MASTER THESIS

Copenhagen Business School

2021

The Impact of ESG Scores on Portfolio Return and Risk

An Empirical Study

Katrina Damtoft Jacobsen | 110463

MSc Economics and Business Administration

Finance and Investments

Supervisor: Teis Knuthsen

May 17, 2021

Number of Pages: 75

Characters: 178,223

Executive Summary

This thesis aims to contribute to the debate on whether Environmental, Social, and Governance (ESG) ratings have an impact on financial performance. The study is based on ESG ratings from Refinitiv and covers the STOXX Europe 600 Index over the sample period 2011 to 2020. The main objective is to investigate whether high portfolios, consisting of ESG leaders, perform significantly different than low portfolios, consisting of ESG laggards. For this purpose, decile portfolios are formed on the basis of companies' relative ESG ratings, which defines the positive screening approach. In order to study the portfolios' comparative performance, several risk and return measures are computed. Furthermore, the returns are tested using popular performance benchmark models, including CAPM, Fama-French 3-, and 5-Factor models.

The findings provide supporting evidence for an outperformance in the high portfolio, according to the ESGC score. In contrary, the remaining scores, ESG, E, S, and G, provide evidence for an outperformance in the low portfolio. However, the results also indicate that the low portfolios generally exhibit higher levels of downside risk. In order to test the robustness of the results the sample period is split into two sub samples. Here, a general positive development is observed. The early sub sample provides clear evidence of an outperformance in the low portfolios, whereas the late sub sample shows outperformance in the high portfolios according to both the ESGC, E, and S scores. Lastly, it is tested whether the results are subject to industry bias. This is done by constructing decile portfolios based on the "best-in-class" screening approach, where companies are assigned according to their relative ESG performance among their industry peers. The analysis shows similar findings, indicating that the results are not only a product of sector displacement.

Overall, the study cannot provide a clear-cut conclusion towards the direction of the relationship between ESG scores and financial performance. However, the findings do show that investors have been able to generate abnormal returns by incorporating ESG ratings into their investment decision-making over the course of the analysis period. Thus, the paper concludes that ESG scores do have an impact on financial performance.

Table of Contents

Chapter 1
Introduction

1.1 Introduction

In recent years, there has been an exponential growth in the number of investors incorporating Environmental, Social, and Governance (ESG) measures into their investment decisions. These investment practices are commonly referred to as Socially Responsible Investing (SRI). The growth has been spurred by an increasing awareness of the environmental and social challenges faced today, which have led investors to demand companies and governments to take action. Especially in the light of the recent COVID-19 crisis, the sustainability concerns have even heightened and caused a record capital inflow into SRI practices. Therefore, investors nowadays are increasingly scrutinizing the ESG performance of companies because they do not want to legitimise unethical behaviour. The approach and motivation behind incorporating ESG measures vary. Some investors refrain from 'sin' investments altogether to uphold certain moral standards, whereas others incorporate ESG measures to possibly identify superior performers in the market. This has caused an increased pressure on companies as it is no longer deemed sufficient to only focus on enhanced shareholder value. Instead, companies need to consider and serve the interests of all relevant stakeholders, which defines the transition from shareholder capitalism to stakeholder capitalism.

The incorporation of non-financial measures in investment decisions, challenges traditional financial theory that defines the optimal portfolio allocation according to the mean-variance rule. Hence, opponents argue that SRI can never become an optimal strategy because it will create a constrained asset universe. Consequently, investors will have to sacrifice returns in order to obtain more ethical portfolio profiles. In counter, supporters argue that the inclusion of ESG measures will lead to outperformance as socially responsible companies exhibit lower levels of risk and higher operational performance. For this reason, there is a heating debate in academic literature on whether the incorporation of ESG measures has an impact on financial returns, and if so, whether it is positively or negatively related. The area of investigation has been widely studied but failed to reach a common conclusion, which is mainly due to the subjective nature of ESG. Even though ESG has become a mainstream phenomenon, and adapted by large institutional investors, there is still no consensus in the terminology and methodology underpinning such practices. This is caused by the difficulty in quantifying ESG measures and the lack of legislative standards. Hence, the question still remains on whether ESG performance has an impact on company returns.

1.1.1 Research Question

This thesis aims to uncover a potential relationship between corporate social performance and corporate financial performance. By applying different SRI screening approaches, this thesis will analyse whether an incorporation of ESG scores has a positive, negative, or neutral effect on financial performance. Due to the increasing demand and rising regulatory pressure, the answer towards this question is found imperative for the future implications of investor returns. The above presented field of investigation and motivation give rise to the following research question:

Do ESG scores have an impact on financial performance?

In order to answer this research question, the following null hypotheses are formulated:

> $H1_0$: Portfolios consisting of companies with strong ESG performance generate higher risk-adjusted returns than portfolios consisting of those with weak ESG performance
>
> $H2_0$: Portfolios consisting of companies with strong ESG performance demonstrate lower volatility in returns compared to portfolios consisting of those with weak ESG performance
>
> $H3_0$: Portfolios consisting of companies with high "best-in-class" ESG ratings are outperforming portfolios consisting of those with low "best-in-class" ESG ratings

1.1.2 Delimitations

The scope of the following thesis is to study whether a portfolio consisting of high-scoring stocks perform better than a portfolio of low-scoring stocks. Therefore, the study is only considering stocks as the relevant asset class. The study has also been limited to the European stock market exclusively. Europe is an interesting subject of investigation, as they are considered the frontrunners both in terms of ESG investing and the green agenda. Furthermore, the delimitation makes it possible to disregard the exposure to exchange rate fluctuations. It is not considered relevant to include all European stocks in the analysis, whereas the STOXX Europe 600 Index has been chosen to constitute the asset universe. More specifically, the relevant stocks are those that have reported prices and ESG scores during the entire period of analysis, which has been limited to a 10-year period from January 1st, 2011 to December 31st, 2020. This results in a total asset universe of 428 stocks.

An essential part of the analysis is the ESG scores, that, together with other relevant market data, has been extracted from Refinitiv's Financial Database Eikon (Refinitiv, n.d.-a). Hereby, the study has been delimited from using different data providers which may implicate the generalizability of results. Moreover, the scores have not been constructed manually, as it would require a significant amount of datapoints to generate the scores for an appropriate population size. The focus of this paper is to construct portfolios and test the ability of using ESG scores as a selection-tool. Therefore, it is considered more meaningful to have the ESG scores provided by a well-known databank for analytical purposes. Furthermore, there is no ‘correct’ or standardized way of measuring companies’ ESG performance, whereas a manual construction is not considered to provide any added value. This also entails that the construction of portfolios and thereby the conclusions reported in this thesis will be highly based on the data quality and scoring methodology applied by Refinitiv.

For the purpose of performance testing, the study has been limited to include three popular benchmark models, namely CAPM, Fama-French 3-Factor and Fama-French 5-Factor. The STOXX Europe 600 Index return has been chosen to represent the market factor. However, the remaining factors applied in the Fama-French 3- and 5-Factor models, have been extracted from the Kenneth French Data Library. More specifically, the factor returns for SMB, HML, RMW, and CMA have been found through the ‘Fama/French European 5 Factors’ data set. The factors have hereby not been constructed manually, as the primary objective of this study is not to test whether the models are correct but to apply them for analytical purposes.

Additionally, the study does not consider taxes or transaction costs, even though it is acknowledged as having an impact on the profitability of investment strategies through the investor’s realised returns. The portfolios are also limited from shorting practices, which is a common approach by previous studies. However, shorting is not considered relevant to answer the problem statement and it is also proven a more difficult practice for the private investor. Ultimately, the reader is expected to have a basic knowledge of mathematics, statistics, and finance.

1.2 Reading Guide

The following section presents a reading guide for the thesis. The guide is included to give the reader a pleasant reading experience, by clearly identifying and explaining the different sections that constitute this thesis. As presented through the illustration below, the paper is divided into six chapters that each serves a specific purpose for the overall cohesion of the study.

Introduction

Chapter 1
The purpose of the first chapter is to set the stage for this study, by introducing the field of investigation, research question and delimitations. Furthermore, this reading guide is included to provide an overview of the thesis structure.

Literature Review

Chapter 2
The second chapter establishes the foundation of the study. It introduces the central concepts of socially responsible investing, including current arguments from supporters and opponents. Furthermore, relevant findings from previously conducted studies will be presented.

Theory

Chapter 3
The third chapter presents the theoretical framework on which this study is based on. Here, the reader will be presented for the fundamental elements within portfolio theory, including risk and return properties, factor models, and performance measures.

Data and Methodology

Chapter 4
The fourth chapter presents the data and methodology that constitute the basis of the analysis, including the construction and testing of portfolios. Moreover, the chapter holds an introduction of some important econometric considerations that are crucial for the reliability of results.

Analysis and Results

Chapter 5
The fifth chapter includes the study's analysis, which is based on the preceding theory and methodology. The chapter includes a presentation of all the findings that constitutes the basis of the following discussion and final conclusion.

Discussion and Conclusion

Chapter 6
The sixth and final chapter provides a critical reflection of the previously described findings. Moreover, how these findings match the constructed hypotheses, previous studies, and theory. Finally, the conclusion will be presented based on the study's research question.

Chapter 2

Literature Review

The purpose of the following literature review is to establish the empirical foundation of this study. The first section will give a comprehensive overview of the basic concepts of Socially Responsible Investing (SRI), including its historical development and definition. SRI is the overarching practice of which ESG investing is derived from, wherefore it is important to understand its origin and basic idea. The part also includes a description of ESG and its three pillars, that constitute the cornerstone of this thesis. This is followed by a description of some of the common strategies applied in SRI, which are negative screening, positive screening, and shareholder activism. Hereafter, arguments from SRI supporters and opponents are presented as their disagreement towards the profitability of SRI is one of the main drivers behind this study. Lastly, relevant findings from previous studies will be presented as well as this study's contribution to the field of research.

2.1 Socially Responsible Investing

Formerly, the only objective for investors were to seek investments that generated the highest expected returns. However, many investors have begun to incorporate non-financial measures into their decision making, such as social and environmental considerations. These practices are also known as Socially Responsible Investing. A formal definition of SRI does not exist, but the following notion is often used to define sustainable practices: "*development that meets the needs of the present without compromising the ability of future generations to meet their own needs*" (Brundtland, 1987, pp. 41). SRI can also be described as an investment strategy where investors carefully consider the Environmental, Social, and Governance (ESG) consequences of their investments. There are several terms mirroring the notion of responsible investing, some of these being: ESG investing, social investing, socially aware investing, green investing, value-based investing, and mission-based investing, which are often used interchangeably (Camilleri, 2020). To avoid any confusion, the terms SRI and ESG investing will be used going forward.

2.1.1 Historical Development

The origins of modern SRI can be led back to the seventeenth century where religious organizations, such as the "Quakers" and "Methodists", were guided by moral beliefs rather than financial motives when spending and investing their money (Wagemans et al., 2013). Later, financially wealthy churches and charities were able to persuade financial institutions to establish so-called "ethical funds". These funds were to exclude investments in companies that engaged in unethical practices, which was originally based on religious beliefs.

The screening criteria often referred to 'sinful' practices, such as pornography, alcohol, or tobacco. Today, this strategy is known as "negative screening" which will be further elaborated upon in 2.2.1.

SRI has its roots in religious movements but has evolved into a much more complex phenomenon. The modern era of SRI traces back to the 1960's, where the first responsible funds became available for private investors. Throughout the 1970 to 1980's more attention came to responsible practices and screening due to various social movements, such as anti-war and civil rights campaigns. At that time additional screening criteria arose, where companies would be excluded if they supported or engaged in practices such as war, apartheid, poor treatment of employees, child labour, etc. (Blowfield & Murray, 2008). Throughout the 1980's more ethical funds were established, including funds specifically focused on companies' environmental impact. The focus on environmental impact was led by incidents such as Bhopal and Chernobyl, as well as an increasing attention towards global warming in international media (Camilleri, 2020).

Entering the 21st century, SRI had become a mainstream phenomenon and various funds were emerging to meet the investors' needs and interests. With the century shift, the political focus was also turned towards SRI and it became the beginning of a more coordinated approach of non-financial performance disclosures (Blowfield & Murray, 2008). These measures were often categorised through the following three factors: Environmental, Social and Governance, which led to the emergence of the term 'ESG'. ESG was officially promoted at the launch of the United Nations Principles for Responsible Investing (PRI) in 2006 (PRI, n.d.). PRI is an international organization working to promote the inclusion of ESG measures into investment decision-making. Since PRI's beginning, their number of signatories has grown from 100 to more than 3,000, with approximately USD 100 trillion assets under management (AuM). By becoming a signatory, the investor publicly commits to investing responsible (Ibid). More specifically, the investor commits to the following six principles defined by PRI.

Figure 2.1: Six Principles for Responsible Investing

UN PRI's Six Principles for Responsible Investments

1. We will incorporate ESG (environmental, social, and corporate governance) issues into investment analysis and decisions-making processes
2. We will be active owners and incorporate ESG issues into our ownership policies and practices
3. We will seek appropriate disclosure on ESG issues by the entities in which we invest
4. We will promote acceptance and implementation of the Principles within the investment industry
5. We will work together to enhance our effectiveness in implementing the Principles
6. We will each report on our activities and progress towards implementing the Principles

Source: Modified version of PRI (n.d.)

Recent numbers, presented by the European Fund and Asset Management Association (EFAMA), illustrate that a total of EUR10.7 trillion AuM incorporate some ESG criteria, which represents 45% of the total AuM in Europe (EFAMA, 2020). This defines a significant milestone for ESG investing. However, the report also states that the number should be interpreted with caution as sustainable investment approaches are being exercised and understood differently.

The recent accelerated growth has been caused by some important market trends and challenges. Overall, the growth has been driven by global sustainability challenges, introducing new risk factors that investors need to incorporate in their decision making (MSCI, n.d.). Climate change remains top-of-mind for many investors as it is still considered to pose the greatest threat to today's society (Norton, 2020). Because of this, a current divestment movement has been observed in the markets, where numerous mainstream investors are divesting from climate sinners, such as fossil fuels and coal (Carlin, 2021). This tendency is expected to continue in the following years, which will result in a major capital reallocation. Another important market trend was seen in the light of the recent COVID-19 crisis that reinforced the importance of ESG measures (Refinitiv, 2021). The ESG funds proved to be more resilient through the crisis as they outperformed most of their conventional counterparts (Whieldon & Clark, 2021). This could also be seen through the record inflows into ESG funds in 2020, surpassing $150 billion in the fourth quarter (Jessop & Howcroft, 2021). The crisis hereby highlighted the importance of incorporating non-traditional risk measures, as the socially responsible companies proved better at absorbing the shock and adjust their business practices (Birkin et al., 2020).

From a political focus, COVID-19 has also been expediting the green agenda. The pandemic has left a severe global economic crisis that needs to be recovered through stimulus packages. In connection to this, the politicians have seen the opportunity to invest in the green agenda. This is done through the EU Recovery Funds that are designed to restore the economy post COVID-19 by investing in the EU climate action program. More specifically, 25% of EUR 750 billion will be invested in long-term projects that meet the energy and climate criteria (Birkin et al., 2020). This will lead to additional pressure on the implementation and regulation of sustainability. Furthermore, EU has just launched the Sustainable Finance Disclosure Regulation (SFDR) (Doyle, 2021). The SFDR encloses various disclosure requirements for financial market participants and came into effect on March 10th, 2021. The purpose of this directive is to provide greater transparency on the degree of sustainability of financial products. Together with recent events, the new regulation is expected to draw even more attention towards SRI strategies and the incorporation of ESG measures.

2.1.2 Defining SRI

To define the area of this thesis, it is found necessary to elaborate upon how SRI is interpreted and understood. Even though SRI has become a worldwide phenomenon and adopted by a large pool of mainstream investors, there is still no consensus in the terminology. This is partly due to the fact that no legislative standard is in place for how a company should measure and report its non-financial performance, which will be elaborated further in section 2.1.4. This outlines the challenges and opportunities for both the companies, investors, and fund managers regarding SRI disclosures.

The areas of SRI lay beyond the traditional framework of portfolio theory as it does not focus on the standard financial performance measures, including risk and return. In contrary, SRI is referring to the non-financial aspects of an investment decision, such as ESG. The practices of SRI are hereby departing from the traditional shareholder capitalism and turning towards stakeholder capitalism. Shareholder capitalism was first presented by Milton Friedman in 1970 in his publication “The Social Responsibility of Business is to Increase its Profits” (Friedman, 1970). The idea was that a company’s only responsibility was to increase shareholder value and thereby did not hold any social responsibility. The publication received considerable attention and became highly influential within corporate governance but is today perceived controversial by most (Tepper, 2020).

In contrary, stakeholder capitalism is about serving the interests of all relevant stakeholders, i.e. employees, customers, suppliers, and communities. Stakeholder theory was first described by R. Edward Freeman in his book “Strategic Management: A Stakeholder Approach” (1984). Freeman’s theory suggests that a company’s real success lies in serving the interest of all relevant stakeholders. The ideology focuses on the long-term value creation and not merely maximizing profits to enhance shareholder value. Although, the theory was first described by Freeman in the 1980’s, the concept has a longer history. Stakeholder capitalism was a popular management theory in the 1950s to 1960s, and after many years of shareholder capitalism being the prevailing ideology, it made a significant comeback (Sundheim & Starr, 2020). The comeback was spurred by the increasing environmental and social challenges. Today, it is no longer acceptable to only focus on financial prosperity and avoid the social and environmental impacts of investments, which underlies the key concept of SRI. However, the enhanced focus on stakeholders, and hereby social obligations, presents a new issue regarding the definition and measurement of relevant performance indicators.

The Europe-based Sustainable Investment Fora (Eurosif) is an organisation with a mission to promote and define sustainable investing (Eurosif, n.d.). Through the years, they have worked intensely to define a common language for SRI investors, as the lack of definition leads to greenwashing and barriers for SRI investing. Greenwashing is when companies provide misleading ESG disclosures in order to appear more transparent and ‘greener’ (Yu, Luu, & Chen, 2020). The main issue in the SRI definition debate is that no specific requirements exist on how sustainability preferences should be included and measured within a portfolio (Eurosif, 2018). Hereby, the inclusion and measurement of a company’s ethical performance become very subjective. The lack of definitions, clear measuring metrics, and legislative standards to govern the practices, also have a negative impact on the already existing information symmetry between clients and their investment advisors. Thus, Eurosif (2018) have developed their own definition of SRI to address the lack of consistency.

> *"Sustainable and responsible investment (SRI) is a long-term oriented investment approach which integrates ESG factors in the research, analysis and selection process of securities within an investment portfolio. It combines fundamental analysis and engagement with an evaluation of ESG factors in order to better capture long term returns for investors, and to benefit society by influencing the behaviour of companies."* (Eurosif, 2018, pp. 12)

This definition combined with the one presented previously from the Brundtland report (1987) illustrate that SRI is an investment strategy that focuses on the longer-term gain, both socially and financially. These two definitions are accompanying the context of this thesis well and have hereby been chosen as the conceptual framework.

2.1.3 ESG

As mentioned in the definition above, ESG is a fundamental part of SRI practices and is often the prominent expression when speaking of non-financial investment measures. ESG is an abbreviation for Environmental, Social and Governance, which typically define the areas that companies are scored within when measuring their social performance. This is also the case in this thesis, as the companies' ESG scores will be used to allocate stocks to their respective portfolios. It is hereby found relevant to elaborate upon ESG; its definition, application, and three fundamental areas, namely E, S, and G.

ESG is a set of criteria standards that socially responsible investors use to screen potential investments. MSCI (n.d.) defines three drivers behind incorporating ESG: Integration, Values and Impact. Integration is about the systematic inclusion of ESG risk and opportunities to enhance the long-term risk-adjusted return. Some investors believe that the inclusion of ESG scoring metrics will help avoid greater financial risk and hereby increase returns, which will be explained further in section 2.3. Values is about investing according to an organisation's or individual's moral beliefs, similar to the origin of SRI. The last objective is impact and refers to investing with the intention to have a positive impact on social or environmental areas.

Environmental performance is often the dominating category when discussing the areas within SRI (Berry & Junkus, 2010). Environmental challenges have gotten more and more attention through the last decade, which may cause its dominating presence when scoring companies' non-financial performance. Furthermore, the environmental measures can be easier to score, compared to social and governance measures, because the nature of environmental measures are typically quantifiable. Common environmental measures include waste, recycling, energy and water consumption, resource use, and pollution (Berry & Junkus, 2010). Thus, the factor mainly refers to how the company responds to climate changes and the risks presented hereof. The main focusses of the social factor are employees and local communities. Examples of social measures are typically human rights, labour and supply chain practices, health and safety, community and local environment, diversity, and minority contracting. Lastly, the governance factor is focused on the rules and principles set forward by the management. Examples of governance measures are reporting, accountability, code of conduct, board independence, transparency, management of ethical issues, structure, and bribery. As visible through the listing of measures, some are quite hard to quantify, especially when no regulatory standards are in place. This automatically leads to subjective interpretations and methodologies generating a consensus gap, which will be addressed in the following section.

2.1.4 Data Issues

One of the main challenges faced by investors, wanting to follow an SRI approach, is the data inconsistency. The problem is mainly due to the variety of methodologies applied by different data providers leading to heterogenous outcomes. In essence, the variations are caused by two parameters, 1) ESG performance is very difficult to measure and 2) the lack of legislative standards. The result of the general inconsistency is that investors do not have a solid foundation on which they can define ESG leaders and laggards and hereby build their investment decision.

ESG data providers play an important role in the investment process. The scoring of companies' ESG performance is challenging work and requires a significant amount of data processing, whereas many investors choose to outsource this process to specialized data providers. The ESG landscape is characterized by a large number of various providers, including non-financial rating agencies, data providers, benchmark providers, credit rating agencies, among others (Scherpenzeel, n.d.). Most well-known global data providers within financial information are reporting ESG scores on a selected list of companies, such as MSCI, Moody's, S&P Global, and Refinitiv. Moreover, a number of specialized research firms have been emerging in recent years, such as Sustainalytics and Matter, whose primary purpose is to define and score companies' ESG performance. The data providers are thoroughly collecting and analysing information on companies' non-financial performance, whereafter the data points are aggregated into overall ESG scores. However, the ESG rankings vary significantly depending on which data provider is chosen (Berg, Koelbel, & Rigobon, 2020).

A recent study shows an average correlation of 0.54 between the scores provided by six of the most prominent scoring agencies, namely MSCI Stats, Sustainalytics, Moody's, S&P Global, Refinitiv, and MSCI (Berg et al., 2020). Another study even reports negative pairwise correlation between some of the data providers' scorings

(UN PRI, 2020). Compared to credit rating agencies that normally have a correlation of 0.99, this finding is highly disturbing. The low correlation is a result of different methodologies underpinning each approach. More specifically, Berg et al. (2020) highlights three distinct sources of divergence: scope, measurement, and weights. Scope divergence refers to the different sets of attributes used by providers to define the concept of ESG performance. Measurement divergence is the appliance of different indicators to measure the same attribute, e.g. labour practices can both be measured through policies or outcomes. Lastly, the difference can be found in the weight divergence, which refers to the relative importance given to different attributes. These factors are highlighting the general problem in ESG investing, which is that ESG performance is extremely difficult to measure due to its 'soft' nature.

Another important issue is that no uniform requirements for reporting ESG information exist (MacMahon, 2020). This indicates that the data being extracted by the scoring agencies are highly incomparable across companies. Consequently, the data inputs are less structured, less complete, and of lower quality than financial reporting data (MacMahon, 2020). Moreover, the ratings are often based on how much and what the company chooses to disclose (Scherpenzeel, n.d.). Besides not having a regulatory standard in place for how the companies should report on ESG matters, the disclosures are most often unaudited. Therefore, a rating assessment exclusively based on company disclosure will lack impartibility and may lead to a misinterpreted ESG performance.

The data issue presents a major problem for investors, especially as the regulatory pressure increases. Furthermore, the inconsistency in the data providers' scoring methodologies sustain the lack of definition and fail to provide an unambiguous conclusion towards the financial prospective of following an SRI strategy (see section 2.4). Investors, and other relevant stakeholders, are hereby calling out for a more standardized reporting guideline that can provide greater transparency and consistency.

2.2 Investment Strategies

Investors apply various strategic approaches to incorporate sustainable measures. Aforementioned, SRI practices involve a high level of subjectivity. Therefore, the investor will need to define the meaning and application. The definition is vital as the chosen strategy ultimately determines the securities available to invest in. In the following sections, the investment strategies, most commonly applied by investors, will be presented.

2.2.1 Negative Screening

Negative screening is the practice of actively excluding certain investments. This strategy is an SRI approach where the investor excludes companies based on certain criteria or ethical beliefs. In practice, the investor will filter out companies that appear to engage in, what is presumed, 'unethical' business practices. These companies are often referred to as ‘sin’ stocks, adopted by the origins of SRI where products were excluded if they were ‘sinful’ (Eurosif, 2018). Common exclusion criteria include alcohol, tobacco, gambling, and weapons. However, in recent years more diversified criteria have emerged, such as animal testing, environment, human rights, labour relations, employment/equality, community investment, and proxy voting. The screenings can also refer to third-party companies, i.e. companies engaging with sin stocks, also known as second-order screening (Ibid). The risk of employing negative screening is that it may constrain the portfolio from certain geographical areas or sectors, resulting in a less diversified portfolio. Conversely, the exclusionary approach can also be used as a reputational safeguard for major investment funds or financial service providers, as they can avoid scandals and criticism for legitimising un-ethical behaviour (Camilleri, 2020). Several studies have been performed to examine the economic effect of excluding sin stocks, which will be addressed further in section 2.4.2. Nonetheless, the motives behind performing negative screening may be other than financial. Some investors see the exclusion of certain products as a direct license to operate based on the interest of beneficiaries (Ibid).

2.2.2 Positive Screening

As opposed to negative screening, positive screening is an inclusionary approach where the investor search for the best performing companies. The strategy is referred to as the second-generation screening, where negative screening represents the first generation. Positive screening practices are often more complex than negative screening, as the answer is not clear-cut and requires intensive analysis of underlying scoring metrics, such as the ESG pillars (Blowfield & Murray, 2008). In practice, the investor following a positive screening strategy will start out by ranking companies according to their ESG performance, whereafter the highest-ranking companies will undergo a conventional financial performance analysis (Eurosif, 2012). By incorporating ESG information, the investor will be able to uncover possible risks and opportunities, and hereby get a better representation of the company’s performance and exposure going forward. Closely related to positive screening is the “best-in-class” screening approach, where the investor will search for companies with the best ESG score within a certain sector or industry. Thus, a company within a ‘sin’ industry may be included in the portfolio, if it is top ranking among its peers. This approach will hereby encourage more companies to implement ESG practices that otherwise would have been excluded, simply because of the industry in which they operate. Furthermore, the “best-in-class” strategy also gives companies an incentive to keep improving their environmental and social performance (Wagemans et al., 2013).

2.2.3 Shareholder Activism

Shareholder activism is an active investment approach that, opposed to the screening strategies, takes place in the post-investment phase. This approach involves, among others, engagement and voting practices, where investors use their shareholder rights to advocate a certain environmental or social agenda. The strategy is hereby taking a proactive approach towards influencing the management of the company to do better by being good. In practice, the shareholders use their voting rights and file proposals for their annual shareholder meetings to pressure the company to improve their environmental and social performance. Enacting this type of strategy requires voting power, whereas the practices are more often used by larger investment funds or organisations rather than by individual investors. The companies targeted by shareholder activism are often large, well-known, manufacturing companies dominating the stock indices. The reasoning behind these firms being targeted is their familiarity among the public, as well as their economic and environmental significance. Some of the typical issues brought up in filed proposals are international conduct, environmental issues, antidiscrimination, corporate governance, productional changes, or reporting disclosures (Wagemans et al., 2013).

2.3 Supporters and Opponents

Based on the literary attention that SRI has been given in the recent years, it is very clear that investors have different opinions towards the economic and social prospective of following this investment approach. In the following sections, the arguments from both the opponents and supporters will be presented.

The criticism of SRI is often referring to Markowitz' (1952) portfolio theory, more specifically his mean-variance optimization rule (Revelli & Viviani, 2015). In the market there exist two types of risks: systematic and unsystematic risk. Systematic risk is referring to market risk, whereas unsystematic risk is firm specific. Modern portfolio theory argues that an investor is able to eliminate all unsystematic risk through diversification (Markowitz, 1952). For this reason, the investor will not be compensated, in terms of return, by undertaking unsystematic risk. Diversification is gained by including multiple assets into the portfolio that are not perfectly correlated. Based on this theoretical perspective, opponents argue that SRI can never become an optimal investment strategy, because it holds a 'diversification cost' (Revelli & Viviani, 2015). The reason hereof, is that following an SRI strategy often involves a constrained portfolio, due to the deselection of un-ethical investments, leading to reduced diversification and hereby risk-adjusted return.

The SRI supporters acknowledge the existence of the diversification effect on portfolio performance, argued by the opponents. However, the diversification cost is claimed to be less or non-existent in the context of SRI, because the excluded investments are assumed to be lower-performing companies (Revelli & Viviani, 2015). One argument is that leading socially responsible companies are less vulnerable to environmental accidents, lawsuits, or other social scandals leading to additional costs and reputational losses. An example hereof is the 'Deepwater Horizon' accident in 2010, which led to a significant pollution of the Mexican Golf and resulted in a 34% decrease in British Petroleum's share price (Amadeo, 2020). By focusing on social performance, companies will constantly be tracking and mitigating risk which will improve governance and management horizon (Revelli & Viviani, 2015). Hereby, it is argued that socially responsible companies are generally facing lower risk than their conventional counterparts, whereas the diversification cost will be non-existent. Furthermore, some find that under-diversification has almost no impact on portfolio performance, because the stock markets are highly vast, liquid, and efficient (Ibid).

Another argument, proposed by SRI supporters, is addressing the operational performance of socially responsible companies. SRI-focused companies are expected to reap the benefits of taking actions towards being more responsible, because they are better at attracting and keeping employees and consumers (Berry & Junkus, 2010). Ethical performance is becoming more and more important in the eyes of consumers, as they are not interested in supporting companies engaging in un-ethical business practices. Employees are also more likely to be attracted to a company that has a social agenda with focus on areas such as treatment of employees, safety and health regulation, and other similar issues. Continued focus on these areas will keep the employees motivated which will result in higher productivity and performance. Lastly, supporters argue that environmental considerations can lead to financial gains through reduced material use and other cost-effective solutions (Schaltegger & Figge, 2000). These arguments are supporting the idea that firms will do well by doing good, while justifying the implementation of ESG metrics in investment decisions (Berry & Junkus, 2010).

2.4 Previous Studies

In the following section, a review of previous empirical studies will be performed. The field of study has been widely examined using various methodologies and scopes, which have led to heterogenous results. The following section will give a comprehended view of the most common approaches. Due to the scope of this thesis, the section will be focused on studies involving portfolio creation and not mutual fund findings. Furthermore, relevant meta studies will be presented, in order to give a more comprehended view on the overall conclusion in the debate. Lastly, this thesis' contribution to the field of study will be presented by comparing its methodology to previous studies.

2.4.1 Positive Screening Studies

Several portfolio-based studies are applying a positive screening approach to identify a potential relationship between corporate social performance and corporate financial performance. More specifically, these studies formulate portfolios based on companies' ESG scores or other identified social measures. After ranking the companies, the researchers will typically allocate companies to high and low portfolios according to their ESG scores. Hereafter, the portfolios are being tested using a variety of theoretical models, such as CAPM, Fama-French 3-Factor, Carhart 4-Factor, and Fama-French 5-Factor. The models are used to measure risk-adjusted returns and to determine abnormal returns, which will be compared across the high and low portfolios or to conventional peers. Even though the theoretical approaches are somewhat similar, the results are highly inconsistent, concluding both positive, negative, and neutral relationships between SRI and financial performance.

Kempf & Osthoff (2007) examines the performance of constructed portfolios based on various socially responsible criteria. The results are based on KLD's SRI ratings of US stocks and covers the period 1992 to 2004. The researchers follow the positive screening approach by forming a high and low portfolio, where the high-rated consists of the top 10% stocks and the low-rated consists of the bottom 10% stocks. Furthermore, they also follow the "best-in-class" screening approach in order to avoid a possible industry bias. For this purpose, companies are divided into ten different industries, according to their SIC code, and hereafter allocated to either a high or low portfolio similar to the positive screening approach. By applying the Carhart model, they find that the high-rated portfolio performs better than the low-rated portfolio. Furthermore, they test a long-short strategy, with a long position in the high portfolio and a short position in the low portfolio, that returns a positive alpha up to 8.7% a year. The highest alpha is obtained from the "best-in-class" screening approach, which indicate that the result is not only caused by a sector displacement. The study hereby concludes a positive relationship between SRI and financial performance.

Similar to Kempf & Osthoff (2007), Statman & Glushkov (2009) also examine the relative performance of socially responsible portfolios and conventional portfolios based on KLD's SRI ratings of US stocks. However, their study has a longer period of analysis, namely 1992 to 2007. The portfolios are constructed based on a positive screening approach that applies 30% cut-off points. The high portfolio is hereby including the top third ranking companies, whereas the low portfolio includes the bottom third. Furthermore, the researchers are applying a long-short strategy similar to Kempf & Osthoff (2007). The portfolios are tested using three different performance benchmarks: CAPM, Fama-French 3-Factor, and Carhart 4-Factor. The results provide strong positive alphas in favour of the high-scoring stocks. However, the study also concludes that the outperformance is offset by the abnormal performance observed in the shunned stocks portfolios, which will be addressed further in section 2.4.2. The overall net effect is hereby supporting the "no effect" hypothesis, stating that "*the expected returns of socially responsible stocks are approximately equal to the expected returns of conventional stocks*" (Statman & Glushkov, 2009, pp. 44).

Halbritter & Dorfleitner (2015) examine the link between social and financial performance based on ESG scores provided by three different sources, namely Asset 4 (Refinitiv), Bloomberg, and KLD. Their results are based on the US market from 1991 to 2012. The authors are constructing two value-weighted portfolios for each respective score, that is ESG, E, S, and G, based on 20% cut-off points. Hereafter, the portfolios' returns are tested using the Carhart 4-Factor model and the Fama-MacBeth regression model. The results provided by the three different data sources were inconsistent. For the long-short portfolio, the score providers Asset4, Bloomberg and KLD provided a positive, negative and neutral alpha, respectively. However, the alphas were not statistically significant. Therefore, the overall conclusion was that the high and low portfolios did not exhibit performance differences. The conclusion was true across all scores and for various robustness checks. Although the alphas proved to be insignificant, the study still found a strong dependence between the chosen rating agency and the outcome.

Mark K. Pyles (2020) has studied the financial performance of high and low portfolios using ESG data from Bloomberg. The period of analysis is 2011 to 2017 and the S&P 500 index constitute the asset universe. The main hypothesis being tested, is whether companies with higher ESG scores have superior returns compared to lower-scoring companies. Similar to previous studies, this study follows a positive screening approach by constructing a high and low portfolio based on 20% cut-off points. Hereafter, the portfolios' returns are tested using the Fama-French 5-factor model. The results show that the 20% highest ranking firms experience lower abnormal returns than the 20% lowest ranking firms. The differences between the portfolios' abnormal returns demonstrate both statistical and economical significance. To examine the results closer, Pyles (2020) chooses to focus the attention on the characteristics of the firms. Here, he observes some common characteristics of the firms in the higher-scoring portfolios, including significantly greater size, higher dividend yields, and lower profitability. After controlling for these elements, the abnormal returns become insignificant. The empirical study (Pyles, 2020) concludes that the ESG scores, disclosed by Bloomberg, show no significant alpha which indicate a neutral standpoint towards SRI and financial performance.

2.4.2 Negative Screening Studies

Another common screening approach is negative screening, which is often applied by researchers to identify possible gains or losses associated with the avoidance of sin stocks. Due to societal norms, institutional investors are often avoiding certain industries, such as tobacco, alcohol, gambling, pornography, and weapons. Hereby, researchers have found it interesting to study whether investors face an additional cost by complying to these norm-constraints. Studies applying negative screening are often starting out by identifying "sinful" stocks involved in some sort of controversial business area. The identified sinners will be placed in one portfolio, whereas the rest of the asset universe will constitute the second portfolio. Hereafter, the portfolios' performances are tested, in a similar manner as previously addressed, using performance benchmark models.

Hong & Kacperczyk (2009) are studying the effect of social norms on returns. The researchers are testing the hypothesis that investors abstaining from sin stocks pay a financial cost. Their analysis is based on US firms in the period 1965 to 2006. In order to test the hypothesis, the study forms a long-short portfolio, with a long position in the identified sinners and a short position in the socially accepted stocks. Hereafter, the performance is tested using the Carhart 4-Factor model. Consistent with the hypothesis, the results show that the sin stocks experience higher returns than the accepted stocks. More specifically, the study shows price effects in the order of 15-20% for investors shunning sin stocks. The authors argue that the observed premium is caused by institutional investors neglecting sin stocks to such a degree where they become depressed relative to their fundamental value. This is also referred to as the "shunned stock" hypothesis. Another argument presented is that sin stocks exhibit higher litigation risk which leads to increased expected returns (Hong & Kacperczyk, 2009).

A similar study has been conducted by Salaber (2007) in the European market over the period 1975 to 2006. This study classifies sin stocks as companies being involved in the following industries: tobacco, alcohol, and gambling. In addition to testing the existence of sin stock premiums, the study seeks to investigate a possible relationship between the premium and certain legal and cultural characteristics. By applying the Fama-French 3-Factor model, the study finds positive abnormal returns in the sin stocks. Furthermore, the study demonstrates that the level of excess returns is highly dependent on locally determined elements, such as religion, taxes, and litigation risk. In a more recent study, Blitz & Fabozzi (2017) seek to uncover the drivers behind the sin premiums observed in previous studies. By applying the Fama-French 5-Factor model, the study finds that the sin premium can be fully explained by the two quality factors, namely profitability (RMW) and investments (CMA). In other words, after controlling for these factors the sin stocks do not provide any premium.

2.4.3 Meta Studies

In a more comprehensive study, Revelli & Viviani (2015) have examined the financial performance of SRI in a meta-analysis that also addresses the inconsistency in results and methodologies of previous studies. A meta-analysis is a statistical technique that examines results from previously performed independent studies to identify overall trends and causality. The analysis consists of 85 studies covering 190 experiments throughout a 40-year time period (1972-2012). The studies included in the analysis are tested according to their concluding size effects which is used to test for heterogeneity in results. The statistics show a high standard deviation of 0.74 and a large cap between the min and max, confirming heterogeneity. The diversity is also confirmed by the experiments' varying conclusions: 26% negative, 53% neutral, and 21% positive. Revelli & Viviani (2015) address different factors that may be causing the diversifying results, such as geographical market, ESG factor focus, investment horizon, portfolio constraints, and financial performance measures. However, one of the key arguments presented is that research within SRI is highly data driven. The lack of result consensus in previous studies may be affected by the use of different data provided by the numerous ESG rating agencies.

To give a comprehensive conclusion of the combined empirical studies, Revelli & Viviani (2015) aggregate the size effects from the 190 experiments. The size effect is tested insignificant illustrating that SRI has no real effect on financial performance (Revelli & Viviani, 2015). As the analysis include studies from a 40-year period, the authors have also tested whether the time element has an impact on the results. The coefficient proved insignificant, indicating that the results would not differ using a different time period. The meta-analysis hereby overcomes the previous lack of consensus by aggregating the results, and it concludes that SRI does not add financial cost nor benefits compared to conventional investments. The conclusion is supported by a more recent meta-analysis performed by Kim (2019), whose results also suggest that the performance of SRI investments are no different than its conventional counterpart. However, the authors are still promoting SRI as being the preferred investment strategy, as the investor can address ESG concerns without compromising financial returns (Revelli & Viviani, 2015; Kim, 2019).

2.4.4 Own Contribution

As illustrated in the sections above, a variety of research studies have been conducted on SRI portfolio performance. However, the results do not provide an unambiguous conclusion towards the profitability of following an SRI strategy. This may be due to the different methodologies applied and the general data issues described in section 2.1.4. Therefore, the area is still found relevant to study, especially looking at the developments in the current COVID-19 environment that may present new findings.

The research conducted has many similarities to the previous positive screening studies, but it also holds some differences. In a similar manner as Kempf & Osthoff (2007), the link between social and financial performance will be studied through constructing high and low portfolios, based on defined cut-off points. More specifically, this study will apply the positive screening approach to construct six portfolios, namely A to F, that in addition to the high and low portfolios also includes the mid-scoring companies. Moreover, a "best-in-class" screening approach will be applied, similar to Halbritter & Dorfleitner (2015), to test if the results are subjects to sector bias. The portfolios' performances will then be tested using various performance benchmarks, similar to Statman and Glushkov (2009), in order to identify possible abnormal returns. The performance benchmarks applied in this study are: CAPM, Fama-French 3-, and 5-Factor models. A main difference in this study, which also represents its contribution, is the thorough analysis of the portfolios' risk and return attributes. This includes the analysis of downside risk, which is an important element of an investment's risk profile. Furthermore, this thesis is only focused on the European market where most previous studies are focused on the US market. Europe is an interesting market to study, as they are considered the frontrunners both in terms of ESG investing and the green agenda. Lastly, this study is found relevant as it includes 2020 in the period of analysis, that contains the first year of COVID-19, where a significant change in the ESG environment was observed.

Chapter 3
Theory

In the following chapter the theoretical framework, on which this empirical study is built upon, will be presented. The purpose of the chapter is to give a fundamental understanding of investment theory and portfolio performance, as these are the main areas of this thesis. The section will present descriptions of basic return and risk properties, factor models, and performance measures, all relevant parameters to conduct the intended portfolio analysis.

3.1 Modern Portfolio Theory

Harry Markowitz' Modern Portfolio Theory (MPT) (Markowitz, 1952) is a framework set out to assist private investors in the construction of efficient portfolios. Markowitz (1952) introduced a diversification model that would select assets based on the trade-off between risk and return. The theory argues that a risk averse investor should seek to construct a portfolio that maximizes the expected return at a given level of market risk. Alternatively, the objective can be to minimize the risk of the portfolio at a given expected return. These objectives can also be described as the "*expected return – variance of return*" rule (Markowitz, 1952, pp. 77). The theory also advocates that risk can be reduced through diversification. By assessing how the investments co-move, i.e. measure the covariance between assets, the investor will be able to adjust the weights accordingly and hereby reduce the risk of the portfolio while maintaining the same level of expected return. Thus, MPT departed from the historical focus of analysing a single investment's risk and return characteristics, to a more holistic view of how the investment would have an impact on the entire portfolio performance.

The fundamental concept of MPT is the risk-return trade-off. Based on this concept, Markowitz (1952) introduced the efficient frontier that illustrates the optimal risk-return trade-offs for an investor. More specifically, it reflects the optimal portfolio combinations that generate the highest expected returns at different levels of risk. At each level of risk, the portfolio that gives the maximized expected return will form the efficient frontier curve and be classified as efficient. Portfolios that fall below the efficient frontier will not be optimal and classified as inefficient. The efficient frontier is formed as a hyperbola where the upward slope will hold the portfolios that are superior to others. The correlation between the securities will determine the diversification gain and shape the efficient frontier curve – the higher positive value the rounder the shape of the curve. The risk averse investors will hereby only invest in the portfolios on the efficient frontier that maximize their utility.

MPT assumes that a risk-averse investor would invest in multiple assets and prefer a less risky portfolio, as opposed to a riskier one, at a given level of expected return. Furthermore, MPT rests on the assumption that markets are efficient which is also known as the efficient market hypothesis (EMH) (Fama, 1970). EMH indicates that all stocks in the market are trading at their fair value, assuming that prices reflect all information. This assumption implies that an investor cannot generate abnormal returns by identifying mispriced assets in the market and trade accordingly, i.e. buy undervalued stocks and sell overvalued stocks. Therefore, it is impossible for an investor to beat the market unless he undertakes higher risks. However, opponents of EMH argue that it is possible to beat the market, because securities do not reflect all information and is often mispriced due to market inefficiencies (Malkiel, 2003).

3.2 Return and Risk Properties

As mentioned in the previous section, one of the basic concepts to understand when dealing with portfolio theory is the risk-return trade-off. It is hereby essential to study both the return and risk properties which will be introduced in the following sections. While return is an unambiguous concept, there are various statistical techniques to measure and quantify risk. Therefore, several risk measures will be described and used in the analysis, these being standard deviation, skewness, kurtosis, and maximum drawdown.

3.2.1 Return

The investment return is measured by its profitability over a given holding period, which can hold a capital gain and possibly a direct payment (Munk, 2019). The capital gain is calculated as the difference in the security's price within the holding period. Moreover, the investor could also receive a direct payment in the form of dividends, which also should be included in the return calculation. This return is known as the "Holding Period Return" and is calculated as follows:

$$R_t = \frac{D_t + P_t}{P_{t-1}} - 1$$

Where:

P_t and $P_{t+\Delta t}$ is the price of the security at the beginning and end of the holding period

$D_{t+\Delta t}$ is the dividends received in the holding period

This computation assumes that the dividends are being paid out and held in cash until the end of the period and ignores the possible gain from reinvesting the amount (Munk, 2019). Alternatively, it can be assumed that the dividends are being reinvested which magnifies the investment position. In this analysis, the return used is the 'Total Return Index' that incorporates the immediate reinvestment of dividends. Hence, the returns are calculated as follows:

$$R_t = \frac{P_t}{P_{t-1}} - 1$$

Where:

P_t and P_{t-1} is the Total Return Index at period t and $t-1$

The return calculated above is for a single security. To find the return of a portfolio, you simply take the weighted sum of the securities' returns, i.e. the proportional value of the investment in each asset multiplied with their respective returns (Munk, 2019). The formula for a portfolio with N assets is the following:

$$R_p = \sum_{i=1}^{N} w_i R_i$$

Where:

w_i is the weight of asset i

R_i is the return of asset i

The dataset obtained from Refinitiv (n.d.-a) reports the returns on a monthly basis. However, for the purpose of the preliminary examination of portfolio returns, it is easier to interpret and compare the annualized returns. The annualized return is found by simply compounding the periods' cross returns. The compounding formula is the following:

$$R_A = \left(1 + R_p\right)^{12} - 1$$

Where:

R_p is the average monthly portfolio return

3.2.2 Variance and Standard Deviation

Risk is often related to investment volatility, i.e. variations in returns, where investments with higher volatility are presumed riskier than investments with lower volatility. Common methods towards quantifying risk are the variance and standard deviation of returns. The variance is calculated as the investment's deviation from its expected mean return, which is the probability-weighted average of the return (Munk, 2019). In practice, this value is often found by the historical mean over the analysed period, which is the case in this thesis. The standard deviation, also referred to as the volatility, is simply the square-root of the variance. The mathematical expression for the standard deviation is the following:

$$\sigma_i = \sqrt{\sum_{t=1}^{T} (R_t - E[R])^2}$$

Where:

R_t is the return for a given date or period

$E[R]$ is the historical mean return

For a portfolio, the standard deviation is measured using the weight and variance of each asset, as well as the covariance between each pair of assets. The covariance term will be addressed in the next section. The formula for calculating the portfolio standard deviation with N assets is the following:

$$\sigma_p = \sqrt{\sum_{i=1}^{N} w_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{j \neq i}^{N} w_i w_j Cov(R_i, R_j)}$$

Where:

σ_i^2 is the variance of asset i

$Cov(R_i, R_j)$ is the covariance between asset i and j

3.2.3 Covariance and Correlation

Aforementioned, the covariance is necessary in order to calculate the portfolio standard deviation. The covariance determines how two assets co-move with each other, or expressed differently, the degree of linear relation between two random variables. If the covariance is positive, the assets will have a tendency to move in the same direction, whereas a negative expression indicate that they move in opposite directions (Munk, 2019). The formula for calculating the covariance over one period is the following:

$$Cov(R_i, R_j) = E(R_i R_j) - E(R_i)E(R_j)$$

Where:

R_i and R_j is the return of asset i and j

$E(R_i)$ and $E(R_j)$ is the historical mean return of asset i and j

As the covariance can assume all values it can be difficult to interpret. Another more intuitive concept is the correlation. Both concepts are measuring the relationship between two assets, the only difference is that the correlation is in a standardized form. Hereby, the correlation can only return a value in the range $[-1; +1]$ (Ibid). The formula for calculating the correlation between two assets are the following:

$$\rho_{ij} = \frac{Cov(R_i, R_j)}{\sigma_i \sigma_j}$$

Where:

Notion as above

The maximum value that the correlation can take is $+1$, which indicate that the two assets have a perfect positive linear relationship. Here, the assets will move in the same direction to the same degree. On the other end of the range is -1, which states that the assets have a perfect negative linear relationship, where the assets will move in different directions to the same degree. A correlation of zero indicate a non-linear relationship, which should not be mistaken for independence. The correlation can only measure linear relationships, whereas non-linearity cannot be detected. By combining assets with low or negative correlation, you can increase diversification and reduce volatility while maintaining the same level of expected return.

3.2.4 Skewness and Kurtosis

The standard deviation will often be supplemented with tests of the return's probability density function. These tests are called skewness and kurtosis which are highly useful in the determination of an asset's risk (Munk, 2019). The skewness is testing whether the returns are sitting symmetrically around the mean, such as the normal distribution curve, or if the distribution is skewed to one side. Any symmetrical or normal distribution will have a skewness of zero. If the return distribution has a positive skewness, it means that more than half of the distribution mass will sit above the mean. Opposite, a distribution with a negative skewness indicates that more returns will sit below the mean. This description enclose that an investor would prefer a positive, or left, skewness in returns. The mathematical expression for the skewness is:

$$Skew(R) = E\left[\frac{(R - E[R])^3}{\sigma_R^3}\right]$$

Where:

Notion as above

The kurtosis, or excess kurtosis, is focused on the tail ends of the return distribution and describes the probability of getting extreme returns (Ibid). The concentration of returns in the tails are compared to the normal distribution. From the notion presented in the formula below, the kurtosis of the normal distribution will be equal to zero. A probability density function with a positive kurtosis will be depicted by having fatter tails, which means that it will have a higher probability of extreme returns than the normal distribution. A negative kurtosis, on the other hand, will show slimmer tails and hereby have a higher probability of showing returns closer to the mean. Holding all things equal, an investment with a negative kurtosis will entail lower risk than a positive kurtosis (Ibid). The formula for calculating the kurtosis is as follows:

$$Kurt(R) = E\left[\frac{(R - E[R])^4}{\sigma_R^4}\right] - 3$$

Where:

Notion as above

3.2.5 Maximum Drawdown

Often the variance and standard deviation estimates are criticized for not being appropriate risk parameters, as they do not detect downside risk (Sortino & Meer, 1991). A portfolio that experiences frequent small losses will have the same variance as a portfolio that normally has stable returns but incur large losses in economic downturns. An investor may prefer the portfolio with frequent smaller losses as it will be easier to recover. Therefore, it is important to study the downside risk. As explained in the previous section, the skewness and kurtosis can illustrate some additional risk beyond the variance and standard deviation. However, other risk measures, that focus more on the downside volatility, also exist.

Many investors are more concerned with the downside risk, i.e. the left tail of the return's probability density function, as opposed to the overall volatility in returns. The downside risk can be studied through the maximum drawdown estimate, which is the largest observed loss from a peak-to-trough decline during a specific period (Pedersen, 2015). Finding the maximum drawdown of a portfolio requires several steps. First step is to find the cumulative return, which is indexed at 1 in period t_0 and hereafter follows the following formula:

$$CR_t = CR_{t-1} * \left(1 + R_p\right)$$

Where:

CR_{t-1} is the cumulative return in the previous period

R_p is the portfolio return

The second step involves the estimation of the High-Water Mark (HWM) which is the maximum value of cumulative return obtained throughout the period (Ibid). The HWM is calculated for each period in time and is written as follows:

$$HWM_t = Max_{t \leq T} CR_t$$

The third step involves the calculation of portfolio drawdown (DD) which can be defined as the downside volatility (Ibid). In practice, the drawdown is calculated as the cumulative loss since the last HWM. The portfolio drawdown can be calculated as follows:

$$DD_t = \frac{CR_t - HWM_t}{HWM_t}$$

The fourth and final step, is to find the maximum drawdown value across the entire period of analysis. The maximum drawdown should be the largest absolute value of DD and as this value is negative, the maximum will be found using the minimum function (Ibid). Henceforth, the maximum drawdown is calculated as follows:

$$MaxDD_T = Min_{t \leq T} DD_t$$

3.3 Factor Models

Besides the basic return and risk properties, portfolio performance can also be studied through factor models. Factor models are the results of economists' attempt to uncover the underlying mechanics that generate variability in returns. In other words, the purpose of factor models is to predict future returns. The results from the numerous studies are plenty, but some of the most widespread performance benchmark models are the CAPM, Fama-French 3-, and 5-Factor models, which will be introduced in the sections below. A theoretical model which incorporates ESG as an implicit factor does not exist. Factors evaluated are typically macroeconomic or quantitative firm-characteristics. However, most empirical studies involving ESG adapt the traditional factor models when measuring the performance of ESG portfolios. This approach is also applied in this thesis.

3.3.1 CAPM

One of the first convincing models trying to uncover the behaviour of financial markets, was the Capital Asset Pricing Model (CAPM) originally developed by William Sharpe (1964). CAPM is built on the concepts of Markowitz' (1952) mean-variance portfolio theory. Hence, one of the basic elements of CAPM is that the investor must be compensated for the risk endured by investing in the asset. Furthermore, the model argues that the risk premium is only defined by the market factor. In other words, the expected return required by an investor will be linearly related to the asset's covariance with the market portfolio. The market portfolio is an indication of a portfolio that holds all risky assets. Furthermore, CAPM introduces the assumption of a riskless market from which investors will be able to borrow or lend capital at a risk-free rate.

The standard CAPM model is known as the Sharpe-Lintner-Mossin CAPM, as it was independently developed by William Sharpe (1964), John Lintner (1965), and Jan Mossin (1966). The model expresses the expected return of an asset as follows:

$$E(R_i) = r_f + \beta_i\left[E(R_m) - r_f\right]$$

Where:

r_f is the risk-free interest rate

β_i is the beta of asset i

$E(R_m)$ is the expected market return

The expected return is hereby an expression of the risk-free rate and its beta multiplied by the market risk premium, which is the market portfolio return in excess of the risk-free rate. The risk-free rate is defined as the investment without risk, i.e. the return you can expect by having money in the bank. Beta is the core element of CAPM, which measures the systematic risk of the asset compared to the risk that the market portfolio holds. Systematic risk is an expression of general market risk which is non-diversifiable, as opposed to unsystematic risk that is company specific and therefore diversifiable. Hereby, beta is a concept that describes the additional expected return required by an investor who invests in asset i instead of the market portfolio. The beta can be written as:

$$\beta_i = \frac{Cov(R_i, R_m)}{\sigma_m^2}$$

Where:

$Cov(R_i, R_m)$ is the covariance between asset i and the market portfolio m

σ_m^2 is the variance of the market portfolio

The model assumes that the market portfolio holds a beta of 1 and the risk-free interest rate has a beta of 0 (Munk, 2019). Hence, if the asset has a beta above 1, it means that the investment has more volatile returns than the market portfolio and the investor would require higher expected returns. Conversely, if the asset beta is between 0 and 1 the investment is said to be less volatile than the market portfolio. The reasoning behind the theory is that, in a market where everyone can borrow and lend at the risk-free rate, an investor would not accept a return lower than that of the market portfolio at a given level of risk.

3.3.2 Fama-French Three-Factor Model

CAPM is a single factor model that assumes that the expected return can be explained by the market portfolio return. However, later theory has challenged its restrictive assumptions. A main critique point is that other factors besides the market factor are relevant in the estimation of expected return. A commonly known and widely adapted multi-factor model is the Fama-French 3-Factor model, introduced by Eugene Fama & Kenneth French (1993). Fama & French (1993) found that the simple CAPM model was not sufficient in explaining asset returns. In their study, they identified two firm characteristics which, in addition to the market factor, were better at explaining the variability in returns. These factor premiums were related to size risk and value risk, namely SMB (Small-Minus-Big) and HML (High-Minus-Low). The model can be written as follows:

$$R_{it} - R_{ft} = \alpha_i + \beta_{iM}\left(R_{Mt} - R_{ft}\right) + \beta_{iSMB}SMB_t + \beta_{iHML}HML_t + \epsilon_{it}$$

Where:

$R_{it} - R_{ft}$ is the return on asset i in excess of the risk-free rate

α_i is the return not explained by the three factors

β_i are the assets' sensitivity to the respective factors: Market, SMB and HML

$R_{Mt} - R_{ft}$ is the market return in excess of the risk-free rate

SMB_t HML_t are the returns of the SMB and HML factors

e_{it} is the error term and represent the idiosyncratic risk

The SMB factor is trying to incorporate a size-effect, based on the anomaly found in empirical research, arguing that small stocks tend to outperform large stocks. Here, the size is referring to market cap and is calculated as the return on a small stock portfolio in excess of the return on a large stock portfolio. The HML factor is trying to incorporate the value-effect presented in earlier empirical studies, stating that value stocks tend to outperform growth stocks. Value is referring to the Book-to-Market (B/M) ratio, which will be high for value stocks and low for growth stocks. The factor is calculated as the return on a portfolio consisting of companies with high a B/M ratio in excess of the return on a portfolio consisting of companies with a low B/M ratio.

3.3.3 Fama-French Five-Factor Model

In more recent years, the Fama-French 3-Factor model has been extended to the Fama-French 5-Factor model (Fama & French, 2015). Beyond the market, size, and value factors, the 5-Factor model incorporates two additional risk factors, namely RMW (Robust-Minus-Weak) and CMA (Constructive-Minus-Weak). The model can be written as follows:

$$R_{it} - R_{ft} = \alpha_i + \beta_{iM}\left(R_{Mt} - R_{ft}\right) + \beta_{iSMB}SMB_t + \beta_{iHML}HML_t + \beta_{iRMW}RMW_t + \beta_{iCMA}CMA_t + \epsilon_{it}$$

Where:

$R_{it} - R_{ft}$ is the return on asset i in excess of the risk-free rate

α_i is the return not explained by the five factors

β_i are the assets' sensitivity to the respective factors: Market, SMB, HML, RMW, CMA

$R_{Mt} - R_{ft}$ is the market return in excess of the risk-free rate

SMB_t HML_t RMW_t CMA_t are the returns of the factors: Market, SMB, HML, RMW, CMA

e_{it} is the error term and represent the idiosyncratic risk

The RMW factor is trying to incorporate the profitability factor, which is proxying the return anomaly that more profitable companies generally generate higher returns than less profitable ones. The theory defines profitability as operating profitability (Fama & French, 2015). Hereby, the factor return is calculated as the return of a robust operating profitability portfolio minus the return of a weak operating profitability portfolio. The CMA factor is incorporating the investment factor, stating that companies following a conservative investment strategy have higher returns than companies following an aggressive investment strategy. Here, investment is defined as the change in assets, where a large change would indicate an aggressive strategy. The factor return is calculated as the return of a portfolio comprising of companies with a conservative investing strategy minus the return of a portfolio comprising of companies with an aggressive investment strategy.

Both the Fama-French 3- and 5-Factor models were attempts to create a model that was better at explaining the variability in returns than the CAPM. In their respective papers, both studies showed higher explanatory power than CAPM in general (Fama & French, 1993; 2015). However, the Fama-French models are not perfect either. The authors acknowledge that the empirical evidence fail to explain why the additional risk factors are performing well and what the underlying mechanisms could be. They also argue that the factors could be proxying other yet-unknown risk parameters. As foretold, neither the CAPM, Fama-French 3-, or 5-Factor models are perfect at explaining returns. It is thereby considered relevant to perform all three factor models, when testing the performance of the constructed ESG portfolios.

3.3.4 OLS

The topic that recurs often throughout the theory section is the application of regression, which is a method used to determine the relationship between two variables. In this thesis, regressions are used to perform the aforementioned performance benchmark models. This underpins the importance of understanding the method used to perform regressions, that is Ordinary Least Squares (OLS) (Stock & Watson, 2015). OLS is a statistical method of analysis used to determine the linear relationship between two variables. The following equation illustrates the simple form with one regressor:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

Where:

Y_i is the dependent variable
β_0 is the intercept
β_1 is the slope of the fitted line
X_i is the explanatory variable
u_i is the error term

The OLS method involves the estimation of the best linear line that fits the observations in a scatter plot. It is uncommon to have a perfect linear relationship between two variables, which is why an error term will always exist in predictions. Thus, the best fitted line becomes the one that minimizes the sum of squared residuals (Ibid). Residuals are measured as the vertical distances between the fitted line and the points on the scatter plot. Hereafter, the residuals are squared to avoid the negative and positive distances to cancel each other out. The following expression illustrates the objective of OLS:

$$u_i = Y_i - \beta_0 - \beta_1 X_i$$

$$Min\left\{\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2\right\}$$

Where:

Notion as above

Taking the first order derivative of the objective function, with respect to the OLS estimators, makes it possible to predict the values of β_0 and β_1. This step gives us the following two equations:

$$\hat{\beta}_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

Where:

$\hat{\beta}_1$ *is the estimated slope of the fitted line*
$\hat{\beta}_0$ is the estimated intercept

Combining the OLS estimators results in the regression population line, which will be a straight line with an intercept equal to $\hat{\beta}_0$ and a slope equal to $\hat{\beta}_1$ (Ibid).

The simple form illustrated in the equation above represents the CAPM model introduced in section 3.3.1, where the dependent variable is the expected return, and the explanatory variable is the market portfolio. However, the model can be extended to include multiple explanatory variables, which are used in both the Fama-French 3- and 5-Factor models introduced above. The OLS equation with multiple regressors can be written as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \cdots + \beta_k X_{ki} + u_i$$

Where:

Notion as above

k is the number of explanatory variables

The model illustrated in the equation above allows you to measure the impact of multiple factors on the expected return. As the model now has more than two dimensions, it becomes difficult to visualize the line in a scatter plot. However, the objective of the regression will rest on the same practices, i.e. minimizing the sum of squared residuals.

The OLS regression is built on four assumptions, which needs to hold in order for the estimated variables to be consistent and reliable (Ibid). The first assumption is that the error term will be equal to zero on average. This assumption relies on the circumstance that some predictions of Y_i will be overestimated, indicating that the true value will fall below the population line. At the same time, others will be underestimated where the true value will lie above the line. However, on average the over- and underestimations will even out, and the error term will be equal to zero. The second assumption states that the dependent and independent variables are not systematically biased. This implies that the variables should be independent and identically distributed. This is an essential part of OLS, as the measured impact of a factor may be manipulated if the variables are not independent. The third assumption are saying that large outliers are unlikely. This is necessary as a large outlier will have a significant impact on the estimation of errors, especially as these are being squared. If large outliers are being identified in the data set, it could be relevant to exclude these from the analysis. The fourth and last assumption states that no perfect multicollinearity should exist. Multicollinearity can occur in a multi-variable regression analysis and describes the phenomena where one of the independent variables are perfectly correlated with another variable. If multicollinearity exists in the dataset, the estimation of regression parameters is not possible (Ibid).

3.4 Performance Measures

The constructed portfolios will be measured according to their risk-adjusted returns using common performance measurement tools. Following the methodology of other studies regarding SRI, this empirical study will apply the Sharpe ratio, Treynor ratio, and Jensen's alpha to analyse the portfolios' performances.

3.4.1 Sharpe Ratio

The Sharpe ratio is developed by, and named after, William Sharpe (1966) and is a common financial metric used to measure the performance of investment assets. The ratio was originally referred to as the *reward-to-variability ratio*, as it measures the average return achieved in excess of the risk-free rate per unit of volatility. After getting considerable attention, Sharpe (1994) revised and renamed the ratio in his paper "The Sharpe

Ratio". The measure is closely related to Markowitz' (1952) MPT, as it argues that you should be compensated for enduring more risk. The mathematical expression for the Sharpe ratio is as follows:

$$Sharpe = \frac{R_p - r_f}{\sigma_p}$$

Where:

r_f is the risk-free rate

R_p is the portfolio return

σ_p is the standard deviation of the portfolio

In theory, the risk-free rate is an investment return with zero risk. Thereby, the investor can isolate the return associated with undertaking risk and compare it to the level of risk endured. The Sharpe ratio can be used to find the most optimal portfolio on the efficient frontier, i.e. the one that yields the highest ratio and thereby the highest risk-adjusted return. The optimal portfolio should not be mistaken for the portfolio that provides the highest return. However, the investors can utilize the risk-free market to adopt the preferred risk-return trade-off and scale the investment up or down depending on risk aversion. Furthermore, the Sharpe ratio can both evaluate the portfolio's past and expected performance, by using either historical or expected measures.

3.4.2 Treynor Ratio

The Treynor ratio is developed by Jack Treynor (1965) and is a performance measure similar to the Sharpe ratio (1994). The difference between the Treynor and Sharpe ratio is the risk measure. The Treynor ratio was originally called the *reward-to-volatility* ratio and measures the average return achieved in excess of the risk-free rate per unit of systematic risk, i.e. the portfolio beta. In other words, it measures the additional risk endured by holding the risky portfolio compared to holding the market portfolio. The relationship between the portfolio and the market is measured in beta, whereby the Treynor ratio can be written as:

$$Treynor\ ratio = \frac{R_p - r_f}{\beta_p}$$

Where:

Notion as above

β_p is the portfolio beta

Similar to the Sharpe ratio, a higher value indicates a better portfolio. The main difference between the two, is that the Sharpe ratio measures the performance compared to the risk-free market, whereas the Treynor ratio measures the performance compared to the market return. The Treynor ratio is also limited, as it is only applicable when measuring portfolio performance and not individual assets, whereas the Sharpe ratio is applicable for both. Another drawdown is that the Treynor ratio is a backward-looking method, where the accuracy of the beta estimate is highly dependent on the correct use of market benchmark. Therefore, the ratio should not be used as a standalone measure.

3.4.3 Jensen's Alpha

Jensen's alpha (henceforth alpha) is a risk-adjusted performance measure published by Michael Jensen (1968), and it constitutes a fundamental part of this thesis. Alpha compares the achieved portfolio return with the predicted return measured by any given performance benchmark model. In other words, the measure states whether you have been 'correctly' compensated for the risk you have endured by holding the risky portfolio. The key feature of performance benchmark models is that they differentiate between two types of risk: 1) systematic risk and 2) non-systematic risk. The former is rewarded in terms of return, where an increase in systematic risk results in higher expected returns. The latter, on the other hand, is not compensated as it is assumed to be diversifiable and therefore has an expected return of zero. However, actual returns are rarely equal to expected returns and performance benchmarks are not perfect, which is where alpha becomes relevant. Essentially, alpha is the intercept in the model. According to CAPM, alpha can be measured as follows:

$$\alpha = R_p - \left(r_f + \beta_p * [R_m - r_f]\right)$$

Where:

r_f is the risk-free rate

R_p is the realized return of the portfolio

β_p is the portfolio beta

R_m is the realized return of the market portfolio

A positive alpha will conclude that the portfolio outperformed the market and generated higher returns than what was predicted by CAPM. Opposite, a negative value indicates that the portfolio has not generated the required return. If the securities in the portfolio are fairly priced, the portfolio return will be the same value as predicted by CAPM which will yield an alpha of zero. In this situation, the market is said to be efficient. The alpha measure can also be applied for multi-factor models where the equation above will be adjusted accordingly. Depending on which model is trusted to provide the best guess on the expected return, the alpha can be used to test whether the investor has been able to beat the market, by generating a positive alpha.

For the purpose of this thesis, alpha is used to assess the performance of the constructed portfolios according to the applied performance benchmark models, that is CAPM, Fama-French 3-, and 5-Factor models. More specifically, it is assessed whether alpha is significantly different looking across the high and low ESG-scoring portfolios.

Chapter 4
Data and Methodology

The purpose of the following chapter is to provide a clear depiction of the data and methodology that constitute the foundation of the subsequent analysis. The first section introduces the data that is used to conduct the analysis, which includes ESG scores and other relevant market data. The following section presents the methodology applied in this study to answer the formulated research question and subsequent hypotheses. The methodology is split into two sub sections, portfolio creation and portfolio testing. The last section includes an introduction of some important econometric considerations that needs to be accounted for when performing this type of quantitative study.

4.1 Data

The first part of the data section presents a comprehensive description of the ESG scores applied in this study. This includes a critical reflection of its source provider and the methodology underlying the scores. Furthermore, this section will also introduce other relevant market data that is needed to conduct the intended analysis. This data includes stock return, risk-free rate, factor model returns, and industries.

4.1.1 ESG Scores

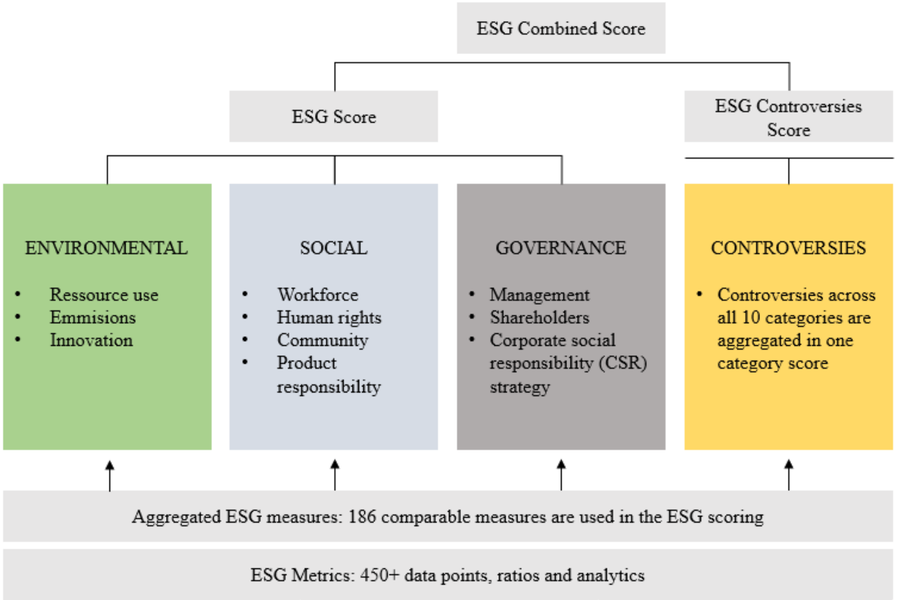
The ESG scores are the fundament of this research analysis, as they are needed to construct and test the performance of high- and low-scoring portfolios. As mentioned in the delimitation section, the ESG scores will not be manually constructed as it would be highly time consuming, require a significant amount of datapoints, and not necessarily provide any added value. Furthermore, the purpose of this thesis is not to construct the scores, but to use them in the construction of portfolios. Therefore, it has been decided to extract the scores from a well-known databank.

Among the myriad of ESG providers are Sustainalytics, Bloomberg and Refinitiv. These data providers are some of the most popular sources in relation to ESG scores (Huber & Comstock, 2017). Furthermore, the datasets are accessible for CBS students and hereby constitute the selection universe. Sustainalytics is a global leader in ESG, corporate governance research, and ratings, and would hereby seem to be the obvious choice. However, their dataset comprises raw scores for indicator level data on policies, programs, initiatives, and disclosures, and not comprehended pillar scores (Sustainalytics, n.d.). They also provide an overall ESG rating, but it has a limited track record. Bloomberg is a global financial market database and has more than 10 years of historic ESG scores (Bloomberg, n.d.). However, their current offerings of ESG scores are highly based on what the companies choose to disclose. Refinitiv is one of the most comprehensive ESG-data providers in the market with data tracking back to 2002 (Refinitiv, n.d.-b). Furthermore, the databank provides the most relevant information for this study, both in regard to investment universe and ESG scores (Refinitiv, 2020). Most importantly, Refinitiv offer overall pillar scores for Environmental (henceforth E score), Social (henceforth S score), and Governance (henceforth G score), as well as two aggregated ESG scores. Therefore, Refinitiv has been chosen as the relevant databank.

Refinitiv's ESG scores are based on more than 450 different ESG metrics gathered from various sources, such as annual reports, company websites, NGO websites, CSR reports, and news. The databank reports ESG scores on more than 10,000 companies worldwide, where approximately 2,100 are located in Europe (Refinitiv, 2020). The scores are based on the companies' relative performance compared to their respective sector and country of incorporation, making the methodology relatively unbiased. The model comprises two overall ESG scores: 'ESG' and 'ESG Combined'. The ESG score encompasses a subset of 186 metrics, based on considerations around comparability, impact, data availability, and industry relevance. The measures are then grouped into 10 categories and reformulate the three pillars (see figure 4.1). The ESG pillar scores are constructed based on the relative sum of the underlying categories, which can vary according to the specific industry. This indicates that more weight is assigned to categories with higher level of transparency. Finally, the ESG score is constructed as the weighted sum of the E, S, and G scores, also according to transparency. All the scores have been normalized to percentages in the range from 0 (ESG laggards) to 100 (ESG leaders), which makes them easily comprehendible.

The ESG Combined score (henceforth ESGC score) is a comprehensive scoring of a company's ESG performance. The score contains both the ESG score, introduced above, and the ESG Controversies score (see figure 4.1). While the ESG score metrics are based on reporting information, the ESGC score also account for any negative media coverage. The ESG Controversies score is based on 23 controversy topics, where a company will be penalised if they are involved in any scandal related to these topics (Refinitiv, 2020). Being involved in such scandals will result in a reduced ESGC score in the latest completed period, and possibly in subsequent years depending on the severity. On the other hand, if the company does not have any controversies their ESG score and ESGC score will be the same.

Figure 4.1: Refinitiv's Scoring Methodology



Source: Modified version of Refinitiv (2020)

Refinitiv updates the ESG-data on a continuous basis which typically follow the corporate reporting patterns. The updates may include new companies, the latest fiscal year update, or inclusion of new controversy events (Refinitiv, 2020). Typically, the scores are updated once a year, whereas it is considered appropriate to also rebalance the constructed portfolios on an annual basis. Furthermore, the scoring metrics are typically disclosed in the companies' annual reports the following year. Therefore, all scores will be displaced one year forward in order to fit the actual announcement of data. Consequently, the financial performance one year is compared to ESG performance the previous year. For simplicity, the scores are assumed to be announced in the beginning of the year, i.e. 1st January, although it varies when a firm publicise their annual reports. This also indicates that the portfolios will be rebalanced in the beginning of each year.

As addressed in section 2.1.4, the scoring methodologies applied by various ESG-score providers vary significantly which lead to general inconsistency and non-comparability in data. This also indicates that the choice of data provider will have an impact on the results obtained in this thesis. The meta-analysis (Revelli & Viviani, 2015) introduced in section 2.4.3 highlights this data inconsistency as being one of the main drivers of heterogeneous results. Therefore, it is important to be critical towards the data provider. Refinitiv's scores are based on publicly available information, exclusively (Refinitiv, 2020). This method can generate some bias as the companies being most transparent also get higher scores. However, their method is highly data driven and makes an effort to assess the companies in the most transparent, consistent, comparable, and objective way possible (Ibid). Finally, Refinitiv is a London Stock Exchange Group (LSEG) business and one of the world's largest providers of financial market data, with more than 400,000 end users in 190 countries (Refinitiv, n.d.-

b). A global platform in this calibre entails a significant amount of expertise and responsibility, that raises their credibility. There will always be limitations when only using a single data provider, but the existing inconsistency in the market makes it impossible to compare results across different data providers. This also entails that the constructed portfolios, and hereby the concluding results reported in this thesis, is highly dependent on the quality and reliability of Refinitiv's ESG scores.

4.1.2 Market Data

Besides the ESG scores, other market data is also required to perform the empirical analysis. As explained in the delimitations section, the asset universe consists of companies from the STOXX Europe 600 Index. Furthermore, to obtain consistency across the time series regressions, only listed companies having reported scores in the entire period (January 1st, 2011 to December 31st, 2020) will be included. In some cases, small adjustments have been made to obtain more sample observations. These adjustments involve a continuation of the scores reported in 2019 for the few companies not having reported 2020 scores at the time of data extraction (January 29th, 2021). The result of the above-mentioned adjustments is an asset universe consisting of 428 companies in the entire period of analysis.

The fundamental element of financial performance research is the dependent variable, i.e. the portfolio return. This implies that the historic returns for all companies within the defined asset universe should be obtained. Refinitiv defines a Total Return Index (RI) (Refinitiv, n.d.-a) which is applied as it accumulates the total growth in capital value. The returns have been extracted on a monthly basis for the entire period of analysis. According to Refinitiv, RI is calculated as follows (Ibid):

$$RI_t = RI_{t-1} * \frac{PI_t}{PI_{t-1}} * \left(1 + \frac{DY_t}{100} * \frac{1}{N}\right)$$

Where:

RI_t and RI_{t-1} is the return index on day t and $t-1$ respectively

PI_t and PI_{t-1} is the price index on day t and $t-1$ respectively

DY_t is the dividend yield

N is the number of working days in the year (260)

As the theory section explains, the returns should be considered net of the risk-free rate. It is common practice to use a short-term interbank offered rate as a proxy for the risk-free rate, because it is assumed to be a good reflection of a risk-free investment (Munk, 2019). An interbank offered rate is a benchmark rate that illustrate the average interest rate at which larger banks lend to one another. In this thesis, the three-month Euro Interbank Offered Rate (Euribor) is used as the risk-free rate due to the geographical focus of the study. The rate has been extracted from Refinitiv (Refinitiv, n.d.-a) on the same dates as the returns. Moreover, the market return is needed which is defined as the return on the applied stock index, namely STOXX Europe 600 Index.

In addition to the market factor, the SMB, HML, RMW, and CMA factors are necessary in order to perform the Fama-French 3- and 5-Factor models introduced in section 3.3. Similar to earlier studies (Blitz & Fabozzi, 2017; Kempf & Osthoff, 2007), this study chooses to obtain the explanatory variables from the Kenneth French Data Library (French, n.d.). As explained in the delimitations section, the factors have not been computed manually as the primary objective of the study is not to test the correctness of the factor models but to apply them for analytical purposes. Kenneth French's dataset holds high credibility, and the factor returns are based on a significant amount of data from prominent sources, which is why the factor-data is considered to be reliable.

Finally, the portfolio analysis also needs to consider if the results are subjects to sector bias. Therefore, the companies' industries have also been extracted from Refinitiv. The industries are classified by a third-party source, being the Dow Jones FTSE's Industry Classification Benchmark, that returns the eleven industry classes listed in table 4.1, alongside their distribution. Based on the classifications, it is possible to test whether the performance is affected when you 'sector neutralize' the portfolio, i.e. apply the same industry weights as the total index.

Table 4.1: Index Industry Weights

STOXX Europe 600 Index	
CIGS Industry	2020
Basic Materials	7%
Consumer Discretionary	14%
Consumer Staples	11%
Energy	6%
Financials	17%
Health Care	14%
Industrials	15%
Real Estate	2%
Technology	6%
Telecommunications	4%
Utilities	5%
Total	100%

Source: Own production

4.2 Methodology

In the following section the methodology, on which the analysis is based on, is presented. The section holds a detailed description of how the high and low-scoring portfolios have been constructed, using various techniques and screening approaches. Furthermore, it describes how the constructed portfolios are being tested and compared using various performance measures.

4.2.1 Portfolio Construction

The fundamental part of the analysis is to study the relationship between ESG scores and financial performance, which is tested through the portfolio approach. Companies are allocated to six different portfolios, namely A to F, based on their ESGC, ESG, E, S, and G scores, respectively. This results in a total of 30 portfolios being tested. Portfolio A will contain the ESG leaders, whereas portfolio F will contain the ESG laggards. By performing this division, it becomes possible to test whether higher-scoring companies perform significantly different than lower-scoring companies. This inclusion approach also portraits the positive screening strategy presented in section 2.2.2. The companies are assigned to the portfolios according to the following rules:

$$A = ESG_t \geq Percentile\left(\frac{5}{6}\right)$$

$$B = A > ESG_t \geq Percentile\left(\frac{4}{6}\right)$$

$$C = B > ESG_t \geq Percentile\left(\frac{3}{6}\right)$$

$$D = C > ESG_t \geq Percentile\left(\frac{2}{6}\right)$$

$$E = D > ESG_t \geq Percentile\left(\frac{1}{6}\right)$$

$$F = E > ESG_t \geq 0$$

The rules stated above indicate that companies will be ranked relatively to the other companies. The percentile rules have been chosen, as they ensure that the number of companies will remain approximately the same within each portfolio each year. This also entails that one portfolio will not benefit from an added diversification effect. Table 4.2 illustrates the number of firms that have been assigned to each portfolio, based on the ESGC score, for the entire period of analysis.

Table 4.2: Number of Firms in the Portfolios

	Number of firms in the portfolios					
Year	A	B	C	D	E	F
2011	72	71	71	71	71	72
2012	73	70	71	71	71	72
2013	72	71	71	71	71	72
2014	72	72	70	71	71	72
2015	72	71	71	71	71	72
2016	73	70	71	71	71	72
2017	72	71	71	71	71	72
2018	72	71	71	71	71	72
2019	72	71	71	71	71	72
2020	72	72	70	71	71	72

Source: Own production

After allocating the firms to their respective portfolios throughout the period of analysis, the firms' weights should be assigned in order to calculate the portfolio return and risk. There are two types of weighting schemes in portfolio theory: 1) equally weighted, where each stock is given the same weight or importance, and 2) value-weighted, where the stock's weight depends on their market value. In this thesis, the portfolio return has been calculated using the value-weighted sum of returns. The primary reason for this, is that the factors extracted from the Kenneth French Data Library (French, n.d.) has been constructed based on value-weighted returns. Thus, the weights have been calculated based on the companies' market value at the time of rebalancing, i.e. the beginning of each year. The market value, as defined by Refinitiv, is the share price multiplied by the number of ordinary shares in issue. Hereby, the equation for the weights is as follows:

$$w_i = \frac{MV_i}{MV_p}$$

Where:

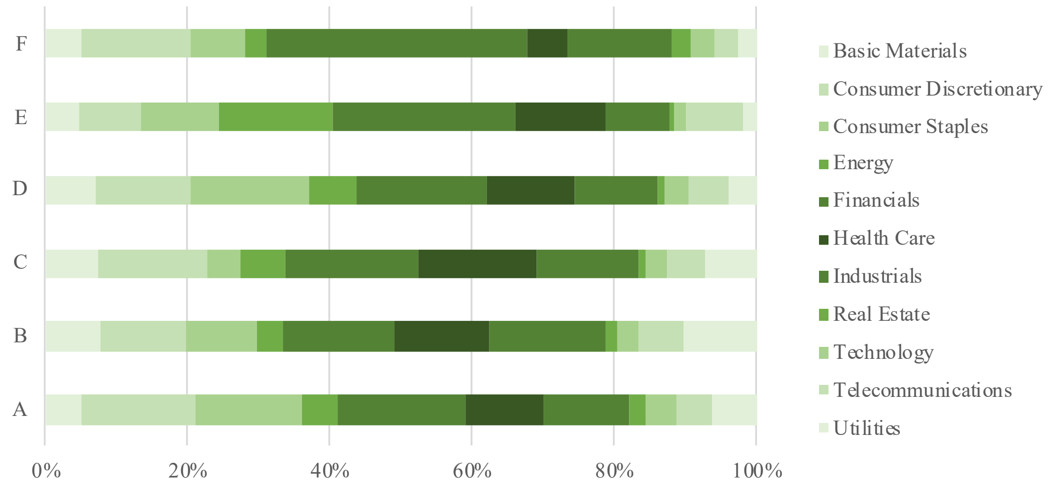
MV_i is the market value for stock i

MV_p is the total market value for the portfolio

In order to detect possible sector displacement within the constructed portfolios, the value-weighted industry weights have been studied. As illustrated in figure 4.2, all industries are represented in the ESGC score portfolios which indicate good diversification, but the composition varies significantly (see remaining scores in Appendix I). Furthermore, it is also observed that the industry weights within the portfolios vary significantly across time (see Appendix II). These findings highlight the importance to construct and test portfolios that are sector-balanced, i.e. portfolios having the same industry weights as the overall index. This approach can be classified as the "best-in-class" strategy, as companies are being ranked according to their scoring among their peers. Thus, companies not scoring in the top of the index can now be assigned to the higher-scoring portfolios, if they have a good ESG score compared to their industry-peers. For each industry, companies have been allocated to the portfolios A to F according to the same rules as previously stated.

Hereafter, the individual stock has been value-weighted according to its industry and constrained by the industry's weight in the total index. The constraints are determined by the industry weights for the applied index in 2020, as illustrated in table 4.1. The results hereof are sector-balanced portfolios where all portfolios have the same industry weights across time and score. Similar to the overall scoring study, the sector-balanced approach provides 30 additional testing portfolios.

Figure 4.2: Average Industry Weights for the ESGC Score



Source: Refinitiv (n.d.-a), Own production

4.2.2 Portfolio Performance

In order to qualify the acceptance or rejection of the constructed hypotheses, the performance of the high- and low-scoring portfolios will be tested and compared. Portfolio performance can be measured in many ways. As introduced in Chapter 3, one can look at the simple descriptive statistics of the portfolio's return and risk properties. For the preliminary examination of the portfolios' performance, the cumulative returns and annualised average monthly returns have been calculated. However, the return itself does not say much without the risk element, which is examined through the standard deviation, market beta, skewness, kurtosis, and maximum drawdown. The calculation of the standard deviation involves the composition of the covariance-matrix, which is a 428x428 matrix that contains the covariances between all the stocks included in the analysis (see subset in Appendix III). Hereafter, the portfolio standard deviation is calculated by using the following vector equation (Munk, 2019):

$$\sigma_p = \sqrt{\pi^T \sum \pi}$$

Where:

π^T is the transposed weight vector

$\sum$ is the 428x428 variance-covariance matrix

π is the weight vector

The market beta, on the other hand, is found by performing linear regressions between the portfolio and market returns. For further interpretation of the portfolios' return distribution and downside risk, the skewness, kurtosis, and maximum drawdown are also studied.

The aforementioned performance characteristics are the standalone risk and return measures. However, as foretold in the theory chapter it is important to study the portfolios' risk-adjusted performance. The risk-adjusted measures used are the ones introduced in the theory chapter: Sharpe ratio, Treynor ratio, and Jensen's alpha. The first two can easily be computed from the above estimates, whereas the alpha requires further analysis. Alpha is examined through the factor models: CAPM, Fama-French 3-, and 5-Factor models, using the statistical programming software Stata (see code in Appendix IV). All the models have two common elements: 1) the independent variable, that is the monthly portfolio return minus the risk-free rate, and 2) the market factor, that is the monthly market return minus the risk-free rate. Moreover, the Fama-French models add the SMB, HML, RMW, and CMA factors. For each portfolio, the alpha is reported and tested for significance. If the test returns a significant alpha, it will be indicated by a star (*) and the number of stars will be determined by the significance level, i.e. the probability of incorrectly rejecting the null hypothesis. The significance is reported on three levels, 5% (*), 1% (**), and 0.1% (***), indicating a confidence interval of 95%, 99%, and 99.9%, respectively. Furthermore, the adjusted R-squared is presented for all performed models. The adjusted R-squared is an unbiased form of the simple R-squared estimator that measures the proportion of the variation in the dependent variable explained by the independent variables. The interpretation of the adjusted R-squared is hereby a bit different as it does not directly signal the proportion of explained variation, but it still gives an indication of the models' explanatory power. Furthermore, the value is more comparable and reliable across different tests.

In order to study the robustness of the results, multiple variations of the tests are performed. For this part, it is only found necessary to report the results for the A and F portfolios, as these are considered most important in relation to the formulated hypotheses. First, the data is split into two sub samples, one containing the early five-year period (2011-2015) and another containing the late five-year period (2016-2020). This division enables the testing of a potential change in the ESG-investing interest. Furthermore, the constructed industry-weighted portfolios are being tested to see whether the obtained results from the value-weighted portfolio analysis is caused by a sector-displacement rather than ESG performance. Lastly, the models will also be performed on a dataset excluding outliers, and the reason hereof is presented in section 4.3.4.

4.3 Econometric Considerations

From the theoretical section of this thesis, it is known that several assumptions must be satisfied for the analytical models to provide unbiased and consistent estimators. It is hereby found critical to test how the data obtained fulfils these assumptions. In the following sections, some of the most critical econometric considerations will be presented, these being: autocorrelation, heteroscedasticity, multicollinearity, outliers, sample selection bias, and errors-in-variables.

4.3.1 Autocorrelation

Autocorrelation, also known as "serial correlation", is a mathematical representation of the degree of similarity in a given series with its lagged version over successive time intervals. If a pattern is observed in the series, such that values can be predicted based on preceding values, the series is said to exhibit autocorrelation. Thus, the autocorrelation test is used to detect non-randomness. The problem arising from autocorrelation is that OLS does not account for the variance between the correlated error terms which causes the computed standard errors and p-values to be misleading (Stock & Watson, 2015). Autocorrelation is most commonly found in time series data which is applied in this thesis. It is hereby necessary to test whether the models exhibit autocorrelation.

A common method to identify autocorrelation is by running the Breusch-Godfrey (1978) test on the regressions performed. The test measures the dependence of the residuals, i.e. the error terms in the regression models, on their lagged values. The standard approach includes the test of the residual's dependence on its one-lagged value, but the test also allows for testing higher order serial correlation. The one-lag test takes the following form:

$$u_t = \phi_1 u_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim IID(0, \sigma^2); t = 1, \dots, n$$

Where:

u is the error terms obtained in the first regression for period t and $t-1$

ϕ is the estimated coefficients in the second regression

The null hypothesis in the above test is stating "no autocorrelation", whereas the alternative hypothesis indicate that the model does exhibit autocorrelation. If the estimated coefficients are significantly different from zero, we reject the null hypothesis indicating that autocorrelation exist. The test is performed by running the relevant regressions and hereafter applying the 'bgodfrey' command in Stata with lags set to one (see Appendix IV).

4.3.2 Heteroscedasticity

Heteroscedasticity refers to the variance of the error terms and happens when the variance is non-constant across different values of an independent variable. Opposite, if the variance of the error terms is constant, they are said to be homoscedastic which is an important assumption for linear regression modelling. Upon visual inspection of the observations, heteroscedasticity will appear when the residual errors tend to fan out over time instead of remaining within the same range. The effect of heteroscedasticity is, similar to autocorrelation, that the variance will be biased and hereby generate misleading standard errors that will impact the validity of the econometric analysis (Stock & Watson, 2015). The Breusch-Pagan (1979) test is a common method used to test for heteroscedasticity. The test is based on the assumption that the error terms are normally distributed, i.e. with a mean of zero and a constant standard deviation. Essentially, the following equation is testing if the coefficients are equal to zero:

$$\hat{u}_i^2 = \delta_0 + \delta_1 x_{i1} + \cdots + \delta_p x_{ip} + \varepsilon_i$$

Where:

$\hat{u}_i^2$ is the predicted squared residuals from the first regression

x_i is the independent variables from the first regression

δ is the estimated coefficients from the second regression

From the above stated regression, the R-squared values are retained to obtain the Chi-Square test statistic and corresponding p-values. The test is performed by first running the relevant regression models and hereafter using the 'hettest' command in Stata (see Appendix IV). The null hypothesis for the test is that the variances of the error terms are equal, whereas the alternative hypothesis is that the variances are not equal. Hence, if the p-values prove significant the models are said to exhibit heteroscedasticity. However, it is possible to overcome the issue by applying more robust standard errors through the 'newey' regression command in Stata (see code in Appendix IV).

4.3.3 Multicollinearity

Another important assumption is that no perfect multicollinearity should exist which is relevant when running multi-factor regression models. As explained in the OLS section, multicollinearity occurs when two or more independent variables are perfectly correlated. This will impose an issue towards the explanatory power of the OLS, as the model will have difficulties in determining which of the independent variables that impacts the dependent variable. The results of having multicollinearity in the model are unstable estimated coefficients and possible widely inflated standard errors (Stock & Watson, 2015). A common method to test for multicollinearity is by using the VIF-test that stands for 'variance inflation factor' (Stock & Watson, 2015). As the name suggests, the VIF test quantifies how much the standard errors, i.e. the variances of the estimated coefficients, are inflated. Essentially, the test measures the ratio between the overall variance of the model and the variance of a model including a single explanatory variable. This is repeated for every explanatory variable in the model and calculated by the following:

$$VIF_i = \frac{1}{1 + R_i^2}$$

Where:

VIF_i is the variance inflation factor for explanatory variable i

R_i^2 is R-square from the

The rule of thumb states that VIF factors above the nominal value of 10 indicate multicollinearity (Stock & Watson, 2015). The VIF-test is performed by first running the relevant regression models and hereafter using the 'vif' command in Stata (see Appendix IV). The tested regression models are the Fama-French 3- and 5-Factor models, as CAPM is a single-factor model and hereby not relevant for testing.

4.3.4 Outliers

It is important to be aware of the existence of outliers in the dataset, as they may have a significant impact on the OLS estimates as explained in section 3.3.4. An explicit definition of an outlier does not exist. Instead, determining an outlier requires a comparison of the observations in the sample, where an outlier will be an observation located far away from the mass (Stock & Watson, 2015). To identify possible outliers in the dataset, the companies' cumulative returns have been plotted against their respective average ESGC scores. If the dataset proves to contain outliers, it is important to test whether the estimated coefficients change significantly if these were removed. This is done by first running the regression analysis on the dataset including outliers and hereafter on the dataset excluding outliers.

4.3.5 Sample Selection Bias

Sample selection bias occurs when a statistical analysis is conducted based on a non-random sample and hereby does not truly represent the population. In practice, selection bias is represented by a correlation between one or several regressors and the error term. The presence of selection bias may distort the statistical analysis and generate biased OLS estimates (Stock & Watson, 2015). Survivorship bias is a common selection bias in financial research, particularly amongst historical performance studies. Survivorship bias is referring to the researcher's biased selection of stocks that are currently available in the end of their research period. In other words, the sample will only consist of 'surviving' stocks and hereby omit those that have been delisted due to liquidation or mergers. In essence, the poorest-performing companies are being excluded from the dataset which may result in a misleading picture of the historical performance. For this reason, a dataset that suffers survivorship bias tends to overstate historical returns.

The sample used for this particular research, consist solely of companies that were listed at the date of data extraction and have been for the entire period of analysis. Moreover, the asset universe is only representing the largest companies in Europe, as they have been selected from the STOXX Europe 600 Index. Therefore, it can be assumed that the dataset is subject to survivorship bias. A solution towards this problem, could be to include all stocks that are listed today and those that have been delisted or merged in the relevant period. However, this solution is not considered appropriate as it would entail the inclusion of smaller companies in the dataset which could potentially generate large disturbances. Furthermore, the issue of survivorship bias is not deemed critical. The purpose of this thesis is to uncover significant return differences between high and low ESG-scoring companies. As the constructed portfolios are all subject to survivorship bias it will not affect the conclusions being drawn. The reported return and alpha estimates may be slightly overstated, but the interpretation of the compared portfolios will remain the same.

4.3.6 Errors-in-Variables

The errors-in-variables (EIV) problem arises when using incorrectly measured variables in regression models. The econometric theory distinguishes between errors in the dependent and independent variables, whereas the latter is considered more critical. The reasoning behind this is that errors in the dependent variable are incorporated in the disturbance term, thus, they do not generate any biased coefficients or harm the regression model (Maddala & Nimalendran, 1996). On the other hand, errors in an independent variable will result in correlation between the regressor and the error term which will lead to biased OLS estimators and inconsistent standard errors.

Errors in the independent variables can be due to several reasons, but two common reasons for errors are typographical errors and measurement errors. In the performed OLS regressions, the independent variables are the risk factor premiums identified in each applied performance benchmark model, that is market, SMB, HML, RMW, and CMA. Aforementioned, the market factor is simply the total return on the index that constitute the asset universe of the study, which is extracted from Refinitiv (n.d.-a). Whereas the remaining factors have been extracted from the Kenneth French Data Library (n.d.). Thus, the applied independent variables are all of secondary character. The data sources could hold both typographical and measurement errors, but it is assumed to be of lower degree. This is based on the fact that both data sources provide highly reliable data that is based on a significant amount of expertise, as mentioned in section 4.1.

Chapter 5

Analysis and Results

The following chapter presents the findings of the conducted analysis. The first section presents the results from the econometric tests, namely autocorrelation, heteroscedasticity and multicollinearity. Furthermore, the sample observations are plotted to identify possible outliers. Hereafter, the results from the value-weighted portfolio analysis are presented, that is the constructed portfolios based on the positive screening approach. The purpose is to study the overall characteristics and test the performance of the high and low ESG portfolios. After follows various robustness tests including sub samples, industry-weighted portfolios, and outliers. The industry-weighted portfolios refer to the constructed portfolios based on the "best-in-class" screening approach which is included to detect a possible sector bias in the results.

5.1 Econometric Considerations Results

As described in section 4.2, some important regression assumptions need to be fulfilled for the models to produce unbiased and consistent estimators. This refers to the previously described econometric considerations, namely autocorrelation, heteroscedasticity, multicollinearity, and outliers. Therefore, the following section will briefly present the results from the performed Breusch-Godfrey, Breusch-Pagan, and VIF tests, as well as a graphic depiction of the dataset to identify possible outliers. The testing results are only reported for the A and F portfolios as these are considered most important for the formulated hypotheses.

Table 5.1 presents the significance of the coefficients from the autocorrelation test across both scores, models, and portfolios. From the results, it can be seen that none of the tested models show significant p-values which indicates that no autocorrelation has been identified. Table 5.1 presents the results from the heteroscedasticity tests for all performed models. It is observed that the A portfolios, based the ESGC score and G score, show heteroscedasticity across all tested models. As explained in section 4.3.2, it is possible to control for heteroscedasticity by applying more robust error terms. Therefore, the A portfolios have been tested using the 'newey' command which resulted in the same outcomes as the standard OLS regressions. Furthermore, the multicollinearity test is performed. The results show that the Fama-French 3-Factor model has a VIF of 1.13, while the Fama-French 5-Factor model has a VIF of 2.59. As described in section 4.3.3, the rule of thumb states that a VIF factor above 10 indicates multicollinearity. Thus, the applied models do not exhibit multicollinearity.

Table 5.1: Autocorrelation

Score	Model	Portfolio	P-value
ESG comb	CAPM	A	0.407
		F	0.381
	FF3	A	0.527
		F	0.909
	FF5	A	0.428
		F	0.837
ESG	CAPM	A	0.092
		F	0.982
	FF3	A	0.690
		F	0.427
	FF5	A	0.698
		F	0.952
E	CAPM	A	0.268
		F	0.275
	FF3	A	0.787
		F	0.366
	FF5	A	0.726
		F	0.989
S	CAPM	A	0.191
		F	0.590
	FF3	A	0.727
		F	0.885
	FF5	A	0.667
		F	0.630
G	CAPM	A	0.577
		F	0.694
	FF3	A	0.132
		F	0.822
	FF5	A	0.116
		F	0.617

H0: No serial correlation

* p<0.05, ** p<0.01, *** p<0.001

Source: Own production

Table 5.2: Heteroscedasticity

Score	Model	Portfolio	P-value
ESG comb	CAPM	A	0.045*
		F	0.101
	FF3	A	0.010*
		F	0.404
	FF5	A	0.006**
		F	0.520
ESG	CAPM	A	0.050
		F	0.284
	FF3	A	0.808
		F	0.338
	FF5	A	0.618
		F	0.273
E	CAPM	A	0.003**
		F	0.916
	FF3	A	0.028*
		F	0.136
	FF5	A	0.627
		F	0.110
S	CAPM	A	0.097
		F	0.025*
	FF3	A	0.397
		F	0.217
	FF5	A	0.851
		F	0.272
G	CAPM	A	0.000***
		F	0.916
	FF3	A	0.000***
		F	0.668
	FF5	A	0.001***
		F	0.748

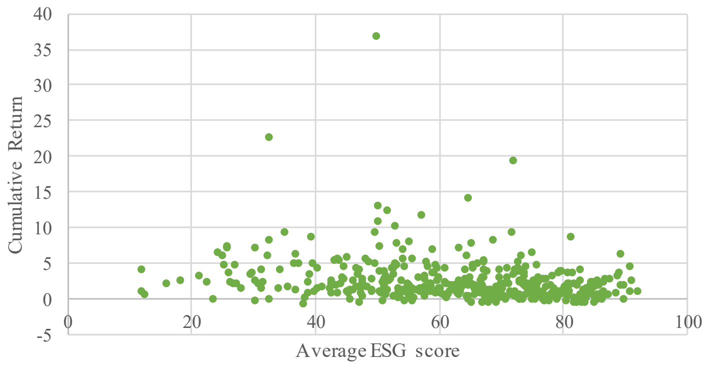
H0: Constant variance

* p<0.05, ** p<0.01, *** p<0.001

Source: Own production

Finally, it is examined whether the dataset contains outliers. To identify possible outliers in the dataset, the companies' cumulative returns have been plotted against their respective average ESGC scores, as explained in section 4.3.4. From the plot (see figure 5.1), it is observed that some companies are located far from the mass. More specifically, three companies have shown a cumulative return above 1500% within the period of analysis. The identified outliers are Neste, Ashtead Group and GenMab. Due to the existence of outliers within the dataset, it is important to report the regression results from both including and excluding outliers in the applied dataset, as explained in section 4.3.4. The results from the exclusion of outliers will be presented later in the analysis to compare them with the results obtained from the complete dataset.

Figure 5.1: Outliers in the dataset



Source: Own production

5.2 Value-Weighted Portfolios

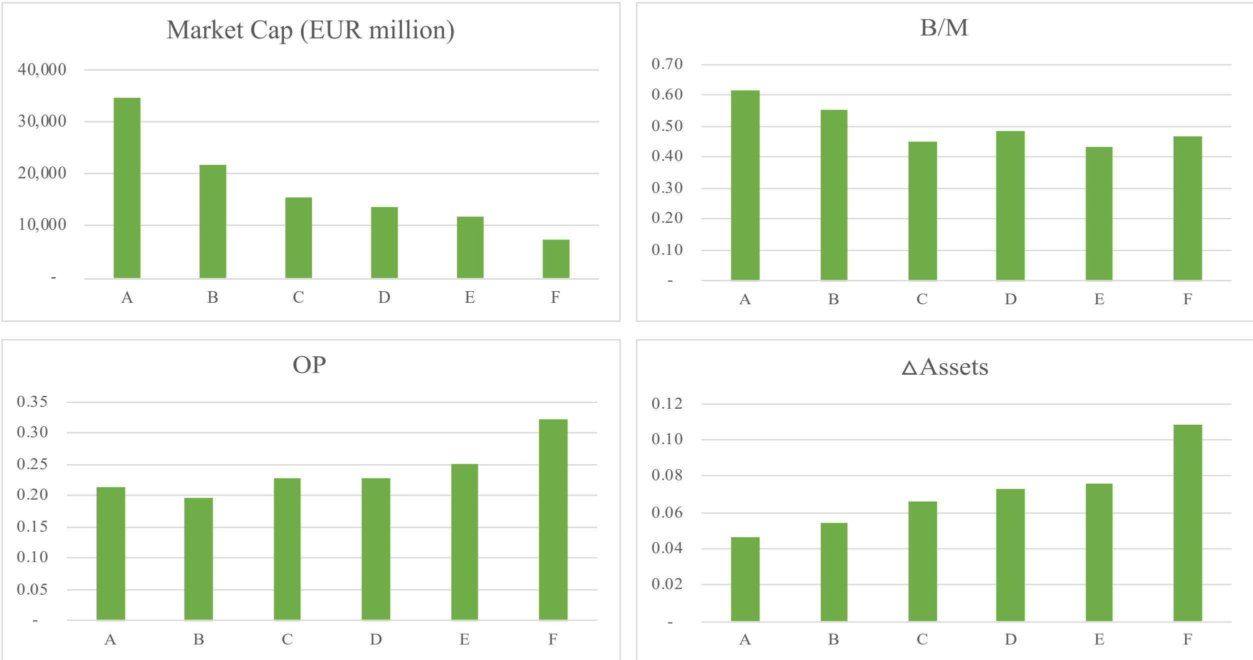
In this part of the analysis, the aim is to uncover whether an investment strategy involving ESG scores has an impact on financial performance. Firstly, the portfolios' financial characteristics will be studied to see if the constructed high- and low-scoring portfolios exhibit significant characteristic differences. Secondly, the portfolios' risk and return properties will be studied to find a preliminary indication of the portfolios' performances. Thirdly, and most importantly, the portfolios will be tested using the performance benchmark models, that is the CAPM, Fama-French 3-, and 5-Factor models.

5.2.1 Financial Performance and Characteristics

The financial performance parameters, illustrated in figure 5.2, are relevant to study as they are proxies for the risk premium factors included in the Fama-French testing models. The risk premium factors are describing anomalies found in the market related to size, value, profitability, and investments. The estimates are based on average firm values across both the period of analysis and the scores, these being ESGC, ESG, E, S, and G, respectively (see details in Appendix V).

The histograms illustrate some clear tendencies in the company characteristics within the portfolios. The firms included in portfolio A generally has a much higher market cap, higher B/M ratio, lower operating profitability, and lower change in assets than the firms included in portfolio F. This indicates that the firms with higher ESG scores are characterized as bigger value companies with low profitability and a conservative investment approach. Whereas the firms with lower scores are characterized as smaller growth companies with higher profitability and an aggressive investment approach.

Figure 5.2: Portfolio Composition by Financial Characteristics



Source: Own production

5.2.2 Descriptive Statistics and Performance Measures

Table 5.3 presents the return and risk properties across the six constructed portfolios for each studied score. The cumulative return is the aggregated return across the entire 10-year period of analysis, January 1st, 2011 to December 31st, 2020. In other words, it represents the total change in the investment price. The annual return is the annualised geometric mean of monthly returns less the risk-free rate, calculated across the entire period of analysis. Similarly, the standard deviation is also the annualised standard deviation of monthly returns.

From the table, it can be seen that the higher-scoring portfolios generally show much lower returns than the lower-scoring portfolios, both looking at the cumulative and annual return columns. This is true for every score, except from the ESGC score. Here, the returns are almost equal for portfolio A and F, with a cumulative return around 136% and an annual return around 8%. Furthermore, the ESGC score illustrates a decreasing return tendency from portfolio B to E. The risk, according to the standard deviation estimate, is lower in portfolio A compared to portfolio F for the two overall ESG scores and the S score. According to the beta risk measure, the ESGC score and S score give an indication of lower risk in the top portfolios. However, the important measure is the risk-adjusted performance which can be studied through the Sharpe and Treynor ratios. The two ratios provide similar results – the higher-scoring portfolio, based on the ESGC score, shows better risk-adjusted performance than the lower-scoring portfolios. In contrary, the remaining scores show opposite results, and their F portfolios even provide higher risk-adjusted performances than the A portfolio in the ESGC score. Another important observation is the return distributions which are examined through the skewness and kurtosis measures. The F portfolios generally exhibit higher negative skewness and higher kurtosis than the A portfolios. This indicates that the lower-scoring portfolios experience more extreme negative results and have fatter tails. The portfolios' downside risk can be studied further through the maximum drawdown estimates. These results show that the F portfolios have generally experienced higher maximum drawdowns, i.e. greater losses, than the A portfolios. This is the case for all scores expect from the G score, and hereby supports the general conclusion drawn from the skewness estimates.

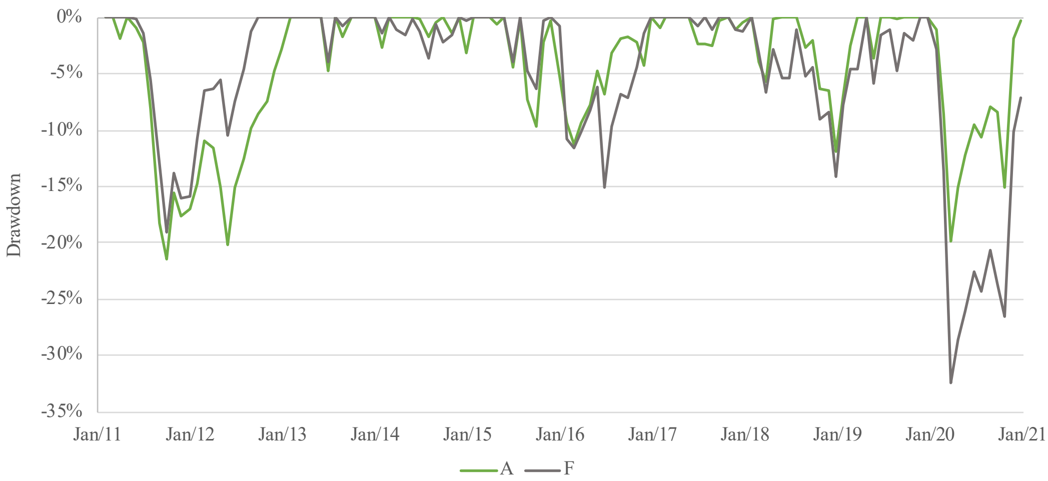
Table 5.3: Descriptive Statistics and Performance Measures

Score	*Portfolio*	*Cumulative Return*	*Annual Return*	*Standard Deviation*	*Beta*	*Sharpe Ratio*	*Treynor Ratio*	*Skewness*	*Kurtosis*	*Maximum Drawdown*
ESG comb	A	136.12%	8.10%	14.39%	0.97	0.56	0.08	-0.26	1.74	-21.39%
	B	157.27%	9.04%	14.45%	0.98	0.63	0.09	-0.21	2.62	-20.53%
	C	117.94%	7.25%	14.20%	0.93	0.51	0.08	-0.30	1.05	-20.63%
	D	108.43%	6.80%	14.36%	0.97	0.47	0.07	-0.19	1.85	-23.06%
	E	77.66%	5.09%	14.58%	0.96	0.35	0.05	0.34	4.49	-24.02%
	F	136.52%	8.15%	15.48%	1.09	0.53	0.07	-0.52	6.99	-32.41%
ESG	A	80.20%	5.23%	14.21%	0.98	0.37	0.05	0.09	2.45	-20.78%
	B	101.57%	6.41%	14.93%	1.04	0.43	0.06	-0.29	3.81	-24.23%
	C	112.26%	6.97%	14.61%	1.02	0.48	0.07	-0.06	4.55	-24.01%
	D	211.92%	11.18%	14.67%	0.99	0.76	0.11	-0.25	2.89	-21.48%
	E	155.33%	8.98%	14.38%	0.92	0.62	0.10	-0.41	2.73	-23.87%
	F	240.63%	12.15%	14.23%	0.93	0.85	0.13	-0.78	3.01	-23.41%
E	A	96.86%	6.15%	15.97%	1.07	0.39	0.06	-0.10	2.51	-22.88%
	B	76.95%	5.05%	13.47%	0.94	0.38	0.05	0.23	4.61	-25.04%
	C	123.65%	7.54%	14.12%	0.95	0.53	0.08	-0.13	2.15	-21.37%
	D	140.47%	8.32%	13.61%	0.94	0.61	0.09	-0.19	2.51	-20.13%
	E	187.89%	10.30%	13.96%	0.96	0.74	0.11	-0.48	3.98	-24.68%
	F	245.65%	12.32%	14.43%	0.94	0.85	0.13	-0.75	2.46	-23.26%
S	A	102.05%	6.45%	13.53%	0.92	0.48	0.07	-0.06	1.61	-18.36%
	B	94.54%	6.05%	15.38%	1.06	0.39	0.06	-0.14	3.64	-24.72%
	C	122.71%	7.49%	15.47%	1.06	0.48	0.07	0.15	4.98	-25.96%
	D	115.65%	7.13%	14.73%	1.02	0.48	0.07	-0.68	4.19	-25.85%
	E	154.36%	8.93%	13.74%	0.95	0.65	0.09	-0.38	3.74	-22.92%
	F	201.02%	10.81%	15.15%	0.97	0.71	0.11	-0.61	5.15	-29.91%
G	A	65.23%	4.32%	14.94%	1.06	0.29	0.04	0.14	5.04	-27.26%
	B	111.19%	6.93%	14.24%	0.98	0.49	0.07	-0.04	2.48	-21.42%
	C	142.84%	8.43%	13.48%	0.97	0.63	0.09	-0.26	1.93	-21.03%
	D	152.31%	8.82%	14.74%	0.97	0.60	0.09	-0.36	2.49	-22.55%
	E	167.84%	9.50%	15.03%	0.96	0.63	0.10	-0.58	2.60	-20.67%
	F	190.98%	10.40%	10.18%	0.95	1.02	0.11	-0.01	1.99	-19.52%

Source: Own Production

As illustrated above, the F portfolios generally show higher maximum drawdowns than the A portfolios. Especially within the ESGC score, a significant difference in drawdowns is observed across portfolio A and F. To further study the downside volatility differences between portfolio A and F, according to the ESGC score, the timeseries drawdown has been plotted. From the graphs presented in figure 5.3, it is observed that the two portfolios generally follow the same downward trends. However, some important differences are also observed. In the beginning of the period (2011-2013), portfolio A generally experiences higher drawdowns and longer time of recovery. The opposite is true after 2015 where portfolio F generally shows higher drawdowns. Especially during the latest COVID-19 crisis, portfolio F experienced much higher drawdowns and have still not recovered from the losses by the end of 2020. This finding is in line with the ESG funds' higher resilience during COVID-19 described in section 2.1.1 and supports the arguments from the SRI supporters (see section 2.3).

Figure 5.3: Drawdown Plot (ESGC Score, Portfolio A and F)



Source: Own Production

The main conclusion drawn from the descriptive statistics and performance measures is that the higher-scoring portfolios based on the ESGC score have a better risk-adjusted performance than the lower-scoring portfolios. For the remaining scores, the F portfolios show higher risk-adjusted performance. However, according to the return distributions a general picture can be observed – lower-scoring portfolios have higher tail-risk than the higher-scoring portfolios. This is supported by the maximum drawdown estimates where the F portfolios experience higher losses than the A portfolios.

5.2.3 Results: CAPM

As the financial theory postulates (see section 3.3), you cannot draw conclusions based on simple descriptive statistics as obtained in the previous section. It is necessary to include risk factors to compare the actual obtained portfolio returns with the expected return according to risk exposures. The first factor-model used for testing is the single-factor CAPM model that assumes the only explanatory factor is the market. The CAPM testing results are listed in table 5.4.

Table 5.4: CAPM

Score	*Portfolio*	*Alpha*	*MKT*	*Adj R^2*
ESG comb	A	0.001	0.979***	0.932
	B	0.002*	0.989***	0.937
	C	0.001	0.938***	0.912
	D	0.000	0.970***	0.924
	E	-0.001	0.966***	0.855
	F	0.001	1.084***	0.831
ESG	A	-0.001	0.978***	0.926
	B	0.000	1.042***	0.902
	C	0.000	1.020***	0.913
	D	0.004**	0.991***	0.914
	E	0.002	0.924***	0.849
	F	0.005**	0.942***	0.851
E	A	-0.001	1.079***	0.917
	B	-0.001	0.945***	0.882
	C	0.001	0.951***	0.914
	D	0.002	0.940***	0.931
	E	0.003*	0.958***	0.879
	F	0.005***	0.951***	0.864
S	A	0.000	0.930***	0.926
	B	-0.001	1.061***	0.906
	C	0.001	1.057***	0.909
	D	0.000	1.033***	0.891
	E	0.002	0.957***	0.907
	F	0.004*	0.957***	0.813
G	A	-0.002	1.064***	0.905
	B	0.001	0.977***	0.904
	C	0.002	0.971***	0.926
	D	0.002	0.987***	0.902
	E	0.003*	0.955***	0.911
	F	0.003*	0.955***	0.864

* $p<0.05$, ** $p<0.01$, *** $p<0.001$

Source: Own production

Looking at the testing results for the ESGC score, portfolio B is the only one returning a significant alpha at the 5% level with a monthly excess return at 0.2%. It is also observed that one of the lower-scoring portfolios, namely portfolio E, returns a negative alpha, although not on a significant level. The market betas are all positive and close to one at a 0.1% significance level, indicating an almost perfect positive correlation with the market. Thus, the market factor has a significant explanatory power in returns. Interesting to notice is that the beta for portfolio A is lower than that of portfolio F, which means that the higher ESG-scoring portfolio exhibit lower market-related risk than the lower ESG-scoring portfolio. Lastly, by looking at the adjusted R-squared values (*Adj R^2*), these are observed to be higher for the higher-scoring portfolios compared to the lower-scoring portfolios. This indicates that the CAPM model has higher explanatory power for the returns in the higher-scoring portfolios which is the general conclusion across all scores.

The results from the ESG score differ from the ESGC as the lower-scoring portfolios, namely D and F, return positive alphas at the 1% significance level with monthly excess returns at 0.4% and 0.5%, respectively. Furthermore, portfolio A returns a negative alpha at 0.1%, however, the value is not statistically significant.

Another difference is that the betas are higher for portfolio A and B, ranging from 0.98 to 1.04, compared to portfolio E and F, ranging from 0.92 to 0.94.

The remaining pillar scores, E, S, and G, show somewhat the same results as the ESG score. They all return positive and significant alphas in the F portfolios, where the E score demonstrate the highest nominal alpha at 0.5% combined with the highest level of significance at 0.1%. Also similar to the ESG score, the A portfolios show negative, but insignificant, alphas in the E and G scores. Lastly, the beta results show higher market-related risk for the higher-scoring portfolios in the E and G scores, while the opposite is true for the S score.

5.2.4 Results: Fama-French Three-Factor Model

The second factor-model used for testing is the Fama-French 3-Factor (FF3) model. In addition to the market factor, the FF3 model includes the size (SMB) and value (HML) factors as relevant explanatory variables. The results show many similarities to CAPM, but they also have some important differences which will be described in the following sections. The testing results appear in table 5.5.

Table 5.5: Fama-French 3-Factor Model

Score	*Portfolio*	*Alpha*	*MKT*	*SMB*	*HML*	*Adj R^2*
ESG comb	A	0.002*	0.944***	-0.186***	0.120**	0.941
	B	0.003**	0.961***	-0.009	0.109**	0.940
	C	0.002	0.910***	-0.074	0.105*	0.915
	D	0.001	0.946***	0.006	0.095*	0.926
	E	0.001	0.890***	-0.122	0.285***	0.883
	F	0.002	1.013***	0.463***	0.320***	0.889
ESG	A	0.001	0.915***	-0.257***	0.224***	0.952
	B	0.002	0.948***	-0.021	0.366***	0.942
	C	0.001	0.972***	0.079	0.192***	0.925
	D	0.004**	0.990***	0.117	0.013	0.915
	E	0.001	0.956***	0.195*	-0.108	0.857
	F	0.003**	0.968***	0.529***	-0.056	0.898
E	A	0.002	0.982***	-0.122*	0.368***	0.957
	B	0.001	0.876***	-0.219**	0.247***	0.910
	C	0.001	0.932***	-0.004	0.071	0.914
	D	0.001	0.958***	0.088	-0.062	0.933
	E	0.003*	0.963***	0.207**	0.000	0.884
	F	0.004**	0.969***	0.514***	-0.025	0.908
S	A	0.001	0.899***	-0.272***	0.096**	0.942
	B	0.002	0.970***	-0.124*	0.343***	0.942
	C	0.002	0.981***	0.152**	0.313***	0.944
	D	0.001	0.981***	0.176*	0.219***	0.910
	E	0.002	0.945***	0.207**	0.061	0.916
	F	0.002	0.987***	0.418***	-0.080	0.840
G	A	0.000	0.970***	-0.087	0.358***	0.943
	B	0.002	0.920***	-0.098	0.214***	0.920
	C	0.002*	0.947***	-0.113	0.085*	0.929
	D	0.003*	0.949***	0.036	0.152**	0.909
	E	0.003**	0.928***	0.044	0.111*	0.915
	F	0.002	0.987***	0.214**	-0.105	0.873

* p<0.05, ** p<0.01, *** p<0.001

Source: Own production

The FF3 model generally increases the nominal value of alpha in the higher-scoring portfolios. The most important finding is observed in the ESGC score. Here, portfolio A and B both return significant alphas at the 5% and 1% levels with monthly excess returns at 0.2% and 0.3%, respectively. For the ESG score, portfolio D and F still show significant alphas at the 1% level. Moreover, the A portfolio alpha has turned positive, although not on a significant level. Similarly, the E score shows significant positive alphas in the lower-scoring portfolios and positive, but insignificant, alphas in the higher-scoring portfolios. The S score does not return any significant alphas for the portfolios which is an improvement from before where portfolio F had a significant positive alpha. Turning to the G score, portfolio F does no longer have a positive alpha, but portfolios C through E all return significant positive alphas.

The market factor is still significant on a 0.1% level in the range 0.89 to 1.01. The interesting observation here is that the higher-scoring portfolios exhibit lower market-related risk for all scores, except from the E score. This is in contrast to the CAPM model where both the ESG, E, and G scores had higher marked-related risk in the higher-scoring portfolios. Besides the market factor, significant values are also observed in both the size and value factors. Both the ESGC and ESG scores return significant SMB coefficients at the 0.1% level in portfolio A and F. Generally, the higher-scoring portfolios exhibit negative exposures to the SMB factor, whereas the lower-scoring portfolios exhibit positive exposures. This implies that the higher ESG-scoring portfolios generally consist of larger companies, whereas the lower ESG-scoring portfolios consist of smaller companies. The increasing trend is also explicitly shown in the S score where all portfolios return significant SMB coefficients. HML also shows highly significant values across all scores. Here, the general observation is that the higher-scoring portfolios return positive values, indicating that these portfolios consist of value-stocks. Opposite, the lower-scoring portfolios have negative values indicating growth stocks. This is true for all scores, except from the ESGC score. The general trends described above also support the findings from the histograms in figure 5.2.

Lastly, the adjusted R-squared values show higher explanatory power for the higher-scoring portfolios compared to the lower-scoring ones, similar to the CAPM model. Furthermore, the general level of adjusted R-squared values is slightly above the values obtained in the CAPM model. Intuitively, this makes sense as both of the newly introduced variables, these being SMB and HML, return significant alphas and hereby add additional explanatory power to the model.

5.2.5 Results: Fama-French Five-Factor Model

The Fama-French 5-Factor (FF5) model is the third and last tested performance benchmark model. In addition to the market, size, and value factors, the FF5 model includes the profitability (RMW) and investment (CMA) factors as relevant explanatory variables. The testing results are included in table 5.6.

Table 5.6: Fama-French 5-Factor Model

Score	*Portfolio*	*Alpha*	*MKT*	*SMB*	*HML*	*RMW*	*CMA*	*Adj R^2*
ESG comb	A	0.002*	0.940***	-0.188**	0.165*	0.075	-0.023	0.940
	B	0.003*	0.964***	-0.002	0.123	0.056	0.035	0.939
	C	0.002	0.893***	-0.109	0.101	-0.148	-0.178	0.917
	D	0.001	0.924***	-0.037	0.111	-0.143	-0.226*	0.929
	E	0.001	0.866***	-0.156	0.399***	0.066	-0.212	0.883
	F	0.002	0.998***	0.446***	0.422***	0.114	-0.120	0.888
ESG	A	0.001	0.909***	-0.264***	0.260***	0.035	-0.050	0.952
	B	0.002*	0.941***	-0.040	0.339***	-0.124	-0.087	0.943
	C	0.001	0.963***	0.076	0.303**	0.184	-0.053	0.926
	D	0.003**	0.954***	0.068	0.190*	0.118	-0.311*	0.918
	E	0.001	0.945***	0.195*	0.050	0.287	-0.046	0.859
	F	0.003**	0.938***	0.483***	0.062	0.025	-0.274*	0.900
E	A	0.002**	0.956***	-0.177***	0.369***	-0.220*	-0.284**	0.963
	B	0.000	0.895***	-0.167*	0.334***	0.365**	0.238	0.919
	C	0.001	0.919***	-0.017	0.177	0.137	-0.103	0.914
	D	0.001	0.943***	0.068	0.021	0.069	-0.128	0.933
	E	0.003	0.952***	0.197*	0.090	0.123	-0.077	0.883
	F	0.004**	0.922***	0.440***	0.136	-0.008	-0.431***	0.915
S	A	0.002	0.897***	-0.281***	0.056	-0.111	-0.035	0.942
	B	0.001	0.954***	-0.138*	0.477***	0.185	-0.113	0.943
	C	0.002*	0.958***	0.118	0.396***	0.009	-0.203	0.944
	D	0.001	0.963***	0.148	0.279**	-0.007	-0.167	0.910
	E	0.002	0.921***	0.178**	0.209*	0.148	-0.196	0.917
	F	0.001	0.983***	0.444***	0.173	0.562**	0.059	0.852
G	A	0.000	0.959***	-0.097	0.458***	0.139	-0.081	0.943
	B	0.002	0.931***	-0.082	0.155	-0.040	0.102	0.919
	C	0.002*	0.941***	-0.119	0.129	0.054	-0.046	0.928
	D	0.002	0.938***	0.029	0.259**	0.169	-0.065	0.909
	E	0.003**	0.902***	0.007	0.222*	0.052	-0.224	0.916
	F	0.002	0.939***	0.140	0.085	0.044	-0.440**	0.880

* p<0.05, ** p<0.01, *** p<0.001

Source: Own production

The alpha values provide similar conclusions to the FF3 model. The ESGC score still reports significant alphas in portfolio A and B, both at a 5% level. In contrary, the ESG score returns significant alphas in portfolio D and F, similar to previous results. In addition, portfolio B also returns a significant positive alpha with a monthly excess return of 0.2% at a 5% significance level. The E score has also shown an interesting development, where portfolio A now exhibits a significant positive alpha. However, the level is still lower than that of portfolio F. In the previously performed model, the S score only returned insignificant alphas, whereas the portfolio C alpha now has a significance at a 5% level. Lastly, the G score still has significant alphas for portfolio C and E, but no longer for portfolio D.

In relation to the market factor, some of the observed betas are slightly lower than those observed in the FF3 model. This may have been affected by the additional risk factors included in the model. In the FF5 model, the higher-scoring portfolios exhibit lower market-related risk in the ESGC, ESG, and S scores. Looking at the SMB factor, the trends are similar to the ones presented in the previous model – the higher-scoring portfolios have negative exposures, and the lower-scoring portfolios have positive exposures. However, the tendency is not as explicit as in the FF3 model and fewer SMB coefficients are significant. The HML factor no longer shows negative coefficients in the lower-scoring portfolios, indicating growth stocks, but the factor exposures are still generally lower in the low portfolios compared to the high portfolios.

Turning to the newly added explanatory variables, that is RMW and CMA, only a few coefficients are significant. RMW is the profitability factor, and the coefficients hereof are generally lower for the A portfolios compared to the F portfolios. This indicates that companies included in the A portfolios generally exhibit lower profitability than those included in the F portfolios. The CMA coefficients are almost all negative, indicating that the portfolios all have a negative exposure towards the investment factor. The values are generally less negative, or higher, for the higher-scoring portfolios, which is typically seen for companies with more conservative investment strategies. Hereby, the RMW and CMA trends match the findings illustrated in section 5.2.1. Based on the several insignificant coefficients observed in the RMW and CMA factors, these factors do not appear to add much explanation towards the variation in returns. This can also be seen through the adjusted R-squared values which are quite similar to the ones presented in the FF3 model, while some are even lower.

5.2.6 Summary Value-Weighted Portfolios

In the examination of the portfolios' characteristics, the high- and low-scoring portfolios show some clear tendencies according to size, value, profitability, and investments. This is also supported by the portfolios' factor exposures in the FF3 and FF5 models. From the preliminary examination of the portfolios' descriptive statistics and performance measures, the ESGC score showed that the higher-scoring portfolios have better risk-adjusted performance than the lower-scoring portfolios. This is supported by the results obtained from the factor models, where the ESGC score report significant positive alphas in portfolio A and B, according to both FF3 and FF5. Furthermore, portfolio A in the ESGC score also exhibits lower market-related risk than portfolio F across all tested factor-models. For the remaining scores, portfolio F seems to be outperforming portfolio A which can be seen through the higher risk-adjusted returns in the descriptive statistics. However, the lower-scoring portfolios also show higher tail-risk and downside volatility. Especially during the recent COVID-19 crisis, portfolio F exhibited more extreme losses than portfolio A according to the ESGC score. After controlling for the various risk factors, only the ESG and E scores report significant positive alphas in the F portfolios across all factor models. This could indicate that the lower-scoring portfolios generally exhibit higher risk and hereby require higher returns.

5.3 Sub Samples

In the following section, the same procedures as presented in the previous sections have been repeated for two sub samples. The first sub sample holds observations from January 1st, 2011 to December 31st, 2015, whereas the second sub sample holds observations from January 1st, 2016 to December 31st, 2020. As mentioned in section 4.2.2, the split has been made to see whether the testing results differ significantly between the two periods. Moreover, if the increasing trend towards ESG-investing can be observed in the results. Similar to the full period analysis, both sub samples will be studied through a preliminary investigation of the portfolios' risk and return properties. Hereafter, the portfolios' returns will be tested according to the three applied factor models. Aforementioned, the descriptive statistics and testing results will only be reported for the A and F portfolios.

5.3.1 Results: Sub period 2011-2015

The return and risk properties have been listed in table 5.7, across the two portfolios, namely A and F, for each studied score. The cumulative return is the portfolios' aggregated return across the early 5-year sub sample January 1st, 2011 to December 31st, 2015. Moreover, the annual return is the annualized geometric mean of monthly returns minus the risk-free rate, and the standard deviation is also the annual estimate.

Table 5.7: Descriptive Statistics and Performance Measures (2011-2015)

Score	*Portfolio*	*Cumulative Return*	*Annual Return*	*Standard Deviation*	*Beta*	*Sharpe Ratio*	*Treynor Ratio*	*Skewness*	*Kurtosis*	*Maximum Drawdown*
ESG comb	A	56.68%	3.37%	13.5%	1.00	0.25	0.03	-0.56	0.36	-21.39%
	F	98.19%	8.36%	12.4%	0.82	0.67	0.10	-0.49	0.05	-19.08%
ESG	A	37.24%	0.67%	13.1%	0.96	0.05	0.01	-0.37	0.24	-19.06%
	F	109.50%	9.57%	12.5%	0.84	0.77	0.11	-0.40	-0.13	-18.85%
E	A	38.89%	0.89%	14.4%	1.05	0.06	0.01	-0.48	0.15	-22.88%
	F	108.36%	9.46%	12.1%	0.84	0.78	0.11	-0.44	0.24	-20.57%
S	A	42.56%	1.45%	12.6%	0.94	0.11	0.02	-0.46	0.34	-18.36%
	F	126.54%	11.34%	12.7%	0.75	0.89	0.15	-0.10	-0.52	-13.03%
G	A	29.16%	-0.55%	13.2%	0.98	-0.04	-0.01	-0.37	0.55	-21.87%
	F	62.42%	4.13%	9.6%	0.90	0.43	0.05	-0.09	0.03	-18.20%

Source: Own production

Compared to the full dataset, the A portfolios have much lower cumulative and annual returns than the F portfolios across all scores. Looking at the G score the annual return is even negative for portfolio A. The risk parameters, i.e. standard deviation and beta, also provide different observations. Here, the A portfolios generally signal higher risk across all scores. The general poor performance of the higher-scoring portfolios is also observed through the substantially lower risk-adjusted return measures. Turning to the return distributions, the skewness and kurtosis measures show contradictory results to the ones previously obtained. In the full dataset larger tail-risk was observed in the lower-scoring portfolios, both indicated by skewness and maximum drawdown estimates, whereas the general conclusion is opposite here. From figure 5.3 in section 5.2.2, it was also observed that portfolio A experienced higher drawdowns in the early period of the analysis which is supported by the findings in this section.

Similar to previous analysis, the returns have also been tested using the three performance benchmark models, these being CAPM, FF3, and FF5. The testing results obtained across all scores, models, and the two relevant portfolios have been summarized and reported in table 5.8. The most important findings are related to the alphas, whereas the remaining factors will not be elaborated upon. From the results, it is clear to see that the F portfolios are performing significantly better than the A portfolios. Across all scores, the alphas in the F portfolios are nominally higher than those of the A portfolios. Moreover, several F portfolio alphas are also significant on a 5% or 1% level. The G score is the only score having a significant alpha represented in portfolio A. However, the value is negative and hereby provides the same conclusion – the lower-scoring portfolios are superior.

Table 5.8: CAPM, FF3 and FF5 (2011-2015)

Score	*Model*	*Portfolio*	*Alpha*	*MKT*	*SMB*	*HML*	*RMW*	*CMA*	*Adj R^2*
ESG comb	CAPM	A	0.000	1.003***					0.923
		F	0.004*	0.840***					0.801
	FF3	A	0.002	0.912***	-0.173*	0.274***			0.946
		F	0.004	0.922***	0.457***	0.035			0.839
	FF5	A	0.001	0.926***	-0.142	0.354**	0.166	0.039	0.945
		F	0.004	0.892***	0.412**	0.057	-0.114	-0.275	0.839
ESG	CAPM	A	-0.002	0.964***					0.926
		F	0.005*	0.859***					0.790
	FF3	A	0.000	0.853***	-0.225***	0.318***			0.962
		F	0.004*	0.951***	0.510***	0.038			0.835
	FF5	A	0.000	0.848***	-0.234**	0.304**	-0.038	-0.025	0.961
		F	0.005	0.926***	0.480***	0.145	0.012	-0.324	0.836
E	CAPM	A	-0.002	1.054***					0.890
		F	0.005*	0.863***					0.822
	FF3	A	0.001	0.911***	-0.209**	0.489***			0.954
		F	0.005*	0.929***	0.449***	0.106			0.865
	FF5	A	0.001	0.890***	-0.242**	0.491***	-0.096	-0.180	0.954
		F	0.005*	0.895***	0.409***	0.261	0.027	-0.450*	0.874
S	CAPM	A	-0.002	0.944***					0.932
		F	0.007**	0.755***					0.762
	FF3	A	0.000	0.861***	-0.210**	0.201***			0.949
		F	0.006**	0.827***	0.259	-0.103			0.771
	FF5	A	0.000	0.868***	-0.208**	0.119	-0.068	0.140	0.949
		F	0.006*	0.825***	0.259	-0.063	0.038	-0.060	0.763
G	CAPM	A	-0.003*	0.981***					0.924
		F	0.001	0.910***					0.798
	FF3	A	-0.002	0.894***	-0.116	0.308***			0.953
		F	-0.001	1.013***	0.302*	-0.207*			0.816
	FF5	A	-0.001	0.870***	-0.150	0.349**	-0.061	-0.243*	0.955
		F	0.000	0.950***	0.218	-0.035	-0.088	-0.713**	0.839

* p<0.05, ** p<0.01, *** p<0.001

Source: Own production

5.3.2 Results: Sub period 2016-2020

For the sample period January 1st, 2016 to December 31st, 2020, the return and risk properties have been listed in table 5.9. Compared to the full sample, the most important observation is that portfolio A has a much higher cumulative and annual return than portfolio F, according to the ESGC score. Moreover, the ESGC score is no longer the only one showing superior returns in the high portfolio. The S score also demonstrates higher returns in portfolio A. Another important finding is that both the ESGC score and S score show better risk-adjusted performance in portfolio A compared to portfolio F, indicated by the Sharpe and Treynor ratios. For the late 5-year sub sample, the return distributions are more similar to the full data period where the lower-scoring portfolios indicate higher tail-risk. Furthermore, the F portfolios exhibit higher maximum drawdowns than the A portfolios for all scores except from G. This was also illustrated for the ESGC score in the late period of figure 5.3 in section 5.2.2.

Table 5.9: Descriptive Statistics and Performance Measures (2016-2020)

Score	*Portfolio*	*Cumulative Return*	*Annual Return*	*Standard Deviation*	*Beta*	*Sharpe Ratio*	*Treynor Ratio*	*Skewness*	*Kurtosis*	*Maximum Drawdown*
ESG comb	A	50.70%	13.05%	14.83%	0.95	0.88	0.14	-0.01	3.08	-19.83%
	F	19.34%	7.93%	18.51%	1.30	0.43	0.06	-0.36	6.19	-32.41%
ESG	A	31.30%	10.00%	15.32%	0.99	0.65	0.10	0.40	3.88	-20.78%
	F	62.59%	14.79%	15.75%	1.01	0.94	0.15	-0.92	4.01	-23.41%
E	A	41.74%	11.69%	17.11%	1.10	0.68	0.11	0.15	4.12	-22.59%
	F	65.89%	15.25%	15.98%	1.02	0.95	0.15	-0.85	2.93	-23.26%
S	A	41.73%	11.70%	14.30%	0.91	0.82	0.13	0.25	2.69	-17.23%
	F	32.87%	10.27%	17.36%	1.14	0.59	0.09	-0.57	4.85	-29.91%
G	A	27.92%	9.44%	16.75%	1.13	0.56	0.08	0.36	6.27	-27.26%
	F	79.15%	17.04%	10.79%	0.99	1.58	0.17	0.03	3.33	-19.52%

Source: Own production

The returns have also been tested using the three factor models and the results are listed in table 5.10. The ESGC score returns significant and positive alphas in portfolio A across all tested models with a monthly excess return of 0.3%. This is in contrast to the full dataset where portfolio A only returned significant alphas in the Fama-French models with an excess return of 0.2%. The ESG score does not provide any significant alphas in the Fama-French models which can be considered a positive development, since the F portfolio had significant alphas in the previous analysis. A more interesting development is that both the E and S scores are returning significant positive alphas in portfolio A in both Fama-French models, with monthly excess returns at 0.3%. For the E score, portfolio F has a significant alpha in the CAPM model, but it disappears when controlling for the other risk factors. Portfolio F even shows negative alphas in both FF3 and FF5 according to the S score, though, not on a significant level. This development signals an increasing focus on the environmental and social aspects of the companies' operations. However, the G score still returns significant positive alphas in portfolio F at high nominal values of 0.5-0.6%, indicating outperformance by the lower-scoring portfolio. Finally, the adjusted R-squared values are generally higher in the sub sample analysis compared to the full period analysis. This indicates that the 5-year period models provide better explanatory power than the full 10-year period models.

Table 5.10: CAPM, FF3 and FF5 (2016-2020)

Score	*Model*	*Portfolio*	*Alpha*	*MKT*	*SMB*	*HML*	*RMW*	*CMA*	*Adj R^2*
ESG comb	CAPM	A	0.003*	0.954***					0.941
		F	-0.003	1.299***					0.897
	FF3	A	0.003**	0.985***	-0.262**	0.005			0.950
		F	0.000	1.118***	0.285**	0.492***			0.952
	FF5	A	0.003**	0.970***	-0.255**	0.100	0.237	0.023	0.951
		F	0.000	1.129***	0.246*	0.384**	-0.399*	-0.158	0.956
ESG	CAPM	A	0.000	0.987***					0.925
		F	0.004*	1.013***					0.901
	FF3	A	0.002	0.997***	-0.421***	0.140**			0.957
		F	0.002	0.984***	0.530***	-0.123*			0.942
	FF5	A	0.002	0.986***	-0.424***	0.206*	0.136	-0.015	0.957
		F	0.002	0.959***	0.489***	-0.014	0.068	-0.187	0.941
E	CAPM	A	0.001	1.096***					0.941
		F	0.004*	1.025***					0.899
	FF3	A	0.003**	1.028***	-0.137	0.284***			0.968
		F	0.003	0.993***	0.537***	-0.115*			0.939
	FF5	A	0.003**	1.021***	-0.199**	0.256**	-0.325**	-0.269*	0.975
		F	0.003	0.957***	0.461***	0.024	-0.015	-0.341	0.941
S	CAPM	A	0.002	0.912***					0.923
		F	0.000	1.137***					0.885
	FF3	A	0.003*	0.963***	-0.432***	0.006			0.951
		F	-0.001	1.119***	0.338*	-0.079			0.895
	FF5	A	0.003*	0.944***	-0.473***	0.080	-0.011	-0.184	0.951
		F	-0.001	1.101***	0.382**	0.079	0.528*	0.176	0.903
G	CAPM	A	-0.001	1.130***					0.900
		F	0.006***	0.987***					0.924
	FF3	A	0.002	1.041***	-0.203*	0.383***			0.946
		F	0.005**	0.983***	0.126	-0.039			0.924
	FF5	A	0.002	1.048***	-0.173	0.372**	0.097	0.128	0.945
		F	0.005**	0.954***	0.082	0.090	0.100	-0.200	0.923

* p<0.05, ** p<0.01, *** p<0.001

Source: Own production

Comparing the descriptive statistics and testing results for the two sub samples, major differences can be observed. The results in the early sub sample gave a clear negative picture of ESG-investing. In the early period, the A portfolios were performing worse than the F portfolios both in terms of return and risk. Moreover, the F portfolios were generating significant positive alphas. However, the later sub sample shows a substantial positive development. According to the descriptive statistics, the A portfolios show superior performance in both the ESGC and S scores. Additionally, both the ESGC, E, and S scores have significant positive alphas in the A portfolios. Moreover, the downside risk difference between the A and F portfolios have also undergone a development across the two sub samples, similar to what was illustrated in figure 5.3. The A portfolios seem to show higher downside risk in the former, whereas the opposite is true for the latter. This development indicates that the ESG focus may have changed over the course of the last 10 years.

5.4 Industry-Weighted Portfolios

The purpose of the following section is to study whether the previous findings have been caused by an industry displacement rather than being a product of ESG performance. Aforementioned, this is done by following a "best-in-class" investment strategy. For this part of the analysis, it is not found relevant to include the descriptive statistics. However, the performance benchmark models will be performed both on the full 10-year period and for the 5-year sub samples. Similar to previously, the results will only be reported for the A and F portfolios.

5.4.1 Results: CAPM, Fama-French Three- and Five-Factor Models

In table 5.11 the results are listed across all scores, models, and the two relevant portfolios, using the complete 10-year period of analysis. Starting from the top, the ESGC score have significant positive alphas in portfolio A across all models, but with the highest level of significance in the FF3 and FF5 models. The alphas signal monthly excess returns at 0.2% and 0.3% respectively. These results are similar, but slightly better, than the results obtained in section 5.2 where the significance level were lower and only present in the Fama-French models. In the value-weighted portfolios, the ESG score had significant positive alphas in portfolio F across all three models. Here, it is only present in the CAPM model and disappears when controlling for additional risk factors. For the E score, both portfolio A and F provide significant positive alphas. However, portfolio F is the superior one as it returns a higher nominal alpha. This is the same conclusion as previously drawn. The S score holds similar results to the E score, where both portfolio A and F have significant positive alphas, but portfolio F show higher monthly excess return. These results differ from previously, where the only significant positive alpha was found in the CAPM F portfolio. The same difference is applicable for the G score.

Turning to the risk factors, some similarities and differences are observed when comparing the industry-weighted portfolios and the value-weighted portfolios. Previously, the market factor indicated a somewhat clear tendency, that the A portfolios exhibited lower market-related risk than the F portfolios. This is no longer the case, as the only score exhibiting this trend is ESGC score. The remaining risk factors display the same trends as the previously obtained results, and it is therefore not found relevant to elaborate further.

Table 5.11: Industry-Weighted: CAPM, FF3 and FF5

Score	*Model*	*Portfolio*	*Alpha*	*MKT*	*SMB*	*HML*	*RMW*	*CMA*	*Adj R^2*
ESG comb	CAPM	A	0.002*	0.996***					0.949
		F	0.002	1.011***					0.891
	FF3	A	0.003***	0.957***	-0.091	0.146***			0.957
		F	0.002	0.992***	0.340***	0.101*			0.913
	FF5	A	0.003***	0.946***	-0.102*	0.227***	0.103	-0.082	0.957
		F	0.001	0.976***	0.344***	0.360***	0.484***	-0.059	0.922
ESG	CAPM	A	0.000	1.064***					0.941
		F	0.004**	0.904***					0.858
	FF3	A	0.001	0.991***	-0.101*	0.275***			0.965
		F	0.002	0.948***	0.467***	-0.130**			0.902
	FF5	A	0.001	0.968***	-0.133**	0.388***	0.074	-0.200*	0.966
		F	0.001	0.922***	0.447***	0.111	0.354**	-0.178	0.908
E	CAPM	A	0.001	1.076***					0.944
		F	0.006***	0.847***					0.846
	FF3	A	0.002**	1.009***	-0.101*	0.253***			0.963
		F	0.004***	0.896***	0.437***	-0.153**			0.892
	FF5	A	0.002**	0.989***	-0.131**	0.333***	0.025	-0.175	0.964
		F	0.004**	0.851***	0.378***	0.095	0.210	-0.377**	0.900
S	CAPM	A	0.000	1.038***					0.946
		F	0.005***	0.949***					0.878
	FF3	A	0.002*	0.974***	-0.147***	0.235***			0.966
		F	0.003**	0.990***	0.433***	-0.122**			0.913
	FF5	A	0.002*	0.961***	-0.169***	0.281***	-0.005	-0.128	0.966
		F	0.003*	0.960***	0.412***	0.170	0.444***	-0.196	0.922
G	CAPM	A	0.000	1.017***					0.933
		F	0.003**	0.900***					0.895
	FF3	A	0.001	0.968***	-0.055	0.185***			0.943
		F	0.003*	0.920***	0.191**	-0.059			0.901
	FF5	A	0.000	0.950***	-0.072	0.338***	0.207*	-0.135	0.945
		F	0.003*	0.879***	0.129	0.103	0.043	-0.369**	0.907

* p<0.05, ** p<0.01, *** p<0.001

Source: Own production

5.4.2 Results: Sub samples

As the results from the 5-year sub samples in section 5.3 showed some interesting developments, it is also found important to test the industry-weighted portfolios in the two period splits. In the following section, the testing results will be presented for both the early (2011-2015) and late (2016-2020) sub sample.

The testing results from the early sub sample are listed in table 5.12, which will be compared to the results obtained in section 5.3.1. In the previous analysis, it became clear that portfolio F performed significantly better than portfolio A across all scores. This conclusion is similar for the scores ESG, E, S, and G. For these scores, portfolio F has higher positive alphas than their counter portfolio A, and several of these are significant. However, the ESGC score show different results. In contradiction to the other scores, the ESGC score show significant positive alphas in portfolio A at a 5% level with monthly excess returns at 0.3%. The nominal alpha is equal to the one obtained for the full period of analysis. Thereby, the results are slightly better for the sector-neutral portfolios compared to the obtained results in section 5.3.1. However, the negative transformation between the results from the full data period and the early sub sample is still visible, similar to the value-weighted portfolio analysis. This can be seen through the higher performance gab between the two portfolios, favouring portfolio F.

Table 5.12: Industry-Weighted: CAPM, FF3 and FF5 (2011-2015)

Score	*Model*	*Portfolio*	*Alpha*	*MKT*	*SMB*	*HML*	*RMW*	*CMA*	*Adj R^2*
ESG comb	CAPM	A	0.002	0.974***					0.937
		F	0.003	0.844***					0.882
	FF3	A	0.003*	0.916***	-0.071	0.209***			0.950
		F	0.002	0.923***	0.358***	-0.041			0.905
	FF5	A	0.003*	0.908***	-0.079	0.258*	0.020	-0.120	0.949
		F	0.002	0.902***	0.334***	0.056	0.020	-0.275	0.907
ESG	CAPM	A	-0.001	1.033***					0.929
		F	0.005*	0.769***					0.811
	FF3	A	0.001	0.950***	-0.086	0.314***			0.956
		F	0.003	0.876***	0.423***	-0.114			0.848
	FF5	A	0.001	0.924***	-0.127	0.326**	-0.112	-0.238	0.958
		F	0.003	0.867***	0.433***	0.122	0.246	-0.316	0.854
E	CAPM	A	-0.001	1.064***					0.929
		F	0.005**	0.798***					0.804
	FF3	A	0.001	0.975***	-0.085	0.342***			0.961
		F	0.004*	0.912***	0.485***	-0.088			0.848
	FF5	A	0.002	0.954***	-0.119	0.343**	-0.102	-0.185	0.961
		F	0.004	0.886***	0.466***	0.136	0.153	-0.445*	0.859
S	CAPM	A	-0.001	1.023***					0.938
		F	0.007***	0.816***					0.848
	FF3	A	0.001	0.943***	-0.105	0.283***			0.961
		F	0.005**	0.923***	0.369***	-0.165*			0.879
	FF5	A	0.001	0.927***	-0.134	0.241*	-0.126	-0.090	0.961
		F	0.005**	0.903***	0.358***	0.039	0.157	-0.376*	0.887
G	CAPM	A	-0.001	0.981***					0.939
		F	0.001	0.878***					0.866
	FF3	A	0.000	0.936***	-0.022	0.195***			0.951
		F	0.001	0.950***	0.274*	-0.086			0.877
	FF5	A	0.000	0.907***	-0.057	0.298**	-0.010	-0.349**	0.956
		F	0.001	0.916***	0.238*	0.090	0.056	-0.464*	0.887

* p<0.05, ** p<0.01, *** p<0.001

Source: Own production

In table 5.13 the testing results have been listed for the late sub sample, which will be compared to the ones obtained in section 5.3.2. The ESGC score show significant positive alphas in both the FF3 and FF5 models, similar to previously, but no longer in the CAPM model. An important difference is that the ESG score also show positive significant alphas in the FF3 and FF5 models. Compared to the late sample in the value-weighted portfolio analysis, where the only significant alpha in the ESG score was present in the CAPM F portfolio. The E score results also differ from the ones previously found, as both portfolio A and F have significant alphas in FF3 and FF5. Moreover, portfolio F has higher nominal alphas which indicate that it is superior. This differs significantly from the previous results, where only portfolio A returned significant alphas. The S and G scores provide the exact same conclusion as previously; the S score having significant excess returns in portfolio A, and vice versa for the G score. Overall, the results are similar for the ESGC, S, and G scores, but different for the ESG and E scores. Similar to the value-weighted portfolio analysis, the late sub sample provide a more positive outlook for ESG-investing, compared to the results obtained from both the full data period and the early sub sample.

Table 5.13: Industry-Weighted: CAPM, FF3 and FF5 (2016-2020)

Score	*Model*	*Portfolio*	*Alpha*	*MKT*	*SMB*	*HML*	*RMW*	*CMA*	*Adj R^2*
ESG comb	CAPM	A	0.002	1.014***					0.959
		F	0.001	1.153***					0.926
	FF3	A	0.003**	1.012***	-0.195**	0.089*			0.966
		F	0.002	1.079***	0.170	0.178**			0.935
	FF5	A	0.003**	0.998***	-0.197**	0.173	0.180	-0.012	0.967
		F	0.002	1.048***	0.189	0.391**	0.546**	0.066	0.944
ESG	CAPM	A	0.000	1.087***					0.951
		F	0.002	1.020***					0.907
	FF3	A	0.002*	1.042***	-0.203**	0.237***			0.974
		F	0.001	1.017***	0.392***	-0.152**			0.933
	FF5	A	0.002*	1.012***	-0.236***	0.387***	0.192	-0.153	0.975
		F	0.001	0.994***	0.393***	-0.006	0.326	-0.007	0.936
E	CAPM	A	0.002	1.082***					0.956
		F	0.006**	0.887***					0.879
	FF3	A	0.003**	1.051***	-0.188**	0.182***			0.970
		F	0.004**	0.898***	0.403***	-0.202***			0.924
	FF5	A	0.003**	1.025***	-0.213**	0.318***	0.191	-0.120	0.971
		F	0.004**	0.832***	0.335***	0.119	0.418**	-0.322	0.933
S	CAPM	A	0.001	1.048***					0.953
		F	0.003	1.067***					0.918
	FF3	A	0.003*	1.026***	-0.268***	0.187***			0.974
		F	0.001	1.049***	0.395***	-0.100			0.938
	FF5	A	0.003*	0.999***	-0.306***	0.306***	0.104	-0.172	0.975
		F	0.001	1.013***	0.388***	0.113	0.441**	-0.044	0.944
G	CAPM	A	0.000	1.046***					0.928
		F	0.005***	0.914***					0.921
	FF3	A	0.001	1.017***	-0.168	0.164**			0.939
		F	0.005**	0.912***	0.125	-0.045			0.922
	FF5	A	0.002	1.006***	-0.140	0.265*	0.337*	0.114	0.943
		F	0.005**	0.869***	0.059	0.143	0.136	-0.302	0.923

* p<0.05, ** p<0.01, *** p<0.001

Source: Own production

5.4.3 Summary Industry-Weighted Portfolios

The factor models CAPM, FF3, and FF5, performed across the entire period of analysis, generally provide the same conclusions as the value-weighted portfolio analysis. The ESGC score show significant positive alphas in portfolio A across all performed models, indicating outperformance by the "best-in-class" investment strategy. However, the remaining scores indicate outperformance in portfolio F, although only present in CAPM for the ESG score. According to the market-related risk, portfolio A only exhibits lower risk than portfolio F in the ESGC score, which is reversed for the remaining scores. Similar to previous analysis, the results are also studied for the two sub samples. The early sub sample give somewhat the same conclusions as the full period, with outperformance in portfolio A for the ESGC score and outperformance in portfolio F for the remaining scores. However, the results are still worse from the pro-ESG perspective, similar to section 5.3.1, as the outperformance in the F portfolios are significantly higher in the early period. For the late sub sample, a positive development is observed similar to section 5.3.2. Here, the higher-scoring portfolios outperform the lower-scoring portfolios in both the ESGC, ESG, and S scores. The F portfolios remain superior according to the E and G scores.

5.5 Excluding outliers

The purpose of this section is to observe whether the identified outliers have an impact on the results obtained in the previous analysis. As illustrated in figure 5.1 in section 5.1, some extreme outliers have been identified in the dataset, namely *Neste*, *Ashtead Group* and *GenMab*. It is hereby important to test whether the results are significantly different when removing the outliers from the sample, as explained in section 4.3.4. After removing the three outliers from the dataset, the normal testing procedure is performed. It is assumed that the 10-year period results will be sufficient to determine whether the outliers have an effect on the results. The results are presented in table 5.14.

Table 5.14: Excluding Outliers: CAPM, FF3 and FF5

Score	*Model*	*Portfolio*	*Alpha*	*MKT*	*SMB*	*HML*	*RMW*	*CMA*	*Adj R^2*
ESG comb	CAPM	A	0.001	0.981***					0.933
		F	0.001	1.085***					0.830
	FF3	A	0.002*	0.945***	-0.184***	0.123***			0.943
		F	0.002	1.013***	0.458***	0.321***			0.887
	FF5	A	0.002*	0.940***	-0.188***	0.166*	0.060	-0.036	0.942
		F	0.001	0.998***	0.441***	0.424***	0.114	-0.123	0.886
ESG	CAPM	A	-0.001	0.977***					0.925
		F	0.004**	0.929***					0.849
	FF3	A	0.001	0.913***	-0.258***	0.226***			0.952
		F	0.003*	0.956***	0.509***	-0.061			0.894
	FF5	A	0.001	0.907***	-0.266***	0.262***	0.033	-0.052	0.952
		F	0.003*	0.927***	0.463***	0.048	0.013	-0.268	0.896
E	CAPM	A	-0.001	1.078***					0.917
		F	0.004**	0.955***					0.868
	FF3	A	0.002	0.981***	-0.122*	0.367***			0.957
		F	0.003**	0.972***	0.510***	-0.020			0.911
	FF5	A	0.002*	0.955***	-0.176***	0.368***	-0.216*	-0.279**	0.962
		F	0.003**	0.929***	0.442***	0.136	0.005	-0.400**	0.918
S	CAPM	A	0.000	0.930***					0.926
		F	0.003	0.960***					0.819
	FF3	A	0.001	0.900***	-0.270***	0.096**			0.941
		F	0.002	0.990***	0.382***	-0.081			0.840
	FF5	A	0.002	0.898***	-0.280***	0.053	-0.114	-0.035	0.941
		F	0.001	0.989***	0.415***	0.157	0.564***	0.098	0.854
G	CAPM	A	-0.002	1.061***					0.904
		F	0.003*	0.953***					0.863
	FF3	A	0.000	0.969***	-0.077	0.352***			0.940
		F	0.002	0.985***	0.207*	-0.107			0.872
	FF5	A	0.000	0.960***	-0.083	0.440***	0.136	-0.057	0.940
		F	0.002	0.937***	0.133	0.080	0.037	-0.439**	0.878

* p<0.05, ** p<0.01, *** p<0.001

Source: Own production

As observed in table 5.14, the overall conclusions are identical to the ones drawn from the value-weighted portfolios based on the complete dataset, i.e. the exact same alphas are significant. However, some of the nominal values differs slightly. In the ESG score, the significant alpha for the CAPM F portfolio is 0.4%, compared to 0.5% before. Similar in the E score, the alpha values are 0.4%, 0.3%, and 0.3%, respectively, compared to the previous results being 0.5%, 0.4%, and 0.4%. These small changes may be caused by one or several outliers being located in the F portfolios, which can adjust the excess returns slightly up or down. However, as mentioned, the overall conclusions have not been affected by the exclusion of outliers, as the same portfolio alphas are proven significant in the results. For the record, the two sub samples have also been tested and does not return contradictory results (see results in Appendix VI).

5.6 Summary

From the value-weighted portfolio analysis, the ESGC score shows that portfolio A is superior to portfolio F, both according to risk-adjusted returns and significant excess returns. The remaining scores show significantly higher risk-adjusted returns in portfolio F. However, after controlling for the various risk parameters only the ESG and E scores return significant excess returns in portfolio F. This could be due to the generally higher tail-risk and maximum drawdowns observed across the F portfolios. Especially during the recent COVID-19 crisis, portfolio F experienced much higher drawdowns compared to portfolio A, signalling higher downside risk during economic downturns. From the sub sample analysis, an interesting development is found from the early to late period of analysis. In the early sub sample, portfolio F is the clear winner across all scores. However, the late sub sample demonstrates a significant positive development, as both the ESGC, E, and S scores provide significant excess returns in portfolio A. In the industry-weighted portfolio analysis, the most important finding is that the ESGC score show significant excess returns in portfolio A across all tested sample periods, implying that the results are not only affected by a sector-displacement.

Chapter 6

Discussion and Conclusion

6.1 Discussion

The following sections holds a critical reflection of the results presented in the analysis. Firstly, the main findings of this study will be summarized and compared to the formulated hypotheses introduced in connection with the research question. Furthermore, the results will be related to the empirical and theoretical foundation of this thesis to discuss the possible explanations behind the obtained results. Hereafter, some of the important delimitations will be revisited to determine the limits of the results' generalizability. Lastly, recommendations for further research are presented as well as future implications for ESG-investing.

6.1.1 Interpretation of Results

Before summarizing the main findings, it is found important to revisit the research question that frames the objective and direction of this study: "Do ESG scores have an impact on financial performance?". In order to make the problem area more tangible, three hypotheses were also formulated in connection with the research question which will be addressed in the following paragraphs.

The study demonstrates strong evidence of superior risk-adjusted returns for high-scoring portfolios as opposed to low-scoring portfolios, based on the ESGC score exclusively. This finding is hereby supporting $H1$ stating that "*portfolios consisting of companies with strong ESG performance generate higher risk-adjusted returns than portfolios consisting of those with weak ESG performance*". This conclusion is true, both for the full data period analysis and the late sub sample. However, the remaining scores, ESG, E, S, and G fail to provide evidence to support the null hypothesis in the full data period. The ESG and E scores are actually providing evidence for the alternative hypothesis that is "*portfolios consisting of companies with weak ESG performance generate higher risk-adjusted returns than portfolios consisting of those with strong ESG performance*". The evidence towards the alternative hypothesis is even stronger looking at the early sub sample, where several scores provide significant abnormal returns in the F portfolios. However, in the late sub sample analysis, both the ESGC, E, and S scores are supporting the null hypothesis, whereas the G score supports the alternative hypothesis.

The results also suggest that the high-scoring portfolios are experiencing lower risk, in terms of standard deviation and beta, than the low-scoring portfolios, according to both the ESGC and S scores. This provides supporting evidence for $H2$ that states "*portfolios consisting of companies with strong ESG performance demonstrate lower volatility in returns compared to portfolios consisting of those with weak ESG performance*". The remaining scores, on the other hand, provide evidence for the alternative hypothesis that is "*portfolios consisting of companies with weak ESG performance demonstrate lower volatility in returns compared to portfolios consisting of those with strong ESG performance*". However, looking at the downside volatility measures exclusively, the findings demonstrate strong evidence to support the null hypothesis across all scores, except from the G score. The evidence is provided by the skewness and maximum drawdown measures, as these are more negative for the low portfolios compared to the high portfolios. This conclusion is true for both the full data period and the late sub sample. However, the early sub sample generally provides evidence for the alternative hypothesis.

The general outperformance observed in the high portfolio based on the ESGC score is also proven robust in a sector neutral environment. This is based on the results found in the industry-weighted portfolio analysis where portfolio A shows significant positive alphas across all tested periods. The findings are hereby supporting $H3$ that states "*portfolios consisting of companies with high "best-in-class" ESG ratings are outperforming portfolios consisting of those with low "best-in-class" ESG ratings*". The opposite is true for the remaining scores which support the alternative hypothesis: "*portfolios consisting of companies with low "best-in-class" ESG ratings are outperforming portfolios consisting of those with high "best-in-class" ESG ratings*". The results from the early sub sample provide similar evidence. However, looking at the late sub sample, both the ESGC, ESG, and S scores provide evidence supporting the null hypothesis, whereas the E and G scores still support the alternative hypothesis.

Similar to previous studies, this study fails to provide an unambiguous conclusion on the relationship between ESG scores and financial performance. As seen through the main findings presented above, the conclusions vary significantly depending on the specific score and period of analysis being tested. However, the ESGC score provide strong evidence towards the financial prosperity of following a positive screening approach. This is in line with the findings provided by Kempf & Osthoff (2007) and Statman & Glushkov (2009) that both observe significant abnormal performance in the high portfolios. Although, the studies are based on different scores, markets, and periods of analysis. The ESG score, on the other hand, generally returns insignificant alphas in the high portfolio, supporting the neutral standpoint proposed by the previous meta studies (Revelli & Viviani, 2015; Kim, 2019). Actually, the ESG, E, S, and G scores all provide evidence, to some extent, that portfolio F is the superior one. This is not due to a poor performance by portfolio A, instead it is due to an abnormal performance by portfolio F. These findings suggest that investors are better off by investing in ESG laggards. This is in line with the results from a recent study conducted by Pyles (2020) who also finds that the low portfolio outperforms the high portfolio. However, the evidence supporting this finding is not as strong as for the ESGC score, because the results are not withstanding across different time periods or screening methods. For instance, the E and S scores show opposite results for the late sub sample and so does the ESG score in the “best-in-class” screening approach, hereby supporting the positive screening approach.

According to the efficient market hypothesis, introduced in section 3.1, all assets in the market are trading at their fair values that reflect all available information. The theory hereby argues that it should not be possible to generate abnormal returns by incorporating ESG scores into your investment decision-making, as these are only reflecting publicly available information. However, in reality most markets display some level of inefficiency due to factors such as information asymmetry, transactions costs, and irrational buying behaviour (Malkiel, 2003). These inefficiencies allow for investors to identify and trade mispriced assets and hereby generate abnormal returns. The significant positive alphas obtained in this study also imply that markets are not efficient and that it is possible to incorporate ESG information to generate abnormal returns.

The identified abnormal returns observed in the high portfolios, according to the ESGC score, could suggest that ESG signal better value. This is in line with the arguments given by SRI supporters, presented in section 2.3. These include, among others, that leading socially responsible companies are better at mitigating certain risks and experience higher operational performance (Revelli & Viviani, 2015; Berry & Junkus, 2010). This can also be supported by several findings from the analysis. From the preliminary investigation of the high and low portfolios’ characteristics, it became clear that the companies in the high portfolios were generally characterized as bigger and more stable value companies (see section 5.2.1). Furthermore, the high portfolio, based on the ESGC score, generally showed lower levels of market beta and downside risk, and proved to be more resilient through the recent economic downturn caused by COVID-19. ESG funds and equities also experienced record inflows during 2020 (Jessop & Howcroft, 2021) which could be proving the supporters’ arguments of higher resilience. Another potential driver of the abnormal returns could be the accelerating growth in SRI, motivated by an increasing demand and regulatory pressure, which has led to higher capital inflows into stocks with higher ESG rating.

In contradiction to the ESGC score, the ESG score does not provide significant abnormal returns in the high portfolios. This suggests that the explanatory element lies within the ESG Controversies score, as it constitutes the only difference between the two scores. As explained in the section 4.1.1, the ESG Controversies score is a factor included in the ESGC score, that penalizes companies’ ESG scores if they are involved in any scandals related to the 23 defined controversies topics. The ESGC score can hereby be considered as the more comprehensive and accurate measure of companies’ ethical performance, as opposed to the ESG score that is essentially based on what the companies themselves choose to disclose. Furthermore, a valid reason for the inconsistency in results across the two different ESG scores, could be the data issue mentioned in section 2.1.4.

More specifically, a reason for not being able to detect abnormal returns in the ESG score, based on the aforementioned drivers, could be the inconsistent methodologies applied by the various rating agencies. For this reason, investors incorporating ESG metrics in their investment decisions do not have the same foundation, whereas it may not be possible to detect the full impact of ESG-investing yet.

Some scores also also showed significant abnormal returns in the low-scoring portfolios. An explanation hereof could be found in the sin stock premium, identified by several previous studies (Hong & Kacperczyk, 2009; Salaber, 2007; Blitz & Fabozzi, 2017). An argument supporting this premium is the "shunned stock" hypothesis where companies involved in controversial industries, such as alcohol, tobacco, and gambling, are being avoided to such a degree where they become systematically under-priced. Although, portfolio F does not specifically represent sin stocks it could be assumed that these would be found in the lower-rated portfolios. Furthermore, a "sinful" company is a relative concept and may refer to operations beyond the three listed industries. As explained in section 2.1.1, investors are currently divesting from industries such as fossil fuels and coal which could cause a "shunned stock" effect. Hong & Kacperczyk (2009) also argue that the premium could be a result of additional risk faced by sin stocks. This is in line with the findings presented in section 5.2.2 where the low portfolios generally exhibited higher tail-risk and maximum drawdowns. Thus, the sin stock premium found in previous literature could be an explanation of the abnormal returns identified in the low portfolios.

6.1.2 Limitations

It should be known that the generalizability of the results obtained in this study is limited by the study's overall research design and applied methodology. The conclusions drawn in this study are directly related to the methodological choices made in the pre-face of the study. Therefore, it is found relevant to discuss some of the limitations that have a significant impact on the results obtained.

This study's results are only covering a small sample of the European market, that is the STOXX Europe 600 Index. Moreover, the asset universe is only including companies that have been listed in the entire period of analysis, which means that the dataset is subject to survivorship bias. Therefore, it is possible that the abnormal returns identified in this paper is highly affected by the restricted investment universe. Producing the same tests using a different or larger dataset could provide varying results. The analysis is also limited to only include ESG scores from a single data provider, namely Refinitiv. This is an important limitation, as previous research show that the results differ significantly depending on which data provider is chosen (Halbritter & Dorfleitner, 2015). Therefore, it is important to note that the results obtained in this study are only valid for the ratings provided by Refinitiv and cannot be applied for other scoring methodologies.

Another limitation refers to the construction of portfolios. In this study, it was chosen to construct value-weighted portfolios, that puts more emphasis on the larger companies. Alternatively, a different weighting scheme could have been applied, i.e. equally-weighted portfolios, which gives the same weighting to each company. Hereby, the smaller companies in the portfolios would have gained higher importance, which could have altered the final results. Finally, the portfolios' returns have not been corrected for any transaction costs or tax considerations. In practice, the implementation of the tested strategies would involve transaction costs, both when entering the investment and when rebalancing the portfolios. Furthermore, tax also has an important impact on the investors realized returns. It would hereby require an analysis of both parameters to determine the actual profitability of entering into the strategies that proved to perform best.

6.1.3 Recommendations and Future Implications

This paper examines whether a connection between ESG and financial performance exist. The results suggest that an investor can generate abnormal returns by investing according to companies' ESG scores. However, as the main purpose of this paper is to measure and compare the financial performance of high and low ESG portfolios, the study cannot justify, with certainty, the driving factors of the obtained abnormal returns. In the interpretation of results, several possible sources of the significant alphas were mentioned. However, it would require further investigation to determine if these were actually explaining the causal relationship. It is hereby recommended that further research is conducted on the driving sources. More specifically, whether the identified abnormal returns are a result from a temporary mispricing in the market, compensate for an additional risk factor, or is an actual sign of higher quality.

A more thorough investigation of the ESG information could also provide additional value to the field of study. As expressed previously, there is a strong inconsistency in the scoring methodologies applied by rating agencies, which leads to low correlations between the companies' final ESG scores. Therefore, it could be interesting to study the proposed strategies across various scores provided by different agencies, to see which perform best. Furthermore, an event study could be conducted to test if the release of new or revised ESG scores are visible in the market. Moreover, whether a positive or negative momentum in a company's score have a consequent positive or negative impact on the company's stock price. This type of study would support the evidence found in this paper that ESG scores have an impact on financial returns.

Finally, it is found imperative to recognize that ESG is a constantly evolving environment, whereas the abnormal returns observed across the period of analysis may be different going forward. It is expected that ESG will keep growing in the coming years, whereas the positive screening strategy could result in abnormal returns. However, the market is becoming increasingly mature and with the recent launch of the SFDR, enclosing various disclosure requirements for ESG investors, it also becomes easier for private investors to compare the available ESG products in the market. Thus, when ESG is adopted by most and incorporated more efficiently into the market, it should no longer be possible to generate abnormal returns.

6.2 Conclusion

Socially Responsible Investing (SRI) defines the practices of incorporating ESG factors into investment decision-making which has been subject for increased attention in recent years. The reasoning can be found in the growing awareness of social issues and climate change which have led investors to demand sustainable practices from companies. As such, sustainable and responsible business conduct has never been more important for the future competitiveness of a company. The increasing popularity of SRI has also drawn the attention of academic researchers, wanting to examine the financial effect of engaging in such practices. The literature review revealed that ESG is still a highly heterogenous space, mainly due to the lack of legislative standards guiding how companies should measure and report non-financial performance. Furthermore, previous studies have yet to reach a common consensus on the impact of ESG ratings on company returns. This paper offers a valuable contribution to the field of study, by performing an intensive contemporary analysis of the relationship between ESG scores and portfolio performance. More specifically, this paper examines whether high and low portfolios experience significant return and risk differences. Thus, the main objective of this study was to answer the research question "Do ESG scores have an impact on financial performance?". For this purpose, three hypotheses were formulated.

The first hypothesis (H1) states "*portfolios consisting of companies with strong ESG performance generate higher risk-adjusted returns than portfolios consisting of those with weak ESG performance*". With the aim of testing this hypothesis, six decile portfolios were formed based on companies' relative ESG ranking. From a preliminary investigation of the portfolios risk and return properties, it was found that the low portfolios demonstrated a clear outperformance. This was true for all scores, except from the ESGC score that signalled outperformance in the high portfolio. To test the portfolios' returns more thoroughly, several performance benchmark models were applied, including the CAPM, Fama-French 3-, and 5-Factor models. After controlling for various risk factors, the results showed similar findings. The ESGC score gave significant abnormal returns in the high portfolios which provided supporting evidence for H1. In contrary, the remaining scores all returned significant abnormal returns in the low portfolios, supporting the alternative hypothesis. The returns were also tested in different sub periods, in order to conduct robustness checks. The early sub sample (2011-2015) suggested a clear outperformance in the low portfolios across all scores. However, the late sub sample (2016-2020) signalled a positive development in ESG investing, as several scores, including ESGC, E, and S, provided significant abnormal returns in the high portfolios. Thus, the late sub sample provided stronger evidence for H1.

The second hypothesis (H2) "*Portfolios consisting of companies with strong ESG performance demonstrate lower volatility in returns compared to portfolios consisting of those with weak ESG performance*" was tested through an examination of the portfolios' risk properties. According to the basic risk measures, that is standard deviation and beta, only the ESGC and S scores provided supporting evidence for H2. The remaining scores either showed similar risk levels across the high and low portfolios, or even higher risk levels in the high portfolios. However, by studying the downside risk measures exclusively, these being skewness, kurtosis, and maximum drawdown, the results suggested that the low portfolios generally exhibited higher tail risk and downside volatility. These findings support H2 as the low portfolio appeared to exhibit higher risk. The late sub sample analysis provided similar results, whereas the early sub sample generally supported the alternative hypothesis.

The third hypothesis (H3) "*Portfolios consisting of companies with high "best-in-class" ESG ratings are outperforming portfolios consisting of those with low "best-in-class" ESG ratings*" was formulated to test whether the results would change significantly when applying sector neutral portfolios. For this purpose, decile portfolios were constructed based on the "best-in-class" screening approach, where companies were allocated according to their relative ESG performance among their industry peers. The ESGC provided strong evidence for H3 since it returned significant abnormal returns in the high portfolio across all performed models. However, the remaining scores indicated outperformance by the low portfolio which is contradictory to H3. Overall, these findings are in line with the findings from H1, indicating that the results obtained are not simply a product of sector displacement. The robustness of the results was also tested using the two sub samples which illustrated the same positive development in ESG investing. Additionally, the outperformance by the high portfolio in the ESGC score proved robust both in the early and late sub sample.

In summary, the findings of this paper do not provide a clear-cut answer to the financial prospective of including ESG metrics into investment decision-making. Overall, significant abnormal returns were observed in several portfolios indicating that markets are not efficient and that it is possible to obtain abnormal returns by incorporating ESG information. However, the direction of the impact is not consistent, as the ESGC score provide significant abnormal returns in the high portfolio, whereas several of the other scores provide significant abnormal returns in the low portfolio. This suggests that the sustainable-conscious investor may be able to generate abnormal returns by following a positive screening strategy based on Refinitiv's ESGC score. Furthermore, the positive trend detected from the sub sample analysis could indicate an optimistic outlook for the future implementation of ESG. Finally, it is concluded that ESG scores do have an impact on financial performance.

References

Amadeo, K. (2020, July 6). *BP Oil Spill Economic Impact.* Retrieved March 5, 2021, from The Balance: https://www.thebalance.com/bp-gulf-oil-spill-facts-economic-impact-3306212

Berg, F., Koelbel, J., & Rigobon, R. (2020, May 17). *Aggregate Confusion: The Divergence of ESG Ratings.* Retrieved February 18, 2021, from SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3438533

Berry, T. C., & Junkus, J. C. (2010). Socially Responsible Investing: An Investor Perspective. *The Journal of Business Ethics*, Vol. 112, 707-720.

Birkin, A., Voigt, I. K., & Stoess, P. (2020, August 11). *Three Ways ESG Can Make Portfolios More Resilient Post COVID-19.* Retrieved January 10, 2021, from EY: https://go.ey.com/3noCMUi

Blitz, D., & Fabozzi, F. (2017). Sin Stocks Revisited: Resolving the Sin Stock Anomaly. *Journal of Portfolio Management*, Vol. 44, 105-111.

Bloomberg. (n.d.). *Sustainable Finance*. Retrieved February 12, 2021, from Bloomberg: https://www.bloomberg.com/professional/solution/sustainable-finance/

Blowfield, M., & Murray, A. (2008). Socially Reponsible Investing. In M. Blowfield, & A. Murray, *Corporate responsibility. A critical introduction.* (pp. 280-303). Oxford: Oxford University Press.

Borgers, A., Derwall, J., Koedijk, K., & Horst, J. (2013). Stakeholder relations and stock returns: On errors in investors' expectations and learning. *Journal of Empirical Finance*, Vol. 22, 159-175.

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

Breusch, T. S. (1978). Testing for Autocorrelation in Dynamic Linear Models. *Australian Economic Papers*, 334-355.

Breusch, T., & Pagan, A. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, 1287-1294.

Brundtland, G. (1987, March 20). *Our Common Future.* Retrieved February 7, 2021, from Sustainable Development: https://sustainabledevelopment.un.org/content/documents/5987our-common-future.pdf

Camilleri, M. A. (2020). The Market for Socially Responsible Investments: A Review and Evaluation. In A. Kuna-Marszalek, & A. Klysik-Uryszek, *CSR and Socially Responsible Investing Strategies in Transitioning and Emerging Economies* (pp. 171-188). IGI Global.

Capital IQ. (n.d.). *Denmark Government Debt*. Retrieved February 2, 2021, from Capital IQ: https://www.capitaliq.com/CIQDotNet/MacroEconomics/InterestRate.aspx?companyId=136585876

Carlin, D. (2021, February 20). *The Case For Fossil Fuel Divestment.* Retrieved March 9, 2021, from Forbes: https://www.forbes.com/sites/davidcarlin/2021/02/20/the-case-for-fossil-fuel-divestment/?sh=3e4df40c76d2

Dorfleitner, G., Halbritter, G., & Nguyen, M. (2015). Measuring the level and risk of corporate responsibility – An empirical comparison of different ESG rating approaches. *Journal of Asset Management*, Vol. 16, 450-466.

Doyle, D. (2021, April 1). *What is the Impact of the EU Sustainable Finance Disclosure Regulation (SFDR)?* Retrieved February 22, 2021, from S&P Global: https://www.spglobal.com/marketintelligence/en/news-insights/blog/what-is-the-impact-of-the-eu-sustainable-finance-disclosure-regulation-sfdr

EFAMA. (2020, November). *Sustainable investment in the European asset management industry: Defining and sizing ESG strategies.* Retrieved March 20, 2021, from EFAMA: https://www.efama.org/Publications/Market%20Insights%20Issue3%20ESG%20strategies.pdf

Eurosif. (2012). *European SRI Study.* Retrieved March 5, 2021, from Eurosif: http://www.eurosif.org/wp-content/uploads/2014/05/eurosif-sri-study_low-res-v1.1.pdf

Eurosif. (2018, February 23). *European SRI Study 2018.* Retrieved Retrieved March 5, 2021, from Eurosif: https://www.eurosif.org/wp-content/uploads/2018/11/European-SRI-2018-Study.pdf

Eurosif. (n.d.). *About Us*. Retrieved February 21, 2021, from Eurosif: https://www.eurosif.org/about-us/

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, Vol. 25(2), 383-417.

Fama, E. F., & MacBeth, J. D. (1973). Risk, Return, and Equilibrium: Empirical Tests. *The Journal of Political Economy* , 81(3), 607-636.

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

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

Freeman, R. (1984). *Strategic Management: A Stakeholder Approach.* Boston: Pitman.

French, K. R. (n.d.). *Developed Markets Factors and Returns*. Retrieved February 15, 2021, from Kenneth R French: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#International

Friedman, M. (1970, September 13). The Social Responsibility of Business is to Increase its Profits. *The New York Times Magazine*.

Gregory, A., & Whittaker, J. (2007). Performance and Performance Persistence of 'Ethical' Unit Trusts in the UK. *Journal of Business Finance & Accounting*, Vol. 34, 1327-1344.

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

Hill, R. P., Ainscough, T., Shank, T., & Manullang, D. (2007). Corporate Social Responsibility and Socially Responsible Investing: A Global Perspective. *Journal of Business Ethics*, Vol. 70, no. 2, 165-174.

Hong, H., & Kacperczyk, M. (2009). The price of sin: The effect of social norms on markets. *Journal of Financial Economics*, Vol. 93, 15-36.

Huber , B., & Comstock, M. (2017, July 27). *ESG Reports and Ratings: What They Are, Why They Matter.* Retrieved February 20, 2021, from Harvard Law School Forum on Corporate Governance: https://corpgov.law.harvard.edu/2017/07/27/esg-reports-and-ratings-what-they-are-why-they-matter/

Janik, B., & Maruszewska, K. (2020). Valuation of the Environmental Effects of Socially Responsible Investments in Europe. *MDPI Sustainability*.

Jarrow, R., & Rudd, A. (1983). A comparison of the APT and CAPM. *Journal of Banking and Finance*, 295-303.

Jensen, M. C. (1968). The Performance of Mutual Funds in The Period 1945-1964. *The Journal of Finance*, 389-416.

Jessop, S., & Howcroft, E. (2021, January 28). *Sustainable fund assets hit record $1.7 trln in 2020: Morningstar.* Retrieved February 15, 2021, from Reuters: https://www.reuters.com/article/us-global-funds-sustainable-idUSKBN29X2NM

Kempf, A., & Osthoff, P. (2007). The Effect of Socially Responsible Investing on Portfolio Performance. *European Financial Management*, Vol. 13, 908-922.

Kim, C. (2019). Can Socially Responsible Investments be Compatible with Financial Performance? A Meta-analysis. *Asia-Pacfic Journal of Financial Studies*, Vol. 48, 30-64.

Lee, L., Giese, G., & Nagy, Z. (2020). Combining E, S, and G Scores: An Exploration of Alternative Weighting Schemes. *The Journal of Impact and ESG Investing*, 94-103.

Lintner, J. (1965). Security Prices, Risk, and Maximal Gains from Diversification. *The Journal of Finance*, 587-615.

MacMahon, S. (2020, October). *The Challenge of Rating ESG Performance*. Retrieved March 5, 2021, from Harvard Business Review: https://hbr.org/2020/09/the-challenge-of-rating-esg-performance

Maddala, G., & Nimalendran, M. (1996). Errors-in-Variables Problems in Financial Models. In G. Maddala, & C. Rao, *Handbook of Statistics* (pp. 507-528). Elsevier Science.

Malkiel, B. G. (2003). The Efficient Market Hypothesis and Its Critics. *Journal of Economic Perspectives*, Vol. 17(1), 59-82.

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

Mossin, J. (1966). Equilibrium in a Capital Asset Market. *Econometrica*, 768-783.

MSCI. (n.d.). *ESG 101: What is ESG?* Retrieved February 20, 2021, from MSCI: https://www.msci.com/our-solutions/esg-investing/what-is-esg

Munk, C. (2019, October 22). *Financial Markets and Investments* [Lecture Notes on Investment Theory]. Retrieved December 20, 2020, from Canvas at Copenhagen Business School: https://cbscanvas.instructure.com/courses/5601

Norton, L. P. (2020, January 30). *13 ESG Investing Trends to Watch for in 2020.* Retrieved from Barron's: https://www.barrons.com/articles/13-esg-investing-trends-to-watch-for-in-2020-51580385602

Pedersen, L. H. (2015). Evaluating Trading Strategies. In L. H. Pedersen, *Efficiently Inefficient: How Smart Money Invests & Market Prices are Determined* (pp. 27-38). Princeton: Princeton University Press.

PRI. (n.d.). *About the PRI*. Retrieved January 20, 2021, from Principles for Responsible Investing: https://www.unpri.org/pri/about-the-pri

Pyles, M. K. (2020). Examining Portfolios Created by Bloomberg ESG Scores: Is Disclosure an Alpha Factor? *The Journal of Impact and ESG Investing*, 39-52.

Refinitiv. (2020). *Environmental, Social and Governance (ESG) scores from Refinitiv.*

Refinitiv. (2021, February 22). *Top 6 ESG investing trends in 2021*. Retrieved from Refinitiv: https://www.refinitiv.com/perspectives/future-of-investing-trading/top-6-esg-investing-trends-in-2021/

Refinitiv. (n.d.-a). *Eikon Trading Software*. Retrieved February 15, 2021, from Refinitiv: https://www.refinitiv.com/en/products/eikon-trading-software/download-eikon

Refinitiv. (n.d.-b). *About us*. Retrieved March 2, 2021, from Refinitiv: https://www.refinitiv.com/en/about-us

Renneboog, L., Horst, J., & Zhang, C. (2008). The price of ethics and stakeholder governance_ The performance of socially responsible mutual funds. *Journal of Corporate Fnance*, Vol. 14, 302-322.

Revelli, C., & Viviani, J. (2015). Financial Perofrmance of Socially Responsible Investing (SRI): what have we learned? A Meta-Analysis. *Business Ethics: A European Review*, Vol. 24 No. 2, 158-185.

Ross, S. A. (1976). The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory*, 13, 341-360.

Salaber, J. (2007, November 1). *The Determinants of Sin Stock Returns: Evidence on the European Market.* Retrieved from SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1071746

Schaltegger , S., & Figge, F. (2000). Environmental shareholder value: economic success with corporate environmental management. *Eco-management and auditing*, Vol 7, 29-42.

Scherpenzeel, A. (n.d.). *Searching for Consistency: Why ESG Ratings Vary*. Retrieved March 10, 2021, from The Sustainability Institute by ERM: https://www.sustainability.com/thinking/searching-for-consistency-why-esg-ratings-vary/

Scholtens, B., & Sievänen, R. (2013). Drivers of Socially Responsible Investing: A Case Study of Four Nordic Countries. *Journal of Business Ethics*, Vol. 115, No. 3, 605-616.

Sharpe, W. (1964, September). Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk. *The Journal of Finance*.

Sharpe, W. F. (1966). Mutual Fund Performance. *The Journal of Business*, 119-138.

Sharpe, W. F. (1994). The Sharpe Ratio. *The Journal of Portfolio Management*, 49-58.

Sortino, F., & Meer, R. (1991). Downside Risk. *Journal of Portfolio Management*.

Statman, M. (2000). Socially Responsible Mutual Funds. *Financial Analysts Journal*, Vol. 56, 30-39.

Statman, M., & Glushkov, D. (2009). The Wages of Social Responsibility. *Financial Analysts Journal*, Vol. 65, 33-46.

Stock, J., & Watson, M. (2015). *Introduction to Econometrics.* Boston: Pearson.

Sundheim, D., & Starr, K. (2020, January 22). *Making Stakeholder Capitalism a Reality.* Retrieved from Harvard Business Review: https://hbr.org/2020/01/making-stakeholder-capitalism-a-reality

Sustainalytics. (n.d.). *Our Solutions*. Retrieved March 3, 2021, from Sustainalytics: https://www.sustainalytics.com/our-solutions/

Tepper, T. (2020, September 16). *Milton Friedman On The Social Responsibility of Business, 50 Years Later.* Retrieved from Forbes: https://www.forbes.com/advisor/investing/milton-friedman-social-responsibility-of-business/

Treynor, J. L. (1965). How to Rate Management of Investment Funds. *Haward Business Review*, 63-75.

UN PRI. (2020, March 27). *ESG rating disagreement and stock returns*. Retrieved from UN PRI: https://www.unpri.org/pri-blog/esg-rating-disagreement-and-stock-returns/5625.article

Wagemans, F. A., van Koppen, C., & Mol, A. P. (2013). The effectiveness of socially responsible investment: a review. *Journal of Integrative Environmental Sciences*, 10:3-4, 235-252.

Whieldon, E., & Clark, R. (2021, April 6). *ESG funds beat out S&P 500 in 1st year of COVID-19; how 1 fund shot to the top.* Retrieved from S&P Global: https://www.spglobal.com/marketintelligence/en/news-insights/latest-news-headlines/esg-funds-beat-out-s-p-500-in-1st-year-of-covid-19-how-1-fund-shot-to-the-top-63224550

Yu, E., Luu, B., & Chen, C. (2020). Greenwashing in environmental,social and governance disclosures. *Research in International Business and Finance*.

List of Tables and Figures

<u>List of Tables</u>

<u>List of Figures</u>

Appendix

Appendix I: Portfolio Industry Weights

ESGC Score

Industry	A	B	C	D	E	F
Basic Materials	5%	8%	7%	7%	5%	5%
Consumer Discretionary	16%	12%	15%	13%	9%	15%
Consumer Staples	15%	10%	5%	17%	11%	7%
Energy	5%	4%	6%	7%	16%	3%
Financials	18%	16%	19%	18%	26%	37%
Health Care	11%	13%	17%	12%	13%	6%
Industrials	12%	16%	14%	11%	9%	15%
Real Estate	2%	2%	1%	1%	1%	3%
Technology	4%	3%	3%	4%	2%	3%
Telecommunications	5%	7%	5%	6%	8%	3%
Utilities	6%	10%	7%	4%	2%	3%

ESG Score

Industry	A	B	C	D	E	F
Basic Materials	7%	6%	6%	7%	5%	5%
Consumer Discretionary	12%	12%	14%	17%	14%	19%
Consumer Staples	14%	8%	10%	11%	19%	4%
Energy	13%	5%	3%	3%	3%	1%
Financials	16%	29%	19%	21%	19%	35%
Health Care	19%	10%	7%	6%	9%	5%
Industrials	7%	12%	17%	19%	22%	19%
Real Estate	1%	1%	2%	1%	2%	4%
Technology	3%	0%	4%	4%	4%	6%
Telecommunications	6%	9%	8%	4%	2%	3%
Utilities	3%	8%	12%	5%	2%	1%

E Score

Industry	A	B	C	D	E	F
Basic Materials	6%	6%	8%	7%	5%	4%
Consumer Discretionary	15%	9%	17%	8%	16%	17%
Consumer Staples	14%	12%	6%	18%	9%	3%
Energy	10%	9%	8%	5%	0%	1%
Financials	27%	16%	15%	14%	23%	42%
Health Care	10%	21%	9%	9%	12%	7%
Industrials	9%	9%	13%	23%	18%	13%
Real Estate	2%	1%	1%	0%	1%	3%
Technology	0%	2%	6%	4%	6%	6%
Telecommunications	2%	10%	12%	5%	3%	2%
Utilities	6%	5%	6%	6%	5%	1%

S Score

Industry	A	B	C	D	E	F
Basic Materials	6%	5%	7%	6%	9%	5%
Consumer Discretionary	13%	16%	14%	9%	9%	16%
Consumer Staples	13%	12%	7%	9%	11%	21%
Energy	8%	11%	11%	2%	2%	1%
Financials	12%	23%	22%	35%	25%	30%
Health Care	24%	8%	5%	5%	8%	2%
Industrials	7%	10%	15%	21%	23%	14%
Real Estate	1%	1%	1%	2%	3%	4%
Technology	5%	2%	2%	2%	2%	3%
Telecommunications	6%	8%	8%	3%	4%	3%
Utilities	5%	6%	8%	7%	4%	1%

G Score

Industry	A	B	C	D	E	F
Basic Materials	8%	5%	7%	4%	4%	5%
Consumer Discretionary	9%	10%	15%	15%	16%	27%
Consumer Staples	10%	13%	15%	12%	10%	7%
Energy	12%	9%	3%	3%	4%	4%
Financials	28%	21%	16%	20%	16%	14%
Health Care	11%	19%	13%	8%	11%	6%
Industrials	10%	8%	12%	16%	18%	23%
Real Estate	1%	0%	1%	1%	2%	4%
Technology	4%	1%	2%	4%	3%	5%
Telecommunications	5%	5%	9%	6%	9%	1%
Utilities	2%	7%	6%	11%	8%	3%

Source: Own production

Appendix II: ESGC Industry Weights Across Time

Portfolio A

Industry	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Average
Basic Materials	4.2%	5.7%	5.4%	4.0%	6.4%	5.8%	2.1%	6.5%	6.4%	3.4%	5.0%
Consumer Discretionary	13.3%	6.2%	17.8%	16.4%	19.5%	14.4%	15.3%	16.5%	16.3%	25.0%	16.1%
Consumer Staples	13.8%	21.0%	16.3%	13.9%	12.9%	20.0%	23.8%	6.9%	11.3%	8.1%	14.8%
Energy	1.5%	18.3%	1.7%	6.4%	6.5%	3.0%	0.6%	7.5%	2.9%	2.4%	5.1%
Financials	32.6%	13.5%	15.5%	17.4%	25.2%	18.3%	20.3%	18.3%	16.3%	5.0%	18.2%
Health Care	3.5%	7.5%	7.4%	11.2%	4.5%	13.8%	10.8%	13.3%	20.0%	16.5%	10.8%
Industrials	12.8%	7.4%	10.8%	10.9%	8.9%	8.6%	10.6%	17.2%	13.1%	21.6%	12.2%
Real Estate	1.8%	1.3%	1.7%	1.6%	2.3%	1.6%	2.9%	2.3%	2.1%	3.6%	2.1%
Technology	0.0%	4.1%	5.5%	4.7%	4.7%	4.3%	5.9%	2.2%	2.0%	11.0%	4.4%
Telecommunications	5.4%	8.2%	7.9%	5.7%	4.6%	5.6%	4.0%	5.5%	3.3%	0.0%	5.0%
Utilities	11.1%	6.7%	10.0%	7.9%	4.4%	4.6%	3.8%	3.8%	6.3%	3.2%	6.2%

Portfolio B

Industry	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Average
Basic Materials	3.9%	4.7%	3.5%	9.2%	7.0%	4.1%	14.0%	6.6%	12.9%	11.9%	7.8%
Consumer Discretionary	7.2%	12.5%	10.1%	13.5%	11.6%	9.4%	9.4%	24.1%	11.7%	10.3%	12.0%
Consumer Staples	10.5%	5.1%	5.3%	10.5%	10.8%	5.0%	12.5%	14.4%	15.6%	8.7%	9.8%
Energy	3.1%	0.4%	2.4%	0.3%	3.6%	8.0%	4.1%	4.9%	7.6%	4.4%	3.9%
Financials	16.7%	9.6%	24.5%	27.2%	16.9%	14.0%	19.1%	12.4%	10.4%	6.3%	15.7%
Health Care	10.0%	25.2%	23.1%	6.1%	1.5%	29.5%	9.9%	8.0%	11.7%	7.4%	13.2%
Industrials	18.1%	19.2%	16.4%	11.4%	24.3%	16.6%	16.0%	9.7%	15.8%	16.9%	16.4%
Real Estate	0.4%	1.2%	1.3%	1.0%	1.4%	1.7%	0.7%	3.1%	2.0%	2.3%	1.5%
Technology	1.7%	0.7%	0.9%	2.6%	0.0%	0.0%	1.5%	3.3%	6.0%	13.1%	3.0%
Telecommunications	13.4%	4.2%	4.6%	3.4%	6.5%	9.0%	6.5%	9.3%	1.4%	7.7%	6.6%
Utilities	15.2%	17.2%	7.9%	14.7%	16.5%	2.7%	6.3%	4.3%	4.9%	11.1%	10.1%

Portfolio D

Industry	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Average
Basic Materials	12.9%	10.2%	14.3%	6.3%	6.5%	2.7%	4.9%	11.1%	1.2%	1.8%	7.2%
Consumer Discretionary	14.4%	7.6%	11.2%	5.0%	6.2%	21.7%	28.9%	11.4%	6.2%	19.4%	13.2%
Consumer Staples	19.4%	24.6%	26.6%	11.1%	7.0%	9.9%	6.5%	25.8%	7.2%	27.2%	16.5%
Energy	9.4%	2.0%	8.6%	20.9%	1.6%	3.0%	7.6%	1.1%	11.8%	1.5%	6.7%
Financials	12.7%	19.7%	9.5%	11.0%	16.0%	30.3%	16.5%	13.6%	29.6%	25.6%	18.4%
Health Care	10.4%	13.4%	2.9%	21.3%	30.0%	3.7%	10.3%	2.6%	23.0%	6.5%	12.4%
Industrials	7.0%	10.0%	10.7%	14.7%	12.6%	7.4%	14.2%	19.4%	9.8%	9.0%	11.5%
Real Estate	0.9%	0.9%	1.0%	1.2%	1.4%	0.6%	0.8%	1.0%	0.8%	0.7%	0.9%
Technology	4.7%	0.9%	1.4%	3.3%	2.4%	5.5%	4.0%	11.0%	0.4%	2.0%	3.6%
Telecommunications	7.8%	6.3%	9.7%	2.1%	10.1%	5.7%	0.0%	0.0%	8.2%	5.6%	5.6%
Utilities	0.2%	4.4%	4.3%	3.0%	6.1%	9.4%	6.4%	2.9%	1.7%	0.6%	3.9%

Portfolio C

Industry	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Average
Basic Materials	10.7%	8.5%	5.7%	9.1%	8.9%	8.2%	5.3%	5.6%	7.6%	3.6%	7.3%
Consumer Discretionary	15.1%	27.5%	16.4%	33.7%	19.9%	14.4%	4.2%	5.7%	9.1%	6.1%	15.2%
Consumer Staples	9.0%	4.2%	2.8%	7.6%	5.1%	7.5%	1.1%	4.1%	2.4%	6.1%	5.0%
Energy	5.1%	2.6%	13.5%	1.4%	7.6%	4.8%	10.6%	14.0%	1.1%	2.0%	6.3%
Financials	12.5%	25.8%	23.5%	13.5%	19.0%	14.6%	13.0%	28.3%	15.8%	20.5%	18.6%
Health Care	19.3%	7.8%	15.1%	8.2%	20.6%	17.7%	27.3%	9.8%	15.1%	26.4%	16.7%
Industrials	10.0%	13.3%	8.9%	16.1%	10.9%	16.4%	15.1%	12.2%	18.6%	22.4%	14.4%
Real Estate	1.1%	0.9%	0.6%	0.8%	0.3%	1.5%	0.6%	0.3%	1.0%	2.0%	0.9%
Technology	1.4%	3.2%	3.1%	0.0%	4.5%	2.3%	4.5%	4.2%	2.6%	2.8%	2.9%
Telecommunications	2.6%	2.5%	5.6%	5.1%	1.5%	3.3%	11.8%	5.1%	15.9%	1.1%	5.5%
Utilities	13.1%	3.8%	4.8%	4.6%	1.7%	9.4%	6.4%	10.7%	10.9%	7.0%	7.2%

Portfolio E

Industry	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Average
Basic Materials	10.4%	8.3%	6.3%	3.4%	1.0%	0.9%	7.9%	3.6%	2.6%	3.1%	4.8%
Consumer Discretionary	2.7%	3.4%	3.5%	3.0%	3.7%	7.8%	12.3%	17.7%	22.2%	9.7%	8.6%
Consumer Staples	1.0%	4.9%	2.7%	12.6%	21.3%	20.2%	3.5%	19.5%	18.7%	7.4%	11.2%
Energy	25.4%	20.4%	20.8%	6.6%	11.5%	14.7%	23.6%	11.6%	9.7%	14.6%	15.9%
Financials	30.6%	23.9%	35.4%	36.6%	26.7%	29.0%	20.5%	14.7%	15.0%	24.3%	25.7%
Health Care	8.8%	11.4%	18.8%	13.0%	16.3%	5.0%	12.1%	19.7%	2.9%	18.2%	12.6%
Industrials	9.0%	8.1%	9.6%	8.5%	5.9%	10.1%	10.4%	7.5%	11.4%	9.5%	9.0%
Real Estate	0.3%	0.3%	0.4%	0.1%	0.6%	1.3%	1.1%	0.8%	1.0%	0.1%	0.6%
Technology	1.3%	0.4%	1.0%	0.5%	0.3%	1.5%	0.2%	1.3%	10.3%	1.3%	1.8%
Telecommunications	10.1%	16.1%	1.0%	15.8%	9.8%	8.4%	7.5%	0.4%	3.5%	6.6%	7.9%
Utilities	0.4%	2.8%	0.3%	0.0%	3.0%	1.2%	0.9%	3.3%	2.7%	5.2%	2.0%

Portfolio F

Industry	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Average
Basic Materials	5.4%	5.9%	6.0%	6.7%	3.9%	3.8%	2.4%	1.8%	2.4%	12.7%	5.1%
Consumer Discretionary	13.5%	24.8%	14.5%	16.6%	21.9%	18.0%	13.4%	8.4%	12.0%	10.8%	15.4%
Consumer Staples	3.2%	6.7%	14.9%	5.2%	1.7%	1.0%	19.3%	3.2%	11.1%	8.4%	7.5%
Energy	1.2%	2.8%	0.0%	0.0%	0.5%	1.4%	0.7%	0.2%	14.4%	10.9%	3.2%
Financials	37.5%	27.7%	32.2%	40.1%	32.5%	46.7%	43.0%	45.5%	34.5%	26.5%	36.6%
Health Care	4.7%	4.2%	4.7%	4.5%	3.1%	3.3%	0.7%	15.0%	6.7%	9.2%	5.6%
Industrials	20.0%	13.1%	18.4%	13.7%	17.4%	16.2%	13.6%	13.3%	10.4%	11.7%	14.8%
Real Estate	5.4%	4.7%	3.8%	4.4%	2.8%	1.7%	0.8%	0.6%	0.7%	1.2%	2.6%
Technology	3.9%	3.9%	2.5%	6.2%	4.8%	5.1%	2.1%	2.2%	2.3%	1.0%	3.4%
Telecommunications	5.4%	5.4%	2.4%	2.0%	2.2%	1.5%	0.7%	9.0%	1.2%	1.3%	3.1%
Utilities	0.0%	0.9%	0.7%	0.5%	9.2%	1.3%	3.1%	0.6%	4.3%	6.4%	2.7%

Source: Own production

Appendix III: Covariance Matrix Subset

Matrix	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Var10		Var428
Var1	0.00416	0.00049	0.00210	0.00172	0.00001	0.00034	0.00146	0.00047	0.00062	0.00127		0.00324
Var2	0.00049	0.00122	0.00094	0.00046	0.00069	0.00091	0.00065	0.00081	0.00087	0.00020		0.00033
Var3	0.00210	0.00094	0.00545	0.00195	0.00068	0.00110	0.00138	0.00083	0.00068	0.00205		0.00256
Var4	0.00172	0.00046	0.00195	0.00434	0.00075	0.00097	0.00106	0.00039	0.00022	0.00111		0.00240
Var5	0.00001	0.00069	0.00068	0.00075	0.00185	0.00126	0.00053	0.00042	0.00073	0.00018		0.00045
Var6	0.00034	0.00091	0.00110	0.00097	0.00126	0.00215	0.00068	0.00072	0.00088	0.00037		0.00068
Var7	0.00146	0.00065	0.00138	0.00106	0.00053	0.00068	0.00185	0.00081	0.00084	0.00068		0.00132
Var8	0.00047	0.00081	0.00083	0.00039	0.00042	0.00072	0.00081	0.00219	0.00103	0.00020		-0.00037
Var9	0.00062	0.00087	0.00068	0.00022	0.00073	0.00088	0.00084	0.00103	0.00333	0.00026		0.00051
Var10	0.00127	0.00020	0.00205	0.00111	0.00018	0.00037	0.00068	0.00020	0.00026	0.00328		0.00192
Var11	0.00218	0.00035	0.00161	0.00165	0.00052	0.00115	0.00123	0.00035	0.00070	0.00164		0.00238
Var12	0.00059	0.00067	0.00097	0.00086	0.00093	0.00114	0.00090	0.00052	0.00115	-0.00009		0.00135
Var13	0.00108	0.00063	0.00121	0.00113	0.00103	0.00123	0.00126	0.00094	0.00098	0.00051		0.00105
Var14	0.00230	0.00016	0.00090	0.00084	-0.00007	-0.00020	0.00070	0.00055	0.00055	0.00056		0.00165
Var15	0.00208	0.00003	0.00204	0.00095	-0.00027	-0.00003	0.00039	0.00011	0.00014	0.00160		0.00241
Var16	0.00144	0.00107	0.00221	0.00202	0.00084	0.00151	0.00153	0.00140	0.00080	0.00111		0.00182
Var17	0.00191	0.00053	0.00194	0.00124	0.00037	0.00082	0.00116	0.00076	0.00090	0.00210		0.00208
Var18	0.00171	0.00065	0.00196	0.00162	0.00077	0.00113	0.00124	0.00073	0.00057	0.00118		0.00171
Var19	0.00136	0.00044	0.00156	0.00108	0.00068	0.00084	0.00088	0.00063	0.00066	0.00125		0.00171
Var20	0.00182	0.00031	0.00148	0.00158	0.00043	0.00107	0.00106	0.00025	0.00033	0.00187		0.00222
Var21	0.00307	0.00047	0.00341	0.00273	0.00107	0.00148	0.00166	0.00058	0.00034	0.00218		0.00411
Var22	0.00069	0.00052	0.00121	0.00070	0.00030	0.00060	0.00077	0.00115	0.00068	0.00079		0.00056
Var23	0.00074	0.00077	0.00095	0.00057	0.00071	0.00098	0.00094	0.00130	0.00151	0.00037		0.00088
Var24	0.00328	0.00041	0.00392	0.00278	0.00077	0.00127	0.00139	-0.00035	0.00057	0.00317		0.00506
Var25	0.00095	0.00038	0.00174	0.00057	0.00038	0.00035	0.00080	0.00069	0.00062	0.00113		0.00093
Var26	0.00275	0.00036	0.00231	0.00205	0.00029	0.00040	0.00136	0.00035	0.00064	0.00124		0.00378
Var27	0.00132	0.00102	0.00158	0.00146	0.00087	0.00125	0.00100	0.00142	0.00114	0.00070		0.00067
Var28	0.00350	0.00036	0.00240	0.00148	0.00008	0.00031	0.00160	0.00076	0.00073	0.00154		0.00298
Var29	0.00214	0.00009	0.00202	0.00113	-0.00018	0.00027	0.00048	0.00043	0.00014	0.00174		0.00249
Var30	0.00278	0.00028	0.00290	0.00257	0.00080	0.00132	0.00097	-0.00059	0.00007	0.00329		0.00468
Var31	0.00122	0.00076	0.00147	0.00091	0.00054	0.00088	0.00110	0.00070	0.00084	0.00051		0.00140
Var32	0.00164	0.00047	0.00145	0.00099	0.00019	0.00061	0.00105	0.00056	0.00064	0.00059		0.00161
Var33	0.00228	0.00033	0.00226	0.00177	0.00065	0.00084	0.00098	-0.00055	0.00012	0.00218		0.00383
Var34	0.00302	0.00022	0.00327	0.00229	0.00059	0.00103	0.00162	-0.00017	0.00051	0.00285		0.00455
Var35	0.00120	0.00039	0.00107	0.00116	0.00045	0.00071	0.00073	0.00022	0.00022	0.00090		0.00158
Var36	0.00146	0.00044	0.00235	0.00112	0.00019	0.00019	0.00056	0.00023	0.00033	0.00086		0.00227
Var37	0.00075	0.00060	0.00166	0.00121	0.00065	0.00094	0.00102	0.00103	0.00109	0.00130		0.00139
Var38	0.00164	0.00071	0.00221	0.00167	0.00077	0.00072	0.00097	-0.00001	0.00039	0.00099		0.00303
Var39	0.00136	0.00078	0.00202	0.00122	0.00069	0.00102	0.00119	0.00118	0.00068	0.00109		0.00145
Var40	0.00038	0.00079	0.00084	0.00024	0.00068	0.00080	0.00077	0.00128	0.00089	0.00017		-0.00013
Var41	0.00241	0.00050	0.00267	0.00152	0.00033	0.00090	0.00138	0.00053	0.00040	0.00193		0.00285
Var42	0.00392	-0.00006	0.00338	0.00279	-0.00002	0.00064	0.00176	-0.00032	0.00043	0.00229		0.00632
Var43	0.00295	0.00039	0.00334	0.00266	0.00077	0.00127	0.00136	-0.00032	0.00073	0.00278		0.00484
Var44	0.00230	0.00065	0.00215	0.00159	0.00051	0.00086	0.00102	0.00059	0.00074	0.00143		0.00247
Var45	0.00227	-0.00066	0.00209	0.00118	-0.00071	-0.00010	0.00002	-0.00023	-0.00056	0.00151		0.00290
Var46	0.00149	0.00084	0.00150	0.00113	0.00056	0.00102	0.00129	0.00117	0.00102	0.00077		0.00161
Var47	0.00162	0.00037	0.00166	0.00166	0.00042	0.00052	0.00096	0.00060	0.00062	0.00047		0.00120
Var48	0.00167	0.00055	0.00180	0.00108	0.00033	0.00057	0.00130	0.00096	0.00086	0.00110		0.00177
Var49	0.00102	0.00056	0.00194	0.00147	0.00081	0.00111	0.00094	0.00023	0.00036	0.00152		0.00205
Var50	0.00355	0.00051	0.00470	0.00267	0.00101	0.00146	0.00136	-0.00012	0.00058	0.00343		0.00559
Var428	0.00324	0.00033	0.00256	0.00240	0.00045	0.00068	0.00132	-0.00037	0.00051	0.00192		0.00957

Source: Own production

Appendix IV: Stata Code

```
***** PREPARE DATASET *****

//Generate Date Variable

gen Date2 = date(date, "DMY")
format Date2 %td
drop date
rename Date2 Date
gen t=_n
tsset t

***** DRAWDOWN PLOTS *****

// 5Year Sub Sample
drop if _n > 60
drop if _n < 61

// Return Index
drop rf mkt smb hml rmw cma
gen t = _n
tsset t

rename a pf1
rename b pf2
rename c pf3
rename d pf4
rename e pf5
rename f pf6

forvalues i=1/6 {
    gen RI`i' = 100 * (1+pf`i') in 1
    replace RI`i' = RI`i'[_n-1] * (1+pf`i'[_n]) if _n > 1
    }

// High-Water Mark
forvalues i=1/6 {
    gen HWM`i' = RI`i' in 1
    replace HWM`i' = max(HWM`i'[_n-1], RI`i') if _n > 1
    }

// Drawdown
forvalues i=1/6 {
    gen DD`i' = (HWM`i' - RI`i')/HWM`i'
    }

// Maximum Drawdown
forvalues i=1/6 {
    gen MaxDD`i' = DD`i' in 1
    replace MaxDD`i' = max(MaxDD`i'[_n-1], DD`i') if _n > 1
    }

// Visualize Drawdown
forvalues i=1/6 {
    gen DDneg`i' = -DD`i'
    }

twoway (area DDneg1 Date) (area DDneg6 Date)
```

```
twoway (line DDneg1 Date) (line DDneg6 Date)

twoway (area RI1 Date) (line DDneg6 Date)

 ***** FACTOR MODEL ANALYSIS *****

 // 5Y Sub Sample (2016-2020)
 drop if _n < 61

 // 5Y Sub Sample (2011-2015)
 drop if _n > 60

 // Return Minus the Risk-Free Rate
 gen pf1 = a - rf
 gen pf2 = b - rf
 gen pf3 = c - rf
 gen pf4 = d - rf
 gen pf5 = e - rf
 gen pf6 = f - rf
 gen MKT = mkt - rf

 drop a b c d e f mkt

 // CAPM, FF3, FF5
 forvalues i=1/6 {
     pf`i' MKT
     estimates store m1_`i'
     reg pf`i' MKT smb hml
     estimates store m2_`i'
     reg pf`i' MKT smb hml rmw cma
     estimates store m3_`i'
     }

 estout m1*, cells(b(star fmt(3)))             ///
 legend label varlabels(_cons constant)        ///
 stats(r2_a)

 estout m2*, cells(b(star fmt(3)))             ///
 legend label varlabels(_cons constant)        ///
 stats(r2_a)

 estout m3*, cells(b(star fmt(3)))             ///
 legend label varlabels(_cons constant)        ///
 stats(r2_a)

 ***** ECONOMETRIC CONSIDERATIONS *****

 // Time Variable
 gen t = _n
 tsset t

 // Return Minus the Risk-Free Rate
 gen pf1 = a - rf
 gen pf2 = b - rf
```

```
gen pf3 = c - rf
gen pf4 = d - rf
gen pf5 = e - rf
gen pf6 = f - rf
gen MKT = mkt - rf

drop a b c d e f mkt

// Breusch-Godfrey Autocorrelation test
reg pf1 MKT
estat bgodfrey, lags(1)
reg pf1 MKT smb hml
estat bgodfrey, lags(1)
reg pf1 MKT smb hml rmw cma
estat bgodfrey, lags(1)

reg pf6 MKT
estat bgodfrey, lags(1)
reg pf6 MKT smb hml
estat bgodfrey, lags(1)
reg pf6 MKT smb hml rmw cma
estat bgodfrey, lags(1)

// Heteroscedasticity Test
reg pf1 MKT
estat hettest
reg pf1 MKT smb hml
estat hettest
reg pf1 MKT smb hml rmw cma
estat hettest

reg pf6 MKT
estat hettest
reg pf6 MKT smb hml
estat hettest
reg pf6 MKT smb hml rmw cma
estat hettest

// Newey-West Correct for Heteroscedasticity
forvalues i=1/6 {
    newey pf`i' MKT, lag(1)
    estimates store m1_`i'
    newey pf`i' MKT smb hml, lag(1)
    estimates store m2_`i'
    newey pf`i' MKT smb hml rmw cma, lag(1)
    estimates store m3_`i'
    }

estout m1*, cells(b(star fmt(3)))       ///
legend label varlabels(_cons constant)  ///
stats(r2_a)

estout m2*, cells(b(star fmt(3)))       ///
legend label varlabels(_cons constant)  ///
```

```
stats(r2_a)

estout m3*, cells(b(star fmt(3)))       ///
legend label varlabels(_cons constant)  ///
stats(r2_a)

// Multicollinearity test
reg pf1 MKT smb hml
vif
reg pf1 MKT smb hml rmw cma
vif
```

Source: Own production

Appendix V: Portfolio Characteristics

Average Market Cap (2011-2020)

Portfolio	ESG comb	ESG	E	S	G	Average
A	25,222	43,762	33,732	37,110	32,683	34,502
B	16,060	20,172	26,515	23,416	22,463	21,725
C	15,929	13,795	15,837	14,927	16,812	15,460
D	17,667	11,892	12,708	13,141	12,750	13,632
E	19,807	9,106	10,578	9,022	10,611	11,825
F	9,918	5,692	5,213	6,884	9,200	7,381

Average Change in Assets (2011-2020)

Portfolio	ESG comb	ESG	E	S	G	Average
A	5.27%	4.83%	3.98%	3.91%	4.98%	4.59%
B	6.76%	4.30%	4.74%	6.42%	4.90%	5.42%
C	6.65%	6.15%	6.39%	7.57%	6.35%	6.62%
D	7.23%	7.64%	7.06%	6.47%	8.22%	7.32%
E	6.46%	8.02%	8.06%	8.18%	7.11%	7.57%
F	9.81%	12.63%	12.58%	9.86%	9.54%	10.88%

Average Operating Profitability (2011-2020)

Portfolio	ESG comb	ESG	E	S	G	Average
A	22.26%	21.87%	18.39%	22.47%	21.35%	21.27%
B	20.55%	18.49%	20.38%	20.48%	18.67%	19.72%
C	21.78%	23.53%	23.06%	20.92%	24.78%	22.81%
D	22.30%	21.50%	23.92%	20.99%	24.89%	22.72%
E	29.14%	26.75%	22.56%	21.93%	24.70%	25.02%
F	27.24%	36.18%	40.50%	36.44%	20.70%	32.21%

Average Book-to-Market Ratio (2011-2020)

Portfolio	ESG comb	ESG	E	S	G	Average
A	0.57	0.73	0.79	0.49	0.62	0.62
B	0.47	0.65	0.66	0.55	0.46	0.55
C	0.43	0.56	0.58	0.44	0.45	0.45
D	0.44	0.56	0.51	0.48	0.43	0.48
E	0.43	0.51	0.45	0.39	0.43	0.43
F	0.47	0.55	0.52	0.45	0.43	0.47

Source: Own production

Appendix VI: Sub Sample Results Excluding Outliers

Sub Sample: 2011-2015

Score	*Model*	*Portfolio*	*Alpha*	*MKT*	*SMB*	*HML*	*RMW*	*CMA*	*Adj R^2*
ESG comb	CAPM	A	0.000	1.003***					0.923
		F	0.004*	0.851***					0.806
	FF3	A	0.002	0.909***	-0.179*	0.280***			0.947
		F	0.004	0.935***	0.461***	0.030			0.843
	FF5	A	0.001	0.924***	-0.147	0.364**	0.173	0.040	0.947
		F	0.004	0.905***	0.415**	0.059	-0.109	-0.290	0.844
ESG	CAPM	A	-0.002	0.960***					0.925
		F	0.005*	0.871***					0.796
	FF3	A	0.000	0.848***	-0.230***	0.317***			0.962
		F	0.004*	0.965***	0.514***	0.032			0.841
	FF5	A	0.000	0.844***	-0.238***	0.301**	-0.038	-0.015	0.961
		F	0.005*	0.939***	0.484***	0.144	0.015	-0.336	0.842
E	CAPM	A	-0.002	1.055***					0.891
		F	0.005*	0.875***					0.825
	FF3	A	0.001	0.913***	-0.204*	0.484***			0.954
		F	0.005*	0.944***	0.453***	0.097			0.867
	FF5	A	0.001	0.893***	-0.236**	0.488***	-0.092	-0.179	0.954
		F	0.005*	0.909***	0.413***	0.259	0.033	-0.461*	0.876
S	CAPM	A	-0.002	0.945***					0.932
		F	0.007**	0.781***					0.787
	FF3	A	0.000	0.861***	-0.207**	0.202***			0.949
		F	0.006**	0.849***	0.229	-0.111			0.794
	FF5	A	0.000	0.868***	-0.204*	0.118	-0.070	0.145	0.948
		F	0.005*	0.852***	0.239	-0.063	0.075	-0.019	0.787
G	CAPM	A	-0.003*	0.971***					0.924
		F	0.001	0.913***					0.801
	FF3	A	-0.002	0.887***	-0.117	0.293***			0.951
		F	0.000	1.019***	0.307*	-0.216*			0.821
	FF5	A	-0.001	0.863***	-0.151	0.329**	-0.069	-0.240	0.953
		F	0.000	0.957***	0.226	-0.034	-0.074	-0.720**	0.845

* $p<0.05$, ** $p<0.01$, *** $p<0.001$

Source: Own production

Sub Sample: 2016-2020

Score	*Model*	*Portfolio*	*Alpha*	*MKT*	*SMB*	*HML*	*RMW*	*CMA*	*Adj R^2*
ESG comb	CAPM	A	0.003*	0.959***					0.944
		F	-0.003	1.299***					0.895
	FF3	A	0.003**	0.988***	-0.256**	0.006			0.952
		F	-0.001	1.116***	0.282**	0.499***			0.951
	FF5	A	0.003**	0.974***	-0.255**	0.093	0.194	-0.002	0.952
		F	-0.001	1.124***	0.238*	0.402**	-0.394*	-0.180	0.955
ESG	CAPM	A	0.001	0.987***					0.925
		F	0.004	0.986***					0.896
	FF3	A	0.002	0.996***	-0.423***	0.145**			0.958
		F	0.002	0.961***	0.516***	-0.128*			0.937
	FF5	A	0.002	0.984***	-0.429***	0.214*	0.127	-0.030	0.957
		F	0.002	0.934***	0.474***	-0.008	0.085	-0.194	0.936
E	CAPM	A	0.001	1.093***					0.940
		F	0.004	1.031***					0.905
	FF3	A	0.003*	1.025***	-0.136	0.286***			0.968
		F	0.002	0.996***	0.535***	-0.103			0.943
	FF5	A	0.003**	1.018***	-0.198**	0.259**	-0.318**	-0.266*	0.974
		F	0.002	0.965***	0.471***	0.017	-0.005	-0.287	0.943
S	CAPM	A	0.002	0.913***					0.922
		F	-0.001	1.127***					0.880
	FF3	A	0.003*	0.965***	-0.434***	0.004			0.950
		F	-0.002	1.110***	0.324*	-0.074			0.889
	FF5	A	0.003*	0.946***	-0.476***	0.079	-0.014	-0.189	0.950
		F	-0.002	1.092***	0.370*	0.085	0.542*	0.186	0.898
G	CAPM	A	-0.001	1.134***					0.899
		F	0.006***	0.982***					0.925
	FF3	A	0.002	1.044***	-0.190	0.383***			0.944
		F	0.005**	0.978***	0.117	-0.035			0.925
	FF5	A	0.002	1.059***	-0.149	0.332*	0.059	0.179	0.943
		F	0.005**	0.950***	0.072	0.088	0.082	-0.205	0.924

* p<0.05, ** p<0.01, *** p<0.001

Source: Own production