Can ESG-Categories portfolios influence financial performance?
Evidence from Europe

Felipe Moreira

Dissertation written under the supervision of Professor Eva Schliephake

Dissertation submitted in partial fulfillment of requirements for the MSc in Finance, at the Universidade Católica Portuguesa, 3rd April 2023

.

Author: Felipe Simões Moreira

Topic: Can ESG-Categories portfolios influence financial performance? Evidence from Europe

Abstract

For the last two decades, Environmental, Social, and Governance (ESG) ratings have emerged as a crucial measure for investors interested in sustainable investments. By aligning their values with ESG factors, investors seek the flexibility to focus on individual ESG metrics while still leveraging potential abnormal returns. The academic literature aims to explore the relationship between financial performance and the different categories of ESG scores. There is contradicting empirical evidence on a positive or negative link. This thesis tries to contribute to the literature by exploring the particular impact of different categories. I analyze financial data and ESG scores from Refinitiv Eikon for the European market from 2005 to 2021. I build three different strategies based on equally-weighted ESG-Categories portfolios with a cut-off of 10%. Abnormal returns are determined by the well-known Carhart (1997) four-factor model. The results suggest that Long-Short ESG-Categories portfolios can outperform European benchmark Indexes with higher Sharpe Ratio and Cumulative returns. However, the study concludes that despite preliminary abnormal gains in Long-Short Human Rights and Product Responsibility categories portfolios, these findings are not robust when increasing the cut-offs nor consistent after splitting into subperiods. Hence, investors should not expect to outperform the market using ESG-Categories portfolios.

Keywords: Sustainable investing; ESG-Categories portfolios; financial performance; ESG scores; Europe

Autor: Felipe Simões Moriera

Tópico: Podem os portfólios de Categorias de ESG influenciar o desempenho financeiro? Evidências da Europa

Abstrato

Nas últimas duas décadas, os ratings relativos aos critérios ambientais, sociais e de governação (ESG) surgiram como uma medida crucial para os investidores interessados em investimentos sustentáveis. Alinhando os seus valores com os princípios de ESG, os investidores procuram flexibilidade para se focarem nas métricas individuais de ESG, enquanto ainda beneficiam de potenciais *abnormal returns*. A comunidade científica procura explorar a relação entre o desempenho financeiro e as diferentes categorias de classificações de ESG, não tendo ainda chegado a um consenso. Esta dissertação tenta contribuir para a literatura académica, explorando o impacto particular de diferentes categorias de ESG na performance financeira. Neste âmbito, construo portfólios de categorias ESG com base em três estratégias diferentes igualmente ponderadas com um corte de 10%. Para este propósito, analiso os dados financeiros e as pontuações de ESG mensais do Refinitiv Eikon para o mercado europeu de 2005 a 2021. A performance é medida de acordo com o modelo de quatro fatores de Carhart (1997). Os resultados sugerem que portfólios *Long-Short* criados com base nas categorias ESG podem ter um desempenho superior ao dos índices de referência europeus, apresentando maior *Sharpe Ratio* e retornos cumulativos. Contudo, apesar dos ganhos anormais preliminares nos portfólios *Long-Short* das categorias de *Human Rights* e *Product Responsability*, estes resultados não são robustos ao aumentar os cortes dos portfólios nem consistentes após a divisão em subperíodos. Deste modo, os investidores não devem esperar ter um desempenho superior ao do mercado utilizando portfólios gerados com base em categorias ESG.

Palavras-chave: Investimento sustentável; portfólios de categorias ESG; desempenho financeiro; critérios ESG; Europa

Acknowledgements

First of all, I am overwhelmed with gratitude as I reflect on my journey to becoming a Master's student in Finance at Católica Lisbon School of Business & Economics. None of this would have been possible without the unwavering support and motivation of my family. I am blessed to have them as role models of my future.

I also want to express my appreciation to Católica Lisbon School of Business & Economics for opening the doors here in Portugal to pursue my academic dreams. I will always cherish the memories and experience that I gained during this time, especially my colleagues who made this journey much funnier.

Moreover, I am blessed to have had the unwavering support of my girlfriend at this point in time, whose love, patience, and belief in me have been a constant source of strength along this road. I am grateful to have her in my life.

Lastly, I would like to extend a special thank you to my supervisor, Eva Schliephake, whose valuable guidance, and patience were fundamental throughout this process.

Table of Contents

List of Tables

List of Figures

List of Appendices

1. Introduction

Environmental, Social, and Governance (ESG) investing has gained momentum in the last few years as investors have come to appreciate the potential financial benefits of investing based on ESG ratings in the long-term horizon (Gibson et al. 2021). Globally, there has been a significant increase in Socially Responsible Investing (SRI) accounting for more than 30% of assets under management (GSIA, 2020), and the prevalence of ESG assets is forecasted to surpass 50 trillion dollars by 2025 (Bloomberg, 2021). This shift in investors' preferences toward sustainable investing has played a significant role in driving this trend.

The goal of this thesis is to answer the following hypothesis question: Can investors gain abnormal returns using ESG-categories portfolios? Investors, when implementing an ESG strategy, still have the principal objective to maximize returns, while alternative social, environmental, and governance goals become important as well. Focusing on the European market, my research tries to disentangle how the ESG-Categories are correlated with financial performance. By deconstructing scores, categories portfolios can support investors' ESG asset allocation strategies to outperform the market by prioritizing ratings of certain environmental, social, or governance classes.

With this intention, I define three strategies for each ESG-Category portfolio to evaluate the impact of their performance: (1) a Best-in-class ESG equal-weighted portfolio going long on the highest ESG-scoring firms and shorting the U.S. 1-month T-Bill, (2) an ESG Short equal-weighted portfolio selling the lowest ESG-scoring firms and buying the U.S. 1-month T-Bill, and (3) another equally-weighted portfolio buying the high ESG-scoring firms and selling the worst ESG-scoring ones.

Empirically, the research follows the ESG-Categories breakdown from Ehlers et al. (2022) to build portfolios, Halbritter and Dorfleitner (2015) for a similar data provider methodology, the weight allocation based on DeMiguel et al. (2009) and Carhart (1997) four-factor model to estimate abnormal returns.

To find the differential impact of the ESG criteria on financial performance of ESG-Categories portfolios, I first define the Refinitiv Eikon database for both ESG data points and firms' financial information. This includes evaluating the three pillars (E, S, G) and the ten categories used as the foundation of the overall ESG score and the ESG Combine. I consider these categorical metrics

from Refinitiv Eikon to build ESG portfolios due to the transparency and availability to attend to the fundamental objectives of investor concerns regarding ESG methodologies (Ehlers et al. 2022). As the paper focuses on Europe, the equity index chosen as the financial benchmark - Europe STOXX 600 Index – is composed of 600 constituents covering a vast spectrum of equities from a range of markets and industries. The sample starts in 2005 until 2021 and includes more than 1000 European firms over the period. I am concentrating on Europe because it has emerged as a dominant force in the global sustainable investment landscape. With a share of over 80% of ESG assets worldwide, European investors have been instrumental in driving the adoption of ESG sustainability metrics into their asset allocation (Morningstar, 2021). When compared to other regions, Europe is also leading the way in terms of measures to support ESG reporting. For instance, the Non-Financial Reporting Directive (NFRD), the EU taxonomy, and the Sustainable Finance Disclosure Regulation (SFDR) are key measures for transparent and accountable ESG metrics across industries (Eurosif, 2021).

As a robustness check, I modify the size and different periods to analyze how these changes affect the performance when hitting specific cut-off points and subperiods. The results suggest a preliminary significant financial influence using a 10 % Long-Short Human Rights and Product Responsibility categories portfolios, indicating up to 1% statistically significant abnormal return gains. However, when increasing portfolio size and evaluating subperiods, the results are not robust, and investors should not expect to obtain abnormal returns using ESG-Categories portfolios.

The following outline delineates the progression of this study. The literature review is described in Section 2, and an in-depth analysis of the data resources, especially delving into Score Data and Financial Data, are shown in Section 3. Section 4 outlines the methodology to build the screened portfolios and evaluate performance. Section 5 reports the empirical results and discussions, and robustness tests are presented in Section 6. Section 7 explains the limitations and potentialities for future research. Lastly, Section 8 provides a summary of the key conclusions of the study. References and Appendixes are present in sections 9 and 10 respectively.

2. Literature Review

This literature review supports the existing literature with an overview of studies that indicate a positive, negative or no relationship between ESG-Categories portfolios and companies' financial performance. The papers concerning this topic focus on global markets using different types of ESG portfolio metrics and data providers. Through an analysis of high-low ESG-rated companies' strategies, I aim to investigate superior outperformance compared to the market and the existence of abnormal returns.

Statman & Glushkov (2009) investigate equally-weighted high and low-rated portfolios using the KLD Research & Analytics (KLD) ESG data provider from 1992 to 2007 U.S. firms. Differently from my research data provider and period, the authors only test social characteristics composed of Community, Corporate governance, Diversity, Employee relations, Environmental, Human Rights, and Products, with a cut-off of 33%. For abnormal returns, the paper uses CAPM and the Carhart (1997) four-factor model for abnormal returns. Their results evidence significantly positive abnormal returns in some categories using a high-low strategy.

Similarly, Kempf & Osthoff (2007) used a best-in-class screening approach and a different data provider (KLD) to analyze ESG high-low rated stocks from the Standards and Poor 500 (S&P 500) and DS 400 from 1992 to 2004. The portfolios were constructed with equal weighting and initial cut-offs of 10%. As robustness tests, they use a 5%, 25%, and 50% cut-off. The Carhart (1997) four-factor model was used to evidence statistically significant portfolio performance, resulting in high abnormal returns when employing a long-short strategy. The authors further discovered high alphas (around 8.7%) when adopting the best-in-class screening, which remained significant even after accounting for transaction costs.

Using a different dataset, Eccles et al. (2014) analyzed two portfolios with equal weighting, consisting of high-rated and low-rated environmentally, socially, and governance (ESG) sustainable firms, taken from a sample of over 180 companies in the United States between 1993 to 2010. The data was obtained from sources such as ASSET4 (Refinitv Eikon), Sustainable Asset Management (SAM), and through executive interviews and personal observations. The authors applied the Carhart (1997) four-factor model and found statistically significant abnormal returns in both portfolios, with the high-rated portfolio yielding 4.8% higher returns than the low-rated portfolio.

In contrast to previous findings, Halbritter & Dorfleitner (2015) also seek to examine the connection between ESG and corporate financial performance. At the time, the authors were the first to utilize three major data providers, such as ASSET4, Bloomberg, and KLD, in the American market and the Carhart (1997) four-factor model to estimate abnormal returns. The study analyzed over six thousand US companies from 1990 to 2010 and only focused on the traditional ESG pillars and ESG scores. Partially aligned with my results, the authors indicate that there is no significant difference between high and low ESG scores, and portfolios based on these metrics should not lead to abnormal returns for investors.

Relatedly, Auer & Schuhmacher (2016) investigate high and low-scoring ESG firms in global markets from 2004 to 2012. Despite variations in ESG criteria, region, and industry, these companies do not exhibit improved risk-adjusted performance. In essence, investing in high or low-ESG-rated stocks does not show any superiority in performance. The authors applied a best-in-class ESG strategy targeting the top 5% of high and low-scoring securities and found that in America and Asia-Pacific, performance was similar to that of the market and in Europe, it may have a negative impact.

It is noteworthy that prior to 2015, the majority of studies concluded that there was a positive relationship between ESG performance and financial performance, depending predominately on the rating provider (Halbritter and Dorfleitner 2015). Comparatively, Berg et al. (2019) provide evidence for a low average correlation among the different data providers. The low correlation can be mainly explained by different methodologies, inconsistent weight allocations, and differing data availability which can compromise and mislead investors when building ESG portfolio strategies (Dimson et al. 2020).

A recent study by Whelan et al. (2021) reports that 59% of the research papers from 2015 to 2020 still demonstrated better performance compared to conventional investments, while only 14% reported negative results or not even a relationship. Investors in general, still encounter challenges to promote ESG ratings as their driver for asset allocation due to uncertainty and distrustful scores methodologies.

In conclusion, there is a shred of conflicting evidence regarding the feasibility of utilizing high-low ESG-rated portfolios to find abnormal returns. Statman & Glushkov (2009), Kempf & Osthoff (2007), and Eccles et al. (2014) found a positive relationship between financial performance and ESG Criteria portfolios and evidenced robust abnormal returns using the Carhart (1997) four-factor

model on specific categories in contrast to my results. Auer & Schuhmacher (2016) and Halbritter & Dorfleitner (2015) despite finding a significant financial influence of some ESG variables, the results were not robust and are more aligned with the majority of my findings.

Despite using a different ESG data provider or more than one, only Auer & Schuhmacher (2016) investigate this relationship with European firms, whereas the rest focus on U.S. equities. This can lead to regional and industry bias depending on the quality of data and the low correlations of scores among providers (Berg et al. 2019). Another key difference is the period, as my thesis focuses on a sample from 2005 to 2021, I can aggregate recent ESG data improvements and focus on periods of major economic and financial turmoil.

In terms of methodologies, I share the same approach as Statman & Glushkov (2009), Kempf & Osthoff (2007), Halbritter & Dorfleitner (2015), Auer & Schuhmacher (2016), when adopting a best-in-class ESG portfolio strategy and evaluating a high-low portfolio performance. However, I analyze all the categories available from Refinitiv Eikon. This method is based on recent research conducted by Ehlers et al. (2022) analysis of the ESG score criteria from Refinitiv Eikon, covering a period from 2010 to 2019, and evaluated best-in-class strategies with cut-offs points of 10%, 25%, and 33%. In contrast to this which has a primary focus on examining the changes in portfolio characteristics and corresponding changes in ESG scores, I focus on the finding empirical evidence of abnormal returns.

3. Data Resources & Research Methodology

Similar to Statman & Glushkov (2009), Kempf & Osthoff (2007), Halbritter & Dorfleitner (2015), Auer & Schuhmacher (2016), and Ehlers et al. (2022) this study uses the Best-in-Class or best-efforts methodologies to build ESG portfolios. In the Morgan Stanley (2018) survey, these strategies are defined by investors who choose leading companies or the most improved in an ESG category when allocating weights to build portfolios, while ESG Short strategies are built to bet against low-rated firms. My rationale is to structure on ESG screening of all categories from Refinitiv Eikon by sorting the equities into deciles based on the monthly highest scores of each category resulting in a Best-in-class Long-only, Short-Only, and Long-Short strategy.

I analyze which category-based strategy implies the highest accumulative returns and Sharpe Ratio over the period. Furthermore, the study evaluates the financial performance of these portfolios with

the Carhart (1997) four factor-model based on the monthly returns to find abnormal returns. I document the relationship between financial performance and ESG-Categories portfolios, limitations proprieties, and other characteristics of the portfolios.

3.1 Score Data

The data on ESG scores and financial returns is retrieved from Refinitiv Eikon. The data provider has a complex and transparent scoring methodology that resumes in aggregated measures for 10 different categories to calculate the three pillars: Environmental (ENV), Social (SOC), and Governance (GOV). Refinitiv uses individual themes and data points to drive critical conclusions to evaluate the scores among industries. The data available is provided from companies' annual reports, news, and other public sources which increases information credibility. Then, the three pillars' scores result from the multiplication of each category's scores by the respective weights according to Refinitiv Eikon analytics data.

The pillar E is composed of the Emission reduction (EMI) category score that assesses the efficacy of a firm's commitment towards reducing environmental emissions in supply chain control and other day-day activities; Innovation (INN) score relates how institutions are using environmental tools and resources in new green technologies, markets, products; and Resource Use (RES) score evaluates the firm's ability to reduce waste, non-renewable resources and be more energetic and water-use efficient in operations (Refinitiv 2022).

Pillar S is constituted of four categories of scores: Community (COM) describes how firms are committed to ethics, citizenship, public health, and environment; Human Rights (HUM) translate the behavior and commitment of companies when addressing human rights themes and rights; Product responsibility (PRO) evaluates the quality of products and services, abording safety, health, data protection, and integrity; and Workforce (WOR) quantify how firms value merit and inclusion, employees satisfaction, career development, safety, diversity, and equal opportunity (Refinitiv 2022).

Lastly, pillar G is formed by three category scores: CSR strategy aboard how institutions integrate environmental, social, and financial practices on daily basis; Management (MAN) measures the corporate governance best practices; and Shareholder (SHA) reflect on how shareholders are treated, valued, and protected regarding takeovers and other legal issues. Based on these metrics, the overall ESG score ranges between 0 and 100 and is obtained by weighting the three pillar

scores. In addition to this, Refinitiv also has an ESG controversy score which in sum is derived from 23 topics that can impact the overall score, resulting in an ESG Combine score which the paper mainly considers in the data analyses (Refinitiv 2022).

The Refinitiv score database is a reliable source of information because of the easy interpretation and wide range of companies available across global markets. Nevertheless, a common and misconducting challenge is how Refinitiv Eikon disclaims the Boolean questions to define indicators ad metrics for scores. Assigning the value zero for both not reporting a relevant metric or providing an answer which has a negative impact on the score, can lead to false interpretation and firm analyses. For instance, categories like Human Rights and CSR strategy are mostly composed of these indicators, which can impact the number of zeros in comparison with others that have different metrics to define scores. To avoid this misorientation and allocation issues, the paper is not going to consider firms that have a zero value in any category, which can be explained later as a limitation of the study.

3.2 Financial Data

As mentioned in the literature review, datasets have a significant influence on ESG strategies and the metrics for scores which can strongly influence the results (Halbritter & Dorfleitner 2015). This study uses DataStream from Refinitiv as the primary source of data for financial and ESG information from companies.

One of the metrics is to retrieve the return index (RI) for individual equities since dividends are reinvested to buy more shares of stock, and these are subsequently considered when shorting equities for instance. Differently from Auer & Schuhmacher (2016) and Ehlers et al. (2022) who investigate global markets, or Statman & Glushkov (2009), Kempf & Osthoff (2007) and Halbritter & Dorfleitner (2015) who explicitly focus on the U.S. market, this study exclusively focuses on the European market constituents from STOXX Europe 600 Index.

The historical monthly data is extracted from December 31, 2004, to December 31, 2021. The period of 17 years shows sufficient data to illustrate how the strategies behave throughout global market events such as the Financial Crisis of 2008 and the Coronavirus Crash of 2020, and specifically for Europe, the European debt crisis from 2009 to late 2010s. The choice for STOXX 600 Index is based on the prevailing belief that it is the most reliable proxy for the European market since is composed of a fixed number of 600 equities from small, mid to large capitalization

securities from the seventeen majorities developed markets across Europe: Germany, United Kingdom (UK), France, Spain, Portugal, Netherlands, Italy, Belgium, Denmark, Switzerland, Norway, Poland, Ireland, Luxembourg, Austria, Finland, and Sweden (Qontigo 2023).

In terms of country representation and industry allocation of the whole sample show the first screening of the raw data of the Index. Every passing year, an overwhelming majority of corporations (roughly 360 firms) – constituting a staggering sixty percent on average – are under the flagship of four major nations: namely, the United Kingdom (UK), Germany, France, and Switzerland, while all the other thirteen countries account for the rest (Appendix 1).

Apart from this, the Index is diversified concerning industry allocation. I use the Refinitiv Business Classification (TRBC) by Thomson Reuters to define the industry. More than 50% of my sample is composed of 3 sectors: Industrials (e.g., Industrial Goods, Industrial Commercial Services, Transportation), Financials (e.g., Banking & Investment Services, Insurance among others), and Consumer Cycles (e.g., Automobile & Auto Parts, Cycle Consumer Products, and Services, Retailers among others) and the rest by Energy, Utilities, and others shown in Appendix 2 (Refinitiv 2022).

The STOXX 600 Europe Index securities are extracted in January of each year to track the changes of the constituents yearly, enhancing the reliability of our findings. To avoid survivorship, bias the paper excludes companies that have been delisted. In the event of a merger, both entities are dissolved if the company keeps its ticker. Concerning the scores, the paper follows Ehlers et al. 2022 data coverage methodology, by collecting the fifteen different scores (ESG and ESG Combine three pillars and ten categories) for each firm from the Index.

The monthly data is chosen because the scores normally are available when firms announce their annual reports which depend on the fiscal year of the company. For example, in Europe, firms do not provide annual reports at the same period of time In England, companies have nine months before the end of their financial year to deliver, while in European Union, public firms provide annual reports within four months of the end of their fiscal year (EC 2023).

The study analyzes that most scores change between the months of March and April or in September. The total number of different firms the study evidence over the period is 1132 but with scores available this number falls to 1045, still increasing the sample size over the period. Bearing in mind that the evaluation of scores and returns is based on the securities present in the STOXX 600 Europe Index for each year.

Table 1 shows the descriptive statistics of the overall sample. Despite using a different Index than Ehlers et al. (2022) our findings share the same evidence in terms of a higher average score in the Environmental ('E') and in the Social ('S') categories in comparison to the Governance category ('G') in European companies. Another piece of evidence is that number of observations for each month is not equal when sorting for each variable because of the lack of scores in certain categories or the number of zeros. Disclosure is relatively low in categories such as Innovation (INN), Human Rights (HUM), and Product Responsibility (PRO), as they have an average 'N' of 369, 429, and 485 respectively in comparison to the rest of the categories.

Ehlers et al. (2022) investigate that this is primarily caused by how difficult it is for companies and Refinitiv Eikon to measure these kinds of categories, particularly before 2015. These scores are often founded on Boolean indicators which can cause confusion and misinterpretation based on the answers from data providers and companies' reports. Nevertheless, after 2015, the number of scoring firms was more than 530 for all categories, representing more than 88% of the available STOXX 600 Europe Index companies. Throughout time, the scores tend to increase because of more data disclosure from firms. New metrics from data providers and governmental guidance, assure a more accurate measure of qualitative and quantitative data Ehlers et al. (2022).

Table 1: Descriptive statistics of ESG ratings per ESG-Categories

The table above illustrates the mean, standard deviation, skewness, kurtosis minimum, and maximum values of the ESG scores labeled into fifteen categories. "N" is the number of total observations for "n" scores available over an average time series of t months from February 2005 to December 2021.

	Average Score	Standard Deviation	Skewness	Kurtosis	Max	Min	Observations	
COMB	54.70	18.56	-0.33	2.54	94.56	1.68	N	570
							t	203
							n	115658
ESG	58.38	19.94	-0.48	2.50	95.77	1.68	N	570
							t	203
							n	115658
ENV	59.79	24.63	-0.48	2.22	99.21	0.32	N	532
							t	203
							n	108083
EMI	67.95	25.37	-0.75	2.62	99.93	0.28	N	534
							t	203
							n	108457
INN	57.44	26.20	-0.14	1.84	99.89	0.85	N	369
							t	203
							n	74830
RES	67.86	26.51	-0.78	2.62	99.93	0.47	N	528
							t	203
							n	107216
SOC	60.47	23.21	-0.47	2.27	98.47	0.63	N	570
							t	203
							n	115654
COM	58.75	29.22	-0.32	1.84	99.85	0.27	N	567
							t	203
							n	115174
HUM	65.17	27.32	-0.67	2.25	99.53	1.39	N	429
							t	203
							n	87117
PRO	63.35	29.22	-0.32	1.84	99.85	0.27	N	484
							t	203
							n	98288
WOR	75.63	21.38	-1.17	3.90	99.91	0.39	N	572
							t	203
							n	116178
GOV	56.68	22.41	-0.30	2.17	99.04	0.56	N	570
							t	203
							n	115646
CSR	59.85	27.38	-0.47	2.09	99.92	0.30	N	510
							t	203
							n	103544
MAN	58.47	28.22	-0.33	1.94	99.93	0.16	N	562
							t	203
							n	114041
SHA	53.39	28.68	-0.13	1.82	99.93	0.11	N	570
							t	203
							n	115649

4. Methodology

4.1 ESG portfolio asset allocation

To construct the Categories ESG portfolios using the STOXX 600 Index data, I define the monthly log returns for each firm. The RI is the return index of each ticker on a trading month, "t" is the month and "i" is the security. The total trading months is 203 and used programming languages for these calculations:

$$Log\ Returns_{t,i} = Ln\left(\frac{RI_{t,i}}{RI_{t-1,i}}\right) \quad (1)$$

At the beginning of each month (t), the tickers of the STOXX 600 Index are ranked in ascending order based on the monthly scores from the corresponding year. To put it clearly, the tickers analyzed in May 2008 are the ones present on the Index on the first trading day of January 2008 and so on. I defined this assumption because of the low turnover ratio of companies that leave/enter the index during the year. The intuition behind adopting a yearly constituent list is simply the high probability the tickers stay in the index throughout the 12 months.

Based on these metrics, the securities are allocated in three different approaches to observe how increasing the size of the portfolio can affect the returns. The first one allocates the tickers into 10 deciles, the second allocates the tickers into 5 percentiles, and the last approach into 4 percentiles. All these allocations equally weigh the stocks among the percentiles (1st decile has the same weight as the 2nd, 3rd, and so on until de 10th for instance. For each month (t), the allocation of the ticker's changes in each percentile according to the score from the previous month (t-1), in other words, the portfolio composition of the Environmental category (ENV) of March 2010 is based on the environmental scores from February 2010 for example. This measure is used in all scores: the ten categories, the three pillars, and the overall ESG and ESG Combine.

The 1/N methodology for the weighting of the securities in each percentile was chosen for two main reasons: easy to implement and most importantly, empirical evidence from DeMiguel et al. (2009) shows that no asset allocation model is consistently better than the 1/N rule. Nonetheless, because this portfolio is built based on scores that regularly change in isolated months, other methodologies could be used in future research.

I test three different ESG trading strategies based on easy implementation and promising results from the related literature (Auer & Schuhmacher 2016; Ehlers et al. 2022; Halbritter & Dorfleitner 2015; Kempf & Osthoff 2007). The first one is a levered, 1-month timeframe long-only portfolio, based on the last month's score-ranked deciles and long the (10 %) highest-scoring stocks financing those trades via shorting the 1-month U.S. T-Bill. The second is a levered, 1-month timeframe short-only portfolio, based on the last month's score-ranked deciles financing those trades via shorting the (10 %) lowest-scoring stocks and long the 1-month U.S. T-Bill. Thirdly a self-financing, 1-month timeframe long/short portfolio, based on last month's score-ranked deciles. I short the (10 %) worst-scoring stocks and long the (10 %) best-scoring stocks. These three strategies are tested for each of the fifteen variables, and the first OOS trading month is February 2005 and the last is December 2021 in all of them.

I also examine the three strategies with some modifications regarding the size of the sample, increasing to 20% and then 25% with the highest (lowest) rated companies in each portfolio. This would increase from around 55 tickers per portfolio to a range of 110 to 119 tickers per percentile at 20% and a range of 137 to 148 tickers per percentile at 25% depending on the category. This measure permits the study to evaluate the assumptions made regarding the selection process, keeping equal weight in all the portfolios.

To compute the returns for the strategies, I consider the returns from the month (t) for the stocks that the strategy went long or short on based on their scores for the previous month (t-1). The strategies are short-term since they only consider the past month's score performance. Under this method, there is a monthly rebalancing of the securities in each percentile, to maintain equal weights (1/N) inside each decile.

As I hold on to the assumption of using the index constituents on the 1st of January for the entire year, new firms enter and leave when considering the following year's allocation, giving this paper a constantly updated view of the actual constituents and escape the biasedness that would occur if still accounted for the companies that left the STOXX 600 Index, the underperformers. In case the score is not available for the company in a category, or the value is zero, the study will not consider the allocation (see section 3.1 Data Scores)

After building the allocation model, I proceed to compute each percentile's monthly return and for all strategies, the first OOS trading month was February 2004 and the last was December 2021. The portfolios are equal-weighted, as DeMiguel et al. (2009) have proven their consistent

superiority over other asset allocation strategies. I compare the strategy with my STOXX 600 Europe benchmark, and also MSCI Europe and MSCI Europe AC which are composed of Developed, Emerging European markets, more than 425 constituents, and capture large, mid-cap firms, with strong financial robustness over the period (MSCI 2023).

Table 2 shows the descriptive statistics of the portfolio built using the strategies presented. The long-short strategy of Human Rights has the highest significant Sharpe Ratio of 0.58 among the categories, also beating the benchmark all the benchmarks STOXX 600, MSCI Europe, and MSCI Europe AC (0.26, 0.28, 0.25 respectively). This ESG-Category portfolio Long-Short also presents the highest cumulative returns over the period in comparison with the other Long-Short portfolios (Appendix 3 & Appendix 4) and Long-Only portfolios (Appendix 5 & Appendix 6). The Short-Only portfolios also provided lower cumulative returns than the two strategies.

In Figure 1, until 2008, the cumulative returns of my strategy were underwhelming compared to its benchmark. Nonetheless, since the Global Financial Crisis in 2008, this strategy steadily increased in outperformance compared to the benchmark, as it was also not heavily impacted by the Covid crisis in 2020. Fees and commissions were not taken into consideration.

Figure 1: ESG-Category portfolio vs Benchmarks Cumulative Returns between 2005-2021

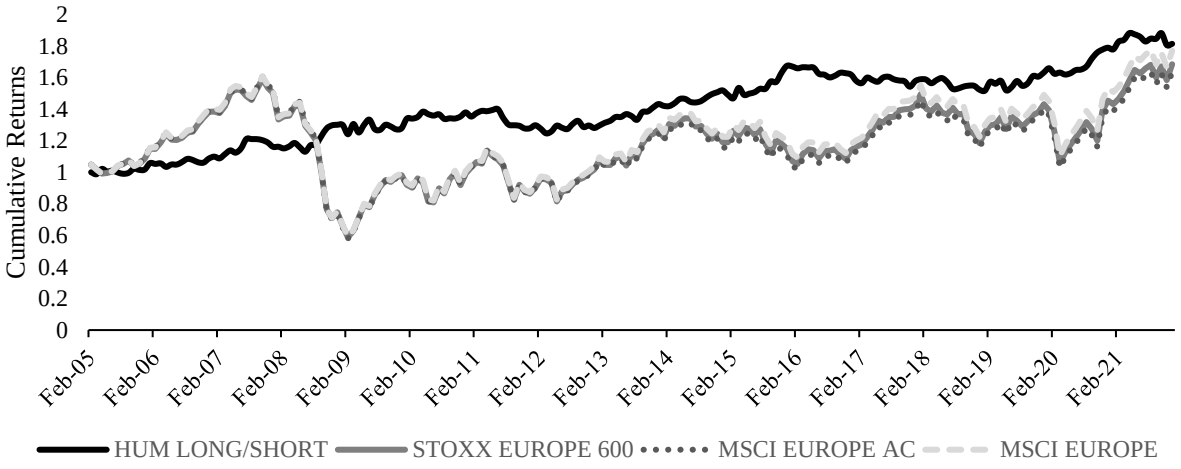


Table 2: Summary Statistics on ESG-Categories equal-weighted portfolios

The table illustrates the annualized average returns, standard deviation, and Sharpe Ratio from February 2005 to December 2021 on monthly basis. The number of months of the sample is 203. The Long-Only strategy is composed of buying the 10% highest scoring firms and shorting the 1-month U.S. T-Bill and the Short-Only strategy is shorting the 10% worst rating firms and going long on the 1-month U.S. The Long-Short portfolio entails buying the highest-rated firms on the long side and selling the worst-rated ones on the short side. All portfolios are equally-weighted using the 1/N DeMiguel et al. (2009). ***, ** and * show a significance level of 1%, 5% and 10%.

	Long Only			Short Only			Long-Short		
	Annualized Avg. Return	Annualized Standard Deviation	Sharpe Ratio	Annualized Avg. Return	Annualized Standard Deviation	Sharpe Ratio	Annualized Avg. Return	Annualized Standard Deviation	Sharpe Ratio
COMB	0.042	17.48%	0.239	-0.030	16.86%	-0.179	0.012	5.65%	0.207
ESG	0.025	19.16%	0.131	-0.013	17.83%	-0.073	0.012	5.98%	0.200
ENV	0.036	18.39%	0.194	-0.053	16.27%	-0.324	-0.017	6.04%	-0.280
EMI	0.037	18.53%	0.202	-0.019	17.08%	-0.113	0.018	5.71%	0.317
INN	0.037	18.65%	0.197	-0.039	18.61%	-0.208	-0.002	6.63%	-0.030
RES	0.030	18.81%	0.160	-0.039	16.87%	-0.234	-0.009	5.66%	-0.166
SOC	0.032	18.16%	0.178	-0.028	17.45%	-0.162	0.004	5.33%	0.074
COM	0.021	18.03%	0.119	-0.017	17.43%	-0.097	0.004	6.37%	0.069
HUM	0.044	18.38%	0.239	-0.007	18.50%	-0.036	0.037	6.42%	0.580 **
PRO	0.041	18.38%	0.225	-0.017	17.75%	-0.095	0.025	5.57%	0.441 **
WOR	0.018	18.33%	0.096	-0.032	16.90%	-0.192	-0.015	5.20%	-0.287
GOV	0.039	17.94%	0.216	-0.025	17.83%	-0.141	0.014	5.57%	0.244
CSR	0.034	18.05%	0.187	-0.031	16.83%	-0.182	0.003	4.98%	0.060
MAN	0.034	18.28%	0.186	-0.033	17.09%	-0.192	0.001	6.10%	0.019
SHA	0.024	18.23%	0.134	-0.017	17.00%	-0.101	0.007	5.49%	0.130

4.2 ESG Asset-pricing model

Further on, after de-portfolio allocation, I use the widely asset-pricing (Carhart 1997) four-factor model to estimate abnormal returns using STATA. This model by accounting for the momentum premium that explains the returns of stocks with a positive momentum relative to those with negative momentum provides a better explanation for the cross-section of expected returns than the three-factor model (Hou et al. 2015). In addition to the market risk, size, and value premiums, the four-factor model has been empirically proven to be useful in portfolio management as a complete approach to asset pricing (Bali et al. 2012), evaluating the performance of mutual funds (Carhart 1997), and constructing more efficient portfolios (Asness et al. 2013). Their study highlights the incorporation of the momentum factor results in higher portfolio returns and lower volatility compared to more value strategies. Nevertheless, when combining value and momentum strategies, they show a better portfolio performance with an improved risk-adjusted return compared to a portfolio that only uses one of these strategies (Asness et al. 2013).

$$r_{i,t} - r_{f,t} = \alpha_i + \beta_i\left(r_{m,t} - r_{f,t}\right) + s_i SMB_{i,t} + h_i HML_{i,t} + w_i WML_{i,t} + u_{i,t} \quad (2)$$

Decomposing the model, i is the portfolios and t months, $r_{i,t} - r_{f,t}$ is the excess returns of the portfolio, the three factors represented by size ($SMB_{i,t}$), value ($HML_{i,t}$) and momentum ($WML_{i,t}$), and the excess return of the market ($r_{m,t} - r_{f,t}$). Then the linear regression provides the coefficients s_i, h_i, w_i αi, βi, and $u_{i,t}$ is the residual. Differently from (Halbritter & Dorfleitner 2015), who uses the factors from Andrea Franzzini's website and methodology (Asness et al. (2013)), this study is based on Kenneth French´s data library for data and methodology from Fama & French (1992). Because of the region specification, the risk and momentum factors used are the monthly Fama French European 3 Factor and European Momentum Factor (Mom). To perform the Long-Short strategies, the 1-month U.S. Treasury bill is also extracted from the library for the risk-free rate.

5. Results

In this section, I evaluate the Category ESG portfolio performance for each of the three different strategies. I therefore analyze the regressions from the Carhart (1997) four-factor model of the high (low) portfolios and find evidence of a relationship between financial performance and ESG-Category portfolios to observe abnormal returns.

The results from the regression for each of the fifteen variables using the Long-only, Short-only, and Long-Short strategies for a 10% cut-off are shown in Table 3. For each strategy, the study is going to evaluate individually the effects of each factor and observe if there are abnormal returns. Primarily observing the alpha indicator, the highest scoring companies of ESG Combine (COMB), Environmental (ENV), Human Rights (HUM), and Product Responsibility (PRO) present relatively small positive alpha, up to zero-point one percent. High-scoring firms on Community (COM), Workforce (WOR), and Shareholders (SHA) present a relatively small negative alpha, up to zero-point one percent as well. The remaining categories present alphas equal to zero for the long-only strategy. Analyzing the short-only strategy, only ENV and Responsible Use (RES) indicate a small negative alpha of zero-point one percent and zero-point three presents, Innovation (INO) and CSR strategy (CSR) indicate a zero alpha, and the rest of the variables indicate positive alphas up to zero-point two percent. Investigating the Long-Short strategy solely, most of the

Table 3: ESG-Categories: Times-series regressions 10% cut-off

The table illustrates the outcomes of the Carhart (1997) four-factor model during the variable sample period from 2005 to 2021 on a monthly basis. Using all of the available data, the regressions are carried out separately for each ESG score and each portfolio based on all the categories. The high strategy is composed of buying the 10% highest scoring firms and shorting the 1-month U.S. T-Bill and the low strategy is shorting the 10% worst rating firms and going long on the 1-month U.S. The high-low portfolio entails buying the highest-rated firms on the long side and selling the worst-rated on the short side. All portfolios are equally-weighted using the 1/N DeMiguel et al. (2009). The table also presents alphas, factors for size, value, and momentum, and the adjusted R2. Fama French (1992) is the study foundation for the explanatory factors. ***, ** and * show a significance level of 1%, 5% and 10%

		Panel A - 10% Best (Worst) Portfolios					
		Alpha	MKT	SMB	HML	WML	Adj. R2
COMB	Long	0.001	0.702***	0.046	0.195**	-0.151**	0.724
	Short	0.001	-0.728***	-0.054	0.004	0.110*	0.698
	Long-Short	0.002	-0.027	-0.008	0.199***	-0.041	0.094
ESG	Long	-0.001	0.766***	0.078	0.185*	-0.198***	0.732
	Short	0.002	-0.744***	-0.106	-0.073	0.116*	0.691
	Long-Short	0.002	0.023	-0.028	0.112*	-0.083**	0.107
ENV	Long	0.001	0.717***	0.070	0.247**	-0.201***	0.742
	Short	-0.001	-0.679***	-0.066	-0.081	0.111*	0.699
	Long-Short	-0.001	-0.679***	-0.066	-0.081	0.111*	0.699
EMI	Long	0.000	0.732***	0.127	0.262***	-0.153**	0.730
	Short	0.001	-0.695***	-0.077	-0.138	0.102	0.683
	Long-Short	0.002	0.037	0.050	0.124**	-0.051	0.107
INN	Long	-0.000	0.746***	0.121	0.192*	-0.122*	0.688
	Short	0.000	-0.755***	-0.124	-0.092	0.170**	0.694
	Long-Short	-0.000	-0.003	-0.020	0.084	0.047	-0.010
RES	Long	-0.000	0.748***	0.161	0.191*	-0.189***	0.728
	Short	-0.003	-0.105	-0.061	0.083	0.123	0.006
	Long-Short	-0.001	0.047**	0.140**	0.134**	-0.065*	0.182
SOC	Long	0.000	0.708***	0.140	0.275***	-0.173***	0.743
	Short	0.001	-0.722***	-0.078	-0.117	0.113*	0.698
	Long-Short	0.001	-0.014	0.062	0.158***	-0.060*	0.100
COM	Long	-0.001	0.699***	0.140	0.231**	-0.119*	0.675
	Short	0.001	-0.699***	-0.135	-0.114	0.147**	0.686
	Long-Short	0.000	0.002	0.001	0.113*	0.028	-0.002
HUM	Long	0.001	0.735***	0.066	0.136	-0.205***	0.720
	Short	0.002	-0.744***	-0.052	-0.092	0.218***	0.718
	Long-Short	0.003**	-0.005	0.003	0.032	0.013	-0.019
PRO	Long	0.001	0.742***	0.112	0.145	-0.171**	0.714
	Short	0.002	-0.724***	-0.165	-0.091	0.142**	0.692
	Long-Short	0.002**	0.023	-0.065	0.043	-0.030	0.016
WOR	Long	-0.001	0.733***	0.155	0.223**	-0.167***	0.740
	Short	0.001	-0.701***	-0.104	-0.099	0.106*	0.696
	Long-Short	-0.001	0.035	0.042	0.115**	-0.061*	0.134
GOV	Long	0.000	0.744***	0.076	0.165*	-0.163***	0.757
	Short	0.001	-0.727***	-0.127	-0.095	0.144**	0.690
	Long-Short	0.001	0.017	-0.051	0.070	-0.019	0.015
CSR	Long	-0.000	0.745***	0.147	0.156	-0.142**	0.733
	Short	0.000	-0.698***	-0.049	-0.108	0.111*	0.698
	Long-Short	0.000	0.048**	0.097*	0.046	-0.031	0.077
MAN	Long	0.000	0.719***	0.073	0.231**	-0.178***	0.729
	Short	0.001	-0.709***	-0.133	-0.047	0.113*	0.675
	Long-Short	0.001	0.014	-0.071	0.173***	-0.066*	0.129
SHA	Long	-0.001	0.740***	0.137	0.164*	-0.163**	0.728
	Short	0.002	-0.709***	-0.054	-0.086	0.111*	0.696
	Long-Short	0.001	0.036	0.068	0.064	-0.053	0.069

categories indicate the same pattern of the other strategies of a very small positive alpha, up to two percent. While ENV, RES, and WOR show relatively small negative alpha and COM and INO have zero alphas.

For all strategies, the abnormal returns are mainly significantly insignificant, except for high-low scoring firms from Human Rights (HUM) and Product Responsibility (PRO) which produce positive significant alpha. These two categories can indicate a preliminary potential relationship between ESG-Categories portfolio performance and returns, although further robustness tests can provide a more reliable and valid of this finding. The rest of the insignificant alphas are in line with Halbritter & Dorfleitner (2015) who also find similar results for ESG, ENV, SOC, and GOV categories using the ASSET4 (Refinitiv), Bloomberg, and KLD data providers for U.S. firms.

Analyzing the coefficients estimated from the model, the beta for market risk premium (MKT) shows highly statistically significant for both Long-only and Short-only, in all the variables returns. The best and worst-scoring companies from every category are influenced by the market. The first strategy presents beta values around zero-point seven and all positive, presenting a lower systematic risk resulting in lower betas. The highest betas of zero-point seventy-seven are from companies that have higher-rated ESG scores, and the lowest betas are from firms with higher-rated Community scores.

Additionally, the second strategy presents all negative betas for every category also around zero-point seven, meaning an opposite direction in comparison to the overall market. The lowest betas are from companies that have lower-rated Emissions scores of zero and the highest betas are from firms that have lower-rated Innovation scores. Regarding the Long-Short strategy, all the betas' values for risk-premium are not significant except for companies with the CSR highest rated scores, the beta is close to zero (0.048).

Furthermore, all the portfolios, independent of the strategy, show that the size component does not considerably change depending on whether a company has a high or low score as betas are not significant for any level. This holds true in all the categories, except the CSR in the long-short which presents a significant relatively small positive number. The coefficient (HML) has a broader impact on the highest-scoring companies' returns. In some categories like Comb, ESG, Social, Emissions, Community, Workforce, and Management, they have the highest statically significance using the Long-only strategy. While using the long-short, most of the categories have a positive statistical significance as well. This means that these portfolios are exposed to more value stocks

than the low-scoring companies. Other categories like CSR, PRO, and HUM for instance do not present statistically significant for long strategies.

Using the short approach, even the highest or lowest-scoring companies from all variables present relatively small negative betas, showing that HML cannot indicate the relationship between value and those returns. Lastly, the coefficient (WML) when applying the long-only strategy, most of the betas are negative statistically significant, meaning that high-rated scoring companies are responsible for the worst-performing stocks of each portfolio. The short-only approach reveals that momentum factor beta is positive and statistically significant in most of the categories, presenting smaller values than the long strategy.

The long-short approach has some categories whose values are not significant but still negative like Emissions CSR, Shareholders, Governance, and Human Rights. Some exceptions like firms from Innovation, Community, and Environmental positive betas in the long-short, conforming that the lowest-rated companies are responsible for the worst returns, despite only ENV being statistically significant to five percent.

6. Robustness: Different cut-offs and sample split

In summary, the results of the three trading strategies indicated that only the 10% cut-off Long-Short Human Rights (HUM) and Product Responsibility (PRO) portfolio categories evidenced significant abnormal returns. The rest did not provide statistically significant findings. Nonetheless, it is important to check if these results are dependent on the assumptions made and improve their credibility.

In order to address the research question of this paper, the first robustness test is focused on increasing the size of the portfolio using different cut-offs. The results from the regression for each of the fifteen variables using the Long-only, Short-only, and Long-Short strategies for a 20% and 25% cut-off are shown in Table 4 and Table 5. The annualized alphas from Carhart's (1997) four-factor model are detailed in Table 6.

In comparison with the 10% cut-off, all alphas for 20% and 25% cut-offs are relatively close to zero and insignificant. There is no evidence of abnormal returns in any of the fifteen categories explored. The Carhart's (1997) four-factor model cannot provide significant alphas when increasing the size of the portfolio as shown in Table 6.

Table 4*:* ESG-Categories: Times-series regressions 20% cut-off

The table illustrates the outcomes of the Carhart (1997) four-factor model during the variable sample period from 2005 to 2021 on a monthly basis. Using all of the available data, the regressions are carried out separately for each ESG score and each portfolio based on all the categories. The Long strategy is composed of buying the 20% highest scoring firms and shorting the 1-month U.S. T-Bill and the Short strategy is shorting the 20% worst rating firms and going long on the 1-month U.S. The Long-Short portfolio entails buying the highest rated firms on the long side and selling the worst rated on the short side. All portfolios are equally-weighted using the 1/N DeMiguel et al. (2009). The table also presents alphas, factors for size, value, and momentum, and the adjusted R2. Fama French (1992) is the study foundation for the explanatory factors. ***, ** and * show a significance level of 1%, 5% and 10%.

		Panel B - 20% Best (Worst) Portfolios					
		Alpha	MKT	SMB	HML	WML	Adj. R2
COMB	Long	-0.000	0.688***	0.082	0.235**	-0.164***	0.738
	Short	-0.000	-0.691***	-0.078	-0.088	0.119**	0.712
	Long-Short	-0.000	-0.003	0.003	0.148***	-0.045	0.120
ESG	Long	-0.000	0.718***	0.047	0.223**	-0.142**	0.741
	Short	0.000	-0.706***	-0.077	-0.069	0.121*	0.703
	Long-Short	0.000	0.012	-0.031	0.155***	-0.021	0.135
ENV	Long	-0.001	0.724***	0.113	0.271***	-0.172***	0.750
	Short	-0.000	-0.700***	-0.090	-0.046	0.111*	0.705
	Long-Short	-0.001	0.024	0.023	0.225***	-0.061**	0.331
EMI	Long	-0.001	0.713***	0.160	0.255***	-0.179***	0.741
	Short	0.000	-0.702***	-0.084	-0.083	0.132**	0.704
	Long-Short	-0.001	0.011	0.076	0.172***	-0.048*	0.185
INN	Long	-0.000	0.731***	0.120	0.161	-0.157**	0.702
	Short	-0.000	-0.716***	-0.143	-0.125	0.149**	0.700
	Long-Short	-0.001	0.021	-0.038	0.020	-0.009	-0.006
RES	Long	-0.001	0.723***	0.128	0.212**	-0.176***	0.732
	Short	0.000	-0.714***	-0.086	-0.020	0.108*	0.700
	Long-Short	-0.001	0.011	0.036	0.186***	-0.069***	0.268
SOC	Long	-0.001	0.709***	0.091	0.241**	-0.164***	0.735
	Short	0.000	-0.710***	-0.062	-0.081	0.091	0.714
	Long-Short	-0.000	-0.001	0.028	0.160***	-0.072***	0.203
COM	Long	-0.000	0.681***	0.111	0.270***	-0.150**	0.733
	Short	0.001	-0.707***	-0.142	-0.054	0.111*	0.699
	Long-Short	0.000	-0.024	-0.035	0.212***	-0.039	0.167
HUM	Long	0.001	0.701***	0.048	0.240**	-0.187***	0.747
	Short	0.001	-0.755***	-0.071	-0.093	0.217***	0.725
	Long-Short	0.001	-0.052**	-0.029	0.142***	0.029	0.029
PRO	Long	0.000	0.719***	0.039	0.275***	-0.178***	0.745
	Short	0.001	-0.716***	-0.133	-0.026	0.155**	0.708
	Long-Short	0.001	0.006	-0.102**	0.240***	-0.023	0.214
WOR	Long	-0.001	0.707***	0.110	0.229**	-0.150**	0.720
	Short	0.001	-0.720***	-0.035	-0.047	0.107*	0.706
	Long-Short	-0.000	-0.009	0.064	0.171***	-0.044	0.142
GOV	Long	0.001	0.708***	0.088	0.219**	-0.164***	0.760
	Short	0.001	-0.716***	-0.060	-0.081	0.134**	0.718
	Long-Short	0.001*	-0.008	0.028	0.138***	-0.030	0.088
CSR	Long	-0.001	0.738***	0.091	0.179*	-0.193***	0.757
	Short	0.000	-0.704***	-0.079	-0.069	0.115*	0.706
	Long-Short	-0.001	0.034*	0.010	0.108**	-0.078***	0.179
MAN	Long	-0.001	0.711***	0.067	0.228**	-0.164***	0.739
	Short	0.000	-0.698***	-0.141	-0.090	0.086	0.695
	Long-Short	-0.000	0.017	-0.086*	0.127***	-0.079***	0.212
SHA	Long	-0.001	0.728***	0.147	0.203**	-0.186***	0.740
	Short	0.001	-0.718***	-0.093	-0.040	0.122*	0.706
	Long-Short	-0.000	0.015	0.039	0.149***	-0.065**	0.178

Table 5*:* ESG-Categories: Times-series regressions 25% cut-off

The table illustrates the outcomes of the Carhart (1997) four-factor model during the variable sample period from 2005 to 2021 on a monthly basis. Using all of the available data, the regressions are carried out separately for each ESG score and each portfolio based on all the categories. The Long strategy is composed of buying the 25% highest scoring firms and shorting the 1-month U.S. T-Bill and the Short strategy is shorting the 25% worst rating firms and going long on the 1-month U.S. The Long-Short portfolio entails buying the highest rated firms on the long side and selling the worst rated on the short side. All portfolios are equally-weighted using the 1/N DeMiguel et al. (2009). The table also presents alphas, factors for size, value, and momentum, and the adjusted R2. Fama French (1992) is the study foundation for the explanatory factors. ***, ** and * show a significance level of 1%, 5% and 10%.

		Panel C - 25% Best (Worst) Portfolios					
		Alpha	MKT	SMB	HML	WML	Adj. R2
COMB	Long	-0.000	0.705***	0.115	0.227**	-0.154**	0.745
	Short	0.000	-0.695***	-0.101	-0.079	0.111*	0.708
	Long-Short	-0.000	0.010	0.014	0.148***	-0.043*	0.169
ESG	Long	0.000	0.706***	0.063	0.228**	-0.147**	0.745
	Short	-0.000	-0.687***	-0.089	-0.077	0.122**	0.696
	Long-Short	0.000	0.019	-0.026	0.151***	-0.025	0.193
ENV	Long	-0.001	0.722***	0.132	0.249***	-0.171***	0.747
	Short	0.000	-0.700***	-0.087	-0.064	0.111*	0.712
	Long-Short	-0.001	0.022	0.044	0.186***	-0.060***	0.340
EMI	Long	-0.000	0.704***	0.126	0.242**	-0.163***	0.739
	Short	-0.000	-0.695***	-0.112	-0.094	0.144**	0.700
	Long-Short	-0.001	0.010	0.015	0.148***	-0.020	0.128
INN	Long	-0.000	0.737***	0.132	0.143	-0.151**	0.714
	Short	-0.001	-0.703***	-0.114	-0.098	0.158**	0.692
	Long-Short	-0.001	0.034*	0.018	0.044	0.007	0.014
RES	Long	-0.001	0.710***	0.117	0.229**	-0.166***	0.731
	Short	0.000	-0.719***	-0.109	-0.034	0.109*	0.707
	Long-Short	-0.000	-0.009	0.008	0.195***	-0.056**	0.251
SOC	Long	-0.001	0.692***	0.086	0.234**	-0.169***	0.731
	Short	0.000	-0.708***	-0.101	-0.093	0.112*	0.711
	Long-Short	-0.000	-0.016	-0.014	0.141***	-0.057**	0.133
COM	Long	-0.000	0.695***	0.100	0.251***	-0.153**	0.739
	Short	0.001	-0.702***	-0.123	-0.061	0.109*	0.701
	Long-Short	-0.000	-0.006	-0.016	0.182***	-0.050*	0.193
HUM	Long	0.000	0.691***	0.067	0.258***	-0.194***	0.743
	Short	0.000	-0.730***	-0.077	-0.079	0.204***	0.717
	Long-Short	-0.001	-0.039*	-0.003	0.172***	0.004	0.056
PRO	Long	-0.000	0.714***	0.066	0.277***	-0.165***	0.739
	Short	0.001	-0.728***	-0.142	-0.038	0.160**	0.710
	Long-Short	0.001	-0.014	-0.076	0.239***	-0.004	0.164
WOR	Long	-0.000	0.703***	0.093	0.209**	-0.176***	0.727
	Short	0.001	-0.709***	-0.072	-0.080	0.108*	0.710
	Long-Short	0.000	-0.005	0.022	0.128***	-0.067***	0.170
GOV	Long	0.000	0.711***	0.097	0.216**	-0.150**	0.752
	Short	0.000	-0.698***	-0.065	-0.086	0.133**	0.716
	Long-Short	0.001	0.013	0.032	0.131***	-0.016	0.133
CSR	Long	-0.001	0.720***	0.065	0.187**	-0.173***	0.752
	Short	0.000	-0.694***	-0.063	-0.085	0.128**	0.705
	Long-Short	0.000	0.048**	0.097*	0.046	-0.031	0.077
MAN	Long	-0.001	0.711***	0.091	0.207**	-0.159***	0.745
	Short	0.000	-0.701***	-0.131	-0.100	0.089	0.697
	Long-Short	0.001	0.014	-0.071	0.173***	-0.066*	0.129
SHA	Long	-0.001	0.709***	0.134	0.212**	-0.171***	0.736
	Short	0.000	-0.704***	-0.093	-0.073	0.137**	0.704
	Long-Short	0.001	0.036	0.068	0.064	-0.053	0.069

Table 6: ESG-Categories portfolios: Alphas within different cut-offs among strategies

The table illustrates the outcomes of the Carhart (1997) four-factor model during the variable sample period from 2005 to 2021 on a monthly basis. Using all of the available data, the regressions are carried out separately for each ESG score and each portfolio based on all the categories. The Long strategy is composed of buying the 10% highest scoring firms and shorting the 1-month U.S. T-Bill and the Short strategy is shorting the 10% worst rating firms and going long on the 1-month U.S. The Long-Short portfolio entails buying the highest rated firms on the long side and selling the worst rated on the short side. 20% and 25% portfolios cut-off were applied to the portfolio All portfolios are equally-weighted using the 1/N DeMiguel et al. (2009). The table also presents alphas, factors for size, value, and momentum, and also the adjusted R2. Fama French (1992) is the study foundation for the explanatory factors. ***, ** and * show a significance level of 1%, 5% and 10%

		Equal-Weighted Portfolios		
		10%	20%	25%
COMB	Long	0.003	0.000	0.000
	Short	0.003	0.000	0.000
	Long-Short	0.007	0.000	0.000
ESG	Long	-0.003	0.000	0.000
	Short	0.007	0.000	0.000
	Long-Short	0.007	0.000	0.000
ENV	Long	0.003	-0.003	-0.003
	Short	-0.003	0.000	0.000
	Long-Short	-0.003	-0.003	-0.003
EMI	Long	0.000	-0.003	0.000
	Short	0.003	0.000	0.000
	Long-Short	0.007	-0.003	-0.003
INN	Long	0.000	0.000	0.000
	Short	0.000	0.000	-0.003
	Long-Short	0.000	-0.003	-0.003
RES	Long	0.000	-0.003	-0.003
	Short	-0.010	0.000	0.000
	Long-Short	-0.003	-0.003	0.000
SOC	Long	0.000	-0.003	-0.003
	Short	0.003	0.000	0.000
	Long-Short	0.003	0.000	0.000
COM	Long	-0.003	0.000	0.000
	Short	0.003	0.003	0.003
	Long-Short	0.000	0.000	0.000
HUM	Long	0.003	0.003	0.000
	Short	0.007	0.003	0.000
	Long-Short	0.010 **	0.003	-0.003
PRO	Long	0.003	0.000	0.000
	Short	0.007	0.003	0.003
	Long-Short	0.007 **	0.003	0.003
WOR	Long	-0.003	-0.003	0.000
	Short	0.003	0.003	0.003
	Long-Short	-0.003	0.000	0.000
GOV	Long	0.000	0.003	0.000
	Short	0.003	0.003	0.000
	Long-Short	0.003	0.003 *	0.003
CSR	Long	0.000	-0.003	-0.003
	Short	0.000	0.000	0.000
	Long-Short	0.000	-0.003	0.000
MAN	Long	0.000	-0.003	-0.003
	Short	0.003	0.000	0.000
	Long-Short	0.003	0.000	0.003
SHA	Long	-0.003	-0.003	-0.003
	Short	0.007	0.003	0.000
	Long-Short	0.003	0.000	0.003

Consequently, the previous positive relationship between ESG-Category portfolio is not visible which can indicate unreliable OLS model outputs in the preliminary findings. The absence of robustness in the OLS estimates can be attributed to multiple factors, including sample size, measurement errors, biased sample selection, and other confounding factors. This research made some key assumptions when treating data which may possible be the cause, but further discussions are necessary to determine the lack of robustness (see section 7. Limitations and Future research). These insignificant alphas are in line with Halbritter & Dorfleitner (2015) who used more cut-offs (1%, 5%, 25%, 50%) and similar results for ESG, ENV, SOC, and GOV categories using the Refinitiv Eikon (Asset4) database. In contrast, Kempf & Osthoff (2007) evidence abnormal returns using 5%, 25%, and 50% cut-offs for some categories.

For the second test, I look at the alphas of the 10 % cut-off strategies to observe if the relationship between ESG scores and financial performance remains constant over the period. Since our first OOS trading is February 2005, my sample divides into four subperiods from February 2005 until January 2009, February 2009 to January 2013, February 2013 to January 2017, and February 2017 to December 2021.

The split covers important macro events related to the financial performance of firms in the financial crises, rebound from crises, and the covid scenario. Table 7 shows the annualized Carhart (1997) four-factor model alphas of the Long-Only strategy, Table 8 of the Short-Only, and Table 9 of the Long-Short for all categories within the subperiods and the full-sample.

Analyzing Table 7, from 2005 to 2009 the majority of the alphas are positive, relatively small, and insignificant. Only in the Human Rights category, an investor would obtain a 2.7% annual abnormal return. Between 2013 to 2017, Emissions (EMI), Product Responsibility (PRO), and Shareholders (SHA) categories are significant to 10% level, achieving a 2.7%, 2.8%, and 2.9% abnormal returns respectively. The other subperiods present insignificant alphas and negative results in most of them. In Table 8, most of the alphas are in line with full sample results, which do not present significant values.

According to Table 9, the majority of the significant alphas are to a 10% level. Between 2005 to 2009, investors could have exploited abnormal returns of 2.1% by investing in the Human Rights (HUM) category, which is the only significant value across the subperiods. Regarding Product Responsibility (PRO), only between 2013 to 2017, investors would outperform the market with a significant alpha of 1.9%. In the last subperiod, investors could explore more abnormal returns,

still below 2%, when investing in the ESG, Emissions (EMI), Community (COM), and Management (MAN) categories.

Overall, the robustness tests demonstrate that the findings did not remain consistent after the adjustments to the standard model. The preliminary financial relationship between ESG-Categories portfolio performance and financial returns did not hold and the abnormal returns are not significant to any category portfolio. These results provide compelling evidence in contrast to previous studies that found abnormal returns using high-low ESG portfolios Statman & Glushkov (2009), Kempf & Osthoff (2007), and Eccles et al. (2014). By dividing the sample into several subperiods, the alphas were not consistent and insignificant for the majority of the categories which was also evidenced by Halbritter & Dorfleitner (2015).

Table 7: Long-Only alphas within various subperiods

The table illustrates the annualized alphas from the long-short strategies of the Carhart (1997) four-factor model over the full sample period and the subperiods starting February until January of the indicating month. The only different subperiod is from February 2017 to December 2021. The strategy is composed by buying the 10% highest scoring firms and shorting the 1-month U.S. T-Bill. Using all the available data, the regressions are carried out separately for each ESG score and each portfolio based on each individual category. All portfolios are equally-weighted using the 1/N DeMiguel et al. (2009). Fama French (1992) is the study foundation for the explanatory factors. ***, ** and * show a significance level of 1%, 5% and 10%

	Equal-Weighted Long-only strategy				
	Full Sample	2005 2009	2009 2013	2013 2017	2017 2021
COMB	0.003	0.008	-0.004	0.027	-0.001
ESG	-0.003	0.009	0.004	0.009	-0.007
ENV	0.003	0.022	-0.013	0.027 *	-0.004
EMI	0.000	0.007	-0.006	0.027	-0.004
INN	0.000	0.017	0.003	0.027	-0.012
RES	0.000	0.003	-0.004	0.025	-0.002
SOC	0.000	0.009	-0.003	0.024	-0.007
COM	-0.003	-0.006	-0.009	0.022	0.001
HUM	0.003	0.027 *	-0.005	0.021	-0.003
PRO	0.003	0.014	-0.005	0.028 *	-0.004
WOR	-0.003	0.010	-0.016	0.023	-0.012
GOV	0.000	0.022	-0.005	0.013	-0.003
CSR	0.000	0.003	0.000	0.019	-0.004
MAN	0.000	0.006	-0.006	0.022	0.005
SHA	-0.007	0.013	-0.007	0.029 *	-0.006

Table 8: Short-Only alphas within various subperiods

The table illustrates the annualized alphas from the long-short strategies of the Carhart (1997) four-factor model over the full sample period and the subperiods starting February until January of the indicating month. The only different subperiod is from February 2017 to December 2021. The strategy is composed of buying the 1-month US T-Bill and shorting the 10% lowest scoring firms. Using all the available data, the regressions are carried out separately for each ESG score and each portfolio based on each individual category. All portfolios are equally-weighted using the 1/N DeMiguel et al. (2009). Fama French (1992) is the study foundation for the explanatory factors. ***, ** and * show a significance level of 1%, 5% and 10%

	Equal-Weighted Short-only strategy				
	Full Sample	2005 2009	2009 2013	2013 2017	2017 2021
COMB	0.003	-0.019	0.008	-0.024	0.011
ESG	0.007	-0.018	0.006	-0.011	0.021 *
ENV	-0.003	-0.018	0.000	-0.021	0.003
EMI	0.003	-0.015	0.003	-0.020	0.019
INN	0.000	-0.006	-0.014	-0.016	0.004
RES	-0.010	-0.019	-0.001	-0.024	0.009
SOC	0.003	-0.018	0.005	-0.022	0.013
COM	0.003	-0.008	-0.001	-0.022	0.016
HUM	0.007	-0.007	0.001	-0.008	0.009
PRO	0.007	-0.006	-0.003	-0.009	0.010
WOR	0.003	-0.022	0.005	-0.023	0.009
GOV	0.003	-0.016	-0.001	-0.025	0.016
CSR	0.000	0.004	0.000	-0.026 *	0.008
MAN	0.003	-0.010	0.001	-0.025	0.010
SHA	0.007	-0.008	0.002	-0.019	0.011

Table 9: Long-Short alphas within various subperiods

The table illustrates the annualized alphas from the long-short strategies of the Carhart (1997) four-factor model over the full sample period and the subperiods starting February until January of the indicating month. The only different subperiod is from February 2017 to December 2021. The strategy is composed of buying the 10% highest-scoring firms and shorting the 10% worst-rated firms. Using all the available data, the regressions are carried out separately for each ESG score and each portfolio based on each individual category. All portfolios are equally-weighted using the 1/N DeMiguel et al. (2009). Fama French (1992) is the study foundation for the explanatory factors. ***, ** and * show a significance level of 1%, 5% and 10

	Equal-Weighted Long-Short strategy				
	Full Sample	2005 2009	2009 2013	2013 2017	2017 2021
COMB	0.007	-0.010	0.004	0.002	0.011
ESG	0.007	-0.009	0.011	-0.002	0.014 **
ENV	-0.003	0.004	-0.013	0.005	-0.002
EMI	0.007	-0.008	-0.003	0.007	0.015 *
INN	0.000	0.011	-0.012	0.011	-0.008
RES	-0.003	-0.016 *	-0.005	0.001	0.007
SOC	0.003	-0.009	0.002	0.002	0.006
COM	0.000	-0.014	-0.010	0.000	0.017 *
HUM	0.010 **	0.021 *	-0.004	0.013	0.007
PRO	0.007 **	0.008	-0.009	0.019 **	0.006
WOR	-0.003	-0.012	-0.011 *	0.000	-0.003
GOV	0.003	0.006	-0.007	-0.012	0.013
CSR	0.000	0.007	0.000	-0.007	0.004
MAN	0.003	-0.004	-0.005	-0.003	0.015 *
SHA	0.003	0.005	-0.006	0.010	0.005

7. Limitations and Future Research

The primary drawback of this study hinges on the fact that the ESG-Categories are exclusively from Refinitiv Eikon scoring methodology which may restrict its applicability to other datasets. Despite the unique exploration of fifteen categories using the Carhart (1997) four-factor model to evaluate abnormal returns and the relationship between ESG-Categories portfolio performance and firms' returns, missing values and insufficient information of several companies could impact the accuracy of the portfolios for the entire time horizon (2005-2021). These issues were treated, but they surely had implications for the precision of the model.

Particularly, several categories' scores are based on Boolean questions, which are part of Refinitiv's methodology to compose the scoring rationale. Product Responsibility (PRO), Innovation (INN), and Human Rights (HUM) were the categories that presented a high quantity of scores equal to a "0" which strongly impacted the asset allocation exercise. This paper assumes not to consider firms with a '0' score, which may exclude companies that could have an impact on the financial returns of the portfolios as results could be biased and drive different conclusions.

Ehlers et al. (2022) faced the same problem and built portfolios that were composed of firms with "0" scoring, but the results were not significantly conclusive. In future research, it would be interesting to compare with other categories from other ESG data providers using a different asset allocation strategy and evaluate if these categories present a smaller number of zeros scores and drive to the new findings.

Another important limitation is that this paper solely investigates European companies. This choice was based on the assumption investors have more transparent and available information regarding ESG scores in Europe, especially looking into categories. Even though this research reached similar findings in some ESG-Categories with Halbritter & Dorfleitner (2015), who solely investigate U.S. companies, there are still incomparable categories that do not guarantee the same results and conclusions. This applies to other geographies under different dynamics and macroeconomic conditions. As ESG data is becoming more reliable, it could be interesting to look into other emerging regions and seek ESG anomalies.

The time horizon could also be a limitation as past studies find abnormal returns before 2005 (Statman & Glushkov 2009; Kempf & Osthoff 2007) which my paper does not cover, as well as other market reactions that were not identified. This paper only took into consideration the

regressions of Carhart's (1997) four-factor model factors and the generated alphas to explain the relationship between ESG-Category portfolios performance and returns. Other metrics could have been applied to improve the model's accuracy. For instance, Halbritter & Dorfleitner (2015) used the Fama and MacBeth (1973) regressions as a statistical method to examine the systematic risk factor beta of returns of companies. In this paper, ESG-categories portfolios were constructed exclusively based on the highest and lowest rated ESG firms from Europe. Therefore, it would be valuable to investigate the risk of the overall market or economy, rather than specific risks unique to particular firms.

Lastly, when analyzing monthly basis trading strategies, one thing to keep in mind is transaction costs. These can significantly impact the portfolio's performance given the high levels of trading activity necessary to rebalance the portfolios according to new scores. Despite the scores usually changing once a year, companies do not report their annual reports at the same, and ESG providers can delay scores availability, which can increase trading equities in specific months. It would be interesting to check if the Long-Short Human Right (HUM) portfolio would still be the best performer category and would outperform the European Indexes.

8. Conclusion

This dissertation investigates the possibility of generating abnormal returns for investors using ESG-Categories portfolios, with a focus on establishing a positive correlation between these portfolio performances and financial company returns. In order to analyze this relationship, this paper built three equal-weighted portfolios based on ESG-Categories using a Best-in-class strategy, ESG Short and Long-Short between February 2005 and December 2021. The portfolios are based on the fifteen categories from Refinitiv Eikon, with a specific focus on European firms. To the best of my knowledge, there are limited empirical studies that assess all the ESG-Categories to explore abnormal returns using Carhart's (1997) four-factor model.

After reviewing the literature, it is evident that there is no consensus among academics regarding the relationship between ESG-Categories portfolio performance and financial returns. Some studies have found a positive link, whereas others have failed to identify any relationship.

After the first screening, the results show that the Long-Short strategy based on the Human Rights (HUM) portfolio provided a higher cumulative and Sharpe Ratio compared to strong European

Indexes before taking into consideration fees and commissions. The evidence provided in this research contradicts Ehlers et al. (2022) results as the authors reported that the Community, Management, and Shareholder categories generated higher Sharpe ratios and annual returns compared to the broader World Index.

I then employ the Carhart (1997) four-factor model to generate abnormal returns using the ESG-Categories portfolios. The preliminary findings suggest a financial influence of some categories' portfolio performance and returns. Using a 10% cut-off, the Long-Short Human Right (HUM) and Product Responsibility (PRO) category portfolios provided abnormal returns of 1% and 0.7% respectively. However, when the cut-off was increased to 20% and 50%, the results were not robust and there was inconsistency observed across the four different subperiods tested. With alphas close to zero and statistically insignificant, the Carhart (1997) four-factor model could not predict excess returns.

The absence of robustness in the OLS estimates may indicate their unreliability, which can be attributed to the limitations of the model. This research assumed that firms with a score of "0" were not considered due to the high prevalence of zeros values that affect the asset allocation. Product Responsibility (PRO) and Human Rights (HUM) are among the categories composed of more "0", which may have led to a biased sample selection, as companies excluded may have a critical impact on the portfolio performance.

Overall, given their non-robustness, the results indicate that investors should not expect abnormal returns when using Long-Short, Long-Only, and Short-Only equal-weighted ESG-Categories portfolios. My findings are aligned with Halbritter & Dorfleitner (2015) and Auer & Schuhmacher (2016) without identifying abnormal returns using ESG criteria. The results go in contrast with Statman & Glushkov (2009), Kempf & Osthoff (2007), and Eccles et al. (2014) abnormal gains' indications using a Best-in-class ESG portfolio up to 2012. Furthermore, Halbritter & Dorfleitner (2015) conclude that the influence of ESG on returns decreases in later subperiods, as worst-scoring firms were outperforming high-scoring firms. My findings presented the opposite conclusion, as some categories provided abnormal returns between 2013-2017 and 2017-2021 using a Long-Short strategy.

In conclusion, the results of this study have the potential to enrich the existing ESG investing literature and have practical implications for investors and scholars interested in ESG portfolio allocation. The study concludes that despite preliminary abnormal gains in some categories, the

results were not robust, and investors should not expect to outperform the market using ESG-Categories portfolios.

9 References

Asness, C.S., & Frazzini, A. (2013). The devil in HML's details. *Journal of Portfolio Management*, 39(4), 49–68

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

Bali, Turan G., Stephen J. Brown, and Mustafa Onur Caglayan. "Systematic risk and the cross section of hedge fund returns." *Journal of Financial Economics* 106.1 (2012): 114-131.

Berg, F., Koelbel, J., and Rigobon, R. (2019). Aggregate confusion: The divergence of ESG ratings. MIT Sloan School, Working Paper No. 5822-19.

Bloomberg (2021). ESG assets rising to $50 trillion will reshape $140.5 trilllion of global AUM by 2025, finds Bloomberg Intelligence. Bloomberg, 21July. Available at https://www.bloomberg.com/company/press/esg-assets-rising-to-50- trillion-will-reshape-140-5-trillion-of-global-aum-by-2025-finds-bloomberg-intelligence/.

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

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal. "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?." *The review of Financial studies* 22.5 (2009): 1915-1953.

Dimson, Elroy, Paul Marsh, and Mike Staunton. "Divergent ESG ratings." *The Journal of Portfolio Management* 47.1 (2020): 75-87.

EC. (2023). European Commission: Company Reporting. Available at: Company reporting(europa.eu)

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

Ehlers, T., Elsenhuber, U., Jegarasasingam, K., & Jondeau, E. (2022). Deconstructing ESG Scores: How to Invest with Your own Criteria. *SSRN Electronic Journal*.

Eurosif. (2021). Responsible Investment Strategies. Available at: Responsible Investment Strategies - EUROSIF

FAMA, E. F., & FRENCH, K. R. (1992). The Cross-Section of Expected Stock Returns. *The Journal of Finance*, *47*(2), 427–465.

Fama, E.F., & MacBeth, J.D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3), 607–636.

Gibson, R., Krueger, P., Riand, N., and Schmidt, P. S. (2021). ESG rating disagreement and stock returns. *Financial Analyst Journal*, 77(4), 104–127.

GSIA, Global Sustainable Investment Alliance (2020). Global sustainable investment review. Biennial Report.

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

Hou, K., Xue, C., & Zhang, L. (2015). Digesting anomalies: An investment approach. *Review of Financial Studies*, *28*(3), 650–705.

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

Morgan Stanley. (2018). Sustainable Signals: Growth and Opportunity in Asset Management. Available at: 2415532_Sustainable_Signals_Asset_Mgmt_L.pdf (morganstanley.com)

Morningstar. (2022). EUROPEAN SUSTAINABLE INVESTMENT FUNDS STUDY 2022

MSCI. (2023). MSCI Europe Index (EUR). Available at: MSCI Europe Index

Quontigo. (2023). STOXX® INDEX METHODOLOGY GUIDE (PORTFOLIO BASED INDICES). March 23. Available at: https://www.stoxx.com/document/Indices/Common/Indexguide/stoxx_index_guide.pdf

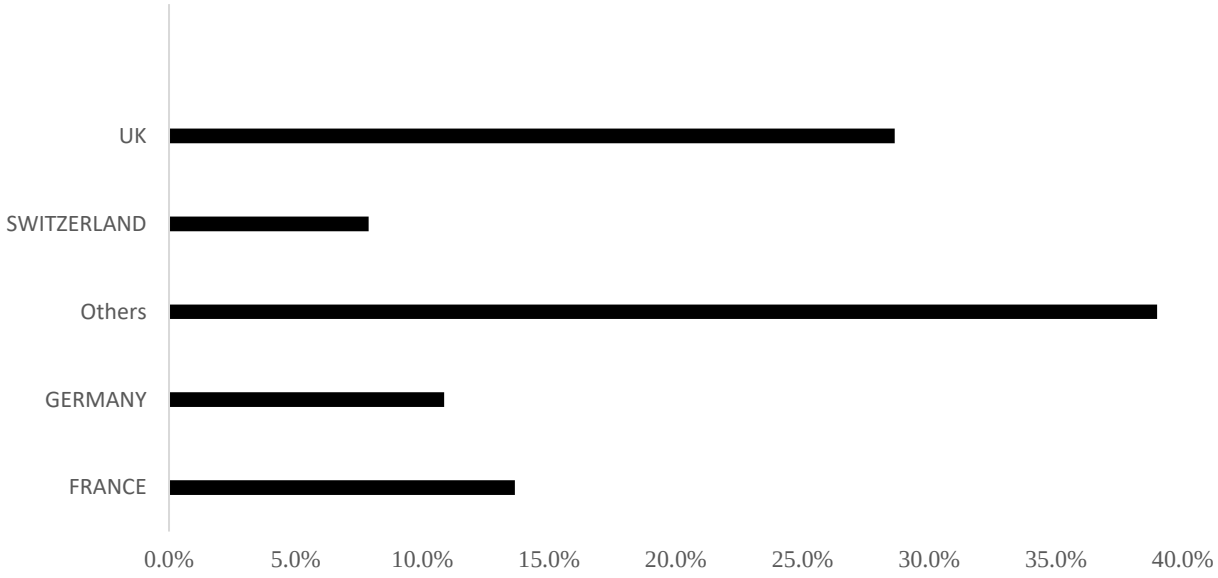
Refinitiv Eikon. (2022). *Environmental, Social and Governance (ESG) Scores from Refinitiv*.

Statman, Meir, and Denys Glushkov. "The wages of social responsibility." *Financial Analysts Journal* 65.4 (2009): 33-46.

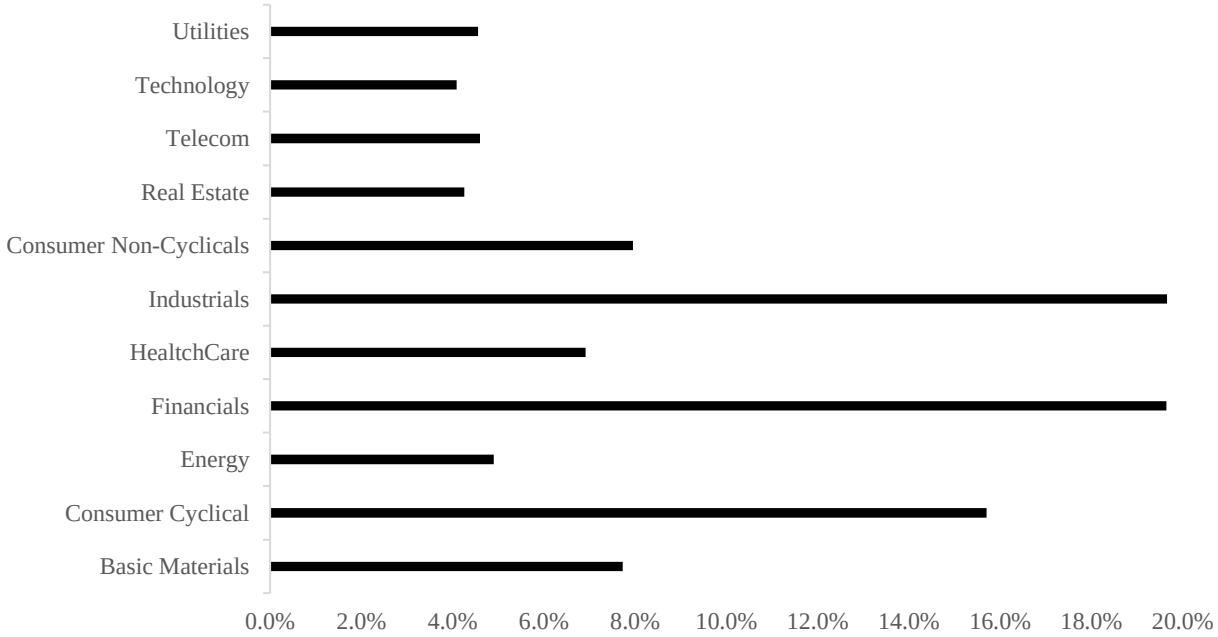
Whelan, Tensie, et al. "ESG and financial performance." Uncovering the Relationship by Aggregating Evidence from (2021): 2015-2020.

10 Appendixes

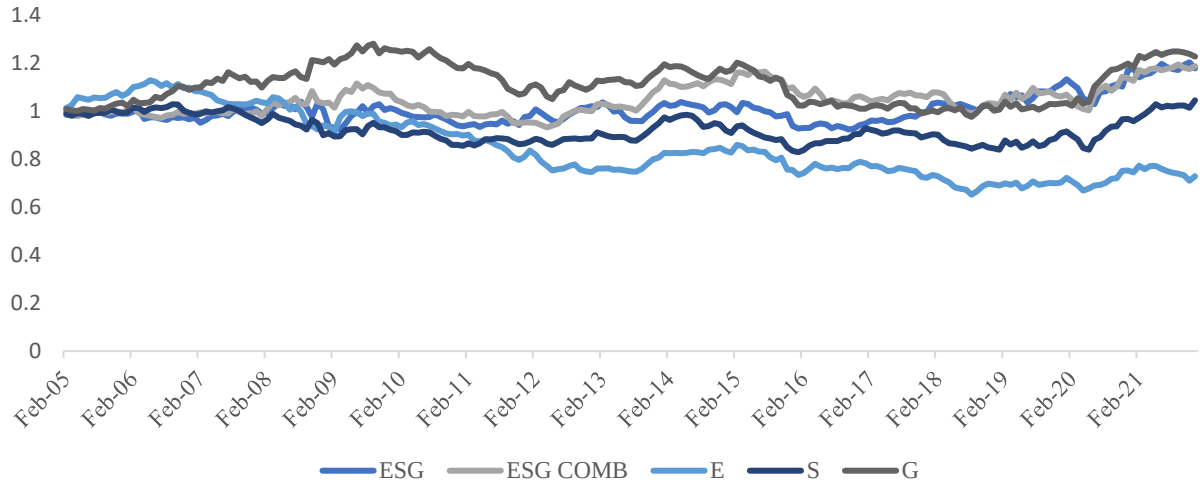
Appendix 1: Country sample representation of STOXX 600 from 2005-2021



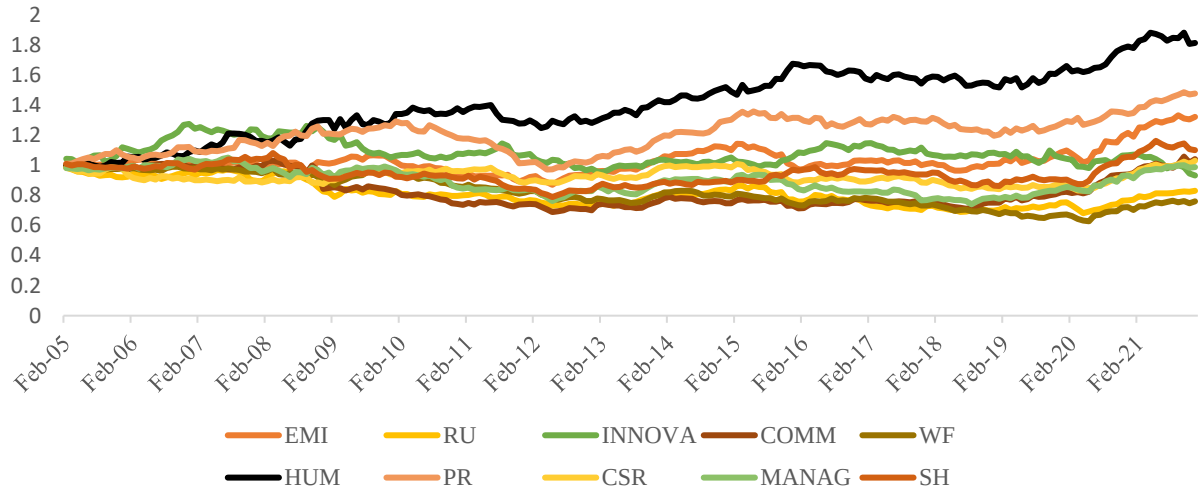
Appendix 2: Industry allocation of the sample from 2005-2021



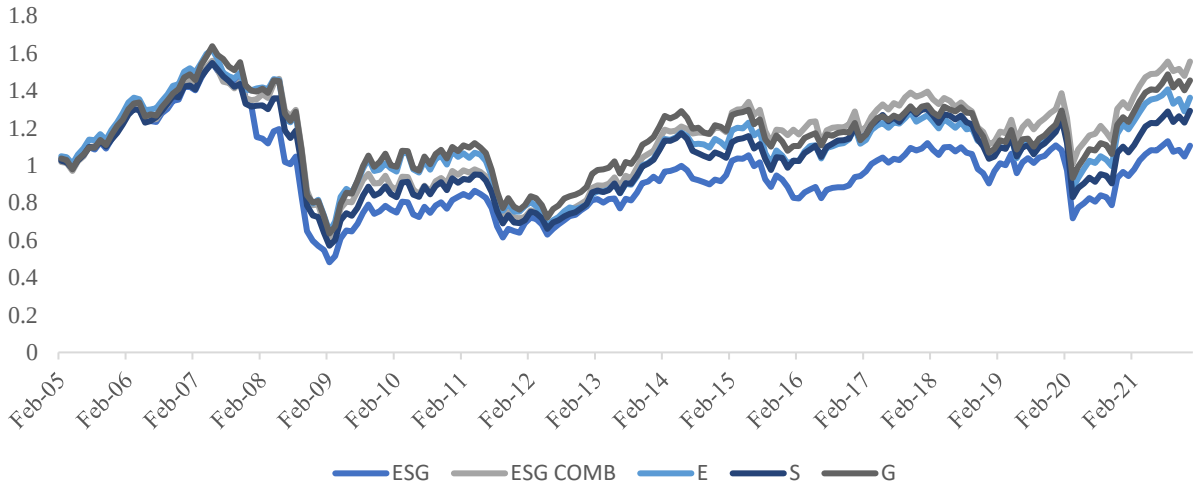
Appendix 3: ESG-Categories: Long-Short portfolios Cumulative Returns between 2005-2021



Appendix 4: ESG-Categories: Long-Short portfolios Cumulative Returns between 2005-2021



Appendix 5: ESG-Categories: Long-Only portfolios Cumulative Returns between 2005-2021



Appendix 6: ESG-Categories: Long-Only portfolios Cumulative Returns between 2005-2021