.3 presents the descriptive statistics of the decile portfolios for the US sample. First, for the equally-weighted US sample, some interesting results appear. What is quite surprising is that for every score and pillar, the highest-rated portfolios have lower annualised mean excess returns when compared with the bottom-rated. For the ESG-rated portfolios, the two highest means of 15.69% are found in the bottom-rated and 4th decile portfolios while the lowest mean of 11.42% is in the top-rated portfolio. As for ESGC, the bottom portfolio has the highest excess return of 14.96% but now the two lowest mean excess returns are found in the 2nd and 3rd portfolios. For the Environmental score, the highest mean of 15.15% can be found in the 8th portfolio while the 9th and 10th contain the two lowest means. For both Social and Governance, the bottom-rated portfolio has the highest mean while the lowest can be found in the 9th portfolio for Social and the top-rated for Governance. This is rather surprising and suggests that good governance and social responsibility does not translate into the best return but actually quite the contrary. Overall, there doesn't appear to be a trend for the means of portfolios as the ratings increase, although the means of the lowest-rated portfolios are consistently above those of the top-rated. On the other hand, the standard deviation of the top-rated portfolios is consistently lower than the bottom-rated, which could give an explanation for the lower excess return.

However, no other clear trend emerges of the volatility decreasing as the ratings increase, with many middle-rated portfolios having higher volatility than the lower-rated. Thus, it appears the difference in volatility is confined only to the top and bottom 10% of ratings. Lastly, the highest Sharpe ratios (SR) can be found in the bottom-rated portfolios for ESG and ESGC. For the top-rated portfolios, the highest SR can only be found for the Social rating while other high ratios can be found randomly throughout, with no clear trend evident between the lower and higher-rated deciles.

Next, for the value-weighted US sample, the mean returns are much lower which could imply that the returns for the equal-weighted sample are tilted toward small stocks which are known to bring higher returns. For the ESG score, the mean ranges from 6.91% to 12.40%. Again, for all scores, each bottom-rated portfolio has a higher mean than the top-rated. However, now the bottom portfolio does not have the highest mean, which can be found randomly in the 4^{th} portfolio for ESG, 9^{th} for ESGC, 8^{th} for Environmental, lowest-rated for Social and the 2^{nd} portfolio for Governance. Similarly to the equal-weighted sample, the top-rated portfolios for each score have lower volatility than the bottom although it is not consistently decreasing as the ratings increase and once again, no clear trend emerges. The highest SR can also be found randomly throughout, for the 4^{th} and 7^{th} portfolio for ESG, 9^{th} for ESGC, 8^{th} for Environmental, lowest-rated for Social and 9^{th} for Governance. It does appear to be the highest in most of the higher-rated firms, although not consistently and again no clear trend emerges.

6.1.3 Europe Sample

Table 6.4 presents the descriptive statistics of the decile portfolios for the European sample, where similar results can be found as with the US. However, mostly all the means are lower than the US sample which implies that the excess returns associated with ESG are lower in Europe. First, for the equally-weighted sample, all of the lowest-rated portfolios again have a higher Sharpe ratio and mean than the top-rated ones. That is, with the exception of the ESGC score, where the top portfolio has a mean of 10.28% and the lowest one has a mean of 9.19%. This may imply that the Controversy score that is added to the regular ESG score has quite an impact on the portfolio's mean return, since the two lowest-rated portfolios actually have the lowest mean return of any other portfolio. Just as with the US sample, no clear pattern emerges and a number of randomly priced portfolios have the highest

TABLE 6.3: Descriptive Statistics for Decile Portfolios for USA

Portfolio	**1 (low)**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10 (high)**
Equal-weighted										
ESG										
mean	0.1569	0.1252	0.1360	0.1569	0.1279	0.1189	0.1428	0.1169	0.1257	0.1142
st. dev	0.1845	0.1847	0.1944	0.1927	0.1976	0.1879	0.1929	0.1914	0.1707	0.1625
t-stat	2.76	2.23	2.29	2.65	2.13	2.09	2.42	2.02	2.42	2.32
SR	0.85	0.68	0.70	0.81	0.65	0.63	0.74	0.61	0.74	0.70
ESG Combined										
mean	0.1496	0.1125	0.1089	0.1473	0.1392	0.1286	0.1431	0.1228	0.1404	0.1299
st. dev	0.1887	0.1818	0.1863	0.1940	0.1851	0.1823	0.1858	0.1751	0.1965	0.1810
t-stat	2.58	2.05	1.94	2.48	2.46	2.32	2.52	2.31	2.34	2.36
SR	0.79	0.62	0.58	0.76	0.75	0.71	0.77	0.70	0.71	0.72
Environmental										
mean	0.1306	0.1377	0.1250	0.1389	0.1392	0.1477	0.1304	0.1515	0.1103	0.1102
st. dev	0.1879	0.1851	0.1913	0.1911	0.1873	0.1899	0.2006	0.1905	0.1661	0.1731
t-stat	2.28	2.44	2.15	2.38	2.43	2.54	2.14	2.59	2.20	2.11
SR	0.70	0.74	0.65	0.73	0.74	0.78	0.65	0.80	0.66	0.64
Social										
mean	0.1415	0.1361	0.1332	0.1331	0.1292	0.1284	0.1338	0.1367	0.1183	0.1302
st. dev	0.1852	0.1845	0.1924	0.1913	0.1897	0.1913	0.1849	0.1926	0.1848	0.1594
t-stat	2.50	2.42	2.27	2.28	2.24	2.21	2.37	2.32	2.11	2.68
SR	0.76	0.74	0.69	0.70	0.68	0.67	0.72	0.71	0.64	0.82
Governance										
mean	0.1603	0.1499	0.1417	0.1451	0.1188	0.1159	0.1282	0.1272	0.1195	0.1148
st. dev	0.2034	0.1922	0.1776	0.2019	0.1854	0.1857	0.1908	0.1778	0.1773	0.1665
t-stat	2.56	2.54	2.61	2.35	2.12	2.06	2.21	2.35	2.22	2.28
SR	0.79	0.78	0.80	0.72	0.64	0.62	0.67	0.72	0.67	0.69
Value-weighted										
ESG										
mean	0.1159	0.0898	0.1010	0.1240	0.1004	0.0691	0.1122	0.0968	0.0956	0.0893
st. dev	0.1606	0.1644	0.1647	0.1671	0.1734	0.1711	0.1522	0.1606	0.1459	0.1350
t-stat	2.39	1.82	2.04	2.44	1.93	1.36	2.44	2.01	2.18	2.21
SR	0.72	0.55	0.61	0.74	0.58	0.40	0.74	0.60	0.66	0.66
ESG Combined										
mean	0.1222	0.0854	0.0720	0.1096	0.0886	0.0926	0.0989	0.0855	0.1325	0.0929
st. dev	0.1602	0.1578	0.1712	0.1611	0.1460	0.1461	0.1409	0.1459	0.1604	0.1558
t-stat	2.51	1.81	1.41	2.25	2.03	2.11	2.34	1.96	2.71	1.99
SR	0.76	0.54	0.42	0.68	0.61	0.63	0.70	0.59	0.83	0.60
Environmental										
mean	0.0973	0.1043	0.0784	0.1057	0.1155	0.1029	0.0892	0.1208	0.0844	0.0904
st. dev	0.1653	0.1483	0.1728	0.1728	0.1598	0.1645	0.1613	0.1440	0.1299	0.1498
t-stat	1.96	2.34	1.52	2.03	2.39	2.08	1.85	2.77	2.18	2.02
SR	0.59	0.70	0.45	0.61	0.72	0.63	0.55	0.84	0.65	0.60
Social										
mean	0.1284	0.1006	0.0883	0.0969	0.0897	0.0948	0.0973	0.1054	0.0935	0.0896
st. dev	0.1686	0.1524	0.1620	0.1665	0.1691	0.1611	0.1619	0.1583	0.1499	0.1345
t-stat	2.50	2.20	1.82	1.94	1.77	1.96	2.00	2.21	2.08	2.23
SR	0.76	0.66	0.55	0.58	0.53	0.59	0.60	0.67	0.62	0.67
Governance										
mean	0.0978	0.1186	0.0910	0.1162	0.0876	0.0811	0.0936	0.0873	0.0970	0.0960
st. dev	0.1690	0.1667	0.1595	0.1758	0.1670	0.1534	0.1552	0.1456	0.1311	0.1412
t-stat	1.93	2.35	1.91	2.18	1.75	1.77	2.01	2.00	2.46	2.27
SR	0.58	0.71	0.57	0.66	0.52	0.53	0.60	0.60	0.74	0.68

This table reports the annualised mean excess return, annualised standard deviation and Sharpe ratios (SR) of deciles portfolios formed by their ESG rating. The t-statistic is obtained by doing a one sample t-test which tests the excess return series for a null hypothesis of zero mean. The statistics are calculated using an equally-weighted sample and also a value-weighted by the market capitalisation of firms. The 1^{st} decile holds the lowest-rated portfolios and the 10^{th} decile holds the highest-rated portfolios.

mean excess return. This includes the 4^{th} portfolio for the ESG score, the 6^{th} for ESGC, the 3^{rd} for Environmental, the 3^{rd} for the Social, and finally, the 3^{rd} for the Governance. What is interesting is that the 3^{rd} portfolio for each individual pillar is the one with the highest mean.

For the value-weighted sample, the same pattern appears. Every mean for the top-rated portfolios is lower than the bottom-rated one. However, once again, the ESGC portfolio has a reversed difference in mean returns that is quite large, with the bottom portfolio having a mean of 4.99% and the top one has a mean of 8.04%. Once again, this shows how important the Controversy score is, particularly in Europe. Except for Environmental and Governance, the top-rated portfolios have a lower standard deviation than the lowest-rated, similarly to the US sample. The highest Sharpe ratios can be found randomly, in the 6^{th} portfolio for ESG, 7^{th} for ESGC, 5^{th} for Environmental, 6^{th} for Social and 4^{th} portfolio for Governance. Again, no clear pattern emerges and the highest means as well as SR's can be found randomly throughout the various levels of ratings.

Table 6.4: Descriptive Statistics for Decile Portfolios for Europe

Portfolio	**1 (low)**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10 (high)**
					Equal-weighted					
ESG										
mean	0.1036	0.1098	0.1197	0.1289	0.0883	0.1010	0.0899	0.1041	0.0995	0.0881
st. dev	0.2149	0.2221	0.2153	0.2205	0.2128	0.2176	0.2252	0.2202	0.2080	0.2153
t-stat	1.60	1.64	1.83	1.92	1.39	1.54	1.33	1.57	1.59	1.37
SR	0.48	0.49	0.56	0.58	0.41	0.46	0.40	0.47	0.48	0.41
ESG Combined										
mean	0.0919	0.0893	0.0999	0.1007	0.1153	0.1276	0.1007	0.0949	0.1102	0.1028
st. dev	0.2192	0.2173	0.2210	0.2050	0.2169	0.2157	0.2151	0.2178	0.2253	0.2151
t-stat	1.40	1.37	1.50	1.63	1.76	1.94	1.56	1.45	1.62	1.59
SR	0.42	0.41	0.45	0.49	0.53	0.59	0.47	0.44	0.49	0.48
Environmental										
mean	0.0989	0.1138	0.1345	0.0958	0.0983	0.1214	0.0976	0.0864	0.0926	0.0948
st. dev	0.2205	0.2268	0.2093	0.2165	0.2135	0.2109	0.2106	0.2180	0.2153	0.2315
t-stat	1.49	1.66	2.11	1.47	1.53	1.90	1.54	1.33	1.44	1.37
SR	0.45	0.50	0.64	0.44	0.46	0.58	0.46	0.40	0.43	0.41
Social										
mean	0.1154	0.0992	0.1177	0.1015	0.1004	0.1119	0.1018	0.1004	0.0956	0.0902
st. dev	0.2162	0.2261	0.2142	0.2197	0.2212	0.2187	0.2195	0.2120	0.2146	0.2129
t-stat	1.76	1.46	1.81	1.54	1.51	1.69	1.54	1.57	1.48	1.41
SR	0.53	0.44	0.55	0.46	0.45	0.51	0.46	0.47	0.45	0.42
Governance										
mean	0.1149	0.1040	0.1156	0.1033	0.0830	0.1085	0.1102	0.0955	0.0995	0.0999
st. dev	0.2073	0.2234	0.2147	0.2213	0.2256	0.2150	0.2176	0.2094	0.2155	0.2168
t-stat	1.83	1.55	1.78	1.55	1.23	1.67	1.68	1.52	1.54	1.53
SR	0.55	0.47	0.54	0.47	0.37	0.50	0.51	0.46	0.46	0.46
					Value-weighted					
ESG										
mean	0.0895	0.0894	0.0883	0.1071	0.0839	0.0937	0.0655	0.0874	0.0665	0.0622
st. dev	0.2068	0.2058	0.2074	0.2155	0.2052	0.1845	0.2106	0.1985	0.1825	0.2020
t-stat	1.45	1.45	1.42	1.65	1.37	1.69	1.05	1.47	1.23	1.04
SR	0.43	0.43	0.43	0.50	0.41	0.51	0.31	0.44	0.36	0.31
ESG Combined										
mean	0.0499	0.0791	0.0540	0.0601	0.0708	0.0949	0.1152	0.0723	0.1024	0.0804
st. dev	0.2131	0.1892	0.1942	0.1887	0.2025	0.1994	0.2071	0.2016	0.2096	0.2070
t-stat	0.80	1.40	0.94	1.08	1.18	1.59	1.84	1.21	1.62	1.30
SR	0.23	0.42	0.28	0.32	0.35	0.48	0.56	0.36	0.49	0.39
Environmental										
mean	0.0747	0.0729	0.1189	0.0848	0.1096	0.0945	0.0836	0.0692	0.0473	0.0668
st. dev	0.2138	0.2095	0.1969	0.2003	0.1797	0.1958	0.1787	0.1923	0.1938	0.2176
t-stat	1.17	1.17	1.99	1.42	2.02	1.61	1.57	1.21	0.83	1.04
SR	0.35	0.35	0.60	0.42	0.61	0.48	0.47	0.36	0.24	0.31
Social										
mean	0.0960	0.0842	0.0784	0.0862	0.0788	0.0942	0.0675	0.0692	0.0732	0.0692
st. dev	0.2109	0.2235	0.2078	0.1987	0.2084	0.2026	0.2007	0.1986	0.2015	0.1896
t-stat	1.52	1.26	1.27	1.45	1.27	1.55	1.13	1.17	1.22	1.23
SR	0.46	0.38	0.38	0.43	0.38	0.47	0.34	0.35	0.36	0.36
Governance										
mean	0.0852	0.1083	0.0867	0.1107	0.0479	0.0724	0.0669	0.0675	0.0622	0.0727
st. dev	0.1987	0.2039	0.2025	0.2036	0.2119	0.1882	0.2026	0.1837	0.2028	0.2014
t-stat	1.44	1.76	1.43	1.80	0.77	1.30	1.11	1.24	1.04	1.22
SR	0.43	0.53	0.43	0.54	0.23	0.38	0.33	0.37	0.31	0.36

This table reports the annualised mean excess return, annualised standard deviation and Sharpe ratios (SR) of deciles portfolios formed by their ESG rating. The t-statistic is obtained by doing a one sample t-test which tests the excess return series for a null hypothesis of zero mean. The statistics are calculated using an equally-weighted sample and also a value-weighted by the market capitalisation of firms. The 1^{st} decile holds the lowest-rated portfolios and the 10^{th} decile holds the highest-rated portfolios.

6.2 Factor Models Regression Analysis

In this section, the decile portfolios are further evaluated by multi-factor models in order to measure the performance of high-rated and low-rated portfolios. In particular, Fama and French's three-factor model and Carhart's four-factor model are applied in order to see the different attributes of the portfolios which vary by their ESG rating and the factor loadings for various levels of scores. It is also interesting to explore whether the high-rated and low-rated firms vary in their loadings to the traditional risk factors in order to further explore the differences in returns. The equally weighted portfolios will be the focus of this section since they give a higher excess return when compared with the value-weighted strategy.

6.2.1 Fama French Three-Factor Model

The first basis for the multi-factor regression analysis is the Fama French three-factor model. Although very popular in ESG literature, it does have some drawbacks and the factors have been shown to not explain the variation in returns as much as some other models which will come later and which will be further tested for the analysis.

6.2.1.1 US Sample

Table 6.5 presents the three-factor model regression results for the decile portfolios for the US sample. The alpha, which represents the annualised abnormal return of the portfolios, is of particular focus since it shows the out or under-performance of the portfolio relative to the market portfolio. Some surprising results are seen, since the bottom-rated portfolios generally have significant and positive annualised abnormal returns while the top-rated are only significant in one case – for the Social score. Thus, the results indicate that the high-rated portfolios do not outperform the low-rated but quite the contrary. The lowest-rated portfolios all have positive and significant alpha values ranging from 3.13% for the Social score to 4.58% for the ESG score, while the only insignificant lowest-rated alpha is observed in the Environmental portfolio. This implies that alpha is statistically different from zero for the majority of the lowest-rated portfolios. On the other hand, for the top-rated portfolios, only one highly significant and positive alpha of 2.75% can be found for the Social rating. Only two other portfolios have positive alphas which are significant below the 1% level and that is for the 4^{th} portfolio for ESG and the 3^{rd} portfolio for Governance. A few other significant alphas with p-values below 5% can be found randomly, such as those in the 5^{th} portfolio for ESGC, the 2^{nd}, 6^{th} and 8^{th} for Environmental, and the 2^{nd} portfolio for Governance. For all the alphas that are not significant, we cannot reject the null hypothesis that the alpha is different from zero and thus cannot conclude anything about them. Additionally, no clear pattern emerges of the high-rated firms having significant alphas and, on the contrary, the lower-rated portfolios are usually the ones with a significant and positive abnormal return. This implies that the lower-rated portfolios generally outperform the higher-rated for all ratings in US, with the only exception being the Environmental score.

The results also show that all portfolios, regardless of the rating, have a highly significant loading on the market factor which is above 1, with p-values consistently below 1%. A market exposure of one implies that the returns move directly in line with the market and thus also their volatility. The higher loading on the market factor implies the portfolios are more volatile than the market while it's also interesting to note that for ESG, ESGC and Environmental portfolios, the market factor is higher for the top portfolio and lower for the bottom-rated portfolio, while the opposite is true for the Social and Governance portfolios. The size factor (SMB) is also highly significant for each rating and all portfolios have p-values below 1%. Although all size factor loadings are positive, they decrease substantially as the ratings go up, implying that the low-rated portfolios are much smaller than the top-rated and thus the reason for the higher returns in the lowest portfolios might be the size premium, first identified by Banz [1981], who found a premium associated with smaller stocks as opposed to larger ones.

The HML loading, otherwise known as the value premium, has mixed results. It is actually significantly positive for many of the ESG portfolios with the exception of the 1^{st}, 3^{rd} and 7^{th} portfolios. The positive loading implies that these portfolios tend to be value stocks with high book-to-market ratios and thus there is a value premium earned for them. For the ESG score, the highest HML factor loading of 0.1515 is actually found in the 5^{th} portfolio, with the 4^{th} portfolio closely behind. Additionally, for the ESGC portfolios, the same pattern appears with the highest HML loadings for the middle portfolios, from 4^{th} to 7^{th}, while the top portfolio also has a significantly positive loading and the bottom-rated is insignificant. This further asserts that the middle and higher-rated portfolios appear to be compromised of value stocks. For the Social and Governance portfolio, the top two portfolios have highly significant loadings. However, for the Environmental portfolios the opposite is true, where the three lowest-rated portfolios have a significantly positive HML loading. Again, mixed results are evident and no clear pattern for the differences in factor loadings between high and low-rated stocks emerges with the exception of the size factor.

Table 6.5: Time-Series Regressions of ESG-Rated Portfolios on the Fama French Three-Factor Model for USA

	1 (low)	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10 (high)**
ESG										
alpha	0.0458***	0.0174	0.0211	0.0419***	0.0126	0.0079	0.0263	0.0023	0.0200	0.0111
R_m-R_f	1.0484***	1.0563***	1.1026***	1.0953***	1.1411***	1.0948***	1.1392***	1.1575***	1.0586***	1.0598***
SMB	0.5935***	0.5099***	0.6393***	0.598***	0.4971***	0.5126***	0.4261***	0.3368***	0.2488***	0.0752**
HML	0.0382	0.1000*	0.0244	0.1202***	0.1515***	0.078*	0.0803	0.1254**	0.0878**	0.0836**
Adj R^2	0.9345	0.9215	0.9333	0.9520	0.9299	0.9400	0.9138	0.9307	0.9399	0.9561
ESG Combined										
alpha	0.0357**	0.0064	-0.0017	0.0330*	0.0276**	0.0201	0.0310*	0.0133	0.0229	0.0188
R_m-R_f	1.0884***	1.0456***	1.1023***	1.104***	1.0944***	1.076***	1.1005***	1.0872***	1.1504***	1.1052***
SMB	0.5553***	0.5245***	0.4867***	0.5280***	0.4412***	0.4152***	0.3741***	0.3356***	0.4550***	0.3202***
HML	0.0502	0.0617	0.0614	0.1132**	0.1338***	0.1532***	0.1167**	0.0154	0.0950*	0.0907**
Adj R^2	0.9342	0.9294	0.9483	0.9133	0.9532	0.9494	0.9159	0.9535	0.9137	0.9374
Environmental										
alpha	0.0224	0.0312**	0.0139	0.0252	0.0240*	0.0319**	0.0117	0.0376**	0.0076	0.0003
R_m-R_f	1.0456***	1.0448***	1.0879***	1.1002***	1.1103***	1.105***	1.1742***	1.1173***	1.0418***	1.1277***
SMB	0.6253***	0.5016***	0.5622***	0.5294***	0.4999***	0.5544***	0.4267***	0.3771***	0.2288***	0.1295***
HML	0.1139**	0.2347***	0.1093**	0.0559	0.0215	-0.0175	0.0715	0.1842***	0.0650*	0.0478
Adj R^2	0.9402	0.9512	0.9321	0.9168	0.9435	0.9236	0.8888	0.9211	0.9456	0.9595
Social										
alpha	0.0313**	0.0270*	0.0209	0.0204	0.0176	0.0124	0.023*	0.0196	0.0062	0.0275***
R_m-R_f	1.0462***	1.0622***	1.0889***	1.096***	1.0994***	1.143***	1.0871***	1.1545***	1.1387***	1.0371***
SMB	0.6192***	0.5163***	0.5878***	0.5711***	0.4858***	0.4404***	0.4482***	0.3986***	0.2696***	0.1013***
HML	-0.0006	0.0910*	0.0960*	0.1053**	0.1332***	0.0412	0.1184***	0.0900*	0.1447***	0.0699**
Adj R^2	0.9239	0.9324	0.9314	0.9475	0.9351	0.9289	0.9419	0.9301	0.9466	0.9586
Governance										
alpha	0.0392**	0.0331**	0.0332***	0.0283	0.0101	0.0052	0.0135	0.0200*	0.0107	0.0128
R_m-R_f	1.1513***	1.1177***	1.0398***	1.1374***	1.0725***	1.1025***	1.1375***	1.0624***	1.0916***	1.0401***
SMB	0.591***	0.5614***	0.5233***	0.5199***	0.516***	0.4293***	0.4241***	0.3831***	0.3118***	0.1788***
HML	0.0542	0.0368	0.0148	0.1549**	0.1047**	0.0764	0.1110**	0.1335***	0.0840**	0.1185***
Adj R^2	0.9008	0.9404	0.9469	0.8967	0.9443	0.9322	0.9426	0.9517	0.9493	0.9350

This table presents the empirical results of the Fama French three-factor regression model. Dependent variable is the monthly excess holding period return. Independent variables are the return on the market portfolio (R_m-R_f), SMB and HML. The annualised abnormal return, factor loadings and the adjusted R^2 are presented. The 1st portfolio holds the lowest-rated stocks while the 10th portfolio holds the highest-rated. All portfolios are equally-weighted for the sample period from December 2015 to January 2018. ***, **, * indicate statistical significance at the 0.01, 0.05 and 0.10 level, respectively.

6.2.1.2 Europe Sample

Table 6.6 presents the three-factor model regression results for the decile portfolios for the European sample. Across the different scores, the general theme is that the alphas are typically positive but mostly insignificant. In fact, there is only a single case of a significant alpha at the 1% significance level, which is observed for the 6^{th} ESGC portfolio. Additionally, no alpha is significant for the lowest-rated portfolios while the highest-rated portfolio has a significant alpha of 1.69% at the 5% significance level for the Social score and of 1.9% at the 10% significance level for ESGC. Additionally, evidence suggesting risk-adjusted outperformance of the higher-rated portfolios can be found randomly within the ratings, including the 4^{th} portfolio for ESG and the 6^{th} for ESGC and Environmental. A few additional positively significant alphas at the 10% can be found randomly, with no clear trend seen between the lower and higher-rated firms.

At first glance, the market factor seems be capturing a lot of the explanatory power. As it was the case for the US sample, the market exposure is without exception statistically significant and above one across all deciles, irregardless of the ESG rating. The difference in market exposures across the sorted portfolios cannot be easily pin-pointed, although the top-rated portfolios consistently have a higher beta than the lowest-rated, implying that they are more volatile. However, a closer look at the classical size factor gives a clearer picture. It can be easily seen that the loading from the size factor is strongly significant across the deciles, with the exception of the top-rated portfolios of the ESG, Environmental and Social scores. Furthermore, the size exposure is clearly decreasing as the deciles go up, implying that the higher the rating the larger the size of firms.

Looking at the HML coefficient, the impact of the value loading is not as prevalent among the different ESG scores as the size factor. For the bottom-rated portfolios, a significantly positive loading can be found for all portfolios except for Governance, implying that these portfolios are comprised of value stocks. The same pattern appears for the top-rated portfolios, where they are all significantly positive except for the Governance score. No clear trend emerges, although the top-rated ESG, Environmental and Social portfolios do have a higher factor loading than the bottom-rated, which implies they are even more comprised of value stocks. It interesting to note that the 9^{th} portfolio for each rating has a significantly positive HML coefficient, implying it is always tilted towards value stocks no matter which rating. Other significantly positive value loadings can be found throughout portfolios in between the top and bottom, with no clear pattern. However, the 3^{rd} portfolio for Environmental and the 4^{th} for Social are the only ones with a negatively significant factor loading, implying they are comprised of more growth than value stocks.

The general verdict from the three-factor model suggests that the cross section of returns sorted on ESG ratings can – to a large extent – be explained by the factor loadings of the model. In most cases, the null hypothesis that the alpha is statistically indistinguishable from zero cannot be rejected, suggesting that no risk-adjusted abnormal return can be achieved in most of the decile portfolios in Europe. However, when relaxing the significance constraint to the 10% level, the significant alphas seem to reside more in the higher-rated portfolios, but only for the ESG, ESGC and Social scores. The adjusted R^2 ranges from around 95% in the lower-rated portfolios and up to around 98% in the top-rated portfolios, suggesting that the model explains the variation in returns quite well. The biggest explanatory factor loading was by far the market and size factor, which both displayed strong significant loadings throughout all portfolios.

Table 6.6: Time-Series Regressions of ESG-Rated Portfolios on the Fama French Three-Factor Model for Europe

	1 (low)	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10 (high)**
ESG										
alpha	0.0128	0.0078	0.0197	0.0284**	-0.0010	0.0079	0.0024	0.0218*	0.0183*	0.0136
R_m-R_f	1.0231***	1.1212***	1.0859***	1.1204***	1.0543***	1.0793***	1.1103***	1.0707***	1.0512***	1.0648***
SMB	0.7232***	0.8023***	0.7864***	0.6801***	0.5586***	0.6147***	0.3419***	0.2277***	0.2023***	-0.0531
HML	0.1986***	-0.0667	-0.0746	-0.0525	0.0314	0.0626	0.1413**	0.2504***	0.081*	0.2088***
Adj R^2	0.9462	0.9565	0.9520	0.9605	0.9345	0.9527	0.9493	0.9659	0.9717	0.9805
ESG Combined										
alpha	0.0010	-0.0039	0.0130	0.0142	0.025*	0.0335***	0.0071	0.0018	0.022*	0.019*
R_m-R_f	1.0482***	1.0695***	1.0621***	1.0295***	1.0873***	1.1003***	1.075***	1.0929***	1.128***	1.0875***
SMB	0.693***	0.696***	0.4609***	0.4779***	0.4332***	0.485***	0.6251***	0.5697***	0.2563***	0.187***
HML	0.2006***	0.0857	0.2267***	0.0438	0.0709	-0.0129	-0.0161	-0.0158	0.1156**	0.082*
Adj R^2	0.9484	0.9556	0.9459	0.9594	0.9604	0.9651	0.9384	0.9393	0.9650	0.9706
Environmental										
alpha	0.0045	0.0093	0.0381**	0.0036	0.0020	0.03**	0.0119	0.0014	0.0169	0.0137
R_m-R_f	1.0747***	1.1251***	1.0679***	1.074***	1.0737***	1.0831***	1.061***	1.0792***	1.0286***	1.1135***
SMB	0.7745***	0.7901***	0.6278***	0.6191***	0.6824***	0.4033***	0.4232***	0.3213***	0.1606***	0.0817
HML	0.1209**	-0.0062	-0.1514**	0.0513	-0.0421	-0.0502	0.0566	0.1744***	0.3055***	0.3217***
Adj R^2	0.9640	0.9495	0.9302	0.9448	0.9511	0.9560	0.9662	0.9676	0.9669	0.9669
Social										
alpha	0.0211	0.0008	0.0204	0.0031	0.0072	0.0225	0.0121	0.0150	0.0125	0.0169**
R_m-R_f	1.057***	1.1173***	1.0554***	1.1173***	1.0821***	1.073***	1.1002***	1.048***	1.0691***	1.0613***
SMB	0.7115***	0.7845***	0.7505***	0.6475***	0.6182***	0.4933***	0.3964***	0.3187***	0.2232***	-0.0589
HML	0.108*	0.0225	0.0434	-0.124*	0.0561	0.113*	0.095*	0.141***	0.1551***	0.1718***
Adj R^2	0.9555	0.9540	0.9434	0.9363	0.9328	0.9383	0.9664	0.9637	0.9717	0.9819
Governance										
alpha	0.0203	0.0121	0.025*	0.0074	-0.0106	0.0195	0.0189	0.0099	0.0136	0.0155
R_m-R_f	1.0217***	1.063***	1.0598***	1.106***	1.1324***	1.0658***	1.1171***	1.0547***	1.0696***	1.0909***
SMB	0.7369***	0.6753***	0.4774***	0.5828***	0.548***	0.5066***	0.3814***	0.4123***	0.3461***	0.219***
HML	0.0379	0.2269***	0.1*	0.0617	0.0555	0.0902	-0.0345	0.0407	0.1198**	0.0822
Adj R^2	0.9514	0.9486	0.9518	0.9600	0.9647	0.9554	0.9676	0.9567	0.9576	0.9616

This table presents the empirical results of the Fama French three-factor regression model. Dependent variable is the monthly excess holding period return. Independent variables are the return on the market portfolio (R_m-R_f), SMB and HML which represent the size and book-to-market risk-mimicking portfolios. The annualised abnormal return, factor loadings and the adjusted R^2 are presented. The 1st portfolio holds the lowest-rated stocks while the 10th portfolio holds the highest-rated. All portfolios are equally-weighted for the sample period from December 2015 to January 2018. ***, **, * indicate statistical significance at the 0.01, 0.05 and 0.10 level, respectively.

6.2.2 Carhart Four-Factor Model

An improvement from the Fama French three-factor model is Carhart's four-factor model, which introduces a new momentum (WML) factor, which has been proven in empirical research to do well in helping to explain the returns of ESG stocks. The Carhart four-factor model is applied in order to measure the performance differences of portfolios which vary by their ESG ratings and it has been the standard model within ESG research in the past. As mentioned previously, this model controls for the impact of market risk, the size factor (SMB), value factor (HML) and the newly added momentum factor (WML) on returns [Kempf and Osthoff, 2007].

6.2.2.1 US Sample

Table 6.7 presents the four-factor model regression results for the decile portfolios for the US sample. The results are rather surprising since they generally indicate that the lowest-rated portfolios outperform the highest-rated. On the other hand, all the annualised alphas are positive, although not all significant. This implies that the overall ESG universe of stocks does not underperform the regression-based benchmarks which is a positive finding, implying that there are positive abnormal returns to be obtained no matter where on the rating scale one invests. For the lowest-rated portfolios, every alpha is significant below the 5% level and thus statistically different from zero. Additionally, every alpha is positive, showing outperformance for even the lowest-rated stocks. The highest alpha can be found in the portfolio with the lowest ESG score, where an alpha of 4.86% is observed. The bottom-rated Governance portfolio had a similar highly significant alpha value of 4.47%. On the other hand, the highest-rated portfolios did not show significant abnormal returns for the ESG, Environmental and Governance scores and thus the null hypothesis of the alpha being zero may not be rejected. The top-rated Social portfolio did have a highly significant alpha of 3.01% while the ESGC score had an alpha of 2.2%, but with a p-value below 10%, implying only slight significance. There are also a few middle-rated portfolios where the highest significant abnormal return can be found randomly, including ESG for which an alpha of 4.46% is found in the 4^{th} portfolio while the 8^{th} portfolio had the highest alpha for the Environmental score at 4.21%.

The value (HML) factor loadings were mostly insignificant in each case, implying that no conclusion can be made about whether the portfolios are tilted towards growth or value stocks and the factor does not appear to explain the returns of ESG portfolios well. A few exceptions can be seen, in the 3rd portfolio for ESG, where the coefficient for HML is -0.086, but with only slight significance below the 10% level, implying that the portfolio is more biased towards growth firms with low book-to-market ratios. The ESGC portfolios have no significant HML coefficients but for the Environmental score, the 2nd and 6th portfolios show highly significant factor loadings for HML. Although, the two differ since the 2nd portfolio has an HML coefficient that is positive and the 6th portfolio is negative. This implies – somewhat surprisingly – that the lower-rated portfolio is more biased towards value stocks than the 6th portfolio, which is biased towards growth stocks. The Social score only shows slight negative significance for the bottom-rated and 6th portfolios. Lastly, the HML loading for the bottom-rated Governance portfolio is significantly negative, implying it is comprised of more growth stocks. This is contrary to the findings of Fama and French [1992], who argue that value stocks outperform growth stocks, since the lowest-rated portfolio has the highest alpha but a bias towards growth stocks.

There was a number of significant overall findings which were consistent for all levels of every ESG score. First of all, the market factor was extremely significant in each case with all p-values below 1%. With two exceptions, it was consistently above one except for the 9th Environmental portfolio and the 10th Governance portfolio, which have a market factor exposure of 0.9996 and 0.9960, respectively. This is interesting since the highly-rated portfolios were the only ones less risky than the market which is in line with prior literature on the topic which has shown higher rated firms to be less volatile. The slightly lower beta for two of the highest-rated portfolios implies they are somewhat less volatile than the market, albeit with a very small difference so it can generally be said they are nearly aligned with the market. However, most of the highest-rated betas were higher than the lowest-rated ones, with the exception of the Social and Governance portfolios. This implies that the top-rated stocks are actually more volatile than the bottom ones. Secondly, the size factor loading was highly significant with all coefficients having a p-value below 1%, with the exception of the top ESG portfolio which was significant at 5%. Generally, the SMB factor loading decreases as the ratings increase, albeit always positive, which means that the returns of the low-rated portfolios are driven by small stocks while the top-rated firms are larger stocks, which is in line with prior literature on the topic. The difference between the size loadings is quite large, for instance for the ESG score, the bottom portfolio had an

SMB loading of 0.5931 while the top one had a loading of 0.0748. Lastly, the momentum (WML) factor is negative and highly significant for each ESG, ESGC and individual pillar portfolio, no matter the level of the rating. This suggests that all portfolios exhibit negative momentum, thus every rated portfolio generally has poorer past performance. The top-rated portfolios do have a somewhat higher WML coefficient than the lowest-rated in each case. Finally, every portfolio has a very high adjusted R^2, ranging from 92.34% to 97.19%, which implies that the model explains the variation in returns for each portfolio quite well.

TABLE 6.7: Time-Series Regressions of ESG-Rated Portfolios on the Carhart Four-Factor Model for USA

	1 (low)	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10 (high)**
ESG										
alpha	0.0486***	0.0204	0.0245*	0.0446***	0.0164	0.0115	0.0302**	0.0062	0.0231**	0.0136
R_m-R_f	1.0121***	1.0169***	1.0578***	1.0607***	1.0918***	1.0466***	1.0884***	1.1059***	1.0181***	1.0267***
SMB	0.5931***	0.5094***	0.6388***	0.5976***	0.4965***	0.5120***	0.4255***	0.3362***	0.2483***	0.0748**
HML	-0.0508	0.0030	-0.086*	0.0354	0.0306	-0.0408	-0.0444	-0.0010	-0.0113	0.0022
WML	-0.1383***	-0.1500***	-0.1708***	-0.1317***	-0.1878***	-0.1839***	-0.1938***	-0.1964***	-0.1540***	-0.1264***
Adj R^2	0.9456	0.9345	0.9487	0.9613	0.9480	0.9593	0.9339	0.9518	0.9563	0.9683
ESG Combined										
alpha	0.0394***	0.0085	0.0003	0.0383***	0.0301***	0.0231**	0.0352**	0.0156	0.0270*	0.0220*
R_m-R_f	1.0416***	1.0182***	1.0752***	1.0359***	1.0613***	1.0365***	1.0464***	1.0568***	1.0917***	1.0621***
SMB	0.5548***	0.5242***	0.4864***	0.5272***	0.4408***	0.4148***	0.3734***	0.3352***	0.4543***	0.3197***
HML	-0.0645	-0.0055	-0.0050	-0.0538	0.0526	0.0562	-0.0159	-0.0591	-0.0493	-0.0151
WML	-0.1782***	-0.1044***	-0.1031***	-0.2594***	-0.1261***	-0.1506***	-0.206***	-0.1157***	-0.2236***	-0.1644***
Adj R^2	0.9521	0.9356	0.9542	0.9494	0.9624	0.9631	0.9406	0.9622	0.9397	0.9539
Environmental										
alpha	0.0252**	0.0321***	0.0172	0.0291**	0.0272**	0.0361***	0.0170	0.0421***	0.0108	0.0024
R_m-R_f	1.0090***	1.0327***	1.0439***	1.0496***	1.0747***	1.0508***	1.1051***	1.0598***	0.9996***	1.1004***
SMB	0.6248***	0.5015***	0.5617***	0.5288***	0.4995***	0.5537***	0.4259***	0.3764***	0.2283***	0.1292***
HML	0.0243	0.2049***	0.0015	-0.0680	-0.0658	-0.1506***	-0.098*	0.0431	-0.0385	-0.0189
WML	-0.1392***	-0.0460*	-0.1675***	-0.1925***	-0.1356***	-0.2067***	-0.2633***	-0.2192***	-0.1607***	-0.1037***
Adj R^2	0.9511	0.9521	0.9474	0.9370	0.9538	0.9474	0.9234	0.9478	0.9645	0.9666
Social										
alpha	0.0342**	0.0295**	0.0250*	0.0235**	0.0210*	0.0162	0.0264**	0.0230*	0.0097	0.0301***
R_m-R_f	1.0089***	1.0254***	1.0416***	1.0556***	1.0526***	1.0938***	1.0444***	1.1068***	1.0922***	1.0030***
SMB	0.6188***	0.5159***	0.5872***	0.5706***	0.4853***	0.4398***	0.4477***	0.3981***	0.2690***	0.1009***
HML	-0.0920*	0.0006	-0.0199	0.0062	0.0186	-0.0800*	0.0139	-0.0266	0.0307	-0.0136
WML	-0.1420***	-0.1404***	-0.1803***	-0.1540***	-0.1780***	-0.1878***	-0.1624***	-0.1815***	-0.1772***	-0.1297***
Adj R^2	0.9355	0.9438	0.9490	0.9604	0.9528	0.9482	0.9574	0.9479	0.9651	0.9719
Governance										
alpha	0.0447***	0.0352***	0.0350***	0.0341**	0.0132	0.0091	0.0167	0.0224**	0.0127	0.0161
R_m-R_f	1.0811***	1.0903***	1.0167***	1.0627***	1.0318***	1.0511***	1.0964***	1.0338***	1.0647***	0.996***
SMB	0.5901***	0.5611***	0.5230***	0.5190***	0.5155***	0.4287***	0.4236***	0.3827***	0.3115***	0.1783***
HML	-0.1180**	-0.0302	-0.0420	-0.0284	0.0048	-0.0498	0.0101	0.0630*	0.0179	0.0103
WML	-0.2675***	-0.1041***	-0.0882***	-0.2847***	-0.1551***	-0.1959***	-0.1566***	-0.1090***	-0.1027***	-0.1680***
Adj R^2	0.9356	0.9460	0.9516	0.9369	0.9583	0.9546	0.9561	0.9591	0.9558	0.9555

This table presents the empirical results of the Carhart four-factor regression model. Dependent variable is the monthly excess holding period return. Independent variables are the return on the market portfolio (R_m-R_f), SMB, HML and WML. The annualised abnormal return, factor loadings and the adjusted R^2 are presented. The 1st portfolio holds the lowest-rated stocks while the 10th portfolio holds the highest-rated. All portfolios are equally-weighted for the sample period from December 2015 to January 2018. ***, **, * indicate statistical significance at the 0.01, 0.05 and 0.10 level, respectively.

6.2.2.2 Europe Sample

Table 6.8 presents the four-factor model regression results for the decile portfolios for the European sample. The results are more varied here and the lowest portfolio does not consistently outperform the highest-rated one. In fact, the top-rated portfolios have a highly significant abnormal return for every score, with p-values below 1%, while the bottom-rated only have significant abnormal returns with a p-value below 5% for the ESG, Social and Governance scores. The consistently significant top-rated abnormal return implies that the alpha is statistically different than zero and thus there is a positive abnormal return associated with investing in highly-rated ESG firms. Moreover, the top portfolio has a higher alpha than the bottom-rated for ESGC, Environmental and Governance portfolios, implying that European investors can achieve higher abnormal returns by investing in the top-rated ESG stocks versus the bottom-rated, although this is dependent on the specific score. Additionally, for the ESG portfolios, the highest alpha was found in the 8$^{\text{th}}$ decile, which was 4.3% and for ESGC the highest abnormal return of 4.06% was in the 6$^{\text{th}}$ portfolio and very close to the alpha of the 9$^{\text{th}}$ portfolio. For the individual pillars, the highest alpha is observed in the 3$^{\text{rd}}$ portfolio for both the Environmental, Social and Governance scores. Again, interesting to note all abnormal returns examined are positive, thus no negative alpha and thus no significant underperformance from the benchmark is observed for the European ESG sample, just as it was the case with the US sample.

The market factor was strongly significant for each rating and score, with consistent p-values below 1%. With the exception of the lowest-rated ESG portfolio, the market factor loadings where consistently positive and slightly above one. Overall, the market factor loadings were shown to be quite random, and thus no meaningful inferences could be drawn with respect to differences between high- and low-rated portfolios. Once again, all of the top-rated portfolios had higher betas than the bottom-rated for each score. Particularly the bottom-rated portfolios of ESG and Governance had market factor loadings which were very close to one, implying they are nearly aligned with volatility of the market. Otherwise, most of the other portfolios had a market factor loading higher than 1, implying they are more volatile than the market and thus have a higher beta.

Consistent results for all ratings were again found in the European sample. The SMB factor loading for most portfolios was strongly and significantly positive with a p-value below 1%. That is, with the exception of the top-rated Environmental SMB loading which was insignificant and the top ESG and Social portfolios, which were only significant at 10% and the only two with negative factor loadings, implying that they were comprised of large-cap stocks. Again, the SMB factor loading generally decreased as the ratings increased with each portfolio implying the lower-rated stocks are perhaps getting the higher returns due to their tilt towards smaller stocks. More specifically the two highest-rated portfolios for each score consistently had a lower size coefficient than the other portfolios. Additionally, the momentum factor was strongly and significantly negative in most cases. This implies the portfolios all have negative momentum, no matter where you are on the scale. The majority of the top-rated firms have a slightly higher momentum coefficient than the bottom-rated, with the exception of the Environmental and Governance portfolios. However, no clear pattern emerges between the various levels of ratings. Lastly, the value factor was more prevalent in the European sample. For ESG, it was statistically highly significant for the 2^{nd}, 3^{rd}, 8^{th} and 10^{th} portfolios. However, the two lower-rated portfolios had a negative HML loading of -0.1724 and -0.1975 respectively while the two top-rated portfolios had a positive HML loading of 0.1334 and 0.1206. This implies that the higher-rated portfolios within ESG were more biased towards value companies with a high book-to-market ratio, while the lower-rated portfolios were more biased towards growth stocks and thus a lower ratio. Similar results were found with the Environmental portfolios, where the middle portfolios showed significant negative loadings while the top portfolios had highly significant and positive HML loadings. The ESGC only showed significance for the 8^{th} portfolio, with a negative loading implying the portfolio is tilted towards growth stocks. Lastly, every portfolio had a very high adjusted R^2, ranging from 93.85% to 98.51%, which implies that the model explains the variation in returns for each portfolio very well.

Table 6.8: Time-Series Regressions of ESG-Rated Portfolios on the Carhart Four-Factor Model for Europe

	1 (low)	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10 (high)**
ESG										
alpha	0.0305**	0.0262**	0.0371***	0.0367***	0.0186	0.0264**	0.025*	0.043***	0.0227**	0.028***
R_m-R_f	0.9955***	1.0951***	1.0577***	1.1081***	1.0218***	1.0498***	1.0816***	1.0396***	1.0433***	1.0431***
SMB	0.7184***	0.7905***	0.7671***	0.663***	0.5529***	0.6055***	0.3264***	0.2226***	0.1876***	-0.062*
HML	0.0953	-0.1724***	-0.1975***	-0.107*	-0.0878	-0.0434	-0.0044	0.1334***	0.0361	0.1206***
WML	-0.1621***	-0.1759***	-0.1785***	-0.0736**	-0.1857***	-0.173***	-0.2061***	-0.1865***	-0.056**	-0.1333***
Adj R^2	0.9542	0.9641	0.9596	0.9621	0.9408	0.9600	0.9601	0.9727	0.9729	0.9851
ESG Combined										
alpha	0.024*	0.0093	0.0323**	0.0303**	0.0388***	0.0406***	0.0252	0.0229	0.0405***	0.0303***
R_m-R_f	1.016***	1.0466***	1.0316***	1.0059***	1.0676***	1.0894***	1.0492***	1.0565***	1.1035***	1.0696***
SMB	0.6856***	0.6825***	0.458***	0.4589***	0.4256***	0.4638***	0.6211***	0.5489***	0.248***	0.1809***
HML	0.0691	0.0113	0.1008	-0.0552	-0.0128	-0.0784	-0.129*	-0.146**	-0.0022	0.0133
WML	-0.2026***	-0.1331***	-0.1831***	-0.1505***	-0.1259***	-0.0879***	-0.1701***	-0.2067***	-0.1667***	-0.1034***
Adj R^2	0.9582	0.9615	0.9542	0.9662	0.9638	0.9663	0.9452	0.9495	0.9717	0.9732
Environmental										
alpha	0.0147	0.036***	0.0553***	0.0185	0.0200	0.046***	0.0241**	0.0163	0.0273**	0.0362***
R_m-R_f	1.0569***	1.093***	1.0418***	1.0473***	1.0514***	1.0564***	1.0391***	1.0555***	1.0136***	1.0798***
SMB	0.7797***	0.7668***	0.6216***	0.6117***	0.6626***	0.3876***	0.4136***	0.3102***	0.152***	0.0668
HML	0.0472	-0.1627***	-0.2599***	-0.0443	-0.1517**	-0.1357**	-0.0216	0.081*	0.2421***	0.1787***
WML	-0.1156***	-0.231***	-0.1647***	-0.1436***	-0.1584***	-0.1388***	-0.1283***	-0.1347***	-0.1074***	-0.2088***
Adj R^2	0.9675	0.9629	0.9385	0.9491	0.9580	0.9597	0.9714	0.9726	0.9699	0.9768
Social										
alpha	0.0339**	0.0151	0.0409***	0.0296**	0.032*	0.0389**	0.0269**	0.0303**	0.022**	0.0249***
R_m-R_f	1.0361***	1.0999***	1.0196***	1.0788***	1.0444***	1.0457***	1.0775***	1.029***	1.0563***	1.048***
SMB	0.7001***	0.773***	0.7418***	0.6271***	0.5997***	0.4904***	0.3806***	0.3213***	0.2054***	-0.066*
HML	0.0334	-0.0918	-0.0654	-0.2887***	-0.0793	-0.0067	0.0063	0.0555	0.0991**	0.1105***
WML	-0.1221***	-0.1566***	-0.1755***	-0.2481***	-0.2314***	-0.1651***	-0.1344***	-0.1303***	-0.0788***	-0.088***
Adj R^2	0.9596	0.9616	0.9491	0.9534	0.9444	0.9441	0.9700	0.9678	0.9735	0.9838
Governance										
alpha	0.0327**	0.033**	0.0444***	0.024*	0.0058	0.0321**	0.0314***	0.023*	0.0317**	0.0369***
R_m-R_f	1.0068***	1.0287***	1.0352***	1.0832***	1.1048***	1.0459***	1.0986***	1.0324***	1.0422***	1.0576***
SMB	0.7167***	0.6777***	0.4606***	0.5725***	0.526***	0.4958***	0.3844***	0.4123***	0.3243***	0.2042***
HML	-0.0271	0.0947	-0.0172	-0.0353	-0.0372	0.0025	-0.1214**	-0.0480	0.0132	-0.0526
WML	-0.1006***	-0.2053***	-0.1712***	-0.1418***	-0.1574***	-0.129***	-0.1265***	-0.1314***	-0.1655***	-0.2024***
Adj R^2	0.9542	0.9583	0.9592	0.9668	0.9698	0.9587	0.9706	0.9619	0.9643	0.9723

This table presents the empirical results of the Carhart four-factor regression model. Dependent variable is the monthly excess holding period return. Independent variables are the return on the market portfolio (R_m-R_f), SMB, HML and WML. The annualised abnormal return, factor loadings and the adjusted R^2 are presented. The 1st portfolio holds the lowest-rated stocks while the 10th portfolio holds the highest-rated. All portfolios are equally-weighted for the sample period from December 2015 to January 2018. ***, **, * indicate statistical significance at the 0.01, 0.05 and 0.10 level, respectively.

6.3 Portfolios Sorted on Size

It was seen in the regression analysis that the size factor plays a significant part in explaining the variation in returns and also that it increases as the ratings decrease. Thus, further analysis is done by neutralising the size of the portfolios and seeing whether the significant abnormal returns are simply coming from the smallest firms, regardless of the rating. As such, the ability to see more clearly the source of returns based on the level of ratings is examined to further investigate and aid in answering the Research Question of this paper.

The analysis utilises Carhart's four-factor model and the overall ESG score for both sample regions in this section. The abnormal returns associated with all the other ratings will be presented afterwards. First, five quintile portfolios are formed by the market capitalisation of the firms. The 1st size portfolio contains firms with the smallest market cap while the 5th size portfolio contains the ones with the biggest market caps. The portfolios are then further double-sorted based on the relevant ESG rating within each quintile formed on size, hence the 1st portfolio contains the lowest-rated firms and the 5th portfolio contains the highest-rated firms. This results in a total of 25 portfolios to examine which vary by their size (market-cap) and level of each ESG rating. Thus, a more thorough examination will be performed in order see the differences in high and low-rated ESG stocks that also vary by their firm size.

6.3.1 US Sample

Table 6.9 presents the regression results of evaluating the double-sorted portfolios with Carhart's four-factor model for only the ESG rating for the US sample. It becomes immediately evident that the lowest-sized portfolios, with the lowest market-caps of firms, achieve the highest significantly positive abnormal return with p-values below 1% in every quintile of the ESG rating. This is very interesting since it suggests that the abnormal returns are really being driven by the small-cap firms regardless of their rating. Even more interesting is that the two top-rated ESG quintiles have the highest positive abnormal return observed of 13.44% and 13.51%, respectively. The 2nd and 3rd quintiles have the lowest abnormal return but still significant and positive and the lowest-rated quintile has a highly significant alpha of 11.40%.

Just as with the decile portfolio regressions, the market factor is highly significant for all 25 portfolios. It is interesting to note that the three top-rated ESG quintiles in the biggest-size portfolio all have betas below one, suggesting that the top-rated and biggest firms are less sensitive to market movements. However, the two lowest-rated portfolios also have a beta below one for the 2nd and 3rd sized portfolios, i.e. those with market caps in the lower and middle-end. The size factor loading is once again significantly positive with p-values below 1% for all size-portfolios except the five portfolios with the biggest market-caps. Since the portfolios were formed by size, it is logical to the SMB factor loading decrease as the average size increases, and given the fact that the five biggest portfolios have negative size loadings naturally suggests that they are comprised of companies with large market caps. The HML loading remains mainly insignificant, thus concrete inferences can not be drawn for most of the portfolios with respect to the value premium. However, the two biggest-sized and bottom-rated portfolios have a highly significant and negative loading, as well as the 4^{th} sized top- and middle-rated portfolios, suggesting that those portfolios tend to comprise of growth stocks. The momentum factor is significant with a 1% confidence level for all of the 10 smallest-sized portfolios, for the top-rated portfolio in the 3^{rd} sized quintile, and for all but the 2^{nd} lowest-rated in the 4^{th} biggest-sized portfolio. It is consistently negative and especially for the five smallest-sized portfolios. This suggests they exhibit negative momentum and thus generally have poorer past performance. The adjusted R^2 is fairly high, ranging from 88.32% to 95.69% and it is the highest for the two highest-rated and biggest-sized portfolios, suggesting that the model explains the variation in their returns quite well and increasingly so with the larger average size of the firms.

Table 6.9: Portfolios Sorted on Size and ESG for US with Carhart's Four-Factor Model

	alpha	**R_m-R_f**	**SMB**	**HML**	**WML**	**Adj R^2**
1st Size Portfolio						
1 (low)	0.114***	1.028***	0.9173***	0.0470	-0.3321***	0.8882
2	0.1068***	1.1509***	0.9659***	0.1042	-0.4057***	0.8865
3	0.1016***	1.171***	0.9515***	0.1188	-0.4665***	0.9106
4	0.1344***	1.0715***	0.9722***	0.1469	-0.6127***	0.8882
5 (high)	0.1351***	1.1885***	1.1745***	0.1033	-0.7724***	0.8853
2nd Size Portfolio						
1 (low)	0.0164	0.9436***	0.5097***	-0.0968*	-0.1667***	0.9065
2	0.0014	0.9807***	0.5572***	-0.0545	-0.0931***	0.9104
3	0.0058	1.0655***	0.613***	0.0636	-0.0668**	0.9187
4	0.0359**	1.0129***	0.6053***	-0.0215	-0.137***	0.9015
5 (high)	0.0236	1.1355***	0.6519***	0.0055	-0.3265***	0.9068
3rd Size Portfolio						
1 (low)	0.0088	0.997***	0.4632***	-0.161**	-0.0208	0.8832
2	-0.0044	0.9623***	0.4945***	-0.0858*	-0.0426	0.9192
3	-0.0293**	1.0927***	0.4036***	-0.0475	-0.0152	0.9263
4	-0.0201	1.1054***	0.4025***	0.0142	-0.0342	0.8918
5 (high)	0.0301	1.0765***	0.5289***	-0.0454	-0.1931***	0.9039
4th Size Portfolio						
1 (low)	-0.0071	1.0405***	0.2233***	-0.0369	0.0828***	0.9021
2	-0.031**	1.104***	0.1832***	-0.0393	-0.0091	0.9082
3	-0.0164	1.0566***	0.2792***	-0.1089**	-0.0758**	0.9110
4	0.0032	1.0661***	0.2076***	0.0200	-0.1041***	0.9296
5 (high)	0.0205	1.063***	0.2472***	-0.1597***	-0.2516***	0.9223
5th Size Portfolio						
1 (low)	-0.0158	1.0428***	-0.0079	-0.1666***	-0.0072	0.9265
2	-0.0078	1.0088***	-0.0683*	-0.1009***	0.0195	0.9397
3	0.0007	0.9913***	0.0186	0.0074	0.0024	0.9447
4	-0.0011	0.9771***	-0.1275***	0.0575*	-0.0543***	0.9569
5 (high)	0.017*	0.9869***	-0.0865**	0.0353	-0.0024	0.9507

This table presents the empirical results of the Carhart four-factor regression model. Dependent variable is the monthly excess holding period return. Independent variables are the return on the market portfolio (R_m-R_f), SMB, HML and WML. The annualised abnormal return, factor loadings and the adjusted R^2 are presented. The 1st size portfolio holds stocks with the lowest market caps while the 10th size portfolio holds the highest market cap firms. Within each size portfolio, the 1st portfolio holds the lowest-rated stocks while the 10th portfolio holds the highest-rated. All portfolios are equally-weighted for the sample period from December 2015 to January 2018. ***, **, * indicate statistical significance at the 0.01, 0.05 and 0.10 level, respectively.

6.3.2 Europe Sample

Table 6.10 presents the regression results of evaluating the double-sorted portfolios with Carhart's four-factor model for the European sample. For the European market, the alpha significance is not as prevalent as in the US market, suggesting the European markets are more efficient. However, the three lowest-rated and smallest portfolios all have highly significant abnormal returns with p-values below 1%. However, these abnormal returns are much smaller than in the US sample, ranging from 6.26%

to 6.92%, suggesting that abnormal returns associated with the small-sized ESG universe in Europe is much lower than in the US. Although the top-rated smallest portfolio is not significant, every other top rated portfolio within the other size quintiles is significantly positive and thus statistically different from zero. However, their alphas are smaller, in the range of 2.4% for the biggest-sized portfolio to 4.40% for the 2$^{\text{nd}}$ biggest-sized portfolio.

Once again, the market factor is highly significant for all 25 portfolios with p-values below 1%. All of the market factor loadings are mostly above one, suggesting they are more volatile than the market. The bottom-rated and smallest portfolio has the lowest beta of 0.9586, suggesting it is less volatile than the market. The two highest-rated and smallest portfolios have market loadings of 1.0031 and 0.9992, respectively, thus they are fairly aligned with the market and neither more nor less risky. The same pattern with the size factor appears just as with the US sample, where all portfolios have a highly significant SMB loading with the exception of the 1$^{\text{st}}$ (lowest) and 3$^{\text{rd}}$ (middle-rated) portfolios with the largest size. Again, all of the significant and five biggest portfolios are the only ones with a negative loading, suggesting, quite intuitively considering the methods applied here, that they are large-cap stocks. Again, the HML factor does not appear to explain the returns very well with a few exceptions. Interestingly the 3$^{\text{rd}}$ (middle-sized) and the two lowest-rated portfolios are the only ones with a positively significant HML loadings of 0.1410 and 0.1457, respectively. This suggests that they are the only portfolios that are tilted towards value stocks. Other significant loadings can be found in the two bottom-rated portfolios within each size quintile and they are all significantly negative. Overall, this suggests the lower-rated stocks have a tendency to be growth stocks while the higher-rated and biggest portfolios have a tendency to be value stocks. Again, a significant and negative momentum factor is observed for the 10 smallest-sized portfolios while the significance decreases as the portfolios get larger. The adjusted R^2 again is the lowest for the five smallest portfolios and the highest for the five biggest, implying that the model does a better job at explaining the variation in returns in larger-cap stocks than in small-cap.

6.3.3 Additional Ratings

Since the previous focus was on the ESG rating only, for additional inferences a further check of the alphas of the remaining ratings for both US and Europe is presented in Table 6.11. This way, the

Table 6.10: Portfolios Sorted on Size and ESG for Europe with Carhart's Four-Factor Model

	alpha	R_m-R_f	SMB	HML	WML	Adj R^2
1st Size Portfolio						
1 (low)	0.0692***	0.9586***	1.1906***	0.0926	-0.312***	0.8752
2	0.0667***	1.0193***	1.2014***	-0.0243	-0.2926***	0.8898
3	0.0626***	1.1401***	0.9636***	-0.0759	-0.3306***	0.9115
4	0.0410	0.9992***	1.1065***	0.0955	-0.3936***	0.8691
5 (high)	0.0288	1.0031***	1.0243***	0.325**	-0.5103***	0.8685
2nd Size Portfolio						
1 (low)	0.0163	1.0445***	0.7513***	0.1222	-0.1654***	0.9232
2	0.0554***	1.0873***	0.8256***	-0.3293***	-0.1776***	0.9387
3	0.0494**	1.0569***	0.6817***	-0.0772	-0.0767*	0.9208
4	0.0150	1.0941***	0.7327***	-0.0984	-0.2162***	0.9101
5 (high)	0.0385**	1.0816***	0.6357***	0.172**	-0.262***	0.9344
3rd Size Portfolio						
1 (low)	-0.0008	1.0677***	0.3666***	-0.1561**	-0.0050	0.9459
2	0.0148	1.0817***	0.4502***	-0.1624**	-0.0977**	0.9410
3	0.0306	1.041***	0.6145***	-0.0190	-0.0528	0.9175
4	0.0482**	1.1084***	0.5508***	-0.0973	-0.1569***	0.9257
5 (high)	0.041**	1.0601***	0.313***	0.0637	-0.1656***	0.9425
4th Size Portfolio						
1 (low)	-0.0073	1.0893***	0.2988***	-0.2245***	-0.0210	0.9470
2	0.0074	1.0271***	0.2006***	-0.1317**	-0.0401	0.9513
3	0.0172	1.0732***	0.2208***	-0.0962	-0.0993***	0.9470
4	0.0497***	1.0277***	0.1426***	0.0836	-0.122***	0.9645
5 (high)	0.044***	1.0559***	0.1787***	-0.0703	-0.1101***	0.9492
5th Size Portfolio						
1 (low)	0.0044	1.0759***	0.0593	-0.2839***	0.0104	0.9486
2	0.0142	1.0296***	-0.1536***	-0.0279	-0.0412	0.9640
3	-0.0008	1.0252***	-0.0548	0.141**	0.0058	0.9623
4	0.0087	1.0178***	-0.1786***	0.0761	-0.0206	0.9676
5 (high)	0.024**	1.0603***	-0.2307***	0.1457***	-0.1622***	0.9710

This table presents the empirical results of the Carhart four-factor regression model. Dependent variable is the monthly excess holding period return. Independent variables are the return on the market portfolio (R_m-R_f), SMB, HML and WML. The annualised abnormal return, factor loadings and the adjusted R^2 are presented. The 1^{st} size portfolio holds stocks with the lowest market caps while the 10^{th} size portfolio holds the highest market cap firms. Within each size portfolio, the 1^{st} portfolio holds the lowest-rated stocks while the 10^{th} portfolio holds the highest-rated. All portfolios are equally-weighted for the sample period from December 2015 to January 2018. ***, **, * indicate statistical significance at the 0.01, 0.05 and 0.10 level, respectively.

ability to examine what firm size and level of ratings is optimal for the highest abnormal returns is made possible and whether this is specific to the type of ESG rating. First, for the US sample, the results indicate high significance of the smallest-sized portfolios regardless of the rating, just as was the case with the ESG rating. Highly significant abnormal returns below the 1% level can be found within all five smallest-sized quintiles, regardless of the rating, and are quite high, ranging from 8.83% to 14.59%. For ESGC, Environmental and Social scores, the highest-rated quintiles outperform the lowest-rated, implying that highly-rated firms do produce higher abnormal returns than lower-rated, however this is true only with the condition that the firms are also the smallest in size of all ESG-rated firms considered. Interesting is that the opposite effect is seen for the Governance portfolios, where the bottom-rated and smallest portfolio outperforms the top-rated, implying that the general conclusion is not upheld for this pillar. A few other significant alphas can be found throughout, although with lower significance at the 5% and 10% level. Interesting to note that the only significantly negative abnormal returns can be found in the two lowest-rated portfolios of the ESGC rating, including the 2nd rated quintiles for the 3rd (middle-size) portfolio and the 4th (larger-sized) portfolios and the lowest-rated for the largest-sized portfolio. One other negatively significant alpha is seen for the Social score for the 3rd (middle-rated) portfolio in the 3rd (middle-sized) quintile. On the other hand, all the other significant alphas are found throughout the 4th and 5th (higher-rated) quintiles regardless of the size. Thus, for US, there appears to be a trend for the higher-rated portfolios outperforming the lower-rated when first dividing the sample universe by size. This is especially prevalent for the portfolios with the smallest size of firms, where the highest-rated have significantly higher abnormal returns than the lower-rated, with the only exception being the Governance score.

Next, for the European sample, the results indicate that the small-size premium does not have as big of an effect on returns here as in the US. For the five portfolios with the smallest market caps of firms, the lowest-rated firms all have significantly positive abnormal returns, while the top-rated are only slightly significant in one portfolio and that is for Governance. Unlike US, all of the significant alphas in Europe are positive, which implies no underperformance is observed, regardless of the size and rating. Interesting to note that for the 3rd (middle-sized) portfolios, all of the significantly positive abnormal returns are found mostly in the two highest-rated portfolios. Slight significance is also found for the 3rd portfolio for ESG and Governance and for the 2nd (lower-rated) portfolio for the Environmental rating. For the 4th (larger-sized) portfolios, all significant alphas below the 1% level are found in the highest-rated portfolios regardless of the type of rating and they all have higher abnormal returns than the other significant and lower-rated portfolios of same size. Lastly, for the largest portfolios with the highest market-caps, significant and positive abnormal returns are found randomly but limited to the higher-rated firms, such as the 3rd (middle-rated) portfolio for ESGC and Governance, the top-rated portfolios of Environmental and Governance and the 4th highest-rated portfolio of Social. In summary, although not entirely consistent, there is some weak evidence that in Europe the larger and higher-rated firms outperform the lower-rated. However, surprisingly for the smallest-sized firms, the lower-rated generally outperform the higher-rated, contrary to the findings in US.

Table 6.11: Alphas of Portfolios Sorted on Size and Ratings for US & Europe with Carhart's Four-Factor Model

	US				Europe			
	ESGC	**ENV**	**SOC**	**GOV**	**ESGC**	**ENV**	**SOC**	**GOV**
1st Size Portfolio								
1 (low)	0.0883***	0.1039***	0.1041***	0.1423***	0.0834***	0.0623***	0.0574**	0.0666**
2	0.087***	0.1129***	0.11***	0.1117***	0.0623**	0.0825***	0.0668***	0.0712**
3	0.1289***	0.1161***	0.1251***	0.0883***	0.0553**	0.058**	0.065**	-0.0001
4	0.144***	0.1172***	0.1131***	0.1369***	0.0438*	0.0444	0.0468*	0.0874***
5 (high)	0.1459***	0.1441***	0.1407***	0.1142***	0.0242	0.0225	0.0324	0.0446*
2nd Size Portfolio								
1 (low)	0.0095	0.0160	-0.0039	0.0108	0.0285	0.0183	0.0132	0.0306
2	0.0140	0.0247	0.0202	0.0125	0.0475***	0.0374**	0.048**	0.0465**
3	0.0015	-0.0020	0.0187	0.031*	0.0392**	0.0567***	0.0356*	0.0282
4	0.0246	0.0378**	0.0200	0.0008	0.0274	0.0344	0.0401*	0.0189
5 (high)	0.0356*	0.0083	0.0273	0.0273	0.0310	0.0271	0.0372**	0.0493**
3rd Size Portfolio								
1 (low)	0.0282	-0.0175	-0.0074	0.0216	-0.0112	-0.0087	0.0017	0.0203
2	-0.0348**	-0.0196	0.0006	-0.0099	0.0038	0.0357**	0.0169	0.0156
3	0.0004	0.0125	-0.0258*	-0.0028	0.0324*	0.0119	0.0222	0.0309*
4	-0.0241	-0.0031	-0.0076	-0.0194	0.0486**	0.0593***	0.0436**	0.0264*
5 (high)	0.0162	0.0126	0.0255	-0.0051	0.0615***	0.0365**	0.0492***	0.0395**
4th Size Portfolio								
1 (low)	-0.0242	-0.0205	-0.0239	-0.0055	-0.0098	-0.0201	-0.0016	0.0277*
2	-0.0265*	-0.0212	-0.0077	0.0033	0.0146	0.0329**	0.0032	0.0159
3	0.0128	-0.0241	-0.0170	-0.0257	0.0328**	0.0110	0.0422***	0.0083
4	0.0013	0.0231*	-0.0179	-0.0002	0.0351**	0.0267*	0.0166	0.0218*
5 (high)	0.0052	0.0119	0.0358**	-0.0044	0.0378***	0.0621***	0.0511***	0.0359**
5th Size Portfolio								
1 (low)	-0.0216*	-0.0083	-0.0146	-0.0119	-0.0083	0.0195	-0.0032	-0.0024
2	0.0037	-0.0035	-0.0048	-0.0154	0.0138	0.0028	0.0179	-0.0007
3	0.0070	0.0126	-0.0102	-0.0085	0.0387***	-0.0060	0.0071	0.0247**
4	0.0085	0.0031	-0.0048	0.0182*	-0.0041	0.0101	0.0226*	0.0080
5 (high)	-0.0048	-0.0106	0.0277***	0.0111	0.0097	0.0228*	0.0044	0.02*

This table presents the empirical results of the Carhart four-factor regression model for portfolios first sorted on size and then sorted on the relevant ESG ratings. Dependent variable is the monthly excess holding period return. Independent variables are the return on the market portfolio (R_m-R_f), SMB, HML and WML. The annualised abnormal return is presented for each portfolio. The 1^{st} size portfolio holds stocks with the lowest market caps while the 10^{th} size portfolio holds the highest market cap firms. Within each size portfolio, the 1^{st} portfolio holds the lowest-rated stocks while the 10^{th} portfolio holds the highest-rated. All portfolios are equally-weighted for the sample period from December 2015 to January 2018. ***, **, * indicate statistical significance at the 0.01, 0.05 and 0.10 level, respectively.

6.4 Portfolio Strategies

For the last part of the analysis, a number of portfolio strategies will be examined. Based on the decile portfolio cutoffs, for each rating, portfolios will be formed by holding long positions in the 20% of the top-rated stocks and also the 20% of bottom-rated stocks for a further robustness check in order to see if prior findings were sensitive to the cut-off rates. Additionally, a long-short strategy will be examined of going long in the high-rated portfolio and short in the low-rated. For the long-short strategy, the alpha represents the return difference between ESG outperformers and underperformers. By applying the Carhart model, which was the best model in previous analysis which explained the variation in returns the most, this section will further examine whether the higher-rated firms outperform the lower-rated with a more relaxed cut-off for the level of ratings.

6.4.1 US Sample

A number of trading strategies involving the 20% of the highest-rated (top) and lowest-rated (bottom) stocks and a long-short strategy with both are examined. Table 6.12 shows the regression results of the trading strategies within each rating with the application of Carhart's model for the US sample. First of all, it is interesting to note that the bottom portfolio has a highly significant and positive abnormal return for every rating with p-values below 1%, except for ESGC which is significant below 5%. The abnormal return of the bottom portfolio is highest for Governance at 3.99%, which is surprising and contrary to prior literature and the lowest for ESGC at 2.38%. The bottom portfolio for ESGC is also the only one that has a lower significant abnormal return than the top portfolios, albeit only with a 0.0010% differential. Otherwise, all bottom portfolios outperform the top ones, which is line with prior findings in the regression analysis. Interesting to note that all alphas for top and bottom are positive, thus suggesting that they outperform the benchmark and thus no negative relation is associated with neither high nor low-rated stocks. However, for the long-short strategy, the only significant alphas for the Environmental and Governance scores are negative, suggesting it is not recommended to buy the high-rated stocks and short the low-rated ones. On the contrary, the opposite should be done if one wants to earn a positive and significant abnormal return. This is interesting since it is contradictory to past prior findings and further contributes to the Research Question that the lower-rated stocks have higher abnormal returns than higher-rated.

Once again, for all top and bottom strategies, the market factor is highly significant and above one, suggesting that both strategies are more volatile than the market. Additionally, for all ratings with the exception of Governance, the top-rated market loading is higher than the bottom-rated one, implying that the top-rated strategies have higher betas, which is rather surprising given that the alphas are generally higher for the bottom portfolios and thus you would expect them to be more volatile. For the long-short strategies, the only negative and significant market factor loading of -0.0544 is seen in the Governance portfolio, implying that this portfolio actually moves in the opposite direction of the market. As with earlier findings, the SMB factor plays a big role and is highly significant with p-values below 1% for all strategies. For each rating, it has a higher loading for the bottom portfolios than the top and thus the bottom strategies are tilted towards smaller stocks, as was found in the earlier analysis. Again, this could be a potential explanation for the higher alpha due to a small stock premium. The long-short portfolios all have negative size loadings, implying that they are tilted towards large-cap stocks and thus can also help explain the negative alphas. Just as with prior analysis, the HML factor is not very prevalent and only significant for the bottom and long-short portfolios for the Environmental and Governance ratings. For the Environmental score, the bottom portfolio is highly significant and positive thus implying it is comprised of value stocks, while the long-short portfolio is comprised of growth stocks. The opposite is true for Governance, where the bottom portfolio is tilted towards growth stocks and the long-short is tilted towards value stocks. Lastly, the momentum factor loading is highly and significantly negative for all strategies with the exception of the ESG and Social long-short portfolios. This suggests poorer past performance of the portfolios. The adjusted R^2 is once again very high for the top and bottom strategies, above 95% for all, and decreases substantially for the long-short portfolios which is expected due to the shorting of stocks in those portfolios.

6.4.2 Europe Sample

Table 6.13 shows the regression results of the trading strategies within each ESG score with the application of Carhart's model for the European sample. In Europe, it is evident right away from the alphas that the bottom portfolio does not consistently outperform the top. In fact, for ESGC, Environmental and Governance ratings the top portfolio has an alpha that is significantly higher than the bottom portfolio. The difference is especially large for the ESGC score, where the top portfolio has a highly significant alpha of 3.46% while the bottom alpha is insignificant and thus we cannot reject that it is statistically different from zero while for the top portfolio we significantly reject the null hypothesis of the alpha being equal to zero. In fact, the alphas of the top portfolios are all highly significant with p-values below 1% regardless of the rating. Likewise, the bottom portfolios all

TABLE 6.12: 20% Cut-off Trading Strategies with Carhart's Four-Factor Model for USA

	alpha	**R_m-R_f**	**SMB**	**HML**	**WML**	**Adj R^2**
ESG						
Top	0.0184**	1.0224***	0.1616***	-0.0045	-0.1402***	0.9749
Bottom	0.0345***	1.0146***	0.5518***	-0.0239	-0.1441***	0.9555
Long-short	-0.0156	0.0078	-0.3902***	0.0194	0.0039	0.3791
ESG Combined						
Top	0.0248**	1.0769***	0.3872***	-0.0323	-0.194***	0.9548
Bottom	0.0238**	1.0301***	0.5397***	-0.0351	-0.1416***	0.9609
Long-short	0.0009	0.047*	-0.1525***	0.0028	-0.0524**	0.0849
Environmental						
Top	0.0066	1.0499***	0.1788***	-0.0287	-0.1322***	0.9763
Bottom	0.0286***	1.0208***	0.5638***	0.1148***	-0.0927***	0.9676
Long-short	-0.0215**	0.0290	-0.385***	-0.1435***	-0.0396**	0.5265
Social						
Top	0.0198**	1.0478***	0.1851***	0.0085	-0.1535***	0.9776
Bottom	0.0319***	1.0171***	0.5674***	-0.0460	-0.1412***	0.9597
Long-short	-0.0118	0.0307	-0.3823***	0.0546	-0.0123	0.3941
Governance						
Top	0.0144	1.0304***	0.245***	0.0141	-0.1353***	0.9668
Bottom	0.0399***	1.0858***	0.5757***	-0.074*	-0.1861***	0.9571
Long-short	-0.0246**	-0.0554**	-0.3307***	0.0885**	0.0508**	0.3548

This table presents the empirical results of the Carhart four-factor regression model of various trading strategies. Dependent variable is the monthly excess holding period return. Independent variables are the return on the market portfolio (R_m-R_f), SMB, HML and WML. The annualised abnormal return, factor loadings and the adjusted R^2 are presented. The top portfolio consists of the top-rated 20% of stocks, the bottom portfolio consists of the lowest-rated 20% of stocks, while the long-short strategy consists of holding a long position in the top portfolio and a short position in the bottom. All portfolios are equally-weighted for the sample period from December 2015 to January 2018. ***, **, * indicate statistical significance at the 0.01, 0.05 and 0.10 level, respectively.

have significantly positive abnormal returns with the exception of ESGC. The alpha's of all long-short strategies show insignificance except for the ESGC, which has a positive abnormal return of 1.9% but with only a 10% confidence level, suggesting it is the only long-short strategy that could potentially bring higher risk-adjusted returns and is the most important for the sustainable investor. However, generally it can be said that the long-short strategy does not bring higher abnormal returns for the European market and thus the top strategy does not outperform the bottom, with the only exception being ESGC.

Just as with the US sample, all top and bottom portfolios have highly significant market factor loadings above one. The top portfolios of the ESG, Environmental and Social scores have market loadings that are lower than the bottom, while the reverse is true for the other portfolios. Thus, no general trend emerges. Just as with the prior analysis, the size factor loading is highly significantly positive for each strategy and is consistently much lower for the top portfolios than the bottom. Thus, the top portfolios are larger than the bottom throughout each rating. All of the long-short strategies have a negative size factor loading, implying the portfolios consist primarily of large-cap stocks. The value factor is insignificant for ESGC and Governance portfolios but it is significant, with p-values below 1%, solely for the top and long-short portfolios of the ESG, Environmental and Social ratings. For the significant loadings, they are all positive implying that the portfolios consist of value stocks and a higher coefficient is consistently seen for the long-short portfolios. This implies the long-short strategy is tilted even more towards value stocks than the top strategy. Once again, the momentum factor is highly significant and negative for all top and bottom portfolios. On the other hand, the long-short strategy has a positive and significant loading for the ESG and Social portfolios, implying that the momentum effect is present and that past winners continue to do well. Lastly, the adjusted R^2 is once again very high for the top and bottom strategies, above 96% for all and even slightly higher than the US sample and decreases substantially for the long-short portfolios.

6.5 Value-Weighted Results

To explore whether the analysis conducted is sensitive to the weighing scheme, a further summarised examination of abnormal returns is presented in this section using the value-weighted approach where the firms are weighted by their market capitalisation in each portfolio. This will help to further examine the Research Question of whether the higher-rated firms have higher abnormal returns than the bottom-rated when constructing value-weighted portfolios. It will also aid in finding the best weighing scheme that will bring the highest abnormal return for an investor wanting to investing sustainably.

6.5.1 US Sample

The annualised alphas using the multi-factor models for the value-weighted sample for US can be seen in Table 6.14. Right away it is observed that the equally-weighted sample had much more significant abnormal returns and thus more inferences were able to be made when compared with the value-weighted sample. A striking difference can be seen in the 6^{th} portfolio for ESG, where the alphas for

Table 6.13: 20% Cut-off Trading Strategies with Carhart's Four-Factor Model for Europe

	alpha	R_m-R_f	SMB	HML	WML	Adj R^2
ESG						
Top	0.0261***	1.0421***	0.0672**	0.0814**	-0.0961***	0.9870
Bottom	0.0281**	1.044***	0.7492***	-0.0451	-0.1691***	0.9687
Long-short	-0.0020	-0.0018	-0.682***	0.1265***	0.073***	0.6077
ESG Combined						
Top	0.0346***	1.0857***	0.2113***	0.0105	-0.1336***	0.9790
Bottom	0.0157	1.0315***	0.6813***	0.0346	-0.165***	0.9727
Long-short	0.019*	0.0542***	-0.47***	-0.0241	0.0314	0.4768
Environmental						
Top	0.0318***	1.0453***	0.1089***	0.2111***	-0.1552***	0.9812
Bottom	0.025**	1.0716***	0.7689***	-0.0553	-0.1711***	0.9773
Long-short	0.0067	-0.0263	-0.66***	0.2664***	0.0159	0.6326
Social						
Top	0.0237***	1.0512***	0.0758**	0.1071***	-0.0852***	0.9856
Bottom	0.0259**	1.0636***	0.7366***	-0.0281	-0.1418***	0.9713
Long-short	-0.0022	-0.0124	-0.6609***	0.1353***	0.0566**	0.5825
Governance						
Top	0.0338***	1.0501***	0.2682***	-0.0190	-0.182***	0.9755
Bottom	0.0321***	1.0175***	0.6943***	0.0334	-0.1499***	0.9662
Long-short	0.0017	0.033*	-0.4261***	-0.0523	-0.0321	0.3791

This table presents the empirical results of the Carhart four-factor regression model of various trading strategies. Dependent variable is the monthly excess holding period return. Independent variables are the return on the market portfolio (R_m-R_f), SMB, HML and WML. The annualised abnormal return, factor loadings and the adjusted R^2 are presented. The top portfolio consists of the top-rated 20% of stocks, the bottom portfolio consists of the lowest-rated 20% of stocks, while the long-short strategy consists of holding a long position in the top portfolio and a short position in the bottom. All portfolios are equally-weighted for the sample period from December 2015 to January 2018. ***, **, * indicate statistical significance at the 0.01, 0.05 and 0.10 level, respectively.

both factor models are highly significant and negative, with p-values below 1%. This suggest that investing in this value-weighted portfolio actually brings a negative abnormal return of over 3.5% and thus underperforms the benchmark. Likewise, negative and significant alphas are observed for the 3$^{\text{rd}}$ portfolio for ESGC with both models and for the 3$^{\text{rd}}$ portfolio for Environmental, although only with a 10% confidence interval. Furthermore, a total of only four positive alphas can be observed in only two of the portfolios. That is for the 8$^{\text{th}}$ portfolio for Environmental and 9$^{\text{th}}$ for ESGC. Thus, it does appear there were negative abnormal returns associated with the lower portfolios and positive abnormal returns for the higher portfolios. However, this was only true for five portfolios in total, implying that for the majority of the sample we cannot reject the null hypothesis of the alpha being zero. Overall, it appears the weighing scheme of the ESG-rated portfolios is extremely sensitive and the resulting abnormal returns are much lower and less significant for the value-weighted approach than the equally-weighted for the US sample.

6.5.2 Europe Sample

The annualised alphas using the multi-factor models for the value-weighted sample for Europe can also be seen in Table 6.14. Once again, much less significant alphas are observed when compared with the equally-weighted sample and thus for majority of the portfolios we cannot reject the null hypothesis that the alpha is zero. Interesting to note that negatively significant alphas are observed in only two portfolios and that is for the bottom-rated ESGC and 5^{th} for Governance, albeit only with the FF3 model. Other significantly positive abnormal returns can be found randomly throughout, including the 3^{rd} and 5^{th} portfolio for Environmental, the 4^{th} for Governance, 6^{th} and 8^{th} for ESG with solely Carhart's model, and 7^{th} and 9^{th} for ESGC. Overall, the significance of alphas shows no clear trend between the high and low-rated firms that are value-weighted for the European market. The only negative abnormal return for all models in the portfolio was found in the bottom-rated ESGC and 5^{th} for Governance, suggesting that the lower-rated portfolios are the only ones that underperform the market. On the other hand, no inferences can be made about the top-rated portfolios and thus the value-weighted sample does not help in answering the Research Question of this paper as much as the equal-weighted. The results also imply that ESG investors in Europe should focus on equally-weighing their portfolios instead of value-weighing in order to achieve higher abnormal returns.

6.6 Statistical Diagnostic Tests

6.6.1 Heteroscedasticity and Autocorrelation

The OLS assumption of constant variance were tested for violations using the Breusch and Pagan [1979] test for heteroscedasticity. Tables containing the full summary of the results are presented in Appendix C, where Table C.1 includes the test results for the US sample, and Table C.4 includes the results for the European sample. The test performed on all models provides evidence of the fact that the residuals of the linear regressions are subject to heteroscedasticity, although in fewer cases when using the Carhart four-factor model.

Table 6.14: Annualised Alphas for Multi-Factor Models for the Value-Weighted US & Europe Sample

	1 (low)	2	3	4	5	6	7	8	9	10 (high)
					US					
ESG										
FF3	0.0144	-0.0078	-0.0053	0.0169	-0.0077	-0.0358***	0.0138	-0.009	0.001	0.003
Carhart	0.0151	-0.0067	-0.0048	0.0175	-0.0058	-0.0356***	0.0142	-0.0084	0.0009	0.0026
ESG Combined										
FF3	0.0184	-0.0131	-0.0357**	0.0050	-0.0028	0.0027	0.0095	-0.0079	0.0289**	-0.0078
Carhart	0.0202	-0.0136	-0.0344**	0.0062	-0.0042	0.0028	0.0101	-0.0087	0.0301***	-0.0081
Environmental										
FF3	-0.0023	0.0137	-0.027*	-0.0018	0.0118	-0.0006	-0.0146	0.026**	-0.0004	-0.0054
Carhart	-0.0014	0.0131	-0.025*	-0.0001	0.0115	0.0008	-0.0145	0.0257**	-0.0001	-0.0057
Social										
FF3	0.0210	0.0060	-0.0125	-0.0077	-0.0156	-0.0100	-0.0046	0.0037	-0.0055	0.005
Carhart	0.0213	0.0065	-0.0109	-0.0070	-0.0132	-0.0100	-0.0045	0.0048	-0.0061	0.0046
Governance										
FF3	-0.0087	0.0117	-0.0080	0.0055	-0.0183	-0.0171	-0.0062	-0.0046	0.0128	0.0071
Carhart	-0.0071	0.0111	-0.0082	0.0088	-0.0175	-0.0154	-0.0067	-0.0054	0.0124	0.0071
					Europe					
ESG										
FF3	0.0066	-0.0024	-0.0052	0.0112	0.0021	0.0282	-0.0081	0.0178	0.0006	-0.0034
Carhart	0.0119	0.0007	0.005	0.0095	0.0007	0.0314*	0.0026	0.0232**	-0.0059	0.0019
ESG Combined										
FF3	-0.0345**	0.0125	-0.0102	-0.0044	0.004	0.0217	0.0269**	-0.0106	0.023*	0.0056
Carhart	-0.0278	0.0098	-0.0114	-0.0029	0.0065	0.0194	0.0384***	0.0006	0.0307**	0.0061
Environmental										
FF3	-0.0139	-0.0194	0.0323**	-0.0011	0.0312***	0.0169	0.02	-0.0006	-0.0118	-0.0005
Carhart	-0.0116	-0.0102	0.0353**	0.0117	0.0253**	0.0184	0.0131	-0.0089	-0.0098	0.0111
Social										
FF3	0.0129	-0.0113	-0.0143	0.0032	-0.0033	0.0189	-0.0025	0.0005	0.0034	0.0065
Carhart	0.0186	-0.0025	-0.0103	0.0196	0.0112	0.0249	-0.0002	0.0066	0.0002	0.0042
Governance										
FF3	-0.001	0.0367	0.0101	0.033***	-0.027*	0.0088	-0.0063	0.0029	-0.0083	0.0038
Carhart	-0.0096	0.0377	0.0134	0.0355***	-0.0146	-0.0023	-0.0111	0.0043	0.007	0.0085

This table summarises for each ESG rating the annualised abnormal return using the Fama French three-factor (FF3) and Carhart four-factor (Carhart) model. All portfolios are weighted by the firm's market capitalisation (value-weighted) for the sample period from December 2015 to January 2018. The 1^{st} portfolio holds the lowest-rated stocks while the 10^{th} portfolio holds the highest-rated. ***, **, * indicate statistical significance at the 0.01, 0.05 and 0.10 level, respectively.

The use of OLS in the presence of heteroscedasticity could result in misleading inferences due to the fact that the the standard errors could be wrong Brooks [2012]. That is, the standard errors from the OLS will be deceptively large for the intercept when the errors are heteroscedastic. The effect of heteroscedasticity on the slope standard errors will depend on its form. For example, if the variance of the errors is positively related to the square of an explanatory variable (which is often the case in practice), the OLS standard error for the slope will be too low. On the other hand, the OLS slope standard errors will be too big when the variance of the errors is inversely related to an explanatory variable.

As suggested by Brooks [2012], a potential remedy to this problem is to use heteroscedasticity-consistent error estimates. A common method used in ESG literature is to employ standard error estimates that have been modified to account for heteroscedasticity following the Newey and West [1987] procedure (see e.g. Halbritter and Dorfleitner [2015]). As a robustness check, the regressions were also estimated using the Newey-West procedure, and the implications of the tests remained unchanged.

The OLS regression model further assumes that the errors are uncorrelated. A test of this assumption is therefore performed using the Breusch–Godfrey test for autocorrelation using a twelve month lag. Full summary of the Breusch–Godfrey test results are presented in Appendix C, where Table C.1 includes the results for the US sample, and Table C.4 includes the test results for the European sample. The test indicates a few cases of autocorrelation. Besides some of the higher-rated ESG portfolios, autocorrelation was generally infrequent – at least not enough to cause problems with the coefficient estimates.

6.6.2 Normality

The test results from the Jarque-Bera test for normality are presented in detail in Table C.3 for the US sample and in Table C.6 for the European sample in Appendix C. The table reveals that the normality assumption is significantly violated in almost all cases for the three-factor model. As for the four-factor model, the null hypothesis that the residuals are normally distributed could not be rejected in approximately half of the regressions. These results where found for both regions.

A closer investigation of the residual estimates suggests that the non-normality is likely caused by substantial outliers in the sample. Outliers are observations that appear not to fit in with the general pattern of the rest of the data, and can have a serious effect on coefficient estimates. For example, the OLS model will receive a big penalty in the form of an increased RSS[1] for points that are very distant from the fitted line. Subsequently, the OLS model will put additional effort into minimising the distances of points that would have otherwise been a long way from the line [Brooks, 2012, p. 166]. The easiest way to identify potential outliers in the data sample is by plotting the residuals over time.

[1] Residual Sum of Squares, i.e. deviations predicted from actual empirical values of data

It is quite common in financial modelling to see one or two extreme residuals cause a rejection of the normality assumption [Brooks, 2012, p. 165], and these are typically observations from extreme market conditions such as the global financial crisis. To shed light on this matter, the monthly residual estimates obtained from both the three- and four-factor regressions on the ESG decile portfolios were plotted separately, and are displayed in Figure 6.2 below[2].

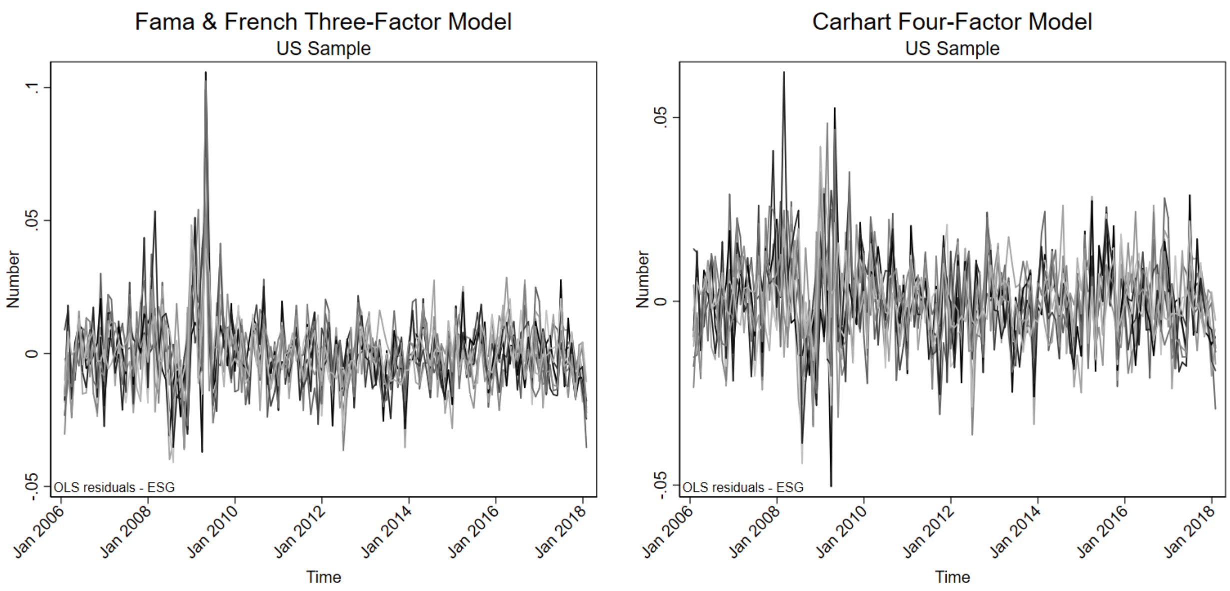


FIGURE 6.2: OLS Residual Estimates

Unsurprisingly, large outliers can be seen in the months following the global financial crisis in 2008, which seem to be the likely cause of the non-normality in many of the regressions. Furthermore, a comparison between the OLS residuals of the three- and four-factor models shows that the outliers were less extreme in the Carhart regressions, which seems to explain why the normality hypothesis was excepted more frequently for that model. A possible solution to this problem involves considering the trade-off between: (a) removing the extreme observations during the financial crisis that could have a disproportionate impact on the OLS estimates and cause residual non-normality, or (b) the belief that each data point represents a useful piece of information on the other, so removing observations will only artificially improve the fit of the model [Brooks, 2012]. It is the authors view that the observations from the financial crisis are quite relevant for the data as whole, and to remove those

[2]The same illustration for the European sample is shown in Figure C.1 in Appendix C. Both graphs give the same implications and thus only the graph for the US sample is included in the main text

observations would make inferences biased, since the four-factor model seemed to be less likely to violate the normality assumption under extreme market conditions.

6.6.3 Multicollinearity

An implicit assumption that is made when using the OLS estimation method is that the explanatory variables are not correlated with one another. Problems arise in a regression when the predictors are highly correlated, a problem known as multicollinearity. In this situation, there may be a significant change in the regression coefficients if you add or delete an independent variable. The estimated standard errors of the fitted coefficients are inflated, or the estimated coefficients may not be statistically significant even though a statistical relation exists between the dependent and independent variables. Seeing that the empirical analysis conducted in this paper involves multiple explanatory variables, it was deemed appropriate for the sake of completeness to also test the models for multicollinearity. Using the variance inflation factors (VIF)[3] and tolerances for each of the independent variables, no evidence of multicollinierity was found.

[3]Following Chatterjee and Hadi [2012], evidence of multicollinearty is found if (1) the largest VIF is greater than 10, and (2) the mean of all the VIFs is considerably larger than 1.

Chapter 7

Discussion and Implication

7.1 Discussion of Factor Models

To begin, the discussion will focus on the factor models used and relate their results to the Research Question of whether the top-rated portfolios had higher abnormal returns than the bottom-rated, the general trends of the abnormal returns and factor loadings and the overall findings that were consistent across all portfolios.

In the previous section, both models were tested for any violations of the statistical assumptions imposed by the OLS model. This was important in order to confirm that the model estimates are reliable enough to conduct fair inferences. As previously discussed, both models were subject to some violations which have been accounted for as much as the scope of this paper allows. Furthermore, it was revealed that the Carhart four-factor model had much more robust residuals and thus fewer violations of the OLS assumptions. As such, the overall implication from the statistical diagnostic tests is that the estimated coefficients obtained from the four-factor model are statistically more reliable than than those obtained from the three-factor model. This notion is supported by the fact that the four-factor model consistently displayed higher adjusted R^2, suggesting that the model has a better so-called "goodness of fit[1]". Given these findings, it was deemed reasonable to focus the attention in the following discussion on the implications behind the findings of the four-factor model.

[1]The goodness of fit of a statistical model describes how well it fits a set of observations

7.1.1 US Sample

For the US sample, although there was no clear trend as the ratings increased, there was a number of significant findings that were observed. A summary of the significant alphas can be found in Table 7.1. As can be seen, there was a trend towards the bottom-rated portfolios having higher abnormal returns and therefore outperforming the top-rated. Moreover, more significant abnormal returns were found in the five lowest-rated portfolios, whereas the three highest-rated portfolios had some of the lowest abnormal returns observed. Thus, it can generally be seen from the analysis in this paper that in the US, investors who invest in firms based on higher ESG ratings underperform those who invest in lower ESG-rated firms. This is contrary to *Hypothesis 1*, which predicted that the higher-rated firms would outperform the lower-rated. An interesting finding was the fact that the lowest-rated portfolios had the highest abnormal return of any other for each score with the exception of Environmental. There could be several explanations for this, one of those can be explained by looking at the size loading of the portfolios. All of the bottom-rated have significantly higher size factor loadings and thus are significantly smaller in size than the top-rated. Thus, one explanation could be the size premium that is associated with smaller stocks, since they tend to be more volatile and thus prone to bigger price movements than larger, well-established firms, which helps explain the higher returns associated with them [Crain, 2011]. What is interesting to note is that most of the market factor loadings (beta) are higher for the top-rated firms, implying they are actually more volatile than the bottom-rated, which is contradictory to the theory and much of prior research. Another explanation can be found in the "shunned stock hypothesis," where lower-rated firms that are associated with controversial industries get shunned by investors and thus are under-priced in the market. There is also evidence of these firms having higher returns, see [Hong and Kacperczyk, 2009, Derwall et al., 2011]. Thus, if these sin stocks are found mostly in the lower-rated firms, it could also be another explanation for the higher abnormal returns and is perhaps a point of discussion for further analysis where the analysis can be improved by dividing the sample by industries. Another interesting finding was that only the bottom-rated and the 4$^{\text{th}}$ portfolios had significantly positive abnormal returns associated with them throughout every rating. This asserts even further the positive returns that were observed in the lower-rated stocks.

Some interesting patterns appeared when looking at the portfolios of the individual ratings. It seems higher ratings for Governance were the least important, where the four bottom-rated portfolios all had fairly high abnormal returns that actually decreased as the ratings increased. This suggests that corporate governance is the least important pillar in US and significant abnormal returns can be achieved by trading in the lower-end of the ratings while mostly insignificant returns were seen in the higher end. One explanation for this could be that in US, the value of good governance is not seen as important or as crucial to the success of firms. This was similar to the findings of [Bhagat and Bolton, 2008, Akbar et al., 2016], who also failed to find a positive relationship between good governance and returns. The Social portfolios had no clear trend evident, where although the bottom-rated portfolio had the highest return, the top-rated had the 2nd highest abnormal return. This suggests that a "values-driven" ESG investor who is willing to give up a small portion of returns in order to invest sustainably could achieve a significant abnormal return by investing in the top-rated Social portfolio. However, the five bottom-rated portfolios for all ratings were significantly positive while only two of the five top-rated were significant. This further confirms the conclusion that generally the lower-rated firms outperformed the top-rated. Lastly, a higher rating appeared to be important for the Environmental portfolios more so than any other. The two highest abnormal returns were found in the 6th and 8th portfolios for Environmental. This is interesting, since it implies perhaps that a higher Environmental rating has a more significant effect on returns in US, whereas for Social and Governance the lower-rated portfolios generally outperform. The higher abnormal return associated with higher Environmental ratings is in line with Derwall et al. [2005], Dixon-Fowler et al. [2013], who found a similar relationship. This could also suggest that in the US, investors value good environmental practices more so than corporate governance and social aspects, which may be due to global warming and increasing pressure on firms to be environmentally friendly and efficient and the demand from consumers for them to do so.

Some overall conclusions can also be made for ESG-rated stocks in US. First is that the factor loadings show that the high- and low-rated firms differ systematically with respect to the size factor. The bottom-rated firms all have significantly higher loadings, implying that they include smaller sized firms, while the top-rated firms all have lower loadings, implying they are larger. This factor is generally decreasing as the ratings increase and thus could aid in explaining the higher abnormal returns that are associated with lower-rated stocks. This is in line with previous studies, such as Humphrey et al. [2012], who find that higher-rated firms are significantly larger than lower-rated. There are a number of reasons for this and the first is that larger sized firms are more well-known to the public and thus have more internal as well as external pressure from all stakeholders to be more sustainable. Moreover, due to their size, they often have more resources to do this than smaller firms and consequently do so [Humphrey et al., 2012]. Additionally, more or less all portfolios have significant momentum factor loadings and what is surprising is that both the lower and higher-rated firms all have negative momentum. Additionally, the negative momentum loadings are slightly lower for the top-rated portfolios than the bottom. This implies that they are past poor performers and thus is actually contradictory to Carhart's theory where past winners earn a premium since it appears that the bottom-rated portfolios, with higher negative momentum, earn higher abnormal returns than the top-rated. Although contrary to Carhart's theory that past losers produce lower returns, it could indicate that future portfolio performance of the losers can become even more significantly positive since, according to theory, past losers become future winners, while past winners become future losers. The negative momentum factor in ESG portfolios was also observed by Humphrey et al. [2012], Bender and Wang [2017], who found that high-ranked portfolios had significantly negative momentum coefficients. The negative momentum also indicates that Asset4 does not give higher ratings to firms simply because they had good performance in the past year, which further asserts the reliability of the ratings.

Overall, the significantly positive abnormal returns associated with the US sample imply that markets are not efficient and that investors can use ESG information in order to gain abnormal returns. However, contrary to *Hypothesis 1* where it was predicted that the higher-rated firms will outperform the lower-rated, the opposite is true for the US sample. Therefore, the findings suggest that sustainable investors may achieve higher abnormal returns in US if they focus solely on the Environmental rating but this only holds if they focus on the middle-rated (4-6) and 8^{th} deciles of the ratings. On the other hand, conventional investors who have no interest in sustainability may invest in the lowest-rated stocks for ESG, ESGC, Social and Governance for the optimal abnormal return.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Min**	**Max**	**% Signif.**
ESG	4.86%		2.45%	4.46%			3.02%		2.31%		2.31%	4.86%	50%
ESGC	3.94%			3.83%	3.01%	2.31%	3.52%		2.70%	2.20%	2.20%	3.94%	70%
ENV	2.52%	3.21%		2.91%	2.72%	3.61%		4.21%			2.52%	4.21%	60%
SOC	3.42%	2.95%	2.50%	2.35%	2.10%		2.64%	2.30%		3.01%	2.10%	3.42%	80%
GOV	4.47%	3.52%	3.50%	3.41%				2.24%			2.24%	4.47%	50%
% Signif.	**100%**	**60%**	**60%**	**100%**	**60%**	**40%**	**60%**	**60%**	**40%**	**40%**			

FIGURE 7.1: Distribution of Significance for US Sample

7.1.2 European Sample

For the European sample, a different trend is seen in the factor models whereas the bottom-rated portfolios did not have the highest abnormal returns anymore. However, once again no clear trend was evident as the ratings increased, as can be seen in the summarised Table 7.2. Interesting to note that more significance was seen throughout all portfolios for each rating, whereas for ESG, Social and Governance only one portfolio had an insignificant alpha, implying that there are positive abnormal returns to be obtained once controlling for the traditional risk factors in Europe no matter where on the rating scale one invests. Moreover, for the top 50% of portfolios (6-10), only three insignificant alphas can be found. This suggests that although the highest returns are not observed consistently for the European investor, nevertheless a values-driven investor may outperform the market by investing in sustainable firms, although it does mean he might have to give up a portion of returns by excluding the lower-rated firms. However, unlike the US sample, where the highest abnormal return was generally found in the lowest-rated portfolio, in Europe one may exclude the two lowest portfolios of any score and still outperform. This is because, rather surprisingly, the most consistent and highest abnormal return was found in the 3rd portfolio. This is important, since a sustainable investor who engages in negative screening by excluding the 20% of lowest ratings may still obtain the highest abnormal return.

When looking at the portfolios of each individual score, a clear pattern is hard to find and thus a generalised trend is not observed. As mentioned, for Environmental, Social and Governance the highest abnormal return was actually observed in the 3rd portfolio with a 99% confidence level. For ESG and ESGC, the highest abnormal return was found in the 8th and 6th portfolio, respectively. The portfolios of ESGC appeared to be the only ones for which a slight trend was seen of the higher-rated portfolios outperforming the lower-rated. Thus, it may be suggested that for the sustainable investor who wants to get higher abnormal returns in Europe to focus solely on the ESGC rating. A logical explanation for this could be that the effect of controversies on firm value in Europe is substantial due to the negative exposure firms get through media and press, in particular because the lowest return was found in the bottom-rated ESGC portfolio. One explanation for not seeing the highest abnormal returns in the highest-rated stocks could be that those firms have lower returns but may still be attractive due to their lower levels of risk. This was confirmed by [Kaiser, 2017], who observed a decrease in standard deviation as ESG ratings increased and it is a good point for further research on the topic to not simply look at the alpha but also the risk associated with it. He further explained that managers integrate ESG due to its focus on risk and long-term value creation and not short-term performance.

The overall conclusions that were made for ESG-rated stocks in US are generally the same for Europe. The size factor plays a big role and is generally decreasing as the ratings increase. One reason why this factor is so prevalent is that the portfolios in the analysis were equally-weighted and so the small stocks were given a much bigger weight than they would if they were value-weighted. Thus, it is logical that the size factor is so significant for each portfolio and for a further check value-weighted results were analysed earlier that produced mainly insignificant returns and thus do not appear to be optimal to implement, as will be discussed later. One general trend for the value loading in Europe was that generally the significant loadings for the lower-rated portfolios were negative, implying they are comprised of growth stocks while the highest-rated portfolios were generally positive, implying they are comprised of more value stocks. As explained by [Bauer et al., 2005], a reason for the high proportion of growth stocks maybe that those portfolios exclude traditional value sectors such as chemicals, energy and basic industries. A growth focus was also found by [Bauer et al., 2005] in SRI funds and that they were more growth oriented than value when compared to conventional funds. Lastly, similarly to the US sample, almost all portfolios had negative momentum, implying they are past losers.

Overall, for the European sample, it was found that although the high-rated firms did not generally have higher abnormal returns than the low-rated, there was a slight trend for more significance of abnormal returns for the top 50% of portfolios. However, a bigger proportion of the lowest abnormal returns was also found in those portfolios when comparing to the lower-rated. It was also found that the highest abnormal return was found by investing in the 3rd portfolio for the individual pillar scores, which equated to scores in the 36-58 range for Environmental, 36-54 in Social and 29-39 for Governance.

	1	2	3	4	5	6	7	8	9	10	Min	Max	% Signif.
ESG	3.05%	2.62%	3.71%	3.67%		2.64%	2.50%	4.30%	2.27%	2.80%	2.27%	4.30%	90%
ESGC	2.40%		3.23%	3.03%	3.88%	4.06%			4.05%	3.03%	2.40%	4.06%	70%
ENV		3.60%	5.53%			4.60%	2.41%		2.73%	3.62%	2.41%	5.53%	60%
SOC	3.39%		4.09%	2.96%	3.20%	3.89%	2.69%	3.03%	2.20%	2.49%	2.20%	4.09%	90%
GOV	3.27%	3.30%	4.44%	2.40%		3.21%	3.14%	2.30%	3.17%	3.69%	2.30%	4.44%	90%
% Signif.	**80%**	**60%**	**100%**	**80%**	**40%**	**100%**	**80%**	**60%**	**100%**	**100%**			

FIGURE 7.2: Distribution of Significance for European Sample

7.1.3 Overall Discussion of Factor Model

Overall, it was generally found that all the significant abnormal returns associated with the ESG-rated universe were positive, even the lowest-rated, which is an important finding. This means that neither low nor high-rated firms underperformed the market, although numerous insignificant abnormal returns were observed where we cannot reject the null hypothesis of the alpha being zero. However, the results of the factor model analysis, particularly for US, did not have the best outcome for the sustainable investor, since it was generally seen that the bottom-rated portfolios outperformed the top-rated. Thus, our results indicate that a sustainable investor in US who perhaps engages in best-in-class screening would hurt their overall return be excluding the lowest-rated stocks. However, for the conventional investor this is an important finding, since they might not limit themselves to a specific rating threshold and thus may invest in the lower-rated firms in order to achieve the highest abnormal return.

The answer to *Hypothesis 3* of whether high-rated (low-rated) ESG stocks from the US sample will outperform high-rated (low-rated) ESG stocks from the European sample was able to be observed. First of all, Europe clearly had a higher percentage of significant abnormal returns, with a rejection frequency of the null hypothesis equal to 80%, implying that more positive returns were found for European portfolios than US, which had a rejection frequency of 62%. When looking at the two extremes, the bottom-rated US portfolio outperformed the European while the top-rated European portfolio outperformed the US. This implies that investing highly on sustainability is more profitable in Europe than US, whereas investing low on sustainability is more profitable in the US. One reason for this could be that European investors have been at the forefront of the SRI movement and thus they value sustainability higher than US investors do. There is also stricter regulations and more standardisation within ESG in Europe, where more formal ESG standards are seen than in US, which may suggest that there is a higher premium associated with highly-rated ESG stocks in Europe and that firm's benefit more from good ESG practices there than in the US [Bender et al., 2017].

7.2 Discussion of Size Portfolios

Some of the most interesting findings were found in the portfolios sorted first by size and then by each ESG rating. By neutralising the size affect, much higher abnormal returns were observed and some clear trends were evident that were not captured with the factor models.

For the US sample, the most positive results were found in the smallest-sized stocks in the ESG-rated universe, which all had significant and positive abnormal returns for all ratings. Moreover, with only the exception of Governance, all the top-rated and smallest-sized portfolios had significantly higher abnormal returns than the bottom-rated. For the ESGC score, this difference was not only significant but also quite large, with the highest-rated portfolio having an annualised abnormal return of 14.59% while the bottom-rated was only 8.83%. Moreover, most of these returns were higher than the ones achieved with the decile portfolios and factor models. This suggests not only that a significant small-size premium is seen for the US ESG universe, regardless of the rating, but also that once the universe is divided by size, the higher-rated firms significantly outperform the lower-rated, which is contrary to the majority of the factor model findings. The only portfolio where this wasn't true was Governance, where the highest abnormal return was found in the smallest and lowest-rated portfolio. This further asserts the findings of the factor models, in which a higher Governance rating appeared to be the least important pillar for higher returns.

Another interesting finding is that the lowest-rated and biggest-sized portfolio for ESGC had a negatively significant abnormal return. This is in line with Khan et al. [2015], who stated that larger firms have a more significant impact from low ratings due to the fact that they are more visible and thus more vulnerable to reputation risk, regulatory and political risk and they are more likely to be targeted by consumer campaigns. Thus, the effect of a low rating is much more evident for bigger firms than smaller ones and this was further confirmed by the findings in this study, where the lowest-rated portfolios within the middle- and bigger-sized quintiles had the only significantly negative abnormal returns that were observed. Moreover, this finding was not observed in the factor models, perhaps due to the small-size premium which overpowered the low-rated large stocks. In summary, the smallest and highest-rated firms generally do have higher returns than lower-rated. This has important implications for the sustainable investor in US, who should focus on not only the higher-rated firms but also the smallest for all ratings except for Governance in order to achieve the highest abnormal returns. Another important implication is that all small-cap stocks within the US ESG universe bring significantly high abnormal returns, ranging from 8.83% to 14.59%. The highest abnormal return was found for the ESGC top-rated and smallest portfolio, suggesting that the focus should be on this rating in order to invest sustainably and not underperform.

The European sample did not appear to have the same significant findings as the US and thus it appears that the small-size premium is not as prevalent for highly-rated European firms as it is for the US. In fact, for the five smallest-sized portfolios, the two lowest-rated and the middle portfolios for all scores (except Governance) had significantly positive abnormal returns. On the other hand, only one top-rated portfolio was significant, which was for the Governance score, which may suggest that Governance is a rather important pillar in Europe, unlike the US, and is further asserted by the fact that the 2^{nd} highest-rated and smallest portfolio for Governance actually had the highest abnormal return when compared with the lower-rated. Moreover, the top-rated portfolios within each sized quintile had consistently positive and significant abnormal returns only for the Governance rating, which further confirms that Governance is particularly important in Europe. For the middle and bigger-sized portfolios, there appears to be a slight trend towards the highly-rated firms having significant and positive abnormal returns, especially in the 3^{rd} (middle-sized) portfolio, implying that a positive effect on returns from a higher rating is especially seen for larger European firms but not for the smallest.

Thus, *Hypothesis 2*, which predicted that a portfolio consisting of stocks with the smallest market caps and the highest-ratings will outperform a portfolio of the smallest and lowest-rated stocks was found to be upheld in US by the results of this analysis. On the other hand, this was found to be incorrect for the European sample, where the lower-rated and smallest firms generally outperformed the top-rated, with the only exception being Governance.

7.3 Discussion of Trading Strategies

The results of the trading strategies by relaxing the 10% portfolio cut-off to 20% still showed similar results to the factor analysis. This was also a further robustness check to see if the results of the analysis change when relaxing the cut-off measure.

For the US sample, the 20% if the lowest-rated stocks consistently outperformed the top 20%, similarly to the findings of the factor models. The only exception was the ESGC score, where the top portfolio had a higher abnormal return than the bottom, albeit with only a 0.001% differential, so essentially the returns were nearly equal. The results of the ESG score were much different, where the bottom portfolio had abnormal returns that were 1.61% higher than the top-rated. Since the only difference between the ESG and ESGC score is the Controversy score, this really shows the significance that controversies have on the returns of the firm. One explanation for this effect is fairly intuitive since controversies, particularly of large and well-known firms, are constantly seen in the media and due to the media storm and public shunning that goes with it, the stock price drops as well. Thus, it appears that although for the ESG score the top-rated firms do not bring higher returns, they do so for ESGC and thus this seems to be a better and more informative measure for sustainable investors to look at than ESG. Thus, it may be concluded that a sustainable investor in US should solely focus on the ESGC score in order to outperform or at least not hurt his return by investing sustainably. Whereas for all other scores, the bottom-rated portfolios outperform the top, even with a more relaxed cut-off. This further confirms the results of the factor model analysis for US.

A number of differences were observed in the European sample that were not seen in the US. For starters, it is clearly seen that the difference between the bottom and top is not as large as the one seen in the US sample. This may imply that although the bottom-rated portfolios bring higher returns in US, in Europe this gap is much smaller and often even negligible. Moreover, the top strategy for ESGC, Environmental and Governance scores brings higher abnormal returns than the bottom-rated. This implies that in Europe, environmental and corporate governance is particularly important in order to achieve higher returns. This is in line with the findings of Gompers et al. [2003], Derwall et al. [2005], Dixon-Fowler et al. [2013], who all found that a higher rating on these two pillars had a significant impact on higher returns. Just as with the US sample, the abnormal return associated with the top ESGC portfolio was particularly substantial and the highest of any portfolio at 3.46% while the bottom portfolio for this rating was the only insignificant one, implying that it is indistinguishable from zero. Just as in the US, this shows the importance of focusing on the ESGC score in particular in order to achieve the highest abnormal return with the highest-rated stocks, while for the ESG score the opposite effect is seen. Again, quite intuitively, controversies have a significant effect on the returns of firms due to much media coverage and public outrage associated with them. Hence, it is a big improvement of this study to use the ESGC score instead of ESG and Thomson Reuters appears to be the only data provider that considers this issue in calculation of scores, which is perhaps for a very good reason when looking at it's effect on returns.

One limitation of the trading strategies is that holding short positions in stocks has considerable costs and drawbacks that must be considered. For some stocks, particularly less liquid and small stocks, it may be difficult or impossible to short them and this thesis recognises that. In this case, the strategy might not be able to be implemented at all. However, accounting for such is beyond the scope of this thesis and the aim of the long-short positions was to see if the differences of returns of the top and bottom-rated stocks were significant.

7.4 Discussion of Value-Weighted Results

It was clear from the value-weighted portfolios that nearly all significant abnormal returns disappear for both the US and Europe sample. This implies that an investor can achieve positive and high abnormal returns with only the equal-weighted sample and not with the value-weighted. One explanation for this, as mentioned previously, could be the size premium associated with the portfolios, which disappears entirely once giving appropriate market-cap weights to the stocks in the portfolios. Thus, a suggestion from the findings in this analysis implies that investors should not value-weight their ESG portfolios as almost no relationship of significant abnormal returns is observed. On the other hand, equal-weighing the portfolios appears to be the optimal choice for ESG investors and also the conventional investor. This is also logical, since it has been generally found in financial literature that equal-weighted samples outperform value-weighted. As an example, Plyakha et al. [2012] compared the performance of portfolios with both weighing methods for stocks in major US equity indices with a sample period of four decades and found that the equal-weighted portfolios outperformed the value-weighted, precisely just as the analysis in this paper has found.

7.5 Delimitations of the Analysis

As stated in the Introduction of this paper, all test results and their subsequent interpretation rely on a set of assumptions made in order to fit the scope of this paper.

7.5.1 Transaction Costs

As mentioned previously, the purpose of this paper is not to simulate or implement any specific investment strategy based on ESG ratings but instead it is to dissect the ESG-rated stocks in order to gain important insight about the underlying drivers and reveal any possible presence of financial performance differences between high-rated and low-rated ESG stocks. Therefore, although transaction costs have a significant effect on any financial study, they are ignored in this thesis. The first reason for this involves the fact that although the analysis utilises monthly re-balancing, the ESG data generally gets updated only once a year. This means that most of the stocks in the portfolios only get re-balanced once a year and so the transaction costs are not as significant as they would be if the portfolios were fully re-balanced monthly. As mentioned by Auer [2014], the frequency of how often the ratings changes occur has a direct impact on the transaction costs of an investment strategy based on ESG scores. Thus, the ratings data utilised in this paper only changes once a year and furthermore, most of the stocks do not experience extreme rating changes in the span of a few years, implying that the transaction costs should not be very high, although they will certainly detract from the abnormal returns observed.

If the investor would like to include transaction costs, round-trip costs between 50 and 200 basis points were found to be appropriate by [Derwall et al., 2005]. Moreover, a number of studies have found that their main conclusions were not affected by introducing transaction costs and thus it was found to be appropriate to exclude for this paper [Auer, 2014, Kempf and Osthoff, 2007].

7.5.2 Additional Delimitations

As discussed in Section 6.6.2, the decision was made not to remove any outliers in an effort to improve the fit of the models. However, under other circumstances the optimal way to address this problem would be to conduct further robustness tests to assess the effects it would have on the analysis. Such robustness tests could, for example, involve splitting the analysed time-period into pre- and post-crisis samples and see if any implications change significantly. Alternatively the use of dummy variables can be used to signal and remove extreme outliers altogether.

One could also argue that a longer time-period should be tested in order to see if the implications of the models would hold. However, the number of ESG observations in the Thomson Reuters database significantly drops as you go further back in time. The empirical analysis conducted in this paper involves the construction of multiple portfolios, and thus relies on a sufficient amount of observation to make valid conclusions. Given that the econometric models used in this paper are based on asymptotic approximations, drawing correct inferences is heavily dependent on the use of large sample statistics [Lo et al., 1997, p. 203]. As was displayed in Figure 6.1, the decile portfolios already contain a minimum amount of observations in the beginning of the sample period in terms of not loosing the effects of diversification. Furthermore, since there was so few companies in the ESG database at the start of 2002, the ability to divide them into 10 portfolios would no longer be justifiable.

Chapter 8

Conclusion

In today's modern age, the landscape of responsible investing is changing rapidly. The effects of global climate change are forcing both governments and corporations to re-think their environmental policies, bad governance was one of the central reasons for the financial crisis of 2008, while the public is becoming increasingly aware of social issues such as child labour and animal testing and increasingly shuns companies that don't uphold certain social standards. This reasoning, along with many other contributing factors, including millennials increasingly stepping into positions of influence and focusing more on ESG issues than older generations, demand for sustainable practices has never been higher. As such, responsible and ethical business conduct has never been more relevant for the future competitiveness of a company and consequently its return. In financial literature, ESG ratings have proved to be a useful metric that helps to quantify all of the relevant non-financial data of a firm, and it has been a subject of strong debate on how these ratings relate to financial performance. Nowadays, ESG investing is not simply a niche market for the sustainable investor, but it is becoming a mainstream method for investors regardless of their objectives. A review of ESG literature revealed that the concept of ESG ratings is still at a fairly early stage and scholars have yet to reach a common consensus on the effect that these ratings have on a firm's return. This paper offers a valuable contribution to the debate by challenging the findings of past research with the use of an improved ESG database offered by Thomson Reuters. The main objective of this paper was, in essence, quite simple – to investigate if significant differences could be observed between the financial performance of high- and low-rated ESG stocks. As such, the Research Question of this paper sought to find out whether firms with higher ESG ratings produced higher abnormal returns than firms with low ratings and whether this differed between the US and Europe.

First, with the aim of testing *Hypothesis 1* where it was hypothesised that higher-rated ESG stocks will outperform the lower-rated, factor model analysis involving Fama French three-factor and Carhart four-factor model was done on decile portfolios which varied by their level of ESG ratings. For the US sample, it was generally found that the lower-rated firms had higher abnormal returns than the top-rated and thus the findings were contrary to the hypothesis. Thus, an investor would hurt his overall return by focusing solely on the highest-rated firms in US. This was particularly true for the Governance rating, where the abnormal return decreased as the ratings increased, suggesting it is not best in terms of returns to invest in firms with higher Governance ratings. On the other hand, the Environmental portfolios appeared to be the only ones that had a slight trend for higher abnormal returns as the ratings increased and thus it may be suggested to only focus on this rating. Therefore, the findings of this paper suggest that in US, investing in higher-rated firms hurts returns and investing in lower-rated firms maximises returns. This conclusion was observed for all ratings with the exception of the Environmental score, which was the only pillar for which a slight improvement in returns was seen for higher-rated firms. As for the European sample, slightly different results were found. For the ESG, Social and Governance only one portfolio for each had an insignificant alpha, implying that there are positive abnormal returns to be obtained once controlling for the traditional risk factors in Europe no matter where on the rating scale one invests. The most consistent and highest abnormal return was found in the 3rd portfolio for the Environmental, Social and Governance pillar. This is important, since a sustainable investor who engages in negative screening by excluding the 20% of lowest ratings may still obtain the highest abnormal return. It was found that although the high-rated firms did not generally have higher abnormal returns than the lower-rated, there was a slight trend for more significance of abnormal returns for the top 50% of portfolios. However, a bigger proportion of the lowest abnormal returns was also found in those portfolios when comparing to the lower-rated. The ESGC portfolios appeared to be the only ones with a slight trend of abnormal returns increasing as the ratings increased. Thus, it may be suggested for the sustainable investor in Europe to focus solely on the ESGC rating in order to invest sustainably and have higher returns.

Secondly, perhaps the most significant findings were found when dividing the portfolios first by size and then by their ratings. This was done in order to test *Hypothesis 2*, where it was predicted that a portfolio of the smallest stocks with the highest ratings would outperform a portfolio of the smallest stocks and lowest ratings for each region. In the US, almost all of the smallest-sized and top-rated portfolios had significantly higher abnormal returns than the bottom-rated and moreover, these abnormal returns were much higher than the ones achieved with the factor models. The only exception to this was Governance and thus it further asserts the conclusion that in the US, a higher Governance rating does not bring a higher abnormal return. However, this was an important finding for the sustainable investor since it was found that highly rated *and* smallest firms significantly outperform the lower-rated and smallest firms. As for Europe, this clear-cut finding was not observed. For the five smallest-sized portfolios, the two lowest-rated and the middle portfolios for all scores, with the exception of Governance, had significantly positive abnormal returns. One significant finding was that the top-rated portfolios within each sized quintile had consistently positive and significant abnormal returns only for the Governance rating, implying that this is the only pillar in Europe where investing higher on the sustainability scale brings higher abnormal returns.

Next, for the portfolio construction and by relaxing the cut-off rate of ratings to top and bottom 20%, the US conclusions found in the factor models were further confirmed. The ESGC portfolio was the only one that slightly outperformed the bottom whereas the bottom-rated portfolios for all other ratings had higher abnormal returns than the top-rated. The same trend was seen for Europe, where the top-rated ESGC abnormal return was the highest of any portfolio while the bottom was insignificant. Furthermore, it appeared that European investors value sustainability more, since the top strategy for ESGC, Environmental and Governance scores had higher abnormal returns than the bottom-rated. Thus, by relaxing the cut-off rate, investing highly on sustainability in Europe brought higher abnormal returns than in the US. This brings us to *Hypothesis 3*, which predicted that high-rated (low-rated) ESG stocks in US would outperform high-rated (low-rated) ESG stocks in Europe.

Lastly, the analysis confirmed that the results found in this paper were highly dependent on being equally-weighted whereas for the value-weighted portfolio sample, nearly all the significance of abnormal returns disappeared for both the US and European sample. This shows the importance of equally-weighing stocks in portfolios based on ESG ratings and it appears to be optimal for the analysis undertaken in this paper.

In summary, the conclusions of this paper were not as black and white as one would hope. A clear trend of higher ESG-rated firms outperforming the lower-rated was generally hard to find and in some cases the opposite was observed. Overall, significantly positive abnormal returns were found in many portfolios, regardless of the rating, implying that markets are not efficient and that investors can use ESG information in order to gain abnormal returns. Nevertheless, some overall conclusions and general trends were found for both the US and Europe sample in the hopes of helping both the sustainable and conventional investor to better base their decisions in order to optimally invest in ESG-rated stocks that bring the highest abnormal return.

Appendix A

Data

A.1 Data Description

Table A.1: Thomson Reuters Percentile and Alphabetical ESG Score Ranks

Score Range	**Grade**
0.0 <= score <= 0.083333	D -
0.083333 <= score <= 0.166666	D
0.166666 <= score <= 0.250000	D +
0.250000 <= score <= 0.333333	C -
0.333333 <= score <= 0.416666	C
0.416666 <= score <= 0.500000	C +
0.500000 <= score <= 0.583333	B -
0.583333 <= score <= 0.666666	B
0.666666 <= score <= 0.750000	B +
0.750000 <= score <= 0.833333	A -
0.833333 <= score <= 0.916666	A
0.916666 <= score <= 1	A +

Appendix B

Methodology

B.1 Thomson Reuters Total Return Index

Thomson Reuters Total Return Index, which can be found under the mnemonic RI, shows a theoretical growth in value of a share holding over a specified period, assuming that any dividends paid are re-invested to purchase additional units of an equity. Using the ex-dividend date, the Thomson Reuters RI is are calculated in the following way:

$$RI_t = RI_{t-1} * \frac{P_t}{P_{t-1}} \tag{B.1}$$

except when $t =$ ex-date of the dividend payment D_t then:

$$RI_t = RI_{t-1} * \frac{P_t + D_t}{P_{t-1}} \tag{B.2}$$

where P_t denotes the price on ex-date, P_{t-1} denotes the price on the previous day, and D_t denotes the dividend payment associated with ex-date t. Gross dividends are used where available and the calculation ignores tax and re-investment charges. Adjusted closing prices are used throughout to determine the price index and hence return index.

Appendix C

Statistical Robustness Tests

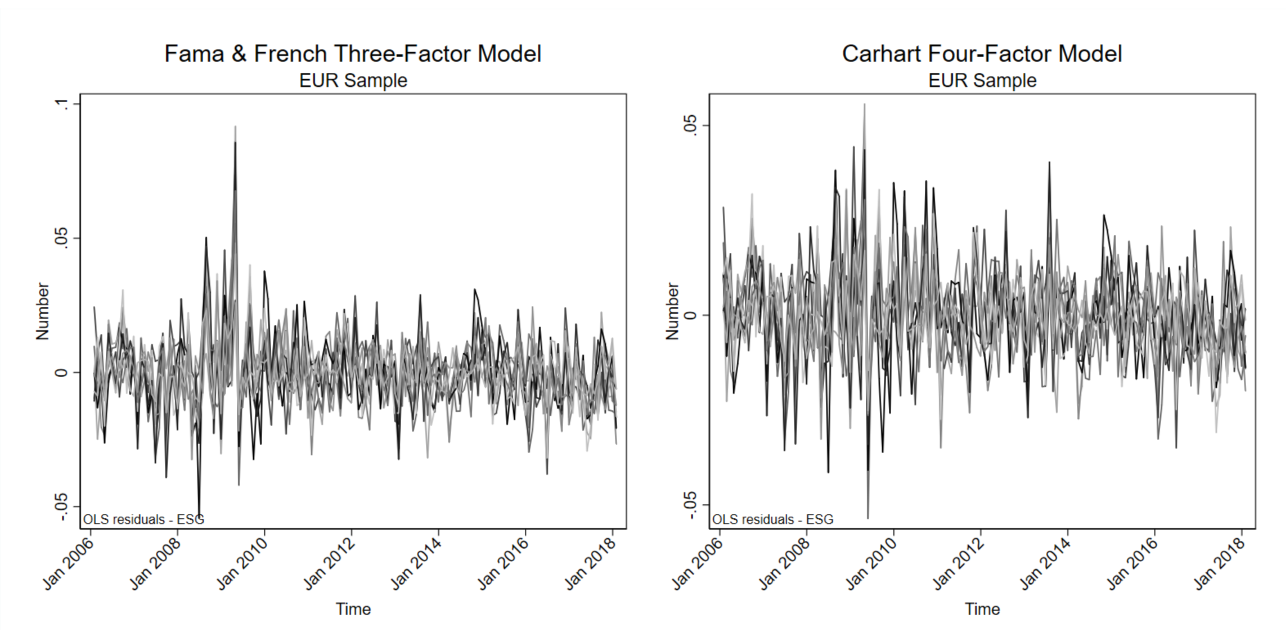


FIGURE C.1: OLS Residual Estimates – Europe

Table C.1: Breusch-Pagan test for heteroskedasticity — US Sample

	Fama & French Three-Factor Model					Carhart Four-Factor Model				
Portfolio	ESG	ESGC	ENV	SOC	GOV	ESG	ESGC	ENV	SOC	GOV
1 (low)	7.07*	26.96***	12.51***	7.79*	38.78***	4.01	8.69*	25.32***	6.73	5.88
2	12.35***	3.16	0.05	16.37***	25.95***	2.40	4.82	0.23	13.38***	21.91***
3	19.89***	21.21***	26.12***	24.23***	8.14**	13.2**	13.37***	18.14***	23.84***	1.97
4	26.81***	67.57***	11.46***	19.87***	114.59***	5.23	29.52***	8.57*	4.58	98.77***
5	87.67***	14.7***	8.76**	65.77***	10.62**	69.77***	3.51	8.6*	16.67***	4.20
6	52.49***	69.68***	39.96***	45.51***	39.82***	21.22***	54.24***	21.25***	36.36***	18.98***
7	28***	93.62***	119.33***	33.52***	66.52***	19.48***	65.89***	83.19***	9.86**	47.18***
8	85.8***	14.26***	57.15***	59.17***	19.65***	43.73***	11.77**	38.28***	20.28***	15.75***
9	9.74**	47.12***	30.84***	46.61***	8.54**	10.81**	26.39***	7.42	20.91***	8.39*
10 (high)	17.53***	49.16***	7.63*	14.26***	33***	22.51***	15.88***	15.47***	15.14***	42.94***

H_0: Constant variance ***, **, * indicate statistical significance at the 0.01, 0.05 and 0.10 level, respectively.

Table C.2: Breusch-Godfrey LM test for autocorrelation — US Sample

	Fama & French Three-Factor Model					Carhart Four-Factor Model				
Portfolio	ESG	ESGC	ENV	SOC	GOV	ESG	ESGC	ENV	SOC	GOV
1 (low)	16.42	27.64***	11.35	13.96	19.74*	12.16	20.19*	15.12	12.97	12.46
2	14.27	16.46	13.03	7.75	15.98	11.15	11.83	14.51	10.81	18.64*
3	14.82	13.78	12.00	15.57	10.23	12.76	11.35	7.53	9.01	8.33
4	11.23	21.05**	15.43	21.32**	19.57*	14.30	18.43	14.21	10.88	11.69
5	15.32	12.75	13.60	18.04	23.29**	14.94	9.10	6.85	15.82	17.07
6	32.8***	10.74	19.96*	17.72	14.30	22.01**	5.78	20.82*	11.41	8.70
7	31.14***	12.91	17.78	19.29*	19.51*	27.04***	15.55	17.93	18.45	17.37
8	20.26*	14.60	29.42***	22.23**	11.70	19.37*	10.00	13.66	19.68*	11.27
9	12.20	28.8***	10.56	9.73	18.10	13.84	19.92*	10.33	22.79**	16.65
10 (high)	33.21***	21.14**	14.54	32.09***	11.10	24.48**	16.88	10.16	23.62**	15.66

H_0: No autocorrelation ***, **, * indicate statistical significance at the 0.01, 0.05 and 0.10 level, respectively.

Table C.3: Jarque-Bera test for normality — US Sample

	Fama & French Three-Factor Model					Carhart Four-Factor Model				
Portfolio	ESG	ESGC	ENV	SOC	GOV	ESG	ESGC	ENV	SOC	GOV
1 (low)	10.48***	25.66***	17.26***	5.72*	49.92***	0.72	2.20	0.71	2.25	0.47
2	10.12***	13.16***	0.73	20.77***	21.06***	1.18	9.2**	1.16	3.65	4.32
3	20.26***	11.29***	28.46***	53.47***	3.55	5.71*	1.90	0.03	6.15**	0.70
4	23.65***	73.47***	28.8***	6.93**	0***	0.73	4.79*	1.29	0.24	11.3***
5	0***	25.09***	8.49**	55.62***	30***	10.58***	6.74**	6**	2.24	1.12
6	36.94***	0***	59.22***	67.2***	40.36***	13.12***	6.97**	3.38	5.1*	1.17
7	64.41***	0***	0***	30.27***	44.87***	21.06***	6.39**	13.53***	1.97	9.5***
8	0***	9.56***	0***	51.26***	8.62**	10.68***	6.65**	8.92**	6.15**	6.06**
9	19.83***	71.28***	48.03***	60.39***	10.54***	4.73*	8.16**	2.41	0.04	8.28**
10 (high)	47.73***	57.96***	19.42***	39.9***	72.09***	30.24***	4.35	10.26***	24.09***	0.43

H_0: Normality. ***, **, * indicate statistical significance at the 0.01, 0.05 and 0.10 level, respectively.

Table C.4: Breusch-Pagan test for heteroskedasticity — EUR Sample

	Fama & French Three-Factor Model					Carhart Four-Factor Model				
Portfolio	ESG	ESGC	ENV	SOC	GOV	ESG	ESGC	ENV	SOC	GOV
1 (low)	54.87***	40.96***	12.57***	27.35***	8.76**	86.64***	63.71***	21.06***	20.07***	21.87***
2	94.52***	68.95***	45.06***	42.89***	101.7***	80.99***	108.16***	16.13***	46.06***	108.86***
3	27.13***	75.32***	25.76***	28.73***	30.44***	15.33***	62.27***	8.26*	25.5***	31.94***
4	3.70	12.61***	26.92***	11.55***	29.15***	3.74	5.12	50.16***	15.1***	8.96*
5	10.69**	26.13***	20.38***	65.64***	3.75	6.84	43.37***	10.03**	42.62***	6.15
6	12.69***	6.5*	24.79***	5.54	24.74***	6.00	10.17**	16.19***	17.68***	18.21***
7	44.45***	2.14	5.63	7.94**	4.37	33.32***	6.58	8.63*	7.09	14.17***
8	44.32***	33.86***	11.38***	27.71***	12.22***	20.97***	13.75***	3.72	18.57***	7.41
9	2.19	16.64***	16.73***	0.57	15.92***	10**	16.23***	42.87***	1.61	8.95*
10 (high)	23.76***	7.79*	60.4***	21.63***	29.61***	21.58***	18.08***	31.78***	34.1***	6.54

H_0: Constant variance ***, **, * indicate statistical significance at the 0.01, 0.05 and 0.10 level, respectively.

Table C.5: Breusch-Godfrey LM test for autocorrelation — EUR Sample

	Fama & French Three-Factor Model					Carhart Four-Factor Model				
Portfolio	ESG	ESGC	ENV	SOC	GOV	ESG	ESGC	ENV	SOC	GOV
1 (low)	6.65	6.84	8.01	16.25	10.26	4.80	6.71	11.79	18.73*	9.87
2	8.33	10.87	13.58	8.36	10.75	8.48	12.72	11.06	9.78	13.25
3	15.57	18.24	13.81	17.33	10.67	14.06	18.24	17.05	13.62	11.65
4	22.49**	12.02	11.24	25.63**	30.14***	20.89*	10.56	14.89	15.99	28.94***
5	21.55**	19*	10.31	19.81*	12.03	28.02***	19.52*	12.33	18.44	8.99
6	12.88	17.97	23.22**	8.56	23.17**	18.58*	16.32	21.53**	17.83	15.68
7	15.32	12.92	22.09**	14.19	11.67	21.12**	19.01*	24.44**	18.81*	9.99
8	17.23	12.71	17.46	8.99	13.68	11.03	21.5**	20.4*	11.76	15.68
9	24.94**	19.41*	10.30	11.69	15.17	26.03**	17.01	15.35	9.73	19.6*
10 (high)	6.49	16.28	18.34	15.56	17.51	5.90	18.73*	17.75	17.79	21.17**

H_0: No autocorrelation ***, **, * indicate statistical significance at the 0.01, 0.05 and 0.10 level, respectively.

Table C.6: Jarque-Bera test for normality — EUR Sample

	Fama & French Three-Factor Model					Carhart Four-Factor Model				
Portfolio	ESG	ESGC	ENV	SOC	GOV	ESG	ESGC	ENV	SOC	GOV
1 (low)	46.95***	43.8***	10.94***	12.93***	11.03***	10.75***	13.23***	3.12	0.62	7.88**
2	69.89***	73.47***	61.48***	42.72***	0***	15.96***	24.97***	13.15***	6.65**	16.97***
3	20.39***	73.47***	16.62***	44.12***	36.84***	3.19	23.21***	2.28	13.69***	11.57***
4	5.69*	4.68*	35.46***	28.99***	15.69***	3.62	6.16**	4.35	8.55**	2.77
5	22.69***	44.02***	21.44***	53.26***	6.63**	3.08	19.21***	2.73	6.52**	2.38
6	12.15***	7.73**	28.27***	9.45***	33.88***	4.10	6.21**	2.72	7.47**	13.56***
7	48.09***	11.49***	20.9***	12.02***	3.65	8.56**	5.12*	6.81**	3.97	11.37***
8	48.51***	20.01***	11.03***	32.03***	10.42***	3.74	1.79	7.38**	6.57**	2.58
9	6.01**	20.53***	13.29***	4.79*	33.2***	3.67	8.27**	21.85***	2.11	4.94*
10 (high)	33.56***	11.29***	42.37***	39.88***	40.32***	2.45	14.64***	2.37	8.37**	6.21**

H_0: Normality. ***, **, * indicate statistical significance at the 0.01, 0.05 and 0.10 level, respectively.

Bibliography

Akbar, S., Poletti-Hughes, J., El-Faitouri, R., and Shah, S. Z. A. (2016). More on the relationship between corporate governance and firm performance in the UK: Evidence from the application of generalized method of moments estimation. *Research in International Business and Finance*, 38:417–429.

Alliance, G. S. I. (2016). Global Sustainable Investment Review. Technical report.

Amel-Zadeh, A. and Serafeim, G. (2017). Why and How Investors Use ESG Information: Evidence from a Global Survey. *Harvard Business School Working Paper*, pages 1–41.

Ammann, M., Oesch, D., and Schmid, M. (2011). *Corporate governance and firm value: International evidence*, volume 18.

Ashwin Kumar, N. C., Smith, C., Badis, L., Wang, N., Ambrosy, P., and Tavares, R. (2016). ESG factors and risk-adjusted performance: a new quantitative model. *Journal of Sustainable Finance & Investment*, 6(4):292–300.

Auer, B. R. (2014). Do Socially Responsible Investment Policies Add or Destroy European Stock Portfolio Value? *Journal of Business Ethics*, 135(2):381–397.

Auer, B. R. and Schuhmacher, F. (2016). Do socially (ir)responsible investments pay? New evidence from international ESG data. *Quarterly Review of Economics and Finance*, 59:51–62.

Aw, E. N. W., Laperla, S. J., and Sivin, G. Y. (2017). A Morality Tale of ESG: Assessing Socially Responsible Investing. *Journal of Wealth Management*, 19(4).

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9(1):3 – 18.

Bauer, R. and Hann, D. (2010). Corporate Environmental Management and Credit Risk. *European Centre for Corporate Engagement (ECCE).*

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

Bello, Z. Y. (2005). Socially Responsible Investing and Portfolio Diversification. *Journal of Financial Research*, 28(1):41–57.

Bender, J., Bridges, T., He, C., Lester, A., and Sun, X. (2017). A Blueprint For Integrating ESG Into Equity Portfolios. *Journal Of Investment Management*, 16(1):44–58.

Bender, J. and Wang, T. (2017). Thematic Indexing, Meet Smart Beta! Merging ESG into Factor Portfolios.

Bhagat, S. and Bolton, B. (2008). Corporate governance and firm performance. *Journal of Corporate Finance*, 14:257–273.

Blitz, D. and Fabozzi, F. J. (2017). Sin Stocks Revisited: Resolving the Sin Stock Anomaly. *The Journal of Portfolio Management*, pages 1–7.

Brammer, S., Brooks, C., and Pavelin, S. (2006). Corporate Social Performance and Stock Returns: UK Evidence from Disaggregate Measures. *Financial Management*, 35(3):97–116.

Breusch, T. S. and Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 47(5):1287.

Brooks, C. (2012). *Introductory Econometrics for Finance.* Cambridge University Press, 2. edition.

Cahan, S. F., Chen, C., and Chen, L. (2017). Social Norms and CSR Performance. *Journal of Business Ethics*, 145(3):493–508.

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

Chatterjee, S. and Hadi, A. S. (2012). *Regression Analysis by Examples.* Wiley, 5 edition.

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

Chen, N.-F., Roll, R., and Ross, S. A. (1986). Economic forces and the stock market. *The Journal of Business*, 59(3):383–403.

Clark, G. L., Feiner, A., and Viehs, M. (2015). From the Stockholder to the Stakeholder. Technical Report March.

Cook, R. D. and Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika*, 70(1):110.

Crain, M. A. (2011). A Literature Review of the Size Effect.

Damodaran, A. (2012). *Investment valuation tools and techniques for determining the value of any asset.* Wiley.

Delmas, M. and Blass, V. D. (2010). Measuring Corporate Environmental Performance: The Trade-Offs of Sustainability Ratings. *Business Strategy and the Environment*, 19(4):245–260.

Derwall, J., Bauer, R., Guenster, N., and Koedijk, K. C. G. (2004). Socially Responsible Investing: The Eco-Efficiency Premium Puzzle. *Financial Analysts Journal*.

Derwall, J., Guenster, N., Bauer, R., and Koedijk, K. (2005). The Eco-Efficiency Premium Puzzle. *Financial Analysts Journal*, 61(2):51–63.

Derwall, J., Koedijk, K., and Ter Horst, J. (2011). A tale of values-driven and profit-seeking social investors. *Journal of Banking and Finance*, 35(8):2137–2147.

Dixon-Fowler, H. R., Slater, D. J., Johnson, J. L., Ellstrand, A. E., and Romi, A. M. (2013). Beyond "Does it Pay to be Green?" A Meta-Analysis of Moderators of the CEP-CFP Relationship. *Journal of Business Ethics*, 112(2):353–366.

Dorfleitner, G., Halbritter, G., and Nguyen, M. (2016). The risk of social responsibility is it systematic? *Journal of Sustainable Finance & Investment*, 6(1):1–14.

Dorfleitner, G., Utz, S., and Wimmer, M. (2013). Where and when does it pay to be good ? A global long-term analysis of ESG investing. *26th Australasian Finance and Banking Conference 2013*, pages 1–29.

Dorfleitner, G., Utz, S., and Wimmer, M. (2018). Patience pays off corporate social responsibility and long-term stock returns. *Journal of Sustainable Finance & Investment*, 8(2):132–157.

El Ghoul, S., Guedhami, O., Kim, H., and Park, K. (2014). Corporate Environmental Responsibility and the Cost of Capital: International Evidence.

Escrig-Olmedo, E., Muñoz-Torres, M. J., Fernández-Izquierdo, M. ., and Rivera-Lirio, J. M. (2017). Measuring Corporate Environmental Performance: A Methodology for Sustainable Development. *Business Strategy and the Environment*, 26(2):142–162.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417.

Fama, E. F. (1991). Efficient Capital Markets: II. *The Journal of Finance*, 46(5):1575.

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

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

Fama, E. F. and French, K. R. (1996). Multifactor explanations of asset pricing anomalies. *The journal of finance*, 51(1):55–84.

Friede, G., Busch, T., and Bassen, A. (2015). ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, 5(4):210–233.

Ghoul, S. E., Guedhami, O., Kwok, C. C. Y., and Mishra, D. R. (2011). Does corporate social responsibility affect the cost of capital? *Journal of Banking & Finance*, 35(9):2388–2406.

Gompers, P. A., Ishii, J. L., and Metrick, A. (2003). Corporate Governance and Equity Prices. *Quarterly Journal of Economics*, 118(1):107–155.

Grossman, S. J. and Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American Economic Review*, 70(3):393–408.

Hagin, R. L. (1990). What practitioners need to knowabout backtesting. *Financial Analysts Journal*, 46(4):17–20.

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

Harper Ho, V. E. (2010). 'Enlightened Shareholder Value': Corporate Governance Beyond the Shareholder-Stakeholder Divide. *Journal of Corporation Law,*.

Hassan, O. A. and Romilly, P. (2018). Relations between corporate economic performance, environmental disclosure and greenhouse gas emissions: New insights.

Hawawini, G. and Keim, D. B. (1995). Chapter 17 on the predictability of common stock returns: World-wide evidence. *Handbooks in Operations Research and Management Science Finance*, page 497544.

Hawley, J. (2017). ESG Ratings and Rankings ESG Ratings and Rankings Introduction and Scope of the Problem. Technical report, TruValue Labs.

Hirigoyen, G. and Poulain-Rehm, T. (2015). Relationships between Corporate Social Responsibility and Financial Performance: What is the Causality? *Journal of Business and Management*, 4(1):18–43.

Hoepner, A. G. F. (2010). Portfolio Diversification and Environmental, Social or Governance Criteria: Must Responsible Investments Really Be Poorly Diversified? *Finans Norge*.

Hong, H. and Kacperczyk, M. (2009). The price of sin: The effects of social norms on markets. *Journal of Financial Economics*, 93(1):15–36.

Huber, B. M. and Comstock, M. (2017). ESG Reports and Ratings: What They Are, Why They Matter? *The Corporate Governance Advisor*, 25(5).

Humphrey, J. E., Lee, D. D., and Shen, Y. (2012). Does it cost to be sustainable? *Journal of Corporate Finance*, 18(3):626–639.

Humphrey, J. E. and Tan, D. T. (2013). Does it Really Hurt to be Responsible? *Journal of Business Ethics*, 122(3):375–386.

Husted, B. W. and Sousa-Filho, J. M. d. (2017). The impact of sustainability governance, country stakeholder orientation, and country risk on environmental, social, and governance performance. *Journal of Cleaner Production*, 155:93–102.

Ince, O. S. and Porter, R. B. (2006). Individual Equity Return Data from Thomson Datastream: Handle With Care! *The Journal of Financial Research*, XXIX(4):463–479.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, 48(1):65–91.

Jensen, M. C. (1968). The performance of mutual funds in the period 1945–1964. *The Journal of finance*, 23(2):389–416.

Jensen, M. C. (1978). Some anomalous evidence regarding market efficiency. *Journal of Financial Economics*, 6(2):95 – 101.

Jiao, Y. (2010). Stakeholder welfare and firm value. *Journal of Banking & Finance*, 34(10):2549–2561.

Jobson, J. D. and Korkie, B. M. (1981). Performance hypothesis testing with the sharpe and treynor measures. *The Journal of Finance*, 36(4):889–908.

Kaiser, L. (2017). Style, Momentum and ESG Integration Motivation.

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

Kerber, R. and Flaherty, M. (2017). Investing with 'green' ratings? It's a gray area.

Khan, M., Serafeim, G., and Yoon, A. (2015). Corporate Sustainability: First Evidence on Materiality. *The Accounting Review*.

Kumar, N., Menou, V., Doole, S., and Nishikawa, L. (2017). Keep It Broad: an Approach To Esg Strategic Tilting. (FEBRUARY).

Kurtz, L. and Dibartolomeo, D. (2011). The Long-Term Performance of a Social Investment Universe is chief investment officer at Nelson Capital Manage- ment in Palo Alto. *The Journal of Investing*, (Fall).

Kuzmina, J. and Lindemane, M. (2017). ESG Investing : New challenges and new opportunities. *Journal of Business Management*, (14):85–99.

Lean, H. H., Ang, W. R., and Smyth, R. (2015). Performance and performance persistence of socially responsible investment funds in Europe and North America. *North American Journal of Economics and Finance*, 34:254–266.

Lee, D. D., Humphrey, J. E., Benson, K. L., and Ahn, J. Y. K. (2010). Socially responsible investment fund performance: the impact of screening intensity. *Accounting & Finance*, 50(2):351–370.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, 47(1):13–37.

Lo, A. W., Mackinlay, A. C., and Campbell, J. Y. (1997). *The Econometrics of Financial Markets*. Princeton University Press.

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, 17(1):5982.

Markowitz, H. M. (1952). Portfolio Selection. *The Journal of Finance*, 7(1):77–91.

Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica*, 34(4):768–783.

Munk, C. (2016). *Financial Markets and Investments*. Copenhagen Business School.

Mnescu, C. (2011). Stock returns in relation to environmental, social and governance performance: Mispricing or compensation for risk? *Sustainable Development*, 19(2):95–118.

Nagy, Z., Kassam, A., and Lee, L.-E. (2015). Can ESG add alpha? *MSCI ESG Research Inc.*, (June):1–23.

Nalebuff, B. J. and Stiglitz, J. E. (1983). Prizes and incentives: Towards a general theory of compensation and competition. *The Bell Journal of Economics*, 14(1):21–43.

Newey, W. K. and West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, 28(3):777–787.

Ortas, E., Moneva, J. M., and Salvador, M. (2014). Do social and environmental screens influence ethical portfolio performance? Evidence from Europe. *BRQ Business Research Quarterly*, 17(1):11–21.

Pedersen, L. H. (2015). *Efficiently Inefficient : How Smart Money Invests and Market Prices Are Determined.* Princeton University Press.

Pelosi, N. and Adamson, R. (2016). Managing the "S" in ESG: The Case of Indigenous Peoples and Extractive Industries. *Journal of Applied Corporate Finance*, 28(2):87–95.

Plyakha, Y., Uppal, R., and Vilkov, G. (2012). Why Does an Equal-Weighted Portfolio Outperform Value- and Price-Weighted Portfolios? *SSRN Electronic Journal*.

Porter, M. E. and Kramer, M. R. (2006). Strategy and Society: The Link Between Competitive Advantage and Corporate Social Responsibility. *Harvard Business Review*.

Przychodzen, J., Gómez-Bezares, F., Przychodzen, W., and Larreina, M. (2016). ESG issues among fund managers-factors and motives. *Sustainability (Switzerland)*, 8(10):1–19.

Renneboog, L., Horst, J. T., and Zhang, C. (2008). Socially responsible investments: Institutional aspects, performance, and investor behavior. *Journal of Banking and Finance*, 32:1723–1742.

Revelli, C. and Viviani, J.-L. (2015). Financial performance of socially responsible investing (SRI): what have we learned? A meta-analysis. *Business Ethics: A European Review*, 24(2):158–185.

Roselle, P. (2016). The Evolution of Integrating ESG Analysis into Wealth Management Decisions. *Journal of Applied Corporate Finance*, 28(2):75–79.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341360.

Sassen, R., Hinze, A.-K., and Hardeck, I. (2016). Impact of ESG factors on firm risk in Europe. *Journal of Business Economics*, 86(8):867–904.

Scholtens, B. (2014). Indicators of responsible investing. *Ecological Indicators*, 36:382–385.

Schröder, M. (2014). Financial effects of corporate social responsibility: a literature review. *Journal of Sustainable Finance & Investment*, 4(4):337–350.

Semenova, N. and Hassel, L. G. (2015). On the Validity of Environmental Performance Metrics. *Journal of Business Ethics*, 132(2):249–258.

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

Sharpe, W. F. (1966). Mutual fund performance. *The Journal of business*, 39(1):119–138.

Sherwood, M. W. and Pollard, J. L. (2018). The risk-adjusted return potential of integrating ESG strategies into emerging market equities. *Journal of Sustainable Finance & Investment*, 8(1):26–44.

Statman, M. and Glushkov, D. (2009). The wages of social responsibility. *Financial Analysts Journal*, 65(4):33–46.

Taylor, J., Vithayathil, J., and Yim, D. (2018). Are corporate social responsibility (CSR) initiatives such as sustainable development and environmental policies value enhancing or window dressing? *Corporate Social Responsibility and Environmental Management*.

Thomson Reuters (2017). Thomson Reuters ESG Scores. (May):12.

Tobin, J. (1958). Liquidity preference as behavior towards risk. *The Review of Economic Studies*, 25(2):65–86.

Treynor, J. L. (1961). Market value, time, and risk. *A slightly edited version available at SSRN: http://ssrn.com/abstract=628187*.

Utz, S. (2017). Over-investment or risk mitigation? Corporate social responsibility in Asia-Pacific, Europe, Japan, and the United States.

Van Dijk-de Groot, M. and Nijhof, A. H. (2015). Socially Responsible Investment Funds: a review of research priorities and strategic options. *Journal of Sustainable Finance & Investment*, 5(3):178–204.

Verheyden, T., Eccles, R. G., Feiner, A., and Arabesque Partners (2016). ESG for All-The Impact of ESG Screening on Return, Risk, and Diversification. *Journal of Applied Corporate Finance*, 28(2).

Wang, Z. and Sarkis, J. (2017). Corporate social responsibility governance, outcomes, and financial performance. *Journal of Cleaner Production*, 162:1607–1616.

Williams, C. A. (2015). Corporate Social Responsibility and Corporate Governance.

Williams, G. (2007). Some Determinants of the Socially Responsible Investment Decision: A Cross-Country Study. *Journal of Behavioral Finance*, 8(1):43–57.

Windolph, S. E. (2011). Assessing Corporate Sustainability Through Ratings: Challenges and Their Causes. *Journal of Environmental Sustainability.*